GREEN CLIMATE FUND | Independent Evaluation Unit

TRUSTED **EVIDENCE**.
INFORMED **POLICIES**.
HIGH **IMPACT**.

LEARNING-ORIENTED REAL-TIME IMPACT ASSESSMENT (LORTA) PROGRAMME

Synthesis Report – Phase I 2018

LEARNING-ORIENTED REAL-TIME IMPACT ASSESSMENT (LORTA) PROGRAMME

Synthesis Report, Phase I

APRIL 2019

GREEN CLIMATE FUND | Independent Evaluation Unit

Center for Evaluation and Development

TABLE OF CONTENTS

TABLES

ACKNOWLEDGEMENTS

The Learning-Oriented Real-Time Impact Assessment (LORTA) synthesis report (Phase I, 2018) was prepared by a large team of staff, consultants, and an external firm lead by the GCF Independent Evaluation Unit (IEU). This work would not have been possible without the full support of the GCF's Accredited Entities and Executing Agencies, the GCF Secretariat, and the GCF's National Designated Authorities and relevant government agencies at the country level. Dr. Jyotsna Puri, Head of the IEU of the GCF, and Dr. Solomon Asfaw, Principal Evaluation Officer of the IEU, provided overall managerial and technical guidance and oversight through the process. The report was jointly drafted by the Center for Evaluation and Development (C4ED), led by Prof. Markus Frölich and Dr. Susan Steiner, and and the IEU team.

We would like to acknowledge the following C4ED team members for their excellent work in contributing to the report by traveling to the GCF's funded projects and working closely with project teams on co-developing impact evaluation designs: Giulia Montresor and Tereza Varejkova (Malawi and Georgia); Nick Barton and Asmus Zoch (Mongolia); Markus Olapade and Clémentine Sardinia (Madagascar); Atika Pasha and Katharina Richert (Uganda); Katharina Richert and Sarah Vasallo (Vanuatu); Esther Heesemann and Michaela Theilmann (Paraguay); and Arne Weiss and Elisabeth Dorfmeister (Zambia). Special thanks go to Babatunde Abidoye (UNDP) who supported the teams in Malawi, Georgia, Uganda, and Zambia, and to Nathan Fiala (University of Connecticut and IEU consultant) who joined the team working on the project in Paraguay. The IEU takes responsibility for all contents of this Synthesis report.

ABBREVIATIONS

AE	Accredited entity
AEDO	Agricultural Extension Development Officer
AWS	Automated weather stations
BNCCC	Bureau National de Coordination pour le Changement Climatique
C4ED	Center for Evaluation and Development
CAZ	Ankenihevy-Zahamena Forest Corridor (Madagascar)
CBCRM	Community-based climate risk management
CBEWS	Community-based early warning schemes
CCT	Conditional cash transfer
CI	Conservation International
CI-M	Conservation International Madagascar
CIS	Climate information services
CLEW	Climate early warnings
COBA	Communautés de base
COFAV	Ambositra-Vondrozo Forest Corridor (Madagascar)
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CT	Control-in-treated
DACO	District Agriculture Coordinator's Office
DAES	Department of Agricultural Extension Services (Malawi)
DCCMS	Department of Climate Change and Meteorological Services (Malawi)
DiD	Difference-in-differences
DMO	Disaster Management Office (Vanuatu)
DoDMA	Department of Disaster Management Affairs (Malawi)
DRR	Disaster risk reduction
DWR	Department of Water Resources (Malawi)
EBD	Eco Banking Department
EBRD	European Bank for Reconstruction and Development
EE	Energy efficiency
EIB	European Investment Bank
EIEC	Environmental Information and Education Center
EPA	Extension planning area
EWS	Early warning system
FAO	Food and Agriculture Organization
FGD	Focus group discussion
FISP	Farmer Input Support Program
GCF	Green Climate Fund
GDP	Gross domestic product
GFCS	Global Framework for Climate Services
GHG	Greenhouse gas

GIS	Geographic information system
GRZ	Government of the Republic of Zambia
HFD	High-frequency data
HH	Household
ICC	Intra-cluster correlation
IE	Impact evaluation
IEU	Independent Evaluation Unit
INDI	Paraguayan Institute for the Indigenous
INFONA	National Forestry Institute (Paraguay)
IPG	Index of geographical prioritization
LORTA	Learning-Oriented Real-Time Impact Assessment
M-CLIMES	Saving Lives and Protecting Agriculture-Based Livelihoods in Malawi: Scaling Up the Use of Modernized Climate Information and Early Warning Systems
M&E	Monitoring and evaluation
MAAIF	Ministry of Agriculture, Animal Industry and Fisheries (Uganda)
MAG	Ministry of Agriculture and Livestock (Paraguay)
MDES	Minimum detectable effect size
MHEWS	Multi-hazard early warning system
MoA	Ministry of Agriculture (Zambia)
MoEPA	Ministry of Environment Protection and Agriculture (Georgia)
MoWE	Ministry of Water and Environment (Uganda)
MSME	Micro, small and medium enterprise
NASFAM	National Smallholder Farmers Association of Malawi
NDA	National designated authorities
NEA	National Environmental Agency (Georgia)
PC	Pure control
PCU	Project Coordination Unit
PICSA	Participatory Integrated Climate Services for Agriculture
PROEZA	Poverty, Reforestation, Energy and Climate Change Project
PSM	Propensity score matching
RCT	Randomized control trial
RDD	Regression discontinuity design
RE	Renewable energy
REDD+	Reducing Emissions from Deforestation and Forest Degradation+
SAS	Ministry for Social Development (Paraguay)
SEAM	Environmental Secretary (Paraguay)
SHF	Smallholder farmer
SLEM	Sustainable Landscapes for Eastern Madagascar
SPREP	Secretariat of the Pacific Regional Environment Programme
SQ	Subquestion
STP	Secretaria Técnica de Planificación del Desarrollo Económico y Social

ToC	Theory of change
UNDP	United Nations Development Programme
UNMA	Uganda National Meteorology Authority
Van-KIRAP	Vanuatu Klaemet Infomesen blong Redy, Adapt, mo Protekt
VMGD	Vanuatu Meteorology and Geohazards Department
VMME	Vice Ministry of Mines and Energy
VNSO	Vanuatu National Statistics Office
WARMA	Water Resources Management Authority
WFP	World Food Programme
ZIAMIS	Zambia Integrated Agriculture Management Information System
ZICTA	Zambia Information & Communications Technology Authority
ZMD	Zambia Meteorological Department

I. THE LORTA PROGRAMME

A. IMPORTANT ELEMENTS OF THE IEU LORTA PROGRAMME

In 2018, the Independent Evaluation Unit (IEU) of the Green Climate Fund (GCF) initiated the Learning-Oriented Real-Time Impact Assessment (LORTA) programme to develop, advise and assist GCF project teams measure their overall impact. The extent to which GCF projects lead to lower greenhouse gas emissions and increase climate resilience can be measured with the help of rigorous impact assessments. This is even more important as the empirical evidence on the impacts of climate-related projects is scarce and presents an opportunity for GCF to contribute to building a body of evidence on what works, for whom, why and under what circumstances

The IEU's LORTA programme aims to:

- Embed real-time measurement systems into funded projects so that GCF project managers can quickly access accurate data on the quality of implementation and likelihood of impact.
- Build capacity within projects to design high-quality data sets for measuring overall impact.

The IEU's LORTA programme measures returns on GCF investments, and assists GCF projects track implementation fidelity. To do so, LORTA helps GCF projects incorporate state-of-the-art approaches for measuring results and informing effectiveness and efficiency into GCF investments. It focuses on encouraging GCF project teams to employ mixed-methods approaches that involve quantitative and qualitative data-collection methods and analyses. Theory-based counterfactual impact assessments are based on experimental or quasi-experimental research designs; real-time measurement systems and qualitative databases that help project teams measure progress in implementation and provide rapid lessons on the early progress of the projects.

With LORTA, it is envisioned that GCF-funded projects will be enabled to increasingly use theory-based impact assessments. The purpose of these assessments is to measure the change in GCF key result areas that can be attributed to project activities. In sum, the objectives of LORTA include the following:

- Measuring the overall change (outcome or impact) of GCF-funded projects and enhancing learning.
- Understanding and measuring results at different parts of theories of change.
- Measuring the overall contribution of the GCF to catalysing a paradigm shift and achieving impacts at scale.

The IEU contracted the Center for Evaluation and Development (C4ED) for consultancy services to develop impact assessment designs of selected GCF projects in collaboration with the IEU and the project teams as well as to provide other relevant technical advice and quality checks. The collaboration between the IEU and C4ED has been close, including weekly calls with updates and constant exchange of ideas and feedback.

B. PHASES OF LORTA

LORTA is organized in three phases:

- Phase I – formative engagement and design: the IEU supports GCF-funded projects to build high-quality, theory-based impact evaluation designs. Formative work includes engagement with project teams, accredited entities (AEs), and GCF staff; designs for theory-based impact evaluations; and developing protocols for database development.
- Phase II – impact assessment: The second phase of LORTA involves the main impact assessment stage (3–5 years) and includes conducting survey pilots, implementing measurement and tracking

systems, collecting baseline and endline data, and continuous monitoring of real-time learning.

- Phase III – data analysis and feedback: The final stage involves analyzing data (both qualitative and quantitative), discussing results and engaging with diverse stakeholders to share results and incorporate feedback as required.

In 2018, eight GCF-funded projects were included under Phase I to undergo the formative engagement and design phase.

II. PHASE I: FORMATIVE WORK

A. SELECTION PROCESS OF PROJECTS FOR LORTA IN 2018

The LORTA Design Workshop in Bangkok, Thailand

The LORTA Design Workshop was organized by the IEU and took place from 24 to 26 July 2018 in Bangkok, Thailand. Participants were representatives from different divisions within the GCF, impact evaluation specialists from C4ED, evaluation specialists from the World Bank and other international organizations, as well as representatives of AEs, implementing partners and project staff from 15 GCF-funded projects. The workshop consisted of different elements and capacity-building measures using various formats, such as panel discussion, presentations, case studies and group work.

The achievements of the workshop were manifold:

- First, dialogue and pathways to partnerships were initiated between all groups of participants.
- Second, the project representatives were given the opportunity to work in groups and critically discuss viable impact evaluation designs for their respective projects, under the guidance of experienced and qualified impact evaluation specialists. They also increased their knowledge about impact evaluation, learned from case studies and were introduced to different impact evaluation methods (randomized and quasi-experimental designs).
- Third, a shortlist was identified of those GCF-funded projects for which impact evaluation designs should then be developed in the remaining part of Phase I of the LORTA programme.

The workshop provided participants engaged in project design and implementation with the opportunity to:

- Reflect upon the importance of including rigorous evidence in the project design process.
- Discuss case studies to learn from impact evaluation experiences in similar work areas.
- Learn about methods of impact evaluation, with a focus on randomized evaluations as well as quasi-experimental designs using mixed methods.
- Develop potential impact evaluation designs by working in groups involving evaluators and project implementers.

Decision-making process

To determine their eligibility for LORTA, the 15 participating projects were reviewed against the following strategic criteria and guiding principles determined ahead of the workshop:

- Innovativeness or importance: A GCF project is eligible for LORTA support if it is innovative but lacks adequate evidence, or if it is a “flagship” GCF project.
- Resource need and/or scalability: The extent to which the project is critical for the overall climate and development objectives of the country and the extent to which there are plans for scaling up – that is, the need for impact evaluation.
- Representativeness of portfolio: Projects are selected so that there is some balance of adaptation and mitigation activities

from both the private and public sectors. However, not every phase of LORTA will contain this representative mix. Furthermore, selected projects should represent the diversity of projects, especially including Africa, least developed countries and small island developing states, as well as including different sectors, such as climate information, food security and livelihoods, REDD+ and the private sector facility.

- Capacity needs: Projects selected depend on the capacity among project staff, including existing capacity in the implementing agency to actualize and deliver designs, and buy-in and support from project staff to help design and implement LORTA.
- Flexibility and adaptability: LORTA will be tailored to the specific project and adapted to the specific institutional context. Buy-in and deep engagement is required in this context, including the willingness to contribute project funds.
- Timing: The timeline of the project and timing of the evaluation will determine what results and outcomes should be focused on. Initially, evaluations should focus on outcomes that are quick to show change. Evaluations of long-term outcomes may span beyond the project cycle.

During the LORTA Design Workshop in Bangkok, staff members of the IEU, C4ED, the United Nations Development Programme (UNDP) and other impact evaluation experts held a meeting to discuss the evaluability and emerging impact evaluation designs of the 15 projects. Following the workshop, the IEU consulted with relevant divisions of the GCF Secretariat to build consensus regarding the most appropriate and eligible projects for the LORTA programme against the criteria above. Each division brought invaluable insight into the projects' details and the broader dynamics within the GCF. Staff members of the GCF echoed the keen interest expressed by workshop participants, and conveyed their continued support for the LORTA programme moving forward. Discussions from these consultations were synthesized to inform the final deliberation of shortlisted projects.

After this comprehensive selection process, the following eight projects were selected to be taken to the next level – that is, to be subject to formative work in preparation of impact evaluations:

i) FP002: Scaling Up the Use of Modernized Climate Information and Early Warning Systems in Malawi.

ii) FP026: Sustainable Landscapes in Eastern Madagascar.

iii) FP028: Business Loan Program for GHG Emission Reduction in Mongolia.

iv) FP034: Building Resilient Communities, Wetland Ecosystems and Associated Catchments in Uganda.

v) FP035: Climate Information Services for Resilient Development Planning in Vanuatu.

vi) FP062: Poverty, Reforestation, Energy and Climate Change Project in Paraguay.

vii) FP068: Scaling-up Multi-Hazard Early Warning System and Use of climate information in Georgia.

viii) FP072: Strengthening Climate Resilience of Agricultural Livelihoods in Zambia.

For each project, a team was formed comprising two researchers from C4ED as well as, in some cases, a member of the IEU or an external consultant. Their task was to develop an impact evaluation design for each project, and all teams went on a field mission of one week. These teams are referred to as "LORTA teams" in this report.

B. ENGAGEMENT WITH PROJECT TEAMS AND KEY STAKEHOLDERS

The task of the LORTA teams was to engage closely with key stakeholders of the selected

GCF-funded projects before, during and after the field missions. Key stakeholders included the National Designated Authorities (NDAs), AEs, implementing agencies, GCF task managers and potential end beneficiaries. The aim of this close engagement was to ensure stakeholder interest in and understanding and sense of ownership of the planned theory-based impact assessments. The cooperation of stakeholders was crucial for the ensuing steps of the LORTA programme. The engagement with stakeholders was initiated and constantly supported by the IEU.

Overall, key for the choice of an appropriate evaluation method was the design and implementation schedule of the selected GCF-funded projects. For example, outcome variables had to correspond to the project timing and mirror the time-horizon (e.g. short-term outcomes can be measured quickly after implementation of a project, whereas long-term outcomes can only be measured a certain time after project finalization). Again, the importance of buy-in and ownership on the side of the implementation partners was taken into account, as was the need to respectfully strive for a balance between strong evaluation designs and requirements for implementation.

C. FIELD MISSIONS

During Phase I of the LORTA programme, the LORTA teams worked out impact evaluation designs for each of the selected GCF-funded projects. They conducted context analyses, examined the existence of appropriate counterfactuals, assessed administrative and secondary data sources and discussed the theory of change. Some of this work was conducted during the field missions (i.e. while the LORTA teams were in the field), although most of it was done remotely, during either the preparation or debriefing phases.

The field mission schedule is presented in Table 1. A timeline of all Phase I LORTA activities is presented in annex I.

Table 1 Field mission schedule (2018)

COUNTRY	LORTA TEAM	TIME PERIOD
Malawi	Giulia Montresor & Tereza Varejkova & Timothy Cha	09/09–16/09/2018
Mongolia	Nicholas Barton & Asmus Zoch	09/09–16/09/2018
Uganda	Atika Pasha & Katharina Richert	07/10–14/10/2018
Paraguay	Esther Heesemann & Michaela Theilmann & Nathan Fiala	18/10–25/10/2018
Madagascar	Markus Olapade & Clémentine Sadania & Jyotsna Puri	21/10–28/10/2018
Vanuatu	Katharina Richert & Sarah Vassallo	04/11–11/11/2018
Zambia	Arne Weiss & Elisabeth Dorfmeister	04/11–11/11/2018
Georgia	Giulia Montresor & Tereza Varejkova & Solomon Asfaw	11/11–18/11/2018

D. SUMMARIES OF EVALUATION QUESTIONS, DESIGNS AND TIMELINES

Malawi

The project "Saving Lives and Protecting Agriculture-Based Livelihoods in Malawi: Scaling Up the Use of Modernized Climate Information and Early Warning Systems" (M-CLIMES) aims at increasing the resilience of rural livelihoods to climate variability. The project plans to install new automated weather stations, build the capacity of farmers and communities to make informed decisions based on climate and weather information, and deliver more accurate customized climate information to vulnerable food-insecure, flood-prone and fishing communities. The project has three components: 1) Expansion of networks that generate climate-related data to save lives and safeguard livelihoods from extreme climate events, 2) development and dissemination of products and platforms for climate-related information/services for vulnerable communities, and 3) strengthening of community capacities to use early warning systems for disaster preparedness.

The AE for this project is UNDP. The project is implemented by the governmental Department of Disaster Management Affairs in collaboration with the Department of Climate Change and Meteorological Services, the Department of Water Resources, the Department of Agricultural Extension Services, the Department of Fisheries, and the National Smallholder Farmers Association of Malawi.

During discussions on the theory of change and implementation plans of the different components of the M-CLIMES project, it became clear that the three components are dissimilar in terms of outcomes and target population. Multiple impact evaluations would thus be necessary; however, the available budget is sufficient for only one. Further discussions focused on component 2, especially the Participatory Integrated Climate Services for Agriculture (PICSA) training. PICSA was the only activity that had been clearly defined and for which a roll-out plan had been determined. PICSA aims at enabling farmers to make informed decisions based on accurate and location-specific climate and weather information. Lead farmers receive PICSA training delivered by extension officers and are expected to replicate the same training for a further group of farmers, known as "contact farmers".

The main evaluation question is whether the PICSA training helps farmers to adapt to climate change, thereby leading to more resilient livelihoods. Sub-questions are as follows:

- Do farmers access tailor-made climate information after the PICSA training?
- Do farmers base their farming plans on seasonal forecasts?
- Are farmers adapting their plans based on changes in forecasts?
- Are farmers indeed better equipped to face climate risks and extreme weather events?
- Does better planning based on accurate and tailor-made climate information lead to better outcomes, in particular to higher and less variable agricultural yields, higher revenues and better food security?

A key limitation for the impact evaluation design was that the selection of lead farmers to participate in the PICSA training in the first year of implementation had already been determined before the LORTA field mission. This fact precluded the use of an experimental approach, which would have provided the most rigorous estimation of the PICSA training's effects. Therefore, the impact evaluation will use a quasi-experimental design (difference-in-differences (DiD) in combination with matching) to estimate the impact of PICSA on beneficiaries by comparing participants with non-participants.

The M-CLIMES project is at a relatively more advanced stage than others in the LORTA project pool. At the time of the M-

CLIMES selection into LORTA, UNDP had already hired a consultancy firm to conduct baseline data collection and scheduled the roll-out date for PICSA. The timeline depicts an endline data collection in 2020, before PICSA will be rolled out in the control districts.

Mongolia

The project "Business Loan Programme for GHG Reduction" in Mongolia, referred to as the "Eco-loan programme", is implemented by XacBank, which is also the AE for this project. The project offers concessionary loans to micro, small and medium enterprises (MSMEs) for investments in capital or processes that reduce CO_2 emissions or energy use by 20 per cent. Loans are granted to applicants running MSMEs who promise to fulfil the reduction criterion and pass the bank's risk assessment. In addition to offering loans to eligible MSMEs, XacBank also runs an awareness campaign, which conveys information about energy efficiency (EE) and renewable energy (RE) as well as information about the Eco-loan programme itself.

The main evaluation question is whether the Eco-loan programme and the related awareness campaign lead to reduced greenhouse gas (GHG) emissions. Sub-questions are as follows:

- Do borrowers change investment plans to access Eco-loan programme funds?
- Does the Eco-loan programme lead to cost reductions and/or increased revenue for borrowers?
- What is the number of loans taken out due to the awareness campaign?
- How does the awareness campaign impact knowledge of and attitudes towards EE & RE?
- Do MSMEs plan to invest in EE & RE in the future?

Serious challenges were noted for conducting an impact evaluation for this particular project. Prime among these was the issue of small sample size. Due to the small number of MSMEs taking out loans (84 loans disbursed among the 110 applicants at time of writing), it seemed unlikely that an impact evaluation of the Eco-loan programme would have enough power to estimate causal effects. The fact that the Eco-loan programme is fully established and has the capacity to reach all eligible applicants further increases the challenge in finding a suitable evaluation framework. Three impact evaluation design options were proposed: 1) A randomized encouragement design based around the awareness-raising pathway, 2) a before–after design, which would be a learning exercise considering the possible effects of the awareness-raising (not an impact evaluation in the true sense), and 3) a propensity score matching (PSM) design to estimate the impact of the Eco-loan programme using a control group from the universe of MSMEs in Mongolia.

However, after several discussions, XacBank decided not to proceed with the LORTA programme as the project managers did not feel the designs would be effective and /or feasible given the structure of the project.

Uganda

The project "Building Resilient Communities, Wetland Ecosystems and Associated Catchments in Uganda" focuses on the adaptation to climate change via an increased resilience of the local ecosystems and communities. The project is a comprehensive undertaking that focuses on improving human living conditions of communities living around wetland areas, while restoring biodiversity in functioning wetland systems and catchment areas. This is to be achieved through the interplay of three project components. The first component addresses physical wetland restoration, the second addresses environmentally compatible alternative agricultural and non-agricultural livelihood trainings, and the third addresses support in climate change adaptation through improved climate information systems.

The AE for this project is UNDP. The project is implemented by several ministries, in particular, the Ministry of Water and Environment; the Ministry of Agriculture, Animal Industry and Fisheries; and the Uganda National Meteorology Authority.

The main evaluation question focuses on the first two components of the project. It refers to whether community members become more resilient against climate change disasters (floods, droughts and mudslides) due to wetland restoration and alternative livelihood trainings. Sub-questions are as follows:

- Are the ecosystems within wetlands restored to their original capacity to recharge groundwater, domicile typical flora and fauna, retain water, etc.?
- Are income levels of community members increased due to alternative agricultural and non-agricultural livelihood trainings?
- Is income volatility reduced due to alternative agricultural and non-agricultural livelihood trainings?

Several impact evaluation design options were proposed (e.g. phased-in randomized control trial (RCT) or matching). In order to finalize an impact evaluation strategy, however, the completion of a mapping of the entire 64,000 ha of wetlands is required. This mapping exercise determines the vulnerability of each wetland system and catchment area and identifies the exact nodes within the system where the restoration efforts should be located. Without this activity, which is still ongoing, the refinement of the final impact evaluation design cannot move forward.

A baseline data collection is scheduled for 2019, while the midline could be at the end of 2022/beginning of 2023 and the endline in 2025.

Paraguay

Deforestation rates in Paraguay are among the highest worldwide, with an annual rate of 1.7 per cent in the past 15 years. Agricultural expansion and the dependency on biofuels have been identified as the main drivers of deforestation in the country, as still more than 40 per cent of the national energy consumption is based on biomass and firewood. The "Poverty, Reforestation, Energy and Climate Change Project" (PROEZA) promotes sustainable agroforestry development and aims at increasing the resilience of highly vulnerable households in the Eastern provinces of Paraguay. The project has three components. The first component is called "Planting for the Future" and targets agricultural households; the second component is "Sustainable Landscapes and Responsible Markets", targeting medium-size private landowners; and the third component, "Good Governance and Law Enforcement", targets public institutions. The first component of the project, which aims to encourage reforestation of land and decrease the demand for firewood, was identified to be suitable for an impact evaluation. Two interventions were designed for this component. Intervention 1 provides training on climate-smart agroforestry production systems to smallholder farmers and intensive technical assistance for the establishment of such systems, including a conditional cash transfer (CCT) on the adoption and maintenance of the new production system. Intervention 2 targets the demand side of biofuel by providing households with efficient cooking stoves that require less firewood or charcoal.

The AE and executing agency responsible for the coordination, technical implementation, and monitoring and evaluation (M&E) is the Food and Agriculture Organization (FAO). The steering and executive committee consists of *Secretaria Técnica de Planificación del Desarrollo Económico y Social*, the Paraguayan Institute for the Indigenous (INDI), the National Forestry Institute, the Ministry for Social Development, the Ministry of Agriculture and Livestock, the Environmental Secretary and the Vice Ministry of Mines and Energy.

There are two main evaluation questions. First, are households benefiting from PROEZA more resilient to weather events? Second, do the efficient cooking stoves help to mitigate climate change? Sub-questions are as follows:

Intervention 1:

- Are farmers changing their agricultural production towards climate-smart plantation systems?
- Does PROEZA increase forest coverage?
- Will PROEZA households experience a lower degree of yield variability and have more stable income flows?

Intervention 2:

- Do more households use efficient cooking stoves?
- Does PROEZA lead to a decrease in firewood consumption?
- Do households have lower expenditure on firewood?
- Do women and children have fewer respiratory problems?

The evaluation design developed by the LORTA team is a phased-in RCT that is clustered at the neighbourhood level, for which the step-wise roll-out of the intervention over the course of five years presents an ideal set up. For the evaluation, PROEZA households in year 2 and year 5 will be considered in order to allow enough time for the agroforestry training to show effects. Neighbourhoods will be randomly assigned to one of the following three groups: 1) no intervention in year 2; 2) agroforestry training and CCT in year 2; and 3) agroforestry training and CCT plus efficient cooking stoves in year 2. This design makes it possible to identify the combined impact of the training and CCTs as well as the additional impact of the improved cooking stove intervention. The comparison group will receive intervention 1 and potentially also intervention 2 in year 5, following the phased-in implementation.

As the proposed methodology was an RCT, obtaining baseline data is desirable but not required if randomization is undertaken with high fidelity. Using data from a government database as baseline has been discussed. If this is eventually judged unsuitable and if the budget allows for it, baseline data could be collected by conducting a household survey in late 2019 or early 2020. The endline data collection will start three years after project implementation (most likely in 2023). Satellite images will also be used for analysis.

Madagascar

The project "Sustainable Landscapes for Eastern Madagascar" (SLEM) aims at implementing sustainable landscape measures to enhance the resilience of smallholder farmers, reduce GHG emissions from deforestation, and make climate-smart investments in agriculture and renewable energy. The sustainable landscape measures consist of a portfolio of activities, among which two will be the focus of the impact evaluation: adaptation and mitigation activities. Adaptation activities include the provision of trainings, inputs and technical assistance to smallholder farmers in order to promote conservation agriculture practices and alternative sources of livelihood (component 1). For mitigation activities, the project plans to provide trainings, per diems and equipment to physically demarcate the limits of protected forest areas and to patrol these areas (component 2).

These two activities will be implemented by Conservation International Madagascar (CI-M), which is one of the AEs for this project. The other AE is the European Investment Bank. The project will be co-executed by Conservation International, Althelia Climate Fund and the *Bureau National de Coordination des Changements Climatiques*, which is part of the Ministry of Environment and Forests.

The main evaluation questions are 1) whether trainings, inputs and technical assistance lead to reduced vulnerability of smallholder

farmers to climate hazards, and 2) whether patrolling interventions help to reduce deforestation in the protected forest areas. Sub-questions are as follows:

Component 1:

- Do households implement alternative livelihood strategies?
- Do households implement conservation agriculture practices?
- Does the adoption of conservation agriculture practices lead to an increase in agricultural production and food security?

Component 2:

- Do patrollers cover a greater distance during patrols?
- Do patrolling interventions lead to a better enforcement of regulations of the protected forest area?
- Does deforestation increase in other areas as a result of an increase in forest surveillance in the target areas?

During the LORTA field mission, two complementary impact evaluation designs were developed: a cluster randomized phase-in and a DiD approach combined with matching. The experimental approach addresses estimation of the short-term effects of the SLEM project. Long-term effects cannot be captured in the experiment because project activities will eventually be implemented in all eligible forest areas. As a result, endline data will have to be collected before the project is rolled out to the control areas. A quasi-experimental design will therefore be implemented as an additional strategy to capture longer-term effects. Under this design, the control group will consist of forest areas outside the study area.

According to the current data-collection timeline, an assessment of the balance at baseline between the selected treated and comparison groups could be performed in May 2019. In May 2021, after one and a half years of complete interventions, a first assessment of the impact of the project could be made. In May 2022 and May 2023, two additional data collection rounds will provide crucial insights on the evolution of the effects of the SLEM interventions over time. During the whole project cycle and parallel to these evaluations, detailed monitoring data and satellite data will be collected and analyzed.

Vanuatu

The project "Climate Information Services for Resilient Development in Vanuatu" aims at increasing the resilience of the population to climate change by developing tailored climate information and disseminating this information to community members across five priority sectors: agriculture, fishing, infrastructure, water and tourism. The key activity in terms of improved climate infrastructure is the installation of a Doppler radar on the main island of Efate, planned for September 2019. The Doppler will cover a radius of 400 kilometres. Its location was chosen such that it will cover the largest number of people possible, which at its current location is slightly less than half of the country's population. In addition to the Doppler, the project will also install eight new automated weather stations at various locations both inside and outside the Doppler's radius, as well as new instruments such as rain gauges, ocean gauges and other weather forecasting tools. While these tools do not have the accuracy and predictive power of the Doppler, they will add to the existing body of climate data used to make long-term forecasts and analyze trends.

The project is jointly implemented by the Secretariat of the Pacific Regional Environment Programme (SPREP), a regional organization based in Samoa, and the Vanuatu Meteorology and Geohazards Department, a national governmental department. SPREP is the AE of this project.

The main evaluation question in this impact evaluation refers to whether the project interventions increase the adaptive capacity of households. Sub-questions are as follows:

- Does the early warning capacity of the Doppler lead to increased resilience of

livelihoods of those living within its radius (measured by less damage, loss of life, loss of crops)?

- Does the use and application of climate information services lead to an increase in overall yields of farmers and less volatility thereof?
- Does the use and application of climate information services lead to less damage, loss of life and injury for workers in fishing?

The expected effect of the Doppler radar is to inform upcoming weather events earlier than is possible with alternative technologies. Given the implementation modalities, a DiD design combined with PSM can be applied to measure the impact of the Doppler radar installation on peoples' behaviour and decision-making, and ultimately their resilience to major climate shocks. As an alternative to the Doppler radar impact evaluation, an assessment of the remaining climate information components with a combination of PSM and DiD was proposed.

The two proposed timelines differ with regard to the timing of the baseline data collection, as this should take place at the latest point possible before project implementation (for the first approach in the first quarter of 2019, and for the second approach in the third quarter of 2019). Endline data collection will be in the third quarter of 2021 in both cases.

Zambia

The project "Strengthening Climate Resilience of Agricultural Livelihoods in Agro-Ecological Regions I and II" aims at increasing the resilience of smallholder farmers in Zambia in view of climate change and climate variability. It takes a value chain approach, addressing barriers to climate-resilient agriculture across key stages of the value chain – planning, inputs, production and post-production – through various activities such as input support, training and infrastructure development in three components. The first component aims at improving the quality of weather/climate-based information and the dissemination thereof, and the second component is directed at irrigation and input support, mostly in form of the government's Farmer Input Support Program (FISP). The third component, on markets and commercialization, helps to drive the production of resilient agricultural commodities and to ensure the sustainability of the first two components. In this way, the project implements targeted interventions to strengthen and promote viable climate-resilient value chains relating to smallholder agriculture.

The AE for this project is UNDP. The main executing entity for the proposed project is the Zambian Ministry of Agriculture, which partners with a range of organizations and governmental bodies including the Zambian Water Resources Management Authority, the Zambia Meteorological Department, FAO and the World Food Programme (WFP).

The main evaluation question is whether the project enables farmers to adapt to unavoidable consequences of climate change through the adoption of climate-resilient practices. Sub-questions are as follows:

- Does the project lead to the adoption of climate-resilient practices?
- If so, do these in turn lead to reduced vulnerability, improved yields and higher income?
- Which messages are most effective in inducing climate- and weather-based decisions by smallholder farmers? How can farmers be nudged through messages to plant climate-resilient seeds/adopt sustainable agricultural practices?
- To what extent do farmers benefit from input support? Does input support lead to more climate-resilient practices?

The evaluation of the overall project impact was proposed through a DiD design combined with PSM. This evaluation would be complemented by RCTs on the effects of farmer input support and on how to deliver effective messages to farmers. The baseline data collection should happen in 2019, and

high-frequency data (phone surveys) are to be collected throughout the whole project cycle. The endline data collection will be scheduled for 2025 (after project implementation).

Georgia

The project "Scaling-up Multi-Hazard Early Warning System and the Use of Climate Information in Georgia" aims at creating a proactive integrated climate risk management approach through the establishment of a countrywide multi-hazard early warning system (MHEWS) and use of climate information in planning and decision-making. The first component of the project aims at the expansion of the hydrometric monitoring network, multi-hazard zone mapping, socioeconomic vulnerability assessments and institutional capacity-building for the production of reliable climate information; the second component targets the development and implementation of MHEWS and the development and delivery of agrometeorological advisory information services for farmers and agricultural businesses; and the third component aims at public awareness and capacity-building on climate risk, institutional capacity-building on risk management and the implementation of community-based activities for early warning and climate risk management.

The AE for this project is UNDP. The implementation is led by the Ministry of Environment Protection and Agriculture, in collaboration with the National Environmental Agency, the Environmental Information and Education Center, the Ministry of Regional Development and Infrastructure, and local governments.

The third component of this project was identified as the most suitable for rigorous impact evaluation. This component contains two activitiethat are community-based and directly target vulnerable communities across the country: 1) implementation of community-based early warning schemes (CBEWS) and community-based climate risk management (CBCRM), and 2) implementation of infrastructural defence measures against floods and mudflows in selected sites.

The main evaluation question is what impact do CBEWS, CBCRM and structural defence measures have on households' resilience against natural hazards. Sub-questions are as follows:

- Are early warning messages beneficial for protecting the community against the effects of climate hazards?
- Are community members aware of the hazard risks they face, and do they receive and understand early warnings?
- Do structural defence measures (flood and mudflow defence structures) contribute to a greater feeling of safety of the beneficiary households?
- Do these households achieve better economic status and enhanced livelihoods as a result of long-term investment plans that will be enabled through increased safety?

Two impact evaluation designs were developed. CBEWS and CBCRM were considered suitable for an RCT design because there seem to be more eligible communities than possible beneficiaries. Due to budget constraints, only 100 communities are planned to be beneficiaries. If this evaluation is pursued, baseline data collection would take place in 2021 and the endline collection in 2025. The design considered for structural defence measures follows DiD coupled with PSM methodology. In this case, baseline data would be collected in 2019 and endline data in 2025.

The full reports for all eight countries are presented in Annex III.

Table 2 Summary of evaluation questions and designs

COUNTRY	QUESTION	DESIGN
Malawi	Does the PICSA training help farmers to adapt to climate change, thereby leading to more resilient livelihoods?	DiD with matching
Mongolia	*Does the Eco-loan programme and the related awareness campaign lead to reduced greenhouse gas emissions?*	*None selected*
Uganda	Do community members become more resilient against climate change disasters due to wetland restoration and alternative livelihood trainings?	Randomized phase-in or matching
Paraguay	Are households benefiting from this project more resilient to weather events? Do efficient cooking stoves help to mitigate climate change?	Randomized phase-in
Madagascar	Do adaptation interventions lead to a reduction of households' vulnerability to climate hazards? Does patrolling in protected forest areas result in a reduction in deforestation?	Randomized phase-in and DiD with matching
Vanuatu	Does the early warning capacity of the Doppler radar and the use and application of climate information services increase the adaptive capacity of households?	DiD with matching and randomization
Zambia	Do the project activities enable farmers to adapt to unavoidable consequences of climate change through the adoption of climate-resilient practices?	DiD with matching and randomized phase-in
Georgia	What is the impact of CBEWS and CBCRM on households' resilience against natural hazards?	Randomization and DiD with matching

III. LESSONS LEARNED AND RECOMMENDATIONS

A. LORTA DESIGN WORKSHOP

The LORTA Design Workshop in 2018 was successful in delivering a diversified programme and introducing the workshop participants to the main ideas and concepts of impact evaluation. It proved effective to alternate presentations in the large group with group work sessions because this enabled the participants to immediately apply the presented topics to their projects. This fostered the understanding of impact evaluation. Given the workshop length of 2.5 days, each group work session was rather short – between one and two hours. In some cases, the group work tasks could not be completed due to time constraints, which might have been disappointing for the project representatives.

Consultation and co-ownership of the design and implementation of measurement systems between the LORTA team and the project team is essential. This means project teams should come on board as early as feasible. Ideally, project teams should engage with the IEU LORTA team before they start planning implementation (and after Board approval).

For some projects, the group work proved to be particularly difficult because limited information was available on project implementation. The evaluation specialists then spent an extended period of the group work time on clarifying questions such as where and when certain project activities would be implemented, who the beneficiaries were supposed to be, how the selection of beneficiaries was to take place, and whether there were more eligible units than potential beneficiaries. In a few cases, these questions were not answered at the end of the workshop – either because implementation of project activities had not yet started and these details not been defined, or because the representatives who attended the workshop did not know such detailed information. In these cases, group work centred around discussing project modalities and only partly succeeded in defining an evaluation method, conducting power calculations and suggesting an evaluation timeline. Ahead of future LORTA workshops, the IEU will request invited project teams to fill in a questionnaire to gather the key project modalities listed above although it is clear that this is not a fail-safe strategy either since in many cases, the projects invited are still not clear on the details themselves. IEU then needs to decide whether the flexibility afforded by having teams come early on, is important, while there is still time to use implementation plans (that will be done later) or, if details are important early on (which also means that many project details are already determined). The composition of the attendees that represent a project could be discussed and analysed a bit more. Although it seems desirable to have high-level representatives (such as the NDA), implementation staff and M&E officers present, it is also clear that these officials are not involved in day to day management. The ideal set of invitees seem to be project managers/leads and M&E officers.

The group work sessions were very helpful to improve participants' understanding of key impact evaluation elements and concepts.

B. PROJECT SELECTION

In the first year of LORTA, projects were selected to achieve a balance of adaptation and mitigation activities from both the private and public sectors. In the future, this representative mix may not always be required. In fact, the IEU plans to focus on private sector projects, primarily in the energy sector, in the second year of LORTA. A thematic focus is a good idea because the formative work in Phase I can benefit from important synergies between the projects – for example, with regard to the elaboration of theories of change or the identification of outcome and impact indicators. This formative work could also be accompanied by a meta-analysis that would help to identify gaps in the literature. In addition, preparation of the evaluation specialists for the LORTA Design Workshop becomes easier because insights from a literature review can be applied to several projects, not only the one or two projects that a specialist is assigned to at the workshop. This allows for better referencing across projects and more effective communication between the evaluation specialists.

GCF-funded projects and investments need to budget early for the cost of undertaking LORTA impact evaluations. This facilitates high-quality designs for building real-time measurement systems and credible measurement.

Optimally, at the time of the workshop, it should be clear that projects need to have part of their own budget dedicated for measuring impact and outcomes in an attributable manner, chiefly for data-collection costs,

because LORTA primarily finances technical advice and quality assurance by evaluation specialists. In several cases, it turned out that projects had either no specific budget line for impact evaluation or that their evaluation budgets were too small to allow for reasonable sample sizes or repeated data collection. Another requirement that may be useful is for project teams to recognize that data collection on the entire population is neither necesssary nor desirable. In many cases, he provision of sufficient funds for impact evaluation had to be discussed between the IEU and the projects on a case-by-case basis. However as the table shows, once project teams are able to consider i) effect sizes; ii) overall targeted outcomes (and associated indicators); iii) measurement methods; iv) real-time data collection; v) sample sizes associated with the key variables of intered; vi) theories of change with associated bottlenecks, then most impact evaluation designs are feasible. To their credit, the UNDP projects did include these costs in their overall budgets before agreeing upon the funding proposal budget with the GCF. This covered three projects (Zambia, Malawi, Uganda). In Madagascar, although Conservation International had not incorporated the measurement of impact/outcomes beforehand, they were able to accommodate it within their overall measurement budgets, without a change. In the case of Paraguay and Vanuatu, there is a high commitment to undertake these measurements and extraordinay efforts have been made to consider how these costs may be accommodated. Overall a key recommendation is that budget be included in the list of eligibility criteria that are applied at the time of selecting the projects for the workshop in order to select from the participating projects those that will enter Phase I of LORTA.

The 15 projects that attended the first LORTA Design Workshop in Bangkok were selected from among the universe of projects funded by the GCF. Thirteen of the participating projects had been approved between November 2015 and March 2018; two projects were still in the application stage at the time of the workshop. The latter could not be selected into Phase I because funding approval was still unclear. We nevertheless believe that the project representatives benefited from attending the workshop because they were introduced to the concept of impact evaluation and could consequently reflect on the feasibility of an impact evaluation in their planned projects. We, however, do not recommend inviting projects that are not yet approved invited to future LORTA Design Workshops to maintain the impartiality and independence of IEU.

The eight projects selected into Phase I were in different phases of project implementation. Some were already in an advanced phase, some in an intermediate phase, and some had not even had their inception workshops. Those projects at an advanced stage of implementation (Malawi, Mongolia) turned out to be challenging in terms of elaborating an impact evaluation design. Project activities had already been implemented, which ruled out their random assignment and made an experimental impact evaluation approach impossible, even though it is the methodologically preferred approach. The ideal timing to include a project into LORTA seems to be when the funding is approved but project activities – or at least those that are to be evaluated – have not yet started to be implemented although it is clear that this means that there will be fewer details available at the time of the workshop on some key activities including implementation plans.

C. PLANNING AND CONDUCT OF COUNTRY MISSIONS

It is imperative that in-person country missions engage the full programme team (especially its senior staff) and have hands-on sessions where project staff can work through the implications of theory of change, measurement systems, survey and analyses plan. These country missions should also

include context analyses and engagement with the ultimate beneficiaries of GCF investments. Thorough engagement of relevant accreditation entity staff cannot be overemphasised.

The timing of the field missions was largely determined by the availability of the project teams. In two cases, the field mission was combined with other meetings held in the country (a GCF readiness workshop in Mongolia and the National Planning Meeting for the GCF-funded project in Zambia), which turned out to make the logistics challenging but also ensured the presence of all key stakeholders at the LORTA capacity-building workshop on impact evaluation.

When the field missions took place in Georgia and Zambia, the project teams were just about to be recruited, and many details about project activities – both regarding design and timing – were still unknown. In consequence, impact evaluation designs were elaborated on the basis of assumptions and will have to be re-assessed after more details become available. In this case, field missions may be postponed to when project teams are in place and inception workshops have been held although admittedly this means that in many cases, implementation plans get determined quickly and there is little space for planning and rolling out impact evaluation designs. Alternatively, a first field mission could take place as early as in the cases of Georgia and Zambia, but a second field mission would become necessary to re-assess the feasibility of the preliminary impact evaluation designs. This alternative approach offers a window of opportunity because during the first field mission the project team can benefit from the LORTA team's expertise in impact evaluation to adjust the implementation of the project activities to the evaluation questions the project team wants to be answered. But this also increases the costs for IEU and it will need to be determined if this is feasible. In the case of Madagascar, the team was already on board, but the LORTA team was able to work with the entire team to discuss and parse key elements of the theories of change and consequently present a combined plan for implementation/real-time measurement/survey data collection in one sheet. It helped immensely in that case that the project lead and the AE representative were champions and were excellent ambassadors in working through key elements of the theories of change and implementation/measurement plans with the visiting team.

The IEU and C4ED elaborated Terms of Reference for the field missions, which contained a preliminary agenda for the week-long mission. The final agenda was usually put together by the project team and/or the AE, sometimes in collaboration with the NDA. In Georgia and Zambia, where no project team had been recruited, the responsibility for organizing the field mission rested with the AE. The final agenda usually deviated somewhat from the preliminary agenda suggested in the Terms of Reference as the availability of key stakeholders dictated the sequencing of meetings and workshop sessions.

The capacity-building workshop on impact evaluation was usually held as a one-day workshop. In Madagascar, all key stakeholders (24 individuals in this case) were not only available for one day but throughout the week, which allowed for the modification of the mission agenda to include group exercises and lively interactions. In fact, the workshop was expanded over three days and informed the participants on the LORTA objectives, analyzed the project's theory of change, defined key indicators of the project, discussed programme implementation details and presented key impact evaluation concepts. By applying the workshop content to the project, the workshop combined pure capacity-building with the elaboration of an impact evaluation design. This interactive workshop style proved very useful and could be applied in any future field missions when key stakeholders can be gathered for more than one day. Detailed documentation (to be

compiled by the respective LORTA team) of the experience in Madagascar would be helpful for future teams to use as an orientation.

The field missions were successful in achieving buy-in for an impact evaluation from the project teams. All teams were enthusiastic about the benefits of impact evaluation and wished to see their project activities being assessed in terms of impact achieved. Most of them therefore willingly shared information and data ahead of and during the field mission and took the LORTA teams to project areas and in some cases the project implementation sites. This close collaboration for one week established a well-functioning working relationship between the project teams and the LORTA teams that is ongoing to date.

Some project teams shared much more documentation with the LORTA teams ahead of the field mission than others did. In one case, the project team refused to share documents because these were still preliminary. While drafts are not official documents, having seen them would have enabled the LORTA team to prepare much better for the field mission. It is definitely important to have project teams be convinced of the importance of sharing information ahead of the field missions. The use of a confidentiality agreement may be helpful in this regard and the IEU will be taking this forward including agreements on collaboration.

In all field missions, the field trips were essential for the LORTA teams to get comprehensive insights and a good level of understanding of the project activities. This, in turn, helped in further sharpening the ideas for impact evaluation designs. The field trips usually took the LORTA teams to selected project sites and enabled them to observe project activities and conduct interviews with project staff and beneficiaries. In the two cases where the projects had not officially started (Georgia and Zambia), the LORTA teams visited pilot project sites or sites of projects similar to the GCF-funded interventions. Given their crucial importance, the field trips should remain an obligatory part of any future field mission.

Ahead of the field missions, C4ED developed a standard set of slides (available in English, Spanish and French) to be used in each of the missions. These slides covered an introduction to LORTA, a session on developing a theory of change, and a session on methods of impact evaluation. These slides were used in all field missions but were adjusted to each project's context (e.g. referring to examples that were similar to the project activities). The slides were illustrative for the participants and effective in conveying the key messages. Further improvement for future use would be possible by including more examples from the area of adaptation to and mitigation of climate change to illustrate the benefits and methods of impact evaluation.

D. EVALUATION DESIGNS

In response to the first draft of an impact evaluation design for Malawi, which was a DiD design, the IEU proposed that it be clearly defined as to which kinds of methods would be acceptable from a quality point of view. There was serious concern about pure DiD designs because they rely on the so-called "parallel trends"1 assumption, which is usually hard to confirm. The IEU elaborated a one-page description of acceptable methods, which C4ED subsequently turned into a graphical representation in the form of a decision tree (presented in annex II). We welcome this definition of quality standards and recommend making the one-page

[1] The parallel trend assumption requires that – in absence of treatment -, the difference between treatment and control group remains constant over time.

description and the decision tree available to all people involved in LORTA.

When the LORTA teams returned from their field missions, they immediately started working on the impact evaluation design reports (presented in annex III). In the course of Phase I, a quality assurance process was developed that proved to be effective. The first stage of this process is internal to C4ED and consists of a presentation of the impact evaluation design in the internal seminar series, a peer-review process and a review of the design report by the Head of Quality Assurance. Only after revisions from these steps were incorporated into the report was it submitted to the IEU and the project teams. The resulting comments and feedback were again taken into consideration. We believe that this quality assurance process worked well.

The process described here was not implemented from the outset. For the first two projects, the LORTA teams did not present their designs in the C4ED internal seminar. It was quickly realized that this was a weakness because LORTA teams did not know about each other's work and could not learn from each other.

Impact evaluation designs were always developed in close collaboration with project teams and AEs. This joint effort was crucial for the stakeholders to gain ownership of the impact evaluation and to get a good understanding of the evaluation needs and the stakeholders' respective roles. This is critical in setting the stage for an ongoing close collaboration throughout the implementation of the projects. The project teams also provided feedback on the impact evaluation design reports, either before submission to the IEU or subsequently. Their feedback was key in designing workable evaluation plans suitable to the constraints of the in-country situation. Collecting feedback from project teams could resolve potential misunderstandings and would ensure that the report correctly reflects the interests of the project team. However it is clear that in all these cases the AE and the IEU need to be part of the process and this needs to be underscored at all times. In many cases, communication gaps were resolved through communication between IEU and the AE.

E. CAPACITY OF PROJECT TEAMS

In all projects that went through Phase I in 2018, the project teams showed genuine interest in measuring the impact of their project's activities. Many project team members, especially those who had attended the LORTA Design Workshop in Bangkok, understood the difference between M&E and impact evaluation and acknowledged the need for a comparison group. However, the conditions required for this comparison group to be valid were often not well understood. Also, the project teams tended to be unaware of the sample size requirements for impact evaluation. All projects will therefore crucially benefit from further technical advice and quality assurance provided through LORTA during Phase II. Those projects that do not progress to Phase II may still conduct an impact evaluation, but given the lack of detailed methodological knowledge, it is clear that they will need external technical advice that would have to be contracted and financed by the projects' own budgets.Currently this independent, credible advise is provided through IEU's budgets.

An additional potentail area support to project teams, is contracting consultancy firms for data collection for M&E or impact evaluation. In two cases, baseline data collection had been scheduled by the project teams before having a good understanding of the implementation of project activities. On the LORTA team's recommendation, data collection was postponed until after the field missions. However, sampling strategy and sample size for Malawi had to be decided under great time pressure because the project team had already hired a consultancy firm for data collection, and this firm was about to go to the field right after the LORTA field mission. In Vanuatu, a research firm had been

hired for a socio-economic analysis including the collection of survey data without specifying the number of respondents to be interviewed. There are three recommendations here: first, that a checklist be developed, for all GCF LORTA projects which outlines the process to be followed in contracting consultancy firms which includes quuestions that contracted teams should be able to answer; second, that a template TOR be provided to LORTA teams; and, third, that there is close collaboration between LORTA teams and project teams in choosing firms for data collection, as this will help ensure data quality.

F. PROJECT DESIGN (FOR FUTURE PROJECTS)

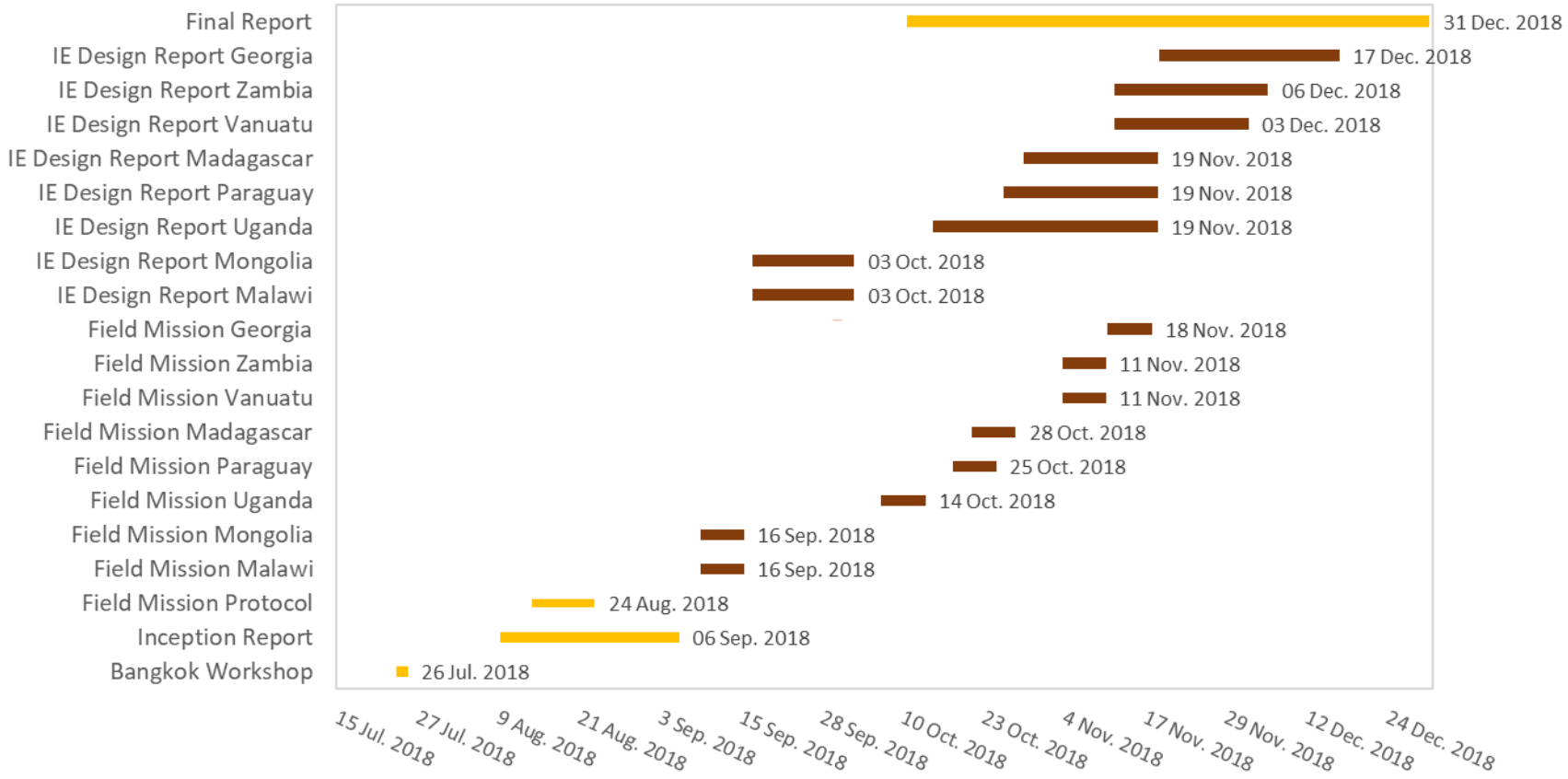
As has become clear in the monthly LORTA Working Group Meetings attended by the IEU, C4ED, the GCF Secretariat and other GCF divisions, one concern important to the GCF Secretariat is to develop a procedure to determine which GCF projects will be likely candidates for impact evaluation in the future. This is a crucial point because projects that undergo impact evaluation should plan for it as early as at the proposal stage and include a sufficient amount of funding in their proposed budgets.

The extent to which GCF projects *cause* lowered GHG emissions and increased climate resilience can only be measured with the help of rigorous impact evaluation. Given its resource intensity in terms of time and money, impact evaluation should focus on research questions that lack evidence and are innovative. In the context of climate change investments, the empirical evidence is very scarce. It is recommended that most, if not all, GCF projects could potentially include an impact evaluation, at least for selected aspects of the project.This will have implications for budget, design and development of theories of change.

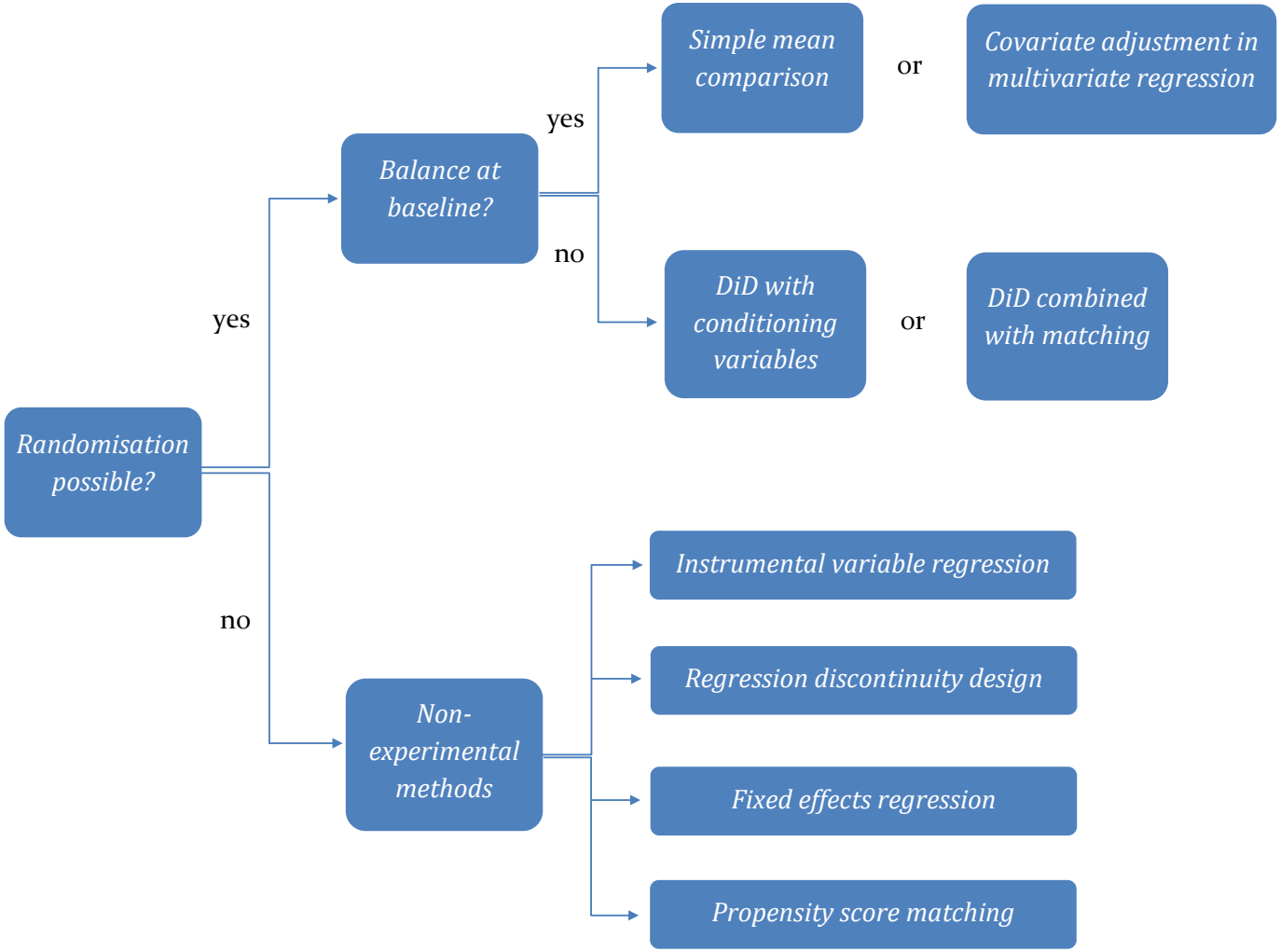
At the same time, not all projects and not all outcomes of projects may be "impact evaluable". The IEU has elaborated a funding proposal review checklist that intends to serve in determining whether proposed projects are constructed to credibly measure the impact caused by the GCF investment. This checklist can guide GCF Secretariat staff in identifying projects that fulfil the quality standards needed for impact evaluation. For example, it can screen out projects that have not clearly defined the type of beneficiaries as well as their eligibility and selection, which would in turn lead to the inability of evaluation specialists to come up with an impact evaluation design.

For teams writing project proposals, the complication with deciding for or against suggesting impact evaluation is that a decision needs to be made on how much money to apply for. While it may be impossible for the GCF Secretariat to advise on exact amounts (which depend on types of data collection, local costs, sample sizes, number of data-collection rounds, etc.), it would be desirable that the GCF Secretariat has a clear understanding of the resource intensity of impact evaluation and does not advise project proposals to reduce the budget lines assigned to this exercise, should budget reductions at the proposal stage be necessary. The amount assigned to impact evaluation in the first wave of LORTA projects ranges from USD 125,000 to USD 450,000. It is to be expected that future projects will require similar amounts.

ANNEX I LORTA OVERALL ACTIVITIES



ANNEX II DECISION TREE ON IMPACT EVALUATION METHODS



ANNEX III IMPACT EVALUATION DESIGN REPORTS FOR THE EIGHT COUNTRIES PARTICIPATING IN THE LORTA PROGRAMME

IMPACT EVALUATION DESIGN REPORT 1: MALAWI

TABLE OF CONTENTS

TABLES

FIGURES

I. INTRODUCTION

A. THE LORTA PROGRAMME

Evaluating the impact of development projects and programmes has gained importance in recent years. Impact evaluation not only allows for increased transparency by measuring the effects of investments, it also provides the opportunity to design and implement development projects more effectively. To contribute to this development, the Independent Evaluation Unit (IEU) of the Green Climate Fund (GCF) has started the Learning-Oriented Real-Time Impact Assessment (LORTA) programme to be able to keep track of GCF projects in terms of performance and results, and to enhance learning within the GCF.

The LORTA programme aims to

- Embed real-time impact evaluations into funded projects so GCF project task managers can quickly access accurate data on a project's quality of implementation and likelihood of impact
- Build capacity within projects to design high-quality data sets for overall impact measurement.

The purpose of the impact evaluations is to measure the change in GCF key result areas that can be attributed to project activities. The LORTA programme will inform on returns on GCF investments and help GCF projects track implementation fidelity. The objectives of LORTA include the following:

- Measuring the overall change (outcome or impact) of GCF-funded projects and enhancing learning
- Understanding and measuring results at different parts of theories of change
- Measuring the overall contribution of the GCF to catalyzing a paradigm shift and achieving impacts at scale

Currently, the LORTA programme is in Phase I (formative engagement and design). In the first year (2018), the IEU will support eight GCF projects to build high-quality, theory-based impact evaluation designs at inception. Formative work will include engaging with project teams, accredited entities (AEs) and GCF staff; designing theory-based impact evaluations; and formulating protocols for database development.

The second phase of LORTA will involve the main impact assessment stage (3–5 years), while the third phase will include analyzing baseline and endline data, discussing results, and engaging with diverse stakeholders to share results and incorporate feedback as required.

B. THE MALAWI PROJECT

The project "Saving Lives and Protecting Agriculture-Based Livelihoods in Malawi: Scaling Up the Use of Modernized Climate Information and Early Warning Systems" (M-CLIMES) is one of the eight projects selected to be part of the inception stage (Phase I) of the LORTA programme. The accredited entity for this project is the United Nations Development Programme (UNDP).

The project aims to increase the resilience of rural livelihoods to climate variability. This is planned to be achieved through scaling up the use of modernized early warning systems (EWS) and climate information in the country. More specifically, the project plans to install new automated weather stations (AWS), build capacity and deliver more accurate and better customized climate information to vulnerable food-insecure, flood-prone and fishing communities.

Malawi remains one of the least developed countries, currently ranking 171 out of 189 in the United Nations Human Development Index (2018). According to the Fourth Integrated Household Survey

(2015/2016), 81 per cent of the population lives in rural areas, with the majority (89 per cent) of households engaged in either agricultural or fishing activities. The agriculture and fishery sectors represent a key source of economic growth for the country and are extremely vulnerable to the highly variable climate of Malawi. The major climatic hazards faced are frequent droughts and floods, as well as strong winds on Lake Malawi.

In early 2015, intensive rainfalls caused flooding that predominantly affected rural households in the Southern Region of Malawi. The floods displaced around 230,000 people and killed 106.[2] Soon after, the population was hit by severe drought, which resulted in the Government of Malawi declaring a state of disaster in April 2016, affecting a total of 24 out of 28 districts.[3]

The M-CLIMES project aligns with the Government's priorities on climate information and early warnings set in the Malawi Growth and Development Strategy and the National Adaptation Programme of Action.

The project will be implemented in 21 districts by the Department of Disaster Management Affairs (DoDMA) in collaboration with a multiplicity of departments and institutions: the Department of Climate Change and Meteorological Services (DCCMS), Department of Water Resources (DWR), Department of Agricultural Extension Services (DAES), Department of Fisheries, and the National Smallholder Farmers Association of Malawi (NASFAM).

The project has three goals:

- Expansion of networks that generate climate-related data to save lives and safeguard livelihoods from extreme climate events
- Development and dissemination of products and platforms for climate-related information/services for vulnerable communities and livelihoods
- Strengthening communities' capacities for use of EWS/climate information in preparedness for response to climate-related disasters

II. MALAWI FIELD MISSION

A. GENERAL REMARKS

An evaluation team, consisting of two consultants from the Center for Evaluation and Development (C4ED) and one IEU staff member (henceforth referred to as the LORTA team), was formed to lead the field mission from 10 to 14 September 2018. The task of the team was to engage closely with key stakeholders of the project – namely, the national designated authority (NDA), accredited entity (AE), implementing agencies, project staff and potential end beneficiaries – to ensure their interest in and understanding and sense of ownership of the planned theory-based impact evaluation.

During the field mission, the LORTA team held meetings and capacity-building workshops with the key stakeholders. Meetings, in the form of expert interviews, were used to acquire the maximum possible information about the GCF-funded project. Stakeholders were interviewed regarding their views about the project's implementation and monitoring strategies, expected impact, challenges and possible solutions. The meetings not only informed the LORTA team about the project but also aimed at fostering collaboration and trust between the team and on-site involved parties. In addition, a capacity-building workshop on impact evaluation was held, targeted at the key stakeholders.

[2] Government of Malawi. (2015). *Malawi 2015 floods post disaster needs assessment report.* Retrieved from https://www.gfdrr.org/sites/default/files/publication/pda-2015-malawi.pdf
[3] World Bank. (2016). *Malawi drought 201522016: post-disaster needs assessment (PDNA).* Washington, DC: World Bank Group. Retrieved from http://documents.worldbank.org/curated/en/640011479881661626/Malawi-drought-2015-2016-post-disaster-needs-assessment-PDNA

Beside conveying technical knowledge, the aim of this workshop was to emphasize the benefit of theory-based counterfactual approaches and real-time learning and measurement.

Under the guidance of the LORTA team, an impact evaluation design was worked out for the Malawi project. The LORTA team conducted context analyses, examined the existence of appropriate counterfactuals (i.e. comparable treatment and control groups), elaborated a theory of change, assessed the availability of baseline administrative and secondary data sources, and acquired budget information. The results from this undertaking are presented in the following sections.

B. THE MISSION AGENDA

The LORTA team received prompt collaboration from the UNDP Project Coordination Unit (PCU). The agenda – shown in appendix I of this design report – was developed to facilitate the joint attendance of all key stakeholders at the LORTA workshop and to allow for field visits.

In particular, the LORTA workshop was condensed into a day, but it contained all the ingredients to inform the key stakeholders on the LORTA objectives, present key impact evaluation concepts, develop a detailed theory of change and discuss programme implementation details with the implementing partners.

The workshop was successful in that it benefited the key stakeholders and the LORTA team. On one hand, the presentations and interactive discussions on theory of change and implementation brought all the key stakeholders together on the same page with respect to ownership and understanding of their contribution within the M-CLIMES project, from the project objectives to the evaluation needs and implementation strategies. On the other hand, the LORTA team benefited from gathering in such a short time a rich set of crucial information to help them design the impact evaluation. They also benefited from gaining understanding and collaboration from the key implementing agencies in adjusting their plans to accommodate the evaluation design.

During the workshop discussions on the theory of change of all three components of the M-CLIMES project, and discussions related to implementation plans of these components, it became clear that the only suitable component of the M-CLIMES project for the impact evaluation was the Participatory Integrated Climate Services for Agriculture (PICSA) programme. The "[PICSA] approach aims to facilitate farmers to make informed decisions based on accurate, location specific, climate and weather information; locally relevant crop, livestock and livelihood options; and with the use of participatory tools to aid their decision making."[4] This conclusion was reached based on the restrictions of the project in terms of budget and timeline. The three components of the project are dissimilar in terms of outcomes and target population, and thus multiple impact evaluations would be necessary to evaluate the whole project. However, the budget is sufficient for only one impact evaluation. Also, PICSA is the only activity that has been clearly defined and for which a roll-out plan has been determined. It would be impossible to prepare a viable evaluation design for the other components as it is not clear exactly what will happen and when. At the same time, the survey company has already been hired, and the baseline survey cannot be delayed any longer. Therefore, PICSA is the best choice to evaluate the intervention that is anticipated to begin soon after the baseline survey. More details on the PICSA roll-out are provided in later sections.

The field visits took place at the Dedza district office and at the Bembeke Extension Planning Area (EPA) office of DAES, and at DCCMS in the town of Blantyre. The visits helped substantially in gaining a clearer understanding of the implementation status of the M-CLIMES project and of the

[4] Dorward, P., Clarkson, G., & Stern, R. (2015). Participatory Integrated Climate Services for Agriculture (PICSA): Field Manual. A step-by-step guide to using PICSA with farmers. Retrieved from https://ccafs.cgiar.org/publications/participatory-integrated-climate-services-agriculture-picsa-field-manual#.W7M0k2gza71.

climate information production process and plans. Appendix II of this design report lists all the people engaged with at the workshop and during field visits throughout the LORTA mission.

During the mission, we were able to establish a tight collaboration with the UNDP M-CLIMES PCU, the UNDP consultant Dr. Babatunde Abidoye and the consultancy firm hired for the baseline survey (Center for Development Management). This enabled us to engage with key stakeholders and to promptly develop an evaluation design and evaluation tools as dictated by the tight M-CLIMES implementation schedule.

C. RESULTS

Discussion of implementation

PICSA implementation plan

The PICSA intervention targets 14 districts that were designated as medium or high food-insecure in the 2015 Feasibility Assessment for the GCF funding proposal.[5] Importantly, PICSA is not implemented only by UNDP within the M-CLIMES project. In fact, the World Food Programme (WFP) has already been piloting PICSA in priority high food-insecure districts as part of the Global Framework for Climate Services (GFCS) programme since 2015. UNDP and WFP are therefore coordinating their efforts, taking into account their respective areas of coverage before deciding in which districts PICSA trainings will be delivered.

The areas of coverage and timeline for PICSA roll-out are presented in

Table 1. The implementation of PICSA is phased in, so that the intervention will not start in all the districts at the same time. The district coverage is divided between DAES and NASFAM.

Table 1 PICSA roll-out

DISTRICT	FOOD SECURITY RISK	UNDP PARTNER	YEAR			WFP GFCS
			2018	2019	2020	
Chikwawa	High	DAES	x			x
Dedza	Medium	DAES	x			x
Chiradzulu	Medium	DAES		x		
Dowa	Medium	DAES			x	
Karonga	Medium	NASFAM		x		
Lilongwe	Medium	NASFAM			x	
Mzimba	Medium	NASFAM			x	
Nkhatabay	Medium	DAES		x		
Ntcheu	Medium	NASFAM	x			
Ntchisi	Medium	DAES		x		
Phalombe	High	NASFAM			x	x

[5] Government of Malawi & United Nations Development Programme. (2015). *Scaling-up early warning systems and use of climate information in Malawi – feasibility assessment.* Risk of food insecurity was measured using the data collected by the Malawi Vulnerability Assessment Committee for their annual vulnerability forecast as the number of years within the decade 2005–2014 during which a given district had any number of households defined as food-insecure. Districts were ranked either low risk (1–3 years), medium risk (4–6 years) or high risk (7–10 years).

DISTRICT	FOOD SECURITY RISK	UNDP PARTNER	YEAR			WFP GFCS
			2018	2019	2020	
Rumphi	Medium	NASFAM	x			
Salima	Medium	DAES		x		
Zomba	High	DAES		x		x

Within each district, a certain number of EPAs were selected by the project implementers as PICSA beneficiaries. The main criterion for selection was proximity to weather stations in order to ensure the availability of reliable weather information. Within each EPA, only some sections will receive the treatment. These sections were selected based on the actual presence of agricultural extension development officers (AEDOs) in a given section, and the specific officer's motivation and ability with regard to the PICSA objectives.

Within each treated district, the M-CLIMES PICSA intervention will cover an average of six EPAs. Within those districts where both UNDP and WFP operate – that is, Chikwawa and Zomba – the respective coverage has been defined at the EPA level.

Staff from the University of Reading conducted a four-day "training of trainers" in preparation for the PICSA roll-out in the targeted communities in 2018. A total of 122 extension agents were trained, of whom 92 were AEDOs and 30 were Agricultural Extension Development Coordinators (AEDCs).

The PICSA implementation strategy requires each trained AEDO, as well as some but not all of the trained AEDCs, to choose two groups of lead farmers[6] – within the agricultural sections for which they are responsible – and to replicate the four-day training as a preparation for the next farming season. Each group should include 25 to 40 lead farmers.

The lead farmers who will receive the PICSA training in 2019 will be responsible for replicating the same four-day training for a group of 25 to 40 contact farmers. An important point is that due to preparation delays for the 2018 roll-out, the training of lead farmers cannot be followed by training of contact farmers before the start of the rainy season. However, it is expected that in 2019 the training of both lead farmers and contact farmers will happen before the next rainy season.

In 2019, each AEDO will be in charge of training two new groups of lead farmers within their respective agricultural section coverage. In parallel, PICSA will be rolled out in six new districts. That is, new AEDOs in these six districts will be trained; these AEDOs will in turn train the lead farmers and the lead farmers will train the contact farmers.

In 2020, PICSA will be rolled out in the remaining four districts with the last round of AEDOs training. In the Year 1 and Year 2 districts, the already trained AEDOs will continue to train more lead farmers. In the subsequent years (2021, 2022, 2023) AEDOs will still continue to deliver the PICSA trainings. The goal is to reach all lead farmers within the treated sections.

Figure 1 illustrates the structure of PICSA trainings. In green are the areas/farmers covered by PICSA. In red are the areas/farmers left out.

[6] Lead farmers are part of the Malawi agricultural extension system. In order to reduce their workload, AEDOs were encouraged to select one or multiple lead farmers in each village who would assist them with training of other farmers. Each lead farmer has a set group of "contact farmers" whom he or she usually trains.

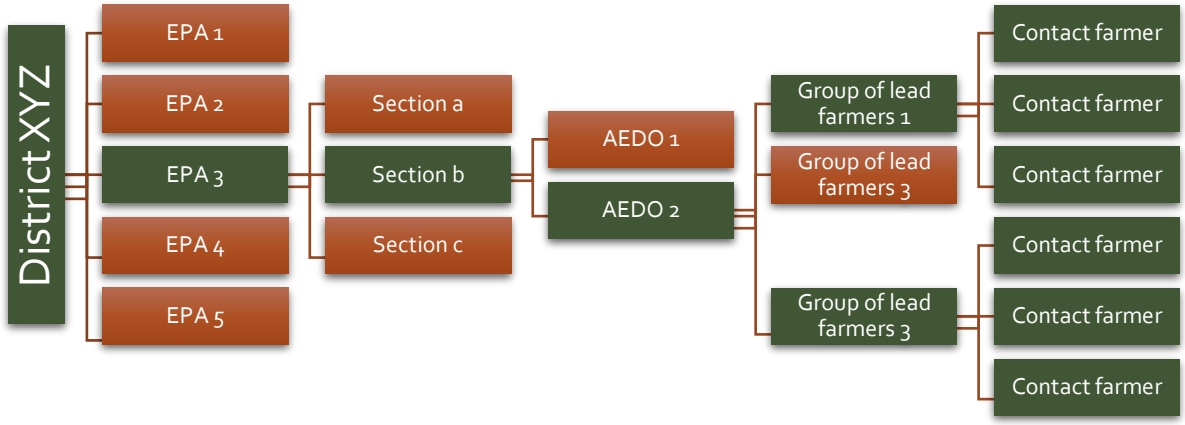


Figure 1 Structure of PICSA trainings

Theory of change

Given the limited budget and the nature of the M-CLIMES project, the impact evaluation will focus solely on component 2. Specifically, we are going to evaluate the impact of PICSA trainings.

The theory of change associated with PICSA trainings is laid out below. It should be noted that the theory of change relies on several underlying assumptions.

i) Inputs

While the budget is provided by UNDP Malawi through GCF funding, the input in terms of staff comes from the implementing partners DAES and NASFAM. The content of the PICSA programme was designed by researchers from the University of Reading, who also carried out the training of extension agents – the AEDOs.

ii) Activities

The actions at the core of the PICSA programme are trainings on interpreting climate information, using historical data to make seasonal forecasts, assigning probabilities to different events, and applying climate-resilient practices based on these forecasts. The training is delivered in the form of participatory activities that teach participants to collectively make decisions on how to respond and/or prepare for imminent climate risks. The trainings are being carried out at three levels:

- The researchers from the University of Reading have trained the AEDOs and other project stakeholders. This training took place in August 2018, and 122 people were trained in total: 92 AEDOs and 30 stakeholders.
- The AEDOs will subsequently train the lead farmers. Each AEDO trains two groups of 25–40 lead farmers in a given year.
- Lead farmers will train their contact farmers (also in groups of 25–40).

i) Outputs

Through the inputs and activities of the PICSA programme a higher exposure to PICSA-related dissemination activities in the treatment group should be achieved. The successful completion of the output stage can be measured in terms of number of farmers who report having participated in PICSA trainings.

i) Outcomes

If the target group benefits from the elements described in the output stage, we should be able to measure first results in terms of enhanced knowledge and practices. In particular, we expect farmers

to have better knowledge of how to use climate information and seasonal forecasts in planning for the next agricultural season, and to apply the recommended steps in line with a specific seasonal forecast.

i) Goals

The main goals behind the PICSA programme are higher agricultural yields, higher revenues, lower food insecurity and enhanced resilience to climate shocks.

These five stages are summarized in Figure 2.

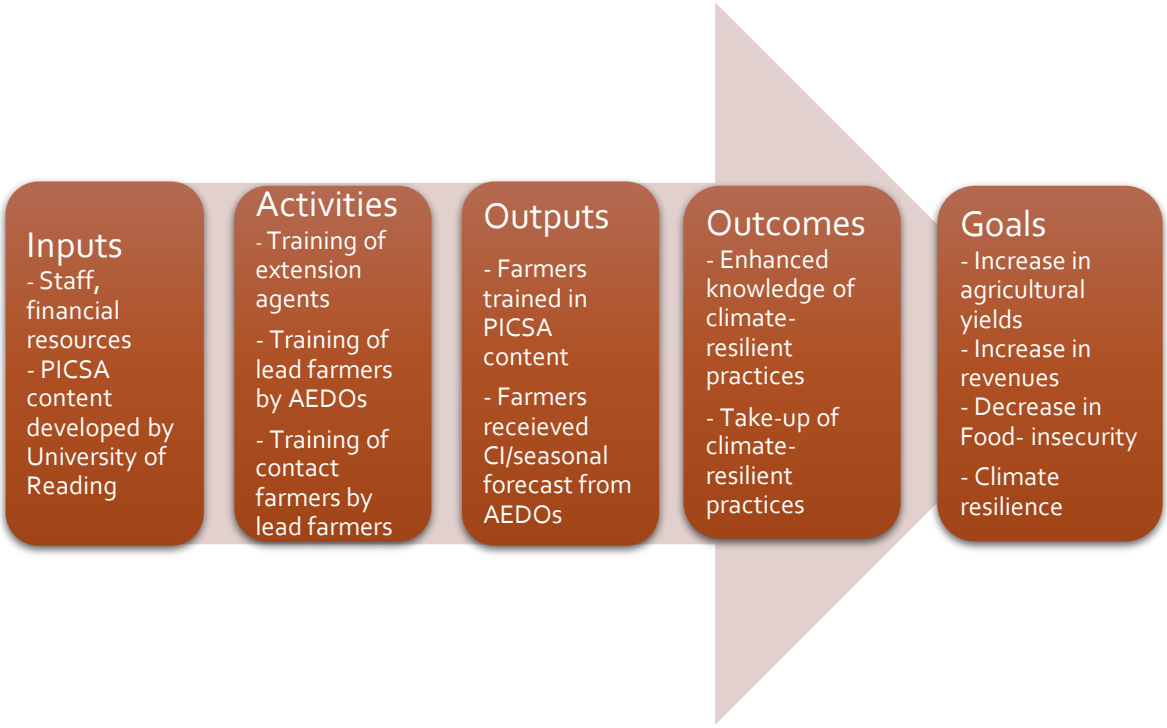


Figure 2 Theory of change

Evaluation questions

The main research questions, derived from the hypotheses underlying the theory of change, are listed below:

- Do farmers access tailor-made climate information?
- Do farmers base their farming plans on seasonal forecasts?
- Are farmers adapting their plans based on changes in forecasts?
- Does the PICSA intervention lead to more resilient livelihoods?
- Are farmers indeed better equipped to face climate risks and extreme weather events?
- Does better planning based on accurate and tailor-made climate information lead to better outcomes, in particular to higher agricultural yields, higher revenues and better food security?

Impact indicators

- Knowledge index constructed from a test on PICSA content
- Number of farmers who accessed climate information/seasonal forecasts
- Number of farmers who declare using climate information in their seasonal plans

- Number of farmers who diversified their livelihoods (i.e. who have started new activities other than subsistence agriculture)
- Maize yields
- Household revenues
- Index of food security

D. PLANS FOR EVALUATION

Methodology

The proposed methodology follows a mixed-method approach that combines quantitative and qualitative data analysis. The quantitative analysis will serve the impact evaluation in the identification of direct and indirect effects of the PICSA programme. Additional qualitative data will be collected in the form of key informant interviews and focus group discussions during baseline and every subsequent wave of data collection. The complementary qualitative analysis will provide a better insight into those outcomes that cannot be solely captured by quantitative data alone and will report issues and experiences from different perspectives.

Impact evaluation design

During discussions throughout the LORTA mission, we discovered that although the roll-out of PICSA in communities had yet not taken place, the AEDOs had already selected lead farmers for the PICSA trainings. In the selection of lead farmers, AEDOs mentioned that – among other criteria – they had considered gender balance, literacy levels, motivation and the ability to successfully reproduce the training for the contact farmers. As a result of such purposive selection, the proposed impact evaluation design focuses on a quasi-experimental approach.

The design uses a difference-in-differences (DiD) technique to estimate the programme impacts on beneficiaries by comparison with non-participants (control groups). The DiD technique enables the estimation of programme effects via the comparison of changes in outcomes and impacts over time between a treated and a control group. The counterfactual being estimated therefore corresponds to changes in outcomes/impacts for the control group. Because of this, treatment and control groups do not need to have the same pre-programme conditions. The estimation of programme effects via differences in outcome/impact changes makes it possible to remove any confounding fixed differences (e.g. better extension service, different historical economic characteristics) between treatment and control group.

In other words, the DiD technique relies on the crucial assumption that differences between the two groups – in the absence of the programme – are fixed over time (known as the "parallel trends" assumption). This is an untestable assumption; however, suggestive evidence can be provided on pre-programme trends on relevant outcomes/impacts between treatment and control group regions – conditional on the availability of such data.

Importantly, time-varying differences are not controlled for and, if present, would undermine the unbiased estimation of the programme effects. Examples of such uncontrolled differences are the construction of some infrastructure or the targeted implementation of some specific agricultural intervention in only one of the two group areas during the period of the programme.

The evaluation design exploits the fact that – in a given EPA of a PICSA district – the PICSA intervention is rolled out only in a given subset of sections. These coincide with the sections staffed by the selected AEDOs. This set up allows for the establishment of control groups of lead farmers and their respective contact farmers in non-treated sections of the same EPAs. Given that the treated

and control sections are located within the same district, we can assume that they would follow the same trend in the absence of the intervention.

Nevertheless, due to the participatory nature of the PICSA approach and the expected high interaction among farmers within and outside PICSA, a high degree of spillovers is assumed. For this reason, the establishment of a second control group in geographically separated districts is considered necessary. This additional control group would be sampled from those four districts that are only planned to be treated from 2020.

Besides providing unbiased estimates for the PICSA programme effect, this addition enables the measurement of spillover effects. More specifically, the comparison of changes between pre- and post-programme outcome levels of treated groups against control groups in non-treated districts will isolate the PICSA programme effect. The comparison of changes in pre- and post-programme outcome levels between control groups in treated districts and control groups in non-treated districts ("pure control" districts) will make it possible to quantify spillover effects of the PICSA programme.

If we let F_T denote treated farmers, F_{CT} denote control farmers in treated districts, F_{PC} denote lead farmers in pure control districts, then the PICSA programme effect on lead farmers is estimated with the following expression:

$$\delta = \Delta F_T - \Delta F_{PC}$$

The PICSA spillovers on lead farmers is estimated as:

$$\gamma = \Delta F_{CT} - \Delta F_{PC}$$

The six districts where PICSA will be rolled out are excluded from the impact evaluation design described above. For those districts, the LORTA team proposed that the DAES and NASFAM implementers consider a randomized selection of lead farmers. This would enable the set up of a one-year randomized control trial (RCT). This would require at a minimum a restricted baseline to check the balance of treatment and control areas. Endline data for the RCT would be collected in parallel to the endline wave planned for the DiD impact evaluation areas.

Sampling

PICSA targets two beneficiary groups: lead farmers and contact farmers. Within each target group, the sampling approach envisages one treatment and one control group in a given PICSA district (control-in-treated (CT)) and one control group in a given non-PICSA district (pure control (PC)). This requires the generation of six random samples for impact evaluation purposes.

The treatment sample of lead farmers within PICSA districts can be generated out of the list of selected lead farmers submitted by PICSA AEDOs. The treatment sample of contact farmers can be generated out of the list of contact farmers of each selected lead farmer.

The CT sample of lead farmers and contact farmers can follow the procedure above in case of non-exhaustive coverage of AEDOs sections before endline. Alternatively, the CT samples will be drawn from non-PICSA EPAs in PICSA districts. Finally, PC samples will have to be drawn from non-PICSA districts.

Given the absence of selected lead farmers in non-PICSA areas, the selection of lead farmers will have to precede the sampling. AEDOs in non-PICSA districts will be requested to replicate the same selection procedure as the one followed in the PICSA EPAs of the PICSA districts.

Power calculations

Power calculations were performed by the LORTA team in collaboration with Dr. Abidoye, the UNDP consultant on the M-CLIMES project.

Power calculations enable us to determine the minimum sample size needed in order to detect the impact of PICSA. To do that, we use the following power formula that relates the sample size to the minimum detectable effect size (MDES) between the mean outcomes of two groups:

$$MDES = (t_{1-\kappa} + t_{\alpha}) \sqrt{\frac{1}{P(1-P)}} \sqrt{1 + \rho(m-1)} \sqrt{\frac{\sigma^2}{N}} \sqrt{1 - R^2}$$

where $t_{1-\kappa}$ and t_{α} are t-statistics representing the required power and level of statistical significance (by convention, we seek a power of 80 per cent and a statistical significance of 95 per cent), P represents the proportion in one of the two compared groups (allocation ratio), ρ is the intra-cluster correlation (ICC), m is the number of individuals per cluster, σ^2 is the variance, N is the total sample size, and R^2 represents the extent to which baseline characteristics predict the endline yields.

We will use a cluster design for the impact evaluation of M-CLIMES. Since the PICSA training is delivered on a group basis, it is likely that there will be some similarities in outcomes between the members of one group. It is important to take this aspect into account in the power calculations. In our case, we consider a cluster to be a lead farmer together with his or her contact farmers.

The above formula allows for a comparison between two groups. However, given the context of our impact evaluation, we need to analyze three different groups:

- Farmers who receive PICSA training (treatment group; T)
- Farmers in Year 1 PICSA districts[7] who do not receive PICSA training and who are located in sections not covered by AEDOs trained for PICSA (control group in treated area; CT)
- Farmers in Year 3 PICSA districts[8] who were selected for PICSA trainings but who will be trained only after the end of our study (control group in control area; PC)

We will then be interested in two comparisons: *T* versus *PC* in order to measure the PICSA impact, and *PC* versus *CT* to estimate the spillovers of the programme. The sample size needed to achieve the desired power has to be determined separately for both comparisons, which means that we cannot consider the entire sample at the same time. The sample size will need to be higher compared to a standard situation with one treatment and one control group. Nevertheless, given the "centrifugal" nature of the PICSA programme, it would be ill-advised to omit the spillover analysis since it promises to be interesting.

Since the *PC* group appears in both comparisons, we should try to maximize the number of households in it. On the other hand, when sample sizes of the two groups are not proportionate – that is, when the ratio is not 50:50 – the power decreases. Therefore, we have to find an optimum size so that the two forces are balanced out.

Impact on yields

[7] Dedza, Chikwawa, Ntcheu, Rumphi.
[8] Lilongwe, Dowa, Mzimba, Phalombe.

Basic descriptive statistics have been obtained from previous studies, as follows:[9]

- The mean yield for maize is 1.8 t/ha.
- The standard deviation is 1.17 t/ha.
- An ICC of 15 per cent is assumed initially, based on similar studies in other countries on agriculture extension services.[10]

Based on different studies on the benefit of climate information, we expect that climate forecast will lead to an increased productivity and higher yields for the farmers in the targeted region. In a study of smallholder farmers in four villages in Zimbabwe, Patt, Suarez and Gwata observed that for farmers who participated in training on the uncertainty that surrounds climate forecasting, farmers who reported changing their management based on forecast information experienced a 19 per cent yield benefit in 2003/04, and a 9 per cent benefit averaged across years, relative to farmers who did not respond to forecast information.[11] Studies with extended interactions between farmers and institutions that provide EWS information have been shown to have reasonably high rates of use and benefits (Hansen, 2011) [12]. Roncoli et al. (2009) state that farmers reported higher yields based on participatory EWS information received and that they were better prepared for the planting season. [13]

Thus, we expect PICSA to have a yield impact of at least 10 per cent. In other words, at endline, we expect to observe a difference between the average yield of treated farmers and the average yield of control farmers to be at least 0.18 t/ha, which corresponds to a 10 per cent change from a baseline yield of 1.8 t/ha. A difference lower than 10 per cent is not economically significant, and therefore we are not interested in proving the statistical significance of such a small impact. We consider 10 per cent as our desired MDES. The standardized MDES is MDES expressed in terms of the number of standard deviations and is calculated below:

$$MDES\ standardized = \frac{MDES}{SD} = \frac{0.18}{0.17} = 0.15\ SD$$

[9] Komarek, A. M., Drogue, S., Chenoune, R., Hawkins, J., Msangi, S., Belhouchette, H., & Flichman, G. (2017). Agricultural household effects of fertilizer price changes for smallholder farmers in central Malawi. *Agricultural Systems, 154*, 168–178. https://doi.org/10.1016/j.agsy.2017.03.016

[10] BenYishay A., & Mushfiq Mobarak, A. (2018). Social learning and incentives for experimentation and communication. *Review of Economic Studies*, rdy039. https://doi.org/10.1093/restud/rdy0390. Also, baseline data from an unpublished C4ED project on Integrated Soil Fertility Management in Burkina Faso.

[11] Patt, A., Suarez, P., & Gwata, C. (2005). Effects of seasonal climate forecasts and participatory workshops among subsistence farmers in Zimbabwe. *PNAS 102*(35), 12623–12628. https://doi.org/10.1073/pnas.0506125102

[12] Hansen, J., Mason, S., Sun, L. & Tall, A. (2011). Review of seasonal climate forecasting for agriculture in Sub-Saharan Africa. *Experimental Agriculture 47(2)*, 205-240. https://doi.org/10.1017/S0014479710000876

[13] Roncoli, C., Jost, C., Kirshen, P., Sanon, M., Ingram, K., Woodin, M., Somé, L., Ouattara, F., Sanfo, B., Sia, C., Yaka, P. & Hoogenboom, G. (2009). From accessing to assessing forecasts: and end-to-end study of participatory climate forecast dissemination in Burkina Faso (West Africa). *Climatic Change 92*, 433. https://doi.org/10.1007/s10584-008-9445-6

Table 2 shows the results of power calculations for the optimal allocation ratio, 41:29:29. Tables for other allocation ratios are in appendix III of this design report. In Table 2, we assume different values for ICC and R^2. The indicator is always maize yield in t/ha, mean value is set to 1.8 t/ha and the standard deviation to 1.17 t/ha. In all cases we are trying to achieve the minimum detectable per cent change of 10 per cent.[14]

[14] Note that the power calculations will be identical for the two comparisons, *PC* versus *T* and *PC* versus *CT*, because all the parameters (sample size, mean, standard deviation) remain the same.

Table 2 Power calculations for ratio 41:29:29 (maize yields)

SAMPLE SIZE (TOTAL)	ALLOCATION RATIO	ICC	R2	FARMERS PER GROUP	# OF GROUPS IN PC	# OF GROUPS IN CT / T	MDES (IN T/HA)	% CHANGE
1,600	0.59	10%	30%	4	166	117	0.189	10.5%
1,800	0.59	10%	0%	4	186	132	0.213	11.8%
1,800	0.59	15%	30%	4	186	132	0.188	10.4%
2,500	0.59	15%	0%	4	259	183	0.191	10.6%

We argue that an overall sample size of 1,800 households is ideal. The 1,800 households should be split into 450 groups of four farmers each. It is better to expect a more conservative value of ICC – 15 per cent. On the other hand, it is reasonable to assume that we will be able to capture some of the variation in yields through baseline covariates (R^2 of 30 per cent). The second consideration is the optimal allocation ratio between the two groups that are being compared. In

Table 3, we show the ratio that allows for the lowest MDES (41:29:29).[15] Thus, 41 per cent of the total sample should be allocated to the *PC* group (186 groups), 29 per cent to the *CT* group (132 groups), and 29 per cent to the T group (132 groups). Splitting the sample equally between the three groups, as shown in Table 7 in appendix III of this design report, is not optimal because it gives us too few observations in each comparison, which weakens the power. Assigning 50 per cent of the sample size to the *PC*, as in Table 8 in appendix III of this design report, is also suboptimal; the disproportion decreases power too much.

The relationship between power and number of four-farmer clusters for the 50:50 allocation ratio is shown in Figure 3. The four curves represent different scenarios for two possible values of R^2 and for two possible values of ICC. The standardized MDES (δ) of 0.16 corresponds to an effect size of 0.191 t/ha.

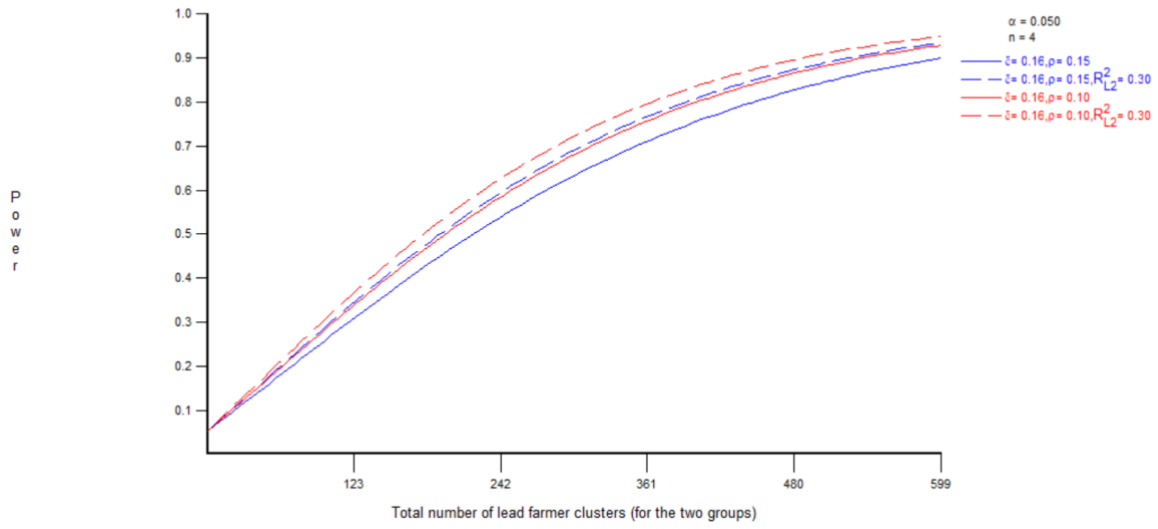


Figure 3 Power versus total number of clusters

[15] Allocation ratio of 0.59 is calculated for the comparison between two groups. This means that when comparing means of *C in C* and *C in T* groups, 59 per cent of the total number of observations would be in *C in C*, and 41 per cent would be in *T* or *C in T*.

Impact on use of climate information in adaptive behaviour

Let us consider another indicator: the use of climate information in adaptive behaviour. We calculate what MDES we can achieve with sample size determined previously using yield as an outcome indicator.

We use the following parameters:

- The mean proportion of farmers using climate information in adaptive behaviour is 6 per cent.
- The standard deviation is 0.24.
- An ICC of 0.5 is assumed, given that for farmers located in the same lead farmer group the likelihood of getting similar climate information, and using it collectively, is high.

Table 3 summarizes the results of power calculations for the indicator of "using climate information in adaptive behaviour" for two different values of ICC (25 per cent and 50 per cent) and two different sample sizes (1,800 and 1,600). The predictive power of baseline covariates (R^2) is assumed to be zero. We see that with ICC of 50 per cent the MDES is 0.061 – that is, 6.1 percentage points – which means that the proportion of farmers using climate information in adaptive behaviour will have to increase by 100 per cent. While an expected increase of 100 per cent might sound sizeable, let us note that the expected baseline level of the variable is rather low (6 per cent). A 100 per cent change means that 12 per cent of households observed at endline should declare using climate information in adapting behaviour in order to estimate a significant programme effect. We find this statistic a reasonable assumption.

Table 3 Power calculations for ratio 41:29:29 (using climate information in adaptive behaviour)

SAMPLE SIZE (TOTAL)	ALLOCATION RATIO	ICC	R2	FARMERS PER GROUP	# OF GROUPS IN PC	# OF GROUPS IN CT / T	MDES (IN T/HA)	% CHANGE
1,600	0.59	25%	0%	4	166	117	0.054	89.6%
1,600	0.59	50%	0%	4	186	117	0.064	107.0%
1,800	0.59	25%	0%	4	186	132	0.051	84.4%
1,800	0.59	50%	0%	4	259	132	0.061	100.9%

The relationship between power and number of clusters of size four for the 50:50 allocation ratio is shown in Figure 4. The two curves represent different scenarios for two possible values of two possible values of ICC. The standardized MDES (δ) of 0.25 corresponds to an effect size of 0.061.

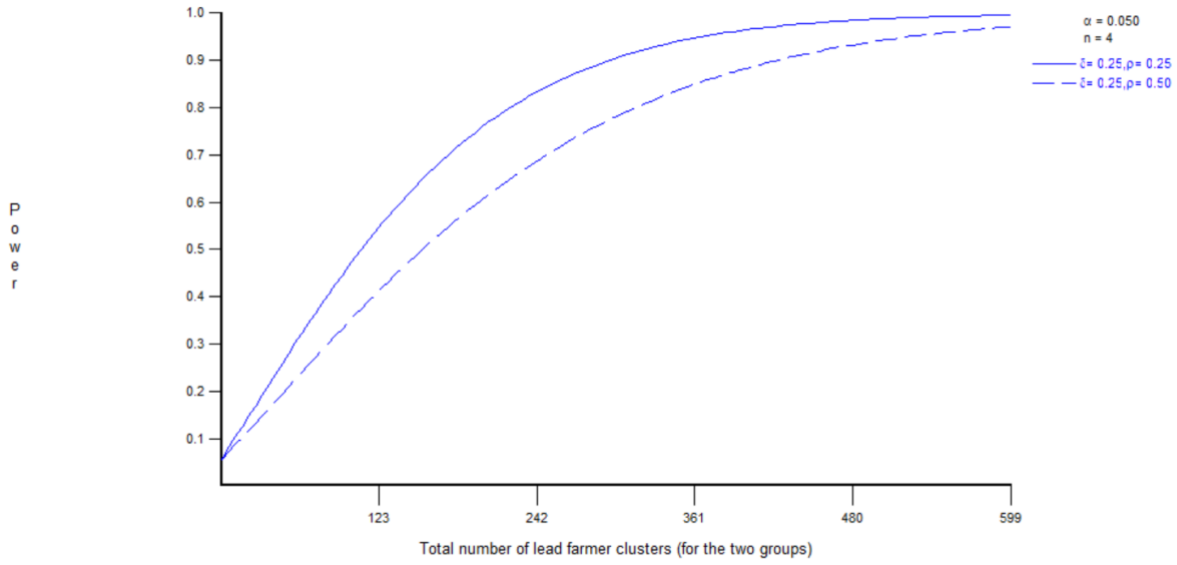


Figure 4 Power versus total number of clusters (using climate information in adaptive behaviour)

Possible risks for impact evaluation

A key limitation of the impact evaluation design is that the selection of lead farmers was determined before the LORTA mission. Although UNDP monitoring and evaluation (M&E) managers held off on the roll-out of PICSA training and further implementation instructions in response to the LORTA mission request in the Terms of Reference of the mission, a number of preparatory activities in the field including lead farmer selections had already taken place. This implies design limitations that can no longer be rectified at this point.

In particular, the purposive selection of lead farmers precludes the use of an experimental approach, which would provide a more rigorous estimation of the programme effects. The proposed impact evaluation design uses the quasi-experimental DiD technique that relies on the satisfaction of the parallel trend's assumption. Such an assumption is not testable, and suggestive evidence based on past trends requires the use of secondary data sources that are yet to be collated.

Furthermore, any time-varying factor differentially affecting treatment and control areas would undermine the unbiased estimation of the programme effects. The estimation of the programme effect is particularly vulnerable to the confounding effect of overlapping development or policy interventions as well as weather-related shocks that may differentially affect the areas.

Another concern relates to the budget for the evaluation. The LORTA team acknowledges that the M&E budget that was indicated in the project proposal for funding was cut down to about one third of its original amount. While the project plan envisions a midline and endline data collection, the M&E budget is currently only able to disburse the payment for the baseline data collection. This flags concerns that need to be clarified before ensuring the M-CLIMES project can enter further phases of the LORTA programme. In the favourable scenario that sufficient funds are procured, the LORTA team proposes to consider an additional RCT impact evaluation, as described earlier. The additional impact evaluation could be funded with those funds initially planned for a midline evaluation.

The last concern is that farmers within the PICSA treated districts will potentially be targeted by a second M-CLIMES intervention – that is, by the delivery of targeted climate-related agricultural advisories via SMS. The implementation strategy for this intervention is yet to be developed and

requires further preparation activities such as mapping of weather stations, further acquisition of weather data from additional private weather stations and discussions with implementers of similar previous pilots (e.g. WFP). Crucially, the implementation of additional interventions for PICSA beneficiaries needs to carefully consider the structure of the PICSA impact evaluation design in order to avoid undermining the identification of the PICSA programme effects.

Qualitative assessments

The qualitative assessments comprise focus group discussions (FGDs) and key informant interviews with stakeholders involved in EWS/climate information.

Importantly – unlike the quantitative survey – the qualitative assessments are planned to evaluate all three components of the M-CLIMES project.

The baseline data collection plans in-depth interviews to cover an exhaustive list of informants at the country, regional and community levels. Key informant institutions are those primarily involved in the implementation of M-CLIMES.

Other key informants are United Nations agencies and development partners who have implemented similar climate-related programmes. Among these organizations are the WFP GFCS, the European Community Humanitarian Office (Disaster Preparedness European Community Humanitarian Office programme), the Malawi Red Cross (Climate Change Project), the Malawi Vulnerability Assessment Committee and the World Bank (Integrated Flood Risk Management Plan). Key community informants will be area/village committees, extension officers, community leaders and representatives.

Finally, private sector stakeholders will also be key in assessing the interest and basis for collaboration on future M-CLIMES components. In particular, operators such as Airtel and Telekom Networks Malawi will be consulted regarding the delivery of climate-related advisories to farmers and fishers via SMS.

The aim of FGDs is to establish baseline information on disaster risk management, understanding and awareness of climate change, access to early warning information, and the use of climate information in order to make informed adaptation choices. FGDs are planned to be conducted with communities in selected areas in all 21 project districts. A larger number of FGDs are planned in the districts not covered by the quantitative surveys.

In those communities where FGDs are planned, the discussions need to take place after the quantitative survey is administered in order not to bias survey responses. A detailed plan for FGDs is yet to be developed by the survey company. Key aspects need to be taken into account, such as gender balance. The selection of the sites will include communities where there is some EWS/climate information equipment and communities with no equipment.

The LORTA team will offer their feedback once a more detailed structure for interviews and respective checklists and interview/discussion guidelines have been submitted by the survey company.

Data collection

The M-CLIMES project envisions baseline, midline and endline data collection. Before the LORTA mission, UNDP contracted a consultancy firm to run the baseline data collection for evaluation purposes.

Given the very tight implementation schedule of the PICSA interventions, the LORTA team discussed and coordinated to make sure that data collection could happen in coordination with the roll-out of PICSA without compromising either the implementation deadline or the quality of the data collected.

Given that the roll-out of lead farmers' training was set to start on the Monday following the end of the LORTA mission, it was agreed to divide the original survey questionnaire into two parts and administer them separately. The first questionnaire contains specific questions directly related to the content of the PICSA training and therefore is subject to a response bias if administered after the training. Such questions mainly relate to knowledge and use of climate information, seasonal preparatory activities and participatory practices. This short questionnaire is administered by the AEDOs to each of their respective lead farmer groups and is filled out on the first day of the training, before training has begun. A second, lengthier and more comprehensive questionnaire is administered to the sampled lead farmers and contact farmers by the consultancy firm.

The LORTA team has agreed to offer any further support needed in sampling and quality assurance activities in respect of the Terms of Reference between UNDP and the consultancy firm.

Timeline of evaluation

As discussed above, the M-CLIMES project is at a relatively more advanced stage than others in the LORTA project pool. At the time of the M-CLIMES selection into LORTA, UNDP had already hired a consultancy firm to conduct baseline data collection and scheduled the roll-out date for PICSA.

As can be seen from

Table 4, multiple activities – namely, desk review, evaluation strategy and design, implementation and data collection – were being performed in parallel during September 2018. The LORTA team closely coordinated with UNDP and the consultancy firm to ensure an effective implementation of programme and evaluation activities.

Table 4 Timeline of evaluation

		2018				2019										2020									
	Activity	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Dec
1	Implementation																								
2	Desk review																								
3	Evaluation strategy and design																								
4	Data Collection																								

As per the described evaluation design, the timeline depicts an endline data collection in 2020, before the PICSA roll-out in the pure control districts. Given the current budget issues, it is not clear whether and when a midline data collection can effectively take place as envisioned in the project proposal.

Importantly, the implementation of future additional interventions – such as climate advisories – in the study districts will require an adequate adjustment of evaluation strategy and might demand additional waves of data collection.

Assessment of institutional capacity for impact evaluation

In the given context, we believe that the LORTA workshop within the mission substantially benefited all stakeholders involved, including the UNDP PCU, the consultancy firm, DoDMA and other implementing governmental bodies, in bringing the needed formative engagement. The UNDP project team and the hired consultants were also very welcoming of the LORTA mission, even if it slightly delayed their tight implementation plan timelines.

The presence and support of Dr. Abidoye in the LORTA mission and overall programme has been crucial in facilitating coordination and understanding among the LORTA team and M-CLIMES implementers and stakeholders.

Overall, we consider the M-CLIMES project to be in need of support for impact evaluation. The LORTA team in coordination with Dr. Abidoye is continuing to provide input and feedback remotely.

Secondary data sets

The evaluation can benefit from the integration of primary data collected with a multiplicity of secondary data sources – both at the baseline and later stages.

One data set that is being consulted by the LORTA team is the fourth wave (2016–2017) of the Integrated Household Survey. This is a national survey conducted by the Government of Malawi to monitor and evaluate the changing conditions of Malawian households. In particular, this data set contains detailed information on farming and fishery activities. Other useful information may be retrieved from the Demographic and Health Survey (latest wave in 2015–2016).

Additional secondary data, such as soil quality, food insecurity and rainfall patterns, may benefit the evaluation in controlling for observed external factors in the econometric analysis.

Finally, monitoring may produce useful information. The M-CLIMES project has an M&E plan that will integrate some inputs from the LORTA team. This will be discussed further in the next subsection.

Plans for monitoring and evaluation

Because of the urgency to address implementation and evaluation design strategies, the plans for M&E were not discussed in detail during the LORTA mission.

The M-CLIMES project incorporated a broad M&E plan within the project proposal. This is currently being examined by the LORTA team. At the moment, only priority monitoring activities have been requested by the LORTA team – that is, the attendance list of lead farmers to the PICSA training sessions, which can inform on compliance and variation in treatment intensity.

III. WAY FORWARD

Overall, we consider that the LORTA mission in Malawi was well received and that it produced promising results.

The success of the LORTA mission has been particularly due to the attentive collaboration of the M-CLIMES UNDP project manager, Rabi Narayan Gaudo, and the M-CLIMES UNDP Knowledge Management M&E Specialist, Ted Nyekanyeka. Their input was crucial in arranging the meetings and making sure all the key informant and decision-making representatives would be consulted. Furthermore, Dr. Babatunde Abidoye was crucial in managing the operational and implementation discussions with the stakeholders as well as in the operational discussions with the consultancy team.

The LORTA team continues its support remotely. Currently, the team is assisting the consultancy firm in finalizing the sampling strategy and is coordinating with the M&E Specialist to collect priority information for the impact evaluation – specifically, farmers' lists, lists of lead farmers' selection criteria and monitoring lists.

However, the future success of this project in the LORTA framework is highly conditional upon a continuous responsive collaboration from the consultancy firm, UNDP and other key stakeholders

involved. Furthermore, the procurement of sufficient funds for the endline data collection at a minimum is essential.

APPENDIX I

Table 5 Agenda of Malawi LORTA field mission

DAY	PROPOSED ACTIVITIES	PARTICIPATION
Day 0: Sunday, September 9	Arrival of team	LORTA team
Day 1: Monday, September 10 Morning	Introductions, overview and discussion of objectives	NDA and DoDMA
Afternoon From 13:00	Presentation and discussions of the project and progress to date by the project team	Project leader and team members especially implementation staff
	Review proposed methodology and the technical proposal provided by the consultant firm selected for impact evaluation work and discuss survey methodology, sample size, etc.	Presentation by the consultant firm followed by discussion
Day 2: Tuesday, September 11 Capacity-Building workshop From 9:00.	Benefits of impact evaluation, and LORTA approaches	By LORTA team Participants were NDA, DoDMA, AE, project leader and team members, particularly the implementation teams, the consultancy firm and all other relevant stakeholders
Workshop on the theory of change	By LORTA team (interactive workshop and discussion)	By LORTA team Participants are NDA, DoDMA, AE, project leader and team members, particularly the implementation teams, the consultancy firm and all other relevant stakeholders
From 13:00	Discussion on implementation strategy to inform the evaluation design	By LORTA team Participants are NDA, DoDMA, AE, project leader and team members, particularly the implementation teams, the consultancy firm and all other relevant stakeholders
Day 3: Wednesday, September 12 Morning	Travel to Dedza on the way to Blantyre (PICSA farmer training field visit)/ Meeting with District Agricultural Officers	LORTA team/ DAES district and EPA officers
Afternoon	Travel to Blantyre for a meeting with DCCMS	LORTA team/ PCU
Day 4: Thursday, September 13 Morning	Meeting with DCCMS technical team (Including Agrometeorologist and PICSA leads)	LORTA team
Afternoon	Travel Back to Lilongwe	

DAY	PROPOSED ACTIVITIES	PARTICIPATION
Day 5: Friday, September 14 9:00	Debriefing: Presentation to the project team; key elements and timeline of design	NDA, AE representatives and other experts, Project leader and team members
From 10:00	Meeting with consultancy firm	LORTA team

APPENDIX II

Table 6 List of stakeholders engaged with during Malawi LORTA mission

LORTA WORKSHOP, LILONGWE		
NAME	POSITION	INSTITUTION
Frank Masankha	Farm Services Coordinator	NASFAM
Pacharo Msowoya	M&E Coordinator	NASFAM
Fannie Tasila Muwa	Farm Services Officer	NASFAM
Amos Ntonya	Chief Meteorologist	DCCMS
Piasi Kaunda	Hydrological Officer	DWR
Chimwemwe Njoloma	Information Officer	Min of Information
Geoffrey Chilombo	Agricultural Communication Officer	DAES
Timothy Cha	IEU	GCF
Yona Phiri	Principal Economist	DoDMA
Tereza Varejkova	Research Manager	C4ED
Dumisani Moyo	Consultant	CDM
Abel Shaba	Consultant	CDM
Annie Mapulanga	Economist	Min of Nat Resources and Env Mgt
Ted Nyekanyeka	Knowledge Management M&E Specialist	UNDP
Babatunde Abidoye	Senior Economist	UNDP
Rabi Narayan Gaudo	Project Coordinator-	UNDP
Hanna Siame	Environmental Officer	EAD (Environmental Affairs Dept)
Precious Mughogho	Senior Meteorologist	DCCMS
Mphanda Kabwazi	Disaster Risk Management Project Officer	DoDMA
Yobu Kachiwanda	Principal Meteorologist	DCCMS
Alexander Phiri	Consultant	CDM
Sam Gama	Principal Mitigation Officer	DoDMA
FIELD VISIT, DISTRICT OFFICE, DEDZA		
NAME	POSITION	INSTITUTION
Osmund Chapotoka	District Agricultural Development Officer	DAES
FIELD VISIT, BEMBEKE, EPA OFFICE		

NAME	POSITION	INSTITUTION
Gladmore Pumbwa	Assistant Veterinary Officer	DAES
Gift Faziri	Agricultural Extension Development Officer	DAES
Martha Phumbwa	Agricultural Extension Development Officer	DAES
Chifundo Chikudzu	Agricultural Extension Development Officer	DAES
FIELD VISIT, DCCMS, BLANTYRE		
NAME	POSITION	INSTITUTION
Amoss Mfonya	Chief Meteorologist	DCCMS
Hussein Milanzi	Principal Meteorologist	DCCMS
Keenness Manganda	Principal Meteorologist	DCCMS
Charles Vanya	Chief Meteorologist	DCCMS
Tasiana Mzozo	GFCS, Project Manager	WHO
Clement Boyce	Chief Meteorologist (Climate)	DCCMS
Adams Chavula	Chief Agrometeorologist	DCCMS

APPENDIX III

Table 7 Power calculations for ratio 33:33:33 (maize yields)

SAMPLE SIZE (TOTAL)	ALLOCATION RATIO	ICC	R2	FARMERS PER GROUP	# OF GROUPS IN PC	# OF GROUPS IN CT / T	MDES (IN T/HA)	% CHANGE
1,600	0.50	10%	30%	4	133	133	0.192	10.6%
1,800	0.50	10%	0%	4	150	150	0.216	12.0%
1,800	0.50	15%	30%	4	150	150	0.191	10.6%
2,500	0.50	15%	0%	4	208	208	0.194	10.8%

Table 8 Power calculations for ratio 50:25:25 (maize yields)

SAMPLE SIZE (TOTAL)	ALLOCATION RATIO	ICC	R2	FARMERS PER GROUP	# OF GROUPS IN PC	# OF GROUPS IN CT / T	MDES (IN T/HA)	% CHANGE
1,600	0.67	10%	30%	4	200	100	0.192	10.6%
1,800	0.67	10%	0%	4	225	113	0.216	12.0%
1,800	0.67	15%	30%	4	225	113	0.191	10.6%
2,500	0.67	15%	0%	4	313	156	0.194	10.8%

IMPACT EVALUATION DESIGN REPORT 2: MONGOLIA

TABLE OF CONTENTS

TABLES

FIGURES

I. INTRODUCTION

A. THE LORTA PROGRAMME

Evaluating the impact of development projects and programmes has gained importance in recent years. Impact evaluation not only allows for increased transparency by measuring the effects of investments, it also provides the opportunity to design and implement development projects more effectively. To contribute to this development, the Independent Evaluation Unit (IEU) of the Green Climate Fund (GCF) has started the Learning-Oriented Real-Time Impact Assessment (LORTA) programme to be able to keep track of GCF projects in terms of performance and results, and to enhance learning within the GCF.

The LORTA programme aims to

- Embed real-time impact evaluations into funded projects so GCF project task managers can quickly access accurate data on a project's quality of implementation and likelihood of impact, and
- Build capacity within projects to design high-quality data sets for overall impact measurement.

The purpose of the impact evaluations is to measure the change in GCF key result areas that can be attributed to project activities. The LORTA programme will inform on returns on GCF investments and help GCF projects track implementation fidelity. The objectives of LORTA include the following:

- Measuring the overall change (outcome or impact) of GCF-funded projects and enhancing learning
- Understanding and measuring results at different parts of theories of change
- Measuring the overall contribution of the GCF to catalysing a paradigm shift and achieving impacts at scale

Currently, the LORTA programme is in Phase I (formative engagement and design). In the first year (2018), the IEU will support eight GCF projects to build high-quality, theory-based impact evaluation designs at inception. Formative work will include engaging with project teams, accredited entities (AEs) and GCF staff; designing theory-based impact evaluations; and establishing protocols for database development.

The second phase of LORTA will involve the main impact assessment stage (3–5 years), while the third phase will include analyzing baseline and endline data, discussing results, and engaging with diverse stakeholders to share results and incorporate feedback as required.

B. THE MONGOLIA PROJECT

The project "Business Loan Programme for GHG Emission Reduction", often referred to as the "Eco-loan programme", run by XacBank in Mongolia has been selected to be part of the LORTA inception stage (Phase I).

The Eco-loan programme can be applied for at any XacBank branch in Mongolia. In 2017 there were 82[16] retail branches nationwide with 38 located in the capital, Ulaanbaatar. All loan officers have been trained on the content and requirements of the Eco-loan product, meaning that micro, small and medium enterprise (MSME) owners can apply independent of their location within Mongolia. MSMEs make up more than 90 per cent of businesses in Mongolia, many of which

[16] XacBank Annual Report 2017.

continue to use old and inefficient equipment. This contributes to the high level of greenhouse gas (GHG) emissions and high energy needs for the outputs produced. Naturally, MSMEs are concentrated in the capital Ulaanbaatar, where roughly half of the population of Mongolia lives. In reality, this means that most of the loans are disbursed in the country's only truly urban setting.

The Eco-loan programme is run by the Eco Banking Department (EBD) of XacBank, a private sector commercial bank that is a subsidiary of the TenGer Financial Group (TFG). TenGer is majority owned by international investors, including the International Finance Corporation of the World Bank and the European Bank for Reconstruction and Development (EBRD); it is therefore held to their standards of governance and transparency. The EBD's focus is to provide financing for sustainable energy, carbon finance, and implementing projects and programmes to improve energy efficiency (EE). The Eco-loan programme fits into the remit of the EBD and had funding approved by the GCF to begin implementation in August 2017. The GCF funding is divided into two parts: a USD 19.5 million loan used to finance the Eco-loans and a USD 500,000 grant used for capacity-building and awareness-raising about the programme. How each of these might affect the outcomes of interest is explained in more detail in the theory of change in section II.C.

The GCF funds supplemented an existing programme that has been running since 2013, when the Global Climate Partnership Fund committed USD 20 million. More funding was made available in 2014, when the EBRD committed USD 15 million under the Mongolian Sustainable Energy Financing Facility programme. Each donor has had slightly different conditions regarding what the funding can be used for, with the GCF requiring the funds to be used for MSMEs. The funds provided by the GCF also enable XacBank to lower the interest rate offered, meaning that the interest rate is now significantly below standard market rates for business loans. The first loan to an MSME from GCF funds was disbursed in November 2017.

The two differently funded parts of the programme both have the goal of reaching as many MSMEs that seek to borrow money to invest in GHG-reducing capital/machinery or processes as possible. The majority of the funds are used to subsidize loans, which are made available to eligible MSME owners. These loans are offered at interest rates varying between 13.8 per cent and 15 per cent. To give some context, a standard business loan from XacBank has an interest rate between 19.2 per cent and 24 per cent. To be eligible, the loan must be used to invest in capital or processes that will save 15–20 per cent of CO_2 emissions or improve EE, whether through the purchase of new capital/equipment, the retrofitting of existing equipment or a change in processes employed by the firm. Those investing in CO_2-reducing or energy-saving products to be sold must make a 20 per cent saving, whereas businesses that are end users of CO_2-saving or energy-saving equipment are required to make a 15 per cent saving.

To be eligible for the Eco-loan programme, an MSME must fulfil the CO_2/energy-saving requirements as well as pass XacBank's usual risk assessment, used for all applicants to their business loans. The proposed investment is assessed by engineers working with XacBank to calculate the level of CO_2/energy reduction forecast according to the current usage and that of the equipment/process proposed in the loan application. To calculate the reductions that will be made, this "pre-evaluation" uses official documents, such as ISO or Mongolian National Standards certificates, on the exact equipment currently in use as well as those to be bought under the proposed loan. If such documents do not exist, laboratory tests will be undertaken. If the proposal meets the 15–20 per cent requirement, the costs of this assessment are covered by XacBank; if a proposal falls just short, it can be amended to make further energy savings. If the GHG requirement is met, then XacBank undertakes a credit risk assessment, which includes any potential damage to the environment or negative spillovers to people or firms in their vicinity. If all requirements are met, then the loan may be disbursed.

The borrower is monitored after the loan has been disbursed to check that the loan was spent on what was proposed and to check that the CO_2/energy use reductions were sufficient to receive the Eco-loan. Documentation with proof of purchase must be provided, and XacBank may visit the MSME.

In the first year of the project, there were 110 loan requests with 84 loans being disbursed by the EBD. The rather small number of beneficiaries reached each year poses the biggest challenge in designing an impact evaluation. The majority of those who did not receive a loan were applications that were inappropriate in the context of the Eco-loan programme because they could not expect to make substantial CO_2 savings.

II. MONGOLIA FIELD MISSION

A. GENERAL REMARKS

An evaluation team, consisting of two consultants from the Center for Evaluation and Development (C4ED), was formed to lead the field mission from 10 to 14 September 2018. The task of the evaluation team (henceforth called the LORTA team) was to engage closely with key stakeholders of the project – namely, the national designated authority (NDA), AE, implementing agencies, project staff and potential end beneficiaries – to ensure their interest in and understanding and sense of ownership of the planned theory-based impact evaluation.

During the field mission, the LORTA team held meetings and capacity-building workshops with the key stakeholders. Meetings with the EBD and compliance team of XacBank were held to acquire the maximum possible information about the GCF-funded project. Borrowers were interviewed during field visits regarding their views about the project's implementation and communication strategy as well as the impacts they experienced. Discussions about the implementation and challenges in evaluation were also held with the EBD team. In addition, a capacity-building workshop on impact evaluation was held, targeted at the key stakeholders. Beside conveying technical knowledge, the aim of this workshop was to emphasize the benefit of theory-based counterfactual approaches and real-time learning and measurement.

Under the guidance of the LORTA team, two potential evaluation designs were discussed with XacBank. The LORTA team conducted context analyses, examined the existence of appropriate counterfactuals (i.e. comparable treatment and control groups), elaborated a theory of change, assessed the availability of baseline administrative and secondary data sources, and acquired budget information. The results from this undertaking are presented in the following sections.

B. THE MISSION AGENDA

The LORTA team experienced a high degree of collaboration from XacBank, who went to great efforts to provide a smooth and productive week's work. The agenda – shown in Table 1 and Table 2 in Appendix I of this design report – was developed to facilitate the joint attendance of the key stakeholders for LORTA workshops and to allow for field visits. Unfortunately, the NDA was unable to attend meetings at the bank due to concurrent workshops held as part of the GCF readiness programme.

After discussing the Eco-loan programme with the project team on Monday morning, we developed a theory of change, enabling us to discuss some potential evaluation questions. This workshop on the theory of change was key to understanding the programme fully. On Tuesday, the introduction to an impact evaluation workshop was held before a presentation of the project by the EBD. On Tuesday afternoon we had time to hold the impact evaluation methods workshop, which

unfortunately could only be attended by the EBD team from XacBank. The employees of XacBank who were present appeared to benefit greatly from the impact evaluation workshops and gained a better understanding of the need for a counterfactual. They were engaged and interested in our work and the plans of the LORTA programme.

Wednesday's visits to three different borrowers gave us a clearer understanding of the eligibility for the Eco-loan as well as the potential impact it has. How these three visits impacted our views on a potential evaluation are described in more detail in the plans for evaluation. Thursday was used to present an introduction to impact evaluation to those attending the GCF readiness workshop, where many of the relevant stakeholders for the Eco-loan project were present, along with potential future GCF partners. Friday included a debrief with the head of the EBD.

During the mission, we established a strong understanding with the EBD team, the head of which spoke supportively of impact evaluations following our presentation at the readiness workshop. The members of the EBD who were present and the other key stakeholders met during the week are listed in appendix II of this design report.

C. RESULTS

Theory of change

The theory of change (ToC) developed with the EBD team considered how both the GCF loan and the GCF grant could help achieve the goals set out by the EBD. Both parts of the programme are implemented with the ultimate goal of reducing CO_2 emissions to protect the environment, but given the private sector nature of XacBank, it is also the goal of the bank to turn a profit, as is reflected in their stated focus on "Planet, People, Profit". Considering the steps discussed in the ToC workshop, we feel the following ToC describes the programme well.

i) Inputs

The grant and the loan from the GCF are used for separate activities and thus are presented in the following figure as leading to the goals along differing paths.

ii) Activities

The grant is used for activities that can be classified under capacity-building and awareness-raising. Capacity-building takes the form of training of the bank's own staff and accountancy 101 for MSMEs taking out business loans. Through this training, which was offered to all branches nationwide, loan officers understand the programme and will be more likely to offer the loan to appropriate businesses seeking a loan. The awareness-raising includes small MSME training sessions to which owners are invited to inform them of the opportunities of the Eco-loan. The EBD staff claimed that few of those attending these small trainings took out a loan. Each year, there is also a larger forum, the Green Finance Forum, where firms are not only informed of the Eco-loan programme but also instructed on the benefits of EE and renewable energy (RE) by the Energy Regulation Commission. The forum attracts 400 attendees, many of whom choose to take the loan afterwards.

The GCF loan is used to finance concessional loans, with interest rates lower than the standard market rates. These loans can be received under the conditions explained in section I.B.

iii) Outputs

Due to the activities in the grant, knowledge of EE and RE business practices should grow, as should the knowledge of the Eco-loan programme of XacBank. This knowledge along with the loans offered at a lower interest rate will enable more MSME owners to take out loans.

iv) Outcomes

By taking out an Eco-loan, firms will invest in more EE and RE equipment/processes.

v) Goals

By investing in more energy-efficient equipment and processes, firms will reduce their emissions and thereby the damage they cause to the environment. This has a wide range of consequences within Mongolia, such as improving air quality in one of the world's most polluted cities. Additionally, it might help the market for EE and RE to grow, reducing the need for concessional loans as prices fall due to the scaling up of the production of EE equipment. This could occur due to information spillovers from MSME owners who invest in such products.

It should also be noted that by reducing energy usage MSMEs will reduce their marginal costs and so might be able to grow due to an improved profit margin. The other goal of the programme, that at least 50 per cent of the beneficiaries are women-led MSMEs, is likely to be achieved more easily than in other countries due to the importance of women in the Mongolian labour market.

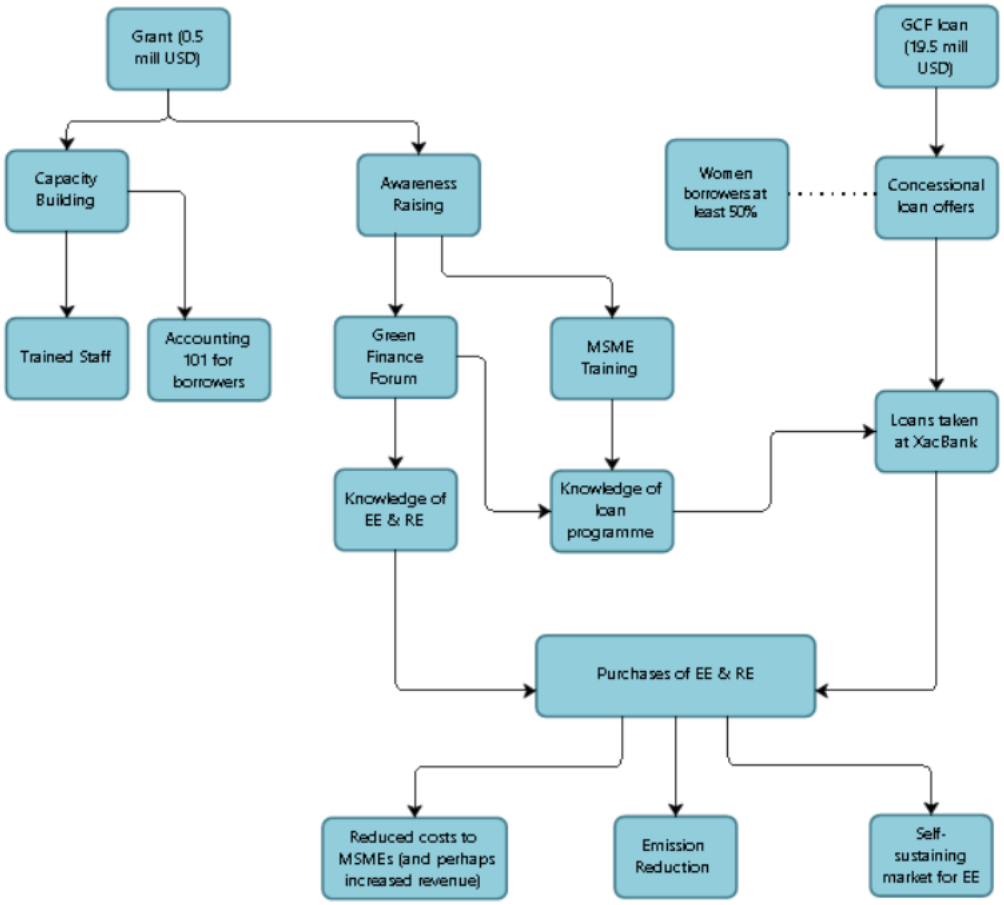


Figure 1 Theory of change

Evaluation questions

- How much were CO_2 emissions reduced/energy savings made?
- What is the number of loans taken out due to training?
- How does training impact knowledge of and attitudes towards EE & RE?
- Do MSMEs plan to invest in EE & RE in the future?

- What is the highest interest rate MSMEs are willing to pay to invest in GHG-reducing technology?
- Do borrowers change investment plans to access programme funds?
- Does the programme lead to cost reductions and/or increased revenue for borrowers?

Impact indicators

- CO_2 emissions
- Number of loans
- A series of questions about EE and RE, which can be used to construct an index. Similar questions could be asked of attitudes towards EE & RE.
- Self-reported investment plans
- Self-reported highest willingness to pay WTP as an interest rate
- Self-reported response to investment plans before and after the forum
- Energy costs (self-reported), revenue earned (self-reported)

D. PLANS FOR EVALUATION

When considering what is often held up as the gold standard of impact evaluations, the randomized control trial (RCT), it was apparent that this approach would not be suitable in this setting, as already clarified during the Bangkok workshop. To receive a loan MSMEs must apply, which already entails a degree of self-selection into the programme. All eligible loan applicants receive the loan if they pass the risk assessment. The programme has already rolled out nationwide and has the funds to cover more loan applicants than it currently deals with. It would therefore be unethical to withhold credit from some MSMEs who are eligible and seeking loans. While the monitoring and evaluation (M&E) run by XacBank can measure how much emissions have been reduced compared to before receiving the loan, this does not necessarily measure the impact of the loan. Some MSMEs might have made these investments even without the Eco-loan's low interest rate, simply paying a higher interest rate for an alternative loan in order to pay for the investment and thus not making any reduction in GHGs due to the programme. Another possibility is that firms increase their investment plans a little due to the lower interest rates (or perhaps even increase their GHG emissions to reach the threshold) and so not all of their GHG reduction is causally due to the programme. Being able to measure the causal difference in GHG reduction is of interest. It would be useful to rule out the possible reason that only firms who would invest anyway are now profiting from the Eco-loan programme, which thereby just raises the profits of those already aware of the benefits of EE and RE. While this is highly unlikely to explain all of the borrowing, it would be helpful to find the extent to which reductions in GHGs are increased due to the programme.

The visits to three borrowers were particularly instructive in this respect. The first visit, to a seller of solar panels and related products to be powered from this decentralized energy solution, was instructive in that it became clear that the lower interest rates available made it possible to expand RE-related business. The owner stated that the difference in interest rates were her profit margin. The second MSME visit was to the owner of a company providing heat to other buildings from large heat-only boilers. This borrower had in fact benefited from other loans in expanding their business and was highly interested in improving efficiency, both to decrease costs and for the environmental benefits. Both here and at the final MSME – a school in the process of building an extension with energy-efficient insulation and solar panels – it seems that the borrowers would have wanted to invest in EE products even without the loan, but the loan increased the extent to which this was possible. This left the impression that the impact of the loan might not be as high as the measured

reductions in GHGs that are stated in the programme's own M&E. For this reason, finding a good counterfactual is particularly important to evaluating the true impact of the Eco-loan programme.

Methodology

During the Bangkok workshop, it was proposed that a regression discontinuity design (RDD) could potentially be applied using the cut-off of 15 per cent or 20 per cent for required emission reductions, comparing those businesses who pass the requirement with those who marginally fail it. For this strategy to deliver robust results, applicants should not be able to manipulate their score after it has been announced. This would mean that there should be a continuity of the density in the number of firms on either side of the cut-off. This would be unlikely to be the case for the Eco-loan product for two reasons. First, firms who know that they are unlikely to fulfil the reductions requirement would be unlikely to apply for a loan, given they can assess what their score will be and know the cut-off in advance. This means that the density of firms below the cut-off would be lower than above. Second, firms who apply for a loan and are assessed as likely to make a saving that falls just short of the cut-off are encouraged to amend their application so that they reach the reductions cut-off. This manipulation, in terms of a score for RDD, is not merely possible in the system but actively encouraged in order to help firms achieve greater CO_2 emissions or energy use reductions. Both these reasons make the use of RDD infeasible with the proposed reduction used as a score.[17]

We therefore considered several possible options to evaluate the Eco-loan programme, which vary in their degree of causal interpretation. Following discussions with the EBD of XacBank, two alternatives were proposed during our visit to Ulaanbaatar (1 and 2), with an additional option (3) discussed with the IEU after returning from the mission:

1. Randomized encouragement design based around the awareness-raising pathway in the ToC
2. Before–after: A learning exercise considering the possible effects of the awareness-raising (not an impact evaluation in the true sense)
3. Matching on observables – propensity score matching

Each of these will be described and explained in turn in the following subsections.

Randomized encouragement design

Given that a traditional RCT approach cannot ethically be implemented in the Eco-loan programme, an alternative approach was considered. In the case of the Eco-loan programme, an awareness campaign is carried out. This targets MSME owners, who might thereby be encouraged to take out a loan from XacBank. If exogenous variation in exposure to this awareness campaign can be generated, and this in turn sufficiently affects the likelihood of taking out a loan, this can be used to estimate the effect of taking out a loan on the outcomes of interest.

Each year, a large event is held, referred to as the Green Finance Forum, to which MSME owners are invited if they belong to a specified sector. During this event, the EBD explains the Eco-loan programme and the Energy Regulation Commission presents EE and RE issues relevant to the sector that is the focus of the forum. This event is publicised through social media and in XacBank branches, with places allocated to MSME representatives first, before other interested parties, such as local academics, receive places. These events have a capacity of 400, and typically some applicants are rejected (though these tend to be non-MSME representatives).

The local branch credit officers identify suitable MSMEs for whom the event could be of interest and invite them specifically. In past years, the branch officers have identified enough participants,

[17] It should be noted that this method was perhaps communicated with too much confidence to the XacBank team following the Bangkok workshop.

such that the forum was full but not oversubscribed in terms of MSME owners. We suggested that the forum could be held for the same sector in two consecutive years and enough candidates could be selected for both years from the start. It would be possible to randomize in which year each eligible candidate is invited.

Estimation

The main outcome of interest to the programme is the reduction in CO_2 caused by the programme. In the following example we use this outcome for illustrative purposes, but the same ideas hold for any potential impact of receiving a loan. Ideally, we would compare the CO_2 emissions of a firm receiving the Eco-loan ($CO2_{i1}$) with the same firm without the Eco-loan ($CO2_{i0}$).

$$Impact_i = CO2_{i1} - CO2_{i0}$$

Of course, this cannot be observed, and thus we need to compare the average emissions of those with the loan to those without. An ordinary least squares regression to estimate the correlation of loans and the change in CO_2 emissions ($\Delta CO2_i$) could appear as in the following equation, in which we are interested in β.

$$\Delta CO2_i = \alpha + \beta Loan_i + \varepsilon_i$$

However, such an estimation will suffer from potential bias – in particular, that firms choosing to take out an Eco-loan might have invested in CO_2 reductions anyway, while those without a loan might not have invested even if given a loan. In order to estimate the true effect β of the loan on the reduction in emissions, we need to impose an exogenous source of variation in MSMEs receiving Eco-loans. We therefore propose the use of instrumental variables (IV) estimation, instrumenting the participation in the Eco-loan programme with the random selection of being invited to the forum. By introducing a first stage in which MSMEs are either invited to the Green Finance Forum or not, exogenous variation in loan take-up can be generated. Here the predicted value for taking out a loan ($\widehat{Loan_\iota}$) is used in the second stage equation.

$$Loan_i = a + b\ Invitation_i + u_i$$

$$\Delta CO2_i = \alpha + \beta \widehat{Loan_\iota} + \varepsilon_i$$

This estimation would enable us to make a causal estimate for the effect of a loan on the outcomes of interest if the first stage is sufficiently strong. This means that the invitation would have to have a strong enough effect on taking out a loan. In the above set up, we would use a reduced form where an invitation is assumed to have a similar effect on taking out a loan as attending, and the variation in the loan is derived from the randomized invitation. This reduced form is likely to be more efficient than a three-stage estimation.

This poses an additional question of whether the attendance has a causal effect on taking out a loan. This can also be considered in a two-stage framework: first predicting attendance with the invitation dummy, and then using the predicted attendance to estimate the impact on whether an MSME takes out a loan. This could also be used to estimate the impact of the forum on EE and RE knowledge and views thereof.

$$Attendance_i = a + b\ Invitation_i + u_i$$

$$Loan_i = \alpha + \beta \widehat{Attendance}_\iota + \varepsilon_\iota$$

Note that if the forum does not have a strong impact on loan take-up, the estimation of the impact of the loan on the final outcomes would suffer from a weak instrument and so would not give robust results for the causal impact of receiving a loan.

Sampling – randomisation procedure

To carry out this randomized encouragement design, various steps need to be carried out (see Figure 2). The group of MSMEs of interest must be identified by XacBank from among the body of all MSMEs. Of those in the target group, it must be decided how many can be interviewed as part of the evaluation sample. This number will depend on the power calculations. Of those in the evaluation sample, we would randomize which MSMEs are invited to the next forum and which would be invited the following year.

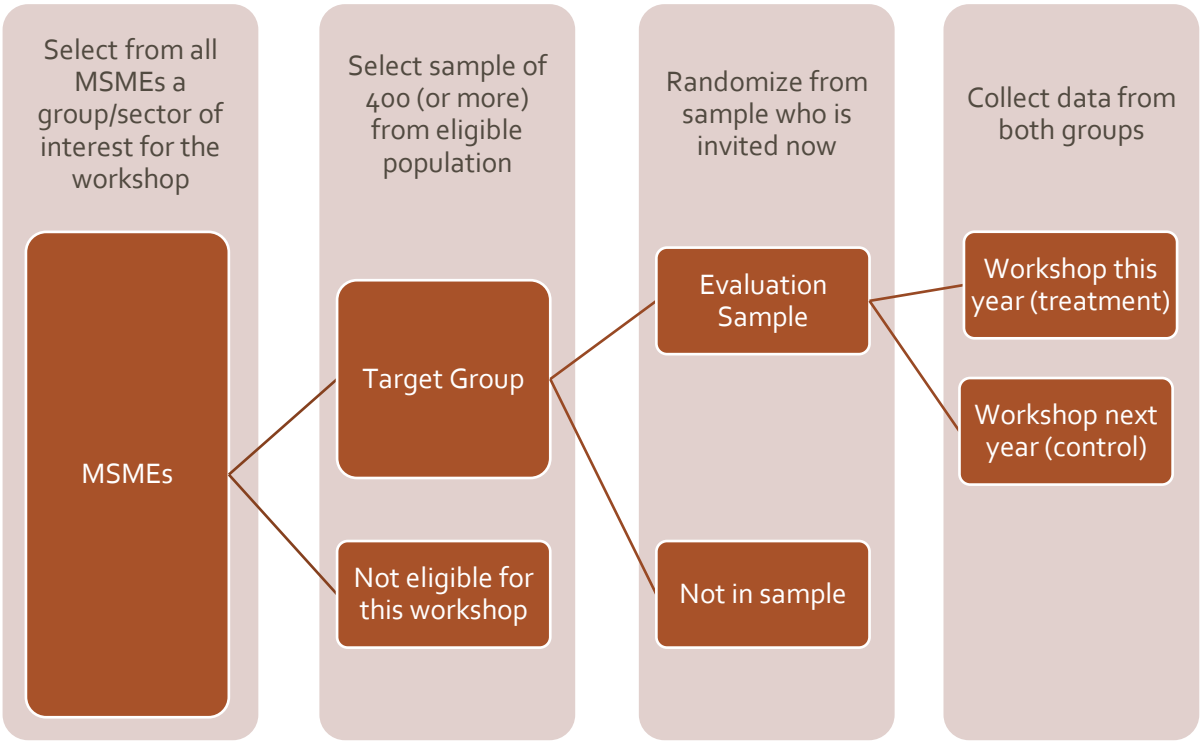


Figure 2 Steps in the randomized encouragement design

Drawbacks and concerns

This method would have the potential for causal interpretation, but would only be relevant for the sector targeted for a particular year's forum. While the results may be generalizable, it is unclear whether the results would be externally valid for other sectors in Mongolia, let alone firms outside of Mongolia. We would also rely on XacBank providing a sample frame from which to randomize, which relies on their knowledge of the pool of MSMEs.

Finding causal effects of the loan also relies on the forum having a strong enough effect on the probability of an MSME taking out a loan and would affect the sample size required (to be calculated in power calculations). Anecdotally, the EBD stated that following the forum they observe a significant proportion of those attending taking out a loan. Not all invited attend, and so this also leads to a concern of self-selection into attendance and thereby a problem of compliance. If enough firms do not comply, and this non-compliance is systematically correlated with the probability to take out a loan ex ante, this would be problematic. In this case, the reduced form would maintain a source of exogenous variation, but effects might be small and insignificant.

This option was viewed as infeasible by XacBank, who felt that they could not change their existing plans nor effectively randomize who is invited without damaging performance of their targeted awareness-raising. Usually, a new sector is selected for the forum each year, meaning that those invited cannot be easily substituted from one year to the next.

Before–after design (with difference-in-differences)

This methodology is the most conservative of the three approaches considered and does not represent an impact evaluation in the true sense, and therefore should be not be named an impact evaluation. As in proposal 1, we would look to exploit differences between those attending the Green Finance Forum and those who did not attend. In this proposal, there is no exogenous variation in who is invited and attends; of those invited, attendance reflects self-selection into the treatment. This evaluation would use data from all those who were invited to the forum, comparing those who attended with those who did not. The questions that can be answered with this technique are considerably more restricted. With data collected prior to the forum and six months to a year afterwards, it may be possible to attempt a difference-in-differences (DiD) estimation on certain aspects, such as knowledge of EE.

First, we can observe which invited MSME owners choose to attend and which do not, using data from the time of the invitation. Here we could consider firm and owner characteristics and investigate how these are correlated with the decision to attend.

For knowledge, we would be more likely to use a DiD design to calculate how knowledge changes with attendance at the forum. Those choosing to attend may have better knowledge prior to the forum, so any difference at the end could not be attributed solely to the forum itself. For this reason, we would subtract the initial difference from the final difference in knowledge.

This requires the standard assumption of parallel trends – that is, that attendees and those choosing not to come would have changed in the same way in the absence of the forum (Figure 3). This assumption may not be true if those who accept the invitation were interested in EE and RE anyway, and so may have informed themselves about EE and RE even without attending the forum. Under a setting where there is no exogenous variation in attendance (e.g. randomising invitations), this may be the best we can do if we rely on XacBank to select the sample from their own knowledge. Finding a more suitable counterfactual would require further data, from which we could potentially match those attending with other similar firms.

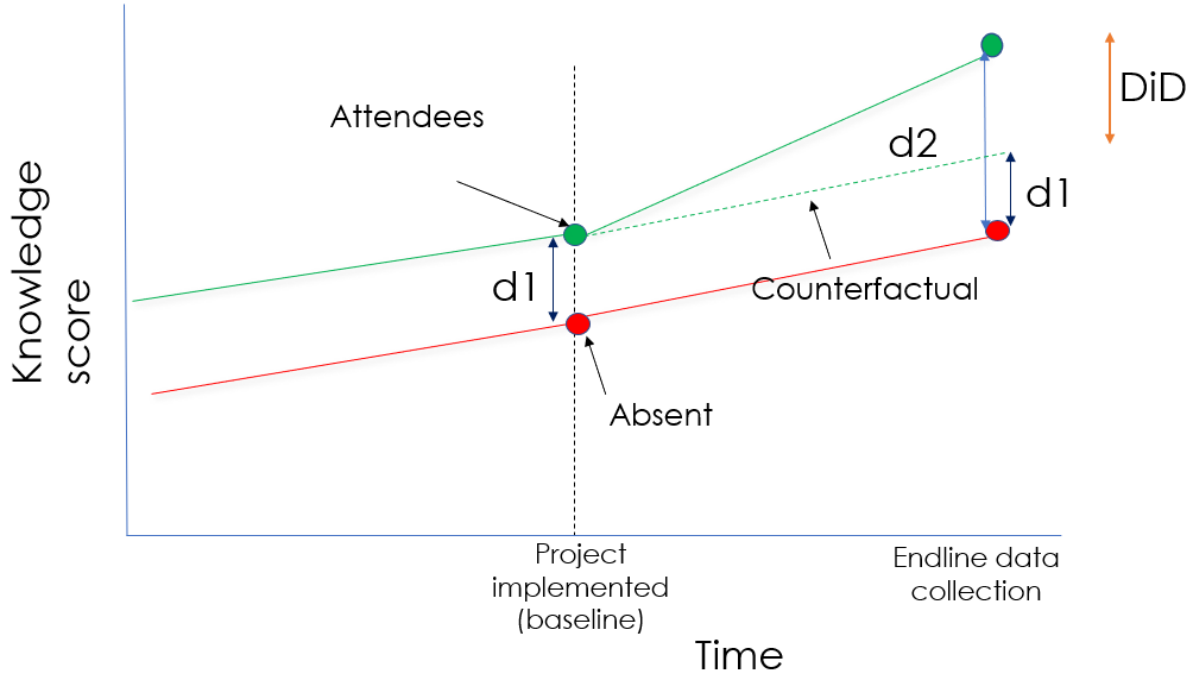


Figure 3 The parallel trends assumption

Drawbacks and concerns

Any method lacking a convincing counterfactual can not be considered an impact evaluation. It seems unreasonable to assume that there would be a parallel trend between those who choose to attend and those who do not. The results may be helpful to understand the differences between these two groups, but no causal interpretation of the differences could be made. This method would rely on enough MSME owners turning down the opportunity to attend the forum. If the majority of those invited indeed attend the forum, the comparison group would likely be too small.

Matching design

This approach would theoretically allow for an estimation of the impact of the Eco-loan programme under some assumptions. By collecting data from both Eco-loan borrowers and other MSMEs, a matching exercise using propensity score matching (PSM) could be attempted. Usually, the firms taking out an Eco-loan have already had some contact with XacBank or are in a network with MSME owners who have. If data could be found from which we could contact MSME owners who are not in the XacBank customer database, we could survey those applying for a loan (or those who attended the Green Finance Forum) as well as other MSMEs who do not. By surveying more firms at baseline who were not among the "treated", we could attempt to match the firms on observable characteristics from a baseline survey using PSM.

In 2014, with EBRD support, the Ministry of Labour in Mongolia carried out a survey of 1,500 MSMEs across the country. These data are managed by the Ministry of Food, Agriculture and Light Industry, whom we have contacted to ask about the sampling strategy used for the survey. In order to identify the MSMEs surveyed, there must exist a list of registered MSMEs in the country. It might be possible to ask for access to the data or at least a sample to be used for research purposes. In this case, we may be able to pre-identify MSMEs in the same sectors as those applying for Eco-loans at XacBank, thus increasing the likelihood that the firms surveyed are similar and thereby comparable. After finding MSMEs that are well matched at baseline, we could then compare the outcomes of interest at endline. If the MSMEs matched were not well balanced in the outcomes of interest at baseline, we could attempt to control for this difference by using a DiD estimation.

For this approach to work, we must first assume that there is common support on the variables we would use for matching. This would mean that there are MSMEs that are similar to those who borrow from XacBank but for whom XacBank has no information. It is not clear how comprehensive XacBank's database of MSMEs near to each branch is. To try to find a good match between borrowers and other MSMEs, a large sample would need to be surveyed at baseline. If there is a strong degree of common support, the sample size for the control group need not be so much larger. But if very little is known about the control group sampled at baseline, a larger multiple will be needed for the sample size. Typically, sample sizes required for the control group under PSM range from 2 to 10 times the size of the treated population.

Drawbacks and concerns

In Mongolia there are roughly 60,000 MSMEs,[18] of which 22,000 have conducted business with XacBank in the past. This suggests that while XacBank's knowledge of local businesses is strong, it is not fully representative of MSMEs in the country. It is unclear whether there are systematic differences between XacBank customers and non-customers. Non-customers could simply have chosen a different bank at normal business loan interest rates but are otherwise similar, or they

[18] As per the project proposal, see https://www.greenclimate.fund/documents/20182/574760/Funding_proposal_-_FP028_-_XacBank_-_Mongolia.pdf/61383ce5-ad1a-44a4-ba19-8007965c3adf

might not have sought bank loans and so be systematically different. This issue is key to the possible success of using matching for an impact evaluation.

Data collection

For all of the above options, baseline and endline data collection would be necessary. Given that the impacts of new investment can be measured immediately after the funds borrowed have been spent on new machinery, the endline does not need to be a long time after programme "treatment" has been implemented. Depending on the evaluation option chosen, the time lag would likely differ. Under option 2, six months would be more than sufficient to see the impacts of the forum, and data could even be collected in the weeks following the forum. For options 1 and 3 in which we also wish to evaluate the impact of the loans, the endline data would need to be collected later. One challenge here is that loans can be applied for and granted throughout the year, so the timing of the "treatment" would vary. In option 1, the endline data could be collected shortly before the second Green Finance Forum. Under option 3, a suitable length of time would need to be defined to see the impact of the Eco-loan programme compared to the control group. Data could also be collected one year later in this case.

XacBank proposed using their own agents to conduct telephone interviews with those in the sample. If the sample frame is taken from XacBank's own records, then this would lighten the load of data collection and reduce costs substantially. However, it raises the question of impartiality and trust. Employees of XacBank may have an interest in manipulating the results so that the programme appears more effective in order to receive future funding. In this case, they may systematically report better results for those receiving the Eco-loan compared to those who do not take an Eco-loan, thus biasing the results. It is also unclear whether the respondents would react differently to employees of the bank compared to a third party hired for data collection.

It would therefore be preferable to use a third party to collect data from the MSME owners if we wish to ensure high-quality data that are less subject to outside criticism.

Timeline of evaluation

The timing of the evaluation would depend on the option chosen. Beginning shortly prior to the Green Finance Forum (usually held around September) would represent a good time to begin the baseline data collection.

If we are only interested in the impact of the awareness-raising campaign funded by the GCF grant, endline data could be collected sooner than if we are also able to evaluate the impact of the loan programme as a whole.

Assessment of institutional capacity for impact evaluation

Within the EBD of XacBank there was no deeper knowledge of impact evaluations present. The concept of evaluation in the experience of the EBD was limited to a before-and-after measurement of CO_2 for each customer. The EBD were encouraging and wanted to be involved in LORTA, with most of the staff very quickly understanding the concept of a counterfactual and the potential of randomisation to do this. Unfortunately given the circumstances of the programme, such an approach was quick to be ruled out, leaving either an encouragement design or quasi-experimental methods available.

Some information is already collected from the clients when they apply for a loan. It may be possible to supplement the information collected at this stage. Ms. Tuul Gazagd, the head of the EBD, also suggested that the XacBank call centre staff could possibly be used to collect data, as they are experienced in asking questionnaires for marketing purposes.

Secondary data sets

The evaluation can benefit from the integration of primary data collected with a multiplicity of secondary data sources – both at the baseline and later stages. Additional data from the Mongolian SME Observatory will be helpful in understanding Mongolian MSMEs and may enable us to calculate better informed required sample sizes.

Given that all data will be collected at the firm level, with firms dispersed widely (mostly across Ulaanbaatar), geographic information system (GIS) environmental data will not be appropriate to use. The wide array of different investment types also means that other outcomes of interest to measure the impact on the environment will likely differ between borrowers. This will mean that different firms cannot easily be compared, so collecting complex environmental data, such as PM2.5 concentration, would not be worthwhile.

Finally, the monitoring may produce useful information. The Eco-loan programme has an M&E plan, which covers some important issues for the firms' investments. This will further be discussed in the next subsection.

Plans for monitoring and evaluation

The M&E systems put in place by XacBank are fairly comprehensive, They ensure that MSMEs plan to make sufficient GHG reductions (pre-evaluation) and check that these have been implemented by every borrower (post-evaluation). The pre-evaluation is carried out by the branch credit officer, with XacBank's corporate banking department becoming involved for loans of more than USD 300,000. XacBank reserves the right to visit each borrower at any time, which is often done during the risk assessment procedure. XacBank reviews each application and, once the loans have been disbursed, ensures funds were spent on the correct item by checking receipts and energy utility bills. In some cases, bank staff members visit the MSMEs to ensure equipment has been installed as promised. There is no continuous long-term monitoring of the use of purchased equipment, as XacBank is mainly interested in whether the firm has invested in the efficient machinery. While the machinery may potentially be sold on, this would mean that another firm is using this more efficient machinery. Further checks as part of our study could include checking that the equipment purchased with the loan is still owned by the MSME at a later date and asking about their use of this equipment. All borrowers at XacBank are required to return an annual environmental and social report related to the risk assessment strategy of the bank.

III. WAY FORWARD

Overall, we believe the LORTA mission in Mongolia was well received, with the EBD keen to see what could be learned from an impact evaluation. Their organization and collaboration enabled the LORTA team to develop a clear understanding of the programme as well as the challenges in establishing a method to be used for an impact evaluation of it. Power calculations will be carried out for the design that is selected. It must be noted that due to the small number of MSMEs taking out loans, it seems unlikely that an impact evaluation of the Eco-loan programme would have enough power to estimate causal effects.

Planning will continue in close cooperation with the IEU to decide upon the best evaluation method. These plans must be agreed upon with the EBD to ensure full cooperation. Ideally, we will be able to gain access to secondary data to locate further MSMEs and then conduct a matching exercise, but the usefulness of this is contingent on the power calculations, which still need to be carried out. If the evaluation is not likely to be able to draw causal conclusions, it is also important to consider the good use of the budget of the programme, and so perhaps reduce the intensity of data collection.

APPENDIX I

Table 1 Agenda of Mongolia LORTA field mission

DAY	TIME	ACTIVITIES	PARTICIPANTS
Day 1, September 10	11:00–11:30	Overview and discussion of objectives of LORTA	NDA, XacBank Eco banking team
	11:30–11:45	Brief introduction of XacBank's MSME loan programme	XacBank, Eco banking team
	11:45–12:30	Project activities plan, stocktaking of documents	NDA, XacBank Eco banking team
	12:30–13:30	Lunch	
	13:30–16:00	Theory of change discussion (including coffee break)	NDA, XacBank Eco banking team
Day 2, September 11	10:00–11:00	The benefits of impact evaluation – discussion	NDA, XacBank Eco banking team
	11:00–12:00	Presentation of the project by the country team	XacBank, Eco banking team
	12:00–13:00	Lunch	
	13:00–15:00	Capacity-building workshop (including coffee breaks)	NDA, XacBank Eco banking team
	15:00–15:30	Discussion with Eco banking team on Main elements of IE design	XacBank, Eco banking team
	15:30	Meeting with Compliance officer	XacBank, Compliance officer
Day 3, September 12	09:00–17:00	Field visits (3 borrowers of MSME loan program): 1. Solar panel seller 2. Heat-only boiler replacement project 3. Construction of EE school building that is installing rooftop solar	XacBank, Eco banking team
Day 4, September 13	09:30–16:45	GCF Readiness Workshop (Full schedule attached below)	XacBank, NDA, consultant team
Day 5, September 14	10:00–12:00	Debriefing: Presentation to the project team; key elements and timeline of design	NDA, XacBank Eco banking team

Table 2 Agenda of Thursday session of Mongolia LORTA field mission, 13 September 2018

TIME	SUBJECT	PRESENTER/ MODERATORS
09:30–10:00	Registration/Tea, coffee	
10:00–10:10	Opening remarks	Batjargal Kh. Director, Climate change and International Cooperation Department, MET
10:10–10:25	The Readiness Project implementation progress and stakeholder engagement	Dr. Dagvadorj D. Lead national consultant
10:25–11:00	Introduction on the draft GCF Country Programme	
11:00–11:20	Coffee break	
11:20–11:40	Introduction on NDA operational manual	Mr. Juerg Klarer International consultant
11:40–13:00	Discussion on the draft GCF Country Programme and NDA operational manual	Dr. Dagvadorj D, Mr. Juerg Klarer, Consultant team
13:00–14:00	Lunch	
14:00–14:45	Session on Developing a Country Programme for GCF Engagement	Ms. Bolormaa E. Associate Professional for Asia Pacific Region, GCF
14:45–15:00	Q&A	
15:00–15:20	Coffee break	
15:20–16:00	Business loan programme for GHG emission reduction in Mongolia – LORTA	Nicholas Barton and Asmus Zoch International Experts, the IEU, GCF
16:00–16:15	Q&A	
16:15–16:45	Conclusion and closing remarks	Dr. Batjargal Z. National Focal Point for GCF

Note: Objective of this session was an **i**ntroduction and discussion on the GCF country programme and NDA operational manual

APPENDIX II

Table 3 Participants in Mongolia LORTA field mission

NAME	POSITION	INSTITUTION
Tuul Galzagd	Director of the Eco Banking Department	XacBank
Greg Zegas	Project Development Officer	XacBank
Batsanaa B.	Project Development Officer	XacBank
Anushay A.	Project Development Officer	XacBank
Anand Vanchin	Project Development Officer	XacBank
Dr. Batjargal Zamba	GCF Focal Point	Ministry of Environment and Tourism
Bolormaa Enkhbat	GCF Associate Professional	GCF
Saruul Dolgorsuren	Managing Officer	Environment and Climate Fund
Jürgen Klarer	Consultant for the readiness workshop	Aequiconsult

IMPACT EVALUATION DESIGN REPORT 3: UGANDA

TABLE OF CONTENTS

TABLES

FIGURES

I. INTRODUCTION

A. THE LORTA PROGRAMME

Evaluating the impact of development projects and programmes has gained importance in recent years. Impact evaluation not only allows for increased transparency by measuring the effects of investments, it also provides the opportunity to design and implement development projects more effectively. To contribute to this development, the Independent Evaluation Unit (IEU) of the Green Climate Fund (GCF) has started the Learning-Oriented Real-Time Impact Assessment (LORTA) programme to be able to keep track of GCF projects in terms of performance and results, and to enhance learning within the GCF.

The LORTA programme aims to

- Embed real-time impact evaluations into funded projects so GCF project task managers can quickly access accurate data on a project's quality of implementation and likelihood of impact, and
- Build capacity within projects to design high-quality data sets for overall impact measurement.

The purpose of the impact evaluations is to measure the change in GCF key result areas that can be attributed to project activities. The LORTA programme will inform on returns on GCF investments and help GCF projects track implementation fidelity. The objectives of LORTA include the following:

- Measuring the overall change (outcome or impact) of GCF-funded projects and enhancing learning
- Understanding and measuring results at different parts of theories of change
- Measuring the overall contribution of the GCF to catalysing a paradigm shift and achieving impacts at scale

Currently, the LORTA programme is in Phase I (formative engagement and design). In the first year (2018), the IEU will support eight GCF projects to build high-quality, theory-based impact evaluation designs at inception. Formative work will include engaging with project teams, accredited entities (AEs) and GCF staff; designing theory-based impact evaluations; and establishing protocols for database development.

The second phase of LORTA will involve the main impact assessment stage (3–5 years), while the third phase will include analyzing baseline and endline data, discussing results, and engaging with diverse stakeholders to share results and incorporate feedback as required.

B. THE UGANDA PROJECT

Climate change, in combination with human and environmental stress factors, has led to a significant degradation of the wetlands in Uganda. This development is particularly deleterious for the approximately four million people living in and around these wetlands, in the so-called catchment areas, as these communities are dependent on the wetland ecosystem and resources for their livelihood and food security. According to Kakuru et al., 80 per cent of the population living adjacent to the wetlands use them to cover household food security needs.[19] Moreover, around 2.7

[19] Kakuru, W., Turyahabwe, N., & Mugisha, J. (2013). Total economic value of wetlands products and services in Uganda. *Scientific World Journal*, 2013. doi:10.1155/2013/192656.

million people in Uganda derive sustenance directly or indirectly from these wetlands.[20]

Climate change is expected to lead to increasingly erratic rainfall and more frequent extreme weather events. The resulting floods, droughts and heatwaves threaten to exacerbate wetland deterioration and make the current livelihood options even more unsustainable. This will undermine the critical role of the wetlands in maintaining water quality and providing ready access to water and other resources for the vulnerable populations in the surrounding areas.

The "Building Resilient Communities, Wetland Ecosystems and Associated Catchments in Uganda" project (henceforth referred to as "the wetland restoration project") is one of the eight projects that was chosen to be part of the LORTA initiative. It focuses on adaptation to climate change via the increased resilience of the local ecosystems and communities. The project particularly focuses on strengthening the resilience of the wetlands and their sustainable restoration, coupled with alternative livelihood trainings for the population living adjacent to them. The total number of individuals expected to gain from this project is 800,000. Women are particularly anticipated to benefit from the project and are partly a focus of project activities. This is because the manual work linked to wetlands is traditionally undertaken by the females in these communities. Consequently, if targeted well enough, women would benefit disproportionately from improved livelihood options or living standards (improved water sources, for instance, which would reduce their water collection chores and time).

The project is in line with a number of ongoing Ugandan national government strategies, such as Vision 2040, the Second National Development Plan, the National Climate Change Policy, the Second National Communication (2014), the National Adaptation Programme of Action (2007) and the National Adaptation Plan, which aim to tackle and mitigate the consequences of climate change. Not only is this subject of predominant interest at the national level, it has also gained traction within the communities. At the community level, wetland restoration and preservation is a vital issue, and Community Conservation Areas, framework management plans and Ramsar Site Management plans have therefore been established to facilitate the process of climate change mitigation and adaptation.

There are several national and subnational entities involved in the planning and implementation of the wetland restoration project. The project will be implemented by several ministries in Uganda, in particular, the Ministry of Water and Environment (MoWE), the Ministry of Agriculture, Animal Industry and Fisheries (MAAIF) and the Uganda National Meteorology Authority (UNMA). There are also several expected institutional beneficiaries, such as the National Forest Authority, National Environment Management Authority, and Ministry of Lands, Housing and Urban Settlements, among others.

The wetland restoration project has three goals or components:

1. Restore critical wetlands to improve ecosystem services such as groundwater recharge, flood control, fishing and agriculture for enhanced livelihoods to the most vulnerable subsistence farming communities.
2. Diversify livelihoods and agriculture to make wetland communities more resilient to climate shocks, by enhancing the skill set of beneficiaries for employability and adaptation.
3. Strengthen communities' preparedness and risk reduction mechanisms to cope with climate-related disasters in sensitive wetland areas. This will be done through participatory and

[20] Ministry of Water and Environment, Climate Change Department. (2014). *Uganda Second National Communication to the United Nations Framework Convention on Climate Change*. Kampala, Uganda: Government of Uganda. Retrieved from https://unfccc.int/resource/docs/natc/uganc2.pdf.

decentralized early warning systems and capacity development for implementing disaster risk reduction measures.

II. UGANDA FIELD MISSION

A. GENERAL REMARKS

An evaluation team, consisting of two consultants from the Center for Evaluation and Development (C4ED) and Dr. Babatunde Abidoye from the United Nations Development Programme (UNDP) (henceforth referred to as the LORTA team), led the field mission from 8 to 12 October 2018. The task of the team was to engage closely with key stakeholders of the project – namely, the national designated authority (NDA), the accredited entity (AE), implementing agencies, project staff and potential end beneficiaries – to ensure their interest in and understanding and sense of ownership of the planned theory-based impact evaluation.

During the field mission, the LORTA team held meetings and capacity-building workshops with the key stakeholders. Meetings, in the form of expert interviews, were used to acquire the maximum possible information about the GCF-funded project. Stakeholders were interviewed regarding their views about the project's implementation and monitoring strategies, the expected impact, challenges and possible solutions. The meetings not only informed the LORTA team about the project but also aimed at fostering collaboration and trust between the team and the on-site parties involved. In addition, a capacity-building workshop on impact evaluation was held, targeted at the key stakeholders. Beside conveying technical knowledge, the aim of this workshop was to emphasize the benefit of theory-based counterfactual approaches and real-time learning and measurement.

In collaboration with the UNDP and project team, an existing impact evaluation design was revised for the Uganda project. Building on previous work done by UNDP and the project team, the LORTA team conducted context analyses, examined the existence of appropriate counterfactuals (i.e. comparable treatment and control groups), built on the existing theory of change, assessed the availability of baseline and secondary data sources, and acquired budget information. The results of this undertaking are presented in the following sections. It should be noted that a monitoring and evaluation plan that included an impact evaluation design had already been developed by UNDP before LORTA started and was used as the basis for this design report. This impact evaluation design and monitoring plan had been submitted to the GCF as part of the project implementation plan.

B. THE MISSION AGENDA

The LORTA team received prompt collaboration from the UNDP project team. The agenda – shown in appendix I of this design report – was developed to facilitate the joint attendance of all key stakeholders at the LORTA workshop, to plan for key stakeholder interviews, and to allow for field visits by the LORTA team.

The LORTA workshop was condensed into one and a half days, but it contained all the ingredients to inform the key stakeholders on the LORTA objectives, present key impact evaluation concepts, develop a detailed theory of change (ToC), and discuss programme implementation details and likely risks to the impact evaluation design with the implementing partners.

The workshop was successful in that it benefited the key stakeholders and the LORTA team. On the one hand, the presentations and interactive discussions on the ToC and implementation brought all the key stakeholders together on the same page, with respect to their sense of ownership and understanding of their contribution within the project, starting from the objectives through to the

evaluation needs and implementation strategies. Simultaneously, the LORTA team benefited by being able to gather, in a short time, a rich set of crucial information to design the impact evaluation. Moreover, the workshop aided in our understanding and collaboration with the key implementing agencies, in adjusting their plans to accommodate the evaluation design and vice versa.

During the mission, the C4ED consultants were able to establish a tight collaboration with UNDP and the UNDP consultant, Dr. Abidoye. This enabled us to engage with key stakeholders and to elaborate on an evaluation design and evaluation tools, as dictated by the current implementation schedule.

C. RESULTS

Implementation plan

The wetland restoration project in Uganda is a comprehensive undertaking, which focuses on improving the human living conditions of communities living around wetland areas, while restoring biodiversity in functioning wetland systems and catchment areas. This is conducted through the interplay of three project components. The first component addresses physical wetland restoration, the second, environmentally compatible alternative agricultural and non-agricultural livelihood trainings, and the third, support in climate change adaptation through improved climate information systems. The intervention targets two Ugandan regions, the Eastern Region (comprising the 10 districts of Pallisa, Kibuku, Bukedea, Namatumba, Buraleja, Budaka, Tororo, Kaliro, Ngoro and Mbale) and the South-Western Region (the 6 districts of Kabale, Kisoro, Kanungu, Rukungiri, Greater Bushenyi and Ntungamo).

Component 1 concerns the physical wetland restoration. Depending on the level of degradation and general geography, wetland restoration may involve very different actions. A healthy wetland in Uganda is flooded with water and is home to flora and fauna such as papyrus and certain types of birds. Wetlands are typically surrounded by features such as trees or natural dams, which facilitate water storage in the wetland. Several climate-related and human-led activities have led to wetlands drying out in Uganda, destroying the aforementioned flora and fauna, and the wetland water retention facilities. This situation has led to severe droughts and floods, which have greatly affected the communities dependent on these wetlands. Finally, communities located in the areas around the wetlands started to move out of the catchment areas into the degraded wetland areas to be closer to the dwindling water sources. Additionally, communities started to use fertile wetland soil for their agricultural activities, such as rice planting, which requires acres of land to be flooded. These agricultural activities within the wetlands have led to further drying out of and damage to the wetland areas.

Wetland restoration programmes, accordingly, require communities to abandon these damaging practices. Additionally, water storage areas that link the wetlands and wetland borders can be put in place, which will enable water to flow from the storage areas to the wetlands and vice versa, depending on the water level in the wetlands. These facilities prevent the wetlands from either drying up or flooding.

Component 1 examines the necessary actions to be undertaken for the restoration of each wetland area and needs to gain buy-in from the communities first. The path to buy-in is conducted through district leaders, who mobilize community leaders through their own communication channels to educate people about the severe state of their wetlands, the actions that can be undertaken to restore the wetlands and, importantly, the benefits restored wetlands would bring to the community. This is also where the interplay between components 1 and 2 becomes crucial, to show community members alternative livelihoods that are compatible with healthy wetlands and ideally economically more profitable. By convincingly conveying the benefits of wetland restoration, communities will be

engaged enough to conduct wetland restorations by themselves – at least, as far as possible and always with the support of the implementing staff.

Component 2 of the project aims to offer viable alternative livelihoods to community members, to ensure their willingness to voluntarily move out of wetland areas with existing housing and agricultural activities. Component 2 is then compatible with the goals of sustainably restoring the wetlands, improving the communities' overall economic living conditions, and facilitating communities' adaptation to climate change – that is, reducing communities' vulnerability to climate phenomena. Accordingly, this component has to be implemented at the same time as the wetland restoration activities under component 1. Training will be conducted at the community level and all community members are eligible to receive it. Priority will be given to those community members who will be directly affected by wetland restoration, as well as those in the catchments.

As outlined above, sustainably restored wetlands require communities to move out of wetland areas and abandon harmful agricultural practices. Consequently, community members have to be offered alternative livelihood options that guarantee their food security and remove their complete dependence on the wetlands, thereby making the idea of moving out of the wetlands rewarding. To identify alternative sustainable and profitable livelihoods for wetland communities, the non-governmental organizations (NGOs) World Vision and Environmental Alert conducted an Alternative Livelihood Study in 2018, on behalf of the UNDP and the MoWE. Communities were surveyed to identify attractive and profitable alternative livelihoods in 11 districts, representative of the whole sample population. Both NGOs are currently employed by the project to train communities in the pilot region in these alternative livelihoods. Every community is offered trainings in the alternative livelihoods identified in the Alternative Livelihood Study, and the training package is modified according to community members' demand. Trainings will therefore include different combinations of livelihoods – for example, beekeeping, fishing or the opening of barber shops. Being contingent on communities' interests, this package will differ across communities. The MAAIF is responsible for this project component.

Component 3 complements climate change adaptation for communities through improved climate information. This undertaking first involves the installation of improved weather stations; second, data analysis produces meaningful weather forecasts; and third, comprehensible climate information is adequately disseminated to the community members. The UNMA is responsible for this component. After the installation of the weather stations, which enable higher-quality data collection, UNMA estimates a minimum of five years of data collection and analysis on time series weather data will be needed in order to create reliable forecasts for a period of several days. In the next step, these forecasts have to be compiled into comprehensible and relevant packages for specific groups in the communities (fishers, farmers, non-agricultural population) and then translated into the different local languages for eventual dissemination. Different modes of dissemination, such as radio, SMS, extension workers or newsletters, are envisioned to adequately transmit the climate information to the relevant population. However, some improved climate information might be reliably available at an earlier stage, before the end of the five-year research period. Currently, the timing of the availability of such information is rarely predictable by project staff. However, UNMA aims at immediate dissemination of improved information to the community members for whom specific information is relevant.

Components 1 and 2 – that is, physical wetland restoration and alternative livelihood trainings – have already been piloted in the Pallisa district. The remaining 64,370 ha of wetlands and 11,630 ha of catchment area will be restored through until the end of the project period in 2025. Project implementation of both components will be phased in over a six-year period as preliminarily outlined in Table 1. Neither the exact amount of restoration hectares per year nor the locations are fixed at the current stage, which explains the preliminary nature of the implementation plan.

Table 1 Project implementation plan (preliminary)

YEAR	ACTIVITY	DATA COLLECTION
2018	Identification of first 10,000 ha	Baseline planned
2019	Implementation starts in 10,000 ha	
2020	Implementation starts in 10,000 ha	
2021	Implementation starts in 10,000 ha	Midline planned
2022	Implementation starts in 14,000 ha	
2023	Implementation starts in 10,000 ha	
2024	Implementation starts in 10,000 ha	
2025	Activities ongoing in all 64,000 ha	Endline planned

The hectares per year are envisioned by the project team to be relatively evenly distributed as proposed in Table 1, but specifics still need to be decided based upon practical implementation considerations, which might not have been foreseen at the current stage. For the first implementation year 2019, however, 10,000 ha of wetland area have already been selected. Selection of the first 10,000 ha was done in a non-random, needs-oriented political process. The implementing agency MoWE discussed with all 16 district leaders which communities in their respective districts would be in greatest need of wetland restoration. The greatest need is related to wetland dependencies with regard to communities' water sources, for instance. The locations of these first 10,000 ha are already geographically mapped, and communities have been informed about project implementation in 2019. For the remaining 54,000 ha, the exact vulnerability of key agroecological and hydrological systems of the wetlands and communities' dependencies is to be identified in a more systematic manner.

At time of writing, procurement is under way for the geographic information system (GIS) and ecological consultants who will map the characteristics of the entire wetland area in the remaining 54,000 ha. This exercise is crucial, as it will result in the project implementation and evaluation team acquiring detailed maps of the wetland and catchment ecosystems, the flora and fauna within and around the wetlands, the vulnerability level of all wetland systems and subsystems (and communities), and – finally – the exact area that will be restored or affected as a result of restoration at one point or one node of the wetland system. Therefore, the role of the mapping exercise would be to determine the vulnerability of each wetland system and catchment area, and to identify the exact nodes within the system (marked bright green in Figure 1) where the restoration efforts should be located. When the intervention is executed at all these exact points, it should ensure the restoration of the entire wetland and associated parishes. Not all these nodes will be tackled simultaneously, and therefore the mapping will enable a clear identification of each area that will immediately or eventually receive intervention. Hence, the mapping will identify the intervention units for components 1 and 2. Currently, this exercise has only been carried out for the first 10,000 ha, a small area map of which is shown in Figure 1.

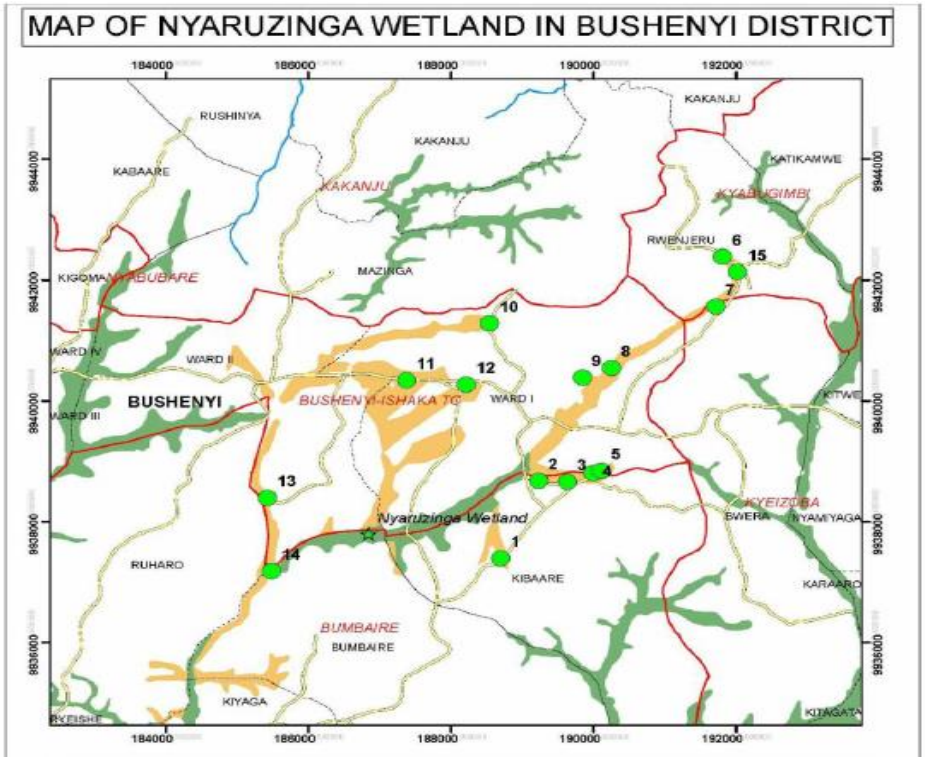


Figure 1 Map from Nyaruzinga wetland in one implementation district

Wetlands consist of interdependent systems. That means that if we restore a wetland area at one specific node, say at node number 2, the recovery of the wetland area around node number 2 might positively influence recovery of wetland areas around node number 3. At the same time, other wetland areas may be "independent" and not affected by restoration of other wetland areas of the same system. Crucially for the design of an evaluation strategy, the mapping will identify these independent restoration nodes. The areas around these independent restoration nodes will accordingly only be affected by the restoration at their specific node, but not by restored wetland areas at another node of the same wetland system.

At time of writing, the baseline data collection for wetland communities in the whole restoration area is planned for late 2018, but since the mapping exercise will only be completed at the end of 2018, our recommendation is that it be moved to early 2019 or later. Wetlands typically need the flooding water from four rainy seasons to recover their flora and fauna, once damaging practices are removed and retention facilities rebuilt. This corresponds to a minimum of two years before the outcomes at wetland level are visible and can be measured. Therefore, the midline should only take place at the start of 2022, after two years of implementation have passed in the areas receiving intervention.

Theory of change

A theory of change is laid out below for each of the three components associated with the wetland restoration project. It should be noted that these theories of change rely on several underlying assumptions, which were extensively discussed during the workshop in Uganda. Component 1 can only be successfully implemented in the case of sufficient political will and support - specifically, support from community leaders and communities themselves. An extensive engagement plan is in place and budgeted in order to realize the needed support. Component 2 requires community members' willingness to adopt alternative livelihoods. At the same time, in order to sustain food security and benefit from the project's target to improve agricultural production, not all community

members should abandon agricultural livelihoods. Component 3 requires meaningful tailoring of climate information to the end users' needs and, ultimately, that beneficiaries change their behaviour according to the new information.

Component 1: Physical wetland restoration

i) Inputs

Input factors are machinery and capacitated staff. The budget is provided by the Ugandan government and UNDP Uganda, through GCF funding.

ii) Activities

The implementing agencies will engage community members in wetland restoration. This activity includes convincing community members to move out of wetland areas and to abandon harmful agricultural practices. Depending on the degree of degradation, and thus restoration efforts, community members will either be offered lunch and free transportation or per diems as compensation. In terms of restoration, community members will build retention facilities in the wetland and catchment areas – for instance, by planting trees. Additionally, small-scale water storage areas will be built near the wetlands, which will serve the dual purpose of reducing flooding during heavy rains and providing water for irrigation systems, especially during dry seasons.

iii) Outputs

After full implementation of component 1, 64,370 ha of wetlands and 11,630 ha of catchment areas will be restored.

iv) Outcomes

As an outcome, the community would sensibly manage the wetland areas, preventing new degradation and sustaining the functioning water storage systems. They will have also moved out of wetland areas and abandoned harmful agricultural practices.

v) Goals

The main goals behind physical wetland restoration are lower food insecurity and enhanced resilience to climate shocks.

Component 2: Alternative livelihoods

i) Inputs

The inputs are capacitated staff and seed funds.

ii) Activities

The MAAIF will lead the training of the communities on economically viable and sustainable agriculture-based and non-agricultural livelihoods and be supported by the district local government and NGOs. These income-generating interventions are introduced, promoted and supported in the wetland and catchment areas, depending on the demand and interest in the communities. Additionally, saving schemes are introduced in the form of revolving funds. These will ensure the availability of funds to repurchase agricultural inputs when the project funding is no more available.

iii) Outputs

Farmers are trained on alternative livelihood options, which could be vegetable farming, for instance, for agriculture-based livelihoods, or training for shop ownership for non-agricultural livelihoods. Sustainable agricultural practices are disseminated – such as the wetland-based irrigation system – which allow for multiple planting times per year and higher yields. Additionally, saving schemes are in place in the form of revolving funds, with members contributing and borrowing without interest, to invest in input factors for their respective businesses.

iv) Outcomes

Farmers adopt the alternative livelihood options and sustainable agricultural practices. The saving scheme is sustainably managed.

v) Goals

The main goals behind the alternative livelihood trainings are again to contribute to higher agricultural yields, higher revenues, lower food insecurity and enhanced resilience to climate shocks.

Component 3: Climate information systems

i) Inputs

The third component has input factors on three different levels. These include hydromat equipment such as weather stations and water sensors, computing capacity referring to hardware and human resources to analyze the data collected through the improved hydromat equipment, and lastly, climate information material that can be disseminated to the end users.

ii) Activities

Water-level sensors are being installed, data are being collected and analyzed, and forecasts are produced and comprehensively translated and tailored to the end users' needs. The information is disseminated through different dissemination modes including radio stations, newsletters, cell phones and extension workers.

iii) Outputs

Water-level sensors are successfully installed, data are constantly collected and analyzed to produce reliable forecasts, and farmers are provided with climate information concerning, for instance, the water level, irrigation times or early warning information.

iv) Outcomes

Farmers apply the climate information to their agricultural production and community practices in general, to increase their resilience against climate hazards.

v) Goals

Component 3 also contributes to increased resilience of communities to climate hazards, higher yields for farmers, higher income and higher food security.

Evaluation questions

The main research questions to be answered by this impact evaluation, derived from the hypotheses underpinning the theories of change, are listed below:

- Are wetlands sustainably restored?
- Are income levels of community members increased?
- Is income volatility reduced?
- Are community members more resilient against climate change disasters (floods, droughts and mudslides)?

Outcome indicators

- Wetland level via satellite data, including vegetation cover (ha) and water cover (ha)
- Adoption of enhanced agricultural practices (planting times, seed varieties)
- Take-up of trainings (share of eligible population that participates)

Impact indicators

- Food and non-food consumption

- Health indicators related to nutrition
- Agricultural production
- Income
- Income volatility per year
- Employment level/status

D. PLANS FOR EVALUATION

Methodology

A comprehensive mixed-methods approach will be adopted for the wetland restoration project, where the results from the analysis of quantitative data (primary and secondary) will be triangulated using information from qualitative interviews and focus group discussions. The quantitative part will consist of the impact evaluation using primary data from surveys, analysis of secondary data available (satellite data, Demographic and Health Survey and Living Standards Measurement Study data) and, finally, regular briefings on project monitoring data. This will be complemented by three rounds of qualitative data, planned at the baseline, midline and endline data collections. Data will be collected from stakeholders in the form of key informant interviews and from beneficiaries using focus group discussions and in-depth interviews. The role of the qualitative analysis will be to provide a deeper insight into the functioning of the project and to explore community, household and individual preferences and behaviours that influence the uptake and sustained effects of the project activities.

Impact evaluation design

This section will present our proposed impact evaluation (IE) strategies building on the IE design that was developed by UNDP. We will focus on the evaluation of components 1 and 2 because treatment timing of component 3 is too unpredictable to set up a robust IE design. In order to finalize the IE strategy for components 1 and 2, however, we require the completion of the mapping project activity (as described in section C), which is planned to be concluded by December 2018. This crucial activity involves the exact mapping and characterization of the targeted wetland areas and benefiting communities. We have incorporated different potential outcomes of the mapping activity in our IE design proposals and will present them in descending order from most to least ideal circumstances with regard to causal identification.

The mapping exercise will yield a sampling frame for the entire 64,000 ha to be restored – that is, it will clearly mark all relevant nodes that are to receive the intervention. Moreover, it will provide information on observable characteristics of each node in the wetland and the vulnerability level of the communities and wetland itself. It is only after we have an established sampling frame of all nodes, and ergo can determine the control and treatment groups, that we can design the exact IE strategy to be implemented.

It is also important to reiterate here that the selection of the first 10,000 ha was purposive, based on the suggestions of and discussions with the district leaders, who have a good overview of the most vulnerable wetland communities. Therefore, in all likelihood, these will be removed from the entire IE exercise, and will only be covered under the project and outcome monitoring analysis.[21] Since

[21] The only case where this would be avoided, and the initial 10,000 ha be included in the IE design, is if the mapping exercise (to be finished in December 2018) determines that these selected 10,000 ha are not significantly different from the remaining hectares of wetlands in terms of vulnerability or degradation. In this case a randomized control trial is possible, to examine the impact of the restoration on these 10,000 ha compared to another 10,000 ha that are part of the project and

these 10,000 ha are to be removed from the evaluation, it is recommended to postpone the baseline survey to the end of 2019, right before the start of the implementation in areas that will become the treatment group for the IE. This is to ensure that the baseline characteristics for the treatment and control groups are not likely to undergo considerable change between the start and end of 2019 (planned baseline and proposed new date for baseline, respectively).

Scenario 1 (ideal conditions)

Based on the staggered nature of the implementation – due to limited resources and capacity – but the assured eventual coverage of the entire population within the intervention area, a phase-in randomized control trial (RCT) can be implemented. As the first 10,000 ha to implement the project in 2019 were non-randomly selected, we propose to randomly select the next 10,000 ha for project implementation in 2020.

From the total 64,000 ha envisioned for project implementation, we will subtract the first non-randomly chosen 10,000 ha and be left with 54,000 ha to draw our random sample from. Our unit of randomisation will be the independent restoration nodes that will not be affected by restoration efforts in other parts of the same wetland system. The mapping exercise will enable us to remove the risk of spillovers. We will randomly select nodes covering 10,000 ha to be treated in 2020, which will serve as our treatment group for the IE.[22] Likewise, we will simultaneously randomly select nodes covering 10,000 ha for our control group, which will receive the treatment only in 2024, after our midline data collection in 2023.

It has been noted from our discussions with the implementation staff that there are systematic differences in the two regions of implementation, and therefore a stratified randomisation will have to be used, where the Eastern and South-Western Regions are the two different strata. In case of differences in population density and size in each cluster, another level of stratification at the population size/density level might be required. Due to the stratified random assignment of the nodes into treatment or control groups, all observable and unobservable characteristics of the two groups will be balanced, some of which we will also ascertain with the baseline data collection in 2019, covering these 20,000 ha. Consequently, the analysis at midline[23] will provide an unbiased impact of the intervention between the control and treatment hectares.

---

have not received intervention yet. To include these 10,000 ha in the design as well, the baseline would have to take place in early 2019, before implementation starts in these wetland hectares. Our discussion with the project implementation team made it clear that implementation will only commence in these first 10,000 ha in 2019, and therefore we anticipate no problems in measuring accurate outcomes within the baseline data collection.
Alternatively, it could be that we find another 10,000 ha of wetlands in the remaining 54,000 ha that are of similar vulnerability and degradation level and therefore comparable on these characteristics with these first 10,000 ha. In this case, after we collect baseline data covering the entire 64,000 ha, a matching-based difference-in-differences approach can then be applied to estimate impact in these initial 10,000 ha and the other 10,000 comparable hectares (matching on degradation, vulnerability, community characteristics, etc., from mapping and baseline data collection). However, the likelihood of finding comparably vulnerable wetland systems in the remaining 54,000 ha is, in the opinion of the LORTA team, highly improbable, making the establishment of a control wetland group extremely unlikely. Therefore, we only suggest an analysis of the project implementation monitoring data for these first 10,000 ha, to establish the trends in measured outcomes.

[22] We only randomly sample 10,000 ha for treatment and 10,000 ha for control because it is likely going to be very hard to convince the project implementation team to randomize implementation in all hectares between 2021 and 2023 as well. Therefore, to avoid more complication and maintain goodwill, we propose to only randomly sample the first and pre-midline 10,000 ha.

[23] This will be in 2023 and is also the effective endline for the impact evaluation since thereafter all regions will have received the treatment.

Scenario 2 (less ideal for impact evaluation design)

Discussions with the project team raised concerns about the practical feasibility of not starting any project activities in the pre-determined control areas before 2024. Therefore, to ensure that we can adapt to unannounced or unintended changes in implementation plans in the assigned control group, which would threaten the internal validity of the IE design, we propose an alternative difference-in-differences (DiD) matching design.

To account for the difference in characteristics between the regions, varyingly sized clusters (population density and size) and difference in community characteristics, we can match the control and treatment clusters on a set of observables, based on data that would already be available after the mapping and baseline data collection in all 54,000 hectares.[24] This will establish the first 10,000 ha, to be assigned as the treatment group, which will cover clusters of differing sizes (number of nodes, population and area covered) but will be balanced in their vulnerability (of wetland and population) levels. Hence, the selection of the first 10,000 ha for the treatment group will not be random, but would be based on all characteristics that make it most likely to match them with the remaining 44,000 hectares.

Relevant criteria for this group-wise balancing will be the region, population size, average population density, vulnerability level of the wetland areas and community within, and community demand for training components. This will imply that at the start of the restoration activity in each year, we will have to match all the observable characteristics of the first 10,000 ha with the remaining hectares of wetland (34,000 ha in 2021 and 20,000 ha in 2022, if following implementation outlined in Table 1). Consequently, the LORTA team would have to remain in close consultation with the project implementation team at the start of each year to develop the implementation plan in a way that the control group area balances with the treatment 10,000 ha in each phase. Due to the flexible implementation needs of the project, with respect to the selection of the exact nodes or the exact 10,000 ha that are to be the control group, this method will, at the very least, enable us to maintain a balance between the two groups on the whole. Accordingly, we would be using a matching-based DiD estimation to establish a causal impact of the intervention, removing any confounding time-invariant differences between treatment and control group.

In other words, the DiD technique relies on the crucial assumption that differences between the two groups – in the absence of the programme – are fixed over time (known as the "parallel trends" assumption). Although this is generally an untestable assumption, evidence can be provided on pre-programme trends on relevant outcomes/impacts between treatment and control group regions – conditional on the availability of such data. Moreover, using monitoring data collected within the programme itself, we can also establish parallel trends on some outcomes within the programme. Finally, for establishing parallel trends on wetland restoration, satellite data can also be useful for periodic assessments of trends in control and treatment regions.

Nonetheless, time-varying differences are not controlled for within a DiD approach and – if present – would undermine the unbiased estimation of the programme effects. Examples of such uncontrolled differences are alternative infrastructural activities in the wetlands or similar livelihoods trainings in the control wetlands from other projects. These could imply that the dissimilarities in the underlying conditions between control and treatment would drive the difference between outcomes.

At the very start, however, we need to be cautious with regard to one issue that can only be clarified after the mapping exercise. Since the matching would occur on a number of different covariates

[24] This implies that we would be oversampling in our control areas, to ensure that some observations from remaining untouched wetlands in 2023 are matching with the first 10,000 ha.

mentioned above, and will likely be stratified according to regions in the first place, a sufficient number of matchable nodes need to exist, and be correctly identified, for the creation of the treatment group, and especially a control group, each year.

Alternative methods for evaluating component 2

As mentioned already, there are two types of livelihood trainings that would be given to the population in the treatment areas:

1. Those dependent on the wetland (farming with wetland irrigation, fish ponds, etc.)
2. Those independent of the wetlands (barbers, small businesses such as grocery shops, etc.)

These trainings provide information to the treatment communities about climate-smart agricultural practices, as well as sustainable and efficient business practices. The difficulty in the evaluation connected with them is the spillover of information on these practices between communities receiving training in 2020 and those that will receive them in 2024 (control communities). The spillovers from the first training type are much less likely, since these are dependent on the wetland restoration itself and need skilled service providers to transfer and cascade the knowledge, as well as to finance/deliver significant inputs for the realization of these livelihood activities, to the beneficiaries. Because of this skills requirement, the interventions are to be implemented by hiring competent personnel (agriculturalists, engineers, etc.) in restoration areas, as well as providing important inputs to the population (fish fingerlings, water storage facilities, seedling and plant nurseries, etc.), which automatically leads to a geographical demarcation for the spillovers from this livelihood option. On the other hand, the livelihood opportunities that are independent of the wetland restoration are more likely to be transferable to other areas of the wetlands, or even other wetland systems, depending on how much mobility and communication there is between communities in different wetlands. There could also be effects on the local economy, which might spillover to the neighbouring control communities.

To account for potential spillovers from the second type of trainings, we suggest two alternative IE designs:

i) Determining the minimum distance between control and treatment nodes: besides having the independent clustering of the nodes, we will establish a minimum distance required between the groups to eliminate the spillover concerns due to training type 2. Therefore, this will follow the same IE design as any of the aforementioned scenarios, while ensuring that intervention activities in intervention wetlands do not affect those in our control group. This, however, makes the matching activity even more complicated in scenario 2, and we will require even more nodes to ensure that we are capturing all effects adequately.

ii) Randomizing the training modality: a second round of "refresher" trainings, besides the normal training, could be included within some of the selected treatment nodes/communities, to assess the impact of these additional trainings. This IE design, however, will no longer answer the same question (i.e. the impact of these trainings) but will attempt to answer the questions about whether these trainings alone are enough or if they need to be repeated in the communities for better uptake, implementation and assessment of final desired outcomes of the communities.

Sampling

The sampling will take place on two different levels: that of the wetland system or subsystem around the independent restoration nodes (in case of the randomized control design or the matching set up) and then that of the households to be sampled within.

For scenario 1, at the first level, all the areas around the nodes that are to be considered as independent sampling units will be identified with the help of the mapping exercise. The mapping exercise will then provide an adequate sampling frame for the selection of the wetland systems/subsystems into control or treatment groups. For the second scenario, we will consider all independent nodes in the 54,000 ha to be included within the sampling frame, to match the 10,000 ha with the remaining independent nodes (covering 44,000 ha). Having established the exact sampling frame for the households, the sampling design and household selection will also be based on the various groups that will be targeted within the project (youth, female-headed household, farmers, etc.). Therefore, there will be a stratified randomized sampling of households from the control and treatment areas. We would like to reiterate at this stage, for scenario 2, that there will be an oversampling of households within the 44,000 ha to ensure a comparable control group over the three years of implementation between 2020 and 2024.

For the sampling frame for all the beneficiaries in the control and treatment, we intend to rely on available census data household listings, lists prepared during agricultural surveys, data on the beneficiaries of the livelihood training from the implementing NGOs, and other available administrative lists (from other programmes or from the schools, farmer trainings, etc.) to generate a sampling frame. Discussions with the project implementation team suggest that there are already existing monitoring data on these wetlands, which we hope will also reveal potential methods to identify communities and households that are dependent on them, in order to develop our sampling frame. It is unlikely that we will conduct a listing exercise over the 54,000 ha[25] since this will be a massive undertaking, for which there is no allocated budget.

Power calculations

Currently the mapping of the wetlands is under way, which is the crucial prerequisite to the IE strategy. As long as we are lacking information on our units of randomisation, which would be the independent restoration nodes, we are unable to perform power calculations.[26] The mapping will identify the number and location of the independent nodes as well as the number and size of communities living around the restoration nodes and benefiting from the restoration. Accordingly, we will be looking at a clustered randomized trial, with one cluster being the communities surrounding the independent nodes and benefiting from wetland restoration.

Power calculations will enable us to determine the minimum sample size needed in order to detect the impact of the wetland restoration project. To do that, we would use the following power formula that relates the sample size to the minimum detectable effect size (MDES) between the mean outcomes of two groups:

$$MDES = (t_{1-\kappa} + t_{\alpha}) \sqrt{\frac{1}{P(1-P)}} \sqrt{1 + \rho(m-1)} \sqrt{\frac{\sigma^2}{N}} \sqrt{1 - R^2}$$

where $t_{1-\kappa}$ and t_{α} are t-statistics representing the required power and level of statistical significance (by convention, we seek a power of 80 per cent and a statistical significance of 95 per cent), P represents the proportion in one of the two compared groups (allocation ratio), ρ is the intra-cluster correlation (ICC), m is the number of individuals per cluster, σ^2 is the variance, N is the total

[25] Potentially 64,000 hectares in case we do end up including the first 10,000 ha as well.
[26] While we have one report that identifies the number of wetlands to be restored in the project (76), without the mapping we cannot be sure how many nodes, and more importantly, how many independent nodes, will be derived from each wetland. Moreover, based on the size of these independent nodes/wetland areas, we can clearly have a large variation in the number of households per independent (cluster of) node(s).

sample size, and R^2 represents the extent to which baseline characteristics predict the endline outcomes.[27]

Since the livelihood trainings are delivered on a group basis and wetland restoration affects all households located in the cluster around one restoration node, it is likely that there will be some similarities in outcomes between the members of one cluster. It is important to take this aspect into account in the power calculations, which is effectively the ICC. Again, this is information that would only be clear after implementation plans are finalized based on the mapping activity.

Possible risks for the impact evaluation design

The capacity-building workshop clearly provided an understanding of the inclination of the stakeholders towards evaluation, as well as the complexities within the evaluation of each component of the project.

We were therefore able to identify and list the following potential risks to the IE design:

i) The first problem arises with the evaluation indicators for each component, and when the targeted results are expected to be visible. For the wetland restoration, a two-year period is the minimum required, provided the weather is favourable over the two years (implying no drought in the middle and two rain cycles per year). If there are weather shocks or other extreme weather events, the wetland restoration itself could be affected and delayed. It is therefore not possible to be sure that the midline can be conducted in 2023, because any severe climate shocks could mean the wetland restoration is stalled or even reversed in the meantime.

ii) Another concern relates to the design of the livelihood training component and the differential trainings that would be given to each of the communities, based on their preferences. If we were to compare the impacts of component 2 on the livelihood options and changes in outcomes for the communities, this might be biased, given the differential effects and market set-ups that will be developed as a result of these varying packages of livelihood trainings. If the diversity of trainings is balanced across groups (as it should be in case of randomisation in scenario 1, where a geographic variable relevant for the choice of training may also be chosen for the stratification), we can still evaluate the impact of the intervention, even if this will not tell us which type of training led to what. This is a larger concern for scenario 2, where it might indeed be that we are unable to account for time-varying differences in the two groups that accompany these trainings.

iii) Finally, the last notable concern relates to the particular political importance of this project. The implementation is driven by the political agenda and political connectedness of each wetland, which could lead to much more digression from the IE plan (in other words, it could be inflexible with regard to the needs of the IE design). A good rapport between the LORTA team, UNDP, MoWE and UNMA is required to ensure that the implementation process remains as flexible as possible, while still maintaining the integrity of the evaluation design. This implies that the LORTA team will have to ensure that the control wetlands in scenario 1 or the balance of the control group in scenario 2 is assured over the implementation between 2021 and 2023.

[27] During the capacity-building workshop, there were multiple outcomes of interest that were listed and considered important for each component by the participants. While estimating an effect for all of these would not be possible (multiple hypothesis testing), when it is certain which component can be included within the final IE design (if we are to leave out component 2, for instance), we can determine the exact outcomes that would be of joint interest to each of the stakeholders. Simultaneously, we can establish all other information, such as the MDES, the ICC and the standard deviation we expect in both groups for those particular indicators, with the help of the project implementation team.

Qualitative assessments

The qualitative assessment would involve interviews and focus group discussions with the beneficiaries and the implementing organizations. With regard to fostering our understanding of change in wetland communities, the exploratory stages, the anecdotal narratives, and all the textual material collected are essential to identify the social dynamics around which the project and all of its components operationalizes.

The baseline, midline and endline interviews will therefore cover themes of uptake, attitudes, perceptions, knowledge and nature of the project implementation and the experiences regarding the project. These will be captured using key informant interviews with the implementing agencies, in-depth interviews with individual household and community members and, finally, focus group discussions with multiple targeted groups within the communities (farmers, women, market vendors, etc.). These qualitative data would be particularly useful in identifying the assumptions that lie behind the results chain of each component's theory of change.

Data collection

Primary data collection

The project envisions baseline, midline and endline data collection. These are recommended to be scheduled in 2019 (mid or end), 2023 (start) and 2025 (end), respectively.

The data collection would involve detailed information that would be gathered from households in both treatment and control areas. This information would include specific questions on the intervention (knowledge and awareness, methods used, implementation timing, perceptions, etc.) as well as the current practices and use of climate information, wetlands and livelihood coping strategies in the communities and households (to distinguish collective action from individual behaviour). Finally, to assess impact on the households in general, there will be an elaborate consumption module, gathering information on food and non-food items, nutrition and health indicators, savings and income level.

The second type of data would be related to the wetland restoration itself. This would be collected using high-quality satellite data (Landsat or Corona are largely used in the literature) for the forest/vegetation cover over the two groups.[28] We will also use satellite data to establish surface water cover in these regions and the impact of the intervention over time. Optical and Synthetic Aperture Radar imagery from satellite platforms provide the means to discretely map surface water and to observe trends in it over time. The advantage of using these satellite images is to be able to provide much more frequent feedback on the wetland restoration to the implementation team, and therefore to transfer knowledge on adaptation of best practices much faster than if only relying on survey data.

Secondary data

Although the workshop did not deal with the monitoring data evaluation, discussions with Stephen Baguma, the UNDP monitoring and evaluation (M&E) specialist, made it clear that the project has developed a monitoring plan in the M&E plan for the project and that the LORTA team can provide support for these monitoring activities. We hope that the analysis of the monitoring data will be another source of "real-time" assessment that the implementing agencies can make use of. This will help establish whether implementation is timely and efficient, has a gender-sensitive approach

[28] These Landsat images capture extremely high-resolution data, with each pixel capturing 30 metres on each side, and are able to detect small changes in the forest cover. For more details, refer to Hansen et al. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853. DOI: 10.1126/science.1244693here: http://science.sciencemag.org/content/342/6160/850.

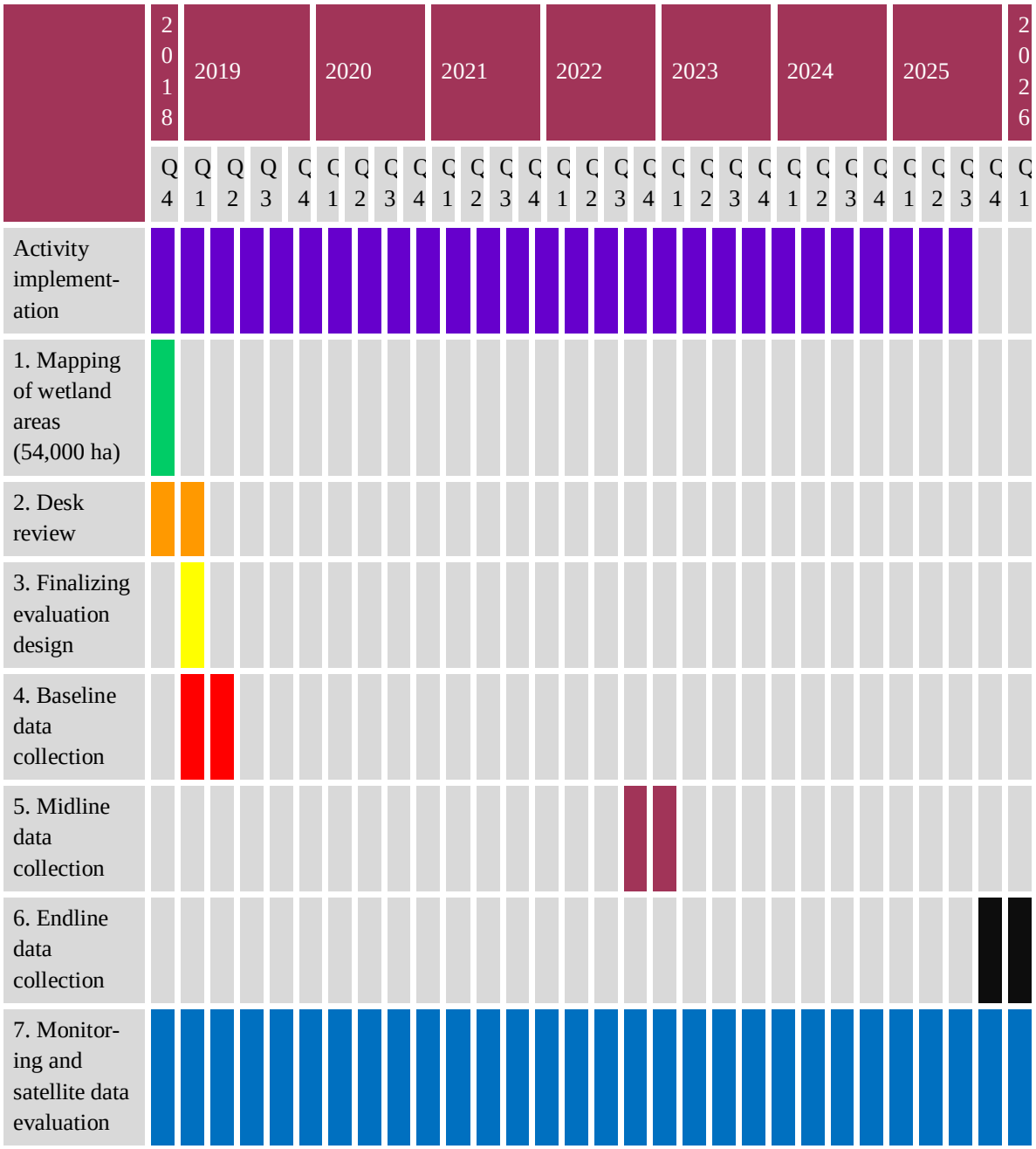
(training of women farmers and female community members), if there are improvements in the outcome levels and, finally, if midterm targets are being met.

Timeline of evaluation

As can be seen from the timeline of the evaluation in .

Table 2, multiple activities – namely, desk review and finalizing evaluation strategy and design (based on the mapping) – are being performed in parallel during the last quarter of 2018. The LORTA team is coordinating with UNDP, MoWE, MIAAF and UNMA to ensure an effective implementation of the programme and evaluation activities – one in line with the IE design. To our knowledge, at time of writing, no consultancy has been hired for the data collection, and these are all details to be left for Phase II of the LORTA programme.

Table 2 Evaluation and implementation timeline



The LORTA team identified the type of support activities needed by the Uganda project team. The team did not have any trained economists or methods experts (of IE tools) in their programme team. Nevertheless, there was a pronounced understanding of the need for IE as a result of the LORTA workshop in Thailand in July and a very good understanding how IE differs from monitoring processes and the concepts of the theory of change.

In such a context, we believe that the LORTA workshop within the mission substantially benefited all the stakeholders involved, including the UNDP, the MoWE, MAAIF, UNMA and other implementing governmental and non-governmental bodies, in bringing the needed formative engagement. It was also our suspicion that the MoWE was sceptical of the success of the IE and might even have misapprehended this to be detrimental to the success of the programme itself. While many of these doubts were allayed during the workshop, it might be that another workshop is necessary to ensure a deeper understanding of how IE will not be detrimental to or hinder progress on the programmatic side.

The presence and support of Dr. Abidoye in the LORTA mission and overall programme has been crucial in facilitating coordination and understanding among the LORTA team and the wetland restoration project implementers and stakeholders.

Overall, we consider the wetland restoration programme to be in need of support for IE. The LORTA team, in coordination with Dr. Abidoye, is continuing to provide inputs and feedback remotely.

III. WAY FORWARD

Overall, we consider the LORTA mission in Uganda to be well received and that it produced promising results for the IE.

The UNDP team was crucial in organizing the workshop and setting up all the meetings with the key stakeholders. Furthermore, Dr. Babatunde Abidoye designed the initial IE design and was crucial in managing the operational and implementation discussions with the stakeholders as well as in the operational discussions with the implementing agencies.

However, the future success of this project in the LORTA framework is highly conditional upon a continuous responsive collaboration from UNDP and other key stakeholders involved. The results from the mapping, for instance, will have to be shared with the LORTA team as soon as they are available, so the LORTA team can work with UNDP and the project team to finalize the IE design.

APPENDIX I

Table 3 Agenda of Uganda LORTA field mission

DATE /TIME	ACTIVITY	VENUE/ LEAD
Sunday, 7 October 2018	Arrival of the C4ED Team at Entebbe, head for Kampala	Check-in into hotel – The Mackinnon Suites, Kampala
Monday, 8 October 2018		
9.00–10.00	Entry meeting at the United Nations Development Programme	Main conference room UNDP
10.00–12.00	Project team meeting, documents appraisal for implementation and evaluation, budget for evaluation, etc.	UNDP
12.00–13.00	Lunch Break	
13.00–14.00	Theory of change discussion with project team	UNDP
14.00–15.00	Discussion on M&E plan	UNDP
15.00–15.30	AOB including discussion for field visit tomorrow	UNDP
Tuesday, 9 October 2018		
Field visit – Limoto, Pallisa District		
Departure at 7 a.m.		
Wednesday, 10 October 2018		
9.00–10.00	Courtesy call to the Permanent Secretary	Ministry of Water and Environment
10.00–12.00	Meeting with the implementing partner and Project Management Unit	Ministry of Water and Environment
12.00–13.00	Meeting with the NDA, Implementing partner, Responsible parties and Project Management Unit	Ministry of Finance, Planning and Economic Development
Thursday, 11 October 2018		
8.00–9.00	Arrival and registration of participants	Ridar Hotel PMU
9.05–9.15	Welcome remarks	Ridar Hotel MoWE
9.15–9.30	Opening remarks	Ridar Hotel UNDP

DATE /TIME	ACTIVITY	VENUE/ LEAD
9.30–10.30	Introduction of LORTA mission and C4ED team	Ridar Hotel Atika Pasha
10.30–11.00	Health Break	Hotel
11.00–13.00	Discussion on the Project Theory of change	Ridar Hotel Katharina Richert (C4ED)
13.00–14.00	Lunch Break	
14.00–16.00	The benefits of impact evaluation – discussion and presentation	Ridar Hotel Katharina Richert (C4ED)
16.00–16.30	Evening Tea	
Friday, 12 October 2018		
8.30–10.30	Impact evaluation design for the wetlands restoration project	Ridar Hotel Babatunde Abidoye
10.30–11.00	Health Break	
11.00–12.15	Impact Evaluation – How can it fail?	Ridar Hotel Atika Pasha (C4ED)
12.15–13.00	Overview of the Project M&E plan	Ridar Hotel Stephen Baguma
13.00–14.00	Lunch break	
14.00–15.00	Debriefing: Presentation to the project team; key elements and timeline of design	

Appendix II

Table 4 List of stakeholders engaged with during Uganda LORTA mission

UNDP meeting first day, October 8	UNDP meeting first day, October 8
Name	Name
Abidoye, Babatunde	Abidoye, Babatunde
Baguma, Stephen	Baguma, Stephen
Kanyike, Tom	Kanyike, Tom
Malwoga, Jascinta	Malwoga, Jascinta
Muhweni, Onesimus	Muhweni, Onesimus
Mujuni, Godfrey	Mujuni, Godfrey
Omodo McMondo, Daniel	Omodo McMondo, Daniel
Zaake, Benon	Zaake, Benon
LORTA mission field trip to Pallisa district, October 9	LORTA mission field trip to Pallisa district, October 9
Name	Name
Samuka, Mohammed	Samuka, Mohammed
LORTA mission meeting at Ministry of Water and Environment, October 10	LORTA mission meeting at Ministry of Water and Environment, October 10
Name	Name
Abidoye, Babatunde	Abidoye, Babatunde
Baguma, Stephen	Baguma, Stephen
Barugahare, Vincent	Barugahare, Vincent
Iyango, Lucy	Iyango, Lucy
Kambalho, Irene	Kambalho, Irene
Malwoga, Jascinta	Malwoga, Jascinta
Mujuni, Godfrey	Mujuni, Godfrey
Mununuzi, Nathan	Mununuzi, Nathan
Nomyondo, Irene	Nomyondo, Irene
Nshemekolwe, Lauben	Nshemekolwe, Lauben
Omodo McMondo, Daniel	Omodo McMondo, Daniel
Stebtoto, Asaddo	Stebtoto, Asaddo
LORTA Workshop, Kampala, Ritar Hotel, October 11 and 12	LORTA Workshop, Kampala, Ritar Hotel, October 11 and 12
Name	Name
Abidoye, Babatunde	Abidoye, Babatunde
Akatwisuka, Rogers	Akatwisuka, Rogers

Amo, Farimu	Amo, Farimu
Anguti, Silas	Anguti, Silas
Arineubabagi, Ave	Arineubabagi, Ave
Awekorlimungu, Margaret	Awekorlimungu, Margaret
Baguma, Stephen	Baguma, Stephen
Barugahare, Vincent	Barugahare, Vincent
Businge, Damien	Businge, Damien
Gokaka, Goeffrey	Gokaka, Goeffrey
Gwebatak, Moses	Gwebatak, Moses
Ituika, Gilbert	Ituika, Gilbert
Iyango, Lucy	Iyango, Lucy
Kabaalu, Deo	Kabaalu, Deo
Kafulu, Twaha	Kafulu, Twaha
Kambercha, Irene	Kambercha, Irene
Katto, Andrew	Katto, Andrew
Kyambadde, Richard	Kyambadde, Richard
Malwoga, Jascinta	Malwoga, Jascinta
Menya, Hakim	Menya, Hakim
Mujabi, Sarah	Mujabi, Sarah
Mujuni, Godfrey	Mujuni, Godfrey
Mununuzi, Nathan	Mununuzi, Nathan
Ojok, Tonny	Ojok, Tonny
Olaya, Collins	Olaya, Collins
Omodo McMondo, Daniel	Omodo McMondo, Daniel
Paira, Christopher	Paira, Christopher
Stebtoto, Asaddo	Stebtoto, Asaddo
Tebandeke, Andrew	Tebandeke, Andrew
Tunyatunga, Patricia	Tunyatunga, Patricia
Waswo, Morton	Waswo, Morton

IMPACT EVALUATION DESIGN REPORT 4: PARAGUAY

TABLE OF CONTENTS

TABLES

FIGURES

I. INTRODUCTION

A. THE LORTA PROGRAMME

Evaluating the impact of development projects and programmes has gained importance in recent years. Impact evaluation not only allows for increased transparency by measuring the effects of investments, it also provides the opportunity to design and implement development projects more effectively. To contribute to this development, the Independent Evaluation Unit (IEU) of the Green Climate Fund (GCF) has started the Learning-Oriented Real-Time Impact Assessment (LORTA) programme to be able to keep track of GCF projects in terms of performance and results, and to enhance learning within the GCF.

The LORTA programme aims to

- Embed real-time impact evaluations into funded projects so GCF project task managers can quickly access accurate data on a project's quality of implementation and likelihood of impact, and
- Build capacity within projects to design high-quality data sets for overall impact measurement.

The purpose of the impact evaluations is to measure the change in GCF key result areas that can be attributed to project activities. The LORTA programme will inform on returns on GCF investments and help GCF projects track implementation fidelity. The objectives of LORTA include the following:

- Measuring the overall change (outcome or impact) of GCF-funded projects and enhancing learning
- Understanding and measuring results at different parts of theories of change
- Measuring the overall contribution of the GCF to catalysing a paradigm shift and achieving impacts at scale

Currently, the LORTA programme is in Phase I (formative engagement and design). In the first year (2018), the IEU will support eight GCF projects to build high-quality, theory-based impact evaluation designs at inception. Formative work will include engaging with project teams, accredited entities (AEs) and GCF staff; designing theory-based impact evaluations; and establishing protocols for database development.

The second phase of LORTA will involve the main impact assessment stage (3–5 years), while the third phase will include analyzing baseline and endline data, discussing results, and engaging with diverse stakeholders to share results and incorporate feedback as required.

B. THE PARAGUAY PROJECT

The Paraguayan "Poverty, Reforestation, Energy and Climate Change Project" (PROEZA), run by the *Secretaria Técnica de Planificación del Desarrollo Económico y Social* (STP) of the Government of Paraguay, has been selected to be part of the LORTA inception stage (Phase I). The Spanish word *proeza* translates into English as "big effort, feat". PROEZA aims at encouraging sustainable agroforestry development and improving the resilience of highly vulnerable households in the Eastern provinces of the country. The project consists of three components.

i) Component 1: "Planting for the Future", targeting agricultural households

ii) Component 2: "Sustainable Landscapes and Responsible Markets", targeting medium-size private land owners

iii) Component 3: "Good Governance and Law Enforcement", targeting public institutions

Component 1 was identified as suitable for a rigorous impact evaluation and was therefore the focus of the mission and this design report.[29]

Deforestation rates in Paraguay are among the highest worldwide, with an annual rate of 1.7 per cent in the past 15 years. Agricultural expansion and the dependency on biofuels have been identified as the main drivers of deforestation in the country, as still more than 40 per cent of the national energy consumption is based on biomass and firewood. The use of firewood is particularly high among the poor. On average 95 per cent of the poor households in the project area rely on firewood and charcoal for cooking and spend around USD 270 per year on the acquisition of such. Most households (88 per cent) still cook on an open fire and do not own a cooking stove, even though stoves generally increase the energy efficiency of cooking. The consequences of deforestation are greenhouse gas emission, soil erosion and disruption of the water cycle, which lead to an increased vulnerability to climate events such as floods and droughts.

Component 1 of the PROEZA project aims at tackling deforestation by 1) encouraging the reforestation of land and 2) decreasing the demand for firewood. Two interventions were designed for this purpose.

Intervention 1 provides smallholder farmers with training on climate-smart agroforestry production systems and intensive technical assistance to establish such systems in follow-up visits. This assistance includes the payment of wages for additional workers and agricultural inputs such as tree saplings, fertilizer, and the like. Six types of climate-smart plantation systems are introduced. One combines native tree species with exotic species that are fast growing and can be used for biomass production much earlier than the native species. Another model mixes native or exotic species with yerba mate plants that benefit from the shade the trees provide. Examples of such plantation plans can be found in Figure 6 in appendix IV of this design report.

The climate-smart plantation systems are expected to generate income to the households after three years. To compensate for the opportunity costs of land usage at the start of the project, participating smallholder farmers receive one cash transfer per year, conditional on the adoption and maintenance of the new production system. The first payment is made after the preparation of the soil, the second upon planting of the trees and the third when at least 80 per cent of the trees have started to grow. The conditions of the fourth and fifth payments are related to the maintenance of the plantation system. These payments are known as environmental conditional cash transfers (E-CCTs) and their size depends on the plantation scheme the farmer decides to adapt, depending on individual preferences and soil quality. At this point, the details on the implementation are not yet specified. It is unclear how the trainings will be given (whether at community or household level), how often the farmers will receive follow-up visits and what the exact payment modalities are.

Intervention 2 targets the demand side of biofuel by providing households with efficient cooking stoves that require less firewood or charcoal. These cooking stoves are built directly in the beneficiary's kitchen and are designed such that the smoke is channelled outside, thereby reducing indoor air pollution.

The beneficiaries of component 1 are poor and extremely poor households in eight departments in Eastern Paraguay: Alto Paraná, Caaguazú, Caazapá, Canindeyú, Concepción, Guairá, Itapúa and San Pedro (see map in Figure 5 in appendix III of this design report).[30] In total, the project aims at

[29] Component 2 entails concessional credit for only about 200 medium-sized landowners. Due to its small sample size, no quantitative impact evaluation can be conducted. Component 3 aims at institutional capacity-building, which is also not suitable for an impact evaluation due to the obvious lack of a counterfactual.

[30] The definitions of poor and extremely poor follow a multidimensional approach called the "Quality of Life Index". See appendix V of this design report for a detailed composition of this index.

reaching 17,100 households for the training and E-CCT interventions and 7,500 households for the cooking stoves.[31] Both interventions will build on an existing social programme called *Tekopora*. *Tekopora* is also a conditional cash transfer (CCT) programme but with a focus on health and education of women and children. Households that currently, or in the future, participate in the *Tekopora* programme and own at least 0.8ha of suitable land for plantation will be eligible for component 1.

The GCF-funded project period is five years. Due to the time-intensive screening for eligibility (to determine the amount of land owned by each *Tekopora* household), the training and E-CCT will be rolled out in five phases. In the first phase, starting in March 2019, 570 households will be treated, of which half will belong to indigenous communities. In the second year, an additional 2,850 households will receive the training, and in years 3–5, 4,560 households will be added each year (see Figure 1). No specific roll-out schedule for the cooking stove intervention has been defined so far.

Figure 1 Timeline of intervention 1 roll-out

While the STP is the national designated authority (NDA), six additional governmental entities form part of the steering and the executive committee: the Paraguayan Institute for the Indigenous (INDI), the National Forestry Institute (INFONA), the Ministry for Social Development (SAS), the Ministry of Agriculture and Livestock (MAG), the Environmental Secretary (SEAM) and the Vice Ministry of Mines and Energy (VMME). The executing agency – thereby responsible for the coordination, technical implementation and monitoring and evaluation (M&E) – is the Food and Agriculture Organization (FAO). One or more private sector companies will be contracted to deliver the training and support to the farmers in the implementation.

II. PARAGUAY FIELD MISSION

A. GENERAL REMARKS

An evaluation team, consisting of two consultants from the Center for Evaluation and Development (C4ED), Esther Heesemann and Michaela Theilmann, and a consultant from the IEU, Nathan Fiala, was formed to lead the field mission from 18 to 24 October 2018. The task of the evaluation team (henceforth called the LORTA team) was to engage closely with key stakeholders of the PROEZA project – namely, the NDA, the AE, implementing agencies, project staff and potential end beneficiaries – to ensure their interest in and understanding and sense of ownership of the planned theory-based impact evaluation.

During the field mission, the LORTA team held multi-stakeholder meetings and capacity-building workshops with the representatives of FAO and the seven government entities involved (INDI, INFONA, MAG, SAS, SAEM, STP and VMME).[32] The multi-stakeholder meetings served to create a common ground of knowledge on the project specifics and to facilitate understanding of the roles and interests of the individual parties. The capacity-building workshop conveyed basic knowledge on impact evaluation (how it differs from M&E, the importance of a counterfactual and power, experimental evaluation methods) and theory of change development. This knowledge built the

[31] It is not specified in the project design whether the 7,500 cooking stove households will be a subsample of the 17,100 households receiving training and E-CCTs or will be taken from a separate sample.
[32] A detailed list of all participants can be found in Table 5 in appendix II of this design report.

basis for a later discussion on PROEZA-specific evaluation designs – namely, simple lottery or phase-in designs. Together with the workshop participants, the LORTA team conducted context analyses, examined the existence of appropriate counterfactuals (i.e. comparable treatment and control groups), elaborated a theory of change, assessed the availability of baseline administrative and secondary data sources, and acquired budget information. The results from this undertaking are presented in the following sections.

In addition, during a one-day field visit, the LORTA team, along with representatives of FAO and STP, met with potential PROEZA participants and learned about their livelihoods, agroforestry and the role of the forest. These potential participants included two farmers and an indigenous community.

B. THE MISSION AGENDA

The LORTA team experienced a high degree of collaboration and support from all involved institutions, especially the STP, who went to great effort to provide a smooth and productive week's work. The agenda – shown in Table 4 in appendix I of this design report – was developed to facilitate the joint attendance of key stakeholders at the LORTA workshops, in particular the presence of an impact evaluation expert from FAO Rome, and to allow for the field visit, all within the week that the mission was meant to take place. The FAO impact evaluation expert, Silvio Daidone, supported the mission in the first three days.

Given that a large number of high-level representatives were present for the opening of the mission, the presentation on the benefits of impact evaluation was held on the first day. The purpose of day 2 was for the LORTA team to gain a better understanding of the planned project activities and the rationale behind them. For this reason, the STP started by presenting the PROEZA project and the contextual background. In preparation for the later discussion on project outcomes and indicators, the LORTA team conducted a brief theory of change workshop. However, the intended interactive development on the theory of change proved to be difficult because the participants were not yet clear on the main outcomes of the project. For this reason, further detailed discussions of the theory of change were postponed to day 4. The field visit took place on day 3. The LORTA team was accompanied by representatives of the STP and FAO and for part of the day by the local chief of the MAG and INDI representatives. The purpose of the field trip was to learn about the livelihoods of the potential beneficiaries, both common agricultural households and the indigenous communities. On day 4, preliminary decisions were taken on the design, theory of change, indicators and evaluation questions. Day 5 started with a discussion about sample size and power calculations, in order to demonstrate the need for a large number of clusters and participants. Furthermore, we discussed the existing M&E system with INFONA – in particular, the current utilization and availability of satellite images. In a later meeting with all stakeholders, the FAO representatives presented their M&E plan for the PROEZA project. On day 6, the LORTA team gave a final summary of the mission results before the mission was officially closed by the Minister of the STP. The last point on the agenda was a small group meeting with the INDI representatives, in which the possibilities for the inclusion of indigenous communities in the impact evaluation were discussed.

C. THEORY OF CHANGE

The theory of change was prepared in intensive group discussion with representatives of all ministries and secretaries involved and FAO Paraguay. We developed separate theories of change, indicators and evaluation questions for each of the two interventions. The results of this discussion are presented below.

Intervention 1: Training and E-CCT

i) Inputs

Inputs to the project will be the funds provided by the government, the GCF and other donors to pay for the saplings, the CCTs and the workforce that delivers the trainings on how to adopt the new planting systems.

ii) Activities

The project implementers will give group and/or individual training on the new planting systems to the farmers, distribute saplings and provide further technical assistance in follow-up home visits. Conditional on complying with the plantation system specific goals, payments will be made to the farmers' bank account or cell phone.

iii) Outputs

The activities will lead to an increased number of farmers trained and equipped to start climate-smart plantations on their fields, and they will have financial incentives to launch and maintain them.

iv) Outcomes

The immediate outcomes of the activities are the planting of trees on farmland and the maintenance of such over the project period. Assuming that the farmland was not used for agroforestry before, the adoption of the new plantation systems will increase the overall forest area in the treatment communities. The new plantation systems will also increase the farmers' agricultural diversification – again, depending on the previous land usage. Furthermore, households' income is expected to increase through the cash transfers and the selling of cash crops grown on their land. Taken together, the effects on income and diversification will improve the food security of the beneficiary households.

The agroforestry plantations are further expected to decrease the crop yield variability by improving or maintaining soil quality over time. This in turn translates into reduced income variability, in particular for those households whose main income source is agriculture.

i) Goals

The ultimate goals are the mitigation of and adaptation to the effects of climate change and the increased resilience of households.

Evaluation questions

- Are farmers changing their agricultural production towards climate-smart plantation systems?
- Does PROEZA increase forest coverage?
- Will PROEZA households experience a lower degree of yield variability and have more stable income flows?
- Are PROEZA households more resilient to weather events?
- Does participation in PROEZA lead to a diversification of agricultural production? [secondary outcome]
- Does PROEZA increase the participants' income? [secondary outcome]
- Does PROEZA increase the participants' food security? [secondary outcome]

Preliminary list of impact indicators

- Share of farmers who grow native or exotic trees on their land
- % decrease in deforestation

- % increase in forest volume
- Variability in yields
- Variability in income
- Impact of weather shocks on household assets, income and expenditure
- Recovery from weather shocks
- Number of crops
- Type of crops
- Household income from agriculture and agroforestry[33] [34]
- Composition of agricultural income
- Food Insecurity Experience Scale
- Household food diversity index

The most important assumption underlying the theory of change is the overall willingness and ability of the smallholder farmers to adapt to the new plantation systems. The targeted farming households own on average only 2–3ha of land. An INFONA extension worker reported a reluctance of the farm owners to plant trees on their land due to their slow growth compared to other crops. Moreover, premature cutting of trees is common to satisfy immediate financial needs. To motivate the farmers to plant and maintain the trees, the financial incentive of the E-CCT must be sufficiently high to overcome those barriers. We will critically assess the existing incentive structure and explore the possibility of conducting microsimulations to define the optimal amount of payment.

Another risk factor to the success of the project is the issue of secure land rights. Without secure land rights, farmers are not in full control of the plantation and thus have less incentive to invest in the long-term benefits of the new plantation systems. Having identified a third of potential beneficiaries who lack formal land titles, the project unit plans support for the households in the formalization process. Moreover, the observation of the entitling process will be of great importance when identifying the mechanisms through which PROEZA impacts agricultural productivity.

Whether PROEZA will in fact lead to increased agricultural diversification and higher income from agricultural production depends strongly on the status quo of agricultural production. If households commonly plant a variety of crops, it is unlikely that an increase in diversification will be detected. It is further assumed that the agricultural outputs of the new plantation systems lead to higher profits than the farmers' previous plantation system. As this is not entirely clear from the project design, we consider income and agricultural diversification only as secondary outcomes of PROEZA.

Intervention 2: Improved cooking stoves

i) Inputs

Inputs to the projects will be the funds provided by the government, the GCF and other donors to pay for the materials for the cooking stoves and the workforce to deliver the intervention.

ii) Activities

The project implementers will build cooking stoves that use firewood more efficiently in the beneficiary's home.

iii) Outputs

[33] It is not clear a priori whether the net income effect from the new plantation scheme will be positive, negative or null due to the alternative land usage and related opportunity costs.
[34] Income from agriculture comprises the revenues that come from selling products of vegetable origin, including crops for human consumption, wood and all sorts of plant-based by-products, but excludes other sources of income such as selling livestock products.

This results in households have functioning, improved cooking stoves in their homes.

iv) Outcomes

The short-term outcome of the intervention is the use of the improved cooking stoves. This will lead to reduced consumption of firewood and hence lower expenditure on firewood. The improved cooking stoves additionally reduce indoor air pollution compared to the traditional open-fire cooking practices, which will lower the incidence of respiratory diseases, especially of women and children, who are most exposed to the pollution.

v) Goals

The ultimate goals are the mitigation of climate change, increased resilience of households and improved population health.

Evaluation questions

- Do more households use efficient cooking stoves?
- Does PROEZA lead to a decrease in firewood consumption?
- Do participating households have lower expenditure on firewood?
- Do women and children in participating households have fewer respiratory problems?

Preliminary list of impact indicators

- Share of households cooking with efficient cooking stoves
- Amount of firewood used for cooking at household level
- Expenditure on firewood at household level
- Symptoms and diagnoses of Acute Respiratory Infection of women and children
- Expenditure on health care

During the discussions, it was pointed out that gender should be an aspect covered in the evaluation. For this purpose, heterogeneous effects of households with a female head of household will be identified using the aforementioned indicators for intervention 1 and 2. Moreover, this exercise will be applied to other vulnerable groups detected during the first phase of implementation, subject to sample size requirements.

D. PLANS FOR EVALUATION

Evaluation design

The evaluation design developed by the LORTA team and the implementing organization is a phased-in randomized control trial (RCT) that is clustered at the neighbourhood level. The interventions to be delivered as part of this project will go to individuals who are currently receiving another CCT from the government. They will thus be familiar with the concept and have experience with receiving funds from the government. A total of 60,000 eligible households have been identified for the programme, but there will only be enough resources for 17,100 to receive the E-CCT and for 7,500 to receive the cooking stove. This provides a large enough population of beneficiaries for an impact evaluation design.

The step-wise roll-out of the intervention over the course of five years presents an ideal set up for a phase-in design (see Figure 1). Furthermore, the local partners preferred the phase-in design over the simple lottery design, in which the control group would not be able to participate in PROEZA throughout the entire evaluation period. Due to the rather long roll-out period, we will be able to observe medium-term effects of the project. This is essential as the aim of the project involves the planting and growth of trees, which can be measured in a meaningful way after three years.

The design as currently discussed is presented in Figure 2. The study population will consist of 300 neighbourhoods in the project area. Ideally, those neighbourhoods will be randomly drawn from the total project population such that the results have a high external validity. The neighbourhoods will then be randomly allocated to the control group, the training and E-CCT treatment group, and the training, E-CCT and cooking stove treatment group.[35] Power calculations for the same are discussed below. This design makes it possible to identify the combined impact of the training and conditional transfer, as well as the additional impact of the improved cooking stove component.

A second design was also discussed, as presented in Figure 3. In addition to being assigned to treatment or control, communities could be randomized by the intensity of treatment. In low-intensity treatment communities, 50 per cent of the eligible households are allocated to the treatment group and the remaining 50 per cent to the control group. This study design would make it possible to determine the impact of spillovers from the interventions, and potential rebound effects. A potential positive spillover effect would be the transfer of knowledge between households – that is, farmers that implement the climate-smart plantation systems without having been directly targeted by the project. We expect such spillovers to be more likely within neighbourhoods than across neighbourhoods. A potential rebound effect might occur as a result of the decreased demand for firewood from the cooking stove intervention. In a functioning market, the reduced demand leads to a decrease in prices, which might trigger the demand for the product again. Having different treatment intensities in place would make it possible to observe such price effects.

Whether or not the treatment intensities will be varied is currently under discussion with the government. During the mission, concerns were raised that the unequal treatment of households within the same neighbourhood could lead to social tensions. There is consensus that this design will work in large communities; however, it might not work in smaller communities. The final design will thus depend on how large the selected communities are (this is discussed more in the data collection section below).

Note that the final sample size under this design requires 500 communities. It nevertheless only differs from the design involving 300 communities in the requirement of having two groups (high intensity and low intensity) within the training and E-CCT treatment arm and the training, E-CCT and cookstoves treatment arm.

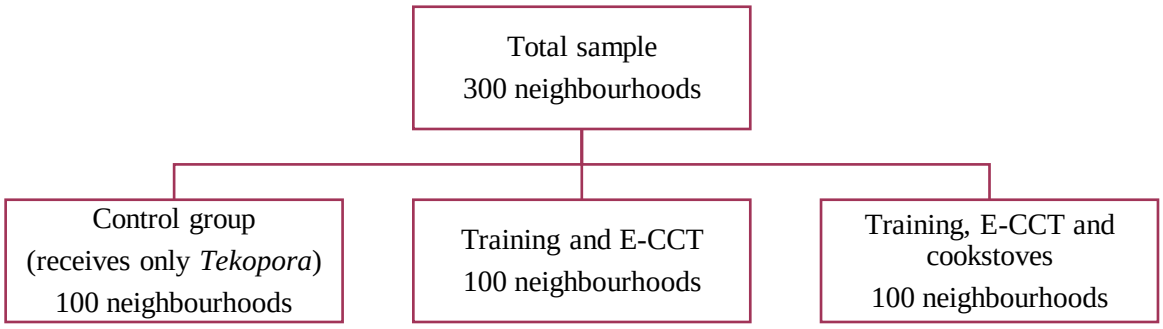


Figure 2 Evaluation design

[35] In total, there are 1,389 distinct localities in the project area. According to the current numbers provided by the government, in each neighbourhood there are on average 5.6 households eligible to participate.

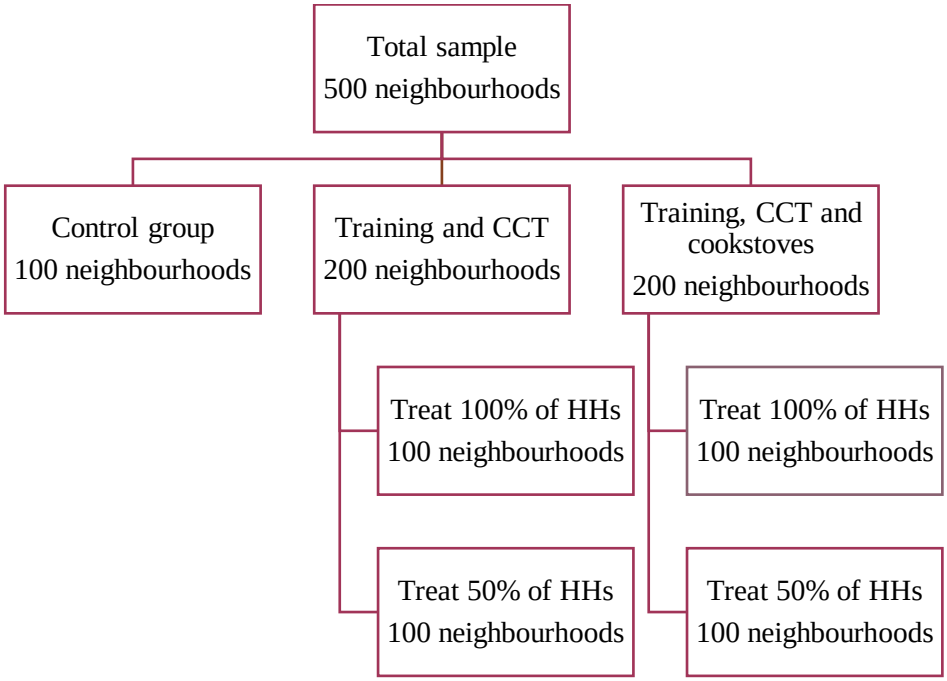


Figure 3 Additional evaluation design

This impact evaluation will comply with rigorous international research standards. We will submit a pre-analysis plan prior to the endline, which will lay out precisely which research question we intend to answer and the methodology we intend to apply. In addition, we will obtain ethical clearance from institutional review boards and register the study at the RCT registry of the American Economic Association. Moreover, if the impact of PROEZA on several outcomes is estimated, we will adjust for multiple hypothesis testing.

Power calculations

Power calculations are to be determined but will consider the number of clusters and heterogeneity in the communities. At this time, we expect there to be up to 100 clusters (neighbourhoods) in each of the treatment and control groups. For the simple model described in Figure 2, this will mean 300 clusters in total.

The first main outcome of interest will be the impact of the interventions on changes in local forest coverage. The project is also interested in other outcomes, described above in the theory of change and outcomes discussion. As we do not have information on these outcomes yet, we present a simple example of power in Figure 4. In this model, we assume 200 clusters (100 control and 100 in each treatment group), an ICC of 0.10 and a sample of 10 households per cluster. Under these assumptions, the study will be able to identify impacts of at least 17 per cent. Discussions with the Government of Paraguay suggest this is a reasonable minimum detectible effect size.

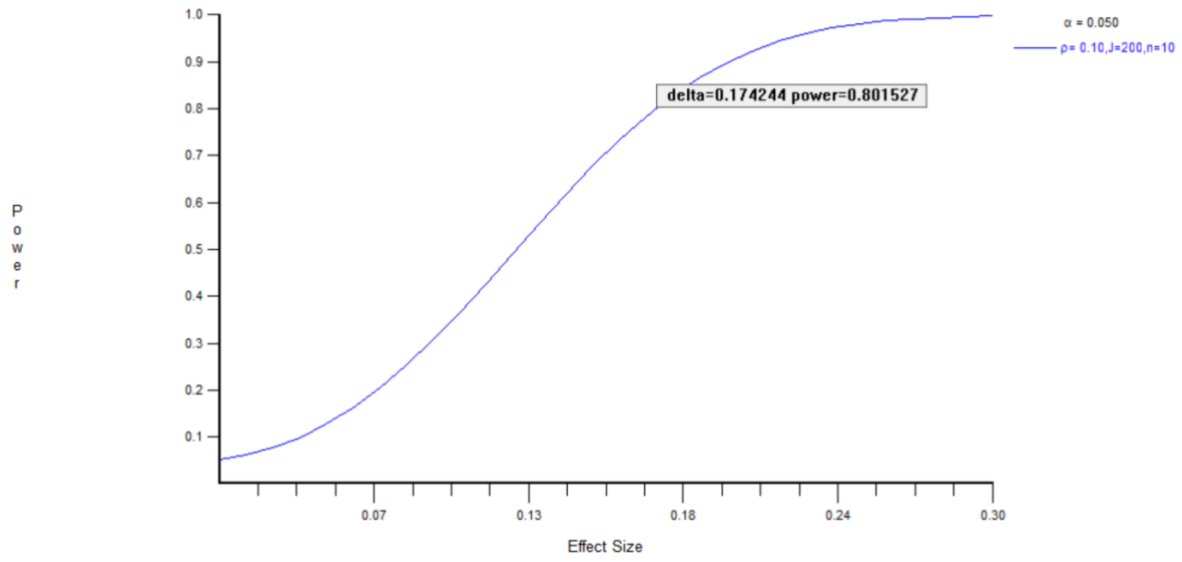


Figure 4 Power calculations for alpha of 5 per cent

Note that these calculations assume a take-up rate of 100 per cent. We do not expect perfect take-up, however. Similar studies have found take-up rates of well below 50 per cent. We believe that the experience of people with the *Tekopora* programme will make take-up significantly higher, but still not perfect. As we describe in the next section, it will be important to use the experience from the first year of the project to identify take-up rates. The numbers in this study will thus be weighted by this take-up rate, and so increased to ensure the same level of power. This will fit with the goals of the project, as PROEZA will need to not just offer the project to 2,850 households but needs a final take-up of 2,850 households. The needs of the evaluation and the project are thus closely aligned.

Timeline of evaluation

The project will roll out starting in early 2019. The plan is to work with 570 households in the first year. This number of households is too few to conduct a rigorous evaluation. We thus plan to use the first year to observe the implementation of the project at a small scale, and pilot the data-collection activities. Take-up is of critical concern for a project like this, and understanding the take-up rate will be important for the evaluation in calculating the final sample size. The first year will thus allow for observing take-up rates.

Furthermore, this "pilot" is convenient for testing and refining survey instruments and outcome measurements. For example, parts of the baseline questionnaire can be applied on these households and medium- and high-resolution satellite imagery can be obtained to observe the feasibility of the observation of changes in forest area. Given that the implementation of the project as well as the impact evaluation will be adapted to suit the livelihood strategies and customs of indigenous communities, the first year will be very valuable in observing the implementation of PROEZA in those particular communities.

The preliminary proposed timeline is presented in Table 1.

Table 1 Proposed timeline of evaluation

2018 Q4	Finalize the designs of the training, eligibility criteria, M&E framework and impact evaluation design. C4ED will also explore the existing secondary data to determine if it can be used for baseline measures.
2019 Q1	The project will start with 570 households. These households will be randomly selected from the data set that PROEZA has compiled. The GCF and C4ED will assist PROEZA in learning about the implementation quality of the project and take-up rates, as well as in piloting the use of satellite imaging for tracking changes in tree coverage.
2019 Q2	Finalize the eligible sample for the full implementation.
2019 Q3	Baseline data collection (if the existing secondary data is not of sufficient quality and coverage to be used as baseline data).
2019 Q4	Randomization to select treatment and control groups.
2020 Q1	Start the interventions in 2,850 treated households.
2023 Q1	Endline data collection on households.
2023 Q2	PROEZA will be rolled out in control communities.

Concerns

Several issues arose during the field visit and discussions with the government. In this section, we discuss the major issues and our proposed solutions.

The M&E plans developed by FAO are quite ambitious. They entail a close forest coverage monitoring via satellite images as well as monitoring of the vulnerability of the participating households with household surveys. For instance, it is planned to geo-reference the plots of each of the 17,100 participating households. Furthermore, in order to make the cash transfer to the complying households, the fulfilment of the conditions needs to be checked regularly and reliably. Despite the large scope of M&E activities, the details of the realization of the plan are not yet decided on. During discussions with the PROEZA team and FAO, several issues arose that need to be resolved. For example, it is not clear yet what kind of satellite images will be used, where they will be sourced from and at what intervals they will be checked. It has also not been decided how the fulfilment of the conditions for the E-CCT will be confirmed and which parties will be involved in this process. The LORTA team emphasized the importance of a common, accessible platform through which all involved entities can upload monitoring data and retrieve information. While the current state of M&E planning is not fully satisfactory, we believe that there is sufficient time in 2019 to resolve these issues. However, the LORTA team should ideally be involved in the further development of the M&E process, especially regarding the use of satellite imaging.

During the field trip we experienced issues with talking to the indigenous community. Many indigenous communities in Paraguay are very closed in order to protect their culture. Conversations with indigenous community advocates also suggest that this group could pose several issues for the evaluation, as land is typically owned by the community instead of by individual households. The variation of treatment intensities as it is currently discussed for the non-indigenous communities is hence not feasible. Without a clear definition of household income and agricultural output, several indicators developed from the theory of change will be difficult to measure. Nevertheless, PROEZA explicitly plans the inclusion of indigenous communities because they belong to the most vulnerable groups in Paraguay and strongly depend on the resources of the native forests. Assessing the impact on this particular social group is therefore particularly interesting and should be part of any

evaluation. However, it is likely that separate methods will be needed for evaluating impacts in these areas.

There are concerns about the state of intervention design because the training has not yet been fully developed. As in most countries, extension services and trainings are not well supported in Paraguay. PROEZA will not work through extension workers and will instead conduct the training using its own project staff. The quality of the training, and the reach of it, will be critical for realizing impacts. For instance, some land may need rehabilitation before planting can be done. It is clear that farmers do not know how to rehabilitate land or use agroforestry methods. The technical assistance for intervention therefore needs to address this issue and build applicable modules into the project design.

Even though the project team has a clear idea on the eligibility criteria for households, it is not yet determined how eligible households will be identified. Screening households for eligibility will be both costly and time-consuming. While sound information for households regarding wealth and land ownership exists, some of the information is likely to be outdated as it was collected in 2014 and 2015. Spot checks on the accuracy of households' socioeconomic characteristics and location information can help to determine the quality of the data and the persistence of the indicators. Whether the size of the land and soil quality are actually suitable for the proposed plantation systems needs to be checked individually either through field visits or via phone calls.

Finally, the government is undergoing personnel changes, which might affect the scope of PROEZA. The new Minister of the STP has shown strong support for the implementation of the project, yet the project unit has raised concerns that certain project characteristics might be adjusted. A stable partnership with and commitment from the government to follow through with the evaluation design, in particular the randomized allocation of neighbourhoods into treatment and control, are crucial for the internal validity of the evaluation. The LORTA team will be in close contact with the local partners, such that relevant changes can be taken into account in a timely manner.

Additional evaluation questions

In addition to the two models described above in Figure 2 and Figure 3, a secondary evaluation question was raised by the LORTA and project teams. There is interest in developing an accountability and feedback intervention for/within the communities. This idea is still under discussion, but the project would like to determine short-run learning about how to increase take-up and utilization of PROEZA. One option that was discussed is to experimentally evaluate the impact of providing information to communities, about how well they are performing in PROEZA (measured through payments to the community), relative to other communities in the area. Previous research suggests that providing information to people about how they are performing relative to others can improve behaviour, especially behaviour that has broader social or environmental outcomes. If this were to happen, it would be conducted as a cross-cutting design across the project and interventions described above.

Data collection

Data sources

There is a variety of data sets than can be used for the impact evaluation. First of all, the STP is currently working on generating a database that includes all households that would be eligible for any intervention aimed at poverty alleviation. For this, they currently merge the database from the Ministry of Social Development, the "Ficha Hogar", with the STP's own database, the "Ficha Social". The *Ficha Social* data contain information on household composition (including education and employment), housing characteristics, asset ownership, agricultural activities, area of land

owned and the GPS location of each household. The current data set covers more than 66,000 households in the project area. Approximately 48,000 of those participate in the *Tekopora* programme, of which approximately 33,000 are categorized as poor or extremely poor according to proxy means testing. Due to adjustments of the questionnaire over time, land data are available for only 60 per cent of the households. This enables the government to identify 10,400 eligible households for now. However, there are many more people that are expected to be eligible for the project. There are thus concerns about completeness of data and representativeness. If this database is deemed suitable by the LORTA team, it can be used as baseline data.

Furthermore, the National Forestry Institute already has access to and works with low-resolution satellite images to monitor forest cover and for carbon monitoring. The monitoring with satellite images is accompanied by field visits to inspect the actual forest coverage on the ground. The Institute currently utilizes 10-square-metre-resolution imaging, which is sufficient to spot decreases and increases of forest area and volume but is insufficient to distinguish the specific types of trees grown. Hence it would not make it possible to detect whether the farmers actually followed the plantation systems, which are based on specific types of trees.

There are two additional primary data collections that are planned. The first uses high-resolution satellite imaging. Discussions with the National Forestry Institute proved positive regarding the ability to identify changes in forest coverage and even to count the number of trees in a given area. For the purposes of this study, we will need 1-square-metre resolution. There will be a cost to obtaining high-resolution data. Rough estimates are that this could cost $100 per square kilometre, but this needs to be confirmed.

For the analysis of the satellite data, the support of external GIS experts is needed. The GCF IEU therefore initiated cooperation with AidData, a research lab at William & Mary's Global Research Institute in Washington, D.C. AidData has extensive experience in geospatial impact evaluations and real-time measurement and can support the evaluation team in the data analysis. Additionally, the Paraguayan Forestry Institute can be engaged in this component as well.

Timeline of data collection

As the proposed methodology is an RCT, obtaining baseline data is desirable but not required if randomisation works. The Government of Paraguay is building a database containing eligible households (see description of the *Ficha Social* in appendix V of this design report). These data can potentially be used as baseline. If they are deemed unsuitable and if the budget permits it, baseline data will be collected by conducting a household survey. The baseline survey would take place in the second half of 2019.

For the impact evaluation, an endline data collection is required. Given that the first impacts of the E-CCT on forest coverage can be observed after three years, the endline data collection will start three years after programme implementation (most likely in 2023).

Budget

There are two phases of this evaluation. The first phase, to be conducted in 2019, will work with the Government of Paraguay as they implement the project on 570 households. The goal is to understand the quality of the existing data and take-up rates for the interventions and to determine what imaging model is most appropriate for the full evaluation. During this time, C4ED and the IEU will need to work closely with the government. The specifications and intensity of the engagement of C4ED in the first phase monitoring and learning exercise depend on the budget made available by the project. In addition to C4ED salary costs, we expect the LORTA team to travel to Paraguay once in 2019.

We also recommend that a junior research manager from C4ED visits the government at least two times in 2019. This would entail two short trips (maximum one week each). This person will be critical to maintaining our relationship with the government and to determine the quality of the data and monitoring systems.

The second phase will be the impact evaluation. The cost of the impact evaluation will be determined by what data we would like to use for the study, and the number of households in the study. As stated previously, take-up rates could significantly impact the sample size needed, and so phase one will enable us to adjust these estimations. We estimate sample sizes assuming an 80 per cent take-up rate. We propose two scenarios for measuring impact and discuss their data-collection costs, which are summarized in Table 2 and Table 3.

Scenario one: satellite data with endline survey

The main goals of the project are related to forest coverage. Assuming that we can easily identify the GPS location of households (to be confirmed from the existing data and examined in phase one), then existing satellite data may be sufficient.

It is not yet clear what level of resolution is needed. The government currently has access to 10m x 10m grids, which offer a relatively good resolution and are free to use. However, given the small size of the plots, higher resolution is likely required. A rough estimate from the government, which needs to be confirmed, is that this high-resolution data could cost approximately USD 100 per square kilometre. If we include 5,000 households in the study, each with maximum 10 hectares of land and a buffer area of an additional 5 hectares, this means 100,000 hectares of land area to inspect at the most, or 750 square kilometres, which is USD 75,000 per round of data collection. We believe this is the upper limit. A quick Google search for high-resolution imaging suggests a cost of between USD 25 to USD 40 per square kilometre, or only USD 18,750 to USD 30,000 per data collection. For the smaller sample, this figure is substantially lower.

In addition to forest coverage, there are secondary outcomes of interest at the household level. These include economic outcomes such as income, consumption of firewood and health, which are particularly relevant for the evaluation of the cooking stove intervention. These indicators have not been collected at baseline by the government, but a full baseline survey may not be necessary because of the randomization. When neighbourhoods are randomly assigned to control and treatment groups, there should be no difference on average in household-level outcomes across the groups before the project activities are implemented. Assuming we utilize only endline data for these outcomes, a household survey will need to be conducted in 2023.

Based on previous data collections in Paraguay, we can expect per survey costs of approximately USD 50 to USD 65 per household, depending on the length of the survey. The sample of 5,000 households will thus cost up to USD 325,000 to survey at endline and USD 195,000 for the smaller sample of 3,000 households.

For the primary data collection, the government's own data-collection structures (as used for the *Ficha Social*) could be used, if the data quality is judged as sufficiently high. The LORTA team would then only accompany the data-collection process by providing feedback on the survey instruments, supporting the training, and piloting and monitoring the data-collection process. Alternatively, an independent third party may be contracted. A potential contractor could be IPA Paraguay. While the service costs of a third party are likely to be higher, the quality of data collected will likely be significantly better. An independent survey firm does not have incentives to manipulate data to push the impact of the project, which is a potential significant threat to the internal validity of the evaluation.

Scenario two: satellite data with baseline and endline surveys

A final option would be to conduct a new baseline survey with the sample. This would be done if the existing data are considered inadequate and we want to collect the full range of potential household outcomes. Under this scenario, survey costs will be twice that of scenario one, so USD 390,000 to USD 650,000, depending on the final sample size.

Table 2 Cost of various scenarios, N=5,000

ITEM	NUMBER	REPEATED	COST EACH	TOTAL COST
Scenario 1				
Satellite imaging (km2)	750	2	$25.00	$37,500.00
Household survey (observations)	5,000	1	$65.00	$325,000.00
				$362,500.00
Scenario 2				
Satellite imaging (km2)	750	2	$25.00	$37,500.00
Household survey (observations)	5,000	2	$65.00	$650,000.00
				$687,500.00

Table 3 Cost of various scenarios, N=3,000

ITEM	NUMBER	REPEATED	COST EACH	TOTAL COST
Scenario 1				
Satellite imaging (km2)	450	2	$25.00	$22,500.00
Household survey (observations)	3,000	1	$65.00	$195,000.00
				$217,500.00
Scenario 2				
Satellite imaging (km2)	450	2	$25.00	$22,500.00
Household survey (observations)	3,000	2	$65.00	$390,000.00
				$412,500.00

Available funds

The current M&E budget amounts to USD 450,000, of which USD 300,000 is earmarked for the FAO specific independent evaluation of the entire PROEZA project (components 1–3). We see several overlaps between the impact evaluation data requirements and the planned M&E activities. First, given the nature of CCTs, the compliance with the condition needs to be verified either through satellite images or through field visits. Given the large number of beneficiaries, the use of satellite images seems to be the more feasible scenario. The LORTA team can make use of the same data towards the end of the project for the treatment communities. Additional costs therefore arise only for the acquisition of satellite images for the control communities.

The assessment of household-level indicators will, however, require primary data collection. Depending on the selected evaluation scenario, this will cover one to two waves in 3,000 or 5,000 households. The collection of more detailed household data regarding agricultural activities, risk

behaviour and income flows can also be of great use to the project coordination team, as it will help to assess the targeting success of the intervention and verify the *Ficha Social* data.

In follow-up budget discussions after the inception mission, FAO Paraguay committed USD 300,000–400,000 for the household surveys and satellite data acquisition. This amount roughly covers scenario 1 for 5,000 households and scenario 2 for 3,000 households. We are confident that a high-quality impact evaluation can be conducted with this amount of money.

Assessment of institutional capacity for impact evaluation

There was a broad range in the knowledge of the participants prior to the LORTA mission. The presentations of methods and the capacity-building workshop were well received. Those who did not know about the benefits of a rigorous impact evaluation, and RCTs in general, quickly understood their relevance and supported them. The Government of Paraguay is in the first stages of building its own impact evaluation unit. While the PROEZA impact evaluation cannot benefit from the unit at this point in time, it is very likely that there will have been progress in its development by the time endline data are collected. The creation of this division emphasizes the government's desire and willingness to invest resources into impact evaluations. In addition to the Government of Paraguay, FAO is also highly committed to accompanying and supporting the LORTA team during the impact evaluation.

Plans for monitoring and evaluation

Each institution will be responsible for the monitoring of their own activities. FAO will hire a monitoring specialist who will be responsible for the coordination of all monitoring activities. It is planned to use an online monitoring information system. The project's funding proposal already contains a detailed outline of indicators to be used for the monitoring. External verification of the data collected by each entity will be undertaken.

III. WAY FORWARD

Overall, we believe the LORTA mission in Paraguay was very well received, with all involved institutions keen to see what could be learned from an impact evaluation. Their organization and collaboration enabled the LORTA team to develop a clear understanding of the programme, as well as of the challenges in establishing the method to be used for an impact evaluation. The immediate next step is to communicate the design and potential budgets to STP and FAO, to finalize their buy-in for the impact evaluation and determine what resources can be made available. In addition to the M&E budget, FAO has expressed an interest in potentially supporting the impact evaluation.

In addition, more precise power calculations using existing data from the *Ficha Social*, an overall assessment of the *Ficha Social* with regard to completeness and data quality, and more detailed discussions with the M&E team on the utilization of the budget and potential synergies are needed.

APPENDIX I

Table 4 Agenda of Paraguay LORTA field mission

DAY	SCHEDULE	ACTIVITY
18 October	14:00–15:15	Opening of the LORTA mission with representatives of each institution involved (SAS, STP, MAG, INFONA, SEAM, INDI, VMME, AFD)
	15:30–17:00	Presentation by LORTA team: "The benefits of impact evaluation"
19 October	7:30–9:45	Project presentation by the STP
	10:00–12:00	Theory of change workshop by the LORTA team
	13:00–16:00	Continuation of theory of change discussion
	16:14–18:00	Presentation by LORTA team: "Impact evaluation methods"
20 October	6:00–9:00	Field visit: Travel to Carayaó
	9:00–9:30	Field visit: Meeting with local authorities
	10:00–13:00	Field visit: Meeting with two rural families of the Tekopora programme
	14:30–17:30	Field visit: Meeting with local indigenous community
	18:00–22:00	Field visit: Travel to Asunción
21 October	11:00–12:00	Meeting with former Minister of STP
22 October	7:30–12:00	Discussion of evaluation design and project timeline
	13:00–15:30	Discussion of programme indicators and outcomes; finalization of theory of change
	16:00–17:00	Discussion of potential evaluation design for component 2
23 October	07:30–10:00	Recap on impact evaluation methods and power calculation
	10:00–11:00	Small group meeting with INFONA on monitoring of forest cover
	11:00–12:00	Small group meeting with SAS regarding the Tekopora programme and payments
	13:00–15:00	Discussion of the M&E strategy lead by FAO
24 October	7:30–8:30	Small group meeting with INDI on evaluation design in indigenous communities
	8:30–10:00	Presentation of the mission's summary by LORTA team
	10:00–11:00	Closing of the mission by the Minister of the STP

APPENDIX II

Table 5 List of participants for Paraguay LORTA field mission

NAME	POSITION	INSTITUTION
Susana Aquino	Adviser	STP
Gisela Dimodica	Adviser	STP
Federico Sosa	Adviser	STP
Ana María Cáceres	Chief DGI	STP
Flavia Sacco	Director	STP
María Teresa Coronel	Director of communication	STP
Jorge González	Expert	STP
Giancarlo Camperi	Director of Poverty Reduction unit	STP
Sonia Garrido	Technical director	STP
Eunice Rivas	Director of Social Services	STP
Rafael Gonzalez Bordon Narcisco	PROEZA Project Coordinator	STP
Fabiola Alcorta	Representative	FAO PY
César Balbuena	Consultant	FAO PY
Aram Cunego	M&E consultant	FAO PY
Jorge Meza	Representative	FAO PY
Muneyuki Nakata	JPO	FAO PY
Silvio Daidone	Impact evaluation expert	FAO HQ
Daniel Puentes	Representative	VMME
Gustavo Cazal	Director DEA	VMME
César Cardozo	Finance director	AFD
Sergio Dávalos	Head of department	MADES
Gustavo González	Head of department	MADES
Federico González	Expert	MAG
Myriam Leiva	Director of M&E	MAG
Francisco Burgos	Director of Cooperation	MAG
Pablo Benítez	Head of department	INFONA
Wilson Jara Rivas	Head of department	INFONA
Antonella Mascheroni	Expert	INFONA
Gustavo Cano	Jefe DP	INFONA
María Teresa Doldán	Environmental expert	SAS
Leila Molas	Head of environmental management department	SAS

NAME	POSITION	INSTITUTION
Carlos Franco	Expert	INDI
Marina Velazco	Adviser	INDI
Faustina Alvarenga	Adviser	INDI
Esther Heesemann	Impact Evaluation Specialist	C4ED
Michaela Thielmann	Research Manager	C4ED
Nathan Fiala	Professor	UCONN

APPENDIX III

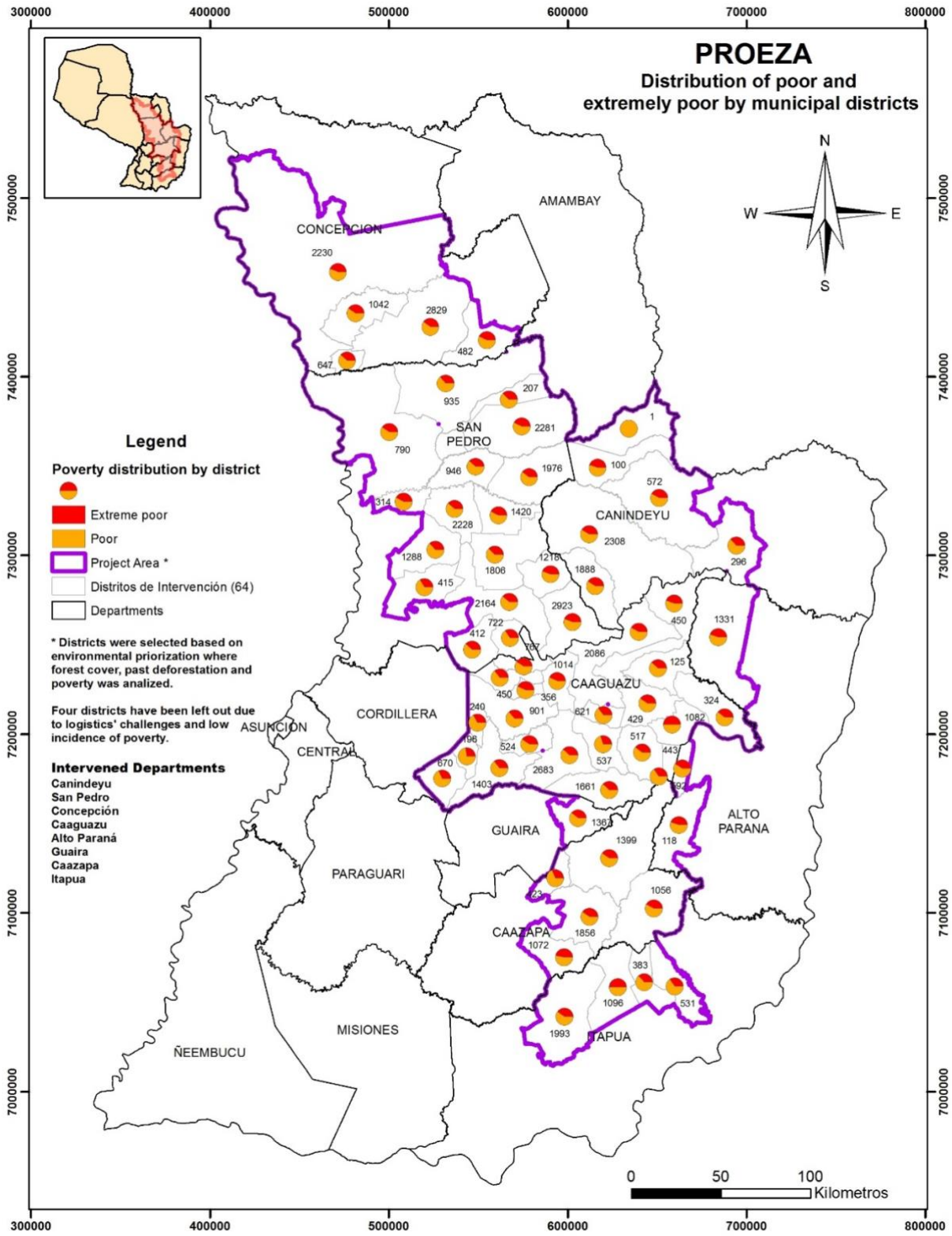


Figure 5 Poverty assessment in PROEZA project area

APPENDIX IV

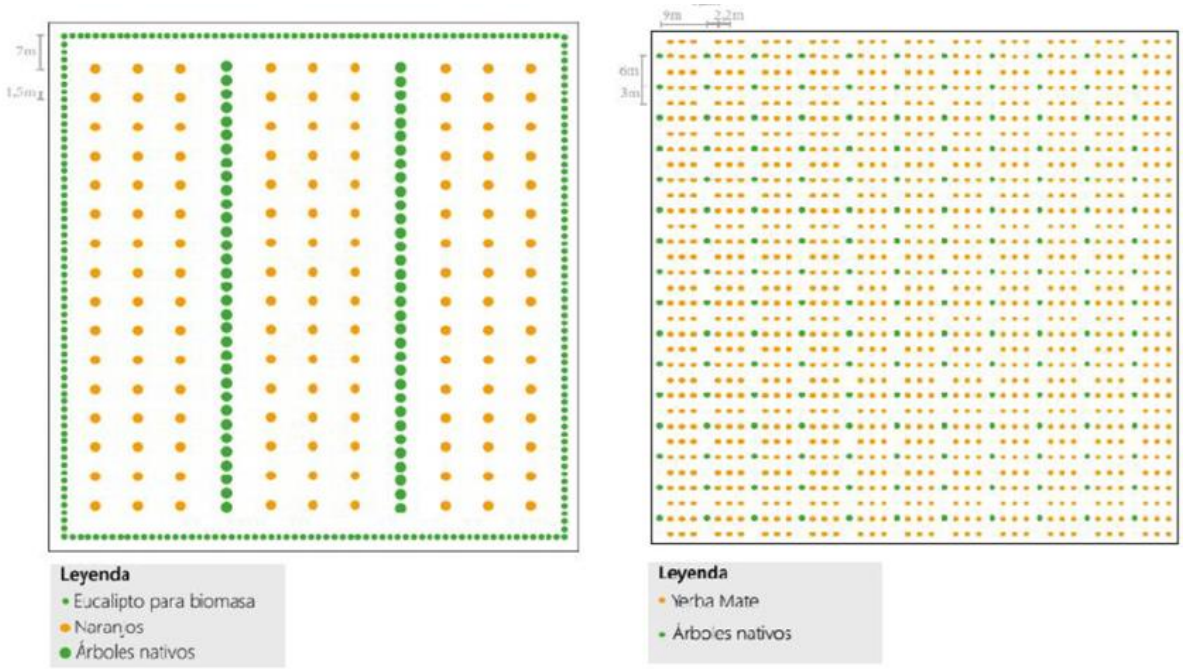


Figure 6 Examples of Agroforestry plantation schemes

APPENDIX V

Ficha Social is an official instrument of the STP, used to identify families in extreme poverty in Paraguay. The information collected is shared with public institutions in order to define the participation of the population into different social programmes. In principle, the instrument was built to provide information for the programme *Sembrando Oportunidades*, created by means of Decree 291 in 2013, as a unique and standardized instrument to identify the families in poverty.

An important feature of *Ficha Social* is that it is focused on the population that resides in dwellings located in the poorest districts of the country, defined by a poverty map called *Mapa de Pobreza* developed by STP. The map identifies the districts according to an index of geographical prioritization (IPG is the Spanish acronym). By August 2017, a total of 279,105 questionnaires had been administered and 60,000 were planned to be administered. From the applied units, 39,818 families were found to be in extreme poverty; 61,273 in poverty; 83,215 not poor but vulnerable; and 94,799 not poor nor vulnerable.

Besides the implementation of *Ficha Social*, other instruments have taken place, such as *Ficha Hogar-Adultos mayores*, which determines the quality of life of older adults, and *Ficha Hogar*, implemented by the Secretary of Social Action (SAS), which determines the quality of life of families that take part in the *Tekopora* programme.[36] From the households considered in *Ficha Hogar*, a quality of life index is constructed to define patterns of eligibility. In the case of the programme *Tekopora*, the activities started in five poor districts in September 2005, with an extension of the activities to 10 other districts following an order of implementation guided by the IPG.

The SAS has two instruments to identify potential beneficiaries of social programmes:

1. The IPG, which enables the identification of districts with the highest share of poverty according to Unsatisfied Basic Needs indexes and poverty in terms of access to the basic basket of goods. The estimations are based on information from the Permanent Survey of Households (EPH) 2011, 2012 and 2013, and Census 2012.
2. *Ficha Hogar*, which consists of a collection of census information and identifies the living conditions of the families. The selection of beneficiaries is based on the application of a quality of life index that takes values from 0 to 100 and considers seven categories as presented in Table 6.

[36] Some tools exist for the monitoring and evaluation of the instruments to measure poverty in Paraguay, such as *Tablero de Control*, RENAF, RENABE, SIIS. *Tablero de Control* is administered by STP, with the information from *Ficha Social* available from July 2015.

Table 6 Categories of the quality of life index

CATEGORIES	DESCRIPTION
1. Children	Number of children 0–5 in the house.
2. Health	• Percentage of members who were sick during the last 3 months with access to a doctor • Percentage of members in the house with health insurance • Number of children (0–5) with vaccination certificate
3. Education	• Number of years of education of the household head and spouse • Percentage of human capital lost because of children (6–24) • Language spoken in the household, mainly by the household head and spouse
4. Income	Occupation category of the household head and spouse
5. Dwelling	• Overpopulation defined with the number of members over the number of rooms in dwelling • Material of the floor, external walls and roof of the dwelling • Existence of bathroom and kitchen • Sanitation
6. Basic services	• Main source of water • Access to electricity • Access to telephone (fixed line or cell phone) • Waste treatment • Type of fuel used to cook
7. Durable goods	Availability of fridge, air conditioner, automobile, motorcycle, washing machine and stove

Source: The World Bank (2010) Paraguay Estudio de Pobreza

The questionnaire of *Ficha Social* is divided into nine main sections: i) geographic identification, ii) field team, iii) result of the interview, iv) dwelling characteristics, v) sociodemographic information, vi) education, vii) health and disabilities, viii) employment, ix) agricultural activities, animals and agricultural crops and x) summary of the household composition The information is provided by the head of the household or a member older than 14 years old. Relevant for the PROEZA project is the identification of households with agricultural inputs. To this end, section ix asks the respondent whether the family owns any plot of land and the number of hectares they own if yes. Table 7 shows the number of *Ficha Social* records that were administered by department between October 2015 to April 2017.

Table 7 Number of Ficha Social records administered between 2015 and 2017

DEPARTMENT	TOTAL
All	147,432
Concepción	5,759
San Pedro	30,485
Cordillera	12,293
Guairá	12,838
Caaguazú	34,020
Caazapá	15,889
Itapúa	14,639
Misiones	2,583
Paraguarí	12,765
Alto Paraná	1,855
Central	257
Ñeembucú	566
Amambay	36
Canindeyú	3,371
Pte. Hayes	12
Boquerón	18
Alto Paraguay	10

Source: Informe final. Identificación de familias en situación de pobreza extrema. Proyecto 96188

As a result of the administration, families and persons were classified in four levels of poverty: extremely poor, poor but not extremely poor, vulnerable but not poor, and not vulnerable or poor. Table 8 shows the composition of the sample.

Table 8 Poverty status of households in the Ficha Social

POVERTY STATUS	FAMILIES		PERSONS	
	Frequency	Percentage	Frequency	Percentage
Extremely poor	15,859	10.9%	98,038	18.4%
Poor not extremely poor	26,647	18.3%	117,282	22.1%
Vulnerable but not poor	42,190	29%	152,566	28.7%
Not poor or vulnerable	60,950	41.8%	163,787	30.8%

Source: Informe final. Identificación de familias en situación de pobreza extrema. Proyecto 96188

IMPACT EVALUATION DESIGN REPORT 5: MADAGASCAR

TABLE OF CONTENTS

TABLES

FIGURES

I. INTRODUCTION

A. The LORTA programme

Evaluating the impact of development projects and programmes has gained importance in recent years. Impact evaluation not only allows for increased transparency by measuring the effects of investments, it also provides the opportunity to design and implement development projects more effectively. To contribute to this development, the Independent Evaluation Unit (IEU) of the Green Climate Fund (GCF) has started the Learning-Oriented Real-Time Impact Assessment (LORTA) programme to be able to keep track of GCF projects in terms of performance and results, and to enhance learning within the GCF.

The LORTA programme aims to

- Embed real-time impact evaluations into funded projects so GCF project task managers can quickly access accurate data on a project's quality of implementation and likelihood of impact, and
- Build capacity within projects to design high-quality data sets for overall impact measurement.

The purpose of the impact evaluations is to measure the change in GCF key result areas that can be attributed to project activities. The LORTA programme will inform on returns on GCF investments and help GCF projects track implementation fidelity. The objectives of LORTA include the following:

- Measuring the overall change (outcome or impact) of GCF-funded projects and enhancing learning
- Understanding and measuring results at different parts of theories of change
- Measuring the overall contribution of the GCF to catalysing a paradigm shift and achieving impacts at scale

Currently, the LORTA programme is in Phase I (formative engagement and design). In the first year (2018), the IEU will support eight GCF projects to build high-quality, theory-based impact evaluation designs at inception. Formative work will include engaging with project teams, accredited entities (AEs) and GCF staff; designing theory-based impact evaluations; and establishing protocols for database development.

The second phase of LORTA will involve the main impact assessment stage (3–5 years), while the third phase will include analyzing baseline and endline data, discussing results, and engaging with diverse stakeholders to share results and incorporate feedback as required.

B. The Madagascar project

The project "Sustainable Landscapes for Eastern Madagascar" (SLEM) is one of the eight projects selected to be part of the inception stage (Phase I) of the LORTA programme. The AEs for this project are Conservation International (CI) and the European Investment Bank (EIB).

The project aims to implement sustainable landscape measures to enhance the resiliency of smallholder farmers, reduce greenhouse gas (GHG) emissions from deforestation, and make climate-smart investments in agricultural and renewable energy. The sustainable landscape measures consist of a portfolio of activities, among which two will be the focus of the impact evaluation: adaptation and mitigation activities. Adaptation activities include the provision of trainings, inputs and technical assistance to smallholder farmers in order to promote conservation agriculture practices and alternative sources of livelihood. Regarding mitigation activities, the project plans to provide trainings, per diems and equipment to physically demarcate the limits of protected forest

areas and to patrol these areas. These two activities will be implemented by Conservation International Madagascar (CI-M). As a result, the impact evaluation of the SLEM project will be done in close cooperation with CI-M, without the involvement of EIB.[37]

Madagascar remains one of the poorest countries in Africa, with 77.6 per cent of the population living on less than USD 1.90 a day in 2012 and a per capita gross national income of USD 400 in 2017.[38] The country is particularly vulnerable to climate hazards and was ranked 22 out of 183 countries in terms of the Climate Risk Index for the period 1995–2014.[39] The severity of climate hazards is expected to increase in the coming years.[40] Madagascar has suffered significant deforestation and land erosion over recent decades, resulting in a need for urgent action. The SLEM project addresses one of the core causes of this alarming trend, namely unsustainable land-use practices. Through its various activities, the project aims at raising awareness on climate-related risks and climate-smart agricultural practices in order to lead to the adoption of practices preventing land and forest degradation while improving smallholder farmers' resilience and food security. The SLEM project aligns with the national REDD+ process.

The project has three goals:

- Improving the production of subsistence and cash crops in order to improve smallholder farmers' food security
- Increasing the options of smallholder farmers to respond to climate-related hazards in order to strengthen their resilience to climate change
- Enforcing the regulations of the forest protected areas in order to reduce deforestation

The project will be implemented in the areas of the Ankenihevy-Zahamena Forest Corridor (CAZ) and the Ambositra-Vondrozo Forest Corridor (COFAV). Both areas are characterized by a climate-vulnerable and food-insecure population relying on forest resources. Frequent cyclones bring heavy rain and strong winds, and often result in flooding.

The project activities target members of *Communautés de Base* (COBA) and of local associations (notably, women producers' and farmers' associations). COBA were created under Decree n°2000-027, in which local communities are responsible for the local management of renewable natural resources. These COBA are created by a group of volunteers either from hamlets of the same village or from several villages. Management transfer contracts over the local forest, made with the government, provide members of the COBA with rights to the sustainable use of forest resources for consumption purposes in exchange for the duty of protecting these resources. In addition, COBA members may benefit from additional ecotourism activities. Most COBA members are involved in agricultural activities and also rely on forest resources.

[37] EIB took part in the initial discussions of the LORTA programme and is fully aware of its objectives and planned activities. However, because EIB's interventions are distinct from those of CI and do not interfere with the impact evaluation strategy, EIB decided not to be involved in these discussions but will remain available for updates on their interventions and implementation plan.
[38] World Bank. (2018). *Poverty & Equity Data Portal.* Retrieved from http://povertydata.worldbank.org/poverty/country/MDG
[39] Kreft, S., Eckstein, D., Lukas Dorsch, L., & Fischer, L. (2016). *Global Climate Risk Index 2016.* GermanWatch, Briefing Paper.
[40] Tadross, M., Randriamarolaza, L., Rabefitia, Z., & Yip, Z. K. (2008). *Climate change in Madagascar: recent, past and future*. World Bank Report.

II. MADAGASCAR FIELD MISSION

A. GENERAL REMARKS

An evaluation team, consisting of two consultants from the Center for Evaluation and Development (C4ED) and the director of the IEU (henceforth referred to as the LORTA team), was formed to lead the field mission from 21 to 26 October 2018. The task of the team was to engage closely with key stakeholders of the project – namely, the national designated authority (NDA), accredited entity (AE), implementing agency, project staff and potential end beneficiaries – to ensure their interest in and understanding and sense of ownership of the planned theory-based impact evaluation.

During the field mission, the LORTA team held meetings and interactive capacity-building workshops with the key stakeholders. Besides conveying technical knowledge, the aim of this workshop was to emphasize the benefits of theory-based counterfactual approaches and real-time learning and measurement. Active interactions during these workshops ensured that the project team went along with the journey. These exchanges were crucial for the project to gain ownership of the impact evaluation. In addition, discussions following group exercises were used to acquire the maximum possible information about the GCF-funded project. Stakeholders shared their views and insights about the project's implementation and monitoring strategies, expected impact, challenges and possible solutions. The presence of multiple stakeholders with various roles and backgrounds brought in key perspectives. These discussions were supplemented by expert interviews and presentations by the project team. All these exchanges not only informed the LORTA team about the project but also fostered collaboration and trust between the team and the on-site parties involved.

Under the guidance of the LORTA team, an impact evaluation design was elaborated for the SLEM project. In collaboration with the project team, the LORTA team conducted context analyses, examined the existence of appropriate counterfactuals (i.e. comparable treatment and control groups), elaborated a theory of change, assessed the availability of baseline administrative and secondary data sources, and acquired budget information. The results from this undertaking are presented in the following sections.

B. THE MISSION AGENDA

The LORTA team received prompt collaboration from CI. An initial agenda – shown in appendix I of this design report – was developed to facilitate the joint attendance of all key stakeholders at the LORTA workshop and to allow for field visits. However, thanks to the availability of all key stakeholders throughout the week (namely, 22 staff members of CI-M, 1 staff member of the Bureau National de Coordination pour le Changement Climatique (BNCCC), Rob Meritt from CI headquarters and Giacomo Fedele from CI Moore Center), this agenda was modified to include group exercises and lively interactions. The final agenda of the week is presented in appendix II of this design report.

In particular, the LORTA workshop was expanded over three days and contained all the ingredients to inform the key stakeholders on the LORTA objectives, analyzing the project's theory of change, defining key high- and low-frequency indicators of the project, discussing programme implementation details and presenting key impact evaluation concepts.

The workshop was beneficial for both the key stakeholders and the LORTA team. On one hand, the presentations and interactive discussions on theory of change, implementation and indicators brought all the key stakeholders together on the same page with respect to ownership and understanding of their contribution within the SLEM project, from the project objectives to the evaluation needs and implementation strategies. On the other hand, the LORTA team benefited from

gathering crucial information to design the impact evaluation and from gaining understanding and collaboration from the AE in adjusting their plans to accommodate the evaluation design.

Presentations by the Chief of Party, the monitoring and evaluation (M&E) director and the geographic information system (GIS) specialist informed the LORTA team on the role of each key stakeholder of the project, current plans for M&E, available data sets for the project and plans for merging household surveys and GIS information.

A courtesy visit to the NDA, Mrs. Heritokilalaina, director of the BNCCC, could be accommodated within the week. This meeting was an occasion for the LORTA team to express our interest in and enthusiasm for this collaboration, present the LORTA objectives and provide the BNCCC with a sense of ownership of this project. BNCCC staff also participated in all workshop activities performed during the mission.

A field visit was organized to one of the COBA of the area of intervention, the Laroka community located in Andasibe. A group discussion with several members of the COBA provided key insights on the internal and external organization of COBA, their interactions with other local communities, their local conditions, patrolling activities and challenges faced during these activities. Some members of the community also took the LORTA team to the forest managed by the COBA.

The mission enabled the LORTA team to establish a tight collaboration with CI and to adapt the impact evaluation design to the needs and concerns of the implementing team.

Appendix III of this design report lists all the people engaged with at the workshop throughout the LORTA mission.

C. RESULTS

Project implementation plan

The SLEM project originally targeted smallholder farmer members of the 230 COBA of the CAZ and COFAV areas. These landscapes are particularly vulnerable to climate-related risks and are characterized by a climate-vulnerable and food-insecure population of smallholder farmers that rely on wood as a primary source of fuel. The original number of 230 COBA had to be reduced to 178, 84 in the COFAV area and 94 in the CAZ area, because of safety concerns in some areas of intervention. COBA in unsafe locations are now excluded from the project. To meet the target number of final beneficiaries, CI is currently considering also allowing local associations that are not COBA (notably, members of women's associations, farmers, producers of seeds, groupings of individuals affected by the protected areas) into the SLEM project.

The SLEM project will be implemented in three phases over three years by CI staff, 250 local trainers in CAZ, 450 local trainers in COFAV and 15 field agents. The collection of baseline data within each COBA will serve as a basis to define the set of adaptation activities that are most relevant to the needs of the local community. Even though these activities will differ across COBA, they will all contain several sessions of trainings, provision of inputs and continuous technical assistance. Field agents are in charge of providing this technical assistance to eligible farmers during the first year of the project for all three phases. Afterwards, technical assistance will be ensured by lead farmers identified during this first year.

The initial timeline for SLEM roll-out is presented in Table 1. Baseline data collection was initially planned for October 2018. Because of delays in the preparation phase, this activity has been pushed back to February 2019, to accommodate the rainy season, the coming elections and harvests. The start of the adaptation and mitigation activities has been postponed accordingly and is expected to be completed by September 2019 for areas selected in the first phase of the project. This new timeline still needs to be discussed within CI, and updates are expected by the LORTA team.

Table 1 SLEM initial roll-out plan

	2018		2019				2020				2021				2022				2023	
ACTIVITIES	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2
ADAPTATION																				
Alternative livelihood training																				
Agroforestry and tree planting																				
Micro-irrigation and drainage																				
Intercropping and multi-cropping systems																				
Off season rice cultivation																				
Mulching																				
No tillage																				
Terracing																				
MITIGATION																				
Provide per diem to community rangers																				
Provide equipment to community rangers																				
Demarcate the limits of protected area																				
Trainings on legislation																				
Establish SMART-SMS database																				

Theory of change

Two theories of change for adaptation and patrolling activities of the SLEM were first developed by CI before the mission, with the support of the LORTA team. These theories of change were discussed at length during the mission with the intent of assessing the plausibility of their underlying assumptions. These theories of change are presented separately in the following subsections.

Component 1: Adaptation activities

i) Inputs

GCF grant and CI funds will be allocated to the hiring of field agents and local trainers and the purchase of inputs. Training modules will be developed.

ii) Activities

Adaptation activities will consist in conducting several trainings, providing inputs and continuous technical assistance on alternative livelihoods strategies (such as cash crops and duck breeding) and conservation agriculture practices. The conservation agriculture practices include agroforestry and tree planting, micro-irrigation and drainage canals, intercropping and multi-cropping systems, off-season rice cultivation, mulching, no tillage and terracing. These activities will be adapted to the needs and characteristics of each COBA. The good conduct of these activities will be measured by collecting information on the number of trainings conducted, the purchase of inputs and the number of field agents and local trainers.

iii) Outputs

The adaptation activities are expected to result in the dissemination of risk reduction practices, an enhanced knowledge on alternative sources of livelihoods, and receiving inputs and technical assistance. The successful completion of the output stage will be measured in terms of number of households participating in the trainings, receiving inputs and assistance, feedback on the clarity of the material presented, and knowledge on climate change and risk reduction strategies.

iv) Outcomes

If the target group benefits from the elements described in the output stage, we expect a certain number of households to implement these new practices, to increase their production of subsistence and cash crops, to use the provided inputs, to rely on assistance when needed, to seek weather forecasts and market information and to adapt their practices accordingly.

v) Goals

The main goals of the adaptation activities are higher production, higher food security and decreased vulnerability to climate-related hazards.

These five stages are summarized in Figure 1.

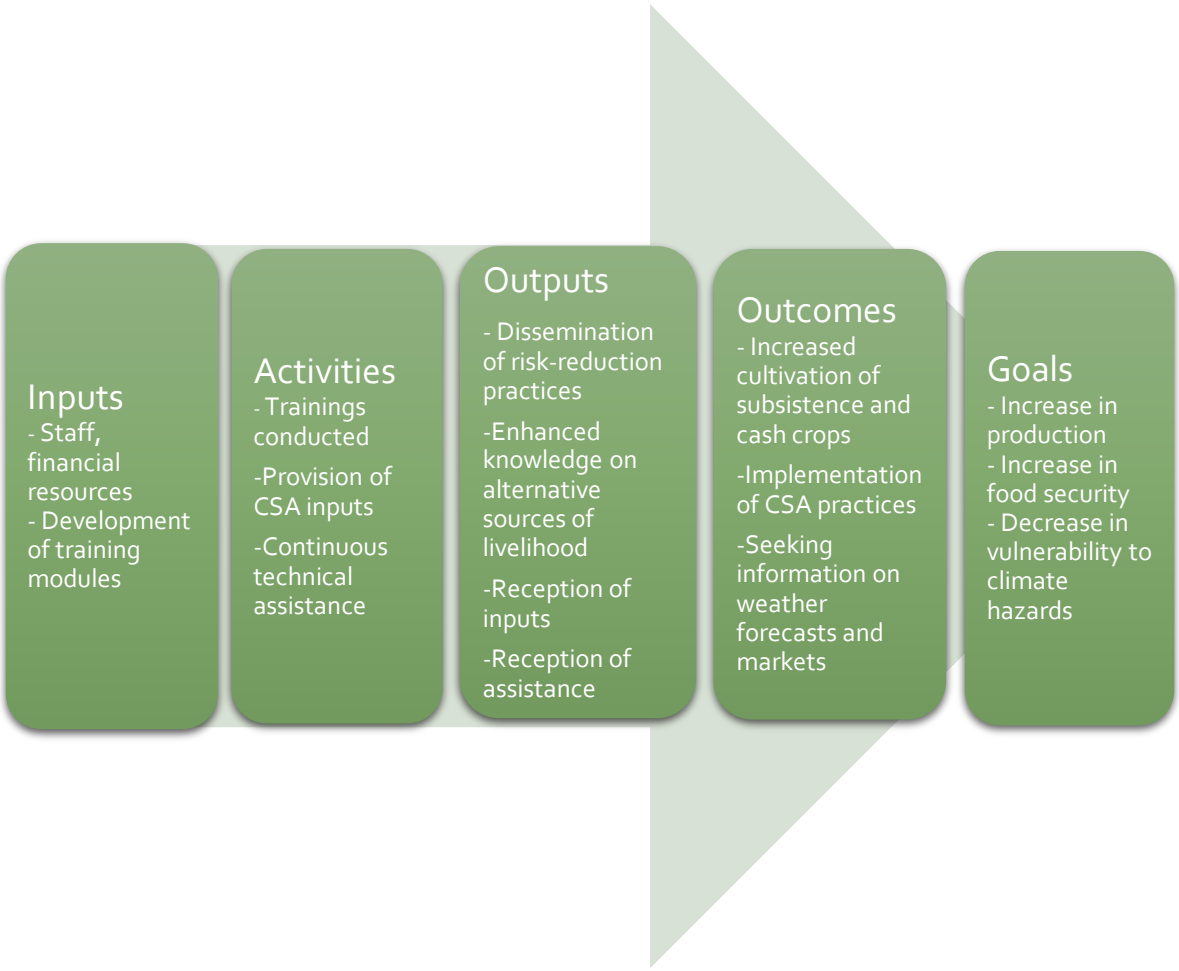


Figure 1 Theory of change of adaptation activities

Evaluation questions

The main evaluation questions, derived from the hypotheses underpinning the theory of change, are listed below:

- Do households implement alternative livelihood strategies?
- Do households implement conservation agriculture practices?
- Do adaptation interventions lead to a reduction of households' vulnerability to climate hazards?
- Does the adoption of new practices lead to an increase in agricultural production and food security?

Impact indicators

- Number of farmers who diversified their livelihoods
- Number of farmers who implemented conservation agriculture practices
- Damages in agricultural, forest and livestock product following climate hazards
- Share of the agricultural production not for household consumption
- Quantities produced of three main crops, animals, forests/tree products
- Food security index based on food consumption, food expenditure shares and the number of strategies to cope with a lack of food
- Number of days members of the household did not eat three meals a day

Component 2: Patrolling activities

i) Inputs

GCF grant and CI funds will be allocated to the hiring of field agents and local trainers and the purchase of equipment. Training modules will be developed.

ii) Activities

Patrolling activities will consist in providing per diems (15,000 Ariary) and equipment to community patrollers, the physical demarcation of the limits of the protected areas, and trainings of patrollers and associations on forest legislation.

iii) Outputs

The patrolling activities are expected to result in patrollers receiving per diems and equipment, enhanced knowledge on forest legislation, increased awareness on sustainable forest management and the visibility of the limits of protected areas.

iv) Outcomes

If the target group benefits from the elements described in the output stage, we expect rangers to increase their surveillance of protected areas and households to extract fewer resources from these areas.

v) Goals

The main goals of the patrolling activities are an increase in the percentage of reported violations that are prosecuted and a reduction in deforestation.

These five stages are summarized in Figure 2.

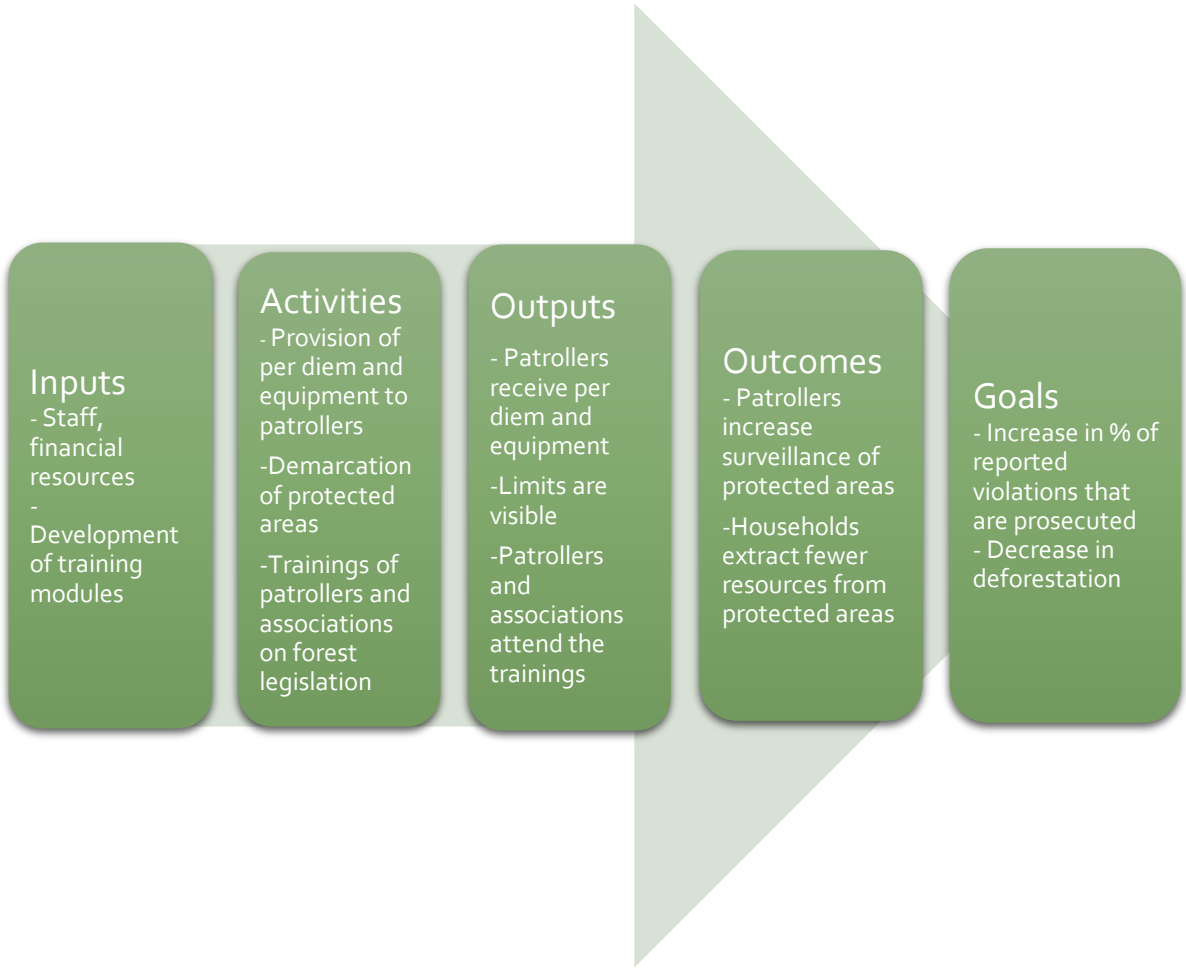


Figure 2 Theory of change of patrolling activities

Evaluation questions

The main evaluation questions, derived from the hypotheses underpinning the theory of change, are listed below:

- Do patrollers cover a greater distance during patrols?
- Do patrolling interventions lead to a better enforcement of regulations of the forest protected area?
- Do patrolling interventions result in a reduction in deforestation?
- Does deforestation increase in other areas as a result of an increase in forest surveillance in the target areas?

Impact indicators

- Distance covered by patrollers
- Number of days per person and per month of patrols
- Percentage of reported violations prosecuted by authorities
- Self-reported inappropriate use of forests
- Deforestation in target and nearby areas, self-reported and measured by satellite images

D. PLANS FOR EVALUATION

Methodology

The proposed methodology follows a mixed-method approach that combines quantitative and qualitative data analysis. The quantitative evaluation will be based on both an experimental and a non-experimental design. Combining the experimental and non-experimental approaches in a complementary fashion will make it possible to increase the timespan that the research covers. A cluster randomized phase-in will serve in the identification of the short-term effects of the SLEM project. For the estimation of longer-term impacts, a difference-in-differences (DiD) approach combined with matching will be used. Additional qualitative data will be collected in the form of key informant interviews and focus group discussions during baseline and every subsequent wave of data collection. The complementary qualitative analysis will help us to further understand for whom and why interventions work or do not work. Qualitative analysis will also be used to assess the gender sensitivity of the SLEM interventions.

Impact evaluation design

During the LORTA mission, two complementary impact evaluation designs have been developed: a cluster randomized phase-in and a DiD approach, combined with matching. The experimental approach represents the most robust estimation of the short-term effects of the SLEM project. A quasi-experimental design is implemented as an additional strategy to capture longer-term effects.

Cluster randomized phase-in

A cluster randomized phase-in is an experimental design that relies on the ranodmization of the order in which each eligible cluster receives the project activities. The unit of assignment of the SLEM project is the COBA, and all eligible COBA of the SLEM project will eventually receive the interventions. In order to avoid social conflicts within COBA, we have opted for a cluster ranodmization, with the COBA being the *cluster* of farmers and patrollers that will be randomly allocated into the different phases of the project. This is opposed to a standard ranodmization where individual farmers and patrollers within COBA would be the unit of ranodmization.

Thanks to a roll-out in several stages, it is still possible to randomly select the order in which the eligible COBA participate in the project. Clusters randomly assigned to a later phase serve as the comparison group until they start receiving the interventions. An additional advantage of a cluster randomized design is to avoid the contamination of the comparison group. Indeed, we expect spillovers to occur within COBA as farmers and patrollers receiving trainings may share new information with individuals who have not participated in the project. Interaction and information sharing with members of other COBA might be rare, as members reported during our field visit of the Laroka COBA. Therefore, contamination of the control group appears less of a concern in a cluster design.

For a rigorous impact evaluation, it is essential that the order of phase-in of the eligible units is determined randomly. According to the law of large numbers, a sufficiently large number of observations is required for the process of randomized selection of beneficiary units in the different phases of the project to effectively produce groups that are on average similar. In other words, if the number of eligible units is large enough, we can be sure that the first beneficiary group and groups subsequently phased in have, on average, the same observable and unobservable characteristics. Hence, ranodmization ensures the greatest degree of comparability between the first project beneficiaries and future beneficiaries – the comparison group – before the roll-out of the project.

Data collection on both groups before the start of the project will enable us to verify that observable background characteristics are balanced between the treatment and comparison group. Then, any differences between the two groups that we observe after the implementation in the treatment group (and before the phase-in of the comparison group) will be uniquely attributable to the project. As an experimental approach, a cluster randomized phase-in is the strongest method of evaluation of the impact of the SLEM project.

The SLEM project will be implemented in three phases. The three stages target a similar number of beneficiaries and are evenly spread over three years. Each implementation phase takes six months. Because the main outcomes of interests to the project – food security, vulnerability and deforestation – are expected to evolve slowly over time, an evaluation at the end of the first phase is not cost-effective. The impact evaluation design will hence focus on the first and third phases of roll-out, with a total of one and a half years between completed intervention and evaluation. The phase-in design will hence measure only the short-term impacts of the project. Due to the limited total project period of five years, spacing between phases could not be increased in order to capture longer-term impacts.

In addition to the velocity of measurable progress in outcomes, limited ranodmization possibilities in certain communities led to the decision to exclude phase 2 from the evaluation. During discussions with the project team, it appeared that some communities could not receive the project in the first phase because of current conflict with CI or current interventions. Therefore, these communities have to be taken out of the impact evaluation sample. By excluding the second phase from the impact evaluation design, these communities can be part of the second phase without undermining the ranodmization of other communities between the first and the third phases.

Despite the likely occurrence of spillover effects within communities, the evaluation strategy does not foresee a quantitative evaluation for budgetary reasons. Instead, qualitative tools such as focus group discussions will be used to investigate how the interventions were perceived by non-participating members of the community, which changes they observed and how they felt affected.

To account for the geographic heterogeneity of the area of intervention and of the size of the forest covered by COBA, after discussions with the project team, we opted for a stratified ranodmization. The stratification will be based on four geographic areas – the north and the south regions of COFAV and the eastern and western regions of CAZ – and three quantiles of the surface of the

forest covered by COBA. Both characteristics may mitigate the impacts of the project. The surface covered by a COBA is independent from the number of patrollers, set to four individuals, which likely affects the patrolling activities. The stratification ensures that the proportion of each of these regions and groups is fairly similar across the phases. Two advantages arise in doing this. First, by gaining control of the composition of the sample, we improve the precision of the estimation of the impacts of the project. In addition, it ensures the representativity of these subgroups in each phase.

To summarize, let H_{P1} denote smallholder farmer members of COBA who were randomly assigned to phase 1 and H_{P3} denote smallholder farmer members of COBA who were randomly assigned to phase 3. The SLEM project's effect on members of a COBA is estimated by the following expression:

$$\beta = H_{P1} - H_{P3}$$

Let F_{P1} denote parcels of forest areas controlled by COBA that were randomly assigned to phase 1 and F_{P3} denote parcels of forest areas controlled by COBA that were randomly assigned to phase 3. The SLEM project's effect on forest areas of a COBA is estimated by the following expression:

$$\gamma = F_{P1} - F_{P3}$$

Both effects will be estimated at the end of phase 2, capturing the impacts of one and a half years of full exposure to the project.

Difference-in-differences with matching

During discussions with the project team, it was clear that greater impacts are expected over time and should be assessed. As already mentioned, spacing between phases cannot be increased and eligible units cannot be excluded from the project. In this context, an experimental setting for impacts of the project after more than one and half year of interventions is not feasible. For this reason, we turn to a quasi-experimental design as an informative, though less robust, complementary strategy.

The DiD approach estimates project effects through the comparison of changes in outcomes over time between beneficiaries and a comparison group. As every COBA of the intervention area will ultimately receive the project, this comparison group will be composed of households in communities where no COBA exists. Because the formation of COBA and membership are not random, we expect that beneficiaries and the comparison group will differ at baseline. Therefore, a pure ex-post comparison of both groups does not enable us to recover the effects of the project. Instead, we will compare changes in outcomes between the two groups, acknowledging potential initial differences.

A DiD design accounts for initial observable and unobservable differences between the treated and comparison groups. As long as initial differences having an impact on the outcomes of interest are constant over time, this method enables the causal identification of the impacts of the project. This approach is also robust to external shocks, as long as these shocks affect both groups similarly. Thus, the crucial assumption of this technique is that the change in outcomes in the treated and comparison groups would have been the same without the intervention. This is called the "parallel trends" assumption.

The parallel trends assumption is illustrated in Figure 3. In this fictitious example, we can see that the agriculture production of the treated group and that of the comparison group evolved in a similar manner before the implementation of the project and would have continued to change in the same way in the absence of the project. This is illustrated by the counterfactual (dotted green line), which represents the change in agriculture production of the treated group in the absence of the project. Based on this assumption, the change in trend observed for the treated group after the project can be uniquely attributed to the project. The effect of the project on agriculture production is then

estimated by computing the difference between d2 and d1, where d2 is the difference in outcome between the treated and comparison group after the project and d1 is the difference in outcome between the same groups at the beginning of the project. Hence, the impact of the project corresponds to the DiD interval (orange arrows).

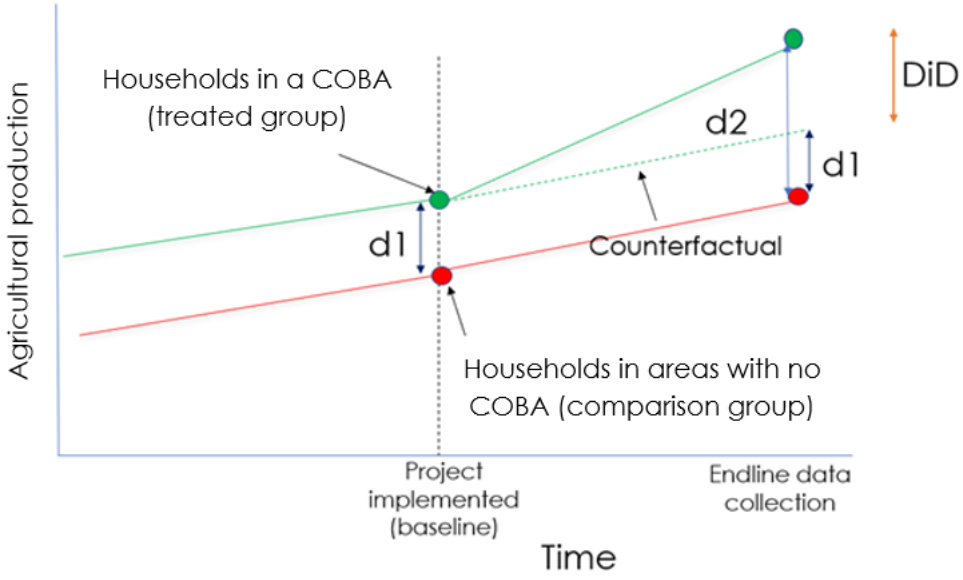


Figure 3 The parallel trends assumption

The DiD design that we suggest for this study will consider COBA randomly assigned to phase 1 as the treated group. The geographical scattering of this group, if reproduced within the comparison group, will reduce the probability that any systematic time-varying difference arises between the two groups due to external shocks. Indeed, because of their geographic dispersion, external shocks are less likely to affect only one of these groups. The validity of this assumption is further reinforced by greater similarity between the groups at the beginning of the project. Therefore, we suggest complementing this approach with matching.

Matching consists in using statistical techniques to construct an artificial comparison group. The idea is to select, for every treated unit, a non-treated unit that has the most similar observable characteristics. Performing matching at the household level requires a data set two to five times larger than the treated group to enhance the quality of the matches and thus of the comparability between the treated and comparison group. The dedicated budget of the SLEM project does not enable us to cover such a large sample. Rather than matching at the household level, we will opt for matching at the community level. We will first rely on the expertise and local knowledge of the project team to select an initial pool of communities with agricultural practices, cultural norms, population density and reliance on forest resources that are similar to the treated areas. From this pool, we will perform statistical matching on distance to forest, population density, distance to the main road, climate and other geographic characteristics available in the project GIS data sets.

Within the identified communities, two groups could be considered. One group would consist of agricultural households. This choice is based on the initial targeting of the SLEM project – that is, smallholder farmer members of COBA. Another group would consist of members of self-help groups, as they represent potential future beneficiaries of CI interventions.

A major drawback of the DiD approach is that the parallel trends assumption during the project cannot be tested and any time-varying difference between the two groups that affects the outcomes of interest would bias the estimation of the impacts of the project. COBA being created on the initiative of individuals, it is possible that communities that implement a COBA care more about the

protection of natural resources and have a higher social capital or social cohesion. Therefore, these communities may invest more effort on average to ensure this protection. Such differences will not be captured by the above-mentioned matching strategy.[41] As an example, farmer members of a COBA could be more likely to implement soil conservation practices than farmers who are not members of a COBA, even in the absence of the project. Soil conservation practices, by improving the soil quality, lead to higher yields over time. In contrast, unsustainable agricultural practices among farmers who are not members of a COBA would lead to a depletion of the soil, resulting in a decrease of yields over time. In such cases, the parallel trend assumption would not hold, as illustrated by Figure 4. We can see that as long as agricultural production does not evolve in the same way for both groups in the absence of the project, the DiD interval will not be equivalent to the effect of the project. In this example, changes in outcomes for the comparison group is not a good approximation of the counterfactual anymore. Therefore, the impact of the project cannot be retrieved.

Even though a lack of knowledge and of access to appropriate inputs represent major obstacles to the adoption of soil conservation practices, in both COBA and non-COBA communities, this risk should not be excluded. Similar examples apply to the exploitation of forest resources. As a result, the credibility of the suggested impact evaluation can be questioned.

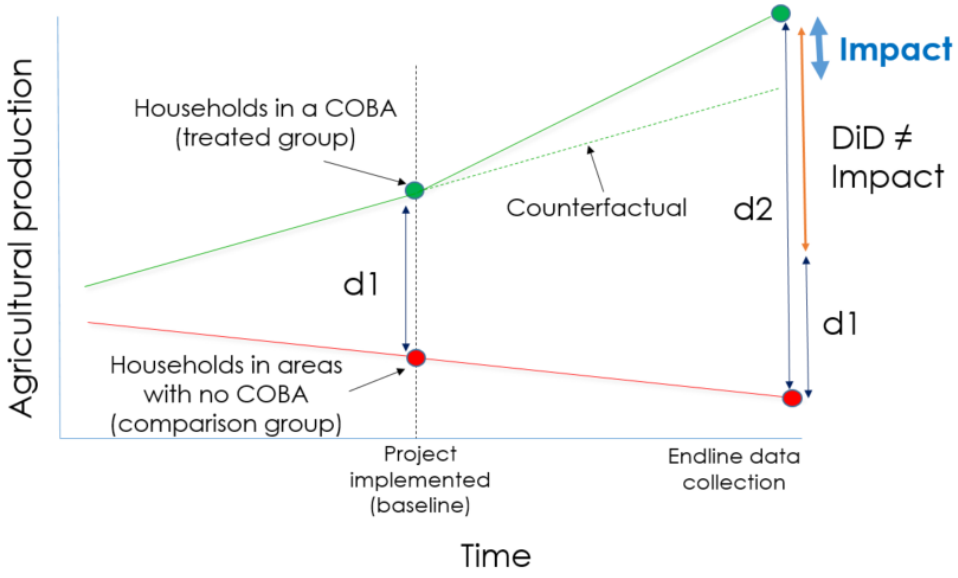


Figure 4 Failure of the parallel trend assumption

The availability of data for both groups or areas for at least two periods before the start of the project would enable us to test the validity of the parallel trends assumption before the project. If this test confirms that this assumption holds, this would significantly enhance the credibility of the chosen strategy. In case of failure of this test, the worthiness of pursing this complementary strategy should be discussed. If maintained, the results may still inform us on the differences between these groups but could not be interpreted as causal impacts of the project. To perform the pre-trends test, we suggest collecting an additional round of data on this comparison group, as will be discussed in the data collection section.

[41] One way of lowering this concern would be to undertake COBA surveys before the baseline data collection in order to identify key characteristics leading to the creation of COBA on which the matching of communities in areas with no COBA will be performed. Doing so would require conducting additional surveys in a sufficient number of communities in areas with no COBA, implying a non-negligible increase in data-collection costs.

Sampling

The evaluation of the SLEM project rests on two units of analysis: households and forest areas. Regarding households, an equal number of members of a COBA – potential beneficiaries of the SLEM interventions – will be randomly selected based on membership lists. Sampling weights will be computed to account for heterogeneity in cluster size.

Regarding forest areas, main analyses will be based on satellite imagery. This approach enables us to divide forest areas into grids and to use each grid as a sampling unit. From these grids, two samples will be constituted: one containing forest parcels far enough from the limits of the demarcated protected area and one composed of forest parcels close to these limits. In both cases, a large number of grids will be randomly selected.

The sampling approach of households for the DiD comparison group will depend on the selected target group and availability of appropriate sampling frames. In the absence of sampling frames, a screening questionnaire may be used to identify relevant households.

Power calculations

Power calculations were performed by the LORTA team in order to estimate the minimum detectable effect size (MDES), considering the constraints of the SLEM project. These constraints consist of the number of clusters in which the project will be implemented during each phase (50 COBA) and in budget constraints for data collection (2,500 households).

We are interested in two bilateral comparisons, between the treated group (COBA members from phase 1) and the comparison group within the area of intervention (COBA members from phase 3), and between the treated group and a comparison group outside the areas covered by COBA (the DiD control group). As a result, the MDES needs to be estimated separately for each of these comparisons. Considering an equal allocation ratio between these three groups, the maximal sample size within budget constraints for each comparison is equal to 1,652 households.[42]

The MDES is calculated with the following formula:

$$MDES = (t_{1-\kappa} + t_{\alpha}) \sqrt{\frac{1}{P(1-P)}} \sqrt{1 + \rho(m-1)} \sqrt{\frac{\sigma^2}{N}} \sqrt{1 - R^2}$$

where $t_{1-\kappa}$ and t_{α} are t-statistics representing the required power and level of statistical significance, P represents the proportion in one of the two compared groups (allocation ratio), ρ is the intra-cluster correlation (ICC), m is the number of individuals per cluster, σ^2 is the variance of the outcome of interest within our population, N is the total sample size and R^2 represents the extent to which baseline characteristics predict the endline outcome.

The MDES was estimated for a power of 80 per cent and a level of statistical significance of 5 per cent. Since we are considering a cluster design, we have to account for the similarity of members within the same COBA. This similarity is measured by the intra-cluster correlation, which compares the variance in outcomes within clusters and between clusters. When the similarity in outcomes within clusters increases and – at the same time – there is heterogeneity across clusters, the variability of the responses of households to the interventions reduces. As a result, the sample size required to detect a significant difference between the treated and the control group increases. Because there is no available data at the COBA level, following the literature we considered four different values of ICC: 0.05, 0.10, 0.15 and 0.20.

[42] Note that in a case where the number of clusters cannot be increased, an equal ratio is optimal. Indeed, although the treated group will be used in two types of bilateral comparisons, the benefits obtained from increasing the size of these bilateral samples are counteracted by the increase in the average cluster size.

To estimate the variance of outcomes of interest, we turned to the Afrobarometer 2017, a nationally representative household survey. This survey contains information on food security, one of the main impact indicators of the SLEM project. An indicator of food insecurity was defined, equal to 1 if the interviewed individual reported that s/he or a member of his/her household lacked food several times or more during the last 12 months. Basic descriptive statistics were computed for the regions including the CAZ and COFAV areas. According to this survey, 55 per cent of individuals residing in the regions that include the COFAV and CAZ landscapes are food insecure. The standard deviation is equal to 0.50.

Table 2 shows the results of power calculations for different values of ICC.[43] In such a setting, we estimate that the MDES varies from 0.077 to 0.139 percentage points, depending on the level of ICC. In other words, we would be able to detect a decrease in food insecurity from 25.3 per cent to 13.9 per cent. This corresponds to a standardized MDES ranging from 0.154 to 0.28. A 2015 systematic review of agricultural input innovations on African smallholder farmers' food security identified improvements in standardized indicators of food security from 0.23 to 1.24.[44] Hence, the minimum effect size we would be able to detect can be seen as a reasonable lower bound of the expected impacts of the SLEM project on food security.

Table 2 Power calculations

INDICATOR	ICC	# OF CLUSTERS PER GROUP (COBA)	TOTAL SAMPLE	R2	SIZE OF CLUSTERS	MDES (IN % POINTS)	% CHANGE IN FOOD SECURITY
Food insecurity	20%	50	2,478	30%	16.52	0.116	21.2%
Food insecurity	20%	50	2,478	0%	16.52	0.139	25.3%
Food insecurity	15%	50	2,478	30%	16.52	0.105	19.1%
Food insecurity	15%	50	2,478	0%	16.52	0.125	22.8%
Food insecurity	10%	50	2,478	30%	16.52	0.092	16.7%
Food insecurity	10%	50	2,478	0%	16.52	0.110	19.9%
Food insecurity	5%	50	2,478	30%	16.52	0.077	13.9%
Food insecurity	5%	50	2,478	0%	16.52	0.091	16.6%

Possible risks for impact evaluation

Potential risks for the experimental impact evaluation design – that is, the cluster randomized phase-in – relate to the possibility of spillovers across COBA and of anticipation effects. For the effects at the level of households, the existence of spillovers will be assessed by exploring impact heterogeneity along two dimensions. First, we will compare the effects of the project between

[43] Power calculations were also performed for alternative number of clusters. The smallest MDES is achieved when the greatest number of clusters is considered.

[44] Stewart, R., Langer, L., Da Silva, N. R., Muchiri, E., Zaranyika, H., Erasmus, Y., Randall, N., Rafferty, S., Korth, M. & Madinga, N. (2015). *The effects of training, innovation and new technology on African smallholder farmers' wealth and food security: A systematic review.* Campbell Systematic Reviews, 11.

control COBA whose managed forest shares a border with a treated COBA and other control COBA whose managed forest does not share such a border. Second, we expect COBA that are more involved in ecotourism activities to have better access to outside information. Indeed, these activities may lead to more travel between different areas and, consequently, to more interactions with members of other COBA. Hence, we will also compare the effects of the project between these COBA and COBA that are less involved in ecotourism activities. This information could be obtained through COBA questionnaires, key informant interviews with COBA treasurers or through focus discussion groups. The presence of spillovers may lead to an underestimation of the impact of the SLEM project on households. However, during our field visit, members of a COBA reported rare interactions with members of other COBA. Although some COBA are gathered in federations, these federations are not very active, and meetings only take place twice a year. The main objective of these federations is to facilitate communications with donors.

While we expect spillovers to be positive at the level of households, negative spillovers may occur at the level of forest areas. Indeed, as a result of an increase in forest surveillance in treated areas, violations could be displaced in nearby, less-monitored areas. In this case, the size of spillovers will be estimated by comparing parcels of forest close to patrolled areas. By comparing parcels of forest sufficiently far from each other, the direct impact of the project can still be retrieved.

A key challenge of the quasi-experimental impact evaluation design – that is, DiD with matching – is the selection of the relevant comparison group. First, this comparison being in areas with no COBA, the unit of assignment of the SLEM interventions does not exist and needs to be substituted by another meaningful unit. As members of a COBA can come from a group of villages, the *foukutany* (the second smallest administrative unit after villages) will be considered. Second, households whose characteristics would have made them likely to be targeted by the intervention need to be selected. One option is to retain smallholder farmers as this corresponds to the initial target of SLEM interventions. Another option is to reach members of self-help groups, as they represent potential future beneficiaries of CI. One advantage of this second option is that listings of members may be available and could serve as a basis for sampling. The main disadvantage is that these households may be reached by CI during the timespan of the project. As CI is currently looking for donors for these other areas, this risk also exists for smallholder farmers.

Another risk lies in the possibility that COBA are created in the comparison areas during the timespan of the project. COBA creations would undermine our strategy only if they became eligible for the SLEM interventions, a point that needs to be discussed with CI.

Qualitative assessments

The plan for qualitative assessments of the SLEM project is still being developed by CI-M. At the time of the LORTA mission, the project team planned for key informant interviews with presidents and vice-presidents of COBA. Local trainers and field agents may also be considered. Focus group discussions will take place with members of COBA, notably with women in order to assess the gender sensitivity of the SLEM interventions. These qualitative assessments are currently planned for all communities of intervention. The LORTA team will query the project team about their updated plan and offer their feedback on the materials developed for this purpose.

Data collection

The SLEM project initially envisioned three rounds of data collection. During the LORTA mission, the LORTA team suggested a reallocation of the budget for data collection across four rounds of data collection, as illustrated in

Table 3. Baseline data collection is currently planned for February 2019, followed by two midline surveys two and three years later. Endline data collection is planned for February 2023.[45] As can be seen in

Table 3, data will not be collected for all groups for each survey round. While households from the treated group will be interviewed four times, the comparison group within the target area will be interviewed at baseline and after two years only, and the comparison group outside the target area will be interviewed at baseline, after three and after four years.

In line with the project's budget constraints, information will be collected on 2,478 households, for a total of 7,434 observations by the end of project. To account for potential attrition across the different survey rounds, the LORTA team recommends adding 10 per cent to the above-mentioned numbers, which will give us up to 8,177 observations.

Table 3 Data-collection timeline

GROUPS	FEBRUARY 2019 BASELINE	FEBRUARY 2020	FEBRUARY 2021 MIDLINE	FEBRUARY 2022 MIDLINE	FEBRUARY 2023 ENDLINE
Phase 1	826 HH	-	826 HH	826 HH	826 HH
Phase 2	-	-	-	-	-
Phase 3	826 HH	-	826 HH	-	-
Outside group	826 HH	-	826 HH?	826 HH	826 HH

Note: HH = households

An additional data point on the DiD comparison group would be valuable in assessing the credibility of the parallel trends assumption. If data are also collected for this group in February 2021, we would be able to compare changes in outcomes between households of COBA assigned to phase 3 with changes in outcomes in the outside group. As both groups would not have received the project, these changes should be on average similar for the parallel trends assumption to hold. In case the budget for data collection can not be extended, a possibility would be to conduct this additional survey instead of collecting data in February 2022. This issue has not yet been discussed with CI.

Data collection will be supervised by the project staff, already experienced in household surveys, and the CI Moore Center. Local enumerators will be hired and trained by the project staff. The household questionnaire has already been developed and is currently being finalized by the project team. This questionnaire was shared with the LORTA team and reviewed before the mission. Discussions between the LORTA and project teams on plans for data collection, training of enumerators, translation and tests of the questionnaire, as well as on high-frequency data quality checks were initiated before the mission and were pursued in the field. Notably, the LORTA team elaborated a checklist for high-frequency data quality checks and suggestions on how to link monitoring and survey data. These documents were shared with the project team. The LORTA team offered to review the enumerators' manual once it has been developed by the project team.

Timeline of evaluation

The timeline of the evaluation is tied to the timeline of the data collection, discussed above. Multiple activities – namely, desk review, review of materials for data collection and elaboration of the

[45] Power calculations were additionally performed under the scenarios of no midline in 2022. However, the subsequent increase in sample size would only lead to a decrease in the MDES of 1 percentage point. Thus, keeping these observations for an additional data point appears to be a better use of available resources.

evaluation strategy – have been undertaken by the LORTA team since September 2018, in close collaboration with CI.

The project team expects to complete each round of data collection in two months. According to the current data-collection timeline, an assessment of the balance at baseline between the selected treated and comparison groups could be performed in May 2019. This assessment will be complemented by an analysis of the profile of target beneficiaries of the SLEM project and how they compare to non-beneficiaries living in fairly similar communities without a COBA structure. In addition to providing further guidance to CI for the selection and implementation of its activities, this baseline assessment will inform the impact evaluation team on the potential adjustments that would need to be made to the evaluation strategy.

The project team needs six months to complete the implementation of all adaptation and mitigation activities. As a result, in May 2021, a first assessment could be made of the impact of the project after one and a half years of complete interventions. First lessons on the success and potential flaws of the SLEM project will provide an opportunity to adjust the interventions of phase 3 COBA.

In May 2022 and May 2023, 2.5 and 3.5 years after completed implementation, two additional data collections will provide crucial insights on the evolution of the effects of the SLEM interventions over time.

In parallel with these evaluations, the availability of detailed monitoring data will enable the LORTA and project teams to make several assessments of the good conduct of the planned activities. Real-time lessons will be drawn in order to continuously seek to improve the implementation of the SLEM interventions and to reach greater impacts.

Finally, the selection of comparable forest areas outside of the intervention area will lend itself to a long-term comparison with beneficiary areas, thanks to the continuous availability of satellite data. An evaluation of the impacts of the SLEM interventions on deforestation could therefore be pursued long after the end of the project.

Assessment of institutional capacity for impact evaluation

The LORTA team identified a high capacity for M&E within CI. CI-M benefits from qualified experts in various domains, including natural and social sciences, GIS data, and monitoring and data collection, and from the technical support of the CI Moore Center. The high capacity of the project staff was demonstrated during the elaboration of the theory of change, indices for food security and vulnerability anchored in the literature, the household questionnaire and current plans for M&E.

However, the LORTA team identified some weaknesses with respect to impact evaluation. Although the project staff acknowledged the need for a comparison group, the conditions required for this comparison group to be valid were not well understood. In addition, the project staff was unaware of the sample size requirements for impact evaluation. Furthermore, the baseline data collection was scheduled before having a good understanding of the implementation strategy of the SLEM activities.

In such a context, we believe that the different LORTA workshops and group activities that took place during the mission were very beneficial to all stakeholders involved. This formative experience provided the project team with enhanced knowledge on the main requirements of impact evaluation. These discussions also contributed to better foresee the details of the implementation, which will ultimately facilitate and improve the delivery of the SLEM interventions.

Overall, we consider the SLEM project to be in need of support for impact evaluation. The LORTA team is continuing to provide input and feedback remotely, and monthly calls are planned with the project team.

Secondary data sets

The evaluation can benefit from the primary data collected being integrated with several secondary data sources – both at the baseline and later stages.

Non-representative household surveys were collected by CI in 2015, by Ecosystem Services for Poverty Alleviation in 2013 in the areas of CAZ, and in the context of social safeguards interventions of REDD+ in the CAZ and COFAV areas.

More importantly, GIS data are available from different sources, including information on climate, topography and deforestation. Notably, satellite imagery data (Sentinel-2) will be used. CI is currently working on the elaboration of an integrated database for the intervention areas, including available household surveys and administrative data. This database will contain information on, among others, health centres, roads, education level, main forest resources used in the village, presence and type of other partners, firewood, charcoal, fires, aggregated agricultural production for various crops and forest fragmentation.

Plans for monitoring and evaluation

CI is currently working on a detailed report on the M&E plan of the project. The current plan benefited from a workshop activity on high- and low-frequency indicators during the LORTA mission. The main outcomes of this activity are presented in

Table 4.

Table 4 High- and low-frequency indicators

ACTIVITY	OUTPUT (HIGH FREQUENCY)	OUTCOME (LOW FREQUENCY)	IMPACT (SUCCESS)
Adaptation			
For all adaptation activities	Trained people materials distributed	HH adopting techniques Diversified sources of food	Food security (HH index) Resilience (HH index)
Agroforestry and fruit trees	Planted survived trees	HH adopting technique Diversified sources of food	Food security (HH index)
Micro-irrigation and drainage	Km of canals Ha of land benefiting HH benefiting # of people building canals	Rice yields # of harvest of rice/y Ha of shifting cultivation	Food security (HH index)
Intercropping & multi-cropping		HH adopting technique Diversified sources of food Ha of shifting cultivation	Food security (HH index)
Off-season rice cultivation	Ha cultivated off season	HH adopting technique Diversified sources of food Rice yields # of harvest of rice/year	Food security (HH index)

ACTIVITY	OUTPUT (HIGH FREQUENCY)	OUTCOME (LOW FREQUENCY)	IMPACT (SUCCESS)
Mulching		HH adopting technique Diversified sources of food Ha covered by mulching	Food security (HH index)
No tillage		HH adopting technique Diversified sources of food Ha concerned by no tillage	Food security (HH index)
Terracing		HH adopting technique Diversified sources of food Ha covered by terracing	Food security (HH index)
Mitigation			
Per diem patrolling	# of patrollers receiving per diem Km patrolled in missions Grid patrolled # transgression recorded	% of deforestation identified (report versus satellite)	Deforestation rate (GIS satellite)
Equipment	# of patrollers receiving equipment	Equipment working and/or replaced	Deforestation rate (GIS satellite)
Demarcate limits of PA	Km of demarcation	% changes in ha of deforestation compared to in nearby areas (outside protected area)	Deforestation rate (GIS satellite)
Train on legislation	# people trained # COBA trained	% of reported violations prosecuted by authorities changes in self-reported inappropriate use of forests	Deforestation rate (GIS satellite) Protected Area Governance (IUCN index)
SMART SMS	# SMS systems established # people trained	# of enforcement thanks to SMS	Deforestation rate (GIS satellite)

Note: HH = households

Routine monitoring on the project activities will include workshop minutes, training reports and community reports. Training reports will contain the list of participants and feedback on the clarity of the training. Several high-frequency indicators on outcomes specific to each intervention will be collected, such as the number of planted trees that survived, kilometres of canals built, number of people building canals and hectares of land, and number of households benefiting from these canals. For mitigation activities, a SMART-SMS database will be developed and frequently updated based on patrollers' reports and a GPS tracking of their activities. This database will contain information on the number of patrollers receiving per diems, kilometres patrolled in missions, the exact grid patrolled and the number of recorded transgressions. In addition, patrollers will be provided with cameras, and drones are being considered for further monitoring of the forest areas. The LORTA

team has offered to review the updated plan on M&E so that the impact evaluation strategy can benefit the most from a real-time assessment of the SLEM activities.

III. WAY FORWARD

Overall, we consider that the LORTA mission in Madagascar was well received and that it produced promising results. The project team has demonstrated a real interest and engagement in the workshop activities. They seemed convinced of the need of a rigorous impact evaluation and of the benefits of this collaboration. After an evaluation of the potential additional costs implied by the suggested evaluation designs, CI confirmed its interest in moving this impact evaluation forward.

The LORTA team offers continuous support to CI in the conduct of the impact evaluation. In collaboration with the project team, the next steps are the consolidation of the impact evaluation strategy – namely, a final ranodmization of the COBA for phases 1 and 3 based on an updated list of COBA and the identification of local communities for the outside comparison group – a supervision of the revised M&E plans, household questionnaire and interviewer manuals, and an assessment of the usefulness of secondary databases for the impact evaluation.

Regular calls will be organized with the project's key stakeholders on at least a monthly basis.

APPENDIX I

Table 5 Planned agenda of Madagascar LORTA field mission

DAY	PROPOSED ACTIVITIES	SESSION LEAD	PARTICIPATION	DESCRIPTION
Day 0: Sunday, October 21	Arrival of team	-	LORTA team	The LORTA team arrives in country, debriefs and prepares for the week. Check-in at Hotel le Louvre
Day 1: Monday, October 22	9:00–10:00 Introductions, overview and discussion of objectives, Presentation of the team	LORTA team	LORTA team, NDA, CI-M, BNCCC representatives, CI representatives and other experts, project leader and team members	Morning of day 1. A combined meeting is most appropriate. Espace Dera Meeting Room
	10:00–11:00 Project activities plans, stocktaking of documents (Cont'd)	Project team	LORTA team, project leader and team members especially implementation staff	One-hour discussion to update and complement information on the project
	11:15–12:30 Project activities plans, stocktaking of documents	Project team	LORTA team, project leader and team members especially implementation staff	One-hour discussion to update and complement information on the project
	13:30–16:00 Theory of change discussion	LORTA team	LORTA team and all project team	Interactive workshop and discussion
Day 2: Tuesday, October 23	9:00–10:00 The benefits of impact evaluation – discussion and presentation with all key stakeholders	LORTA team	LORTA team, Conservation International Madagascar, BNCCC, CI, NDA, key representatives, project management team and other relevant stakeholders	Invite all implementing agencies, AEs, NDA, key representatives, project management teams and key stakeholders, especially those who oversee high-level decisions, e.g. regarding changes in operational timelines, activities and their modifications, budget lines for M&E and overall project budgets Duration: 1 hour including coffee breaks and discussion
	10:00–11:00 Presentation of the project by the country team	LORTA team	Same as above	Same as above Duration: 1 hour
	11:15–13:00 Capacity-building workshop	LORTA team	LORTA team, participants are project leader and team members, particularly the	What does impact evaluation do? Why is it important? What implications for projects? What can we learn? 1/2 day.

DAY	PROPOSED ACTIVITIES	SESSION LEAD	PARTICIPATION	DESCRIPTION
			implementation teams	This contains some more technical discussion on implementation timelines, comparison groups, e.g. whether a randomized phased-in design could be made operationally feasible.
	13:00–14:00 Lunch Break			
	14:00–15:00 BNCCC courtesy visit	LORTA team, CI headquarter, CI-Moore and Chief of Party, CI-M	Held at the BNCCC Office in Ampandrianomby	30 min to 1 hour
	15:00–16:00 The benefits of impact evaluation – discussion and presentation with all key stakeholders (Cont'd)	LORTA team	LORTA team, CI-M, BNCCC, CI, (NDA, country-specific), key representatives, project management team and other relevant stakeholders	Invite all implementing agencies, AEs, NDA, key representatives, project management teams and key stakeholders, especially those who oversee high-level decisions, e.g. regarding changes in operational timelines, activities and their modifications, budget lines for M&E and overall project budgets Duration: 1 hour including coffee breaks and discussion
3: Day Wednesday, October 24	9:00–11:00 Discussion with project team	Project team	LORTA team, CI representatives, project team, advisers etc.	Discussion of key documents, field visits, identification of targeted beneficiaries, mapping of key stakeholders, geography, timeline of key activities, discussion of anticipated bottlenecks in the project's theory of change. Stocktaking of planned data collection and monitoring information system activities and monitoring and tracking activities.
	11:15–12:30 Discussion with M&E specialist and finance specialist, Cost scenarios Main elements of impact evaluation design	LORTA and project teams	LORTA team, CI representatives, M&E specialist and finance specialist	Discussion on budget lines, needs for sample size, discussion on procurement options for baseline data
	13:30–16:00 Discussion with M&E specialist	LORTA and project teams	LORTA team, CI representatives, M&E specialist	Discussion on budget lines, needs for sample size, discussion on

DAY	PROPOSED ACTIVITIES	SESSION LEAD	PARTICIPATION	DESCRIPTION
	and finance specialist, Cost scenarios Main elements of impact evaluation design (Cont'd)		and finance specialist	procurement options for baseline data CI-M will present the M&E plan and logistics for the household survey
Day 4: Thursday, October 25	9:00–11:00 Debriefing: Presentation to the project team; key elements and timeline of design, discussion of next steps	LORTA team	NDA, CI and BNCCC representatives and other experts, Project leader and team members	
	11:00–11:15 Coffee Break			
	11:15–12:30: Debriefing: Presentation to the project team; key elements and timeline of design, discussion of next steps (Cont'd)	LORTA team	NDA, CI and BNCCC representatives and other experts, Project leader and team members	
Day 5: Friday, October 26	7:30–11:00 Driving to Andasibe 11:30–13:30 Check-in at Feon'ny Ala hotel Lunch at Feon'ny Ala hotel 13:30–17:00 Visit of the Iaroka community ecotourism site 18:00–19:00 Night walk at VMMA Forest	Project team	LORTA and project teams	Visit of the Iaroka community ecotourism site. A CAZ target community Nocturnal visit of the VOI MMA site in Andasibe (optional)

APPENDIX II

Table 6 Actual agenda of Madagascar LORTA field mission

DAY	PROPOSED ACTIVITIES	SESSION LEAD	PARTICIPATION	DESCRIPTION
Day 0: Sunday, October 21	Arrival of team	-	LORTA team	The LORTA team arrives in country, debriefs and prepares for the week. Check-in at Hotel le Louvre
Day 1: Monday, October 22	9:00–11:00 Introductions, overview and discussion of objectives Presentation of the team	LORTA team	LORTA team and CI team	Location: Espace Dera Meeting Room Presentation by Jo Puri of the IEU and LORTA programme, presentation by Markus Olapade of C4ED and impact evaluation, 2*2 matrix game by Jo Puri: initiation to DiD
	11:30–13:00 Project activities plans, stocktaking of documents	Project team	LORTA team and CI team	Presentations by Zo, Clarck and Ando on the overall project objectives, secondary data sets and GIS data, M&E plan and the role of the different stakeholders
	14:00–16:30 Theory of change discussion	LORTA team	LORTA team and CI team	Interactive workshop and discussion Presentation by Clémentine Sadania and group exercises on discussing the underlying assumptions of the existing theory of change of the project
Day 2: Tuesday, October 23	9:00–13:00 The benefits of impact evaluation (including coffee break)	LORTA team	LORTA team and CI team	Presentation by Clémentine Sadania and discussion Presentation of factorial designs by Jo Puri Role-play on Pros and Cons of impact evaluation (Two groups: IE team and members of the Government)
	14:00–15:00 BNCCC courtesy visit	LORTA team, CI GCF	LORTA team and CI team	Meeting with the Director of BNCCC at the BNCCC Office Meanwhile, the rest of CI team worked on the timeline and eligibility criteria of the project, and established a list of other projects funded in the area.
	15:00–16:00 Presentation by CI team	CI team	LORTA team and CI team	Presentation of the above-mentioned exercise
	16:00–16:30 Workshop on selection bias	LORTA team	LORTA team and CI team	Presentation by Clémentine Sadania on selection bias and game on auto-selection versus ranodmization

DAY	PROPOSED ACTIVITIES	SESSION LEAD	PARTICIPATION	DESCRIPTION
3: Day Wednesday, October 24	9:00–11:30 Workshop on selection bias and finding a valid comparison group	LORTA team	LORTA team and CI team	Presentation by Clémentine Sadania Group exercises on finding the appropriate comparison group of the project activities led by Jo Puri
	12:00–13:30 Capacity-building workshop	LORTA team	LORTA team and CI team	Presentation by Markus Olapade on ranodmization and randomized phase-in
	14:30–16:00 Discussion on IE designs and M&E activities	LORTA teams	LORTA team and CI team	LORTA teams Parallel activities: Discussion with M&E specialists on possible IE designs (with M. Olapade and C. Sadania) Group exercises on identifying indicators of success for M&E and IE (led by Jo Puri)
	16:00–16:30 Presentation of main M&E indicators by the project team	Project team	LORTA team and CI team	Presentation of high- and low-frequency indicators and indicators of success (impact indicators) for each of the adaptation and mitigation activities
Day 4: Thursday, October 25	8:30–11:30 Debriefing: Presentation to the project team; key elements and timeline of design, discussion of next steps	LORTA team	LORTA team, CI GCF and M&E specialists	Discussion led by Jo Puri
	11:45–12:30 Debriefing with the all team and closing	LORTA team	LORTA team and CI team	Presentation of the selected design by Jo Puri in answers to the concerns expressed by CI team Closure of the workshop by Zo
Day 5: Friday, October 26	7:30–11:00 Driving to Andasibe 11:30–13:30 Check-in and lunch at Feon'ny Ala hotel 13:30–17:00 Visit of the Iaroka community (COBA) 18:00–19:00 Night walk at VMMA Forest (COBA)	Project team	LORTA and project teams	Visit of the Iaroka community ecotourism site: A CAZ target community. Discussions, Q&A. Visit of the forest area covered by the COBA. Nocturnal visit of the VOI/COBA MMA site in Andasibe

APPENDIX III

Table 7 List of stakeholders engaged with during Madagascar LORTA mission

LORTA WORKSHOP, ANTANANARIVO		
NAME	POSITION	INSTITUTION
Jyotsna Puri	Head of the IEU	GCF
Markus Olapade	Managing director	C4ED
Clémentine Sadania	Junior Impact Evaluation Specialist	C4ED
Rob Merritt	Senior Project Manager	CI GCF Agency
Giacomo Fedele	Scientist	CI Moore Center
Zo Lalaina	Chief of Party	CI-M, Tana office
Bruno Tsing Yat Rajaspera	Director of field projects	CI-M, Tana office
Andoniaina	M&E manager	CI-M, Tana office
Soloson	Adaptation lead	CI-M, Tana office
Mialy	Finance Manager	CI-M, Tana office
Jeannicq	Mitigation lead	CI-M, Tana office
Clarck	GIS Specialist	CI-M, Tana office
Mialy	Finance Manager	CI-M, Tana office
Johari	Grants Manager	CI-M, Tana office
Harison	Ecological Monitoring Manager	CI-M, Tana office
Ando Rabearisoa	Socio-Economist	CI-M, Tana office
Holy	Finance Senior Manager	CI-M, Tana office
Hanta	CAZ Field Director	CI-M, CAZ office
Harinaina	Socio-economist	CI-M, CAZ office
Jean Michel	Forestry Manager	CI-M, CAZ office
Joro Ramorasata	Finance Manager	CI-M, CAZ office
Phillipe	Grant Manager	CI-M, CAZ office
Rejela	COFAV Field Director	CI-M, COFAV office
Zo Zotovonirina	Socio-economist/ Gender specialist	CI-M, COFAV office
Michel	Forestry Manager/ Biologist	CI-M, COFAV office
Haingo	Grant Manager	CI-M, COFAV office
-	-	BNCCC

IMPACT EVALUATION DESIGN REPORT 6: VANUATU

TABLE OF CONTENTS

TABLES

FIGURES

I. INTRODUCTION

A. THE LORTA PROGRAMME

Evaluating the impact of development projects and programmes has gained importance in recent years. Impact evaluation not only allows for increased transparency by measuring the effects of investments, it also provides the opportunity to design and implement development projects more effectively. To contribute to this development, the Independent Evaluation Unit (IEU) of the Green Climate Fund (GCF) has started the Learning-Oriented Real-Time Impact Assessment (LORTA) programme to be able to keep track of GCF projects in terms of performance and results, and to enhance learning within the GCF.

The LORTA programme aims to

- Embed real-time impact evaluations into funded projects so GCF project task managers can quickly access accurate data on a project's quality of implementation and likelihood of impact, and
- Build capacity within projects to design high-quality data sets for overall impact measurement.

The purpose of the impact evaluations is to measure the change in GCF key result areas that can be attributed to project activities. The LORTA programme will inform on returns on GCF investments and help GCF projects track implementation fidelity. The objectives of LORTA include the following:

- Measuring the overall change (outcome or impact) of GCF-funded projects and enhancing learning
- Understanding and measuring results at different parts of theories of change
- Measuring the overall contribution of the GCF to catalysing a paradigm shift and achieving impacts at scale

Currently, the LORTA programme is in Phase I (formative engagement and design). In the first year (2018), the IEU will support eight GCF projects to build high-quality, theory-based impact evaluation designs at inception. Formative work will include engaging with project teams, accredited entities (AEs) and GCF staff; designing theory-based impact evaluations; and establishing protocols for database development.

The second phase of LORTA will involve the main impact assessment stage (3–5 years), while the third phase will include analyzing baseline and endline data, discussing results, and engaging with diverse stakeholders to share results and incorporate feedback as required.

B. THE VANUATU PROJECT

Climate change is one of the most consequential events facing humanity as the world moves through the twenty-first century. For small developing countries and island states, the threat is particularly acute: rising temperatures, rising sea levels, and the increased frequency and intensity of tropical cyclones are the greatest threats to their sustainable development.

Located in the South Pacific, Vanuatu is one of the most vulnerable countries in the world to these climate shocks.[46] With a population of 277,554 and an economy heavily dependent on subsistence

[46] As per the project proposal: https://www.greenclimate.fund/documents/20182/574760/Funding_proposal_-_FP035_-_SPREP_-_Vanuatu.pdf/59266edc-7e5b-4068-a1fe-d4cd3d7437e1

agriculture and fishing,[47] addressing these climate risks is an urgent priority.

The project "Climate Information Services for Resilient Development in Vanuatu", referred to by the abbreviation of its local name, Van-KIRAP (Vanuatu Klaemet Infomesen blong Redy, Adapt, mo Protekt), is one of the eight projects selected to be part of the inception stage (Phase I) of the LORTA programme.

The project is jointly implemented by the Secretariat of the Pacific Regional Environment Programme (SPREP), a regional organization based in Samoa, and the Vanuatu Meteorology and Geohazards Department (VMGD), a national government department. Its overall aim is to increase the resilience of the population against climate change by developing tailored climate information and disseminating this information to community members across five priority sectors: agriculture, fishing, infrastructure, water and tourism. These sectors have been prioritized based on their economic importance to the country's gross domestic product (GDP). The tailored information, referred to as climate information services (CIS), will then be used by community members – also referred to "end users" – to mitigate the negative effects of climate shocks, improve productive decision-making and ultimately lead to more resilient livelihoods.

The project consists of many components, which we divide into three major categories:

Enhancing existing climate infrastructure in order to improve forecasting: These activities will involve installing new and updating existing climate equipment to help generate an even greater body of climate data and improve forecasting. The project will ensure that the new weather and climate infrastructure enhances the development and delivery of CIS throughout the country.

Tailoring the information such that it is sector specific: Although a body of robust climate knowledge currently exists within the VMGD and regionally, the information needs to be repackaged and operationalized in the form of tailored CIS for the specific application of the end users within their respective sectors. Climate information is currently general and not broken down into manageable action recommendations for end users. For example, in the agricultural sector, farmers would benefit more from information on what crops to plant for the coming season based on predicted rainfall patterns as opposed to simply being presented with the season's predicted rainfall.

Improving the communication channels between the providers of CIS and the end users: In addition to the fact that much of the climate information that currently exists is not sector specific, end users do not apply this information because they lack awareness about it and education on how to use it. Further, the current communication structures are one-directional in the sense that they flow from the producers of CIS to the end users but do not allow feedback or clarification from the end users. The result is that a large body of information exists, but is not properly disseminated, understood and applied. Thus, the Van-KIRAP project aims to improve the channels of information dissemination from the VMGD to the end users, and train end users on how to interpret and apply CIS to their daily lives through informal trainings, workshops and capacity-building activities.

II. VANUATU FIELD MISSION

A. GENERAL REMARKS

An evaluation team, consisting of two consultants from the Center for Evaluation and Development (C4ED) (henceforth referred to as the LORTA team), was formed to lead the field mission from 5 to 9 November 2018. The task of the team was to engage closely with key stakeholders of the project – namely, the national designated authority (NDA), AE, implementing agencies and project staff – to

[47] As per the project proposal: https://www.greenclimate.fund/documents/20182/574760/Funding_proposal_-_FP035_-_SPREP_-_Vanuatu.pdf/59266edc-7e5b-4068-a1fe-d4cd3d7437e1.

ensure their interest in and understanding and sense of ownership of the planned theory-based impact evaluation.

During the field mission, the LORTA team held meetings and a capacity-building workshop with the key stakeholders. Meetings, in the form of expert interviews and discussions, were used to acquire the maximum possible information about the GCF-funded project. Stakeholders were interviewed regarding their views about the project's implementation and monitoring strategies, expected impact, challenges and possible solutions. The meetings not only informed the LORTA team about the project but also aimed at fostering collaboration and trust between the team and on-site involved parties. In addition, a capacity-building workshop on impact evaluation was held, targeted at the key stakeholders. Besides conveying technical knowledge, the aim of this workshop was to emphasize the benefit of theory-based counterfactual approaches and real-time learning and measurement.

In coordination with the project team, the LORTA team worked out an impact evaluation design for the Vanuatu project. The LORTA team conducted context analyses, examined the existence of appropriate counterfactuals (i.e. comparable treatment and control groups), elaborated a theory of change, assessed the availability of baseline administrative and secondary data sources, and acquired budget information. The results from this undertaking are presented in the following sections.

B. THE MISSION AGENDA

The LORTA team received earnest collaboration from the SPREP Project Coordination Unit (PCU) and the VMGD. The agenda – shown in appendix I of this design report – was developed to facilitate the joint attendance of all key stakeholders to the LORTA workshop and to allow for the field visit.

In particular, the LORTA workshop was condensed into a day, but contained all the ingredients to inform the key stakeholders on the LORTA objectives, present key impact evaluation concepts, elaborate on the theory of change and discuss programme implementation details with the implementing partners.

The workshop was successful in that it benefited the key stakeholders and the LORTA team. On one hand, the presentations and interactive discussions on theory of change, impact evaluation methods and potential impact evaluation designs brought the key stakeholders together on the same page with respect to ownership and understanding of their contribution to the impact evaluation for the Van-KIRAP project. However, even though the theory of change presentation at the workshop yielded useful information, which helped the LORTA team refine the theory of change, the presence of only one out of five sector coordinators (simply because they had not been not hired yet) made it difficult to crystalize the specific outcomes at the sector level. On the other hand, the LORTA team benefited from gathering in such a short time a rich set of crucial information to design the impact evaluation and benefited from gaining understanding and collaboration from the key implementing agencies for the requirements of an impact evaluation.

The field visit took place at the climate and weather division of the VMGD office, at the envisioned location of a Doppler radar installation close to Port Vila, and at a resource centre that is planned to be enhanced for the project's purposes. The visit helped substantially in gaining a clearer understanding of the implementation status of the Van-KIRAP project and of the climate information production processes and plans. Appendix II of this design report lists all the people engaged with at the workshop and during the field visit throughout the LORTA mission.

During the mission, we were able to establish a tight collaboration with the SPREP Van-KIRAP PCU and the VMGD project team. This enabled us to engage with key stakeholders and to develop evaluation designs in close coordination and feedback loops with the project team.

Discussions throughout the mission week highlighted two focal points of the Van-KIRAP project, which we consider of interest for impact evaluation. The first focal point is concerned with the installation of a Doppler radar, which absorbs almost half of the project's overall budget. The installation of this new Doppler will enable the production and dissemination of Climate Early Warnings (CLEW), which will reach the people of Vanuatu several hours, and in some cases several days, in advance of a weather disaster. While CLEW systems did exist in Vanuatu prior to the project, the Doppler will deliver more accurate predictions farther in advance than was previously possible. The radius of the Doppler radar will not cover all of Vanuatu, however. Quantifying the effect of the Doppler will therefore not only be relevant to inform the investment decisions of other disaster-prone countries, it could also support the decision to install a second Doppler radar reaching the remaining population within Vanuatu in the future.

The second focal point addresses improved dissemination modes of climate information. The field visit made it very clear that a large variety of climate information products is already provided by the VMGD. However, the current system has deficiencies in conveying the climate information to end users. This is where the Van-KIRAP project steps in and develops a multifaceted approach of information dissemination through different communication modes. The Van-KIRAP project makes use of existing government communication structures whenever possible, such as the current chain of communication between the VMGD and island-level area administrators described below in section C. The project will build on these communication structures by adding feedback loops from end users to information providers and incorporating options for reassurance and clarifications. An impact evaluation of this multifaceted approach would also be relevant to facilitate learning in other disaster-prone areas, as well as within Vanuatu to cover sectors beyond the scope of the current project.

Even though we consider both focal points equally relevant, budget constraints will most likely only allow for one impact evaluation. Data-collection processes for the impact evaluation will have to align with the data-collection efforts of a socioeconomic survey (conducted through the Commonwealth Scientific and Industrial Research Organisation (CSIRO)), which is already planned and budgeted for with USD 125,000 (for baseline, midline and endline data collection). In our understanding, the project team was similarly interested in measuring the effect of the Doppler radar and in measuring the effect of the remaining components addressing improved communication channels for climate information. However, the project team was more sceptical about the feasibility of the second approach of CIS dissemination.[48] More details on the Van-KIRAP implementation plan and suggested designs for impact evaluation are provided below.

C. RESULTS

Project implementation plan

Van-KIRAP will span all six provinces of Vanuatu, and is estimated to benefit 70,000 people directly (roughly 30 per cent of the population) through access to improved climate information and training on how to interpret and apply it. It aims to indirectly benefit the remaining 70 per cent of the population through broad stakeholder engagement and delivery of communication products.[49] While the project aims to target the whole country, the activities will necessarily affect some parts more

[48] As explained in detail below, this approach builds on using too-remote areas that cannot be reached by the project as control areas for the impact evaluation. However, it is difficult for the project team to define these "too remote", unreachable areas beforehand, as they would still elicit all possible efforts to reach the whole population. Accordingly, feasibility concerns were higher with this approach and are further described below.

[49] As per the project proposal: https://www.greenclimate.fund/documents/20182/574760/Funding_proposal_-_FP035_-_SPREP_-_Vanuatu.pdf/59266edc-7e5b-4068-a1fe-d4cd3d7437e1

than others, in different ways and at different times. However, there is no conscious plan on the part of the project implementers to stagger the timing of the activities, particularly the dissemination activities, such that certain provinces or certain islands are targeted first.

The key activity in the first project category of improved climate infrastructure is the installation of a Doppler radar on the main island of Efate, planned for September 2019. The Doppler will cover a radius of 400 kilometres. Its location was chosen so that it will cover the largest number of people possible, which at its current location is slightly less than half of the population. Note that the population within this radius may or may not overlap with the 30 per cent of the population who will receive tailored climate information via the sector coordinators and climate champions. See Figure 1 for an approximate illustration of the Doppler radar location and its radius.

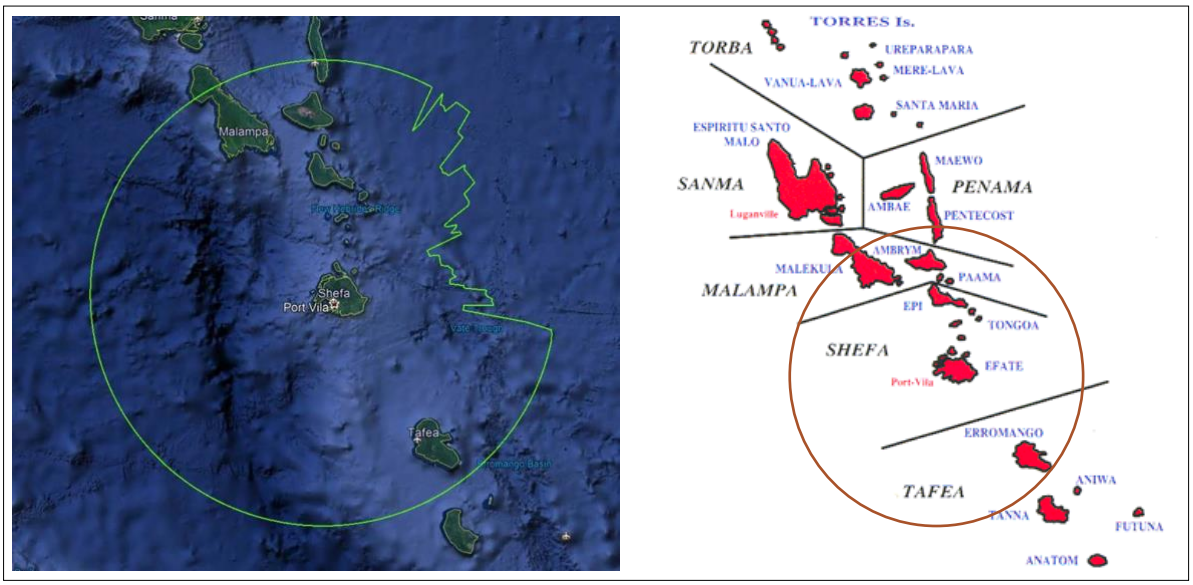


Figure 1 Approximate Doppler radar location and radius in Vanuatu

The advantage of the Doppler over other existing climate hardware is that it has greater predictive power and is able to detect cyclones and storms hours to days in advance. Information will be transmitted to the population via sirens and text messages as major channels, but also through all other potential communication modes such as websites, TV or radio.[50] The Doppler thus has the potential to greatly improve disaster preparedness for the people living within its radius. Outside of this radius, however, information on impending weather disasters cannot be predicted with the same accuracy, and those people will not receive radar-informed messages.[51] Therefore, people living within the radius of the Doppler will receive climate information via the activities of the champions and sector coordinators *plus* text message warnings from the Doppler, whereas those living outside of its radius will only receive the former.

In addition to the Doppler, the project will also install eight new automated weather stations (AWS) at various locations both inside and outside the Doppler's radius, as well as new instruments such as rain gauges, ocean gauges and other weather forecasting tools. While these tools do not have the accuracy and predictive power of the Doppler, they will add to the existing body of climate data used to make long-term forecasts and analyze trends.

Once climatologists and other technical experts have repackaged and tailored the CIS such that it is sector specific, the project will utilize a twofold approach to disseminate the information to end

[50] The text messages are sent out via SMS blasts.
[51] People in the control area will, however, receive personalized, sector-specific (i.e. non-blast) text messages with climate information based on forecast methods not involving Doppler radar information.

users, employing five sector coordinators and 20 selected "climate champions". The climate champions will again appoint two to three "resource men or women", who help them spread information. In total, 47 resource men and women are envisioned to help spread information and train end users across the country. The sector coordinators, climate champions and resource men and women will raise awareness, build capacity and effectively transmit CIS to the end users.

The sector coordinators will be trained by the VMGD on CIS and act as the go-betweens for the VMGD and end users. They will translate climate information from the VMGD into actionable items for end users, help tailor the information and gather feedback from end users on potential changes to the structure of communication.

The community climate champions will be chosen among prominent people from the communities such as village chiefs and other influential people. Together with their respective resource men and women, they will educate community members via meetings and workshops, and build on existing communication channels such as SMS, Facebook, brochures and posters. In addition to these means of communication, they will also chair meetings and workshops at climate centres across the country, 10 of which will be enhanced as part of the project. Unlike the sector coordinators, their work will not necessarily be sector specific but will overarch sectors. Moreover, community champions and resource men and women will work on a voluntary, unsalaried basis.

The training of the sector coordinators is set to take place in early 2019, and that of the climate champions beginning in mid-2019. The sector coordinators will travel to the communities, while the champions and their resource men and women will be stationary but based strategically such that their activities will span the country. However, there are areas that will not be reached due to their remoteness. Further, islands without Internet or strong network coverage will not be reached by certain modes of dissemination such as Facebook and SMS. Therefore, there will necessarily be islands that receive dissemination and training (through workshops and community meetings held by the sector coordinators/climate champions) *and* via various media modes; islands that receive information via one or the other; and islands that receive none at all.

The key beneficiaries of these activities, the end users, differ depending on the sector. In the agricultural sector, the beneficiaries are farmers, who comprise two thirds of the population of Vanuatu and contribute to a significant portion of the country's GDP.[52] In the fishing sector, the final beneficiaries are the fishers. Subsistence fishers are extremely important to the local economy for household income and food security.[53] According to figures from 2009, the Vanuatu National Statistics Office (VNSO) estimates that 77 per cent of households are engaged in some form of fishing activity.[54] This entails that the agricultural farmers and fishing groups overlap – most households do both agriculture and fishing. In the tourism sector, tourism companies are the principal beneficiaries. Tourism is the fastest-growing sector and the country's main foreign exchange earner.[55] It is of great national importance to the economy's development. In the infrastructure sector, the Department of Public Works and private construction companies are the beneficiaries and will be able to implement better decision-making and planning based on CIS. Finally, the end beneficiaries in the water sector are public and private water providers.

[52] Vanuatu GDP share of agriculture – data, chart. TheGlobalEconomy.com. Retrieved from https://www.theglobaleconomy.com/Vanuatu/Share_of_agriculture/
[53] FAO Fisheries & Aquaculture – Fishery and Aquaculture Country Profiles – The Republic of Vanuatu. Retrieved from http://www.fao.org/fishery/facp/VUT/en#CountrySector-Statistics
[54] Bowman, C., Cutura, J., Ellis, A., & Manuel, C. (2009). *Women in Vanuatu: Analyzing challenges to economic participation.* Washington D.C.: World Bank. Retrieved from https://openknowledge.worldbank.org/handle/10986/2624
[55] FAO Fisheries & Aquaculture – Fishery and Aquaculture Country Profiles – The Republic of Vanuatu. Retrieved from http://www.fao.org/fishery/facp/VUT/en#CountrySector-Statistics

As the project is still in the process of hiring staff, the main project activities have not yet begun. The VMGD is in the process of hiring two additional sector coordinators and selecting the community climate champions. The installation of the Doppler is set for September 2019.

The project team conducted five participatory sector workshops for each sector in October 2018. These took place with community representatives for each sector. For each sector, the workshops identified which specific climate information products would be required to support the resilience of end users and which communication modes would be most effective to reach them. The results of the workshops, in the form of Action Plans, will be made available through the project team by early December.

Theory of change

The LORTA team developed two versions of the theory of change: one that focuses on the first category of activities, the installation of climate infrastructure and specifically the Doppler radar, and one that focuses on the second and third categories of activities, the repackaging and dissemination of climate information to end users.

The following theory of change focuses on the first category of activities, specifically the installation of the Doppler radar.

i) Inputs

Approximately USD 10 million will be committed to install and maintain the Doppler. Climatologists and other technical staff at the VMGD are needed to interpret the data from the Doppler.

ii) Activities

The activities will involve the actual installation of the Doppler on the main island of Efate, just outside the capital city. Climatologists at the VMGD will then receive the Doppler information, analyze data and disseminate alerts to end users via sirens and SMS accordingly.

iii) Outputs

The direct outputs of these activities will be that the new Doppler radar is installed and can capture information on cyclones, landslides and floods in a more precise and accurate manner than was previously possible. All end users within the Doppler's 400-kilometre radius will receive data on impending weather disasters via SMS or sirens.

iv) Outcomes

The outcomes will be sector specific. In agriculture, farmers will use the Doppler's warnings to take immediate measures to protect their crops, such as early harvest, if possible, or covering crops with banana leaves. In the tourism sector, tour operators will plan ahead and schedule their activities based on forecasts. In the infrastructure sector, workers will close bridges and roads in advance of weather disasters to avoid accidents. In the fishing sector, fishers will secure their vessels to avoid damage and avoid fishing during rough weather events to decrease injury and loss of life. In the water sector, water companies will close and protect their tanks.

v) Goal

The goal is that end users across all five sectors experience increased resilience of livelihood. For farmers, this may mean higher yields and more stable incomes; for tour operators, this can mean increased usage of tourist attractions due to planning based on CIS, ultimately leading to a higher number of tourists; for infrastructure workers and fishers, this can mean more sustainability of the sector and thus their livelihoods via reduced injury and loss of life – not only for the workers in these sectors, but also for the general population; and for end users in the water sector, this can mean less damage to tanks and therefore more revenue and sustainability.

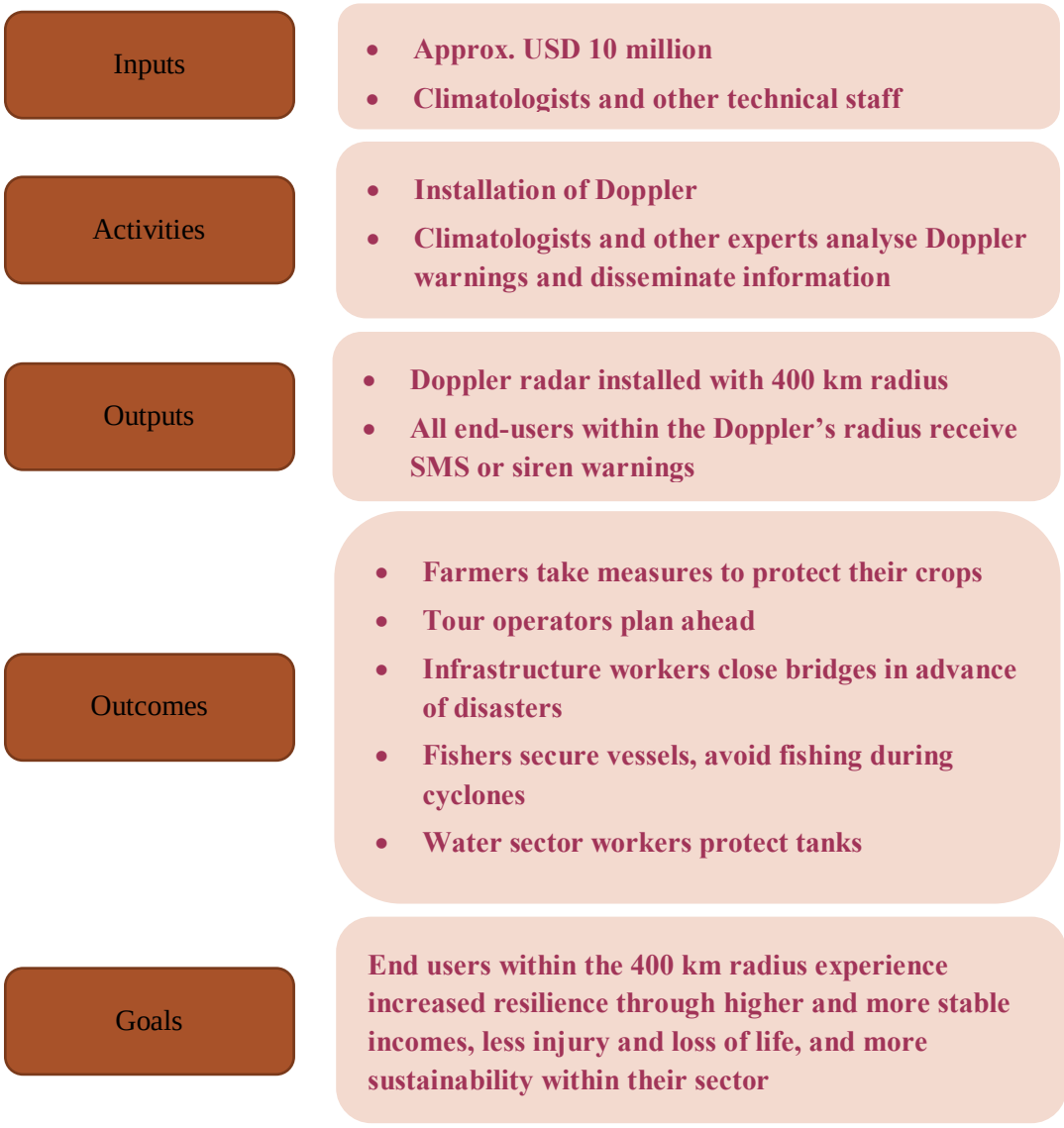


Figure 2 Theory of change for the Doppler radar technology

Below is the theory of change for the second and third categories of activities.

i) Inputs

The budget will be provided partly from the GCF and partly from the Government of Vanuatu. Besides the budget, additional inputs will include the project team, comprising members of SPREP and the VMGD. The core team will include one project manager, one project implementation support officer, one climate information services officer, and one monitoring and evaluation adviser from SPREP, as well as one project implementation manager, and one administrative and finance officer from the VMGD. Additionally, the VMGD will provide five sector coordinators and other technical climate experts.

ii) Activities

The core activities will centre around selecting and training the community climate champions, enhancing climate centres where champions can meet with community members and disseminate information, training end users on how to interpret CIS through meetings with climate champions and sector coordinators, and utilizing cell phone apps, SMS, Facebook and other media to further disseminate climate information. The current structure of communication between the VMGD and the end users in the communities is as follows:

VMGD → Public Works Department → Provincial offices → Area administrators → Communities

However, there is a break in communication between the area administrators and the communities. The information moves seamlessly through government structures but is not effectively getting from the final layer of government to the people. This is the gap the climate champions and the sector coordinators will fill. In addition to their work via workshops, one-on-one training and general word-of-mouth dissemination, the project will make better use of media such as Facebook and SMS, as well as flyers and pamphlets, to effectively communicate information.

iii) Outputs

The result of these activities will be that 20 climate champions are trained in the use of CIS and equipped to train end users, at least 10 climate centres are enhanced and established as places of knowledge exchange and information dissemination, and at least 30 per cent of end users receive sector-specific data on climate trends that they can use and apply to their everyday lives.

iv) Outcomes

The outcomes will be sector specific, and very similar to those above, save for the agriculture sector. In agriculture, farmers will use CIS to change their planting and harvesting schedules based on rainfall patterns, improve their irrigation and drainage systems, and change/diversify the crops they grow in order to harvest more quickly. In the tourism sector, tourism operators will plan ahead and schedule their activities based on forecasts. In the infrastructure sector, workers will close bridges and roads in advance of weather disasters to avoid accidents. In the fishing sector, fishers will secure their vessels to avoid damage and avoid fishing during rough weather events to decrease injury and loss of life. In the water sector, water companies will close and protect their tanks.

v) Goal

The goal is that at least 70,000 end users across all five sectors experience increased resilience of livelihood. For farmers, this may mean higher yields and more stable incomes; for tour operators, this can mean increased revenues via a higher number of tourists; for infrastructure workers and fishers, this can be more sustainability of the sector and thus their livelihoods via reduced injury and loss of life; and for end users in the water sector, this can mean less damage to tanks and therefore more revenue and sustainability.

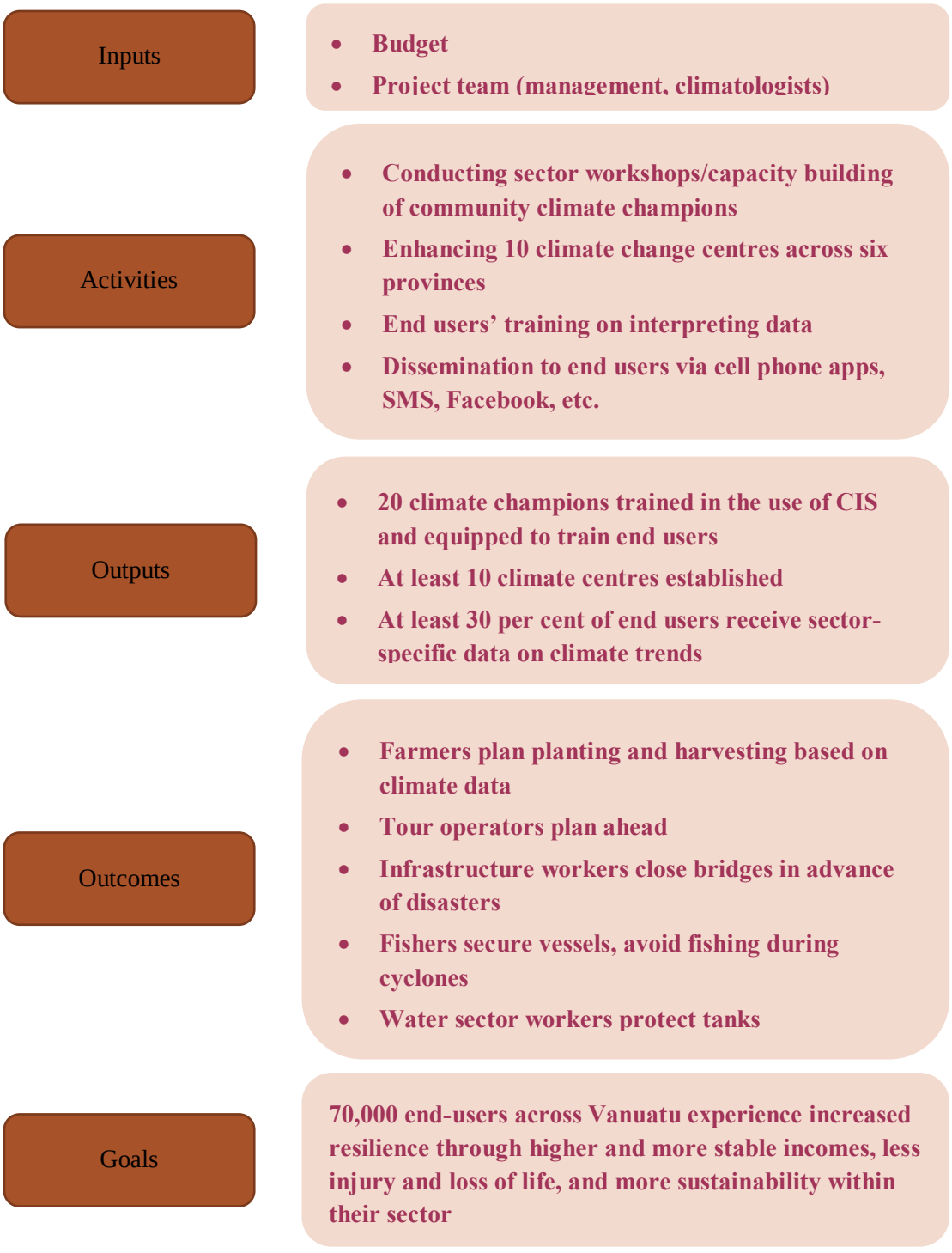


Figure 3 Theory of change for the climate information services dissemination

Based on the theory of change, we propose evaluation questions for which we consider impact evaluation feasible, as outlined below.

Evaluation questions

- Does the early warning capacity of the Doppler lead to increased resilience of livelihood of those living within its radius (measured by less damage, loss of life, loss of crops)?

- Does the use and application of CIS lead to an increase in overall yields of farmers and less volatility thereof?
- Does the use and application of CIS lead to less damage, loss of life and injury for workers in fishing?

Impact indicators

- A) Evacuation numbers during disasters (Disaster Management Office (DMO))

  B) Number of injuries and deaths related to weather disasters (hospital and DMO data)

  C) Farmers: Level and volatility of crop yields (collected via surveys with retrospective yield data per month for the last 12 months)
  - C1) Probability of yield loss
  - C2) Volatility of monthly income (phone surveys)
  - C3) Crop diversification measured by number of new crops planted per season
  - (survey data)
  - C4) Number of high-risk cash crops planted per season to measure perception of
  - risk (survey data)
  - C5) Asset building

  D) Farmers: Climate preparedness perception (collected via surveys asking how prepared farmers feel for weather events and disasters)

  E) Uptake of information (survey data)

  F) Timing of early warnings (i.e. are they indeed earlier, and if so, by how much?)

  G) Tourism: Number of tourists

  H) Fishing: Injuries, loss of life, damage

  Infrastructure: Injuries, loss of life, damage

  J) Water: Damage to water tanks
- A) Average level of crop yields (collected via surveys with retrospective yield data per month for the last 12 months)

  B) Volatility of crop yields (collected via surveys with retrospective yield data per month for the last 12 months)
  - B1) Volatility of monthly income

  C) Number of planting times per farmer per year (collected via survey data)

  D) Preparedness perception of farmers

  E) Uptake of information (survey data)
- A) Number of injuries in fishing (DMO data)

  B) Number of lives lost in fishing (DMO data)

  C) Monetary value of damage in fishing (DMO)

D. PLANS FOR EVALUATION

Approach and methodology

As outlined above, we discovered during our mission that the major benefits of the project, as also perceived by the project team, were twofold. The first major benefit is to improve climate infrastructure to generate earlier and more accurate early warning data, the centrepiece of this being

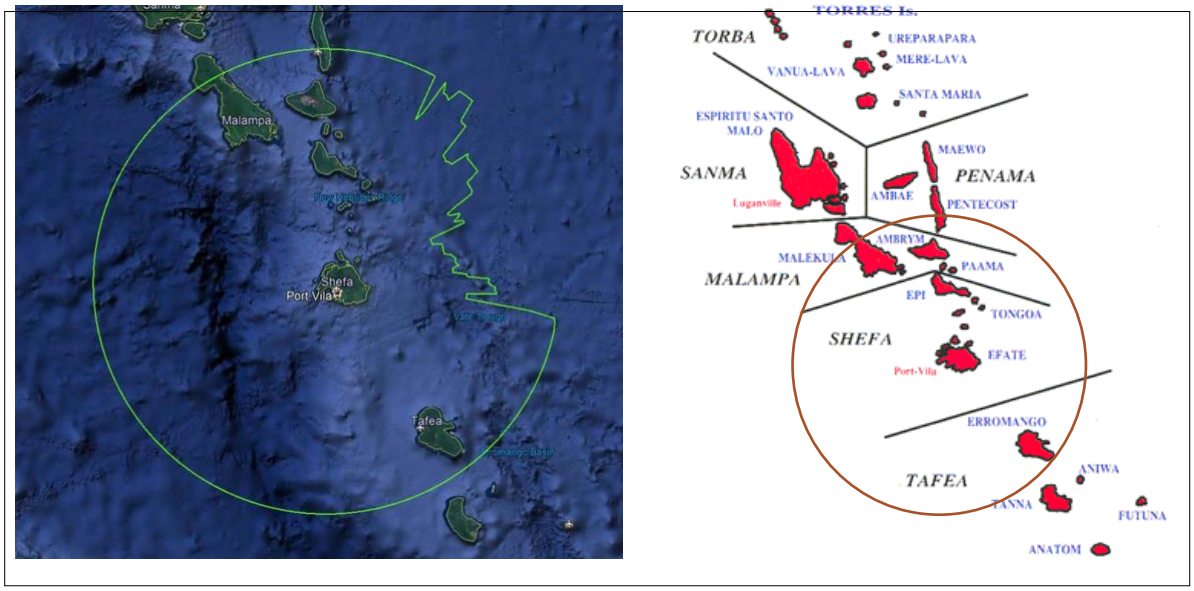
the installation of a Doppler radar. The second important impact the project is aimed at is improving the transmission of climate information to end users. The VMGD visit made it very clear that a large variety of climate products already exists in Vanuatu. However, the uptake of these climate products by end users is low due to several barriers identified through a baseline of the current CIS stage in Vanuatu, conducted by the project team.[56] As described in the discussion of implementation above, the project focuses on improved communication channels to ensure the information 1) actually reaches the end users and 2) is accessed in a comprehensible and actionable format. Accordingly, we developed an evaluation strategy for both focal points. However, budget will most likely only be sufficient for one impact evaluation design.

Both impact evaluation designs focus on a quantitative analysis measuring the impact of the treatment. However, the impact evaluation strategy incorporates joint efforts with the M&E project advisers to additionally gather qualitative information through focus group discussions and semi-structured interviews. This mixed-methods approach makes it possible to quantify the treatment effect on the one hand and explain channels or barriers for the treatment to result into impact on the other hand.

Impact evaluation design

Doppler radar

As noted above, the additional effect of the Doppler radar is to inform earlier about upcoming weather events than is possible with alternative technologies. The information based on these climate early warning systems will reach the people of Vanuatu between an hour and several days prior to the event. Hazards that can be detected by the Doppler are severe weather events such as cyclones, floods and landslides, or less severe but more frequent ones such as heavy rainfalls, wind and hail storms. The fact that the 400 km radius of the Doppler radar will not cover all Vanuatu creates a natural control group, which will not receive the Doppler radar treatment. The approximate Doppler location and its radius are illustrated in Figure 1.



Since the location for the Doppler radar was already pragmatically chosen, a randomized approach is not feasible. Given the implementation design, a difference-in-differences (DiD) design combined with propensity score matching (PSM) can be applied to measure the impact of the Doppler radar

[56] The term "baseline" here refers to the analysis of existing information as well as consultation with national, provincial, sectoral and non-government based stakeholders. This baseline information was collected by the project team as outlined in the GCF funding proposal.

installation on peoples' behaviour and decision-making, and ultimately their resilience to major climate shocks. The location of the Doppler radar was mainly chosen based on coverage of the largest possible parts of the population of Vanuatu and should accordingly not be linked to outcomes, which is crucial to fulfil the DiD prerequisites. Parallel trends, however, can and should still be checked with pre-treatment data.[57]

Pending final implementation decisions by early 2019, the Doppler radar will be installed close to the capital Port Vila on Efate Island in September 2019. Baseline data on outcome variables will therefore have to be collected prior to September 2019. In order to capture the longest possible effect, we propose endline data collection to be scheduled for September 2021. With the project termination envisioned for April 2022, this will hence give sufficient time to produce impact evaluation results by conclusion of the project.

We propose the use of primary survey data as well as secondary administrative data to analyze the impact of the Doppler radar. Concerning administrative data, the DMO[58] in Vanuatu collects data on disaster outcomes, which could be positively affected by earlier warnings through the Doppler technology. These outcomes could be evacuation numbers after cyclones, damage to infrastructure and injuries or deaths. According to the technical advisers in the project team, most climate hazards Vanuatu faces occur at least once per year.[59] Therefore, outcomes should be compared over the period of one year before Doppler implementation (July 2018–July 2019) and one year after Doppler implementation (July 2020–July 2021). The strategy of comparing average outcomes over two 1-year periods before and after, and with and without treatment should result in similarly frequent weather occurrences in the treatment and control region, and thus comparable outcomes, even though the regions were non-randomly selected. To increase precision, we will match households with similar baseline characteristics and similar meso-level indicators.

The administrative data from the DMO focuses, however, on severe disaster events only. To capture the full effect of the Doppler radar, which benefits the population not only during severe disasters but also during moderate weather hazards like heavy rainfall, wind and hailstorm, we propose to complement the analysis with survey data. Baseline survey data could capture perception-based disaster preparedness measures[60] and yield volatility, and compare these measures before and after Doppler installation, between treatment and control areas. Formally, the DiD-estimator (δ) reads as follows:

$$\delta = (Outcome_{t}^{T} - Outcome_{t}^{C}) - (Outcome_{t-1}^{T} - Outcome_{t-1}^{C})$$

The estimator corresponds to the coefficient of the interaction term between the dummy variable for the treatment group and the dummy variable for the second period in a basic regression without additional control variables. $Outcome_{t-1}^{T}$ refers to the outcome of the treatment group before Doppler implementation, $Outcome_{t-1}^{C}$ captures the outcome for the control group before Doppler implementation, while $Outcome_{t}^{T}$ refers to the treatment group's outcome during one year after Doppler implementation and $Outcome_{t}^{C}$ to the control group's outcome during one year after Doppler implementation. Please note that because the treatment and control groups are both

[57] A potential source of heterogeneity could be stemming from the capital region, which is the centre of the treatment region and might be more prone to technology development and dissemination. In case this is so, and pre-trends validate this concern, we propose to exclude the capital island Efate from the exercise. However, a qualitative analysis of existing literature, data sources (see FAO, 2008, and VNSO, 2017, details provided elsewhere in this design report) and discussions with the project team do not support the concern of systematic differences potentially leading to non-parallel trends.
[58] See the website of the National Disaster Management Office of Vanuatu: https://ndmo.gov.vu/.
[59] Cyclones are potentially the most severe weather event and occur once to twice per year, for instance.
[60] Survey questions on preparedness could include how prepared households feel for extreme weather events (measured on a Likert scale) and a binary measure on whether people immediately prepared once they heard a warning for the last occurring weather event.

receiving the remaining project components (improved information channelled through sector coordinators and champions, which is independent of the Doppler radar), we will only capture the additional Doppler effect in our DiD analysis.

Additionally, the DiD analysis could include different treatment arms. The chosen dissemination mode of climate information based on the Doppler radar is primarily via SMS blast. However, other possible communication modes will be mobilized, such as sirens, TV or Internet. A large literature shows dissemination modes of climate information to be crucial for actually reaching people and evoking action.[61] [62] However, the best ways to increase take-up and effectiveness of CIS are still under research. Since the most effective dissemination mode is also unclear in the current setting, the evaluation could additionally explore a behavioural component and assess the effectiveness of different message modifications. Self-evidently, these modifications will not affect the transmission of the crucial warning content of the SMS, but other components could be modified. The not-for-profit organization CARE assessed the importance of a trusted source of climate information in a qualitative analysis in Vanuatu after the Tropical Cyclone Pam had hit the country in 2015.[63] Since a trusted source appears to be relevant to believe in the message and act accordingly, the SMS messages could include the signature of the VMGD or the district leader, or another respected authority involved in CIS production. Moreover, the depth of the message could be modified. One treatment arm might for instance include the most crucial warning only, while another involves warning plus more specific action recommendations per each of the five sectors.

Climate information dissemination

As an alternative to the Doppler radar impact evaluation we propose the assessment of the remaining climate information components with a combination of PSM and DiD approaches. Specifically, we will assess the effect of the installation of new equipment (completed by December 2019), the engaging and informing activities by the sector coordinators (trained by March 2019) and the set up of climate champions (trained by mid-2019). The challenge with this approach is the identification of an adequate control group because the project aims at reaching all actors in the five sectors and a phase-in approach is not feasible. However, the project team considers it impossible to reach very remote areas. For instance, the project team estimated that the Torres Islands in the very far north-western part of Vanuatu will most likely be unreachable due to their geographic distance and lack of universal cell phone coverage. The Torres Islands form part of the Torba Province and host 1,022 inhabitants or 10 per cent of Torba's overall population.[64] The impact evaluation study could find matches between households on the non-treated area of the Torres Islands and the remaining islands of Torba Province with propensity score matching. However, as there might still be differences between the remote Torres Islands and treated remaining islands, with parallel trends, PSM would have to be combined with a DiD approach. Existing census data from the Mini Census from 2016 should be used to approximate the existence of parallel trends.[65] Similar to the identification of the Torres Islands, other remote areas unlikely to be reached by the project would need to be identified in collaboration with the project team to cover a population large enough for impact evaluation. Should the project manage to reach the identified remote areas contrary to prior expectations, the

[61] Cole, S., & Sharma. G. (2017). *The promise and challenges in implementing ICT for agriculture.* Mimeo. Preliminary draft.
[62] Aker, J. (2011). Dial "A" for agriculture: a review of information and communication technologies for agricultural extension in developing countries. *Agricultural Economics, 42*(6), 631–647.
[63] CARE International. (2017). Does gender responsive disaster risk reduction make a difference? A comparative study of category five tropical cyclone Pam in Vanuatu. CARE International.
[64] VNSO. (2017). *2016 Post-TC Pam mini census report. Volume 1.* Port Vila, Vanuatu: Vanuatu National Statistics Office. Retrieved from https://vnso.gov.vu/index.php/2-uncategorised/153-mini-census.
[65] VNSO. (2017). *2016 Post-TC Pam mini census report. Volume 1.* Port Vila, Vanuatu: Vanuatu National Statistics Office. Retrieved from https://vnso.gov.vu/index.php/2-uncategorised/153-mini-census.

impact evaluation strategy has to find a coping mechanism. We propose to focus the evaluation on sectors that are most prevalent among the population of Vanuatu and that are therefore most likely to provide the highest density of households involved. With regard to power considerations, this strategy makes it possible to find the highest possible sample size for impact evaluation. We propose to oversample the control group households in order to be left with a sufficient number of households for matching in case parts of the control group are eventually reached by the project. We therefore suggest focusing this evaluation design on households engaged in agriculture, since they represent the largest part of the population, spanning approximately two thirds of it.

For the impact evaluation, a baseline data collection in early 2019 would first need to collect all necessary characteristics of a sample of farmers that could influence the outcomes, most importantly crop yields and their volatility. The characteristics would be used to find comparable matches between treatment and non-treatment households. Second, baseline data collection would measure the outcomes the impact evaluation focuses on – namely, yield level and volatility, planting times and weather preparedness for the agricultural analysis. During the endline data collection in July 2021, the same survey questions would be assessed. Again, the DiD approach allows for differences in baseline characteristics, as long as parallel trends for the development of outcomes are given. In this regard, we propose to find matches within the same province, which should provide similar political circumstances and spread of new technologies. Additionally, geographical closeness speaks to presumably similar susceptibility to weather events.

Sampling

Household listings including households' occupation are available from the VNSO. The VMGD team will be able to procure the listings of farming households, which are needed for the evaluation. For the Doppler approach, we need to sample households working in agriculture. The fishing, tourism, infrastructure and water sectors will be analyzed with administrative data only. Through the listings and additional information on cellular network coverage, we will identify the total population of households in agriculture. From the full population with cell phone coverage, we will randomly select households for the evaluation sample based on power calculations presented below. In case the listings provided by the VNSO are insufficient, the sector coordinator for agriculture who works across Vanuatu will be able to produce the listing for his sectors.

Similarly, for the second impact evaluation approach concerning climate information dissemination, we would need household listings for households engaged in farming. To receive the household listings, we can follow the same approach as for the Doppler radar impact evaluation. However, we would also include households without cell phone coverage but would take note of this characteristic for the analysis of heterogeneous effects with and without access to a cellular network. We will then again randomly select households according to the power calculations. We should oversample the control group to guarantee adequate matches can be found for the envisioned PSM approach.

Power calculations

Power calculations enable us to determine the minimum sample size needed in order to detect the impact of the project. To do that, we use the following power formula that relates the sample size to the minimum detectable effect size (MDES) between the mean outcomes of two groups:

$$MDES = (t_{1-\kappa} + t_{\alpha}) \sqrt{\frac{1}{P(1-P)}} \sqrt{\frac{\sigma^2}{N}} \sqrt{1 - R^2}$$

where $t_{1-\kappa}$ and t_{α} are t-statistics representing the required power and level of statistical significance (by convention, we seek a power of 80 per cent and a statistical significance of 5 per cent), P represents the proportion in one of the two compared groups (allocation ratio), σ^2 is the

variance, N is the total sample size, and R^2 represents the extent to which baseline characteristics predict the endline yields. We will not follow a clustered approach as our treatment is not clearly delivered on a group basis.

The above formula allows for a comparison between two groups. In the impact evaluation strategy for the Doppler radar effect above, we proposed to add a behavioural component and assess the effectiveness of different message modifications. We proposed the inclusion of a trust-generating component in the messages. Accordingly, the impact evaluation would analyze three different groups:

- Farmers in the treatment area of Doppler radius, who receive warning messages with an additional trust-creating component (treatment group; T)
- Farmers in the treatment area of Doppler radius, who receive warning messages with regular content (control group in treated area; CT)
- Farmers in the control area, who do not receive SMS warnings with higher precision based on the Doppler radar information (control group in control area; PC)

We will then be interested in two comparisons: *T* versus *PC* in order to measure the impact of improved Doppler information transmitted through SMS messages with a trust component, and *PC* versus *CT* to estimate the impact of improved Doppler information transmitted through a regular SMS message. The sample size will need to be higher compared to a standard situation with one treatment and one control group.

Since the *PC* group appears in both comparisons, we should try to maximize the number of households in it. On the other hand, when sample sizes of the two groups are not proportionate – that is, the ratio is different from 50:50 – the power decreases. Therefore, we have to find an optimum so that the two forces are balanced out.

For the second approach, in which we propose to analyze the effect of the remaining CIS components, we will consider the following comparisons:

- Farmers in the treatment area, who receive CIS from sector coordinators, champions and other communication modes (treatment group T)
- Farmers in the control area of remote islands, who are not reached by the project components (control group C)

As we only have two groups here, we will be able to apply the optimal sample allocation ratio of 50:50 – that is, equal sample size in treatment and control group.

Impact on yields

Data on yields for the major crops produced in Vanuatu are available from the Food and Agriculture Organization (FAO) database.[66] However, to the best of our knowledge, available databases and studies do not include precise information on standard deviations. We will therefore build our power calculations on assumptions based on standard deviations found in other countries. As 82 per cent (45,195) of the households in Vanuatu are engaged in banana farming, we will focus on banana yields for our power calculations.[67] The mean yield for banana is 9.81 t/ha. We will borrow estimates for standard deviations from other studies focusing on agricultural production. The (arguably strong) assumption is that agricultural productions vary to a similar extent over different

[66] FAOSTAT. Retrieved from http://www.fao.org/faostat/en/#compare
[67] VNSO. (2017). *2016 Post-TC Pam mini census report. Volume 1.* Port Vila, Vanuatu: Vanuatu National Statistics Office. Vol. 1. Retrieved from https://vnso.gov.vu/index.php/2-uncategorised/153-mini-census.

crops in different countries. Komarek et al.[68] performed a study on maize yields in Malawi and found average maize of 1.8 t/ha with a standard deviation of 1.17 t/ha, which relates to 65 per cent of the mean value. We assume the same relation of the standard deviation to the mean for banana production in Vanuatu. We therefore estimate the standard deviation to be at 6.37 t/ha, which is 65 per cent of the mean value of 9.81 t/ha.

We also rely on studies conducted outside Vanuatu when it comes to the expected effect size. Patt et al.[69] found yields for farmers in Zimbabwe increased by 19 per cent when applying forecast information. Their effect was compared to farmers not benefiting from the provided training and improved climate information. After several years, the effect was still a 9 per cent increase.

For our first impact evaluation approach of the Doppler radar effect, circumstances are different from Patt et al. in two regards. On the one hand, the (Doppler) treatment does not contain training, which would reduce the expected effect, but on the other hand, the improved and earlier early warning information may have a larger impact on farmers than general forecasting information. We therefore believe that a conservative estimated effect size of 10 per cent is adequate to draw conclusions on the sample size required. An MDES of 10 per cent implies that the calculated sample size will be sufficient only to detect significant effects of the treatment if these correspond to an increase of yields by at least 10 per cent.

Table 1 shows results for the MDES at different sample sizes, values of R^2 and standard deviations.

Table 1 Power calculations for the impact evaluation on the Doppler radar effect for the indicator "Banana yield"

	MEAN	BASELINE STD. DEVIATION	TOTAL SAMPLE	R2	SAMPLE SIZE IN C PURE	SAMPLE SIZE IN C IN T OR T	ALLOCATION RATIO (C PURE)	MDES	% CHANGE
1	9.81	6.37	3,000	30%	1,242.6	878.7	0.59	0.66	7%
2	9.81	6.37	3,000	0%	1,242.6	878.7	0.59	0.79	8%
3	9.81	9.81	3,000	30%	1,242.6	878.7	0.59	1.01	10%
4	9.81	9.81	3,000	0%	1,242.6	878.7	0.59	1.21	12%
5	9.81	6.37	2,000	30%	828.4	585.8	0.59	0.81	8%
6	9.81	6.37	2,000	0%	828.4	585.8	0.59	0.96	10%
7	9.81	6.37	1,500	30%	621.3	439.35	0.59	0.93	9%
8	9.81	6.37	1,500	0%	621.3	439.35	0.59	1.11	11%

We aim to reach at least a percentage change (last column, Table 1) of 10 per cent. To be on the conservative side and to take potentially larger actual standard deviations into account (rows 3 and 4), we suggest a total sample of 3,000 households. Of these, 1,242 households will be part of the pure control group, and 879 households will be surveyed for each of the two treatment groups.

[68] Komarek, A. M., Drogue, S., Chenoune, R., Hawkins, J., Msangi, S., Belhouchette, H., & Flichman, G. (2017). Agricultural household effects of fertilizer price changes for smallholder farmers in central Malawi. *Agricultural Systems 154*, 168–178. https://doi.org/10.1016/j.agsy.2017.03.016

[69] Patt, A., Pablo Suarez, P., & Gwata, C. (2005). Effects of seasonal climate forecasts and participatory workshops among subsistence farmers in Zimbabwe. *PNAS 102*(35), 12623–12628.

For the second impact evaluation approach of the effect of the remaining CIS components, which include capacity-building, treatment circumstances are comparable to the study of Patt et al.[70] In order to remain conservative, we again aim for an MDES of at least 10 per cent. Table 2 shows results for the MDES at different sample sizes, values of R^2 and standard deviations for the second proposed impact evaluation.

Table 2 Power calculations for the impact evaluation on the climate information effect for the indicator "Banana yield"

	MEAN	BASELINE STD. DEVIATION	TOTAL SAMPLE	R2	SAMPLE SIZE IN C	SAMPLE SIZE IN T	ALLOCATION RATIO	MDES	% CHANGE
1	9.81	6.37	3,000	30%	1,500	1,500	0.50	0.55	6%
2	9.81	6.37	3,000	0%	1,500	1,500	0.50	0.65	7%
3	9.81	9.81	3,000	30%	1,500	1,500	0.50	0.84	9%
4	9.81	9.81	3,000	0%	1,500	1,500	0.50	1.00	10%
5	9.81	6.37	2,000	30%	1,000	1,000	0.50	0.67	7%
6	9.81	6.37	2,000	0%	1,000	1,000	0.50	0.80	8%
7	9.81	6.37	1,500	30%	750	750	0.50	0.77	8%
8	9.81	6.37	1,500	0%	750	750	0.50	0.92	9%

Power calculations presented in Table 2 suggest a total sample size of 2,000 households. Rows 5 and 6 show that an increase of 7 per cent or 8 per cent can be detected based on 1,000 control households and 1,000 treatment households. However, we recommended oversampling of the control group to guarantee the identification of adequate matches, even if parts of the control group might be reached by the intervention. We thus propose oversampling of the control group by 30 per cent, leading to a final sample size of 2,300 households.

Possible risks for impact evaluation

Possible risks linked to our proposed impact evaluation strategies are outlined below.

Doppler radar

A potential limitation linked to the Doppler radar evaluation approach is that weather disasters during our evaluation period will not be the most severe possible for Vanuatu and thus the maximal potential benefit of the Doppler technology may not be revealed. The severity of occurring disasters has to be qualitatively captured in the impact evaluation report and evaluation results interpreted accordingly. However, by analyzing and comparing two 1-year periods in treatment and control groups, the likelihood of not capturing severe disasters is low. As explained above, this is because most climate hazards Vanuatu faces occur at least once per year, according to the technical advisers in the project team. The potentially most severe weather event of a cyclone occurs once to twice per year, for instance.

We might face the risk of non-parallel trends between the treated area within the radius of the Doppler radar and the non-treated area outside the radius. This concern might be valid because the

[70] Patt, A., Pablo Suarez, P., & Gwata, C. (2005). Effects of seasonal climate forecasts and participatory workshops among subsistence farmers in Zimbabwe. *PNAS 102*(35), 12623–12628.

treatment area includes the capital island, where spread of technology might be faster than in other areas. Based on our baseline data collection, we will be able to capture this possibility. If we indeed find non-parallel trends due to the capital island, we can exclude it from the final evaluation. However, we do not consider this possibility likely, as the control area includes the island Espiritu Santo, which is as developed and economically important as Efate, the capital island. Also, disaster preparedness seems to be comparable, because Espiritu Santo is the only island beside Efate to possess sirens.

A third concern is related to parts of the control area in the northern part of the archipelago being covered by the radius of a Doppler radar from New Caledonia. However, we can identify these areas and either exclude them from our DiD analysis or control for them with a binary indicator.

Climate information dissemination

The largest concern for the second proposed approach is that the a priori defined control areas could eventually be reached by the project's activities, even though the project team currently estimates that they will not be able to reach them. To account for this risk, we will have to oversample our control group to make sure a sufficient number of non-treated households will remain even after potential contamination of the control.

Qualitative assessments

There are plans within the existing M&E structure to conduct focus group discussions (FGDs) at least twice per year. The LORTA team met individually with the SPREP M&E adviser to determine how efforts for this activity could be combined. A collaboration is possible, and discussions on the content of the FGDs are ongoing. The M&E adviser's first field visit for qualitative data collection will be in early 2019. Ideally, the activities will be scheduled before the impact evaluation's baseline survey is conducted and will inform the questionnaire development (described in the following section). However, this is open to discussions with the M&E team. The current plan is that there will be three interviews conducted at the VMGD level and 25 FGDs with end users distributed across the islands. Overall, we will accordingly have information from at least six interviews and 50 FGDs per year.

From the impact evaluation perspective, the purpose of the FGDs will be twofold. First, they will provide information on behavioural changes and coping mechanisms for climate disasters that may not have been apparent until then; they will therefore serve as a starting point for developing and refining the baseline questionnaire. Second, FGDs will add context and substance to the quantitative data such that we can better understand the choices people make to respond to disasters and find potential explanations for why we do or do not see particular changes. An additional benefit of FGDs will be that they will enable us to speak with end users across all sectors, whereas the quantitative survey evaluation will necessarily be restricted to only agriculture.

Given that SPREP already plans to conduct FGDs, there would be no additional cost for the qualitative assessment informing the impact evaluation.

Data collection

There will be two quantitative data-collection rounds. Their timing depends on the chosen evaluation design. For the impact evaluation of the Doppler radar, a baseline would be conducted shortly before the radar implementation, thus in July 2019. For the impact evaluation of the remaining CIS components, the baseline would take place before the start of all project activities, thus in February 2019. Endline would – in both cases – take place at the latest possible time, in July 2021, about two to two and a half years later. The baseline and endline surveys should be combined with CSIRO's socioeconomic survey, which has three planned rounds of data collection. However,

the exact potentials for cooperation still need to be discussed with CSIRO directly. According to the project team, it should be possible to include questions for the impact evaluation in the socioeconomic survey.

Timeline of evaluation

The Van-KIRAP project is at a relatively early stage, so it is still possible to incorporate requirements for the impact evaluation in the project timeline.

Table 3 depicts a proposed timeline of the Doppler radar impact evaluation, and Table 4 shows the timeline of the impact evaluation of the remaining climate information components. The two proposed timelines differ mainly with regard to the timing of the baseline data collection, as this should take place at the latest point possible before project implementation.

Table 3 Evaluation timeline Doppler impact evaluation

		2019				2020				2021				2022
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
Doppler	**Activity**													
1	Desk study													
2	Survey development													
3	Baseline													
4	Introduction of CI project components													
5	Doppler implementation													
6	M&E activities													
7	Endline													
8	Data analysis													
9	IE results dissemination													

Table 4 Evaluation timeline CIS impact evaluation

		2019				2020				2021				2022
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
CI	**Activity**													
1	Desk study													
2	Survey development													
3	Baseline													
4	Introduction of CI project components													
5	Doppler implementation													
6	M&E activities													
7	Endline													
8	Data analysis													
9	IE results dissemination													

Assessment of institutional capacity for impact evaluation

The project team was thoroughly convinced of the necessity of impact evaluation and eager to add it to their existing M&E strategy, even if the LORTA team would not support them. The team was very helpful in formulating a theory of change and an impact evaluation design, and it was clear that although M&E planning was still in its early stages, they have a high capacity to develop sound monitoring systems for all project activities. The M&E adviser from SPREP brings a wealth of multi-country experience to the project. As the primary competency of the M&E advisers lies, evidently, in monitoring activities, we would recommend having an impact evaluation conducted or advised by impact evaluation specialists.

Secondary data sets

The following secondary data sets will be used for the analysis. This list is not exhaustive, and other sources may be added.

- Data on disaster outcomes from the DMO (https://ndmo.gov.vu/resources/downloads)
- Mini Census data on population characteristics in 2016, after Cyclone Pam in 2015 (https://vnso.gov.vu/index.php/2-uncategorised/153-mini-census)
- Data on household lists and locations from the VNSO (https://vnso.gov.vu/)
- Data to track disaster-related injury and death from hospitals (discussed with project team that access would be granted as long as the project team can set up a memorandum of understanding with hospitals)
- Historical data on agricultural production from FAO (http://www.fao.org/faostat/en/#compare)
- Historical climate data with records dating back to 1973 for the capital island Efate and the second urban centre of Vanuatu, the island Espiritu Santo (http://www.fao.org/docrep/pdf/011/i0530e/i0530e02.pdf)

Plans for monitoring and evaluation

The LORTA team discussed the current plans for M&E in-depth with the SPREP M&E adviser. The following are monitoring sources that the project team could implement to inform an impact evaluation or access from existing M&E plans:

1. Lists that track the number of people who enter the climate centres and on which days. These data would make it possible to see the number of people who are directly exposed to the climate information and teachings from the climate centres.
2. Sector coordinator tracking sheets. This would be a record of all the sector coordinators' interactions with end users, including workshops held, information distributed, communities visited, etc.
3. Data on Internet access throughout the country. Particularly for dissemination modes like Facebook, access to the Internet is crucial. Adding a dummy variable to the data to capture whether or not the village has Internet access will be helpful in interpreting the results of different dissemination modes. The VMGD agreed to support the access to village-level Internet coverage data.
4. Records that track where paper-based information such as brochures and pamphlets are sent. To the extent of our knowledge, a clear record of this is not currently kept; however, it can be implemented. Another dummy variable can be added to the data based on these records to capture whether a household received paper-based information or not.

5. Records that track the number and content of warnings sent out.
6. VMGD requests for data. The VMGD keeps track of all requests for additional climate data they receive. The use of these data would make it possible to see if the number of people inquiring about further data increases over the course of the project.

III. WAY FORWARD

Overall, we consider that the LORTA mission in Vanuatu was well received and that it produced promising results. The success of the LORTA mission has been particularly achieved thanks to the attentive collaboration of the core project team.

The LORTA team continues communication remotely, especially regarding access to existing data sources and procurement of information necessary to plan the impact evaluation, such as a clear mapping of Internet and cellular network coverage in Vanuatu. We are also awaiting the finalization of the Sector Action Plans, which at time of writing are being prepared by the project team and shall be shared in early December. Moreover, the future success of this project in the LORTA framework is highly conditional upon a thorough collaboration with the contracted survey organization, CSIRO, which will conduct a socioeconomic survey. At time of writing, the project team has established contact with CSIRO for us, and a first talk is scheduled for the second week of December 2018. As no additional funds are available, the impact evaluation baseline and endline surveys have to be fully incorporated into CSIRO's socioeconomic survey.

APPENDIX I

Table 5 Agenda of Vanuatu LORTA field mission

DAY	PROPOSED ACTIVITIES	PARTICIPATION
Day 0: Sunday, November 4	Arrival of team	LORTA team
Day 1: Monday, November 5 Morning	Introductions, overview and discussion of objectives	LORTA team, SPREP, VMGD
Afternoon	Presentation and discussions of the project and progress to date by the project team	Project leader and team members especially implementation staff
Day 2: Tuesday, November 6 Morning	Field trip: visit VMGD offices; tour through the climate, weather and forecasting divisions with technical explanations about the work	LORTA team, SPREP, VMGD
Afternoon	Travel to resource centre, envisioned to be used for project purposes and location of Doppler radar installation	LORTA team, SPREP, VMGD, Potential climate champion
Day 3: Wednesday, November 7 Workshop Morning	By LORTA team: Benefits of Impact evaluation, and LORTA approaches. By LORTA team: introduction into impact evaluation methods	By LORTA team Participants are SPREP and VMGD
Workshop Afternoon	By LORTA team: interactive workshop and discussion of the theory of change By LORTA team: discussion of potential evaluation designs for Van-KIRAP	By LORTA team Participants are SPREP and VMGD
Day 4: Thursday, November 8	M&E discussions: current stage and potential to collaborate	M&E officers
	Budget discussions: discuss available budget for impact evaluation	SPREP project leader and VMGD project manager
Day 5: Friday, November 9	Debriefing: Presentation to the project team and discussion of key elements and timeline of impact evaluation design	LORTA team, SPREP, VMGD

APPENDIX II

Table 6 List of stakeholders engaged with during Vanuatu LORTA mission

MONDAY, LORTA INTRODUCTORY MEETING, VMGD OFFICE, PORT VILA		
NAME	POSITION	INSTITUTION
Johanna Johnson	Project Manager Vanuatu, SPREP	SPREP Van-KIRAP
Moirah Matou	Manager VMGD Van-KIRAP	VMGD
Pakoa Leo	Sector Coordinator Agro-Meteorology	Van-KIRAP
Rebecca Polestico	Monitoring and Evaluation Adviser	SPREP
Sunny Seuseu	Climate Information Technical Adviser	SPREP Van-KIRAP
Vitolina Samu	Monitoring and Evaluation Specialist	SPREP
TUESDAY, FIELD VISIT, VMGD OFFICE, PORT VILA		
NAME	POSITION	INSTITUTION
Fred Jockley	Manager of Weather Forecast	VMGD
Melinda Natapei	Acting Manager Climate Service	VMGD
Johanna Johnson	Project Manager Vanuatu, SPREP	SPREP Van-KIRAP
Moirah Matou	Manager VMGD Van-KIRAP	VMGD
Rebecca Polestico	Monitoring and Evaluation Adviser	SPREP
Sunny Seuseu	Climate Information Technical Adviser	SPREP Van-KIRAP
Vitolina Samu	Monitoring and Evaluation Specialist	SPREP
TUESDAY, FIELD VISIT, RESOURCE CENTRE, EFATE		
NAME	POSITION	INSTITUTION
Kenneth	Champion Van-KIRAP	Van-KIRAP
Johanna Johnson	Project Manager Vanuatu, SPREP	SPREP Van-KIRAP
Moirah Matou	Manager VMGD Van-KIRAP	VMGD
Rebecca Polestico	Monitoring and Evaluation Adviser	SPREP
Sunny Seuseu	Climate Information Technical Adviser	SPREP Van-KIRAP
Vitolina Samu	Monitoring and Evaluation Specialist	SPREP
WEDNESDAY, LORTA WORKSHOP, VMGD OFFICE, PORT VILA		
NAME	POSITION	INSTITUTION
Fred Jockley	Manager of Weather Forecast	VMGD
Raysen Vire	Sector Coordinator Infrastructure	Van-KIRAP
Sunny Seuseu	Climate Information Technical Adviser	SPREP Van-KIRAP
Moirah Matou	Manager VMGD Van-KIRAP	VMGD
Pakoa Leo	Sector Coordinator Agro-Meteorology	Van-KIRAP
Vitolina Samu	Monitoring and Evaluation Specialist	SPREP

Johanna Johnson	Project Manager Vanuatu, SPREP	SPREP Van-KIRAP
Rebecca Polestico	Monitoring and Evaluation Adviser	SPREP
Melinda Natapei	Acting Manager Climate Service	VMGD

IMPACT EVALUATION DESIGN REPORT 7: ZAMBIA

TABLE OF CONTENTS

TABLES

FIGURES

I. INTRODUCTION

A. THE LORTA PROGRAMME

Evaluating the impact of development projects and programmes has gained importance in recent years. Impact evaluation not only allows for increased transparency by measuring the effects of investments, it also provides the opportunity to design and implement development projects more effectively. To contribute to this development, the Independent Evaluation Unit (IEU) of the Green Climate Fund (GCF) has started the Learning-Oriented Real-Time Impact Assessment (LORTA) programme to be able to keep track of GCF projects in terms of performance and results, and to enhance learning within the GCF.

The LORTA programme aims to

- Embed real-time impact evaluations into funded projects so GCF project task managers can quickly access accurate data on a project's quality of implementation and likelihood of impact, and
- Build capacity within projects to design high-quality data sets for overall impact measurement.

The purpose of the impact evaluations is to measure the change in GCF key result areas that can be attributed to project activities. The LORTA programme will inform on returns on GCF investments and help GCF projects track implementation fidelity. The objectives of LORTA include the following:

- Measuring the overall change (outcome or impact) of GCF-funded projects and enhancing learning
- Understanding and measuring results at different parts of theories of change
- Measuring the overall contribution of the GCF to catalysing a paradigm shift and achieving impacts at scale

Currently, the LORTA programme is in Phase I (formative engagement and design). In the first year (2018), the IEU will support eight GCF projects to build high-quality, theory-based impact evaluation designs at inception. Formative work will include engaging with project teams, accredited entities (AEs) and GCF staff; designing theory-based impact evaluations; and establishing protocols for database development.

The second phase of LORTA will involve the main impact assessment stage (3–5 years), while the third phase will include analyzing baseline and endline data, discussing results, and engaging with diverse stakeholders to share results and incorporate feedback as required.

B. THE ZAMBIA PROJECT

The project "Strengthening climate resilience of agricultural livelihoods in Agro-Ecological Regions I and II" aims to increase the resilience of smallholder farmers in Zambia in view of climate change and climate variability. Zambia is divided into three major agroecological regions, and the project focuses on two of them – Agro-Ecological Regions I and II – which are shown in Figure 1 below. Region I, in the southern portion of the Southern and Western provinces, is one of the hottest, driest and poorest regions in Zambia. Region II has three subregions (IIa1 and IIa2, and IIb) and is a medium-rainfall belt running East–West through the centre of the country. It is an area with relatively good soils and receives more rainfall than Region I. Region IIb, while often considered a part of Region II, is differentiated from the other parts of the region and can be characterized as a low-rainfall area.

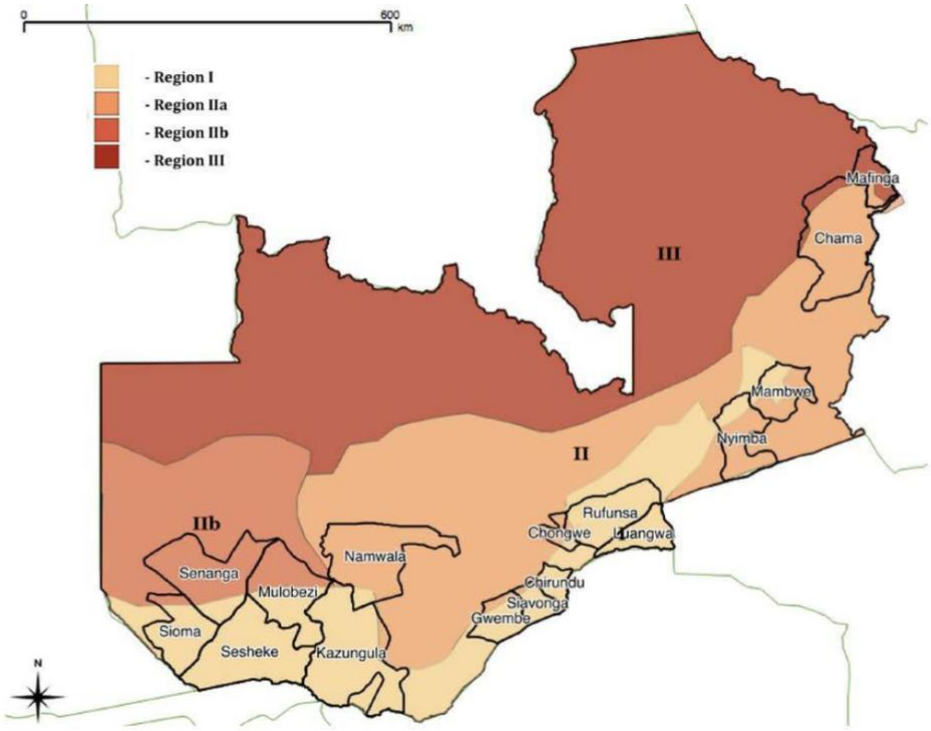


Figure 1 Map of major agro-ecological regions in Zambia

Smallholder farmers in Regions I and II face increasing risks as a result of climate change, primarily due to variability of rainfall and increased frequency of droughts, which directly affect agricultural production. Furthermore, these regions have a very high poverty incidence, rain-fed agriculture is predominant and there is a lack of crop diversity (e.g. Regions I and II mostly grow maize, which is not very resistant to climate change, whereas they grow the more resistant cassava to only a limited extent).

The project aims to achieve increased resilience by taking a value chain approach, addressing barriers to climate-resilient agriculture across key stages of the value chain – planning, inputs, production and post-production – through various activities such as input support, training and infrastructure development. To achieve this the project will implement targeted interventions to strengthen and promote viable climate-resilient value chains relating to smallholder agriculture, specifically targeting value chains that are gender sensitive and provide viable economic opportunities for women. This includes three interrelated components, as presented in Figure 2.

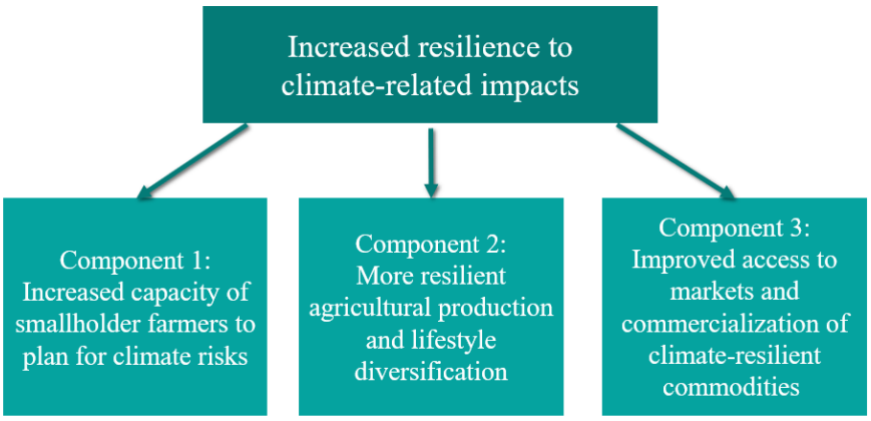


Figure 2 Components of the project Strengthening climate resilience of agricultural livelihoods in Agro-Ecological Regions I and II

While the first component aims at increasing the quality of weather/climate-based information and the dissemination thereof, the second component is to a large extent directed at irrigation and input support, mostly from the Farmer Input Support Program (FISP). The third component on markets and commercialization will help to drive the production of resilient agricultural commodities and to ensure the sustainability of the first two components.

A total of 157,000 farming households in 220 camps (camps are the smallest administrative unit in Zambia, similar to villages or communities) in 16 districts will eventually benefit from the interventions. These 16 districts are in five provinces spread across Agro-Ecological Regions I and II (namely, Eastern, Lusaka, Muchinga, Southern and Western provinces).[71] They were selected given their specific vulnerability to climate change risks – primarily increasing droughts, variability of rainfall and occasional floods, coupled with a high incidence of poverty. Target beneficiaries currently have little resilience to cope with climate impacts or sustain livelihoods in the face of climate change. The project officially started in October 2018 and is planned to run for seven years. The inception workshop will take place at the beginning of 2019. The main executing entity for the proposed project is the Zambian Ministry of Agriculture (MoA). The MoA will partner with a range of organizations and government bodies including the Zambian Water Resources Management Authority (WARMA), the Zambia Meteorological Department (ZMD), the Food and Agriculture Organization (FAO) and the World Food Programme (WFP). Further support for the project and quality assurance is undertaken by the United Nations Development Programme (UNDP), which is also the accredited entity (AE) of the GCF.

II. ZAMBIA FIELD MISSION

A. GENERAL REMARKS AND MISSION AGENDA

An evaluation team, consisting of two consultants from the Center for Evaluation and Development (C4ED) (henceforth referred to as the LORTA team), was formed to lead the field mission from 5 to 9 November 2018. The task of the team was to engage closely with key stakeholders of the project – namely the AE, implementing agencies, project staff, the national designated authority (NDA) and potential end beneficiaries – to ensure their interest in and understanding and sense of ownership of the planned theory-based impact evaluation.

[71] The districts are Mambwe, Nyimba, Chongwe, Luangwa, Chirundu, Rufunsa, Chama, Mafinga, Kazungula, Siavonga, Gwembe, Namwala, Shangombo, Senanga, Sesheke and Mulobezi.

The LORTA team experienced a high degree of collaboration and support from all involved institutions, especially UNDP, who put great effort into organizing a productive week's work. The agenda – shown in appendix I of this design report – was developed to facilitate the joint attendance of all key stakeholders to the LORTA workshop, coordinate with the National Planning Meeting for this project – which took place at the same time as the LORTA mission – and allow for field visits.

As part of the field mission, the LORTA team held meetings and a capacity-building workshop with the key stakeholders. Since a full project team is not in place yet – only one person has been hired so far – the communication and coordination mostly took place via UNDP. The field mission had to be coordinated with the National Planning Meeting, which had the purpose of consolidating all regional workplans for the project into one first-year national workplan. This meeting took place from 6 to 9 November. Although the meeting made coordination and scheduling of appointments more challenging, it also ensured the presence of all key stakeholders involved in the project.

The capacity-building workshop on impact evaluation, which aimed at emphasizing the benefit of theory-based counterfactual approaches and real-time learning and measurement, was successfully held with all 64 participants of the National Planning Meeting attending. While the capacity-building workshop was condensed into one day, it contained all the ingredients to inform the key stakeholders on the LORTA objectives, present key impact evaluation concepts, develop a detailed theory of change and discuss programme implementation details with the implementing partners. The workshop was successful in that it benefited the key stakeholders and the LORTA team. On one hand, the presentations and interactive discussions on theory of change and implementation brought all the key stakeholders together on the same page with respect to the evaluation needs and possible implementation strategies. On the other hand, the LORTA team benefited from gathering in such a short time a rich set of crucial information to collect ideas for the design of the impact evaluation. They also benefited from gaining understanding and collaboration from the key implementing agencies in adjusting their plans to accommodate the evaluation design.

Furthermore, individual meetings with representatives from MoA, WFP and other key stakeholders could be arranged during the mission, including during the National Planning Meeting. The meetings not only informed the LORTA team about the project but also aimed at fostering collaboration and trust between the team and the on-site parties involved.

In collaboration with UNDP and under the guidance of the LORTA team, the theory of change, possible evaluation questions and possibilities for randomising elements of the project were discussed with the participants during the workshop. The results of these discussions were then used by the LORTA team to define the final evaluation questions and develop an impact evaluation design. The team – building on previous work done by UNDP and other key stakeholders – conducted context analyses, examined the existence of appropriate counterfactuals (i.e. comparable treatment and control groups), assessed the availability of baseline administrative and secondary data sources, and acquired budget information.

In addition to the workshop and the individual meetings, two field visits were conducted to districts where similar programmes had previously been implemented, because the current project has not yet started. These visits helped substantially in gaining a clearer understanding of the project activities and in further sharpening the ideas for impact evaluation designs. Appendix II of this design report lists all the people engaged with at the workshop, during individual meetings and on field visits throughout the LORTA mission.

B. IMPLEMENTATION PLAN

The project targets over 157,000 farmers and their families in 220 camps in 16 districts in Agro-Ecological Regions I and II. As mentioned previously, the project is still in its early stages, and the

overall workplan is not yet available. It will likely only be finalized during the project's inception workshop at the beginning of 2019.

For component 1, the installation of manual rain gauges in every camp is planned, as well as the installation of automated weather stations in selected sites within the targeted districts. These stations are installed to increase the density of weather and climate data. To further increase the availability and usability of data, hard-copy historical data will be digitized. Furthermore, the capacity of ZMD staff on generation, analysis and modelling of climate information will be strengthened through training and university cooperation. To disseminate the improved weather- and climate-related information, various channels will be used such as radio, television, field extension services and print media; however, the main channel will be SMS text messages. To increase the capacity of farmers to use the improved and disseminated information, trainings at farmer field schools will be held in every camp.

Component 2 includes the introduction of water storage and irrigation equipment; the construction of boreholes, weirs and irrigation canals; and the training of smallholder farmers and district officers in implementing and maintaining irrigation infrastructure. Another main part of component 2 is the distribution of seeds, soil kits and tools, which is part of the FISP. Seeds will be distributed through cooperatives in one of two ways: either a farmer gets a bag of seeds directly (direct input support), or the farmer receives an e-voucher, which he or she can use to purchase the seed or any other agricultural input of his or her choice. An important choice farmers have to make is between open-pollinated seed varieties, which have the advantage of only having to be bought about once every three years, and hybrid seed types, which are more climate resilient and have a higher yield but need to be bought every year.

Another important part of component 2 will introduce and strengthen farmer field schools, which are then used for trainings on sustainable agricultural practices and improved seeds (i.e. aiming at higher yields and more resistance to droughts). Farmer field schools have between 20 and 30 members and are managed by extension workers. The training participants of farmer field schools will be the same for components 1 and 2. These schools will also receive seeds but in very small quantities and only for training purposes. In addition to the farmer field schools, a learning centre of excellence will be established in each district to scale up and disseminate good practices on climate-resilient agriculture. Here, lead farmers are the ones to receive the first-hand training, and follower farmers then learn about the training content from lead farmers. Lastly, every camp will receive training and input to adopt alternative livelihoods, which includes the distribution of 1,520 bee hives, 14,000 goats – which will add up to 50,000 beneficiary households through passing on of the animals – and 68,000 fish ponds.

Component 3 mostly works through cooperatives. Across the 16 districts, 71 multipurpose processing centres will be established and training on processing will be provided by cooperatives. Storage and transport facilities will be strengthened – for example, by making toyo cycles (tractor-cycles with 0.5-ton cargo capacity) available on a loan basis. Furthermore, a marketing platform will be established, and access to finance and insurance products will be strengthened.

C. THEORY OF CHANGE

The theory of change was prepared in intensive group discussions with representatives of all government bodies, UNDP, WFP and FAO. We developed separate theories of change for each of the three components. The results of this discussion are presented below.

Component 1: Increased capacity of smallholder farmers to plan for climate risks

i) Input

Inputs to the project will be the funds provided by the government, GCF and other donors, for the equipment and the workforce to deliver the intervention.

ii) Activities

The activities of component 1 include the installation of automated and manual weather stations, training of ZMD staff to strengthen their capacity for generation and modelling of climate information, and updates of weather and climate models and testing them against reality. Another key activity is the dissemination and use of tailored weather/climate-based forecasts, for which smallholder farmers (SHFs) will be trained. Agreements will be established with cell phone companies to send out messages to farmers.

iii) Outputs

These activities result in precise, timely and dense weather forecasts, which are available to SHFs (mostly via SMS, but also other channels such as radio). Furthermore, SHFs are trained and receive SMS with climate/weather information on a regular basis.

iv) Outcomes

As a short-term outcome, SHFs know about the services of weather forecasting and accept these services (e.g. they read the SMS). The long-term outcome is that SHFs use these services and are able to plan for and manage climate risk.

v) Goals

This will result in an increased climate resilience of SHFs.

Component 2: More resilient agricultural production and lifestyle diversification

i) Input

Inputs to the project will be the funds provided by the government, the GCF and other donors for the materials and the workforce to deliver the intervention.

ii) Activities

The first subcomponent of component 2 includes the purchase of water storage and irrigation equipment; the construction of boreholes, weirs and irrigation canals; and the training of SHFs and district officers in implementing and maintaining irrigation infrastructure. Further subcomponents include the provision of improved seeds, tools and soil kits; the training of existing cooperatives on sustainable agricultural practices; and the organization of farmer meetings to encourage the establishment of cooperatives. In addition, farmer field schools will be introduced, through which all training modules of components 1 and 2 are carried out. Also, inputs and training on alternative livelihoods (e.g. goat-keeping or beekeeping) will be provided.

iii) Outputs

Ponds, weirs, boreholes and irrigation canals are built, SHFs are trained to use the water system, and lead farmers and district officers are trained on how to manage catchment areas of water and irrigation systems. Furthermore, seeds are bought and distributed to cooperatives, and SHFs are trained on the use of sustainable agricultural practices and use of improved seeds. In addition, farmer field schools are set up in every camp, and every district will have a learning centre of excellence.

iv) Outcomes

As an outcome, SHFs will have improved access to irrigation and will adopt irrigation farming and fish-farming. Furthermore, cooperatives will distribute the seeds to SHFs once a year, who plant the improved seeds. Cooperatives improve their practices and new cooperatives are established. A long-run outcome is an increase of SHF's yields in case of droughts. In addition, the lead farmers adopt

sustainable agricultural practices, then train followers, who in turn adopt these practices as well. Also, alternative livelihoods are adopted.

v) Goals

This will result in an increased climate resilience and a more stable income for SHFs.

Component 3: Improved access to markets & commercialization of climate-resilient commodities

i) Input

Inputs to the project will be the funds provided by the government, the GCF and other donors for the materials and the workforce to deliver the intervention.

ii) Activities

The activities include the establishment of multipurpose processing centres (MPCs), training of SHFs on processing, purchase of cycles and storage facilities, and training of SHFs on storage and business skills.

iii) Outputs

MPCs are established across districts, SHFs have received training, toyo cycles and facilities are bought, warehouses are built and a market information platform is established.

iv) Outcomes

As an outcome, SHFs know about processing and offered services (such as the "Dial-A-Load" project, which provides transporters with a supply-and-demand information platform), they process their products and use these services. This will lead to an increased adoption of alternative livelihoods and sustainable agricultural practices due to the facilitation of market access.

v) Goals

This will result in more stable income for SHFs.

During the workshop, we also discussed the activities of all components in groups, including the participants, regarding the formulation of possible evaluation questions and the feasibility of impact evaluation design ideas. The inputs and results from these discussions were then used by the LORTA team to develop both the evaluation questions, which are listed below, and the impact evaluation design, which will be described in more detail in section D.

Evaluation questions

- Does the project overall lead to adoption of climate-resilient practices, reduced vulnerability, improved yields and higher income?
- Which messages are most effective in inducing climate and weather-based decisions by SHFs? How can SHFs be nudged through messages to plant climate-resilient seeds / adopt sustainable agricultural practices?
- To what extent do farmers benefit from input support? Does input support lead to more climate-resilient practices?

Impact indicators

- Knowledge level of climate information / climate-resilient agriculture
- Number of farmers who declare adopting sustainable as well as climate-resilient agricultural practices
- Indicators of sustainable agricultural practices such as average declared tillage depth
- Number of farmers who declare using climate information in their seasonal plans

- Number of farmers who diversified their livelihoods (e.g. who diversify from maize and who have started new activities other than subsistence agriculture)
- Purchase decisions on agricultural inputs, seeds in particular
- Average level and volatility of crop yields
- Share of the agricultural production not for household consumption
- Household revenues / income
- Income volatility per year
- Food security index based on food consumption, food expenditure shares and the number of strategies to cope for a lack of food

D. PLANS FOR EVALUATION

Methodology

A comprehensive mixed-methods approach will be adopted, where the results from the analysis of primary and secondary quantitative data will be triangulated using information from qualitative interviews and focus group discussions. The complementary qualitative analysis will provide a better insight into those outcomes that cannot be captured by quantitative data alone and will report issues and experiences from different perspectives.

The impact evaluation's quantitative analysis will entail using primary data from surveys, analysis of available secondary data (such as satellite data; administrative data from the farmer registration module of the Zambia Integrated Agriculture Management Information System (ZIAMIS); the Living Conditions Measurement Survey; and the Rural Agricultural Livelihood Survey) and regular briefings on project monitoring data. An aim of the LORTA programme is to regularly update the stakeholders about the progress of and results from the project implementation. Therefore, importantly, we plan high-frequency phone calls to stakeholders between baseline and endline as a means for both monitoring and a real-time impact evaluation of the overall project impact.

Impact evaluation design

This section will present our proposed impact evaluation (IE) strategies. We will focus on the evaluation of the overall project impact through a difference-in-differences (DiD) design. This evaluation will be complemented by randomized control trials (RCTs) within the project on the effects of farmer input support and on how to deliver effective messages to farmers. While the first two strategies were discussed during the LORTA mission, the evaluation questions on effective message delivery were developed during later discussions with the GCF IEU and Dr. Babatunde Abidoye of UNDP. The proposed IE strategies have been shared with UNDP, without any objections received. We believe that the budget permits this comprehensive IE approach, which makes it possible to not only evaluate the overall project impact and its potential drivers but also investigate questions of wide relevance beyond the particular case of this project.

We propose to focus on the overall project instead of only one component for three reasons. First, it is so far unclear who will benefit when from which project activities. Once beneficiary selection is clearer, possibly in the lead-up to the inception workshop, a randomized phase-in design for components or subcomponents is a further option, which can still be built into the evaluation of the overall project impact. Second, a crucial part of the evaluation of the overall project impact is the high-frequency data (HFD) obtained through phone calls (more information in the sampling section). The rich panel-data set these create will enable us to study the drivers behind any project impact, which would otherwise remain a black box. For instance, the spatial and temporal variation in project roll-out can be used to shed light on the relative importance of different components for

the overall effect. Furthermore, the collection of HFD is planned to start before project implementation, which will enable further plausibility tests of the parallel trends assumptions underlying causal identification through a DiD design. Third, learning about the overall project impact and its drivers is important for the Government of the Republic of Zambia (GRZ), which contributes the largest share of the budget.

The project and the availability of a large number of cell phone numbers of farmers make it possible to test rigorously, with relatively cheap interventions, how to tailor messages effectively. The goal of these behavioural interventions through several treatment arms is to nudge farmers towards adopting climate-resilient crops. Learning about this goal is important in itself, but the potential implications are more far-reaching. By testing intervention components based on the behavioural sciences, lessons can be applied to many different settings.[72]

The largest part of the GRZ's co-financing goes into the FISP. FISPs are used in many countries (e.g. Malawi, Kenya, Tanzania, Zambia, Mali, Nigeria, Ghana, Senegal and Ethiopia[73]). There are several studies aimed at estimating the effects of a FISP on individual farmers,[74] the economy as a whole[75] and whether targeting of farmers could improve efficiency.[76] However, an evaluation based on experimental methods has not been conducted yet. The buy-in of the MoA into randomized designs opens a unique opportunity to do just that and evaluate to what extent farmers causally benefit from input support and whether input support leads to more or less climate-resilient practices.

Design of overall project impact evaluation

It is currently not foreseen by the project partners that budgetary or logistical constraints necessitate a phased-in roll-out of the entire project. This rules out a randomized phase-in design. Since programme areas (the camps within the selected districted) are already pre-selected, a randomized assignment of the programme is not possible either. This leaves DiD, possibly combined with matching, as the most feasible option. The decision to be taken for the IE design is therefore only whom to select for the IE study in order to estimate the project impact. We are referring to this decision when we speak, somewhat loosely, of selecting treatment and control farmers, not about assignment to the project. In this section we will describe the basic idea of our selection strategy, and in the following section we will go into the details of the sampling strategy.

Since selection of individual beneficiaries is not entirely clear yet and the project's theory of change also includes spillovers of information and training knowledge within the programme areas, the proposed IE design focuses on the average impact for a farmer in the project areas. Since we focus on the overall project outcomes, all indicators listed above are relevant.

DiD relies on the parallel trends assumption for causal inference of the project's impact. Since the beneficiaries are farmers, an important time-varying factor is weather. As the likelihood of different weather shocks grows with geographical distance, the control farmers need to be chosen in

---

[72] For information technology in agriculture see, for instance, Cole, S., & Sharma, G. (2017). The promise and challenges in implementing ICT for agriculture. Precision Agriculture for Development. Retrieved from http://precisionag.org/uploads/cole-sharma-july1-2017.pdf

[73] Kanyamuka, J. S., Jumbe, C. B., & Ricker-Gilbert, J. (2018). Making agricultural input subsidies more effective and profitable in Africa: the role of complementary interventions. In A. Obayelu (Ed.), *Food Systems Sustainability and Environmental Policies in Modern Economies* (pp. 172–187). Hershey, PA: IGI Global. doi: 10.4018/978-1-5225-3631-4.

[74] E.g. Chibwana, C., Fisher, M., & Shively, G. (2012). Cropland allocation effects of agricultural input subsidies in Malawi. *World Development*, *40*(1), 124–133.

[75] Arndt, C., Pauw, K., & Thurlow, J. (2015). The economy-wide impacts and risks of Malawi's farm input subsidy program. *American Journal of Agricultural Economics*, *98*(3), 962–980.

[76] Asfaw, S., Cattaneo, A., Pallante, G., & Palma, A. (2017). Improving the efficiency targeting of Malawi's farm input subsidy programme: Big pain, small gain? *Food Policy*, *73*, 104–118.

proximity to the treatment farmers. Farmers' livelihoods and vulnerability to climate change determine how farmers react to weather events. Therefore, the parallel trends assumption is the more plausible the more farmers are similar in their livelihoods and vulnerability to climate change.

We therefore propose to choose the sample of treatment and control farmers from pairs of districts that were recently separated due to an administrative reform and in which one newly founded district is part of the treatment and the other one is not. The separation of these districts was enacted by the President of Zambia, then of the Patriotic Front, with the stated goal of effective and efficient public service delivery through decentralization.[77] The separation happened about three years before the next scheduled presidential election, which makes it unlikely that the true political motivation was to cater to the voter bases of the Patriotic Front. Systematic differences between the newly separated districts are therefore less likely.

According to the vulnerability assessment by the Zambia Disaster Management and Mitigation Unit, the newly created pairs of districts lie in the same livelihood zones and are bordering each other. This is illustrated in the map below (Figure 3), where the districts and the livelihood zones are displayed. The districts of interest are Shangombo, Senanga and Sesheka, which are part of the Western Province and were all divided into two new districts in 2012.[78] However, since the district map is from 2010, the separation is not yet shown, but what is visible is that each district falls into one livelihood zone – that is, Shangombo and Sesheke fall into livelihood zone 2 (South-western cereal, livestock, and timber) and Senanga into livelihood zone 1 (Zambezi Plain rice, livestock and fishing).

[77] Zambia: President Sata creates 7 more districts. (6 December 2012). LusakaTimes.com. Retrieved from https://www.lusakatimes.com/2012/12/06/president-sata-creates-7-districts/

[78] Shangombo divided into Shamgonbo and Sioma; Senanga divided into Senanga and Nalolo; Sesheke divided into Sesheke and Mwandi.

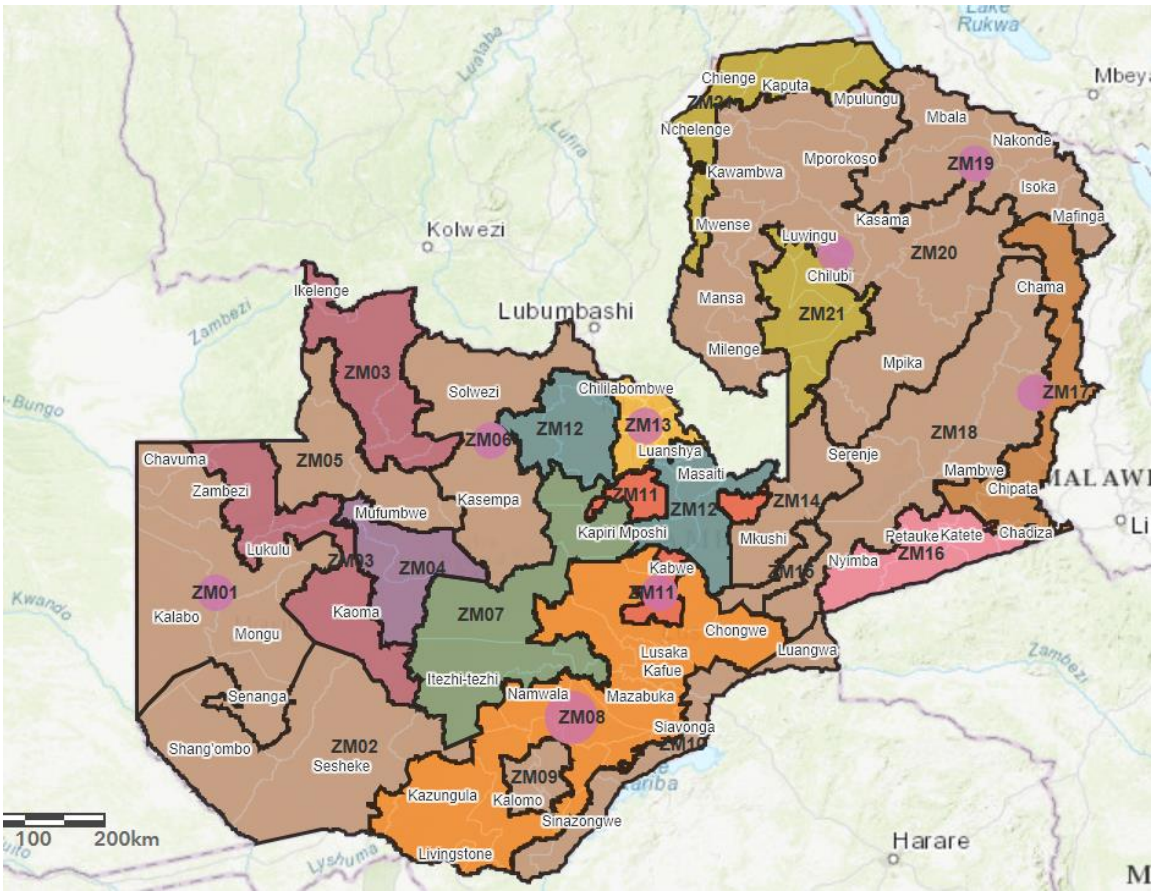


Figure 3 Map of districts and livelihood zones, Zambia (2010)

The parallel trends assumption cannot be directly tested during the intervention period, but its plausibility can nevertheless be assessed through placebo tests. Before deciding on the selection of control farmers, we propose to test for parallel trends in historical data in the years before project implementation. The following data sources are available to do so: satellite data for yields; weather data; and data from ZIAMIS and the Rural Agricultural Livelihood Survey for farmer characteristics. These tests should be carried out at the smallest level possible – that is, at the farmer level for farmer characteristics and possibly yields, and on the smallest geographical unit available for weather data (this still has to be discussed with ZMD). If the placebo tests do not reject that treatment and control districts moved in parallel pre-intervention, treatment and control farmers can be sampled randomly from the respective districts.

In case the placebo tests fail, the DiD approach can be combined with matching. More specifically, the sample of treatment and control farmers can be chosen such that they are matched in terms of time trends in available variables that are likely to influence project outcomes, directly through weather or indirectly through farmers' capacity to cope with climate change and adverse weather events. Examples of these variables are weather and climate trends, agricultural practices and cultivation decisions, household income and farm size. The most suitable matching procedures are coarsened-exact matching and entropy matching, as they are designed to improve balance in covariates.[79] The most important covariates for our purposes, given the DiD design, are the time

[79] King, G. & Nielsen, R. (2015). Why propensity scores should not be used for matching. Retrieved from http://gking.harvard.edu/publications/why-Propensity-Scores-Should-Not-Be-Used-Formatching

trends in the mentioned variables. Whether this exercise is indeed necessary will only be clear once the placebo tests have been run.

Once control groups are chosen and after collection of baseline data but before project activities are implemented in the treatment areas, the HFD can be used for further assessment of parallel trends. Once treatment areas receive the project, the HFD enable us to test for commonality in shocks (e.g. weather). If major differences in shocks are observed, the HFD may be used to control for these differences through several possible methods: by restricting the sample to farmers with common shocks, by explicitly matching farmers in the control and treatment groups based on the realization of shocks, or simply by including presence and severity of shocks as control variables into the regression framework. Whether the latter approach is necessary and feasible will only be clear once shocks have been observed and it can be assessed whether there is sufficient overlap (in terms of sample size for adequate statistical power) between treatment and control for different types of shocks.

To summarize, the proposed IE design based on HFD and with the availability of some historical data enables us to go far beyond the standard DiD design of an overall project impact. It enables us to get much closer to the ideal of an RCT in terms of internal validity, while at the same time permitting us to monitor and look into the black box of why and how the project impacts farmers' lives.

Of course, many things in life come with trade-offs. In this case, the restriction to sampling from certain control districts or even to a matched sample and the HFD collection through phone calls compromises the representativeness of the results compared to if the sample of treatment farmers was sampled randomly from all smallholder farmers. The same data sets that are used to select the control and treatment farmers can, however, be used to assess to what extent these farmers are different from the average farmer in the project areas and also from farmers outside them.

Impact evaluation design for the effectiveness of messages

An important part of the project is informing farmers. In component 1, farmers receive weather and climate information as well as training on how to respond to this information. In component 2, farmers learn and are encouraged to adopt sustainable agricultural and climate-resilient practices. Since it is unclear how to best convey these messages, the IE design proposes to run two RCTs, each with several treatment arms, on how to design messages such that farmers a) respond effectively to climate information and b) buy and plant climate-resilient seeds. The messages should be sent for at least two consecutive seasons in order to test the sustainability of effects.

As one can think of many behavioural components to be included in the messages, we propose to first run a behavioural diagnosis survey[80] via phone with a sample of 500 farmers before designing the different treatment arms. We propose a large sample of 500 farmers in order to have sufficient heterogeneity in farmer characteristics. To give a first idea how this can be applied to the project, think of the case of the still low adoption rate of climate-resilient seeds. According to information from Dr. Abidoye, only 58 per cent of farmers in Zambia (and only 35 per cent of the smallholder farmers who receive FISP) use the more drought-resilient hybrid seeds, despite the fact that these are producing higher yields. The main behavioural bottleneck hindering higher adoption is unclear. Possibly, farmers lack the knowledge that they have to purchase new hybrid seeds every season in order to achieve high yields. This would call for a well-designed information message. Perhaps, however, many farmers are aware of the benefits but are simply lacking access or are struggling to

[80] For a description of the behavioural diagnosis approach, see Datta, S., & Mullainathan, S. (2014). Behavioural design: A new approach to development policy. *Review of Income and Wealth*, *60*(1), 7–35.

overcome the obstacles of purchasing new hybrid seeds. The latter bottleneck would instead call for an intervention that helps farmers act on their intentions – for example, through plan-making or self-commitment.[81]

It is equally important to learn what might stop farmers from appropriately reacting to weather information. There are many potential bottlenecks: lack of understanding of the weather information or how to react to it, lack of a clear plan with concrete steps of action, lack of trust in the accuracy of the information, and the like. Once component 1 is ready to send out tailored weather information, a diagnostic survey should therefore be run to diagnose potential bottlenecks. This survey should include some face-to-face interviews and possibly focus group discussions in order to discuss challenges to understanding and adoption.

Once the content of messages is decided based on the insights of the behavioural surveys, one RCT on weather information and one RCT on nudging farmers to buy and plant climate-resilient seeds could be run, each with several treatment arms. In order to look at long-term effects, tailored messages should be sent out for two consecutive years – and even one year longer for the RCT on nudging farmers – with data collection running in parallel. We propose to collect data on the outcomes through phone calls.

Impact evaluation design for farmer input support

In some areas, there are more SHFs eligible for receiving the FISP than funds and local capacity available to provide them. Furthermore, there have been reports about mis-targeting. According to the Minister of Agriculture and Livestock at that time, besides helping farmers to diversify from maize production, eliminating corruption was a reason for introducing the e-voucher system.[82] This system should enable the MoA to directly target individual eligible farmers. Through this system, which was eventually rolled out nationwide in 2017 and scaled back again in 2018 due to administrative challenges, farmers receive electronic codes through SMS, either directly or through cooperatives, and can redeem their voucher at agro-dealers. There are, however, still reports about FISP benefits not being phased out as planned, probably related to cases when e-vouchers are not directly given to individual farmers.[83]

Here, a randomized rotation design of FISP benefits among eligible farmers in a pilot district might increase public support to the distribution procedure and at the same time enable a rigorous impact evaluation of the effects of FISP on climate resilience and social and economic outcomes. For example, according to the ZIAMIS database, about 24,800 SHFs are registered in Chongwe, of which only about 17,800 currently receive FISP. We therefore assume that 17,800 is the maximum number of farmers that can receive FISP benefits. According to the proposed IE design, the 17,800 farmers receiving FISP rotate among the 24,800 eligible farmers from one year to the next. We would in the first year randomly select 17,800 farmers to receive FISP (treatment group 1), whereas the remaining 7,000 farmers would serve as a control group (control group 1). In the following year, the 7,000 farmers who did not receive FISP before (control group 1) would now get FISP and become treatment group 2. Among the 17,800 farmers who received FISP in the first year, 10,800

[81] For an example of how to implement these interventions through phone calls and messages in rural parts of Africa as well for administering diagnostic surveys, see Rockenbach, B., Tonke, S., & Weiss, A. R. (2018). *Using behavioral insights to decrease non-payment for public utilities.* University of Cologne Working Paper.
[82] Zambia: Government introduces e-voucher system for farmers. (24 November 2012). LusakaTimes.com. Retrieved from https://www.lusakatimes.com/2012/11/24/government-introduces-evoucher-system-farmers/
[83]Zambia: Remove all civil servants from FISP beneficiaries list – Lusaka PS. (14 May 2015). LusakaTimes.com. Retrieved from https://www.lusakatimes.com/2015/05/14/remove-all-civil-servants-from-fisp-beneficiaries-list-lusaka-ps/; Zambia: Plan by Government to phase out "old" FISP beneficiaries gets support. (14 January 2018). LusakaTimes.com. Retrieved from https://www.lusakatimes.com/2018/01/14/plan-government-phase-old-fisp-beneficiaries-gets-support/

are randomly selected to continue receiving FISP, whereas the remaining 7,000 do not receive FISP. The rotation can be continued in the following years.

This design not only ensures that over time all farmers will have received equal benefits on average, it also makes it possible to study dynamics: Do any effects of FISP rely on continuous support, or are there lasting benefits of a one-time assistance? This is not only interesting for research but also highly policy relevant, since the FISP is not meant to provide long-term support. This randomized rotation design has already been discussed with the leader of the District Agriculture Coordinator's Office (DACO) in Chongwe, who approves of the plan but stressed the importance of receiving buy-in from farmers. In order to study long-term effects, we propose to collect data through phone surveys for three years.

Sampling

Data collection for evaluating the overall project impact

In order to have some time for HFD collection in the treatment (T) and control (C) areas before project implementation starts, data analysis of historical trends between treatment and potential control areas should commence as soon as possible. Based on this analysis, control districts should be chosen. Based on our current knowledge, we propose to focus on three pairs of districts, all of which were recently separated (in 2012): Sioma (T) versus Shangombo (C); Senanga (T) versus Nalolo (C); Sesheke (T) versus Mwandi (C). Further analysis is necessary to assess whether these districts are sufficiently similar to other programme districts in order to draw broader conclusions.

Once the samples of control and treatment farmers are chosen based on the available secondary data, a baseline household survey will be conducted to set the basis for the HFD (verifying phone numbers, availability for frequent calls and providing a basic training into the questions that will be asked over phone) and in order to elicit a larger set of variables than can be asked during necessarily short phone calls. The baseline household survey will also be used to assess whether availability for phone calls is related to outcomes and, crucially, different between treatment and control farmers. Whereas the former would on it is own only reduce the generalizability of the estimated project impact, it would in combination with the latter compromise the internal validity of the DiD design (see below how this issue may then be addressed).

The statistics available to us give reason for optimism that phone availability may not be a severe concern: cell phone ownership is widespread in Zambia. According to the CIA World Fact Book, the subscription rate to cell phones in Zambia is 84 per 100 inhabitants.[84] In a sample in the ZIAMIS database of about 14,000 farmers coming from the district of Rufunsa, about 85 per cent of farmers are registered with a cell phone number. Further verification of cell phone coverage is needed (a network coverage map has been requested from the Zambia Information & Communications Technology Authority (ZICTA)) before embarking on the HFD through phone calls, but the currently available statistics make this plan promising.

Shortly after baseline and before project activities start, phone calls should already start among treatment and control farmers in order to assess parallel trends in variables that are otherwise unavailable (most importantly, knowledge and adoption of climate-resilient practices).

We propose to have around five phone calls per year for the first six years of the project, with more phone calls before and after the start of the rainy season than in the rest of the year. The phone calls are planned to take, on average, five minutes, assuming longer phone calls at the beginning than at the end of the six years. They are short in order to minimize survey fatigue and will only assess

[84] Central Intelligence Agency. (n.d.). Africa: Zambia – World Factbook. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/geos/za.html

essential indicators of programme monitoring and evaluation, such as participation in programme activities, farming decisions (e.g. tilling (method), planting, applying fertilizer, harvesting) and social-economic outcomes (e.g. experience of food insecurity, income). The number of phone calls and their length may have to be changed once quotations for phone calls are available, power calculations have been done and first pilot phone calls have been done.

In the seventh and last year of the project, an endline household survey will be conducted. The baseline and endline household surveys will also be used to measure project impact for those farmers who cannot be reached by phone. We can thereby test for differences in impact between farmers with phone access and a larger, more representative sample (including farmers who cannot be reached) and possibly extrapolate the results from the phone surveys to this larger population.

Data collection for randomized control trials on effective messages

For all RCTs, ranodmization should be stratified by agricultural camps and balanced according to farmer characteristics that are available from secondary data, in particular from the farmer registry (such as income, plot size, main types of agricultural production), in order to improve statistical power. A collection of primary data at baseline is not necessary.

Sampling for RCT on nudging farmers to buy and plant climate-resilient crops: Farmers will be randomly sampled from individual recipients of e-vouchers. The reason for sampling from e-voucher recipients is that farmers' redeeming decisions will automatically be captured in the FISP module of the ZIAMIS database. This provides an easy way to capture a main outcome variable: farmers' purchasing decisions. Furthermore, all farmers who receive individual e-vouchers are registered with a cell phone number. Therefore, follow-up phone surveys are possible. During follow-up phone surveys, farmers' planting decisions, as the most important outcome variable, will be elicited. In order to test for the sustainability of any effects, phone surveys and analysis of the ZIAMIS data should be running for at least two consecutive planting seasons.

Sampling for RCT on effectiveness of weather messages: The sample of farmers for tailored weather messages will be randomly sampled from recipients of weather messages once it is clear who will receive messages. A likely high intra-cluster correlation in weather shocks means that farmers should be sampled from all 220 camps.

Sampling for randomized control trials on effects of farmer input support

Farmers will be sampled from a district with over-subscription (this information has been requested from the MoA) and among farmers with registered phone numbers. Outcomes will be collected via phone interviews. As for the RCTs on effective messages, ranodmization should be stratified by camps and balanced according to the previous purchase decisions and farmer characteristics that are available from secondary data, in order to improve statistical power. It is not necessary to collect primary data at baseline.

Power calculations

Power calculations enable us to determine the minimum sample size needed in order to detect with a high probability a meaningful impact if there is one. Which impact can be considered meaningful is highly context specific. It can be approached from a practical perspective through the question "Below which effect size would the programme be considered a failure?"[85] This discussion still has to take place between the project team (once it is in place) and the implementing partners. We have also not received access to secondary baseline data (most importantly from the farmer registry) on

[85] Or "minimum desired effect", see Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2011). *Impact evaluation in practice*. Washington, D.C.: World Bank Publications.

key outcome variables in order to estimate the variance and intra-cluster correlation (within camps). Since the main evaluation question on the project's overall impact will be done through a DiD design with several observations before and after intervention, we also need to estimate the within-farmer variation in outcomes (over time). As will be highlighted below, power calculations for the proposed designs are complex and go beyond simple between-group comparisons. Therefore, power calculations still have to performed once the discussions on minimum acceptable level of impacts have taken place and baseline data have been received.

Overall project impact

Generally speaking, a DiD design with several pre- and post-intervention data points has higher statistical power than a purely cross-sectional design.[86] This makes it promising that the available budget will allow for a high-powered design.

Since some interventions will likely be rolled out at the camp level (training activities) and outcomes are likely to be correlated more strongly within camps than across camps, it is important to include a sufficient number of camps (as statistical clusters) in the sample. As a rule of thumb, Gertler et al. recommend a minimum level of 30 clusters in both the treatment and control groups.87 This level would be reached with the proposed six treatment and control districts.

Impact of nudging farmers and of different weather-based messages

For both types of interventions, farmers from all 220 camps can be targeted such that the number of clusters would not be a concern. Intra-cluster correlation is less of a concern for the nudge messages than for weather-based messages, since the former targets individual farmers whereas weather-based messages and weather shocks are location specific.

Since both the intervention and the data collection are relatively cheap, a sufficient sample size should be reached within the budget. However, power calculations should still be conducted once baseline data are available in order to not spend more than the required amount of money on evaluating this question. An important element of the power calculations is adjusting alpha-levels because of multiple hypothesis testing when different treatment arms are compared to one another.

Impact of the Farmer Input Support Program

Key outcome variables of the FISP are yields and income, among others. Since these outcome variables generally are likely to be noisy, a large sample size is needed to account for that. Power calculations still have to be performed to determine the minimum sample size.

Possible risks for the impact evaluation design

Beneficiary selection and "power" of the project interventions

The proposed DiD design estimates the average impact for a farmer in the project areas, rather than at the level of direct beneficiaries. Through this design, spillover effects of training and information beyond the direct beneficiaries would be captured, but at the same time the average effect is likely much smaller than the effect on the direct beneficiaries. A more powerful design could be chosen if the selection of direct beneficiaries is clearer (possibly in time for the inception workshop) and is based on available secondary data. In this case, the treatment and control farmers could include a matched sample based on the beneficiary selection criteria, which would require a revision of the

---

[86] Juras, R., Comfort, A., & Bein, E. (2016). *How study design influences statistical power in community-level evaluations.* OAH Tier 1B Evaluation technical assistance brief No. 3. HHS Office of Adolescent Health. Retrieved from https://www.hhs.gov/ash/oah/sites/default/files/ash/oah/oah-initiatives/assets/ta-tppevalbrief-studydesignstatspower.pdf
87 Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2011). *Impact evaluation in practice*. Washington, D.C.: World Bank Publications.

proposed sampling strategy. By including the HFD collection, we can go some way towards the more powerful design even without knowing the exact selection criteria because we can observe which farmers (with which characteristics) become direct beneficiaries at which time. A requirement for this is that the sample of treatment farmers includes enough farmers who eventually become direct beneficiaries.

Cell phone coverage

A key assumption behind the feasibility of the impact evaluation design is that a sufficient number of farmers can be reached regularly by phone. The aggregate statistics for Zambia and the sample of farmers from Rufunsa district make this assumption plausible. A network coverage map requested from ZICTA should provide further information. For an example of HFD collection through cell phones and its challenges, see Hoogeveen et al.[88] However, the proof of the pudding lies in the eating. The diagnostic survey by phone will show real-life availability in the different agricultural camps. In case a high percentage cannot be reached, the HFD approach as well as the phone-based data collection for the RCTs will have to be reconsidered. Unless substantial additional funds become available, abandoning phone-based data collection would necessarily mean substantially reducing the scope of the IE design. This may ultimately lead to a standard DiD with baseline, midline and endline delivered through household surveys.

Support for randomized assignment of the Farmer Input Support Program

While we have found broad buy-in into randomized methods with respect to FISP among technocrats, DACO leaders and the Deputy Director of the MoA, a randomized rotation of FISP benefits will have to be approved by high-level leaders in the MoA and must not experience strong resistance by farmers. If the randomized assignment is limited to one district as a pilot, then the LORTA umbrella would provide a unique opportunity to implement an RCT on the effects of farmer input support.

Availability of tailored weather messages

It is unclear when tailored weather messages will be available during the project. It will likely be several years after the project inception since weather and climate models first need to be built, based on yet-to-be-scanned historical weather information, and tested with actual weather information. This might not leave much time for evaluating the effectiveness of messages. Therefore, if during the first project years progress on tailored weather information is stalling, IE priorities may be better moved elsewhere, given the relatively tight IE budget of USD 200,000.

Qualitative assessments

The qualitative assessment would involve interviews and focus group discussions with the beneficiaries and the implementing organizations. With regard to fostering our understanding of change in farmer practices, the anecdotal narratives and all the textual material collected are essential to identify the social dynamics around which the project and all of its components operationalize.

The baseline and endline interviews will therefore cover themes of uptake, attitudes, perceptions, knowledge and the nature of the project implementation and experiences regarding the project. These will be captured using key informant interviews with the implementing agencies, in-depth interviews with individual household and community members, and, finally, focus group

[88] Hoogeveen, J., Croke, K., Dabalen, A., Demombynes, G., & Giugale, M. (2014). Collecting high frequency panel data in Africa using mobile phone interviews. *Canadian Journal of Development Studies/Revue canadienne d'études du développement*, *35*(1), 186–207.

discussions with farmers. These qualitative data would be particularly useful in identifying the assumptions that lie behind the results chain of each component's ToC.

Timeline of evaluation

The LORTA team is coordinating with UNDP, MoA and other stakeholders involved in the project to ensure an effective implementation of the project and evaluation activities, to be in line with the IE design. Since the workplan of the project is not yet finalized, the dates for the activities in the timeline (Table 1) are only tentative.

Table 1 Draft timeline of evaluation

Q1 2019	Project inception workshop & finalization of workplan, eligibility criteria, M&E framework and impact evaluation design
Q1 2019	Project inception workshop & finalization of workplan, eligibility criteria, M&E framework and impact evaluation design
Q1 2019	C4ED will explore the existing secondary data for data analysis of historical trends between treatment and potential control areas
Q2 2019	Diagnostic survey through phone with a sample of farmers
Q2 2019	Randomization of control and treatment households for FISP and "nudging" messages
Q2–3 2019	Baseline household survey
Q3 2019–Q4 2024	HFD collection (around five phone surveys per year)
Q4 2019	Project implementation starts
Q1 2020–Q4 2021	Analysis of ZIAMIS data and follow-up phone surveys (for FISP)
Q2 2025	Project implementation ends
Q3 2025	Endline data collection

Budget

For the first part of the IE, it will be important to understand the quality of the existing data and the final design of the programme, which will be finalized during the first months of 2019. During this time, C4ED and the IEU will need to work closely with the government and UNDP. The second part of the IE will be its implementation. The cost of the IE will be determined by what data we would like to utilize for the study and power calculations for sample size. At our current level of knowledge, we estimate minimum sample sizes to be 1,500 for the baseline and endline survey and 3,000 for the phone survey. As the IE contains three different strategies, a cost estimate for the data collection for each evaluation is given separately in the tables below. Two cost estimations are provided (estimations for data collection only, *excluding* costs for training/pilot/pre-test and materials): one with the minimum previously mentioned sample size and only messages for adoption of climate-resilient seeds (minimum cost,

Table 2), and one with a larger sample size and messages for adoption of climate-resilient seeds plus messages for reacting to weather information (advanced cost, Table 3). The current budget for IE amounts to USD 200,000. Depending on the final scope of the evaluation and the necessary adjustments to the evaluation cost once quotations are made, this amount may or may not be sufficient to conduct a high-quality IE.

In addition to these data-collection costs and C4ED salary costs, we expect (at least one member of) the LORTA team to travel to Zambia twice in 2019 (once for the final refinement of the IE design and once for the baseline data collection and the set up of the phone surveys).

Table 2 Cost of the different impact evaluation designs – minimum cost

ITEM	OBSERVATIONS	REPEATED	COST EACH (USD)	TOTAL COST (USD)
Overall impact				
Phone survey (5 min., 6 years)	1,260*	30	$1.25	$47,250
Household survey (baseline and endline)	1,500	2	$35.00	$105,000
Key informant interviews	6	1	$600.00	3,600
Focus group discussions	6	1	$50.00	300
				$156,150
Effective messages to farmers				
Diagnostic phone survey (20 min.)	500	1	$5.00	$2,500
Sending out seed messages (2 years)	3,000	2	$0.10	$600
Phone survey seed messages (20 min., 3 years)	500	3	$5.00	$7,500
Key informant interviews	16	1	$600.00	9,600
Focus group discussions	16	1	$50.00	800
				$21,000
Farmer input support				
Phone survey (20 min., 3 years)	1,500	3	$5.00	$22,500
				$22,500
Total sum				**$199,650**

Note: *Assumption that 84 per cent of respondents from baseline survey can be reached via phone

Table 3 Cost of the different impact evaluation designs – advanced cost

ITEM	NUMBER	REPEATED	COST EACH (USD)	TOTAL COST (USD)
Overall impact				
Phone survey (5 min., 6 years)	2,520*	30	$1.25	$94,500
Household survey (baseline and endline)	3,000	2	$35.00	$210,000
Key informant interviews	6	1	$600.00	$3,600
Focus group discussions	6	1	$50.00	$300
				$308,400
Effective messages to farmers				
Diagnostic phone survey (20 min.)	500	1	$5.00	$2,500
Sending out seed messages (2 years)	3,000	2	$0.10	$600
Phone survey seed messages (20 min., 3 years)	500	3	$5.00	$7,500
Phone survey weather messages** (20 min. 3 years)	500	2	$5.00	$5,000
Key informant interviews	16	1	$600.00	$9,600
Focus group discussions	16	1	$50.00	$800
				$26,000
Farmer input support				
Phone survey (20 min., 3 years)	3,000	3	$5.00	$45,000
				$45,000
Total Sum				**$379,400**

Notes: * Based on the assumption that 84 per cent of respondents from baseline survey can be reached via phone
** Cost for weather messages is not budgeted based on the assumption that these are already part of the project budget.

Assessment of institutional capacity for impact evaluation

The LORTA team identified IE within Zambia project team at this time. Hiring for the project management unit has only just started. No trained economist or methods expert (of IE tools) has yet been hired. It is planned to hire an M&E expert, but her or his level of expertise cannot yet be assessed. However, given the fact that this role means overseeing the monitoring of a large project with a budget of USD 130 million, it is difficult to see that this person would have the additional capacity to design and implement a complex IE study.

Nevertheless, there was a pronounced understanding of the need for IE as a result of the LORTA workshop in Thailand in July and a very good understanding how IE differs from monitoring processes and the concepts of the theory of change. Furthermore, previous trainings by Dr. Abidoye clearly showed that some workshop participants already had knowledge of key IE terms. The LORTA workshop was able to build on this foundation by deepening their knowledge and training a large number of people who have not previously been trained on IE.

We believe that the LORTA workshop within the mission substantially benefited the stakeholders involved in bringing the needed formative engagement. The presence and support of Dr. Abidoye in the LORTA mission and overall programme has been crucial in facilitating coordination and understanding among the LORTA team and the project implementers and stakeholders.

Overall, we consider the project to be in need of support for IE. The LORTA team, in coordination with Dr. Abidoye, is continuing to provide input and feedback remotely.

III. WAY FORWARD

Overall, we consider that the LORTA mission in Zambia was well received and that it produced promising results.

The success of the LORTA mission has been particularly achieved thanks to the attentive collaboration of the UNDP Zambia, especially Eric Chipeta, and the senior economist and consultant of the UNDP HQ, Dr. Babatunde Abidoye. Their input was crucial in arranging the meetings and making sure all key informants and decision-making representatives would be consulted. Furthermore, Dr. Abidoye was crucial in supporting the operational and implementation discussions with the stakeholders.

However, the future success of this project in the LORTA framework is highly conditional upon a continuous responsive collaboration from UNDP and other key stakeholders involved.

APPENDIX I

Table 4 Agenda of Zambia LORTA field mission

DAY	PROPOSED ACTIVITIES	PARTICIPATION
Day 0: Sunday, November 4	Arrival of team	LORTA team
Day 1: Monday, November 5 09:00–10:00	Introductions, overview and discussion of objectives	LORTA team, UNDP
10:00–17:00	Presentation and discussions of the project	LORTA team, UNDP, FAO, WFP, ZMD, WARMA
Day 2: Tuesday, November 6 8:00–10:00	Field Trip to Chongwe Meeting and discussion with the district representatives of the MoA	LORTA team, UNDP, MoA,
10:00–13:00	Meetings with beneficiaries from previous project, visiting of project sites (beehives, goat-keeping, multi-processing units, irrigation plants)	LORTA team, UNDP, MoA, farmers
17:00–19:00	Discussion on impact evaluation design ideas and possibilities	LORTA team, UNDP, WFP, ZMD
Day 3: Wednesday, November 7	Impact Evaluation Design Workshop Introduction to Impact Evaluation Impact Evaluation Methods Theory of Change Evaluation Designs	LORTA team, UNDP, MoA (including district officers), FAO, WFP, ZMD, WARMA (64 participants)
Day 4: Thursday, November 8	Field trip to Chirundu Meeting with the district representatives of the MoA and former beneficiaries, visiting of former project sites (irrigation plants, crop diversity, dam)	LORTA team, UNDP, MoA, farmers
Day 5: Friday, November 9 10:00–12:00	Meeting with the MoA (Deputy Director)	LORTA team, UNDP, MoA
15:00–19:00	Individual meetings with MoA and WFP to answer questions on evaluation design possibilities and further deepen understanding of the programme	LORTA team, UNDP, MoA, WFP
Day 6: Saturday, November 10	Debriefing and discussion of next steps	LORTA team, UNDP
Day 7: Monday, November 12	Individual meeting with NDA and MoA	LORTA team, UNDP, MoA

APPENDIX II

Table 5 List of stakeholders engaged with during Zambia LORTA mission

LORTA WORKSHOP, LEGACY LODGE LUSAKA (SELECTION)		
NAME	POSITION	INSTITUTION
Eric Chipeta	Programme Analyst – Energy & Environment	UNDP
Babatunde Abidoye	Global Technical Adviser & Senior Economist	UNDP
Arthur Asumani	NAP Technical Officer	UNDP
Essayas Tatek	Livelihood Specialist – Project Management Unit	UNDP
Chongo Simpasa	Programme Associate – Environment	UNDP
Winnie Musonda	Environmental Adviser	UNDP
Simon Banda	Inspector	WARMA
Hartley Muchese	Hydrogeologist	WARMA
Misael Kokwe	Technical Coordinator of the Climate-Smart Agriculture project	FAO
Derrick Ndimbwa	Programme Officer	WFP
Martin K. Swaswa	Assistant Director	ZMD
Lyson Phiri	Agricultural Programme Meteorologist	ZMD
Mutau Mutau Christopher	Engineer	ZMD
Chisakuta M. Stanislaus	Deputy Director	Ministry of Agriculture
Mwamba Malata	Agribusiness and Marketing Department	Ministry of Agriculture
+ c. 50 representatives of Ministry of Agriculture from all target districts (e.g. from District Agriculture Coordinator's Offices)		
INTRODUCTORY MEETINGS, UNDP OFFICES LUSAKA		
NAME	POSITION	INSTITUTION
Mandisa Mashalogu	Country Director	UNDP
Eric Chipeta	Programme Analyst – Energy & Environment	UNDP
Babatunde Abidoye	Global Technical Adviser & Senior Economist	UNDP
Arthur Asumani	NAP Technical Officer	UNDP
Essayas Tatek	Livelihood Specialist – Project Management Unit	UNDP
Winnie Musonda	Environmental Adviser	UNDP
Frank Nyoni	Senior Environmental & Water Quality Officer	WARMA
Martin K. Swaswa	Assistant Director	ZMD
Lyson Phiri	Agricultural Programme Meteorologist	ZMD
Mutau Mutau Christopher	Engineer	ZMD
Derrick Ndimbwa	Programme Officer	WFP

Misael Kokwe	Technical Coordinator of the Climate-Smart Agriculture project	FAO

MEETING WITH MINISTRY OF AGRICULTURE		
NAME	POSITION	INSTITUTION
Chisakuta M. Stanislaus	Deputy Director	Ministry of Agriculture
Eric Chipeta	Programme Analyst – Energy & Environment	UNDP
Reynolds K. Shula	Senior Technical Officer	Ministry of Agriculture
Henry	Farm Management Officer	Ministry of Agriculture
Dominik Myamayungo	Advisory Extension	Ministry of Agriculture

FIELD VISIT, CHONGWE		
NAME	POSITION	INSTITUTION
Eric Chipeta	Programme Analyst – Energy & Environment	UNDP
Babatunde Abidoye	Global Technical Advisor & Senior Economist	UNDP
Chintu Chintu	District Agriculture Coordinator Office – Chongwe	Ministry of Agriculture
Peter Daka	Dept. of Fishery	Ministry of Agriculture

FIELD VISIT, CHIRUNDU		
NAME	POSITION	INSTITUTION
Eric Chipeta	Programme Analyst – Energy & Environment	UNDP
Valentine Kitubi	Senior Agricultural Officer	Ministry of Agriculture
Ireen Matambo	Extension Officer	Ministry of Agriculture
Wiseman Mulenca	Extension Officer	Ministry of Agriculture
Oscar Mutumfu	Extension Officer	Ministry of Agriculture

IMPACT EVALUATION DESIGN REPORT 8: GEORGIA

TABLE OF CONTENTS

TABLES

FIGURES

I. INTRODUCTION

A. THE LORTA PROGRAMME

Evaluating the impact of development projects and programmes has gained importance in recent years. Impact evaluation not only allows for increased transparency by measuring the effects of investments, it also provides the opportunity to design and implement development projects more effectively. To contribute to this development, the Independent Evaluation Unit (IEU) of the Green Climate Fund (GCF) has started the Learning-Oriented Real-Time Impact Assessment (LORTA) programme to be able to keep track of GCF projects in terms of performance and results, and to enhance learning within the GCF.

The LORTA programme aims to

- Embed real-time impact evaluations into funded projects so GCF project task managers can quickly access accurate data on a project's quality of implementation and likelihood of impact, and
- Build capacity within projects to design high-quality data sets for overall impact measurement.

The purpose of the impact evaluations is to measure the change in GCF key result areas that can be attributed to project activities. The LORTA programme will inform on returns on GCF investments and help GCF projects track implementation fidelity. The objectives of LORTA include the following:

- Measuring the overall change (outcome or impact) of GCF-funded projects and enhancing learning
- Understanding and measuring results at different parts of theories of change
- Measuring the overall contribution of the GCF to catalysing a paradigm shift and achieving impacts at scale

Currently, the LORTA programme is in Phase I (formative engagement and design). In the first year (2018), the IEU will support eight GCF projects to build high-quality, theory-based impact evaluation designs at inception. Formative work will include engaging with project teams, accredited entities (AEs) and GCF staff; designing theory-based impact evaluations; and establishing protocols for database development.

The second phase of LORTA will involve the main impact assessment stage (3–5 years), while the third phase will include analyzing baseline and endline data, discussing results, and engaging with diverse stakeholders to share results and incorporate feedback as required.

B. THE GEORGIA PROJECT

Georgia is a young country. After the collapse of the Soviet Union, it experienced civil war and economic chaos but started to recover quickly during the 2000s with a number of reforms that enabled the country to position itself as an emerging market. Despite recent economic growth, poverty levels remain high, especially in rural areas, where 41.4 per cent[89] of the population live. Furthermore, 56 per cent of people are working in the agricultural sector, which is particularly vulnerable to climate change.

[89] Central Intelligence Agency. (n.d.). Middle East: Georgia – World Factbook. Retrieved December 3, 2018, from https://www.cia.gov/library/publications/the-world-factbook/geos/gg.html.

The country is characterized by a complex mountainous topography and climate, which make the population vulnerable to a large set of natural hazards including floods, flash floods, mudflows, landslides, avalanches, drought and strong winds.

According to the disaster records of the National Environmental Agency (NEA), over the 21-year period 1995–2015, 152 lives were lost and total damages from hydro-meteorological and geological hazards were 2.8 billion Georgian Lari, or about USD 1.2 billion. Floods and landslides jointly make up 50 per cent of these damages and 83 per cent of loss of life, as shown in Figure 4 in appendix 1 of this design report. Figure 5 (appendix 1 of this design report) shows the damage per event and per year for each main hazard during the same period. Landslides are the most frequent hazard but are associated with relatively less damage per event compared to the others. Flood and drought happen less frequently but cause relatively more damage per event. To give a visualization of the geographical spread of these hazards across the country, Figure 6, Figure 7 and Figure 8 show the 2016 regional maps for flood, drought and landslide exposure, respectively. The maps for flood and drought depict a clear climatic pattern: the western regions are mostly affected by flood, whereas the eastern regions are more exposed to drought. In general, around 70 per cent of the country's territory is exposed to geological disasters such as landslides and avalanches. The map for landslides shows that areas with the highest number of landslide events overlap substantially with flood-prone areas.

There are currently several barriers preventing Georgia from addressing vulnerability to climate-induced hazards and risks. First, a lack of adequate hydro-meteorological monitoring networks, forecasting models, and human and financial resources prevents comprehensive forecasting and early warning. At the institutional level, there remains a need for both a national protocol clearly defining roles and responsibilities for an early warning system (EWS), and an adequate legal and regulatory framework. Existing climate risk information is not being systematically used to inform national, sectoral and local planning, mainly due to the lack of a comprehensive and definitive national hazard and risk mapping. At the community level, there is no experience with community-based early warning schemes (CBEWS). Community-based organizations are neither informed about risks nor empowered for action. Some community-based climate risk reduction approaches (e.g. agroforestry, restoration of floodplain zones) have been piloted by various international projects, but these have not been scaled up.

To overcome these barriers, the GCF-funded project "Scaling-Up Multi-Hazard Early Warning System and the Use of Climate Information in Georgia" aims at creating a proactive integrated climate risk management approach through the establishment of a countrywide multi-hazard early warning system (MHEWS) and the use of climate information in planning and decision-making. This project is one of the eight projects selected to be part of the inception stage (Phase I) of the LORTA programme. The AE is the United Nations Development Programme (UNDP) Georgia, and its Energy Efficiency unit is responsible for the oversight of the implementation of the project. The project is aligned with the national governmental plans to strengthen the country's EWS and hazard risk management as per the National Strategy on Disaster Risk Reduction (DRR), adopted in 2015. The government provides the highest share of co-financing (cash and in-kind) of USD 38 million, followed by the GCF (USD 27 million) and the Swiss Agency for Development Cooperation (USD 5 million).

The implementation spans a period of seven years (2019–2025) and is led by the Ministry of Environment Protection and Agriculture (MoEPA), in collaboration with a multiplicity of stakeholders: NEA, the Environmental Information and Education Center (EIEC), the Ministry of Regional Development and Infrastructure, and local governments.

The project builds upon a prototype project implemented in the Rioni River basin, where an almost real-time EWS for floods and flash floods was developed. More specifically, the project focuses on

all 11 major river basins in Georgia (for a map, see Figure 9) and consists of three components, each of them comprising several activities:

- Component 1: Expanded climate-induced natural hazard observation network and modelling capacities to secure reliable information on climate-induced hazards, vulnerability and risks
  - Activity 1.1: Expansion of the hydrometric network
  - Activity 1.2: Floodplain zoning based on hazard and risk maps for all major river basins and hazard and risk maps for key climate-induced hazards
  - Activity 1.3: Introduction and implementation of methods and tools for the systematic gender-sensitive socioeconomic vulnerability assessment for decision-making for prioritization of resilience investments
  - Activity 1.4: A centralized multi-hazard disaster risk information and knowledge system
- Component 2: MHEWS and new climate information products supported with effective national regulations, coordination mechanisms and institutional capacities
  - Activity 2.1: Institutional and legal frameworks, public–private partnerships and associated institutional capacity-building for the MHEWS and for the enhanced use of climate information by the public and private sectors
  - Activity 2.2: Development and implementation of the MHEWS covering all river basins
  - Activity 2.3: Enhancing access to and the use of weather and climate information and agrometeorological information services by farmers and agricultural enterprises
  - Activity 2.4: Multi-hazard risk management planning platforms
- Component 3: Improved community resilience through the implementation of the MHEWS and priority risk reduction measures
  - Activity 3.1: Implementation of CBEWS and community-based climate risk management (CBCRM)
  - Activity 3.2: Public awareness and capacity-building programme at all levels to effectively deliver climate risk information and training to communities and local first responders
  - Activity 3.3: Implementation of risk reduction interventions that would significantly reduce the risks against which the MHEWS will operate

The project objective is to reduce exposure of Georgian communities, livelihoods and infrastructure to climate-induced natural hazards through a well-functioning nationwide MHEWS and risk-informed local action. The project proposal submitted to the GCF specifies that Activity 3.1 shall undergo an impact evaluation. This seems timely and relevant because the literature on the impact of CBEWS and CBCRM interventions on the adaptive capacity of communities and households is very scarce. However, the effectiveness of the MHEWS crucially depends on the "last mile" communication and delivery of the warnings to local communities and enhanced community-based risk reduction.

As of time of writing, the project is in the very early stages of implementation. The project team has not yet been recruited and the inception workshop, where project activities will be defined in detail, is planned for mid-February 2019. The impact evaluation design below should thus be regarded as tentative and subject to refinement once more details become available.

II. GEORGIA FIELD MISSION

A. GENERAL REMARKS

An evaluation team, consisting of two consultants from the Center for Evaluation and Development (C4ED), one global UNDP consultant and one IEU staff member (henceforth referred to as the LORTA team), was formed to lead the field mission from 12 to 16 of November 2018. The task of the team was to engage closely with key stakeholders of the project – namely, the national designated authority (NDA), AE, implementing agencies, project staff and potential end beneficiaries – to ensure their interest in and understanding and sense of ownership of the planned theory-based impact evaluation (IE).

During the field mission, the LORTA team held meetings and a capacity-building workshop with the key stakeholders. Meetings, in the form of expert interviews, were used to acquire the maximum possible information about the GCF-funded project. Stakeholders were interviewed regarding their views about the project's implementation and monitoring strategies, expected impact, challenges and possible solutions. The meetings not only informed the LORTA team about the project but also aimed at fostering collaboration and trust between the team and the on-site parties involved. In addition, a capacity-building workshop on IE was held, targeted at the key stakeholders. Besides conveying technical knowledge, the aim of this workshop was to emphasize the benefit of theory-based counterfactual approaches and real-time learning and measurement.

Under the guidance of the LORTA team, an IE design was worked out for the Georgia project. The LORTA team conducted context analyses, examined the existence of appropriate counterfactuals (i.e. comparable treatment and control groups), elaborated a theory of change, assessed the availability of baseline administrative and secondary data sources and acquired budget information.

B. THE MISSION AGENDA

The UNDP Energy Efficiency unit, which is in charge of the GCF-funded project, collaborated with the LORTA team. The agenda (see Table 4 in appendix II of this design report) was designed to ensure the attendance of all key stakeholders at the LORTA IE capacity-building workshop. For the exhaustive list of participants, please consult Table 5. Field visits were also an integral part of the LORTA mission to help the team to better understand the underlying characteristics of the project.

The LORTA workshop was condensed into one day, which proved sufficient to inform the key stakeholders on the LORTA programme objectives, present key IE concepts, develop a theory of change[90] and discuss project implementation details with the implementing partners. Unfortunately, some of the invited participants did not attend. Besides the UNDP personnel, the representatives of the NEA, MoEPA and EIEC were present. Notably, the Emergency Management Agency and local governments were not represented.

The field mission was successful in that it benefited both the key stakeholders and the LORTA team. On one hand, the presentations and interactive discussions on the theory of change and implementation brought all the key stakeholders together on the same page with respect to ownership and understanding of their contribution within the GCF project, the project objectives, the evaluation needs and implementation strategies. On the other hand, the LORTA team gathered information crucial for the IE design and gained understanding and collaboration from key implementing agencies, who committed to adjusting their plans to accommodate the IE design.

[90] Although the theory of change developed during the workshop was very general due to the project being at a preliminary stage, we expect to get more detailed information on the planned activities once they are defined in early 2019.

Since the project proposal for the GCF-funded project already envisioned an IE of Activity 3.1, the discussions during the field mission mostly focused on this activity, although other activities were also discussed. This applies, for example, to the implementation of structural risk reduction measures, such as wire mesh mat linings, gabions, embankments, concrete regulation walls, and riverbed and channel cleaning (Activity 3.3).

Field visits to the sites of the previous prototype project in the Rioni basin were organized. The LORTA team was able to learn about the practical aspects of both structural and non-structural measures, as agroforestry and flood defence structures were part of the itinerary. Meetings with representatives of local municipalities and communities were arranged, during which the details on the implementation of the prototype project were discussed.

The mission concluded with a final debriefing in the presence of the LORTA team, the UNDP team and the representatives from NEA and MoEPA. The LORTA team suggested that the IE be focused on CBEWS and CBCRM measures in flood-prone river basins (Activity 3.1) and on flood and mudflow defence structures (Activity 3.3).[91] The UNDP team generally agreed on the proposed scope of the study; however, they requested that more hazards be considered in the evaluation of Activity 3.1, such as drought and landslides. The LORTA team agreed to consider the multi-hazard dimension of the project in the proposed IE design, and it was tentatively proposed to focus on two flood-prone river basins and one drought-prone river basin. Details of the resulting IE design follow below.

The UNDP team also raised the question of whether an IE of agrometeorological advisory services (Activity 2.3) would be possible. This activity plans to develop new climate information products, such as agri-climate maps, calendars and advisories, and deliver those to farmers. In particular, information on extreme events, precipitation, temperature, soil and growing conditions will be shared through SMS, radio, TV, web pages and agricultural extension centres. The dissemination will potentially benefit from public-private partnerships with cellular network companies. Farmers can then incorporate such information into their farming decisions (time of planting and harvest; time of applying fertiliser, pesticides and irrigation; choice of technologies). This intervention is meant to cover all farmers nationally; however, the project proposal envisages a pilot in year 4 (2022). The national scale-up will follow at the end of year 6 (2024).

The LORTA team considered an IE of this activity after returning home and concluded that while an IE of the pilot would be possible in principle, it is difficult to combine it with an IE of Activity 3.1, which seems of primary interest for the UNDP team.[92] We, therefore, do not elaborate in this report on evaluating agrometeorological advisory services.

C. THEORY OF CHANGE

Activity 3.1 of the project aims at improving resilience at the community level through securing the delivery and the use of CBEWS and the implementation of non-structural intervention measures (CBCRM). CBEWS and CBCRM will be implemented in 100 of the most vulnerable communities across the 11 major river basins. Communities will be chosen based on high hazard risk, short lead time of extreme events, and remoteness or connection problems to be serviced by the national MHEWS.

[91] The LORTA team originally proposed to focus on floods only, because floods are the most prevalent risk in terms of damage and loss of life.

[92] It would be possible to consider Activities 2.3 and 3.1 as two different treatment arms within one IE, but difficulties of implementing such an IE arise due to the different timelines and different target groups (beneficiaries) of these activities. Alternatively, two different IEs could be conducted, but this would have major consequences for budget needs.

The CBEWS will complement the national MHEWS. All chosen communities will receive warning communication and response tools, such as generators, sirens, boards and evacuation route signalling, to facilitate the seamless communication of warnings to remote areas and fast evacuation in case of natural hazards. This is to ensure that community members are aware of the risks they face and that national warnings are easily available at the local level. Training will be provided on how to interpret this information. Upstream communities, which are affected by very short lead time, will additionally receive risk monitoring devices, such as water level monitoring sensors and/or staff gauges, to be able to produce their own warnings.

Under CBCRM, communities will be trained to plan, implement and maintain non-structural intervention measures necessary to manage climate-induced risks. The planning and implementation of such measures will happen through a participatory process, paying special attention to gender mainstreaming. Examples for CBCRM measures are agroforestry activities, the establishment of locally controlled and managed flood zones, or watershed restoration.

Activity 3.3 will implement priority structural intervention measures in high-risk areas (based on sound cost-benefit analysis) to reduce the risks that the MHEWS will be designed to address. The aim is to provide infrastructural measures, such as wire mesh mat linings, gabions, embankments, concrete regulation walls, and riverbed and channel cleaning, to protect the most exposed communities against floods and mudflows. Thirteen sites across the country have been pre-selected where such measures are to be built. An increased level of physical protection can then lead communities to make more livelihood-enhancing, long-term investment plans that would otherwise not be possible in anticipation of frequent damages.

The important underlying assumption of both activities' theory of change (Figure 1 and Figure 2) is that communities understand and engage with the early warning schemes and risk management measures. The project thus needs to ensure a complete buy-in from vulnerable communities. If communities do not trust the early warnings put in place, are not well informed and ready to apply pre-defined response procedures, or do not participate in the planning and implementation of non-structural measures, an increased resilience cannot be achieved.

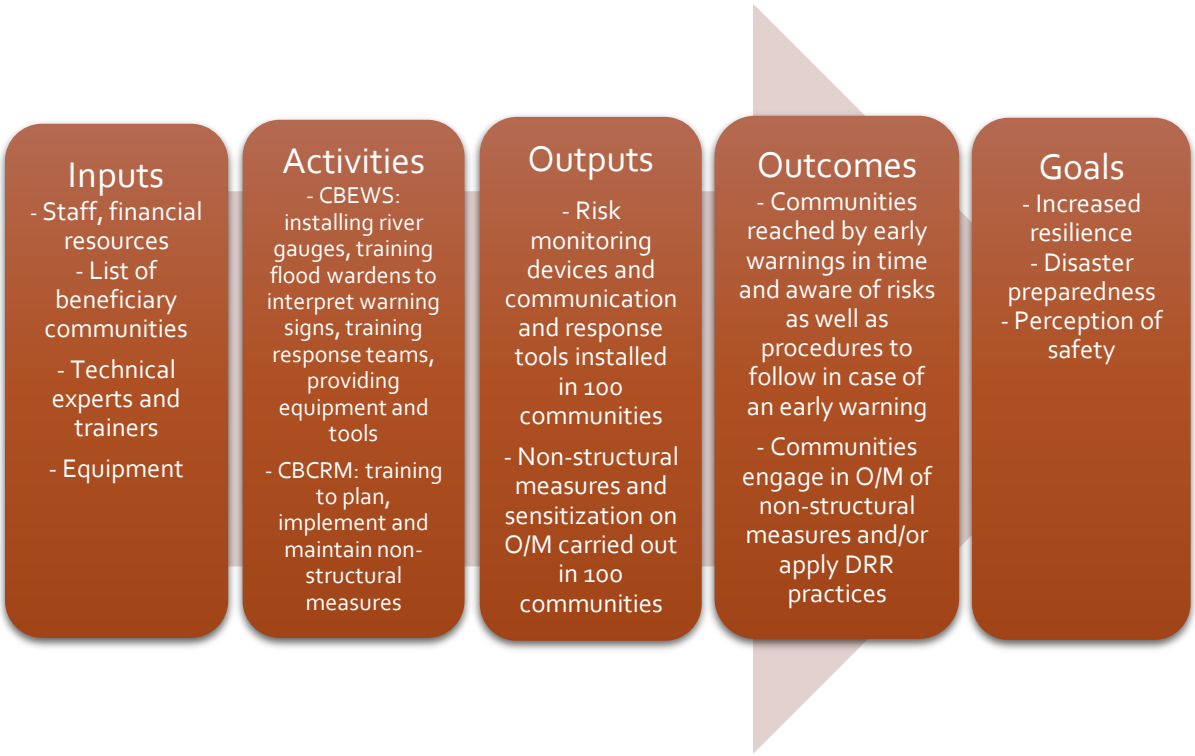


Figure 1 Theory of change – CBEWS and CBCRM non-structural measures

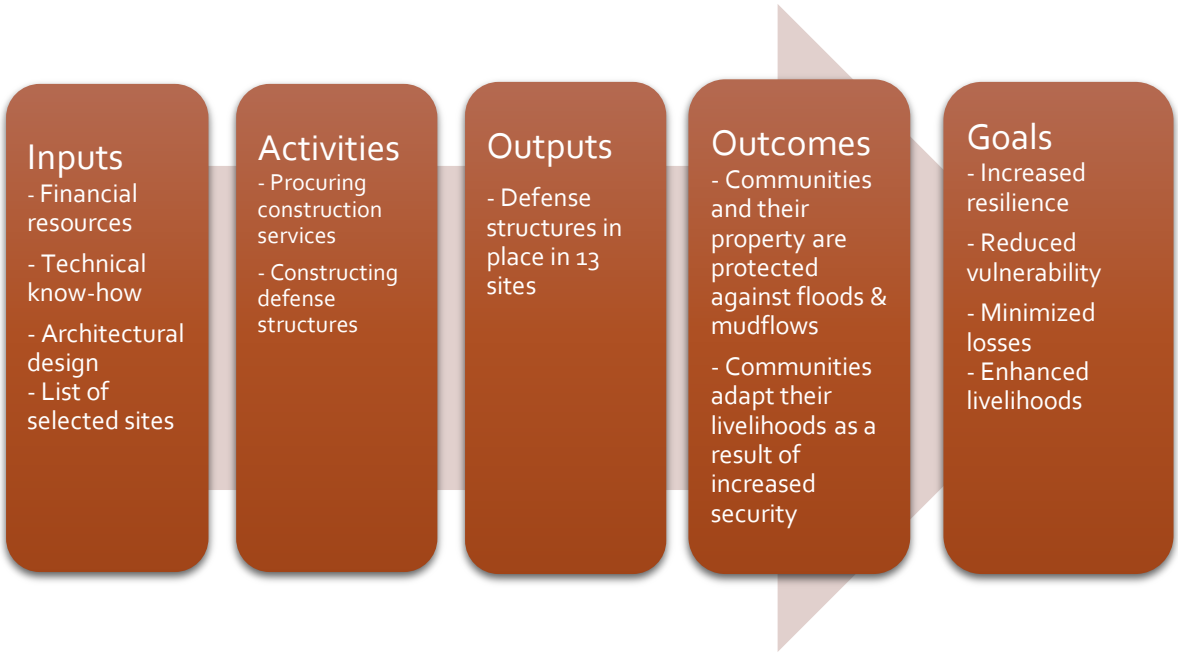


Figure 2 Theory of change – structural measures

Evaluation questions and indicators

The proposed IE will attempt to answer the following questions:

Evaluation Question 1: What is the impact of CBEWS and CBCRM on households' resilience against natural hazards?

Sub-questions (SQs) are as follows:

1. Are the warning messages beneficial for protecting the community against the effects of climate hazards?
2. Are communities provided with risk monitoring equipment, communication and response tools?
3. Is a community risk committee (responsible for reading and interpreting climate information and disseminating early warnings) formed and trained?
4. Are community members aware of hazards and risks?
5. Do community members receive and understand early warnings?
6. Do they trust the warnings?
7. Does the community engage in developing guidelines and procedures to follow in case of natural hazards?
8. Do community members know which guidelines and procedures to follow?
9. Do they feel better prepared for natural hazards?
10. Are the special needs of women, older adults, people with disabilities and children considered in the CBEWS?
11. Does the community have a participatory planning process for non-structural intervention measures?
12. Are women sufficiently represented in this planning process?
13. Does the community implement and maintain non-structural measures?

Indicators:

- Physical damage and loss of life in the aftermath of natural hazards SQ 1)
  - Number of injuries
  - Number of deaths
  - Crop loss / livestock loss / destroyed buildings
  - Monetary value of damages
- Resilience index (e.g. extent to which household depends on only one climate-sensitive livelihood activity, number of alternative activities that are less climate sensitive, wealth status, access to weather forecasts and climate information, savings, food reserves, availability of support network, effectiveness of community-level leadership) (SQs 1 & 13)
- Number of climate monitoring devices and communication and response tools installed (SQ 2)
- Evacuation routes marked (SQ 2)
- Existence of community risk committee/centre (SQ 3)
- Number of persons trained in CBEWS measures (SQ 3)
- Knowledge of relevant natural hazards and climate risks (SQ 4)
- Number of communication tools installed (SQ 5)
- Timing of warnings (SQ 5)
- Website established (SQ 5)
- Number of evacuated people (SQs 5 & 6)
- Number of community members in risk committee (SQ 7)
- Number of community risk meetings held (SQ 7)
- Emergency preparedness index (correct answers to questions on emergency procedures, test understanding of standardised warning signs in form of pictures → set of warning icons should be developed in the project) (SQ 8)
- Number of evacuation maps distributed/installed (SQ 8)
- Reported perception of own safety (SQ 9)
- Number of women involved in EWS (risk committee, maintenance workers, EWS trainings) (SQ 10)
- Evacuation plans account for specific characteristics of vulnerable groups (e.g. dress code of women, limited ability to move) (SQ 10)
- Number of meetings held for planning non-structural measures (SQ 11)
- Number of participants in these meetings (SQ 11)
- Share of female participants (SQ 12)
- Number of maintenance workers employed/assigned (SQ 13)

Evaluation Question 2: Do flood and mudflow defence structures contribute to greater resilience of beneficiary households?

Sub-questions (SQs) are as follows:

1. Do defence structures reduce the direct effect of natural hazards (e.g. effectively hold back floods)?
2. Are beneficiary households aware of the defence structures?
3. Do they feel better protected?

4. Do their long-term investment plans change due to an increased feeling of safety?
5. Does their economic status improve in the medium to long run?

Indicators:

- Physical damage and loss of life in the aftermath of natural hazards (SQ 1)
  - Number of injuries
  - Number of deaths
  - Crop loss/ livestock loss/ destroyed buildings
  - Monetary value of damages
- Awareness of defence structures (SQ 2)
- Reported perception of own safety (SQ 3)
- Crop diversification measured by number of crops planted per season (SQ 3)
- Number of high-risk cash crops planted per season to measure perception of risk (SQ 3)
- Plans of long-term investments and/or realized investments (SQ 4)
- Asset index (SQ 5)
- Income (SQ 5)

D. PLANS FOR EVALUATION

In this section, we propose two different IEs, one for measuring the impact of CBEWS and CBCRM measures (Activity 3.1 of the project) – henceforth referred to as intervention 1 – and one for measuring the impact of risk reduction structural measures (Activity 3.3 of the project) – henceforth referred to as intervention 2.

As we elaborate below, we find intervention 1 suitable for a randomized control trial (RCT) design, with ranodmization clustered at the community level. This IE will isolate the additional effect of intervention 1 as opposed to interventions with national coverage, such as MHEWS and public awareness campaigns, which will also affect the communities in the control group. The additional effects of intervention 1 will be estimated via the comparison of outcome levels between the treatment and the control group some time after the intervention activities have been implemented. Given the scarce evidence on the effect of early warning systems and disaster risk reduction measures, we find that the possibility of conducting an RCT is an extremely valuable opportunity to generate rigorous policy-relevant evidence. Furthermore, the multi-hazard dimension of the project may make it possible to provide evidence on multiple hazards within the same IE.

The IE of structural measures will answer the question of whether defence structures against floods and mudflows will contribute to improved livelihoods in the areas concerned. For this IE we propose a non-experimental design – namely, difference-in-differences (DiD) coupled with propensity score matching (PSM). Any experimental design for this intervention would be out of the question since all 13 sites for implementation of defence structures have already been selected.

Intervention 1 (Activity 3.1): randomized control trial

One hundred of the most vulnerable communities across the 11 major river basins will be identified for the implementation of CBEWS and CBCRM measures. In order to compute communities' level of vulnerability and hence to identify beneficiary communities for intervention 1, the project plans to conduct a risk zoning of all major river basins, resulting in hazard maps, and a socioeconomic

vulnerability assessment of all communities in these basins.[93] The communities lying at the top of the vulnerability distribution will then be defined as eligible. It is currently assumed by the project team that considerably more than 100 communities would be eligible. However, the project has the capacity and the funds to implement CBEWS and CBCRM measures in only 100 of them, which lends itself as a suitable setting for an RCT, for which there was broad buy-in from UNDP and MoEPA during the field mission. From among all eligible communities, 100 can be randomly assigned to intervention 1 and the others can constitute the control group for the IE.

Because of implementation constraints (with regard to budget and capacity), the risk zoning and socioeconomic vulnerability assessment are going to be phased in over time. According to discussions with UNDP and key government stakeholders during the field mission, we understand that the mapping will be completed for three river basins by the end of year 2 (2020), for an additional four basins by the beginning of year 4 (2022), and for the remaining four basins by the end of year 4. The order in which the basins will be mapped has not yet been established. Implementation of CBEWS and CBCRM measures will start in each of these respective groups after the mapping and socioeconomic vulnerability assessment have been completed.

In the field mission, it was agreed to focus the IE only on the three river basins that will be mapped first (and that these should be two flood-prone areas and one drought-prone area). However, this would crucially limit the number of communities that undergo treatment and could narrow down part of the data-collection sample to around 30, given the target of 100 beneficiary communities in total. We, therefore, propose to include all 100 treatment communities in all 11 river basins in the IE. The definition of eligibility, ranodmization and treatment of communities would be phased in over time, but an IE that considers all 100 communities would nevertheless be possible.[94] We fully acknowledge that this needs to be agreed upon with the project team.

Figure 3 visually represents the RCT strategy. In each of the three river basins portrayed, communities will be ranked according to their multi-hazard and socioeconomic vulnerability. The most vulnerable communities will be defined as eligible for the treatment. Out of these, approximately 10 communities per basin would be randomly selected as beneficiaries for intervention 1, while the remaining communities would serve as the control group.

[93] The risk zoning and the socioeconomic vulnerability assessment are part of component 1 of the project.
[94] The incorporation of additional basins and/or hazard strata at different years can be accounted for in the regressions by including time dummies and their interaction terms with stratum dummies.

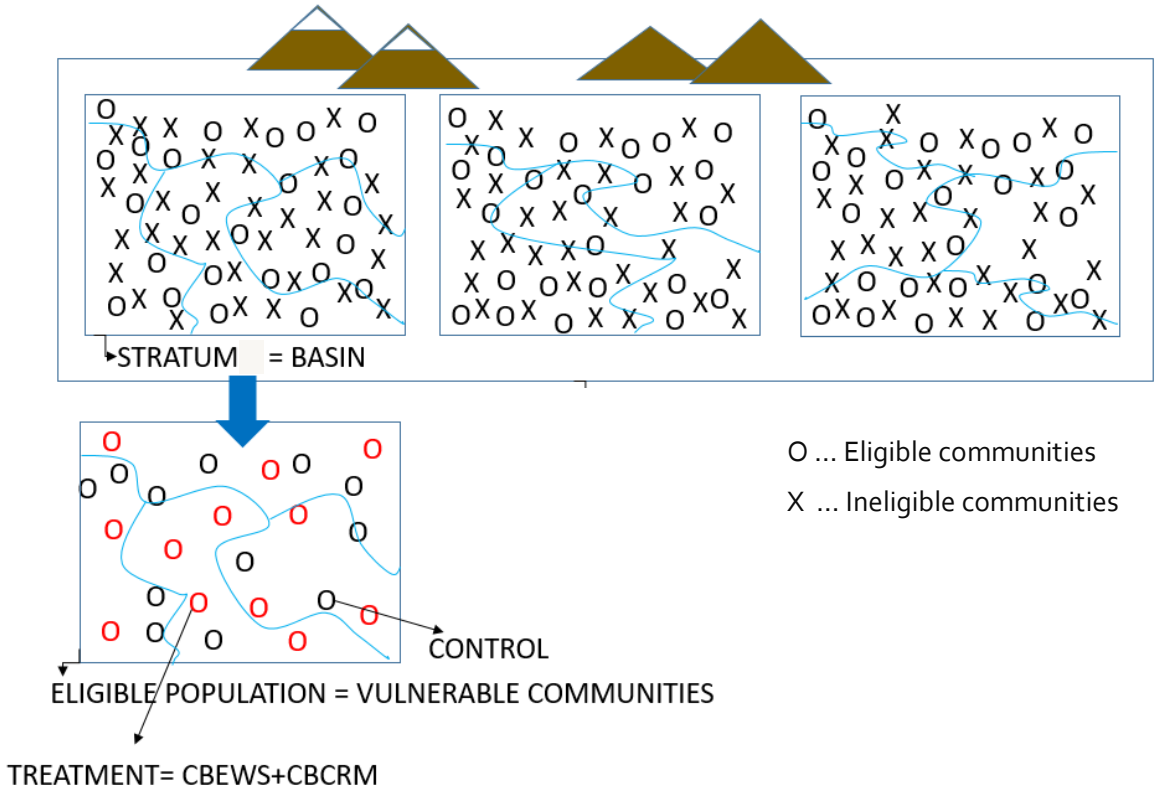


Figure 3 Sketch of the randomized control trial strategy

Since different river basins are affected by different natural hazards, it will be necessary to treat the river basins, or groups of river basins, with the same hazard risk as separate strata in this IE and in the sampling for data collection. This stratification ensures that treatment status is not correlated with differences across river basins, which could bias the estimated effect of intervention 1. It also ensures that there will be treatment as well as control communities from each river basin, or group of river basins, in the final sample.

Let us stress that a key need for the feasibility of the RCT strategy is for the eligibility threshold to be established in a way that it allows for a sufficiently large eligible population (ideally, 300 or more communities are eligible). At the time of the field mission, it was still unclear how exactly the vulnerability of communities would be defined and where the threshold for eligibility would be set. This needs to be discussed further with the project team.

One challenge and exciting particularity of the IE is that the CBEWS and CBCRM measures aim at increasing communities' resilience against multiple natural hazards. While some communities tend to be affected by flood or landslides, others face the risk of drought, and still others suffer from further hazards. Some communities are even affected by several hazards at the same time. Hence, which communication and response tools will be provided to the community and which non-structural intervention measures will be implemented will be tailored to the communities according to the risks faced. However, it is our understanding that all communities do receive the same *kind* of activities – namely, communication and response tools as well as non-structural intervention measures – and can, therefore, be evaluated together.[95]

[95] In the later data analysis, it is possible to include interaction terms for treatment status and hazard faced to investigate whether the same kind of activity is more or less effective for certain hazards.

As noted above, some communities, namely those from upstream communities, will receive risk monitoring tools and facilitation to produce their own early warnings, in addition to the standard set of interventions. This difference in the kind of activities received could be taken into account by not only stratifying the communities according to river basin (or groups of river basins) but additionally according to their upstream or downstream location. Whether this additional stratification is feasible and required needs to be decided when the first list of eligible communities is completed.

Furthermore, heterogeneous analyses might be relevant within the scope of this IE. In particular, the project claims to use gender-sensitive approaches. A gender-specific analysis would be possible through distinguishing between female-headed and male-headed households. For selected outcome variables, it might also be useful to administer a specific questionnaire module to women in each surveyed household. Other heterogeneous analyses, based on income levels or any other relevant dimensions, can be performed, provided that sample size is sufficient. For instance, the sample could be split into a low-income and a high-income group based on the median income (50 per cent below the median and 50 per cent above the median). We take into account such disaggregation in our power calculations.

As discussed during the field mission, it may be necessary to make communities near the new structural sites (intervention 2) ineligible for intervention 1 in order to avoid confounding the two separate interventions of the project in one IE. We propose another IE design for the structural measures alone.

Overall timeline for the randomized control trial

The proposed timeline for the first IE (the RCT) is described in Table 1. At the end of 2020, the risk zoning and socioeconomic vulnerability assessment will be completed in the first three river basins, called group 1. In early 2021, the identification of eligible communities in this group can thus be completed, and eligible communities can be randomly assigned to either treatment or control. We suggest collecting baseline survey data in the sampled communities just before intervention 1 is rolled out. If groups 2 and 3 form part of the RCT, the same sequencing of activities will be implemented once the risk zoning and vulnerability assessments are completed.

To limit the costs as much as possible, there is the possibility of avoiding baseline data collection in river basins mapped by year 4 (groups 2 and 3). A wealth of information will be collected from each community during the hazard risk mapping and socioeconomic vulnerability assessments, which could be used to check balance at baseline. While such information is in principle also available for group 1, we nevertheless highly recommend the collection of baseline survey data in these basins as this can prove very useful. Importantly, at this stage we do not know the distribution of communities within the basins or the size of the eligible population, which could affect statistical power. Baseline covariates can be used to improve statistical power and – in case ranodmization fails – allow for the use of quasi-experimental methods such as DiD and PSM. In principle, it would be possible to forgo any baseline survey data collection. If ranodmization was successful, a baseline survey would not be required.

While it would be desirable to conduct a midline survey to be able to observe differences between treatment and control communities in the short term (for group 1; for example, in early 2023), this is not a necessity and, given the restricted IE budget (see Budget section below), it does not seem to be a viable option. Endline data could be collected as late as 2025.

Table 1 Timeline for the randomized control trial impact evaluation

YEAR AND QUARTER	MAIN ACTIVITY
2020 Q4	Finalised hazard risk maps and socioeconomic vulnerability assessments for group 1 (i.e. first three river basins)
2021 Q1	Identification of eligible communities for intervention 1 and ranodmization to treatment and control in group 1
2021 Q2	Baseline data collection in group 1
2021 Q3	Roll-out of intervention 1 in group 1
2022 Q1	Finalised hazard risk maps and socioeconomic vulnerability assessments for group 2 (i.e. additional four river basins)
2022 Q2	Identification of eligible communities for intervention 1 and ranodmization to treatment and control in group 2
2022 Q3	Roll-out of intervention 1 in group 2
2022 Q4	Finalised hazard risk maps and vulnerability socioeconomic assessments for group 3 (i.e. last four river basins)
2023 Q1	Identification of eligible sample for intervention 1 and ranodmization to treatment and control in group 3
2023 Q2	Roll-out of intervention 1 in group 3
2025 Q1	Endline data collection in all three groups

Power calculations

Power calculations enable us to determine the minimum sample size needed in order to detect the impact of a given intervention. To do that, we use the following power formula that relates the sample size to the minimum detectable effect size (MDES) between the mean outcomes of two groups:

$$MDES = (t_{1-\kappa} + t_{\alpha}) \sqrt{\frac{1}{P(1-P)}} \sqrt{1 + \rho(m-1)} \sqrt{\frac{\sigma^2}{N}} \sqrt{1 - R^2}$$

where $t_{1-\kappa}$ and t_{α} are t-statistics representing the required power and level of statistical significance (by convention, we seek a power of 80 per cent and a significance level of 5 per cent), P represents the proportion in one of the two compared groups (allocation ratio), ρ is the intra-cluster correlation (ICC), m is the number of individuals per cluster (community), σ^2 is the variance, N is the total sample size, and R^2 represents the extent to which baseline characteristics predict the endline outcome variable.

Given the context of our IE, we provide calculations for the above described IE sample of three river basins. We consider a cluster RCT with two arms:

- Households of beneficiary communities of CBEWS and CBCRM (treatment group, *T*)
- Households of communities that are eligible for CBEWS and CBCRM but that are not beneficiaries (control group, *C*)

We will then be interested in the comparison of outcomes between *T* versus *C*. We apply the optimal sample allocation ratio of 0.50 – that is, equal sample size in treatment and control group.

We consider both a sample consisting of three basins (group 1 only) and a sample of 11 basins (all three groups). We also consider the possibility of disaggregating them into 50%–50% subgroup

analyses. The subgroups can be created based on any category of interest – for instance, based on low- and high-income levels.

Evaluation question 1: Impact on standardized outcomes

Evaluation question 1 refers to the additional impact of CBEWS and CBCRM on households' resilience against natural hazards and on their disaster preparedness as opposed to national scale programme components. Currently, there is very scarce empirical evidence on impacts of EWS and DRR interventions. We were only able to find one effectiveness review of disaster risk management programmes, prepared by Oxfam;[96] we found no articles published in scientific journals. The outcome in this report is measured via an index prepared ad hoc, and standard errors are not reported, which renders the statistics difficult to be applied to power calculations in our context.

Table 6 shows the power calculations for the RCT. We consider a standardized index, which could serve to measure preparedness, resilience or any other relevant outcome. The use of standardized indices fits this project particularly well, as it makes it possible to compare outcomes across different hazards. We assume different values for ICC and number of clusters (30 or 100 communities in both treatment and control). We set the R^2 to either 0 per cent or 30 per cent. The second value considers the possibility that we will be able to capture some of the variation in outcomes through household baseline covariates. The upper panel shows results for the full sample of either 60 or 200 communities, while the bottom panel disaggregates the sample into half to allow for 50%–50% heterogeneity analysis. In all cases we are trying to achieve an MDES of 0.20 of a standard deviation. This is the standard MDES commonly sought by researchers; any change lower than 0.20 is usually considered economically insignificant.

Focusing on the results in the upper panel of Table 6, we prefer to consider the more conservative scenario where the ICC is set at 20 per cent. For instance, resilience to shocks is very likely to be highly correlated among households living in the same community. The actual size of the eligible population, and therefore the number of available clusters, will be key in determining statistical power. We achieve an MDES of 0.20 using 200 clusters that are equally split into treatment and control groups in all scenarios considered. With 60 clusters, the MDES of 0.20 is never achieved. Counting with 25 households per cluster, the overall required sample size for the RCT therefore amounts to 5,000 households.

Moving to the bottom panel, the target MDES of 0.20 is only achieved for a sample of 5,000 households and an ICC of 10 per cent. While we can expect some gain in power due to a decrease in the standard deviation when looking within subgroups, we cannot provide any reliable estimate about the magnitude of this gain at this point. Therefore, we can consider the bottom panel in the table as a conservative estimate, in which we would be underpowered for conducting a 50%–50% heterogeneity analysis for most scenarios.

Intervention 2 (Activity 3.3): Structural measures

This second proposed IE aims to answer the question of whether providing communities with supplementary protection against extreme natural hazards such as floods and mudflows, by constructing robust defence structures, actually changes communities' perception of their safety and resilience, and changes their attitudes and plans towards long-term planning and investment.

Structural risk reduction measures will be implemented in 13 sites selected from a list of 21 eligible sites across the whole country. The selection had already been made before the start of the project.

---

[96] Oxfam GB. (2012). *Community-based Disaster Risk Management and Livelihoods Programme. Effectiveness review – summary report.* Retrieved from https://policy-practice.oxfam.org.uk/publications/effectiveness-review-community-based-disaster-risk-management-and-livelihoods-p-247231

Construction work on six sites will be completed by the end of year 4 (2022), and the remaining seven sites will be finished by the end of year 7 (2025).

The treatment group will consist of households living in the proximity of the above-mentioned selected sites. Based on field visits and discussion with key stakeholders, it is assumed that a minimum of 80 households will benefit directly from a given structure as it will protect their property and agricultural land from floods and mudflows. This would sum up to 1,040 direct beneficiary households. An economic analysis underlying the project proposal came up with a larger number: 3,500 properties would be protected, thus benefiting 6,500 people. As a conservative approach, we estimate that these would represent 3,000 households. In the power calculations below, we consider two scenarios: 80 beneficiary households per site and 230 beneficiary households per site.

To construct the control group, we can use the other sites that were shortlisted for structural measures but will not be covered by the project due to limited funds. However, based on the current shortlist, we have only eight control sites as opposed to 13 treated ones. Given the PSM design, it would be desirable to ensure a higher sample size in the control group to allow for observations not being on common support with observations in the treatment group and thus not being usable in the data analysis.

We see scope for two possible approaches. First, we could sample more households in the proximity of control sites than in the proximity of treatment sites. However, since the propensity to be affected by floods and/or mudflows is strongly correlated with the physical proximity to the defence structure, sampling more households per control site could be problematic. The level of risk faced by these additional households (geographically further away from the control sites) would not necessarily be comparable to the treated households. Including these households does not likely lead to more observations on common support. While we rule out this option at the moment, a further discussion of the radius of influence of defence structures with the project team appears useful. Second, if it was possible to identify additional vulnerable sites, with the help of UNDP and the project's chief technical adviser, we could sample observations from them to be used as control units.

In Table 2, we present our proposed timeline for baseline, midline and endline data collection, based on the planned project implementation as laid out in annex E of the UNDP GCF proposal. As the main outcome of interest is households' subjective feeling of safety and resilience, it is crucial that baseline data be collected before the start of detailed design and construction works so that people's expectations about their vulnerability to future flood and mudflow risks are not biased. This is important especially since the proposed DiD approach will make use of baseline indicators.

Table 2 Timeline for the propensity score matching impact evaluation

2019 Q1	Identification of extra vulnerable sites to be added to the control group (optional)
2019 Q2	Baseline data collection
2019 Q4	Detailed design of structural measures prepared by local engineering/construction company for two sites
2021 Q4	Detailed design for 6 (in total) flood protection structures
2022 Q4	6 structural flood protection investments implemented
2024 Q4	13 (in total) site-specific structural flood protection measures implemented
2025 Q1	Endline data collection

Power calculations

For the power calculations we again consider the standardized index as an outcome indicator. As above, such an index can be used to measure resilience or households' feeling of safety. To estimate the sample size necessary for achieving the desired MDES of 0.2 with the power of 80 per cent and significance level of 5 per cent, we perform regular sample size calculations assuming an experimental design. However, in order to ensure that the final sample size does not shrink too much after the PSM procedure (that is, after trimming the observations that lie outside of the common support), it would be desirable to oversample the control group.

Table 6 shows power calculations for a standard RCT, considering a clustered design with 13 clusters in the treatment group. We provide scenarios for having 13 treatment and 8 control sites as well as for having 13 treatment and 13 control sites. As the design of the IE is non-experimental and we plan to match treatment and control units, it is necessary to make the control group as comparable as possible to the treated group. In other words, we need to ensure a high degree of overlap in the propensity score distributions of the two groups. Considering only eight control sites for 13 treatment ones might prove problematic, since we need to achieve equal sample sizes in *C* and *T*, which would imply a higher number of households per site in the control group. As noted above, this would probably imply selecting households from a wider radius around each respective site. But the households further away are likely to be less vulnerable, and, therefore, less similar to the households living closer to the sites in the treatment group. Therefore, we suggest increasing the number of control sites to at least 13, to avoid sampling more households per site in the control group than in the treatment group. If possible, we should strive to find an even higher number of control sites, as it would increase the chances of achieving a well-matched sample. In the best scenario, we should have 26 control sites for the 13 treated ones as we would ideally sample two control units per one treatment unit given the same distance to a defence structure.

We suggest sampling either 80 observations per site or 160 (possible only if the number of direct beneficiaries is 230 rather than 80). From the table, it is obvious that the desired MDES of 0.2 can be achieved only by assuming a very low ICC (2.5 per cent or lower). Due to the lack of relevant studies or data sets on the topic of disaster resilience that would inform us on the ICC that we should expect, we cannot confirm whether this is a reasonable assumption. However, in most cases when data are missing, the ICC is usually assumed to be at least 10 per cent, and so the 2.5 per cent seems unreasonably low. The inability to achieve a low MDES is due to the small number of clusters – we have only 13 sites per arm.

Given the 2.5 per cent ICC, the recommended total sample size for an RCT with 13 treatment sites and 13 control sites would be 2,080 observations. Sampling 160 rather than 80 observations per site does not increase the statistical power and would therefore not be advisable. As we rely on matching rather than an RCT, the standard procedure is to increase the sample size in the control group in order to ensure that the treatment units are well matched, and that the sample size remains sufficiently large, even after the observations outside of the common support are dropped. We suggest sampling two control units for each treated unit, which implies a total sample size of 3,210. This procedure would only be possible if the number of direct beneficiaries per site was 230 rather than 80, as we would otherwise run into the problem of sampling households that are further away from given sites and that are thus less comparable.

Qualitative assessments

During the field mission, we did not perceive a strong interest from the UNDP team or the UNDP partners (NEA and MoEPA representatives) in conducting qualitative assessments, such as focus group discussions (FGDs) and key informant interviews with stakeholders involved in the interventions. One reason for their low interest may be the restricted budget available for evaluation.

If it turned out that the ICC to be expected in communities where structural defence measures are to be constructed, is considerably larger than 2.5 per cent, we would propose considering the option to entirely assess the structural measures via qualitative research methods. Clearly, this would not allow for a rigorous, quantitative IE but it could help to uncover perspectives and changes in attitudes and behaviour that would be difficult to investigate with quantitative data. We are open to discussing this option further.

Monitoring & evaluation

UNDP showed a strong commitment to the monitoring and evaluation (M&E) activities throughout the project timespan.

Specifically, the project proposal mentions six domains of impact for M&E:

- Physical and financial assets
- Social capital, empowerment, change of behaviour
- Food security
- Environment
- Institutions, policies, regulatory framework
- Gender

Within the M&E activities, both community and household survey data collections are planned. Based on the discussions held during the mission, we understand that the M&E household survey target is about 200 households every two years. At this stage, the LORTA team is awaiting more specific details.

During the mission, the UNDP team showed some availability and flexibility in coordinating the M&E activities with the IE. For example, the UNDP team indicated the possibility of applying M&E to the beneficiary communities in the IE sample. The LORTA team expects to follow up on the M&E planning and to collaborate on the development of M&E assessments.

Budget

The budget assigned to the IE in the project proposal amounts to about USD 100,000 in total for baseline, midline and endline data collection. According to the estimates that UNDP Georgia obtained from one of the survey companies that previously assisted them in evaluating projects, the survey cost covering 3,000 households mostly residing in rural or remote areas should be USD 50,000 given that the interviews last approximately one hour. The estimated price is significantly lower than in other LORTA countries and might ensure financial feasibility of both baseline and endline data collection. Table 3 reports the estimated costs for each of the two IE designs, considering different scenarios. The costs related to the RCT are USD 133,000 in the case of three waves of data collection and USD 108,000 in the case of two waves of data collection. Conducting endline data alone would result in a cost of USD 83,000. Costs of the DiD approach are comparable with the two-wave RCT, but as mentioned earlier, there are concerns about an insufficient statistical power due to a low number of clusters.

Cost estimates were also obtained for qualitative data collection. One FGD with 10 participants is estimated to cost about USD 700 to USD 850 in rural locations and USD 600 to USD 750 in urban locations. The average cost estimate for one in-depth interview ranges from USD 50 to USD 60.

A part of the project's budget has been assigned specifically to M&E activities. The LORTA team asked UNDP to consider budget reshuffling as well as cost sharing options in order to secure sufficient funds for the IE. Some M&E activities could possibly coalesce with the IE data collection.

We are currently awaiting more details on the exact M&E budget as well as on the planned M&E activities.

Table 3 Cost estimations for various impact evaluation alternatives

ALTERNATIVE	NUMBER OF OBSERVATIONS	UNIT COST PER INTERVIEW (USD)	COST PER WAVE OF DATA COLLECTION (USD)	TOTAL COST (ROUNDED TO THE NEAREST HUNDRED USD)
CBEWS and CBCRM				
Scenario 1 (only endline):				
RCT	5,000	$16.67	$83,350	$83,400
Scenario 2 (baseline for group 1 + endline for groups 1,2,3):				
RCT	5,000 (1,500 baseline)	$16.67	$83,350 endline $25,005 baseline	$108,400
Scenario 3 (baseline and midline for group 1 + endline for groups 1,2,3):				
RCT	5,000 (1,500 baseline, midline)	$16.67	$83,350 endline $25,005 baseline, midline	$133,400
Structural defence measures				
Scenario 1 (baseline and endline):				
DiD with PSM	3,160	$16.67	$52,677	$105,400
Scenario 2: Qualitative assessment				
FGDs in 13 sites	13	$850	$11,050	$11,050

Main challenges for the impact evaluation

A key challenge is that the RCT IE would identify the additional effect of CBEWS and CBCRM on the knowledge of hazards, disaster preparedness and resilience on top of the nationwide interventions. The latter include MHEWS, agrometeorological climate information, pervasive public awareness campaigns, capacity-building at all levels on climate risk, and training of local first responders. More information on the CBEWS and CBCRM activities is then crucial in order to understand the real potential for additionality of these subcomponents. An important remark is that while the project proposal indicates that the capacity-building of communities and local trainings of first responders would happen on a national scale, the discussions during the LORTA mission did not confirm this clearly. If local capacity-building and training efforts eventually target only, or with higher intensity, the vulnerable communities treated by CBEWS and CBCRM, this would work in favour of the RCT IE.

A second related point is the correct definition of the unit of analysis, which crucially affects the IE design. In this report – following the discussions during the LORTA mission – we consider households to be the unit of analysis. However, given that CBEWS and CBCRM are community-level interventions, it is still unclear whether all households in a given treated community will actually receive any treatment. If not, a random sample of households will not enable us to detect an impact. It might be that the community-level activities will rather target restricted groups of households or individuals, such as rescue groups or disaster risk management committees, in which

case it would be more interesting to focus on these subgroups for the baseline and endline surveys. However, using rescue groups or disaster risk management committees as a unit of analysis would certainly have negative implications on the maximum possible sample size,[97] which might preclude a rigorous quantitative analysis.

Potential risks also emerge from the implementation deadlines set up by the project. Since the RCT cannot be implemented without the risk zoning and socioeconomic vulnerability assessment being completed, data collection – if a baseline will take place – cannot start before 2021, or later in case of delays. Baseline data collection for the IE of the structural measures would, of course, occur much earlier, in 2019.

An additional threat to the IE derives from the communities' and households' potential knowledge of the implementation timeline and the criteria of selection. For example, in the case of structural measures for which preparatory activities have already started, the LORTA team enquired whether the related beneficiary communities and households are currently aware of these construction plans. If that is the case, households could be anticipating the results of the intervention and provide biased responses in the baseline survey. Another threat would arise if communities were aware of specific vulnerability criteria for their selection as beneficiaries of the CBEWS and CBCRM intervention. Community members would be tempted to overestimate their self-reported survey measures of vulnerability during socioeconomic assessments in order to increase their chance of being selected as beneficiaries.

Further, the availability of future donor funding may contaminate the IE control group. This issue was discussed during the LORTA mission, and the LORTA team made the case that any additional interventions must be carried out in areas outside of the IE sample, or at least should be postponed until endline data have been collected if donors intend to target the IE sample areas.

A general concern relates to the budget constraint for the IE. The LORTA team understands that the budget allocated for IE is about USD 100,000. This is not a large budget and does not allow for two IEs to be implemented. We are therefore considering the possibility of cost sharing with the planned M&E activities, even though we must acknowledge that surveys to be conducted under M&E are planned to target rather few respondents. The potential for cost sharing is therefore rather small. Nevertheless, if the cost sharing were to be undertaken – and therefore M&E activities were to be applied to the IE sample – a potential challenge is an adequate coordination such that the M&E activities do not undermine the IE activities or contaminate areas assigned to the control group.

Secondary data sets for the impact evaluation

IE can benefit from the integration of primary data collected with a multiplicity of secondary data sources – both at baseline and at later stages.

Secondary data may be retrieved from activities within the project implementation, project M&E and external sources. The following list of sources is considered. This list is not exhaustive and other sources may be added later:

- Baseline hazard vulnerability maps
- Baseline and M&E community socioeconomic assessments
- Community vulnerability score databases
- M&E household assessments
- Disaster databases from the Ministry of Internal Affairs and the State Security and Crisis Management Centre

[97] As there is probably only a small number of rescue groups per community.

- Frequency and timeliness of MHEWS and CBEWS warnings per community from NEA
- Climate and weather data from NEA
- Soil and environmental quality assessed by the project experts within the M&E activities
- M&E records on community participation in the project activities and trainings
- M&E and CCTV records of evacuated people
- Performance records of emergency drills

The LORTA team has so far requested a few data sets from the prototype project. This may help us understand the content of the mapping and vulnerability assessments, which will be similarly implemented in the scale-up project. The LORTA team will coordinate with the UNDP project team on the need for additional data sources.

Assessment of institutional capacity for the impact evaluation

Georgia project team possesses moderate capacity to successfully carry out rigorous IEs. While the UNDP Energy Efficiency team and the M&E specialist broadly understand the purpose of an IE, there are no experts with training in quantitative methods who would be able to correctly apply the IE tools. We did not come into contact with the consultancy firm that was being considered for the baseline data collection, and so we cannot conclude on their capacity.

Nevertheless, the UNDP team, project team, as well as the representatives from the NEA and MoEPA, appreciated the LORTA mission and showed a genuine interest in the methods for rigorous IE. During the capacity-building workshop they understood the crucial need for a credible counterfactual and why a before-and-after comparison does not enable us to assess the impact. They also understood the necessity of adjusting implementation plans in order to accommodate the IE needs.

Overall, we consider it to be reasonable to provide support to the UNDP Georgia in carrying out a rigorous IE. UNDP Georgia could also benefit from C4ED's presence during the inception workshop in February 2019; it is necessary to ensure that the IE is designed to answer a meaningful evaluation question (i.e. that the intervention packages are comparable across different hazards), and that the planned activities are on a level sufficiently close to the households, so that we can observe an impact at the household level. C4ED could also convey insights from the literature and discuss with the project team regarding how to render the interventions more effective. Furthermore, it would be optimal if C4ED could be involved in the baseline data collection to ensure quality of the collected data as well as of the outcomes of interest used in the later analysis. In particular, attention must be paid to developing a solid survey instrument.

III. WAY FORWARD

In summary, we proposed two possible impact evaluation designs: a two-arm RCT and a DiD with PSM. The crucial next step for UNDP Georgia and the GCF IEU is now to decide which IE design should be accepted and put into action. It seems unlikely that the budget will be sufficient to cover both options. For statistical and feasibility reasons listed earlier, C4ED recommends focusing on the two-arm RCT and to conduct qualitative assessments for the other project's components (structural measures).

Overall, we consider that the LORTA mission to Georgia went smoothly and that there is a strong potential for a rigorous IE. The success of the LORTA mission has been particularly achieved thanks to the attentive collaboration of the UNDP Energy Efficiency team leader, Nino Antadze, and her colleagues. The LORTA team also benefited greatly from the presence of the previous project's chief technical adviser, Margaretta Ayoung, who also contributed to the proposal writing of the

current LORTA project. Their input was crucial in helping us understand the details of the project, arranging the meetings and making sure that all the key stakeholders were consulted.

The LORTA team continues its support remotely. Currently, we are communicating with UNDP Georgia in order to receive all the crucial information that is still missing (especially vulnerability assessments, vulnerability scores, etc.). We are also reviewing the literature for evidence on how to make CBEWS and CBCRM interventions more effective in increasing communities' resilience, so that we can lead informed discussions with the UNDP project team during the upcoming inception workshop.

Depending on which IE options are selected for implementation, it will be crucial to ensure that there are sufficient funds for all the necessary waves of data collection. We can further continue discussions on the budget lines with UNDP Georgia once it becomes clear what IE options we will go forward with.

The future success of this project within the LORTA framework is highly conditional upon a continuous responsive collaboration from UNDP Georgia and other key stakeholders involved.

APPENDIX I

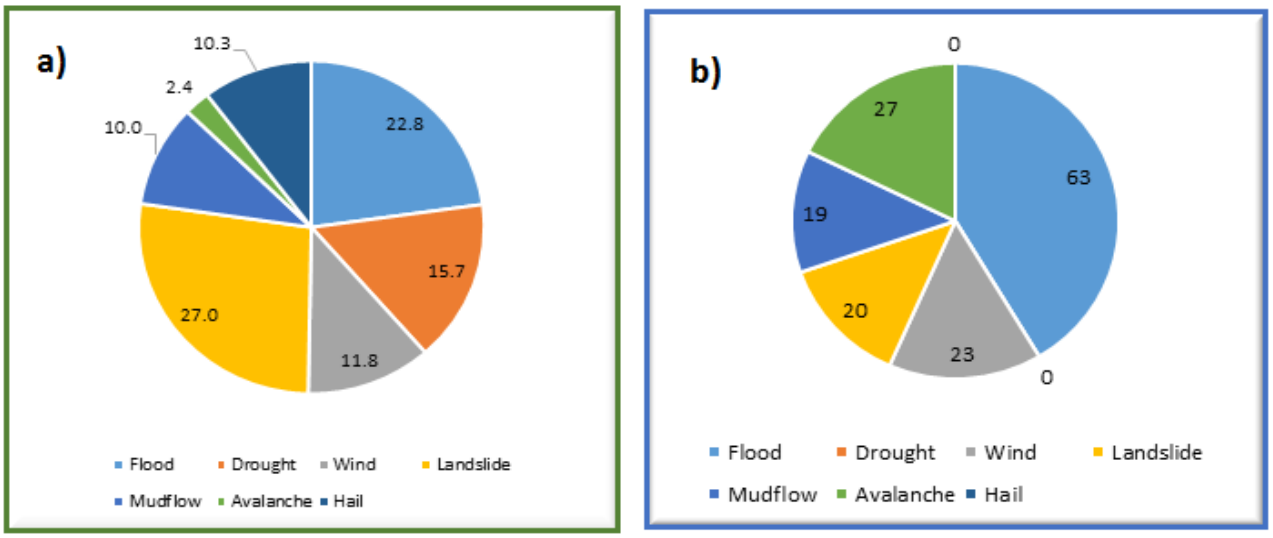


Figure 4 Statistics for all hazards, 1995–2015: a) Damage, b) Life loss

Source: Annex II of Feasibility Study (2017)

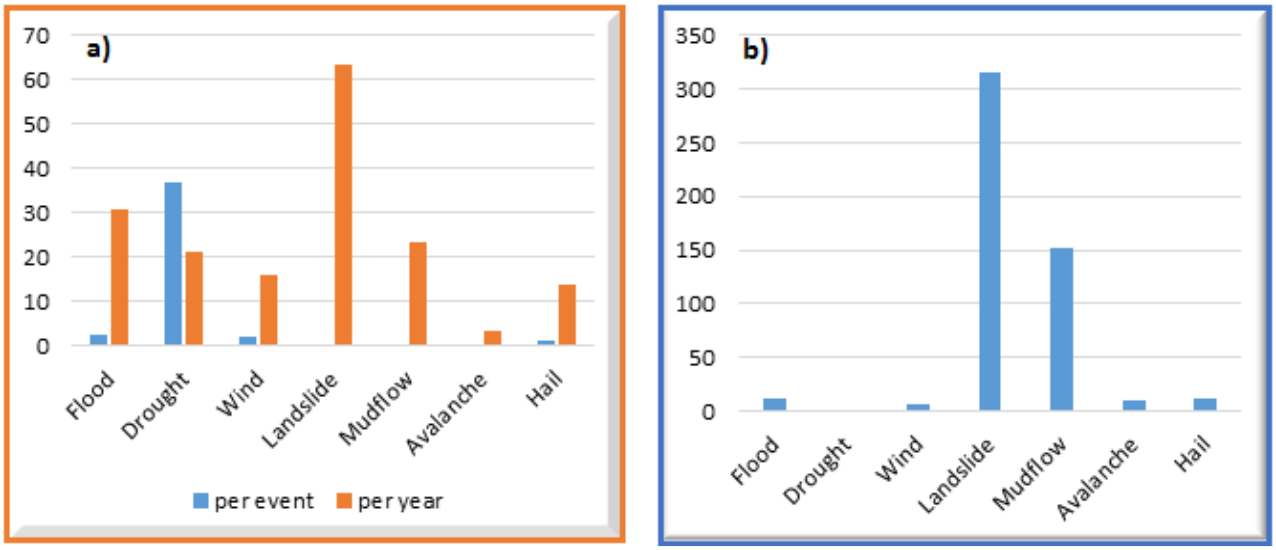


Figure 5 Statistics for all hazards, 1995–2015: a) Damage per event and year, b) Frequency per year

Source: Annex II of Feasibility Study (2017)

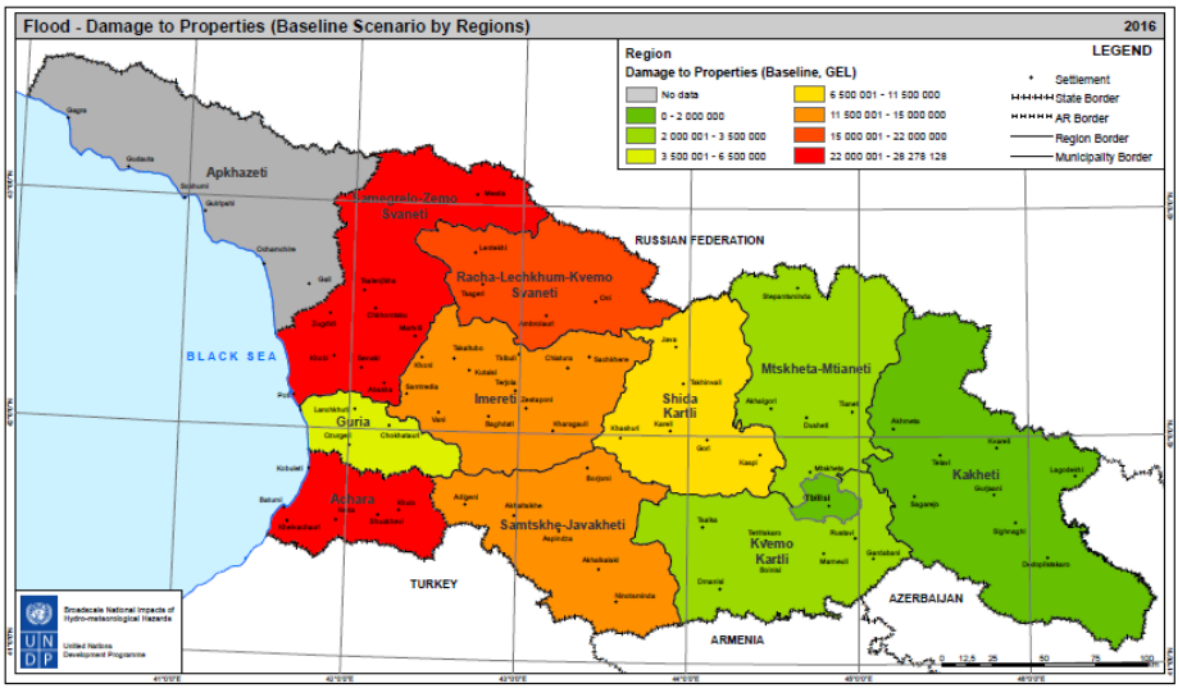


Figure 6 Regional scenario for flood (2016)

Source: Annex II of Feasibility Study (2017)

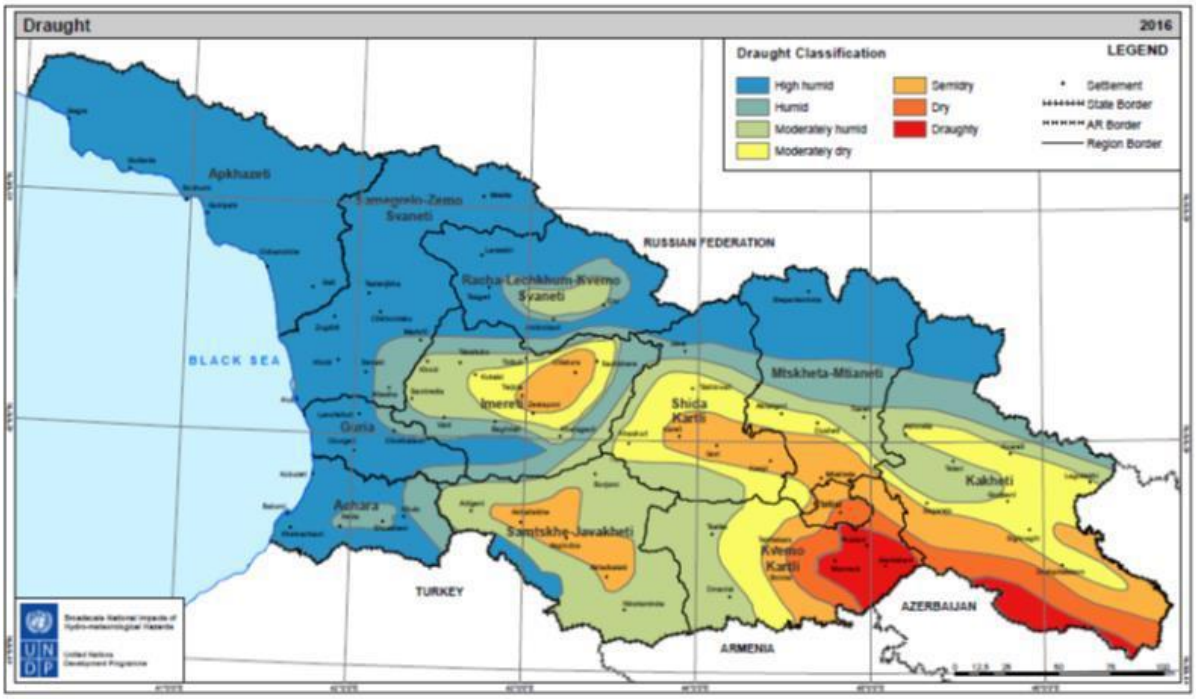


Figure 7 Regional scenario for drought (2016)

Source: Annex II of Feasibility Study (2017)

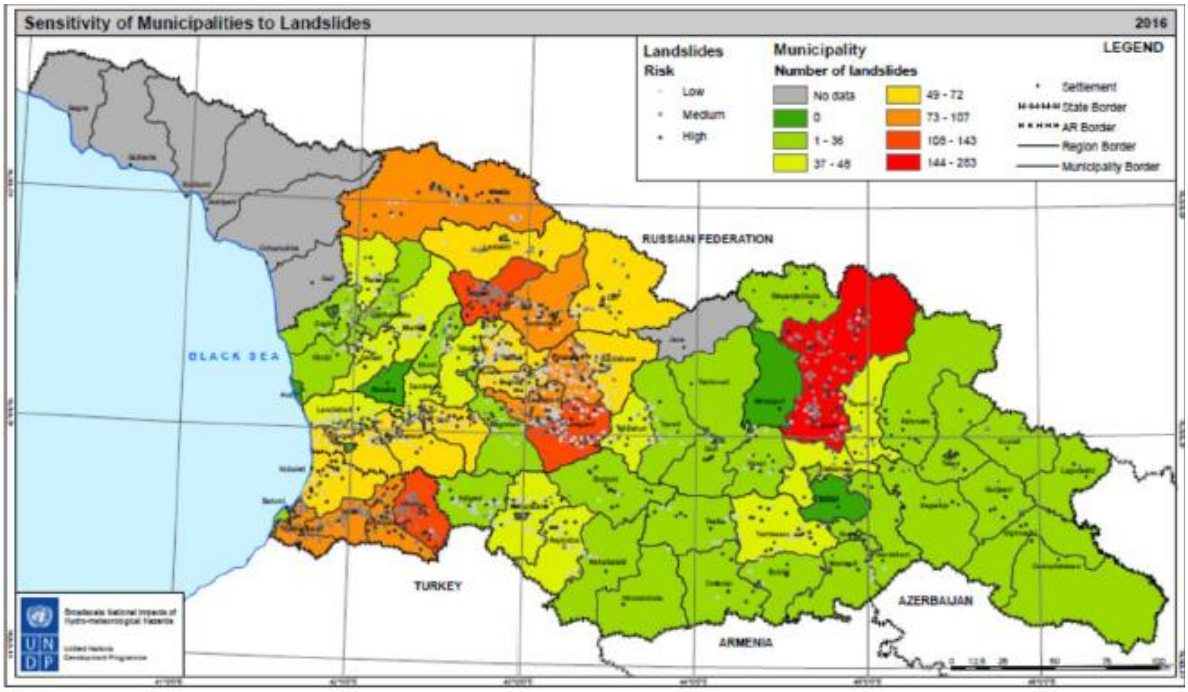


Figure 8 Municipality scenario for landslides (2016)

Source: Annex II of Feasibility Study (2017)

Figure 9 Major river basin groups

Source: Annex II of Feasibility Study (2017)

APPENDIX II

Table 4 Agenda of Georgia LORTA field mission

DAY/TIME	ACTIVITIES	PARTICIPATION
Day 0: Sunday, November 11	Arrival of team	LORTA team
Day 1: Monday, November 12 Morning	Introductions, overview and discussion of objectives	UNDP Energy Efficiency Unit and NDA
Afternoon From 15:00.	UNDP/ Project Activities Plans, Stocktaking of documents	UNDP Energy Efficiency Unit
	Project Presentation (by N. Antadze, M. Ayoung) – present on the provisional timeline, specific activities and who is responsible for each activity	UNDP Energy Efficiency Unit
Day 2: Tuesday, November 13 Capacity-Building workshop From 10:00	Session 1: Project Presentation (15 min) Session 2: Benefits of Impact Evaluation (1h) Session 3: Impact Evaluation Methods (1h 45 min)	By LORTA team Participants are MoEPA, NEA, EIEC and UNDP EE
From 14:00 Workshop on the theory of change	Theory of Change workshop (interactive workshop and discussion)	By LORTA team Participants are MoEPA, NEA, EIEC and UNDP EE
From 17:30	Discussions on the best IE design in alignment with the project implementation	By LORTA team Participants are MoEPA, NEA, EIEC and UNDP EE
Day 3: Wednesday, November 14 Morning	Meeting in Ianeti agroforestry site / Meeting with representatives from Samtredia municipality / Visit flood defence structure and agroforestry site in Sajavakho	LORTA team, UNDP, local community and municipal representatives
Afternoon	Meeting with Tskhaltubo municipality / Visit of meteorological station in Kutaisi	LORTA team, UNDP, NEA, West Georgia Branch
Day 4: Thursday, November 15 Morning	Tsageri municipality / Lentekhi municipality / Visit flood defence structures on Kheledura and Tskhenistskali rivers	LORTA team, UNDP, Tsageri municipality infrastructure unit, Lentekhi municipality, local community representatives
Afternoon	Internal meeting to discuss findings of the mission, deciding on best IE design and practical aspects for its implementation	LORTA and UNDP teams
Day 5: Friday, November 16 10:30	Presentation of possible IE designs by the LORTA team, continuing discussion on the main elements of IE design, procurement options for baseline data	LORTA team, UNDP
Afternoon From 14:30	Final debriefing: Presentation to the project team; key elements and timeline of IE design; Agreement on the next steps	LORTA team, UNDP, NEA, MoEPA

Table 5 List of stakeholders engaged with during the LORTA mission

LORTA WORKSHOP AND MEETINGS, TBILISI		
NAME	POSITION	INSTITUTION
Tuya Altangerel	Deputy Resident Representative	UNDP Georgia
Nino Antadze	EE Team Leader	UNDP Georgia
Nestan Khuntsaria	EE Portfolio Associate	UNDP Georgia
Khatuna Chanukvadze	M&E Specialist	UNDP Georgia
Salome Lomadze	SDC Project Manager	SDC
Margaretta Ayoung	Chief Technical Adviser	Independent
Nino Tandilashvili	Deputy Minister	MoEPA
Maia Tskhvaradze	Chief Specialist, Climate Change Department	MoEPA
Nino Tkhilava	Ministerial Employee	MoEPA
Tereza Varejkova	Research Manager	C4ED
Vano Tsiklauri	Former CTA for AF Floods Project	UNDP Georgia
Gia Tsagareishvili	Deputy Head	NEA
Giorgi Kordzakhia	Deputy Head of Hydro-Meteorological Department	NEA
Ramaz Chitanava	Head of Hydro-Meteorological Department	NEA
Babatunde Abidoye	Senior Economist	UNDP
Irakli Megrelidze	Deputy Head of Hydro-Meteorological Department	NEA
Merab Gaprindashvili	Head of Geological Department	NEA
Tariel Bedidze	Deputy Head of Hydro-Meteorological Department	NEA
Elene Didebulidze	Adviser to the Head	EIEC
Solomon Asfaw	Principal Evaluation Officer	IEU GCF
FIELD VISITS		
NAME	POSITION	INSTITUTION
Paata Kokhreidze	Ianeti community representative	Ianeti community
David Bakhtadze	Head of Division, former head of infrastructure unit	Samtredia Municipality
Aleko Dadunashvili	Deputy Gamgebeli	Tskhaltubo Municipality
Avtandil Qvachakidze	Representative	NEA, West Georgia Branch
Iuri Dartsuliani	Representative	Tsageri Municipality Infrastructure Unit
Germane Qurasbediani	Representative	Lentekhi Municipality Economic Development Unit

Table 6 Power calculations for the randomized control trial

ICC	# CLUSTERS C	# CLUSTERS T	TOTAL SAMPLE SIZE	R^2	# HOUSEHOLDS /CLUSTER	MDES
10%	30	30	1,500	30%	25	0.223
10%	30	30	1,500	0%	25	0.267
10%	100	100	5,000	30%	25	0.122
10%	100	100	5,000	0%	25	0.146
20%	30	30	1,500	30%	25	0.292
20%	30	30	1,500	0%	25	0.349
20%	100	100	5,000	30%	25	0.160
20%	100	100	5,000	0%	25	0.191
10%	15	15	1,500	30%	25	0.316
10%	15	15	1,500	0%	25	0.378
10%	50	50	5,000	30%	25	0.173
10%	50	50	5,000	0%	25	0.207
20%	15	15	1,500	30%	25	0.413
20%	15	15	1,500	0%	25	0.493
20%	50	50	5,000	30%	25	0.226
20%	50	50	5,000	0%	25	0.270

Table 7 Power calculations for the DiD with PSM

ICC	# SITES C	# SITES T	TOTAL SAMPLE SIZE	ALLOCATION RATIO	R^2	# HOUSEHOLDS/ SITE	MDES
5%	8	13	1,680	0.62	30%	80	0.262
5%	8	13	3,360	0.62	30%	160	0.294
5%	13	13	2,080	0.50	30%	80	0.229
5%	13	13	4,160	0.50	30%	160	0.257
2.5%	8	13	1,680	0.62	30%	80	0.203
2.5%	8	13	3,360	0.62	30%	160	0.216
2.5%	13	13	2,080	0.50	30%	80	0.177
2.5%	13	13	4,160	0.50	30%	160	0.189
0%	8	13	1,680	0.62	30%	80	0.118
0%	8	13	3,360	0.62	30%	160	0.083
0%	13	13	2,080	0.50	30%	80	0.103
0%	13	13	4,160	0.50	30%	160	0.073

Independent Evaluation Unit
Green Climate Fund
175, Art center-daero, Yeonsu-gu,
Incheon 22004, Republic of Korea
Tel. (+82) 032-458-6428
ieu@gcfund.org
https://ieu.greenclimate.fund

GREEN CLIMATE FUND | Independent Evaluation Unit

Center for Evaluation and Development