UNIVERSITÀ DEGLI STUDI DI MILANO

DOTTORATO DI RICERCA IN AGRICOLTURA, AMBIENTE E BIOENERGIA

DIPARTIMENTO DI SCIENZE AGRARIE E AMBIENTALI - PRODUZIONE, TERRITORIO, AGROENERGIA

PARTICIPATORY EVALUATION OF AGRICULTURAL DEVELOPMENT NEEDS

AS A TOOL FOR IMPROVED SUSTAINABILITY IN DEVELOPMENT COOPERATION.

AGR/01

Pietro De Marinis

TUTOR: Prof. Guido Sali

COORDINATORE DEL DOTTORATO: Prof. Daniele Bassi

PhD course year: 3rd   Cycle: 32°

2019/2020

ACRONYMS AND ABBREVIATIONS

ACP - Augmenter la conservabilité des produits agricoles or improving the shelf life of agricultural products

ADB – African Development Bank

AFS - Augmenter la fertilité des sols or improving soil fertility

AHP – Analytic Hierarchy Process

a.k.a. – also known as

AM - Accès au marché or access to market.

ARDST – Appui au Retour des Réfugiés et Déplacés par le biais de la Sécurisation des Terres (project)

ARP - Augmenter la rentabilité des produits agricoles or improving the rentability of agricultural products

A.S. – Agricultural System

AVG – Average

BOND - Bond is the UK network for organisations working in international development.

CAP – Common Agricultural Policy

CI – Consistency Index

CIAT – International Centre for Tropical Agriculture

CIRGL – Conférence Internationale de la Région des Grands Lacs africains

CMEF – Common Monitoring and Evaluation Framework

CMOC – Context-Mechanism-Outputs Configuration

CR – Consistency ratio

DAC – Development Assistance Committee

DC – Development Cooperation

DEC - Durabilité économique or economic sustainability

DEM – Digital Elevation Model

DEN - Durabilité environnementale or environmental sustainability

DES - Durabilité sociale or social sustainability; RDI - Réduire la dépendance de l'extérieur or reducing external dependencies

DFID - Department for International Development (UK)

DGs – Directorates General (CE)

DISAA – Department of Agri-Environmental Science

DMs – Decision Makers

DP - Diversifier la production or diversifying production; APC - Augmenter le rendement des cultures or improving crops yield

DRC – Democratic Republic of Congo

DSA – Durabilité du système Agricole

DSS – Decision Support System

EC – European Commission

EDP – Externally Displaced People

ESA – European Space Agency

EU – European Union

FAO – Food and Agriculture Organisation

FTA - Formation et appui Technique Agricole or technical agricultural training

GD – Goma Diocese

GDP - Gross Domestic Product

GIS – Geographic Information System

GMM – Geometric Mean Method

ICGLR - International Conference of Great Lakes Region

IDP – Internally Displaced People

IEG-WB – Independent Evaluation Group of the World Bank

IFAD – International Fund for Agricultural Development

ILAC - Institutional Learning and Change Initiative of the CGIAR Centre Bioversity International

IPES – International Panel of Experts on Sustainable Food System

ISA - Introduction des Semences Améliorées (et connaissances) or introduction of improved seeds (and knowledge)

ISBP - Intervention de Structuration de la Base Productive or interventions in farmers' organization building

IT / ICT – Information Technologies / Information and Communication Technology

ITC - University of Twente. Faculty of Geo-Information Science and Earth Observation, Netherlands.

JRC – Joint Research Centre (EU)

LatCrit - 'LatCrit theory' is a relatively recent genre of critical 'outsider jurisprudence' - a category of contemporary scholarship including critical legal studies, feminist legal theory, critical race theory, critical race feminism, Asian American legal scholarship and queer theory.

LE - Lutter contre l'érosion des sols or preventing soil erosion

LGBTQ - Lesbian, Gay, Bisexual, Transgender, and Queer (or Questioning)

M&E – Monitoring and Evaluation

MCA – Multi-Criteria Analysis

MCDM / MCDA – Multi-Criteria Decision Making / Multi Criteria Decision Aiding

MCE – Multi-Criteria Evaluation

MODM – Multi-Objectives Decision Making

MONUSCO - Mission de l'Organisation des Nations Unies pour la stabilisation en RD Congo

NASA - National Aeronautics and Space Administration

NGO – Non Governmental Organisation

NIR – Near Infra-Red (radiation)

NK – Nord Kivu

OCHA - United Nations Office for the Coordination of Humanitarian Affairs

3

ODA – Official Development Assistance

OECD-DAC - Organization for Economic Co-operation and Development

PAHP – Participatory Analytic Hierarchy Process

PCM - Pairwise Comparison Matrix

PP – Public Policy

PriEsT – Priority Estimation Tool

PWR – Pair-Wise Ranking

R&D – Research and Development

RBM – Result Based Management

RCT – Randomized Control Trial

ROARS – Reliable, Objective, Available, Realistic, Specific

RS – Remote Sensing

RUAC - Réduire l'utilisation des agrochimiques or reducing the use of agrochemicals

SAM – Spectral Angle Mapper

SDGs – Sustainable Development Goals

SDSS – Spatial Decision Support System

SES – Socio Ecological System

SMART - Specific, Measurable, Achievable, Realistic, Time-bound

SPOT - Satellite Pour l'Observation de la Terre

SPWR – Simplified Pair-Wise Ranking

ToC / TOC – Theory of Change

TPA - Transformation des Produits Agricoles or processing of agricultural products

UN - United Nations

UNDESA – United Nations Department of Economic and Social Affairs

UNDP - United Nations Development Programme

UNDP-IEO - United Nations Development Programme - Independent Evaluation Office

UNEG/UNDP – United Nations Evaluation Group of the UNDP

UNESCO - United Nations Educational, Scientific and Cultural Organization

UNHCR - United Nations High Commissioner for Refugees

US / USA – United States of America

USAID – United States Aid for International Development

WHO – World Health Organisations

ABSTRACT

'*Every human conduct that is not totally impulsive or just routine mechanics seems to be guided by an evaluation process*' (Dewey, 1939). Switching away from this individual definition of evaluation, the thesis acknowledges the definition given by OECD-DAC and focus the experimentation of well-known multicriteria decision aiding methodologies for ex-ante assessment in a specific development cooperation case study.

Chapters 1 and 2, resume the structured bibliographic that has been conducted in order to identify an updated sample of literature about evaluation in development cooperation. During the evolution of the sector, several approaches and methodologies have been conceived and are continuously developing in order to match the variety of *evaluanda* and contexts. So, three main approaches can be defined, still keeping in mind that they are interlaced in their mutual evolution: the "Positivist-Experimental", the "Pragmatist-Quality" and the "Constructive-Social Process" approaches group the contributions of several authors, ranging from the more theoretical philosophical basis to the development of specific methodologies and practical techniques. The reality of human development today, as defined by UN through SDGs, requests a multifaceted approach that is capable of mixing the existing approaches, methodologies and techniques. Pluralism of values and focus on the real use of evaluation results through participatory knowledge management are the key elements of the "Constructivist-Social Process approach" which is gaining attention in development cooperation evaluation systems. The literature review took in consideration the state of art of governmental and non-governmental organizations for worldwide and brought to the conclusion that ESs are nowadays in the middle of a renovation process which is far from being concluded.

For this reason, we worked in order to provide new insights about sound methodologies able to integrate the different paradigms and methodologies in order to keep methodological robustness and allow a transformative evaluation process to take place in a specific case study.

Chapter 3 frames our case study that is located in the Diocese of Goma, Nord Kivu, Democratic Republic of Congo. The framework for this case study is provided by a 3-year EU-funded project called ARDST 'Appui au retour de réfugiés et déplacés par le biais de la sécurisation de terres en Diocèse de Goma', led by Caritas Development Goma NGO. The project started in February 2016 and ended in June 2019. In this context our research group was asked to lead a strategic sectoral evaluation. The aim of the evaluation was to inform the choice of interventions meant to trigger sustainable agricultural development in the Diocese of Goma.

Chapters 4 to 6 report the pre-print version of three scientific papers issued from the research. Chapter 4 deals with the sketching of a spatial decision support system aimed at Multi-Criteria Evaluation of potential pilot sites for agricultural development and refugees' resettlement. Land disputes are considered both key sources and perpetuating factors of conflict in the eastern Democratic Republic of the Congo. Existing literature demonstrates that remote sensing is a useful tool for systematically monitor the spatial-temporal land use/land cover dynamics in many regions of the world. For this reason, in this Chapter we propose a methodology for the integration of different

sources of information, namely satellite imagery and census information, in order to set up multi-method approach for evidence-based decision-making in development cooperation.

Chapter 5 deals with the use of a simplified multicriteria decision-making methodology and namely the application of a simplified form of pairwise ranking (SPWR). International Aid initiatives involve complex real-world decision-making problems in all their relevant phases. Agri-environmental and generally landscape-scale issues are the typical target of participative decision-making procedures as they involve community resources planning. To successfully solve the complex real-world problems multi-criteria group decision-making approaches are recognized as reliable and effective. Main research questions are : What are the priorities for intervention in order to achieve sustainable development of the agricultural sector in the case study? Is SPWR a sound technique to prioritize different alternatives for intervention while improving group consensus in real-world complex cases? SPWR was found to be a useful technique for eliciting discussion among a multidisciplinary group for pointing out and discuss inconsistencies in decision makers' preferences. Moreover, the final group ranking matches the existing guidelines for sustainable agricultural development in the region demonstrating that SPWR is a sound technique for prioritization of alternatives in the fields of agricultural development and international aid.

Chapter 6 deals with the use of a modified form of analytic hierarchy process, namely the participatory analytic hierarchy process (PAHP), as a tool for choice criteria elicitation and resource allocation in the framework of our case study. In the field of international aid the participatory approaches to assessment, research, management and budgeting have been widely studied and applied in the last decades, mostly because international aid initiatives, in all their relevant phases from planning through implementation, monitoring and evaluation, are faced with the problem of identifying initiatives that could be successfully and sustainably implemented. During our research, we were able to make use of most frequent inconsistencies in pairwise comparison matrixes in order to stimulate the debate, to adjust local preferences and to build consensus across the group. From an operational point of view, the PAHP methodology was also suited for training the project team and for identifying a shared resource allocation pattern, which matches the existing international guidelines for agricultural development in the region.

Finally, in Chapter 7, we use the results of our research to support the proposal of the wider use of the experimented evaluation approaches and methodologies as tools for fostering the diffusion of evaluation culture in development cooperation, which in turns is a key objective to foster the impact of interventions aiming sustainability. In fact, in development cooperation, the dissemination at all levels of a culture of evaluation and the construction of solid, adaptive and inclusive feedback and decision-making systems are seen as possible solutions to the everlasting doubt about development cooperation impact. In our research, we experimented a methodological approach and some specific multicriteria participatory techniques that should be taken into account when trying to link in a completely interconnected chain, the cooperation results on different scales and to evaluate, both ex-ante and ex-post, its overall contribution towards the SDGs.

1   INTRODUCTION

1.1   Definition of evaluation

During the following discussion we will make frequent reference to the concepts of 'evaluation', 'public policy', 'development interventions', 'sustainability', 'socio-ecological system', 'agroecology', 'theory of change', 'decision support systems', 'multicriteria decision-making'.

First of all, then, it is necessary to clarify the term 'evaluation'. The word in fact lends itself to very varied interpretations: in the first half of the twentieth century Dewey J., philosopher and pedagogy expert, stated that 'every human conduct that is not totally impulsive or just routine mechanics seems to be guided by an evaluation process' (Dewey, 1939). Switching away from this individual definition, other authors over time have given other interpretations and for the purposes of the present discussion, we report the definition given by OECD-DAC (2010): an evaluation is the systematic and objective assessment of an upcoming (ex-ante evaluations), ongoing (ad-interim) or completed project (final and/or ex-post), programme or policy, its design, implementation and results. In general terms, evaluation helps in understanding the worth and merit of a policy programme or project or single actions in order to improve it. The specific aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. Anyway, other evaluation criteria can be identified and used according to specific cases. As it is widely reported by evaluation providers in the field of development cooperation, evaluations should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of all stakeholders.

Conceptually it is important to note that, when dealing with ad-interim or ex-post evaluations, literature tends to prefer the term 'evaluation', although appraisal and assessment also figure in it. The 'ex-ante' evaluative action is instead normally called 'appraisal' or 'assessment', and hardly ever 'evaluation'. There is no strong semantic reason for this choice. According to the Oxford online dictionary, 'appraisal' is 'an act of assessing something or someone', while 'evaluation' is defined as 'the making of a judgement about the amount, number, or value of something'.

It might be argued that 'evaluation' includes more of a (final) judgement call, and therefore seems more naturally ex-post. However, both ex-ante appraisal and ex-post evaluation mechanisms stress that they only provide the evidence for decisions and judgement calls to be made at a more political

level afterwards. At the same time, any gathering of information, whether ex-ante or ex-post, will include some assessment and judgement on the relevance of information. From this perspective, during this dissertation, we use the concepts of evaluation and appraisal as synonymous, preferring a broad definition of evaluation and appraisal as including both ex-ante and ex-post processes. A similarly broad use of the concept of evaluation is proposed in the OECD's new Framework for Regulatory Policy Evaluation (2014). Tables 1.1 and 1.2 resume the definition of several other recurrent concepts in the evaluation practice.

Table 1.1 -Resume of the definitions of other useful concepts recurrently used in the present dissertation. From Mertens et al., 2019.

Term	Definition	Example
Evaluand	The person, program, idea, policy, product, object, performance, or any other entity that you are evaluating	Support services for students with disabilities at the local secondary school
Merit	The absolute or relative quality of the evaluand, either ♦ intrinsically or ● in regard to a particular criterion	♦ More children with disabilities are completing course requirements. ● The school is in compliance with federal standards for supporting students with disabilities.
Worth	The evaluand's value *in a particular context*	▪ Parents of students with disabilities are highly satisfied with their children's progress. ▪ Employers are happy to have students they can hire who have good skills.
Monitoring	Observing and reviewing the progress of a program over a period of time to see if it is achieving its objectives	In a program designed to teach about global historical sites, virtual reality glasses are used by the students. Are the teachers following the strategies provided in their training to incorporate the use of virtual headsets in the classroom?
Evaluation	"Evaluation is a *profession* that uses *formal methodologies* to *provide useful empirical evidence* about *public entities* (such as programs, products, performance) *in decision-making contexts* that are inherently *political* and involve multiple often-conflicting *stakeholders,* where resources are seldom sufficient, and where *time-pressures are* salient" (Trochim, 1998, p. 248, emphasis added).	In the virtual reality evaluation, to what extent have the students gained in knowledge about the global historical sites after using the virtual headsets to independently tour those sites?

Table 1.2 – Examples of Evaluands. From Mertens et al., 2019.

- **Program:** The Native American community in North Carolina wanted a program to reduce the number of teenagers who smoked tobacco. The Not on Tobacco intervention was designed to help American Indian teens stop smoking (Horn, McCracken, Dino, & Brayboy, 2008).

- **Project:** The W. K. Kellogg Foundation (WKKF) funded a project to determine how to improve the accessibility of courts for deaf and hard-of-hearing people across the United States (Mertens, 2000).

- **Policy:** The Los Angeles Unified School District wanted an evaluation of its Title I[1] Achieving Schools initiative, which was implemented in the K–12 school system to support the use of promising school-level practices in elementary schools in high-poverty areas (Barela, 2008).

- **Product:** An evaluation proved that portable music players, such as Apple's iPod, do not interfere with cardiac pacemakers (BioMed Central, 2008).

- **Idea:** The United Nations Foundation–Vodafone Foundation Partnership (2009) evaluated the idea of improving the health of people in the developing world by sending them crucial health information through mobile technology, such as smartphones, personal digital assistants (PDAs), laptops, and tablet PCs.

- **Program:** A comprehensive reading reform model called Success for All was evaluated in 35 schools over a 3-year period (Borman et al., 2007).

- **Program:** Mertens, Harris, Holmes, and Brandt (2007) conducted an evaluation of a master's program in one university to prepare teachers who are deaf or hard of hearing and/or from ethnic/racial minority groups to teach children who are deaf and who have an additional disability.

- **Program:** A midsized Midwestern city implemented a local program to reduce and prevent homelessness and chronic unemployment (Coryn, Schröter, & Hanssen, 2009).

- **Program:** A program for people with developmental disabilities was in place for 5 years and had centers in multiple locations across one state. The program was designed to provide individualized services for people with developmental disabilities in order to improve their quality of life (Fredericks, Deegan, & Carman, 2008).

- **Program:** The World Bank initiated an evaluation of a program to reduce corruption in Uganda and Tanzania (Marra, 2000).

- **Program:** Mareschal, McKee, Jackson, and Hanson (2007) conducted an evaluation of a program to reduce youth violence in four communities.

- **Project:** The Infant Feeding Research Project was evaluated to determine its effectiveness in reducing the rate of pediatric HIV/AIDS in southern Africa via interventions to enhance infant feeding by counseling the mothers of at-risk babies (Buskens & Earl, 2008).

- **Program:** Hovey, Booker, and Seligman (2007) measured the impact of a theatrical program on knowledge and attitudes about HIV/AIDS among an audience of Mexican farm workers who were at elevated risk for contracting HIV/AIDS and were in need of HIV/AIDS-related education.

It is then necessary to define policies, programmes and projects. In the following discussion we take in consideration several scales of action, ranging from the wider public policy (PP) level to the smallest local intervention, and we focus on the Development Cooperation (DC) cycle. A public policy is defined as a set of strategies and tools aimed at improving the conditions relating to a topic of collective interest. This definition is shared by many authors (Martini et al., 2009) in the field of evaluation and is based on some additional concepts: the binomial 'strategy and tools' refers to the need to evaluate the 'theory' and the 'practice' of the policy under examination, meaning its implementation process and its measurable effect. The concepts related to 'improvement' and 'collective' refer to the objective of the undertaken actions which is to increase well-being at collectivity level. We will see later how the different interpretation of these concepts has created different schools of thought and different approaches to the practice of evaluation. Interventions in

the framework of DC may be seen as examples of policy tools on a smaller scale and in a specific context. A <u>development intervention</u> is defined by fundamental concepts that are similar to the ones used to define a PP: a development intervention, which can be shaped as a simple project or a composite programme, is defined as a set of activities aimed at translating a planned strategy into actions that are supposed to have real and positive impact on the daily life of the concerned subjects, a.k.a. stakeholders.

Evaluating a public policy or intervention, according to the most influential US organizations in the sector[1], means expressing an opinion on its implementation process and its effects. In order to be able to do this, a systematic gathering of information is necessary, a term of comparison is needed and normally it is necessary to set an improvement goal concerning the policy itself and its ability to modify the reality on which the policy insists.

The variability of possible contexts and sectors in which the evaluative action may be required determines the need to adapt to the cognitive needs from time to time, but still the two main questions are: 'how is a policy/intervention implemented?' and 'what are the consequences, in terms of changes in the status quo, brought forward by the policy/intervention as it is applied?' These two questions well cover the itinerary that goes from checking the consistency of the choice of an intervention to be carried out to the judgement on the validity and dimension of its impacts (bringing forward consequent considerations towards its replication, modification, or elimination). This exposes equally well the cognitive space which hosts the evaluation process: between what is written in the strategy (policy, programme or project) and its effects on reality, there are the actors who conceived it, those who implement it, those who receive it and all the individual and collective modes of action that can exist given a place and a time of application. Clearly, this reasoning leads to recognizing the need to separately evaluate the process of implementing a policy / intervention and the effects it has on the state of things. Equally clearly, we realize that the context in which a policy / intervention is carried out, shapes the final result (effect) depending on the interaction between the policy / intervention and the context itself.

---

[1] MDRC, New York; Mathematica Policy Research, Princeton; ABT Associates, Boston; Rand Corporation, Los Angeles; Urban Institute, Washington.

1.2   Other useful definitions

Here, it seems important to clarify the concept of causal linkages between a policy/intervention and its supposed impacts. Since the middle of '90, the leading approach for mapping these casual linkages is known as <u>Theory of Change (ToC).</u> The evolution in the field of evaluations that leads to the use of the theory of change can be reconstructed starting from the late 1950s when Kirkpatrick theorized the model of evaluation of the four levels of learning (Kirkpatrick, 1954). Further important steps in this evolution are the use of the CIPP model (context, inputs, processes and products) by Stufflebeam (1971) and its widely used logical structures or models (logical frameworks or logframes) that define causal chains usually consisting of objectives, inputs, activities, results and long-term impacts. Methods of analysis and conceptualization of the hierarchy of development policies / interventions such as logframes represented significant progress, providing a framework through which the relationships between the components of a programme could be traced and articulated. However, American authors such as Weiss, Chen and Patton have increasingly highlighted the challenges brought about by the increasingly complex programmes themselves and by contexts of intervention (Mertens et al., 2019). The book 'New approaches to evaluate complex community initiatives', by Weiss (1998), is perhaps the key moment after which the use of the concept of the ToC has spread more widely. The ToC can be therefore essentially defined as a comprehensive description and illustration of how and why a desired change is expected to happen in a particular context. It is focused in particular on mapping out or 'filling in' what has been described as the 'missing middle' between what a programme or initiative does (its activities or interventions) and how these lead to desired goals being achieved (Centre for Theory of Change, 2019). At this point it is mandatory to define the <u>chain of results</u> that constitutes the backbone of the theory of change: a chain of results is the logical and physical set of causally interconnected inputs, actions, outputs, outcomes and impacts of a project, programme or policy. This chain can be divided in the implementation phase (input and actions) and the results phase (outputs, outcomes and impacts). Globally, the chain of results explains through which resources, activities, expected results the programme seek to improve or change reality.

Nowadays, the 'desired change' or 'improvement goal' that underlies both the public policy being evaluated and the evaluation itself converge on the concept of sustainability (Kareiva et al., 2012) that lends itself to becoming the ultimate goal of all reflections on development initiatives. If considered within the workspace of the evaluation of development policies or interventions,

sustainability is simply 'the permanence of the benefits on the recipients and on the territory even after the conclusion of the programme' (Bezzi, 2010). Translating this definition onto the wider earth system, understood as we will see later in the sense of a socio-ecological unity, the sustainability of the human presence on the planet, which is often referred to by the term 'sustainable development' is defined by the FAO in 1989 as 'the conservative management of the quality and quantity of existing natural resources and the consequent orientation of institutional policies and technologies so as to guarantee the satisfaction of human needs for present and future generations' (FAO, 2010). However, the definition of sustainable development requires clarification about the term 'development'. If it literally means transformation of a state or characteristic over time, in the following discussion it is understood in terms of 'human development'. This concept was developed by the economist Mahbub Ul Haq (Haq et al., 1986) and finds its conceptual basis in the work of Amartya Sen on the human abilities of 'being' and 'doing' desirable things in the course of one's life (Sen, 1999 and 2011). So, the concepts of sustainability and human development begin their wider diffusion in the 1990s some years after the publication of the first United Nations Report on Human Development which had the objective of overcoming the use of economic development as an indicator of the evolution of the quality of human systems on the planet and therefore of human well-being (Brundtland Commission, 1987; UNDP, 1990, 2013, 2014). In this perspective, human development is to be understood as the goal of the collective work of human beings on planet Earth to be achieved through the widening of the possible options of choice, an improvement in living conditions, intellectual capacities and useful practices for determining this higher number of possible future choices. This concept has remained central in the debate about the evaluation of public policies since they, taken as a whole, have the objective of obtaining or maintaining a desired state of human development, understood in its elementary components of economic, ecological and social possibilities (UNDP, 1990). Concatenating the concepts described so far and bringing them back into the framework of the evaluation of public policies and development interventions, it is finally necessary to define the scale, or rather the range of objects and scales on which this evaluative research can be performed. Leaving to the following paragraphs the analysis of the dimensions and targets to which / on which the evaluation can work, we introduce here the concept of socio-ecological system understood as a unitary system composed of people and nature, in turn understood as an indivisible unit of which the human being is an integral part (Berkes et al., 1998 and 2003). It is a coherent, complex and adaptive system composed of biophysical and social factors that interact regularly on different spatial, temporal and organizational scales, more or less hierarchically

structured. A socio-ecological system identifies a set of interconnected strategic resources (natural, socio-economic and cultural) whose flow is regulated by its own dynamics (Redman et al., 2004). In this light it is possible to affirm that any existing resource on the planet and potentially usable by man is integrated in a complex socio-ecological system (SES). Thus, the very communicative image used by E. Ostrom takes shape, describing a SES as an organism composed of a multi-level series of subsystems and variables internal to each subsystem, similarly to how a living organism is composed of organs, tissues, cells, macromolecules and so on (Ostrom, 2009).

Dealing with agriculture we need to define the use of agricultural system and agricultural sector. Given the SES standpoint we keep during the dissertation, the agricultural systems is defined by the complex network of economic, social and environmental relationships existing in rural areas and that is linked to urban areas through the food system. In this framework the agricultural sector can be considered a sub-system of the agricultural system, referring to the economic and technical issues emerging from the management of agricultural activities in the framework of the existing agricultural system.

Finally, when dealing with agricultural systems, the approach to development that internalize the concepts of sustainability and socio-ecological approach may be called 'agroecological approach'. Agroecology is defined as a complex approach: it contains ethical values, working methodologies and practical and multidisciplinary techniques that follow the ecological dynamics and a participatory approach. Agroecology tends to plan and use existing resources for the good sake of improving human living conditions in the full respect of sustainability (Altieri, 1987; Wezel et al., 2009; OECD, 2019).

Paradigm will be another recurrent word in the present dissertation. According to Guba (1989) and Denzin et al. (2005) paradigms are broad metaphysical constructs that include sets of logically related philosophical assumptions. Different paradigms can be defined by four sets of philosophical assumptions as it is shown in Table 1.3. As we will explain in chapter 2, this definition is important because the different approaches to evaluation refer precisely to these different sets of philosophical assumptions.

Table 1.3 - Four sets of Philosophical Assumptions in Paradigms. From Mertens et al., 2019.

Philosophical assumption	Guiding question	As experienced in life
Axiology	What is the nature of ethics?	Your societies, cultures, and religion taught you morals of what is right and wrong. You developed your own values, which are the principles or ideas that you believe in and are willing to support and defend. Both guide your ethics in how you conduct your lives. Lisa, Martina, and Kent all morally object to murder. Lisa values justice, so she ethically supports capital punishment. Martina values a person's right to self-defense, so she ethically denies capital punishment. Kent values equal rights of appropriate social services for all, so he ethically rejects capital punishment.
Ontology	What is the nature of reality?	Is there one reality that I can discover? Or are there multiple realities that differ, depending on the experiences and conditions of the people in a specific context? Lisa's brother, a prison guard, tells her stories of the awful behavior of the prisoners. Martina's brother received addiction rehabilitation while in prison and now has a job and family. Kent has seen documentaries and read articles about the cruel treatment of prisoners in mass incarceration in the United States. Whose reality is real?
Epistemology	What is the nature of knowledge, and what is the relationship between the knower and that which would be known?	Another wording for the epistemological question might be: How should the evaluator relate to the stakeholders? Do you, as the evaluator, objectively stand apart from the stakeholders, or do you engage with them in deep conversation and in their activities? How could you learn what constitutes reality in the case of a woman murdering her boyfriend?
Methodology	What are the systematic approaches to gathering information about what would be known?	Do you need to compare two groups, or can you document progress by intensively studying one group? Should you use a quantitative, a qualitative, or a mixed methods approach? Do you send surveys to the prison, go interview the woman, and interview pertinent workers at social service agencies? Do you collect data about domestic violence and gather information about the local and state laws?

1.3   Scope of the PhD thesis: main research questions

The present dissertation focuses on ex-ante evaluation research (also called 'appraisal') applied to development interventions in the agricultural sector. We aim at experimenting the use of a constructivist/transformative evaluation approach and participatory methodologies as a useful tool for widening the diffusion of the so-called 'evaluation culture' in the DC sector. In order to reach the goal we focus on a specific case study concerning agricultural development and the choice of priority interventions for improving sustainability in a complex context. Therefore the present dissertation describes the path composed by the mixed-method evaluation and the participatory decision-making process with the aim of:

- Outlining the potential effect of the use of such methodologies at the specific stage of ex-ante evaluation;
- Inferring the potential effect of the use of such an approach on the other stages of the DC cycle and evaluation management (ex-ante, ad-interim, final, ex-post).

Consequently, our study begins with a general review about evaluation theory and practice, useful to frame the evaluative action in respect of the existing different evaluation paradigms. The dissertation goes on outlining evaluative research domains and focusing on decision support systems used in development cooperation initiatives. After drawing the state of the art, the research turns operational and tackles some specific questions about suitable ways of approaching the ex-ante sectorial and strategic evaluation with the aim of identifying relevant interventions for the sustainable development of the agricultural sector in a specific area. Then, the results of the experience are used to propose some potential effect of the use of transformative and participatory approaches to other types of evaluation.

The dissertation is therefore articulated on the following main research questions:

- QUESTION A) What is evaluation and what does it mean to evaluate a project, a programme or whatever?
- QUESTION B) What are the existing approaches, models, methods and techniques of evaluation and to what objectives are they addressed to?
- QUESTION C) How does the evaluation system work in the DC sector? What is and why is 'evaluation culture' so important to improve the DC system?
- QUESTION D) How a constructivist and participatory approach towards evaluation and decision-making can be applied to the field of development cooperation?
- QUESTION E) How to run an ex-ante strategic sectorial evaluation with participatory approach in order to improve the sustainability of the agricultural system in NK?
- QUESTION F) What would be the advantages of using a transformative and participatory approach in other types of evaluations (ad-interim, final, ex-post)?

1.4 The structure of the thesis

This PhD thesis is organized in eight chapters: the discussion begins (present chapter 1) with the definition of several concepts used in the present dissertation and goes on giving a first-hand answer to research question A.

In the second chapter, the state of the art of the existing evaluation theories and methods is drawn, completing the answer to question A and answering research questions B and C.

The third chapter represents the cornerstone of the dissertation. It contains the description of one case study and contains the explanation of how a constructive evaluation approach can inform decision-making in the field of development cooperation, so answering research question D.

Here, we describe the whole-research workflow and we present graphically how the chapters 4, 5 and 6 contribute to the detailed description of specific steps in the overall workflow. Given the importance of the third chapter in order to understand the sense of the whole thesis, we included a brief and maybe redundant description of the experimental setting and methodologies that are further detailed in the following chapters. *Repetita iuvant*.

Chapters 4, 5 and 6 individually report the pre-print version of the scientific papers issued from the implementation of the mixed method evaluation-based and participatory decision-making process, therefore answering research question E. A short seventh chapter contains the conclusion, answer to question F, and hints at further research perspectives. Chapter 8 contains the acknowledgments.

1.5 References

Altieri, M. A. (1987). *Agroecology: the scientific basis of alternative agriculture*. Boulder: Westview Press.

Barela, E. (2008). Title I Achieving Schools Study. *American Journal of Evaluation*, *29*(4), 531–533. https://doi.org/10.1177/1098214008328405

Berkes, F., Folke, C., & Colding, J. (1998). *Linking social and ecological systems : management practices and social mechanisms for building resilience.* (F. Berkes, C. Folke, & J. Colding, Eds.). Cambridge: Cambridge University Press.

Berkes, F., Colding, J., & Folke, C. (2002). *Navigating Social-Ecological Systems: Building Resilience for Complexity and Change.* (F. Berkes, J. Colding, & C. Folke, Eds.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511541957

Bezzi, C. (2010). *Il nuovo disegno della ricerca valutativa*. Milan: Franco Angeli.

BioMed Central. (2008). iPods do not interfere with cardiac pacemakers, study shows. Retrieved November 27, 2019, from www.sciencedaily.com/releases/2008/01/080131214602.htm

Borman, G. D., Slavin, R. E., Cheung, A. C. K., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final Reading Outcomes of the National Randomized Field Trial of Success for All. *American Educational Research Journal*, *44*(3), 701–731. https://doi.org/10.3102/0002831207306743

Bruntland Commission. (1987). *Our Common Future*. (V. Hauff, Ed.). Oxford: Oxford University Press.

Buskens, I., & Earl, S. (2008). Research for change: Outcome mapping's contribution to emancipatory action research in Africa. *Action Research*, *6*(2), 171–192. https://doi.org/10.1177/1476750307087051

Center for Theory of Change. (2019). What-is-theory-of-change. Retrieved September 10, 2019, from https://www.theoryofchange.org/what-is-theory-of-change/

Coryn, C. L. S., Schröter, E. C., & Hanssen, C. E. (2009). Adding a time-series design element to the success case method to improve methodological rigor: An application for non-profit program evaluation. *American Journal of Evaluation 30*, 80–92.

Dewey, J. (1939). *Theory of Valuation. International encyclopedia of unified science* (Vol. 2). Chicago: The University of Chicago Press. https://doi.org/10.1086/286595

FAO. (2010). What is meant by the term 'sustainability'? Retrieved November 11, 2019, from http://www.fao.org/3/ai388e/AI388E05.htm

Fredericks, K. A., Deegan, M., & Carman, J. G. (2008). Using System Dynamics as an Evaluation Tool: Experience From a Demonstration Program. *American Journal of Evaluation*, *29*(3), 251–267. https://doi.org/10.1177/1098214008319446

Guba, E. G. (1989). *Fourth generation evaluation*. (Y. S. Lincoln, Ed.). Newbury Park: Sage Publications.

Denzin, N. K., & Lincoln, Y. S. (2005). *The SAGE handbook of qualitative research*. Thousand Oaks: Sage Publications.

Haq, K., & Kirdar, U. (1986). *Human development: The neglected dimension*. Islamabad: North South Roundtable, UNDP.

Horn, K., McCracken, L., Dino, G., & Brayboy, M. (2006). Applying Community-Based Participatory Research Principles to the Development of a Smoking-Cessation Program for American Indian Teens: 'Telling Our Story.' *Health Education & Behavior*, *35*(1), 44–69. https://doi.org/10.1177/1090198105285372

Kareiva, P., & Marvier, M. (2012). What Is Conservation Science? *BioScience*, *62*(11), 962–969. https://doi.org/10.1525/bio.2012.62.11.5

Kirkpatrick, D. L. (1954). *Evaluating Human Relations Programs for Industrial Foremen and Supervisors*. Madison: Kirkpatrick Publishing.

Mareschal, P. M., McKee, W. L., Jackson, S. E., & Hanson, K. L. (2007). Technology-Based Approaches to Preventing Youth Violence: A Formative Evaluation of Program Development and Implementation in Four Communities. *Youth Violence and Juvenile Justice*, *5*(2), 168–187. https://doi.org/10.1177/1541204006295148

Marra, M. (2000). How Much Does Evaluation Matter? Some Examples of the Utilization of the Evaluation of the World Bank's Anti-Corruption Activities. *Evaluation*, *6*(1), 22–36. https://doi.org/10.1177/13563890022209091

Martini, A., & Sisti, M. (2009). *Valutare il successo delle politiche pubbliche*. Bologna: Il Mulino.

Mertens, D. M., & Wilson, A. T. (2019). *Program evaluation theory and practice*. New York: The Guildford Press.

OECD. (2014). *Framework for Regulatory Policy Evaluation*. Paris: OECD Publishing. https://doi.org/10.1787/9789264214453-en

OECD. (2019). Agroecology definition. Retrieved September 10, 2019, from https://stats.oecd.org/glossary/detail.asp?ID=81

OECD-DAC. (2010). *Glossary of Evaluation and Results Based Management (RBM) Terms*. *Working Party on Aid Evaluation, OECD*. Paris Cedex: OECD Pubblications.

Ostrom Elinor. (2009). A General Framework for Analyzing Sustainability of Social-Ecological Systems. *Science*, *325*(July), 1962–1980. https://doi.org/10.1002/2016JE005088

Redman, C. L., Grove, J. M., & Kuby, L. H. (2004). Integrating Social Science into the Long-Term Ecological Research (LTER) Network: Social Dimensions of Ecological Change and Ecological Dimensions of Social Change. *Ecosystems*, *7*(2), 161–171. https://doi.org/10.1007/s10021-003-0215-z

Sen, A. (1999). *Development as Freedom*. New York: Alfred Knopf.

Sen, A. (2011). *The idea of justice*. London: Belknap Press.

Stufflebeam, D. L. (1971). The Relevance of the Cipp Evaluation Model for Educational Accountability. In *Annual meeting of the American Association of School Administrators*. Ohio State Univ., Columbus. Evaluation Center.

Trochim, W. M. K. (1998). Donald T. Campbell and Research Design. *American Journal of Evaluation*, *19*(3), 407–409. https://doi.org/10.1177/109821409801900316

UNDP. (1990). *Human Development Report 1990*. (American Writing Corporation, Ed.). New York: Oxford University Press.

UNDP. (2013). *Human Development Report 2013 The Rise of the South: Human Progress in a Diverse World*. (C. D. Incorporated, Ed.). Kanata: Gilmore Printing Services.

UNDP. (2014). *Human Development Report 2014 Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience*. (C. D. Incorporated, Ed.). Durham: PBM Graphics.

Weiss, C. H. (1998). *Evaluation : methods for studying programs and policies* (2nd ed.). Upper Saddle River, N.J.: Prentice Hall.

Wezel, A., Bellon, S., Doré, T., Francis, C., Vallod, D., & David, C. (2009). Agroecology as a science, a movement and a practice. A review. *Agronomy for Sustainable Development*, *29*(4), 503–515. https://doi.org/10.1051/agro/2009004

2 Framing evaluation: state of the art

2.1 Evaluation theory

2.1.1 History of evaluation and social research

Evaluation, as a need coming from 'the community', can be tracked back to the 1800s, when the U.S. government first asked for external inspectors to evaluate public facilities such as prisons, schools, hospitals, and orphanages (Stufflebeam et al., 2002). However, many authors report that professional evaluation starts to exist only in the 1960s, with the booting of Lyndon Johnson's Great Society initiatives. Indeed, it is difficult to define precisely when evaluative research or evaluation profession are really born and many other authors (Rossi et al., 2004) state that, from a social science perspective, social scientists were conducting studies of major social programmes related to education, public health, and employment even before the World War I. Social research is indeed somehow the mother of evaluation as we conceive it now, because social scientist have always focused the evolution of human society under the effects of external or internal drivers. In these terms, social scientist were evaluators before evaluation was founded as discipline, and origins of social research are to be sought in ancient Arab and western philosophers (Mitchell, 1970). Anyway, social science origins can be traced back to the birth of social inquiries and with the philosophical positivist school of thoughts in the early 1950'. After that, during the 1960s, the USA War on Poverty brought to the final marriage between social research and evaluation, fuelling the need for a systematic approach to evaluating the social programmes and thus strengthening the need for specialized professionals and organisations. It is for this reason that in 1976, two professional associations related to evaluation were founded in the United States: the Evaluation Research Society and Evaluation Network.

During this early period, evaluation was essentially devoted to assessing educational innovations, resource allocation problems and antipoverty programmes.

Nowadays the two associations work together in the AEA (American Evaluation Association - www.eval.org), regrouping more than 7,000 members from all 50 U.S. states and from over 86 foreign countries. Before 1995, there were only five regional and/or national evaluation organizations in the world (Mertens, 2005). After only seven years the International Organisation for Cooperation in Evaluation (IOCE – www.ioce.net) was founded as an alliance of evaluation networks and associations

who already formed other alliances with other international organisation such as the United Nations Children's Fund (UNICEF) and several other organizations interested in improving the capacity of civil society to influence policy makers and other decision makers (Mertens et al., 2019).

This historical emphasis on the USA does not mean that evaluation was entirely absent from Europe or other parts of the world. However, it is from the 1970s onwards that evaluation began to take root in European countries such as Scandinavia, UK or France, where traditional and strong commitment to democratic governance demands for evaluation to build accountability (European Commission, 2013).

As a matter of fact, since European Structural Funds exist, the central European authority has been interested in evaluation and it is the evolution of these structural funds that determines the spreading of evaluation throughout the EU. At every stage of the programming cycle (ex-ante, ad-interim or on-going, final, ex-post), structural funds have clearly stated aims and responsibilities. It is commonly acknowledged that the introduction of evaluation into many countries in Southern Europe occurred as a result of the requirements of Structural Fund regulations. From modest beginnings in 1988, there is now an elaborated Structural Fund evaluation approach in EU.

Based on the experience of the 2000-2006 period, the major innovations of the 2007-2013 evaluation provisions were the introduction of the principle of proportionality and the encouragement by the Commission of an approach to ongoing evaluations based on the needs of Member States and regions.

For the 2014-2020 period, there has been a thorough revision of the articles concerning evaluation. There is a stronger focus on results, a much stronger emphasis on the intervention logic and on concentration as a tool for widening appropriation. Evaluation of impact has been introduced in the role of evaluation with requirements for managing authorities to undertake impact evaluations in line with an evaluation plan. Impact is no longer regarded as the long-term evolution of a statistic which may or may not have been influenced by policy, but the net contribution of the policy to observed change.

2.1.2 Paradigms and approaches to evaluation

Different approaches to evaluation exist since the beginning. They stem from different visions of which is the most effective way to influence policy making and implementation processes through evaluation.

One of the most recognized metaphor to frame the existence of different approaches to evaluation goes under the name of 'Christie and Alkin's three of evaluation'. Figure 2.1 shows a modified version with four branches (Mertens et al., 2019).

Figure 2.1 - Evaluation three according to Mertens et al., 2019.
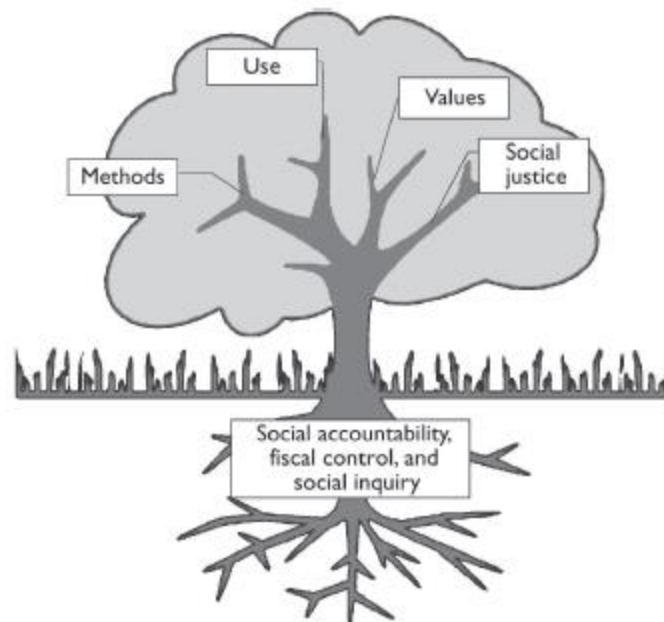
The roots of the tree represent the causes behind evaluation. The foliage is divided in four branches which represent four possible focuses in evaluation practice. These four focuses represent different dimensions of evaluation. Different paradigms attribute different relative weight to these dimensions.

Therefore four main interconnected paradigms for evaluation exist: the post-positivist, the pragmatic, the constructivist and the social-justice-oriented (or transformative). Evaluators referring to the post-positivist paradigm approach evaluation by focusing on methods. The pragmatic evaluators focus on the use of evaluation results. Supporters of the constructivist paradigm point out the diversity of values and the importance of building a consensus as key elements informing evaluation results. Social-justice-oriented (or transformative) paradigm's supporters focus the

importance of minorities' point of views and the ability of evaluation to change the very same implementing actors taking part in the evaluation processes. Table 2.1 regroups the main contributors according to each evaluation paradigm.

Table 2.1 - Regrouping main authors according to evaluation paradigms (adapted from Mertens et al., 2019).

Branch / Paradigm	Main authors		
Methods branch (Post-positivist paradigm)	Ralph Tyler, Donald Campbell, Thomas Cook, William Shadish, Robert Boruch, Peter Rossi, Gary Henry, Mel Mark, Huey-Tsyh Chen, Stuart Donaldson, Donald Kirkpatrick, Robert Brinkerhoff and others		
Use branch (Pragmatic paradigm)	Daniel Stufflebeam, Carol Weiss, Joseph Wholey, Eleanor Chelimsky, Michael Patton, Hallie Preskill, David Fetterman, Abe Wandersman, Jean King, Brad Cousins, Marvin Alkin, Coryn, Noakes, Westine, Schröter,Morrow, Nkwake, Bartholomew, Parcel, Kok and others		
Values Branch (Constructivist paradigm)	Michael Scriven, Eliot Eisner, Robert Stake, Ian Shaw, Malcolm Parlett, David Hamilton, Egon Guba, Yvonna Lincoln, Sharlene Hesse Biber and others		
Social justice branch (Transformative paradigm)	Deliberative democratic evaluation	LatCrit	Indigenous
	Barry MacDonald, Saville Kushner, Ernest House, Kenneth Howe, Jennifer Greene, Katherine Ryan	Dolores Delgado Bernal, Lilia Fernandez, Tara Yosso, Daniel Solórzano	*Maori* Fiona Cram, Linda T. Smith
	Human rights	Disability/ deaf rights	*African American/CRT* Stafford Hood, Rodney Hopson, Asa Hilliard, Aaron Brown, Leander Boykin, Reid E. Jackson, James Scheurich, Gloria Ladson-Billings, Henry Frierson, Veronica Thomas
	Thomas Schwandt, Donna Mertens, Marco Segone, Karen Kirkhart	Donna Mertens, Martin Sullivan, Carol Gill, Raychelle Harris, Heidi Holmes	
	Feminist	LGBTQ	
	Kathryn Sielbeck-Bowen Sharon, Sharon Brisolera, Cynthia Dillard, Denise Seigart, Bessa Whitmore, Saumitra Sen Gupta, Sharlene Hesse-Biber	Jeffrey Todahl, Sarah Dodd	*African* Bagele Chilisa, John Bewaji *Native American* Marie Battiste, Cheryl Crazy Bull, Joan LaFrance

Despite this apparent subdivision, the different paradigms and the relative approaches to evaluation are not statically divided and opportunities exist for intermingling of ideas. The pragmatic approach for instance was born over and share elements of both the two conflicting visions brought forward by the post-positivist and the constructivist paradigms. The pragmatic vision is intermediate: it accepts that reductionism oversimplifies complexity but at the same time refuse the possibility to

deal with unfathomable complexity carried by the systemic vision of constructivists. In the same way, the constructivist paradigm and the social-justice-oriented paradigm share the vision of evaluation as a process being part of the policy or intervention and consequently, they share the assumption that participation is a key element in evaluations.

Given this existing interconnection between the paradigms, other metaphors have been used to frame evaluation, such as the 'river estuary metaphor' or the 'ocean currents metaphor', visualising the different paradigms as intermingling streams of water. This last metaphor not only allows for intermingling of waters; it also demonstrates that many forces come into play to determine the nature and effects of different ocean currents (Mertens et al., 2019).

These different approaches initially evolve separately with different timing but soon they reach a phase of strict co-evolution.

Keeping in mind that the 4 paradigms and the relative 'evaluation schools' are not isolated boxes nor branches but instead are intermingling flows, one could affirm that the difference between the mentioned paradigms can be described as a fading gradient in which differences stems from the different mind-sets that are typical of the so-called 'hard' and 'soft' sciences. In fact, if all evaluation theorists agree on the fact that the systems which are targeted by policies or interventions are complex systems, a first subdivision of the approaches to evaluation can be drawn on different attitudes toward complexity.

Figure 2.2 illustrates how the four paradigms can be placed according to their proximity to opposite attitudes towards complexity and other important aspects of evaluation such as preferred methodologies, role of participation in evaluation and of the evaluator in society.
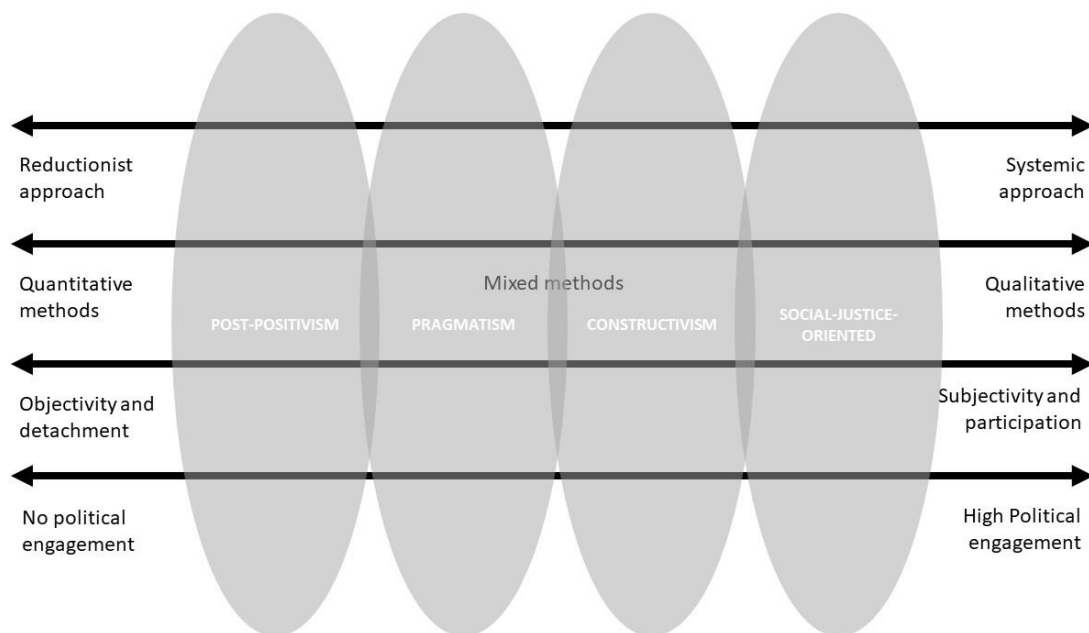
In front of the unquestionable complexity of socio-ecological systems on which both public policies and punctual development intervention are applied, there is a gradient of different reactions framed by two opposite visions.

On one edge, the reductionist vision is the historically most deeply rooted in scientific thought (Bacon, 1620). It assumes that any system can be split into separate parts whose disjunctive analysis is able to provide in an additive way a complete description of the entire system (Kemeny et al., 1956). Complexity is therefore seen as a source of bias in evaluations that should be broken down in its simpler and measurable components. On the opposite edge, the systemic vision comes from system thinking and relativity studies. Systemic vision assumes that a complex system is evaluable only in its entirety (Arnold et al., 2015; Forrester, 1961; Frandberg, 2003; Ulrich, 1987). According to this view, the sum of the individual parts in which a system can be broken down and studied does not restore its entirety. Aristotle put it as 'the whole is greater than the sum of its parts'. This is because there are characteristics emerging from the relationships between the parts that contribute significantly to the system as a whole and that cannot be discarded when evaluating the effects of a programme. Complexity is seen as a resource to be valorised in order to access full information and to grant evaluations their full role in improving programmes.

The four evaluation paradigms can be placed along the gradient between these two opposite visions and, depending on their relationship with complexity, they justify the choice of different methods, different uses of programme theories and different roles for evaluations and evaluators (see also figure 2.3 below for the different roles of evaluators according to each paradigm).

Figure 2.3 – Resuming the role attributed to the evaluator by the four different paradigms of evaluation research. From Mertens et al., 2019.
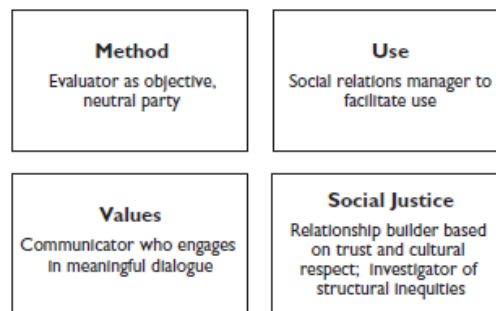
Post-positivist and pragmatic paradigms refer, to a different degree, to the reductionist vision.

Post-positivist authors usually apply the experimental and counterfactual design to their evaluations through which they try to reduce the complexity of the system under examination, thus defining the 'treatment' brought by the programme and identifying and measuring differences in indicator variables for the treated and the un-treated subjects. The randomized control trial (RCT) is the preferred method for evaluating differences with counterfactual approach and evaluation design is drawn mainly on quantitative experimental methods. Post-positivist evaluators refer to the programme theory as a tool for selecting measurable result indicators to be used in order to generalize the results of the evaluation. Conclusions are based on the numbers produced by the methodology, without paying too much attention to the interrelations that exist between the actors of the system nor to the different effect of the programme for different stakeholders.

The more reductionist the background paradigm is, the more the ultimate aim of the evaluator is to identify some specific causal links considered representative of the programme functioning and to focus the evaluation on those. In this way the internal validity of the evaluation and of the identified causal links is put before the external validity (Campbell et al., 1979; Campbell et al., 1963; Shadish, et al., 2002a): methodology robustness and integrity are the key worries of the post-positivist evaluator (McIntosh, 2008) who is a technician whose role is to implement robust methodologies in

order to produce neutral data and objective evaluation results. Evaluation in this vision is a scientific tool for evidence-building in decision and policy making.

Along the gradient between reductionism and holism, the pragmatic paradigm shares similarities both with post-positivist and constructivist visions so, pragmatic authors move from post-positivist position towards system thinking, opening the door to complexity but still refusing it as a resource and therefore trying to reduce it. They focus evaluation user's utility in selecting useful and measurable indicators on the base of the programme theory. Logframes are the preferred schematization of the programme theory and are seen as mandatory tools for planning, indicators selection, monitoring and evaluation. Inside the pragmatic stream, the realist approach to evaluation (Pawson, 2002; Pawson et al., 1994, 2004; Tilley et al., 2000), built a practical compromise between the constructivist and the post-positivist paradigm, accepting mixed methods for collecting and analysing information but still preferring quantitative, experimental or quasi-experimental methods. From the pragmatic point of view the evaluator is a social relation manager who follow the explicit needs of the stakeholders and collect only the data that are useful to demonstrate stakeholder thesis. Evaluation is therefore seen as a practical tool for supporting a more 'political' decision and policy making.

Along the gradient between reductionism and holism, the constructivist paradigm shares similarities both with pragmatic and transformative visions. Constructivists refer to complexity as a positive source of different perspectives. Different value systems brought-in by stakeholders are key elements in the evaluation (Mingers et al., 2010): stakeholders' participation to the evaluation process grants the possibility to select agreed-upon indicators whose measurement is to be achieved with mixed methods. Indicators are selected while elaborating with all the stakeholders the Theory of Change, which is a specific programme theory achieved with stakeholders' participation. Closed Loop Diagram and other conceptual graphic frameworks are used in order to reconstruct and to illustrate the existing complex causal links among the chain of results that explain how inputs transforms into impacts through activities, output and outcomes. Noteworthy experiences are to be found in in the case of complex health programmes evaluations (Butland et al., 2007; WHO, 2009; Leischow et al., 2008).

The most important elements in the systems under evaluation are the relationships among stakeholders, the relationship between the stakeholders and the context and finally the relationship

between the programme and the context itself. For this reason, evaluation is seen as an intrinsic part of the programme and a useful process to build a constructive cultural environment for the programme implementation which in turns create a positive environment for the use of evaluation results.

The more systemic the approach of the evaluator is, the more qualitative the methods for data collection and analysis are, preferring non-experimental methods such as case studies and natural experiments. In fact, methodology flexibility and sensitivity (Best et al., 2010), is felt as the most important features of the evaluator's tools in complex contexts. Constructivist evaluators are no more technicians but communicators or animators whose role is to engage in meaningful existing dialogue between stakeholders and to guide the group toward a sort of self-evaluation for better decision and policy making. Evaluation is no more a scientific nor practical tool for decision and policy making but instead it is intrinsic cultural part of the decision or policy making process.

Transformative paradigm shares the constructivist vision, but it looks at programme implementation and results from the standpoint of existing minorities. A kind of ethical and political view is incorporated in the evaluation, assuming that programmes are expression of the leading social classes and therefore intrinsically bring social injustice for minorities if not carefully planned, monitored and evaluated. Relationships between stakeholders gain even more importance than in the constructivist setting, especially those with minorities. The programme effects emerging from these relationships and from the relationship existing between the programme and the context are seen as key elements to be explained by the evaluation. Preferred methods are mainly qualitative because they allow to point out specific point of views. Transformative paradigm therefore sees the evaluator as a relationship builder whose specific attitude towards cultural respect and sensitivity about structural inequities grants the possibility to point out achievements and failures in decision and policy making. In this perspective evaluation become a social empowering process for communities.

All the mentioned paradigms had an important role in the evolution of evaluative research and are still applied in specific field of development (Mertens, 2019).

Post-positivist and pragmatic paradigms have the merit of using programme theories in order to construct rigorous assessments that systematically eliminate the contextual factors that make causal attribution more difficult. In the primal work of Chen et al. (1983) programme theory was largely

28

based on attempts to identify and quantify those variables that would have an impact on programme outcomes. Following post-positivist authors clearly paved the way for the 'pragmatic compromise' (Chen, 2016) exactly as the pragmatics paved the way for the constructivists by using some specific concepts such as Joint Effects (synergistic effects), Adjuvants (adjuvant factors), Viability and Transferable Cogency[2] (Chen, 2016; Tilley et al., 2000).

The more reductionist the background paradigm is, the more the evaluation practice identifies some individual components in each target programme on which to measure overall performance and effectiveness (Chen, 2016). This occurs in the objective-oriented evaluation (Tyler, 1942) and in its different forms (goal attainment model, Fitz-Gibbon et al., 1996). Noteworthy experiences regarding the use of this approach are present in the field of educational policies (Fitzpatrick et al., 2004).

However, the more reductionist the evaluation is, the more the intrinsic effect of complexity is ignored, thus losing evaluation external validity and often ignoring the peculiar value that those directly affected by the intervention reserve for particular programme effects. Although today more reductionist approaches are criticized in front of the overwhelming complexity of globalized socio-ecological systems, it is to be recognized that the reductionism-oriented paradigms to evaluation played an important role in bringing people to understand, accept and spread the evaluation practice, especially in the context of the evidence-based policy movement. In fact, this approach is still widely used in assessing the 'impacts'[3] of policies and development interventions as evidenced by the many textbooks on the subject (Mertens, 2012; Rossi et al., 2004b; Wholey et al., 2010).

---

[2] Joint Effects (synergistic effects), Adjuvants (adjuvant factors), Viability and Transferrable Cogency are respectively understood as effects determined by the coexistence of several factors, as a configuration of factors that determines a certain effect on the effectiveness or performance of the whole programme (Chen, 2016), as the potential of the policy/intervention to "really realize" and bring about the desired change, as a way to compare the results of different evaluations, or the information useful to produce such results, on the basis of equal configurations of factors. In particular, the realist school of evaluation proposes the formalization of evaluations in CMOC configurations, that are configurations of Context, Mechanism, Output and Context-Mechanism-Output-Pattern (Tilley et al., 2000), in order not only to evaluate the specific case, but to be able to compare it with "similar" cases. From the analysis of these factors emerges the concept of viability. On this basis, the pragmatic approach aims to adapt the evaluation practice to the context and in this direction requires a greater recognition and acceptance, obtainable through a more widespread use (Chen, 2016; Flygare et al., 2013).

[3] Here we use the word "impacts" but objective-oriented evaluations mostly measure output and outcomes while directly attributing these achievements to the expected impact. Other schools of thoughts strongly criticize this attitude by arguing that the configuration "intervention-context" can't be ignored in translating outcomes into impacts.

From the opposite standpoint, one main acknowledgement to more systemic paradigms is that they helped to bring out the importance of the policy implementation process on its final effects. In fact, effects are determined precisely by the relations among involved actors, between the actors and the context, and by the features emerging from these relationships. In order to highlight these complex networks of relationships, it is necessary to use a multidisciplinary and multi-methods approach that responds more comprehensively to the cognitive needs of evaluation. This adaptation of the evaluator to the specific complexity of the case under evaluation, on the one hand grants the possibility to point out the real effect of a programme, on the other hand it determines the risk of an overload of information that is difficult to manage in methodological and practical terms. This is the reason for the existing critics to the applicability of the holistic approach. Although some contributions already exist regarding the attempt to organize the enormous amount of methods and models that can be used within the systemic approaches (Cabrera et al., 2008; Patton , 2010; American Evaluation Association, 2007), further research is needed to improve its applicability (Chen, 2016).

2.2 Purposes of evaluations

Purposes of evaluations have been framed and organized by a wide number of authors and organisations. In general terms evaluations serve a desire for description, judgement and argumentation (Martini et al., 2009). Narrowing the scope of present dissertation to the evaluation of programme and policies, two big families of evaluation exist in terms of purposes: Scriven (1967) first suggested a distinction between formative evaluation and summative evaluation. Formative evaluation is intended to foster development and improvement within an ongoing activity (or person, product, programme, etc.). Summative evaluation, in contrast, is used to assess whether the results of the object being evaluated (programme, intervention, person, etc.) met the stated goals.

'*All assessments can be summative (i.e., have the potential to serve a summative function), but only some have the additional capability of serving formative functions.*' - Scriven (1967)

 '*Formative evaluation is primarily a building process which accumulates a series of components of new materials, skills, and problems into an ultimate meaningful whole*.'- Guyot (1978).

From the definition of evaluation given in paragraph 1.1 (OECD-DAC, 2010) the evaluative action is a goal-oriented process: the evaluation outcome, which is a reasoned judgment with sound

argumentations, is sought precisely because it informs a reflection that in turn helps in takin reasoned decision and actions on the given process. It may be the case to decide what to do in a specific context or whether to start a policy/intervention (ex-ante evaluation), carry it forward (ad-interim evaluation) or decide if to propose it again at its end (ex-post evaluation). The purpose of an evaluation process is therefore first of all to understand the evolution of the system itself, secondly to understand the causal links between the voluntary intervention on the system and its evolutions, and last but not least important, to make recommendations or provide support elements for decisions (or judgement). In other words, it is a learning process aimed at improving a decision-making process. Nowadays, international organisations state that evaluation has the purpose of informing a result-based management of the activities: it means looking at DC as a value chain which should have a structural result framework against which it is mandatory to monitor and evaluate progress and achievements.

In the course of history, however, the evaluation has taken on numerous specific purposes.

In the 2002 book 'Evaluation models' Stufflebeam summons at least five different purposes for an evaluation. Starting from the '50 and passing through the Cold War period, evaluations were used in the USA primarily for improving the efficiency of military strategies and assets, during the '60 and '70 evaluations were used to test or prove several social and educational policies against criteria such as accountability, efficiency, effectiveness and quality (merit and worth). During the '80 effectiveness and efficiency goals converged in measuring the excellency of (US) international business performance while during the '90 the main purpose of evaluation seemed quality, competitiveness, goal attainment and equity of delivery. Since the '90 another purpose of the evaluation emerges: it is the purpose of evaluations with constructive approach that can be called 'transformative' or 'consensus-building'.

In order to understand this very last purpose for evaluations, we have to go back to our definition of evaluation and understand that if the quest for arguments, as a basis for justifying the evaluation outcomes, is something fundamental to give value and weight to the evaluation itself, on the other hand the process of arguments building is inscribed in a negotiating and discursive social process. This means that (1) produced arguments are always susceptible to counter-arguments, and that (2) the evaluation process itself can have strong impacts on the participants and on their environment, which in the end is the very environment in which the evaluated policy/intervention is implemented.

This second feature is of fundamental importance to understand the last evolutions of the concept and of the evaluation techniques.

Evaluation outcome in no more only the reasoned judgement which is supposed to influence decisions about a policy/intervention. The most important outcome becomes the very impact that participation triggers on participating stakeholders. In fact, the evaluation process is an experience that is able to change participant stakeholders themselves and their social environment. In this perspective, evaluations are no longer implemented for the purpose of producing a final output but instead they are integrated in the development policy/intervention as constitutive and transformative processes in order to influence directly the course of things.

In order to be used in this new perspective, the evaluation must be able to recognize the existence of a certain degree of flexibility since its own planning depends on it. This kind of flexibility is not just a negative aspect as one might think superficially. In fact this feature of evaluations is determined by the co-presence, in the context of development and in the evaluation process, of different actors and different points of view that, if on the one hand complicate the process, on the other hand can enrich the attribution of a reasoned judgment with innumerable nuances and connotations which would be otherwise impossible. Diversity is to be seen, in this case as in many others, not as an inconvenient difficulty but as a tangible richness in which to find new solutions.

The dose of relativity, that constructive evaluations imply by their own nature, is addressed by evaluative research in the same way other sciences challenge the variability and unfathomability of the natural world, for instance under the name of 'system biology': with rigorous method and rationality. It is therefore with the aim of making the assessment exercises 'open to inspection' and possibly repeatable that the research establishes standardized procedures and schematizes the objects of the evaluation. With this approach we assume the possibility to create an adaptive research design that open the door to mixed methods as a key element.

Having said this, it is possible to reduce the purpose-related complexity to four broad and interconnected classes of evaluation purposes (see Table 2.2) that include the often-mentioned couple of purposes 'accountability and learning'.

Accountability is defined as the need to make the decisions accountable from the point of view of both the results obtained and the resources invested (internal efficiency and external effectiveness,

merit and worth). In the middle lays the internal consistency of the programme and its relevance (cyclic consistency between intentions, actions, results and impacts).

'Learning' purpose means the will to understand what works and doesn't work in order to decide how to improve the ongoing intervention or to modify future implementations.

Figure 2.4 illustrates how these two broad classes of purposes can be (not exhaustively) further divided in more specific scopes.

Figure 2.4 - Scheme of the evaluation purposes.
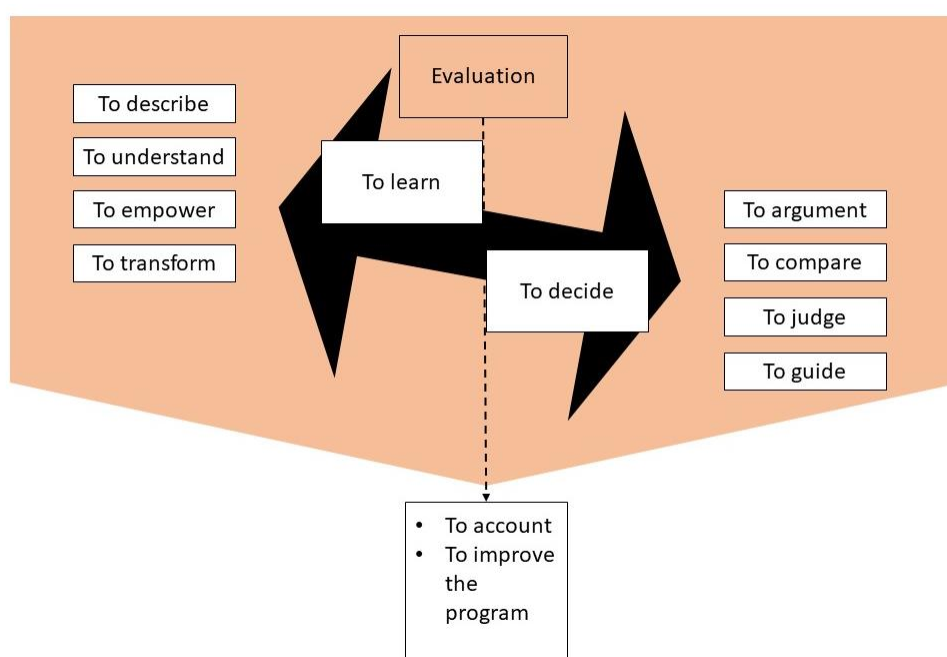
Finally, in other cases, as already mentioned and in a more constructivist and transformative perspective, the evaluation purpose is to 'support the change', by creating consensus among the actors of a given system and therefore to simultaneously 'evaluate progress' and 'improve collaboration in order to improve progresses' in a virtuous spiral.

Table 2.2- Regrouping evaluation types (not-exhaustive list) according to four broad classes of purpose. (Adapted from Mertens et al., 2019, original source: Based on Centre for Disease Control and Prevention (CDC). (1999).

Purposes	Types of evaluation
To gain insights or to determine necessary inputs. For example: • To assess and build capacity in the community. • To assess needs, desires, and assets of community members.	• Context evaluation • Capacity building • Needs and assets assessment • Organizational assessment • Relevance evaluation

Purposes	Types of evaluation
• To identify needed inputs, barriers, and facilitators to programme development or implementation. • To determine feasibility of methods to describe and measure programme activities and effects.	
To find areas in need of improvement or to change practices. For example: • To refine plans for introducing a new service. • To characterize the extent to which intervention plans were implemented. • To improve the content of educational materials. • To enhance the programme's cultural competence. • To verify that participants' rights are protected. • To set priorities for staff training. • To make midcourse adjustments to improve participant logistics. • To improve the clarity of communication messages. • To determine whether customer satisfaction rates can be improved. • To mobilize community support for the programme.	• Implementation evaluation • Responsive evaluation • Participatory evaluation • Process evaluation • Monitoring • Formative evaluation • Developmental evaluation
To assess programme effectiveness. For example: • To assess skills development, knowledge gain, and/or attitude and behaviour changes by programme participants. • To compare changes in provider behaviour over time. • To compare costs with benefits. • To find out which participants do well in the programme. • To decide where to allocate new resources. • To document the level of success in accomplishing objectives. • To demonstrate that accountability requirements are fulfilled. • To aggregate information from several evaluations to estimate outcome effects for similar kinds of programmes. • To gather success stories.	• Outcome/impact evaluation • Summative evaluation • Policy evaluation • Replicability/exportability/transferability evaluation • Sustainability evaluation • Cost analysis • Participants perception analysis

Purposes	Types of evaluation
To address issues of human rights and social justice. For example: • To broaden consensus among coalition members regarding programme goals. • To support organizational change and development. • To determine inequities on the basis of gender, race, ethnicity, disability, and other relevant dimensions of diversity.	• Indigenous evaluation • Culturally responsive evaluation • Disability- and deaf-rights-based evaluation • Feminist evaluation • Gender analysis • Transformative participatory evaluation

2.3   Evaluation in development cooperation

Taken from Sasaki (2006), the following Table 2.3 briefly resumes the evolution in Official Development Assistance (ODA) evaluation from the '40 to the beginning of the new millennium.

In the field of development cooperation first evaluations were carried out exactly when structured DC started to be implemented in the 1940s with the 'father' of all DC initiatives, the Marshall Plan (OECD, 1996). In the mid-1940s, the World Bank and several UN agencies were established (1944 and 1945, respectively). In that period donors were focused with reconstruction after the World War II. Economists provided the main theoretical basis for understanding how aid works (Clements, 1996). Development seemed a simple matter to be planned according to the best achievable economical results, with mild worries about environmental or social issues as criteria for guiding the planning and implementing efforts. In this period, even though no report including the word 'evaluation' in its title can be found, the effectiveness of DC was discussed within the framework of development economics. One good example, the so-called 'Lewis report', with the formal title Measures for the Economic Development of Under-Developed Countries, was submitted to the UN by Nobel Prize economist Lewis, A. (1951).

In 1950s, development aid suffered from ideological confrontation between the Capitalist and the Communist paradigm. In terms of implementing strategy we notice a shift from 'programme aid' to 'project aid'. In this period, preliminary work focusing evaluation methods was conducted (Rossi, Lipsey & Freeman, 2004, p.9) and the first evaluation report, Evaluating Development Projects (Hayes, 1959), was prepared for UNESCO.

In the 1960s, the establishment of bilateral aid agencies boomed. They found their roles in providing technical assistance, budget support and funding for project-type aid, whereas multilateral agencies had supported the large-scale programme-type aid. In this period, economic cost-benefit and cost-efficiency analyses for project appraisal gained some methodological sophistication (Mckean, 1966; Krutilla & Eckstein, 1964; Dorfman, 1963; Mishan, 1971). Hirschman's Development Project Observed (1967) was considered as the true preliminary trial of evaluating aid projects (Cracknell, 2000).

In the 1970s three main changes occurred: first, the main focus of aid shifted to agricultural development and basic human needs (social sectors), reflecting an ideological shift in the aid community, from economic development to poverty reduction.

Second, bilateral and multilateral aid agencies set up evaluation units one after another. These units started conducting evaluation focusing on short and medium-term aspects of process (design and implementation), (Valadez & Bamberger, 1994) and used methodologies were found to be inappropriate for programmes and projects in the social sector. Rossi, Freeman, and Wright works, *Evaluation: A systematic Approach (1st ed.)* (1978) and *Doing Evaluation* (1980), were therefore published as an answer to the need of more systematic evaluation approaches and determined that evaluation started its separation from the purely economic analysis in aid sector.

Third, during this period, some large-scale aid evaluations in Colombia (1971-1975) and Nicaragua (1974-78) employed the experimental design for the first time (Campbell and Stanley, 1966; Rawlings, 2003) determining the first wave of criticism about the extreme difficulty in implementing such an approach, especially in respect of more rapid and economical ways (Valadez & Bamberger, 1994).

The 1980s are known for 'Reaganomics' in the USA, 'Thatcherism' in the UK, and 'Nakasoneism' in Japan: large market-based reforms took place in response to the general distrust towards the government institutions. DC quickly shifted from a project-based aid perspective to the structural programme aid perspective which involves deep reliance on market mechanisms. This wave also stimulated the rise of non-governmental organizations (NGOs) and the movement of participatory/empowerment approaches (Chamber 1988), both of which tried avoiding the use of governmental mechanisms. While evaluation units of aid agencies tried detecting impacts and effectiveness, their evaluation methods varied from very simple and low-cost means, such as expert review (Kumar, 1987 & 1989), to more systematic or formal ways, such as the experimental design. A landmark study, Cassen's (1986) *Does Aid Work?* was produced in this period. In the same period

European donors and the EC (now called EU) developed the Project Cycle Management (Eggers, 2002). Recording the methodological developments in aid evaluation in this period, OECD published *Methods and Procedures in Aid Evaluation* (1986).

During the 1990s two dynamic trends were simultaneously pursued among donor agencies: (i) *the diffusion of the DAC five evaluation criteria* and (ii) the '*Managing Results*' *approach.* In 1991, the OECD-DAC Working Party on Aid Evaluation approved the *DAC five evaluation criteria* (i.e., relevance, effectiveness, efficiency, impact, and sustainability), which are still widely accepted by many aid agencies, especially European donors, and by U.N agencies, as well as Japan. In the 1990s, the Project Cycle Management system took DAC evaluation criteria and became widely used as a management tool among European donors and Japan. The aid community remained anyway committed to market-based development and continued applying its philosophy for their aid. Conversely, the 'Wapenhans report' (World bank, 1992) hinted for the first time in history to the need of moving toward new approaches that neither represented traditional project-type aid nor simple market-based programme loans. Another landmark of the 1990s was the book *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector* (Osborne & Gaebler, 1993). It proposed the concept of 'managing results' by introducing strategic planning and performance measurement. Among aid agencies, the system based on the concept of 'managing results' became generally known as Result Based Management (RBM).

From this period and up to the present, evaluation research is implemented increasingly with mixed approaches and methods: the two main approaches being the DAC five evaluation criteria and performance measurement, or a mixture of theme. This variety of approaches may represent a chance for the system to develop new frameworks and methodologies suitable to appropriately evaluate the progress of inter-sectoral programmes.

Table 2.3 - Resuming the evolution of evaluation in ODA, from the 40s until the beginning of the new millennium. From Sasaki, 2006.

	Dominant or rising institutions	Donor ideology	Donor focus	Types of aid	Trends in aid evaluation	Aid evaluation reports / Influential books
1940s	Marshall Plan and UN system (including World Bank).	Planning.	Reconstruction.	- Marshall Plan (largely program aid)		
1950s	United States, with Soviet Union gaining importance from 1956.	Anti-communist, but with role for the state.	Community Development Movement.	- Food aid - Projects aid		Lewis (1951) Hayes (1959)
1960s	Establishment of bilateral programs	As for the 1950s, with support for state in productive sectors.	Productive sectors (e.g. support to the green revolution) and infrastructure.	- Bilaterals gave TA & budget support - Multilaterals supported projects.	(The boom in establishment of bilateral aid agencies.) - *Sophistication of economic analysis for project preparation*	Campbell & Stanley (1966) Hirschman (1967)
1970s	Expansion of multilaterals especially World Bank, IMF and Arab-funded agencies).	Continued support for state activities in productive activities and meeting basic needs.	Poverty, taken as agriculture and basic needs (social sectors).	- Fall in food aid - Start of import support.	The boom in establishment of evaluation units. - *Focus on the process of aid* - *Large-scale experiments & disappointment.* - *Log-frame by USAID*	Tendler (1975) Rossi, Freeman & Wright (1978)
1980s	Rise of NGOs from mid-1980s.	Market-based adjustment (rolling back the state).	Macroeconomic reform.	- Financial program aid - Debt relief.	Focus on aid effectiveness by various approaches - *Rapid, low-cost methods* - *Participatory approach* - *Training & empowerment*	Scriven (1980) Casley & Lury (1982) Cassen (1986) OECD (1986) Riddell (1987) Chambers (1988)
1990s-	Eastern Europe & FSU become recipients; emergence of corresponding institutions.	Market-based adjustment continued	Environment and gender (but passed quickly).	Financial program aid & debt relief continue	Diffusion of DAC's evaluation criteria vs. Adoption of "Managing results" approach. - *Sector level, country level, and thematic evaluation.*	OECD-DAC (1991) "DAC evaluation criteria" Osborne & Gaebler. (1993) World Bank (1992) "Wapenhans report"
(Late '90s – 2000s)	Aid coordination forum at local level as well as headquarter level	Move back to role of the state. (Balance the market & the state)	Poverty and then governance	Move toward sector support.	New agenda: Sector program evaluation New agenda: Evidence-based evaluation in aid	World Bank (1998) Wolfenson. (1998) Hatry (1999)

*Note.* Entries are main features or main changes, there are of course exceptions.

Moving to present days, international development evaluation has shifted from donor-agency-controlled evaluations to partnership evaluations and country-led evaluations. This shift in evaluation perspective is due to the still emerging consciousness, nested mostly in the work of nongovernmental organizations (NGOs), that top-down evaluations miss the complex, diverse realities of the in-country beneficiaries and focus more on the evaluation tools reliability than on the accomplishments of the programme objectives (IEG-WB, 2009). What is happening today is a decentralization of evaluation culture, knowledge and expertise and a centralization of achievement indicators. The United Nation's 2030 Agenda for Sustainable Development announced that country-led evaluations of the SDGs will be used to inform a global follow-up and assessment of achievements toward sustainable development. The debate is ongoing about how to improve the impact of international development assistance and while several countries are still struggling for the creation of effective international development agencies with declared evaluation policies (OECD, 2016), practitioners are divided between the need of strengthening evaluation robustness and the need to make evaluation more participated and grounded to reality (Bhatkal et al., 2015; Chambers, 2005; Picciotto, 2009). As development initiatives become more complex, conventional evaluation approaches are no longer able to fully evaluate how composite interventions funded, designed and implemented by multiple stakeholders, and operating in complex environments, contribute to observed changes in multiple

(intended and unintended) outcomes. Under these increasingly common scenarios, it becomes necessary to find new evaluation approaches that are complexity-responsive, equity-focused and gender-responsive (EvalNetwork, 2016).

2.3.1 Users of evaluations

Keeping in mind the definition of evaluation given in paragraph 1.1, we summarize here the users of evaluation or the actors interested in implementing evaluation. Leaving behind individual evaluations on personal scale, evaluations can be interesting operations for the private sector, for the public sector and for the wide range of civil society organizations (Figure 2.5).

Figure 2.5 - General non-exhaustive scheme of evaluation fields of application.
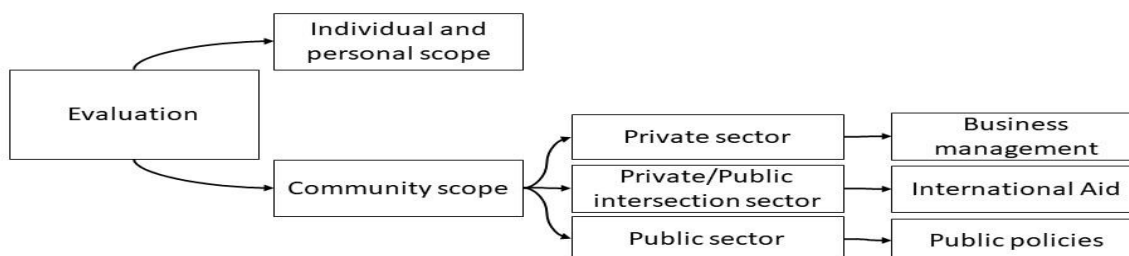

Narrowing our focus to the field of DC, the main users of evaluation are the national agencies for DC, the intergovernmental organizations, the banking foundations and international banks, the NGOs and civil society in general. The different actors of DC are interested in different scales and different types of evaluations and most of all they are interested in running evaluations for different purposes. As it is shown in Figure 2.5 (and also after-on in Figure 2.6), in general terms the interest in results information is different at every level of the DC chain because at lower scale (project or programme scale) it is difficult to measure the contribution of small initiatives on global indicators / criteria. Result information must be aggregated back along the DC chain in order to assess the contribution of DC 'bricks' (small scale initiatives) to the final and global 'building' that is the sustainable development goal. In fact, if accountability is a shared concern, the position along the hierarchy of DC system influences heavily the way that organisation will be able to use evaluation results to correct the behaviour of the whole chain. Nevertheless, evaluation has the role of impacting directly who performs it. Therefore, results information can be used at each level to improve performance and attitudes at the specific level.

Unfortunately today, the reality of a more and more fragmented DC system, alongside increasing pressure for domestic accountability and communication, mean that for most of the DC providers, information about development co-operation results (outputs and immediate outcomes) and performance information (inputs) (tiers 2 and 3 in Figure 2.6) are often favoured over information about development change or long term outcomes (tier one in Figure 2.6). In other words, DC providers are somehow forced to concentrate their evaluation effort to measure and communicate output and outcomes, therefore privileging more criteria such as efficiency and effectiveness of intervention, than relevance, impact and sustainability. This is a bottleneck in the improvement of evaluations information use (OECD, 2017).

Figure 2.6 - Main users of evaluation results along the DC chain. Different positions in the chain inform different interest in results and different uses of results information.
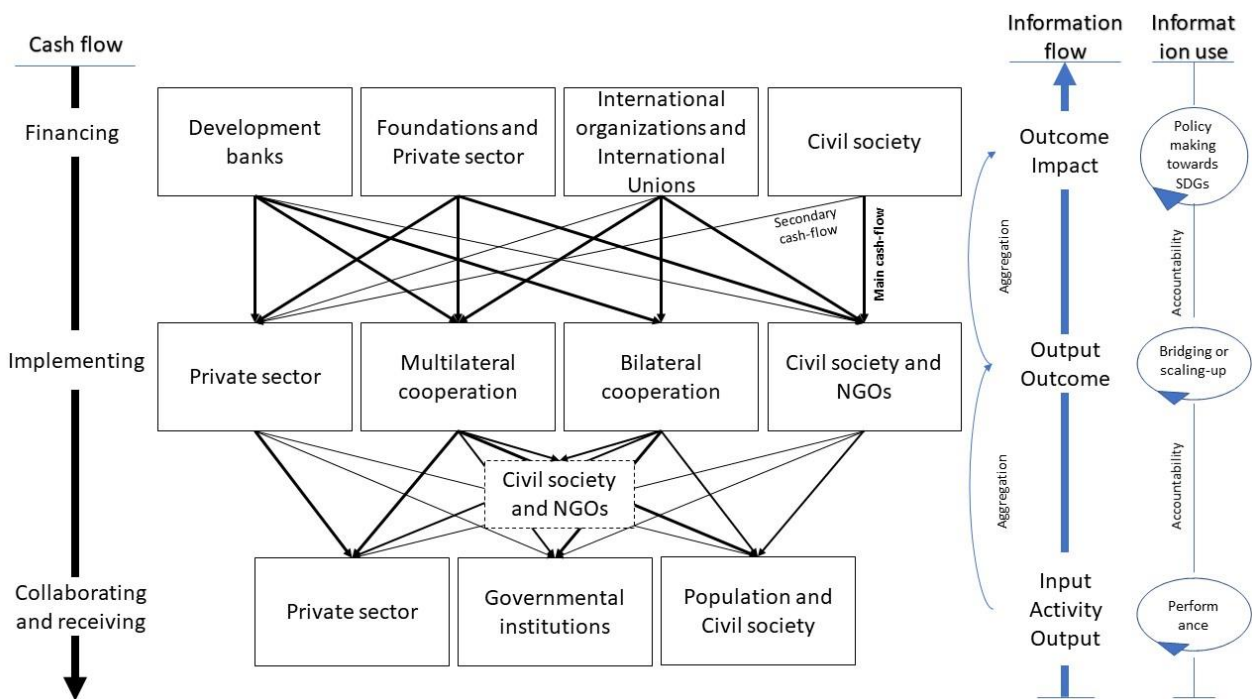
2.3.2 The Evaluand

Evaluations can be framed according to at least three others dimension which are implementation scale, timing and type of evaluand. An evaluand is the object of evaluation.

From a scale perspective evaluation can target local communities and related context and interventions or can address regional, national or multi-country evaluands such as international programmes or policies. SDGs are an example of global result framework for sustainable development evaluation. Evaluations can be implemented before a programme starts (ex-ante),

during the programme implementation or after its conclusion (ad-interim, final or ex-post evaluations). Finally, a variety of evaluands exist. In the framework of DC, several agencies use a list of potential 'objects of evaluation' (evaluands) which creates a list of evaluation typologies. Figure 2.6, from OECD (2010), shows the main typologies of evaluation carried out by national development agencies of OECD member-countries.

Figure 2.7- Showing types of evaluation conducted by national agencies of OECD member countries. From OECD, 2010.
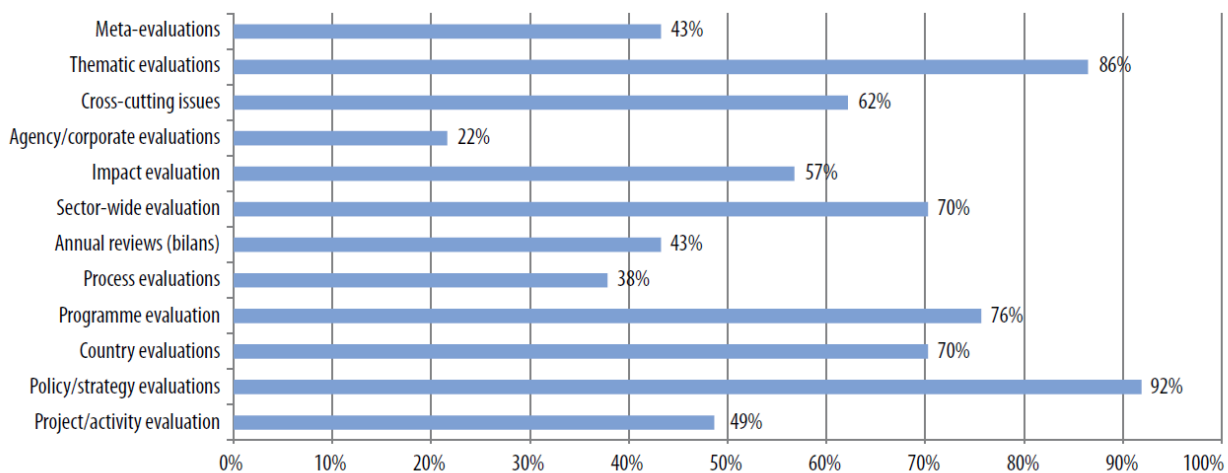
Evaluations within these typologies range in terms of scale and timing of implementation and most of all they target different evaluands. Being national agencies, it is evident that the most targeted scale is regional or national, and that most mentioned evaluands are programmes, policies, strategies, sectors or themes while 'smaller' evaluands such as projects or processes are left to other actors, often to NGOs or specialized independent organisations.

In any case, the evaluand can be investigated in terms of 'processes' or 'results' according to the specific evaluation question. If the main question is 'how does the evaluand.....?' then it is an evaluation of implementation, if the question is 'what does the evaluand.....?', then it is an evaluation of results.

This difference can be seen also by looking at OECD definition of chain of results and of result-based management in DC. Figure 2.7 is taken from the definition of chain of results in one of the last reports (Gertler et al., 2016) of the World Bank Group and it illustrates how evaluations can be defined in terms of implementation or results evaluations.
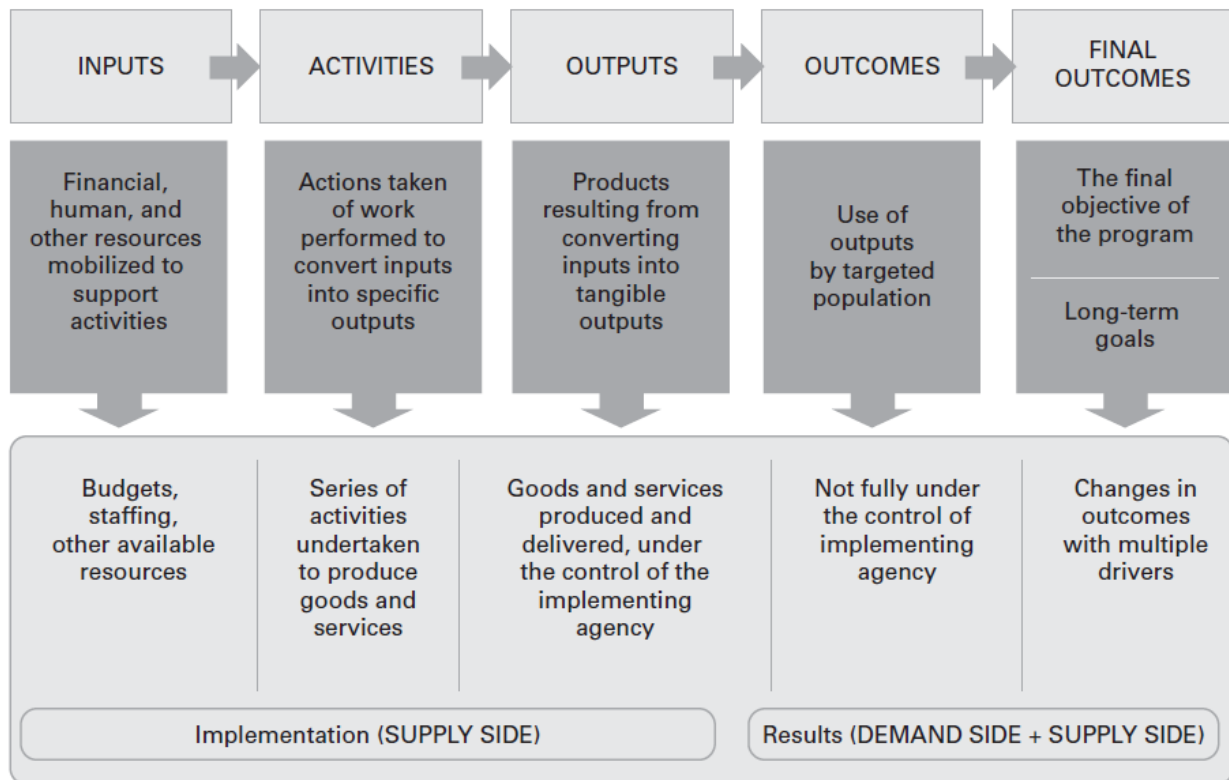
In order to evaluate, it is necessary to investigate (and where possible measure) the context, the theory of change, the implementation process and the effects. We underline here the complementary importance of all these elements and namely of the analysis of the implementation process and of the effects obtained. These two 'aspects of interventions' are of complementary utility both for the outcome of the policy/intervention itself and for the construction and carrying out of an evaluation of the same. The greatest diversity of opinions from different evaluation paradigms and in terms of identifying the most coherent approach to the improvement of policies and development interventions can be found in the evaluation of the effects and in the extent to which the TOC is valorised. In fact, as regards to the evaluation of the implementation process, there is a greater homogeneity of approaches across the evaluation paradigms.

Moreover, although the skills needed to analyse the two sides are slightly different (in the field of political and social science in the first case, more economic-statistical in the second, even if the difference is determined by the methods that have come to be affirmed throughout history more that from a real difference of problems to be faced ..), it is important that the two analyses are integrated. In practice we can observe that these are practically always carried out by the same

organization, with complementary staff focused on both sides. This happens above all if the evaluation petitioner/buyer is sufficiently qualified to understand the organic nature of the evaluative mission.

2.3.3   Evaluation of the context

Context evaluation is a mandatory and very important part of any kind of evaluation. The main reason is that the stakeholders of any programme, the relationships among them, their actions and the programme itself can produce different effects according to different contexts. Context is defined as the part of a socio-ecologic system which is not directly linked to or touched by the activities brought by the programme itself, but that still has important role in shaping the reaction of the system to the intervention. Introducing an interesting special issue on the context's role in evaluations Fitzpatrick (2012) explains that evaluators attitudes towards the context has changed along the evolution of the discipline.

The four different paradigms described so far have completely different perception and reactions to the context of a target policy or intervention. Passing from the post-positivist to the transformative vision and through the intermediate positions, evaluators shift from trying to delete, ignore or set-a-side the context to summoning the role of the context in shaping every little feature of the system under analysis, thing that implies the need for a careful context analysis. In between the pragmatic paradigm and namely the realist current postulates the identification of context indicators as mandatory for reconstructing the context-mechanism-output configuration (CMOC) that is the only way to really attribute causality to what happens during the implementation of a policy or programme (Pawson et al., 1997).

Again, in Rog et al. (2012), the authors develop a framework meant to make context evaluable. The framework is composed of five dimensions: 'the context of the problem or the phenomenon being addressed, the context of the intervention being examined, the broad environment or setting in which the intervention is being studied, the parameters of the evaluation itself, and the broad decision-making context' (p. 27). Each of the five context areas should be examined from different perspectives: physical, organizational, social, cultural, traditional, political, and historical.

2.3.4 Evaluation of implementation[4]

When the evaluation question insists on 'the way or the process by which the policy/intervention is implemented, we talk about the evaluation of the implementation (Rossi et al., 1999).

The need for evaluating the implementation can occur in all the phases of the policy/intervention cycle itself: during the drafting of the policy (through analysis of constitutive scenarios), during implementation (to adjust the aim) or at the end (to learn good practices and avoid bad).

In general, the reasons behind an evaluation of the implementation are:

1. The need to understand what is happening when 'everything changes' meaning when the policy changes not only the behaviour of the beneficiaries but also that of the implementing parties.
2. The need to understand if the actions of the implementing parties are valid and relevant and in any case in order to keep this information also at central buyer level.
3. The need to understand the potential positive or undesirable effects of an experimental policy in order to adjust it before a possible greater diffusion.
4. The need to explain a good or bad functioning of a policy of which the outcomes are already known.

These four reasons for the analysis of the implementation are relatively important, depending on the characteristics of the policy under consideration and namely on the degree of complexity / uniformity and the degree of maturity (experiment or long-standing policy) of the policy/intervention (Martini et al., 2009).

In their cornerstone book Martini et al. (2009) identify four relevant elements that are usually correlated to uncertainty in the evaluation of implementation and that therefore should be carefully analysed during an implementation assessment:

1. The level of organizational innovation/complexity. From this standpoint, the analysis of the supposed performance of the organizational structure of the implementing team is to be assessed and compared with existing knowledge of organizational structures.

---

[4] In this paragraph, where no other specific references are provided, the text is a personal reinterpretation of Martini et al., 2009.

2. The number of actors involved. Obviously larger implementing teams need stronger organizational structures and possibly more innovative solutions for decision making and progress monitoring.

3. The existence of conflicts. It is strictly related to the typology and number of stakeholders and it should be pointed out and treated with specific care because conflicts are both dangerous elements for the achieving of any developmental result and potential richness for societal confrontation, sharing and consensus building. In this perspective the evaluator should carefully analyse the typology of conflict and decide whether a sound strategy to valorise the existing conflict in terms of stakeholders' implication exist and can be applied.

4. The burden imposed on policy recipients. It refers to the fact that the participation to certain policies/interventions is subordinate to a specific behaviour or investment by the recipient. In this case it is simple to identify new obstacles to the implementation process that should be pointed out by the evaluation because of their hampering impact on the implementation process (and on outcomes production).

The evaluation of the implementation should reconstruct both the implementation process itself (who does what, when, and how?) and the mechanism supposed to bring changes (why should the scheduled actions bring the expected change?). In other words, it is necessary to share the TOC with all the implementing partners and see how actors' attitudes toward the TOC match or collide.

An evaluation of implementation must describe at least the following elements of the TOC:

1. Motivation: includes the description of the problem to be addressed and the state that one would like to reach with the policy/intervention (the purpose or definition of the desired improvement of the status quo).

2. The Intervention: includes the description of the recipients, the resources to be used and the activities to be put into practice to transform (or connect) the aims into expected results.

3. The Results: are divided into outputs, short and long terms outcomes, expected and unexpected impacts. The formers are identified by the 'as-is' products of the activities (i.e. number of trained persons).
The outcomes are instead defined by the effect that the realizations have on beneficiaries' lives (what does the greatest number of trained persons entail?) and impacts are the desirable or

undesirable changes in the development path. Results should be expressed in measurable dimensions.

Concerning Motivation, only the stakeholders can truly describe their own reasons, except in the case of hidden interests which should anyway be pointed out by the evaluator. The intervention description should be found on existing documents and can be improved by collecting interpretations by the stakeholders. Concerning result indicators there are two great schools of thought (Martini et al., 2009): the first, which may be called 'top-down approach', sees in the written texts of the policy and in the indicators identified by existing logframes the only yardstick and therefore the only source of valid indicators. In this case the evaluation focuses efficiency and deals with the transformation rate of inputs, through the implementation system, into outputs, often into measurable terms such as funds / results, which are finally taken as indicators. The second school of thought, on the other hand, opposes this approach (and may be therefore called the 'bottom-up approach') and states that the only valid questions can come from the eyes of those who actually applied policies/interventions (street level bureaucracy) and who in the everyday work has made countless micro-decisions on how to apply the written policy, programme or project, coping with endless and unpredictable operational difficulties. In this case the indicators can still be referred to the input / output ratio but generally new quantitative indicators, that were not initially included in the policy / intervention design scheme are used, along with qualitative indicators that emerged from the characteristic bottom-up point of view.

According to numerous authors (Martini et al., 2009; Morris et al., 1978), the implementation evaluation has three main cognitive objectives: to describe, to judge and to explain. The description of the TOC underlying the policy and of the implementation process constitutes the basic elements sine qua non any comparison (factual and judgment approach) or explanation of the results is not possible.

When the description is completed, in order to judge the state/quality of the implementation, it is useful to have a term of comparison which, as previously mentioned, can be defined by examining the intentions of the policy as reported in the official texts or, if not available, in the vision of the policy makers. Also, other similar experiences can serve as a term of comparison with the due precautions. Finally, the expectations of the beneficiaries can be the reservoir from which to draw the term of comparison for the policy in question. And finally, mostly in the case of constructivist or

transformative evaluations, the evaluation can also be concluded without a judgement because description and explanations could be enough for achieving the result of identifying critical aspects and learn upon them.

It should be pointed out that at this stage we are once again on the border between the analysis of the implementation and the effects, while remaining within the first area because the object of evaluation (and recommendations) is always the evaluation process, not the effect on the society / reality.

2.3.5   Evaluation of results[5]

When the evaluation question is 'what does the evaluand achieve?' we talk about the evaluation of the results.

As already mentioned, results are part of a chain connecting inputs to society level impacts. Figure 2.8 shows the chain of results with examples. We underline that the boundary between evaluation of the implementation and assessment of the results is permeable and that the two evaluation phases trespass one into the other at the output/outcomes frontier. In fact, it is easy to understand that the way a programme is implemented, here comprised the attitudes of implementing partners, has consequences on the 'quality' of outputs that immediately translate in intermediate outcomes and medium terms outcomes.

---

[5] In this paragraph, where no other specific references are provided, the text is a personal reinterpretation of Martini et al., 2009.

Each phase of the result chain is usually described by making use of one or more indicators; therefore, we talk about performance indicators or result indicators. Indicators are the preferred way of measuring results. An indicator is a quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor (OECD, 2017). Performance indicators refer to inputs and activities and generally measure efficiency or effectiveness at output level so that also output count can be used in computing performance indicators. Result indicators refer to output, outcome and impact and are used to measure achieved results and to compare them against a threshold that may be internal or external to the project. It is internal if we talk about output or outcome and it is usually set at the beginning of the project as expected output or outcomes in the logframe. Also expected impacts can be stated a priori but usually impact indicators are used to describe the contribution towards an external goal such as one or more SDGs.

Literature (Save the Children, 2019; EUROPAID, 2019) says that a good indicator should be 'SMART' or that it 'ROARS'. The two are acronyms summarizing main features of good indicators:

SMART means:

- Specific: The measured changes should be expressed in precise terms and suggest actions that can be taken to assess them.
- Measurable: Indicators should be related to things that can be measured in an unambiguous way.
- Achievable: Indicators should be reasonable and possible to reach, and therefore sensitive to changes the project might make.
- Replicable: Measurements should be the same when made by different people using the same method.
- Time-bound: There should be a time limit within which changes are expected and measured.

ROARS means:

- Relevant: It measures an important part of an objective or output.
- Objective: If two people measure the same indicator using the same tool, they should get the same result. The indicator should be based on fact, rather than feelings or impressions (another way to say this is to say that it should be Measurable).
- Available: Indicators should be based on data that is readily available, or on data that can be collected with reasonable extra effort as part of the implementation of the (sub-) project.
- Realistic: It should not be too difficult or too expensive to collect the information (related to the next one in the list).
- Specific: The measured changes should be attributable to the project, and they should be expressed in precise terms.

These features should be understood as guiding criteria for choosing good indicators that, in the end, have to satisfy the opposing needs of being specific and useful for generalized comparisons. In the end, project managers must make decisions and select indicators in a pragmatic perspective, meaning that indicators are tools to reconstruct a description of reality that serves the improvement of the reality itself. If an indicator is good for this objective, it will be a good indicator.

As anticipated, indicators can be divided according to the specific phase of the results chain they refer to, so for instance, the EU identifies several levels of indicators when designing the Common

Monitoring and Evaluation Framework (CMEF) for its own Common Agricultural Policy (European Commission, 2015):

- Context indicators describing general information relevant to the policy (such as the amount of agricultural land available or information on the average age of farm managers).

- Income support and market measure output indicators, which provide information on, for example, the number of beneficiaries of CAP income support.

- Output indicators monitoring EU policies on rural development, for example, on the public expenditure for investment.

- Results indicators for the income support elements of the CAP measuring the direct and immediate effects of interventions (for example the percentage of farmers income which came from income support).

- Rural development results indicators assessing the effect of rural development policy, such as preventing soil erosion and improving soil management. Most of these indicators are also target indicators. In addition, rural development complementary result indicators aim to assess the net effect of CAP intervention.

- Target indicators used to set quantified objectives at the beginning of the programming period for the rural development policy (some of which correspond to result indicators).

- Impact indicators measuring the impact of policy interventions for the longer term and when there are effects beyond the immediate period (of which some are also included in the context indicator set).

2.3.5.1   Impact assessment[6]

Impact evaluations target specifically the higher level of the results chain and namely the results affecting society. Impact analysis is therefore a component of the policy or programming cycle in any public management, where it can be implemented as ex-ante analysis (European Commission, 2017) or ex-post assessment (OECD, 2019).

---

[6] In this paragraph, where no other specific references are provided, the text is a personal reinterpretation of Martini et al., 2009.

The nature of impact assessments as part of the policy cycle also means that they are inherently not 'neutral measuring instruments'. They can cause 'observer effects' where the fact of observation leads to changes in the things being observed.

Impact assessment is a theory-based activity: first, it involves establishing a TOC. There are multiple terminologies, but the essence of impact analysis is to establish such a supposed chain of causation ('theory') from intervention to impact and to measure or describe the changes induced along that chain. This approach has the advantage of specificity and focus but also tends to limit observed effects to categories that have been predicted, omitting unexpected effects, including 'perverse' or undesirable effects that had not been anticipated by the designers of the intervention.

Second, a theory of change itself builds on theoretical preconceptions (OECD, 2011) depending on the evaluation background reference paradigm and therefore impact assessments can unwillingly and unconsciously focus on one set of potential effects with trade-offs in respect of the others.

Anyway, impact assessment is aimed at determining and measuring the causal links underlying the practical effects of the policy on its objectives.

The answer to the basic impact evaluation question 'what is the impact or causal effect of a programme (P) on an outcome of interest (Y)?' is given by the basic impact evaluation formula:

$\Delta = (Y \mid P = 1) - (Y \mid P = 0)$.

This formula states that the causal impact (Δ) of a programme (P) on an outcome (Y) is the difference between the outcome (Y) with the programme (in other words, when P = 1) and the same outcome (Y) without the programme (that is, when P = 0). To this end, it is first necessary to accurately define the 'effect of what' and the 'effect on what' which are to be measured. In the typical terminology of the evaluation with counterfactual and experimental approach we talk about 'definition of the treatment' of which to measure the effect and of the 'dimensions' on which to measure the effect. In this simplified view of policies/interventions, complex policies can also be traced, consisting of several interventions with the awareness of operating a huge simplification by reducing the complicated web of causal links between interconnected interventions to a simplified sum of disconnected and individually measured parts. In this fold of the counterfactual discourse are inserted the first strong criticisms brought by other schools of thought like the one that follows the realistic approach to evaluation. Remaining within this conceptual simplification path, the treatment

51

assumes, in the formalization of the method, the form of a 'treatment variable' which is usually binary. In more complex cases we talk about 'treatment dosage' and instead we require the use of continuous treatment variables.

The dimensions on which to measure the effect instead take the form of continuous or discrete 'result variables' which must be observable and measurable.

Looking once more at the results chain in Figure 2.8, these result variables are often the results indicators used to compare the programme achievement with internal (against expected impacts through achieved outputs and outcomes) or external standards or progress thresholds concerning society level development (such as the SDGs).

As already mentioned, a widely used rule of thumb (Gertler et al., 2016) to ensure that the indicators used are good measures is summed up by the acronym SMART. Indicators should embody the following features:

- Specific: To measure the information required as closely as possible.
- Measurable: To ensure that the information can be readily obtained.
- Attributable: To ensure that each measure is linked to the project's efforts.
- Realistic: To ensure that the data can be obtained in a timely fashion, with reasonable frequency, and at reasonable cost.
- Targeted: To the objective population.

2.4   Methods for programme evaluation[7]

Methods used in evaluation can be divided in data-collection methodologies and data-treatment methodologies (UNESCO, 2014), but the two spheres are interlinked because several methods may be used in both phases of evaluation, mostly in the case of a mixed method approach to evaluation (USAID, 2013).

Methods for gathering and processing information can be further divided into the two large families of 'quantitative' and 'qualitative' research methods (Newman, 1998).

---

[7] In this paragraph, where no other specific references are provided, the text is a personal reinterpretation of Martini et al., 2009.

Quantitative research is used to quantify the problem by generating numerical data or statistics. It is used to quantify (variables) states, attitudes, opinions, behaviours. Although the topic is debated in the academic community, quantitative methods are normally used to generalize the results of a sample to a larger population. Quantitative research uses measurable data to describe facts through variables and actively investigates models or statistical functions that can describe the evolution of these variables (at least one dependent variable with respect to one or more independent variables). Quantitative methods put internal validity of identified relations before the external validity and therefore the results of quantitative methods are more difficult to be generalized or externally compared with other results. Quantitative data collection methods are by nature structured. They include various forms of inquiry: online surveys, paper surveys, surveys on mobile media, face-to-face interviews, telephone interviews, longitudinal studies, website interceptors and systematic observations.

Qualitative research is mainly exploratory and for this reason the external validity of its results is higher than for quantitative methods. Conversely, qualitative methods do not investigate causal linkages with statistics and do not provide models for running projections. It is used to understand the reasons, opinions and motivations underlying a particular aspect under investigation. It often provides insights into the problem or helps develop ideas or hypotheses for potential quantitative research. Qualitative research is also used to discover trends in opinions within populations of individuals and to explore the reasons and values that determine them. Qualitative data collection techniques are normally unstructured or semi-structured and do not produce numbers directly prone to statistical analysis. These include the focus group technique (group discussions), individual or collective interviews and participatory observation. The sample size is generally small, and respondents are selected to meet a certain set of criteria.

In general terms, and according to the preferred paradigm, evaluation research can be structured according to various strategies. Figure 2.9 is taken from Bezzi (2010) and shows how strategies for evaluative research can be divided in management-oriented and social-research-oriented. Social-research oriented strategy can be further divided in experimental and non-experimental strategies according to the design and preferred methods for data collection and treatment. Experimental strategy usually relays on the Randomized Control Trial method while methods such Non-equivalent Groups Design (NEGD), Propensity Score Matching (PSM), Regression Discontinuity Design (RDD) and Reflexive Comparisons constitute the range of quasi-experimental methodologies among which non-

experimental strategy can choose. Non-experimental strategy must be designed according to the need of tackling one-criterion or multicriteria evaluations. In the case of mono-criterion evaluations methods can be chosen upon the quasi-experimental methods, in the case of multicriteria and specific complex systems evaluators must rely only on qualitative, non-experimental methodologies, such as case studies and natural experiments.

Figure 2.10 – Sketching the different strategies for evaluation planning which determine the use of different methods. From Bezzi, 2010.
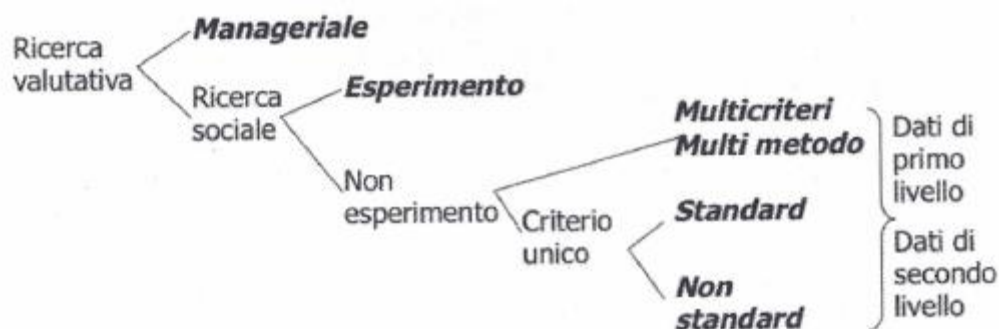
Methodologies for evaluation can be framed in three broad strategies: the experimental, the non-experimental strategies and the descriptive strategies. The difference stems from the overall structure of the methodologies and from the type of data: in experimental strategies the evaluation designs is structured and data comes from artefact groups of subjects while in non-experimental designs data are observational and comes from real-world groups, not completely controlled by the researcher, and the structuration of the design is somehow simpler. Non-experimental methods can be further divided according to the degree in which the researcher can select groups on observable or unobservable variables and according to the use of more quantitative or qualitative methods. Once more the different classes should not be understood as completely separated groups but instead as a gradient of different methodologies that share some feature. On one edge of the gradient lays the experimental method that represents the warhorse of the most reductionist and purely counterfactual approaches. It derives from the medical pharmacological field, it uses to simplify the system and to control in detail the selection of 'treated' and 'untreated' subjects: it is normally applied at the scale of pilot interventions and is usually applied in order to inform the decision over a possible extension of the treatment. The truly experimental method uses quantitative method for data production and analysis and implement the Randomized Controlled Trial design. The experimental method is rarely applicable since, unlike pharmacological experiments, policies hardly

lend themselves to the voluntary creation of a group of people who are denied the possibility of benefiting from the policy itself. Despite of this, there are numerous evaluations conducted with this method, especially when the evaluation is integrated in the policy making process from early phases[8].

Non-experimental methods constitute a broad group of methodologies that can be divided according to the mechanism of selection of the policy/intervention's beneficiaries. Non-experimental methods in fact are used by evaluators in case they cannot control the formation of treated and untreated subjects groups, and therefore they have to find an existing subdivision among the subjects group that can be used for developing the counterfactual comparisons aimed at highlighting the effect of the treatment (in these cases data are observational and therefore not experimental). In the case a policy selects its beneficiaries based on observable characteristics it is possible to use methods like Regression Discontinuity Design and Propensity Score Matching. If the policy selects its beneficiaries on the basis of characteristics that are unknown, for example in the case of volunteer participation to the programme, the evaluator can use methods such as the 'difference in differences' one, the 'instrumental variables' one, the 'analysis of the interrupted time series' one, the' detection of perceived effects' method and finally the 'natural experiment' method.

Moreover, non-experimental strategies fade into descriptive strategy which include quantitative and quantitative, normally non-counterfactual, methodologies and all the qualitative unstructured or semi-structured methods for data collection and analysis among which participatory methodologies are increasingly diffused. Case study method, participatory observation, Delphi-method, focus groups, brain storming, unstructured or semi structured interviews are among the main non-experimental and descriptive methods used in evaluative research. Figure 2.10 provides a graphical framing of mentioned methods: different methodologies are placed in a two-dimensional gradient determined by qualitative-quantitative and counterfactual-non-counterfactual features.

The different timing and scope of evaluation process determines the selection of different methods. Figure 2.11 shows the methods in a bi-dimensional space framed by the different steps of the results

---

[8] We refer to the amount of American and British policies on which the experimental method was used, including the American experience of the National Supported Work Demonstration conducted by the MDRC around 1980 (Maynard et al., 1979), the English experience of Employment Retention and Advancement Demonstration in the '90s (Greenberg & Morris, 2005) and the Mexican experience of the PROGRESA programme launched in 1997 (Behrman, 2010).

chain, which corresponds also to the timing of the evaluation (ex-ante to ex-post), and the qualitative or quantitative quality of the different methods.

Figure 2.11 - Scheme showing the positioning of mentioned data collection and analysis strategies in a double gradient between counterfactual and quali-quantitative features.
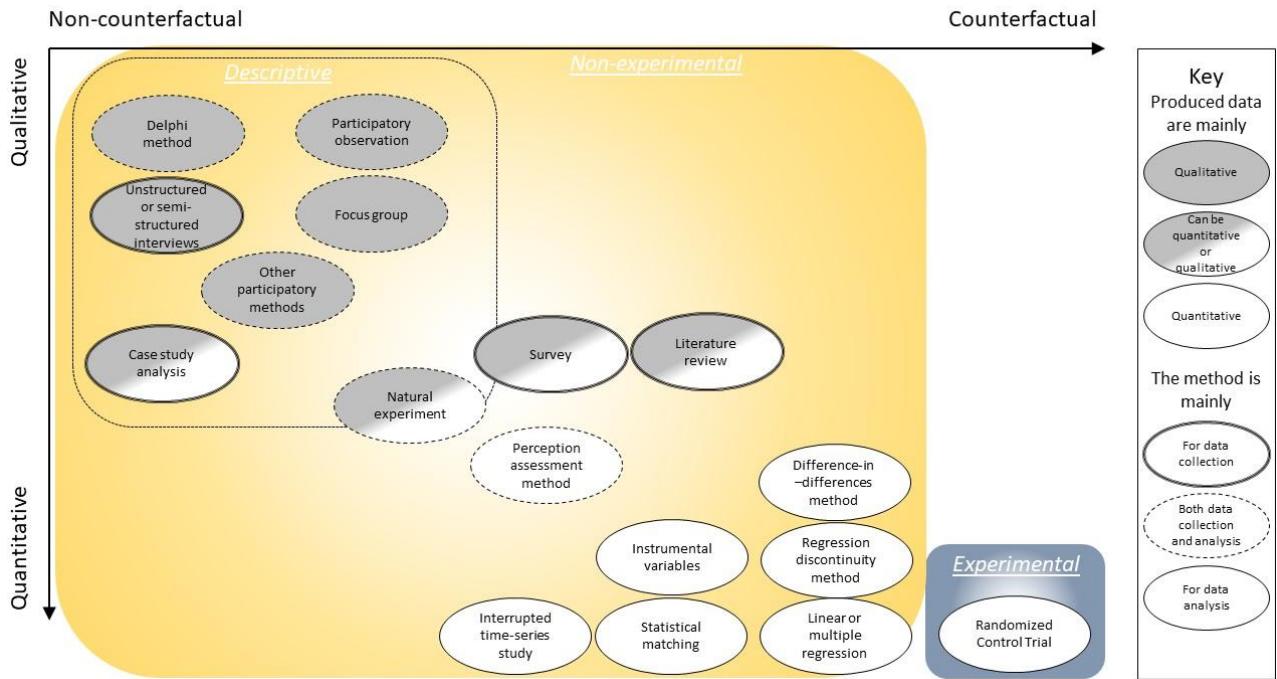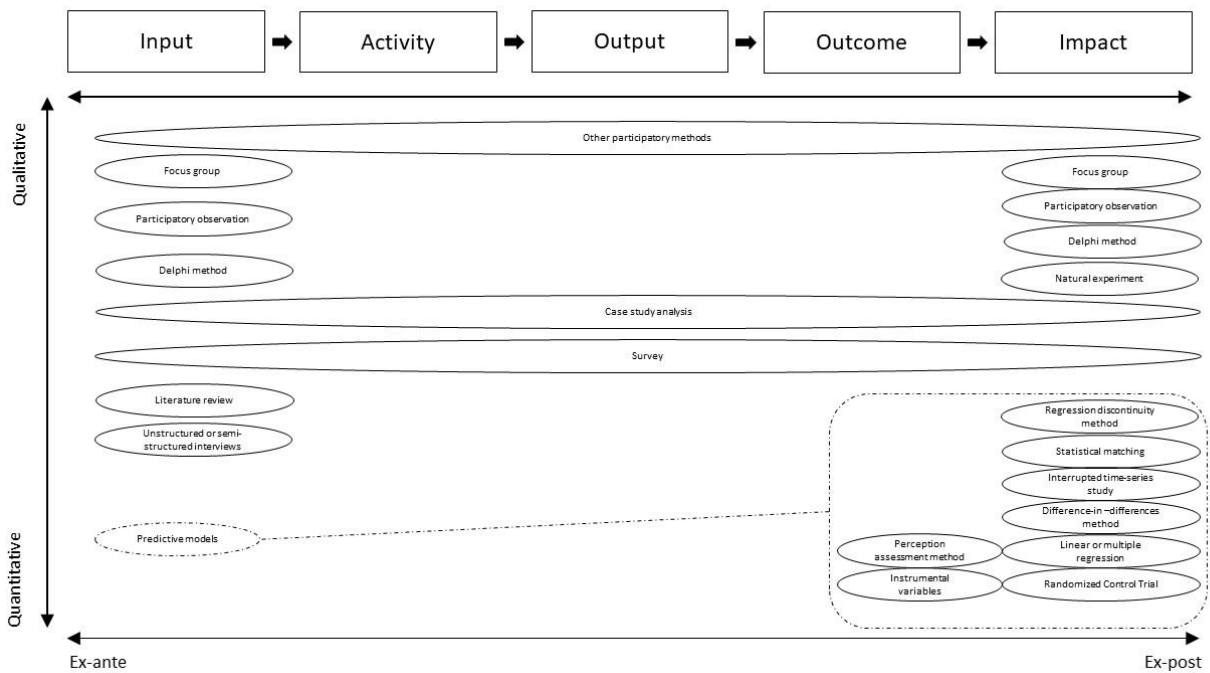

Figure 2.12- Showing evaluation methods framed by the results chain and the quantitative or qualitative quality of the methods.

As it appears clear, some of the mentioned methodologies are useful both in ex-ante and ex-post evaluations. It's the case of most participatory methodologies, of the case study approach and of the surveys which are evergreen methods. Other methods are specifically conceived to collect or analyse information in some phases of the results chain. In particular, ex-post impact assessments are often conducted with a counterfactual approach and experimental or quasi-experimental methods. Some of these methods, such as the regression analysis or time-series analysis, can be used also as parts of predictive models which are useful tools for ex-ante impact analysis.

2.4.1 Choosing evaluation methods

There is no definitive way to frame the selection of proper evaluation methodologies without knowing the specific programme and context. The 'best' method or combination of methods is dependent on the evaluation questions, intended uses and attributes of the intervention and evaluation process. For this reason, several organisations produced their own framework for classing methods and choosing the appropriate one depending on evaluation circumstances (BOND, 2019; WHO, 2008; UNDP-IEO, 2019; Better Evaluation, 2014; Kellog Foundation, 2017). In general terms, evaluation methodologies should be chosen after a rapid 'evaluation of evaluability' of the specific intervention in the specific context. Data availability, implementation cost, time availability, available human capital and other really practical constraints may constitute the most important factors when dealing with real-life choices of evaluation approach, strategy or method.

2.5 Criteria for programme evaluation

Criteria to be used when implementing DC evaluations have been formalized in 1991 by the OECD-DAC and since then international guidelines for DC result-based management agree on the fact that national evaluation policies should include a clear statement of evaluation criteria. Table 2.4 shows the evaluation criteria used by some important providers of DC and confirm the widespread use of the five core OECD-DAC criteria: relevance, effectiveness, efficiency, sustainability ad impact.

Table 2.4 . Showing the results of the assessment concerning the evaluation criteria used by some main DC providers. For each provider the used criteria are marked with the 'x' symbol. Sources of information concerning each provider is reported at the bottom of the table. (* five specific performance criteria on agricultural development).

	OECD-DAC	DFID	EU	IFAD	WB	UNEG/UNDP
OECD-DAC STANDARD CRITERIA						
Effectiveness	x	x	x	x	x	x
Efficiency	x	x	x	x	x	x
Impact	x	x	x	x *	x	x
Relevance	x	x	x	x	x	x
Sustainability	x	x	x	x	x	x
OTHER CRITERIA						
Acceptability			x			
Adaptation to climate change				x		
Coherence / Complementarity		x	x			x
Coordination		x	x			
Coverage		x				
Environment and natural resources management				x		
Equity			x			
EU-added value			x			
Gender equality				x		
Innovation and scaling up				x		
Utility						
Women empowerment				x		
Source	(OECD-DAC, 2010a)	(DFID, 2013)	(EC-European Commission, 2014)	(IFAD, 2015)	Gertler et al., 2016	(UNEG, 2016)

The effort by the OECD-DAC since 1991 has brought evaluation providers to converge on the use of standard evaluation criteria that may be enriched by some specific indicator for thematic assessment such as the case of agricultural development related indicators used by the IFAD (2015). The standard criteria are:

- Relevance: Are we doing the right thing? How important is the relevance or significance of the intervention regarding local and national and funding requirements and priorities?

- Effectiveness: Are the objectives of the development interventions being achieved? How big is the effectiveness or impact of the project compared to the objectives planned (Comparison: result –planning)?

- Efficiency Are the objectives being achieved economically by the development intervention? How big is the efficiency or utilisation ratio of the resources used (Comparison: resources applied – results)?

- Impact: Does the development intervention contribute to reaching higher level development objectives (preferably, overall objective)? What is the impact or effect of the intervention in proportion to the overall situation of the target group or those effected?

- Sustainability: Are the positive effects or impacts sustainable? How is the sustainability or permanence of the intervention and its effects to be assessed?

As it is shown in Table 2.4, some DC providers have created their own specific criteria. First, it has to be noted that, even if some criteria are different, the majority overlaps in their significance: for instance, Coherence, Complementarity and Coordination criteria overlaps in the sense that they all aim at fostering the integration of the target programme into broader policies, strategies or other programmes. Other criteria such as the ones generically related to equity also overlap because they aim at a specific focus on disadvantaged classes when evaluating the effects of a programme. Finally, some provider has criteria that are specific for their scope such as the IFAD that, usually working in the agricultural sector, has a criterion related to environment and natural resources and specific criterion also for woman empowerment, or the EU that inserted the EU-added value criterion in order to assess whether the target programme is generating some specific pro-Europe effect.

2.6   Result-based management in development cooperation

2.6.1   The results chain

In parallel with the diffusion of public sector management among the public administrations that first occurred in the USA and in England during second part of the 19th century, the application of results-based management emerged in the field of DC (OECD, 2014; OECD-DAC, 2000). High level agreements[9] to manage for development results[10], which were originally spurred by commitments to measure progress towards the MDGs (UNDESA, 2010; OECD, 2019), were signed during those

---

[9] High-level agreements to boost development effectiveness through enhanced focus on results were made at Paris (2005), Busan (2011) and Nairobi (2016) (OECD, 2010a, 2019).

[10] Development results are defined as the output, outcome or impact (intended or unintended, positive and/or negative) of a development intervention (OECD, 2010).

years. These agreements have been instrumental in shaping the behaviour of development co-operation providers with respect to results.

For the DC providers, the so-called result agenda took the form of internal results frameworks meant to collect and aggregate result data coming from lower stages in the chain of implementation and results. Several types of frameworks have been created at different levels and scales (agency-wide/corporate, country or sectoral programmes, or individual project). These operational frameworks were meant to collect and integrate basic results from the different levels of the result chain which links inputs, to activities, outputs and ultimately to outcomes and impact[11] (See Figure 2.12). At each level of the results chain, results information (in different forms) is used to demonstrate achievement, or not, of results. Results information can be used for accountability, communication (both internal and external), direction (steering and decision making) and learning.

Results information or data about the programme can be categorised into three tiers (Figure 2.12):

- Tier 1: development results: global results, country results (impacts and outcomes)
- Tier 2: development co-operation results: direct results of interventions (outputs and outcomes)
- Tier 3: performance information: financial and performance information (inputs and management information)

Tier 1 refers to the information about programme outcomes (short and medium terms effects) and impacts (long term effects). If the evaluation works at this level, it is the case of a results evaluation that usually has the form of an impact assessment and the main answer provided addresses the extent to which the desired change is happening in reality. Tiers 2 is still referred to results of the programme but on a lower level along the results chain. In fact, the information produced by the evaluation at this level is about the outputs and sometimes the outcomes of a programme. These are normally useful in order to judge how the implementation is achieving the expected results and to answer accountability questions. Tier 3 refers to the information about the implementing structure. It can take the form of a description of the implementation process or it can be performed as a quality monitoring or efficiency assessment against some standard threshold.

---

[11] Impact in the results chain is defined as: "positive and negative, primary and secondary long-term effects produced by development interventions, directly or indirectly, intended or unintended" (OECD, 2010).
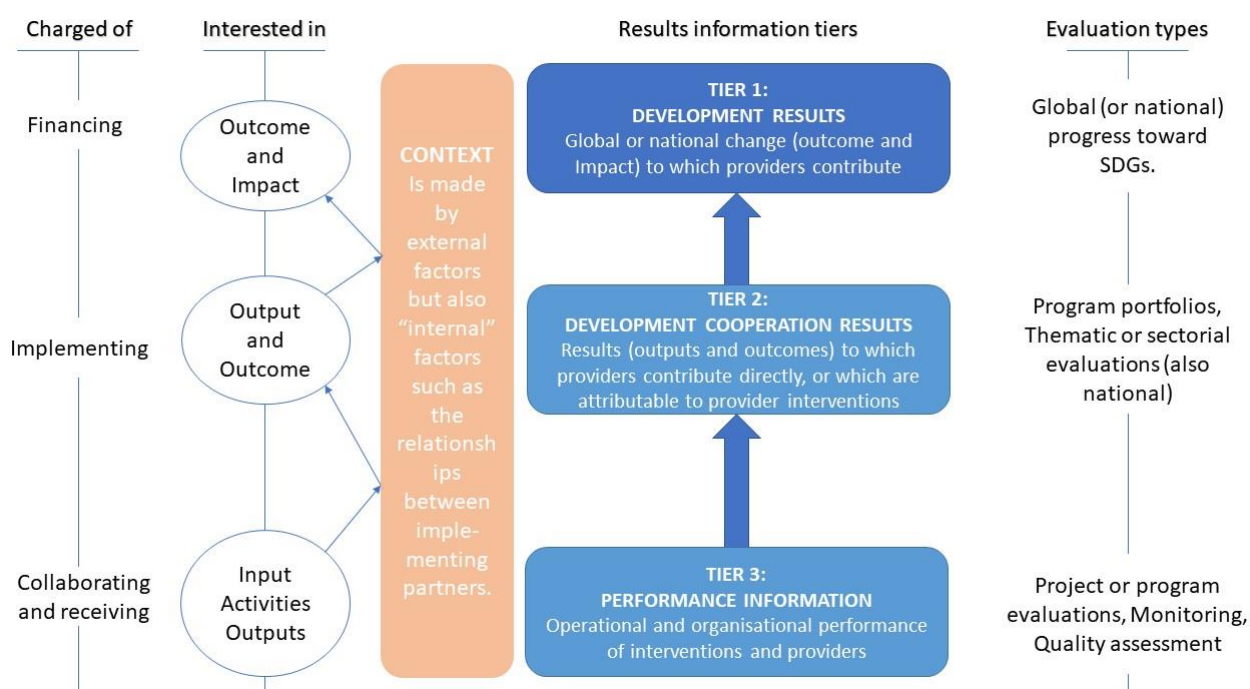
The final goal of DC is supporting the achievement of sustainable development results (outcomes, and tangible impacts or changes). In the framework of DC, the results-based management has been introduced to support the actors to achieve the ultimate goal.

Nowadays DC is changing and is more fragmented with a greater number of actors, channels and instruments involved (OECD, 2017). Many OECD-DAC members are in fact reducing bilateral aid, replacing this channel with a more 'projectized' approach with increased use of non-governmental channels, private sector instruments and multilateral organisations (OECD-DAC, 2010a). Adapting their action both to specific local needs and to global sustainability objectives, DC providers now utilise a range of results-based management approaches ranging from top-down agency-wide results (and performance) frameworks to more context-driven and country-led approaches.

DC contributes to tangible and sustainable improvements in the lives of beneficiaries in partner countries and at the same time constitutes a workbench for development strategies in diversified contexts. Generating and using evidence of the results of DC is essential both for improving its own impact and for learning. Figure 2.13 shows how a result-based decision making in the framework of DC can foster the attainment of sustainability and accountability of interventions.

Results-based decision-making leads to effective development co-operation by:
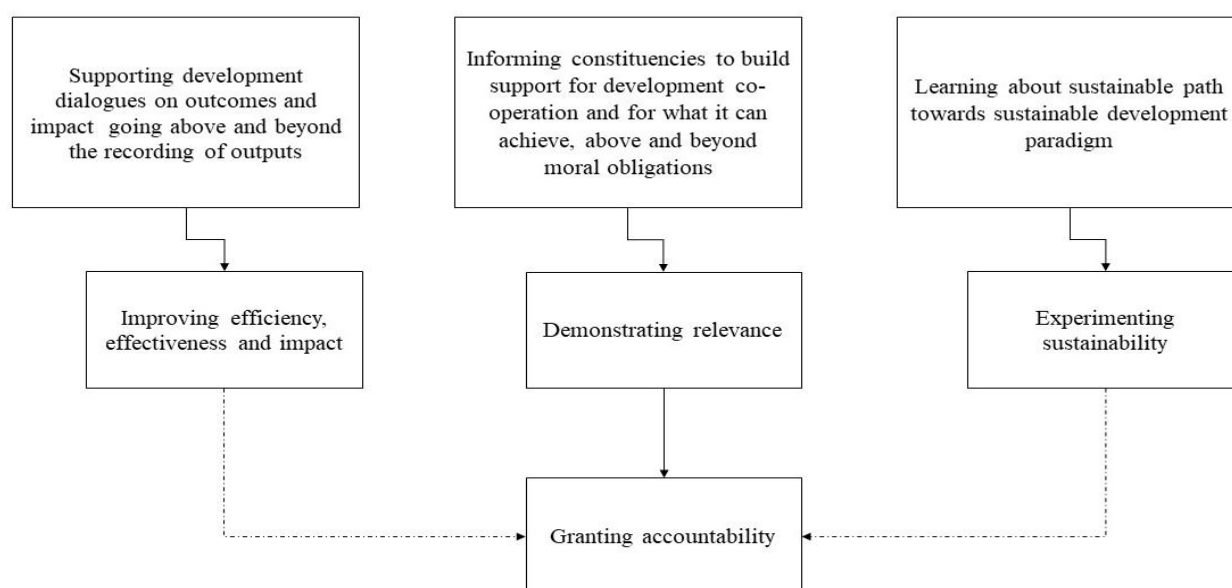
Looking at this matter with the lens of SDGs, we need results information to ensure that the decisions taken by DC providers can really foster the achievement of sustainable development and the expected results scheduled by the 2030 Agenda for Sustainable Development. This concept will be further detailed in the following paragraph also thanks to Figure 2.15.

Managing for development results remains a priority of providers and partner countries as it is witnessed by several studies conducted by the OECD-DAC and other actors of DC (OECD, 2016; OECD-DAC, 2013). Results information is now widely used for accountability and communication. However, there are challenges in more purposeful use of results information for direction, including strategic policy making, and learning, including quality assurance.

The OECD (2017) identifies six interrelated challenges to be tackled by DC providers:

1) Building new narratives able to link project level results with overall sustainability goals is challenging but worthy, as the consequence would be an improved knowledge and communication about the needed path towards global sustainability.

2) Constantly innovating and integrating the RBDM systems at each level of DC results chain is costly but it paves the way for M&E systems that are fit for their purpose and scale.

3) Being realistic about attributing and aggregating results along the results chain, in fact the generation at each level of data and the aggregated use is methodologically challenging. Therefore, providers should be realistic about aggregation and attribution, focusing more on the contribution of development co-operation and streamlining and harmonising indicators among providers.

4) Enabling country ownership of results information without constraining the ability of partner countries to improve data collection and use by diverting attention from country statistical systems. A new generation of provider-specific results frameworks at country level represents an opportunity for a shared platform for results based management. The Sustainable Development Goals and their targets and indicators can serve as a basis for this.

5) Linking results and performance information is important and used by all providers to demonstrate how well aid is delivered.

6) Enhancing resources for RBDM is a fundamental step towards the diffusion of a learning culture where evaluation is no more external judgement but instead a community driven process able to improve performance in delivering impacts.

A we will see in the next paragraph, in order to tackle these 6 challenges it becomes very important to find a compromise between reductionist and holistic approaches to evaluation: safeguarding the ownership of knowledge and systems by the whole community of stakeholders becomes fundamental in order to build solid attributions of development progresses to the action of DC.

The principles of effective DC emphasize ownership by developing countries and the use of their results frameworks in evaluation. Today, however, the results frameworks of individual providers are often built around their own goals and priorities. Aligning with the results frameworks of partners requires country insights, knowledge of good practice and willingness to be held accountable for results.

The universality of the SDGs enhances the potential for mutual accountability in development co-operation. DAC members are working to integrate the SDGs into their individual development co-operation objectives, and to support their partners to deliver and demonstrate results against this comprehensive and demanding global framework.

Development co-operation needs to move up the results chain towards impact; to strengthen its results frameworks, and to give substance to the results dialogue about real, tangible change on the ground.

2.6.2   Evaluation culture

The situation described so far is the main reason why several organizations call for the need to spread an 'evaluation culture' in order to improve the impact of evaluations on DC learning system and consequently to tune DC action.

In its simplest form, an evaluation culture can be defined as a commitment to roles for evaluation in decision-making within an organization (Owen & McDonald 1999). Other authors have further developed the concept by focusing alternatively on structural policies (Roxana Melenciuc, 2015) or on the DC sector (Link & Social Value Italia, 2007; Mayne, 2008).

In an interesting brief of the Institutional Learning and Change (ILAC) Initiative, Mayne (2008) clearly identify the features of an organization showing a true evaluation culture in its activity. First of all, the author argues that, if many organizations have systems of results, this does not mean they possess and enact a true evaluation culture. In order to understand this, it is important to state that all organizations have their own existing culture, which, as Kim (2002) notes, '… conveys a sense of identity to employees, provides unwritten and, often, unspoken guidelines on how to get along in an organization…an organizational culture is reflected by what is valued, the dominant leadership styles, symbols, the procedures, routines, and the definition of success that make an organization unique'.

Now, the fact that an organization has a results-based planning system with results frameworks for programmes, it has results monitoring systems in place generating results data, it has an evaluation unit and a reporting system in place, it does not mean that an 'evaluation culture' is in place.

Conversely, the existence of an evaluative culture in a system or organization would determine the evidence of certain specific features or routines (Mayne, 2008). In Table 2.5 we report the features proposed by Mayne, we slightly adapt them to the DC system and we point out seven propaedeutic steps towards the objective of spreading an evaluation culture.

Table 2.5 - Effect of Evaluation Culture on organizations or systems and propaedeutic results in order to spread evaluation culture in the DC system.

ID	Features of an organization (or existing routines)	Features for the DC system (a system of different stakeholder organizations)	Expected results in the DC system in order to spread an evaluation culture
1	Structured learning events routinely held to discuss future directions, using available results data and information	Structured learning fora routinely held to discuss future directions, using available results data and information.	1. A strategic central organization coordinating DC (and evaluation within) worldwide is operative.
2	Senior managers regularly stressing the importance of credible results information for good management, and asking for results information at programming meetings	Senior managers regularly stressing the importance of credible results information for good management, and asking for results information at programming meetings.	2. The central organization routinely organizes fora. 3. The central organization produces and makes available consistent multi-sector and multidisciplinary evaluation methodologies and tools for all the different scales and contexts.
3	Organizational units accountable for demonstrating that they are learning	Central organizational unit accountable for demonstrating that learning is occurring	
4	Participation in measuring results occurring throughout the organization	Participation in measuring results occurring throughout the system	4. Scientific and general public communication occurs and entails reactions from the target plethora.
5	Decisions on design and deliver routinely and visibly informed by results	Decisions on design and deliver routinely and visibly informed by results	5. Participation is widespread and provides feedback from all the different stakeholders.
6	Good results management showcased	Good results management showcased	
7	Results information widely shared around the organization	Results information widely shared across the different stakeholders	
8	Honest mistakes tolerated and seen as opportunities to learn and improve rather than as opportunities to blame or penalize	Honest mistakes tolerated and seen as opportunities to learn and improve rather than as opportunities to blame or penalize	6. A 'constructive environment or atmosphere' is in place, granting the possibility to all the different stakeholders to profit from the evaluation process.
9	Training on 'results matters' integrated into regular manager and staff training, supplemented with specific results management	Training on 'results matters' integrated into regular manager and staff training, supplemented with specific results management.	
10	Managers able to adjust their activities and outputs to reflect what is being learned	Managers able to adjust their activities and outputs to reflect what is being learned	7. Managing bodies, main groups of interests and other leading actors in the DC sector is committed to accept evaluation results and namely the results that do not fit the leading vision (or the most powerful lobby).

The expected and propaedeutic results we proposed in Table 2.5 overlaps largely the six OECD challenges described in the previous paragraph (OECD, 2017) and the '12 lessons learned' by the OECD-DAC (2013). These are challenges for the DC providers in order to improve their global result based management system, which in turn is supposed to foster the positive impact of DC. These challenges may be condensed in three scale-based classes of problems by crossing them with the existing typologies of research being implemented on the topic.

In facts, an interesting work by Adelle et al. (2012) reviews existing literature on how evaluation affects policymaking. The author distinguishes between four types of research that can contribute to the increase of evaluation effects in policymaking processes. These research typologies also apply to the DC sector.

Type 1 research deals with the design of appraisal/evaluation, focusing on tools and methods; type 2 aims to assess the performance of ex-ante and ex-post appraisal. Type 1 and type 2 constitute by far the bulk of literature on both ex-ante and ex-post evaluations, and are normally produced by practitioners and consultancies, often with the aim to advise policy-making on the best way to implement an appraisal.

Type 3 research focuses on evidence utilization and whether appraisal leads to policy change via processes of learning.

Type 4 takes the investigation further by addressing the real motivations of policy-makers and the interests at stake in policy evaluation. Types 3 and 4 are focused on measuring the effect of evaluation on policy making and on identifying external drivers, concurrent to the evaluative action, that determine policy choices.

Crossing the six OECD challenges, or the expected and propaedeutic results we proposed in Table 2.5, with these areas of research helps in narrowing the target and focus on three possible pragmatic areas of study and scales of intervention.

According to our opinion these three research directions are:

1. Evaluation approaches and strategies
2. Other drivers of policy-making processes

3. DC system infrastructure: implementation methodologies and learning systems (focus on diversity of stakeholders and evaluation methods)

The first point, and the second to a lower degree, refer to the higher scale (tier 1 in Figure 2.12). Accepting complexity in evaluation means a particular challenge for both ex-ante and ex-post perspectives. It also means a scale shift from project and programme assessments to broader policy appraisals and therefore invites reflection on a new synthesis of approaches, as the actually most diffused may not be sufficient to draw broader policy conclusions.

Literature witness the existing debate on paradigms and evaluation approaches, which translate into different evaluation designs and methodological choices. Different 'evaluation schools' have bad-placed reciprocal prejudices about the results of evaluations implemented with any specific approach or methodology that is different from the ones consolidated in that specific 'school'. We need to get over the dividing differences in approaches to realize that in order to tackle complexity we need all the different methodologies working together in a systemic way.

In order to understand this point, we can bring the example of SDGs framework. The SDGs are designed as a global framework to guide development towards sustainability. They are signed and declared as 'interconnected and indivisible' and need addressing in a systematic or holistic way. No one of the existing evaluation paradigm is able to answer this call on its own: for instance, qualitative perception and estimation and precise quantitative measurements are to be used in synergy if we want to capture complexity and design interventions able to interact with it. We need a new evaluation framework that is able to harmonize and use the different features expressed by the different approaches in order to adapt to and to valorize diversity and complexity of contexts and actions.

Still at the higher scale, another important challenge is about granting to the evaluation systems the real ability to impact decision-making at policy, programme or project level. Until there is even one little doubt about the possibility that policy decisions are made for any different reasons other than the results of the evaluative action, evaluation systems will be useless (Canover, 2985; Weiss, 1993).

The increased focus on policy level appraisal, as well as the embedding of evaluation within the DC agenda, make evaluation intrinsically more political. This raises questions for the DC system about

what it really aims to achieve through evaluation and how it can translate this institutionally. For academic research it raises increasingly type 3 and type 4 research questions (Adelle et al., 2012)

Systems and organization showing a real evaluation culture could tackle these challenges because their action would imply accumulating knowledge and 'learning' about 'complex systems' reactions to complex interventions'. This entails the need for approaches and methodologies that are able to assess complexity in a systemic perspective. A real systemic perspective in evaluation should grant the possibility to trace any existing driver, other than evaluative action, of policy choices and to describe their role and weight them according to their impact on society and environment. The evolution of evaluation paradigms that we traced in paragraph 2.1 witness that needed transformations are one the way, but that they need further research, experimentation and use.

The third research direction refers especially to the second and third tiers in Figure 2.12. Literature hints at the lack of evaluation standards as the source of communication problems across the DC system and among the DC providers at the different scales and timings of the policy/project cycle, translating in the impossible use of evaluation results in comparative and learning terms across the whole system.

If SDGs are to be a global and shared framework of indicators, the system measuring the progress towards them should be shared and agreed in the same way. Unluckily, DC stakeholders at different levels have different understanding of the role and sense of evaluation, translating into different connotations of evaluative action unable to communicate and build knowledge. This is witnessed for instance by the difference in the understanding of evaluations at different levels of international NGOs. At the bottom scale of DC system, NGOs and field DC practitioners perceive evaluation as an external action, often in its simplified form of judgement on their work. The main problem here can be resumed as a misunderstanding of the role of these bottom level players in the DC system and consequently as their inability to contribute to the overall learning process.

Mebrahtu (2002) perfectly resumes such a belief by reporting the statement of one DC practitioner in Ethiopia, about his own perception of M&E actions: 'We collect most of the data necessary but we don't see where or how it is used... we write reports, collect them, and pass them on to the sector manager who writes more reports and sends them off-we don't learn anything from this process, then the whole thing starts again...!'

In the same study the author analyses the different perception among the different levels in international NGOs, from the central headquarters through the country main and field offices. Collected information confirms how at the third tier level the evaluation practice is diffusely implemented but poorly understood and maybe inefficiently linked to higher scales. The findings then support Oakley's (1996) general observation that a large gap exists between international NGO assertions that M&E is a necessary and valuable activity and evidence of good quality practice in these areas.

Still concerning the third research direction, also the different timing and perspectives in evaluation create a communication problem and consequent inefficiencies in the result chain: some authors (Smismans, 2015) still point out the need of <u>strengthening the link between ex-ante and ex-post evaluation.</u> The analysis of the reports from the main international organizations shows how different practitioners and academic communities have dealt with ex-ante and ex-post evaluation in a 'siloed' fashion while using similar concepts in different ways and different methods for the same purpose.

This is confirmed for example in the official EU discourse about DC evaluations. Here 'evaluation' normally refers to ex-post (i.e. retrospective) evaluation, which can be interim (i.e. at the mid-term of an initiative), final (at its conclusion), or ex-post in the strict sense (which can take place several years after the intervention has finished). The EU Commission talks about ex-ante evaluation only in relation to expenditure programmes and, if we focus on DC, only in relation to the programme or project selection that happens against a call, or at the needs assessment step in project cycle management. The divide between ex-ante impact assessments and ex-post evaluation is also partially reflected within the organizational structure of the European Commission, both at the level of the Secretariat General and the Directorates General (DGs), and of several other international organizations. This discursive and institutional divide at the international DC provider level reflects a wider gap between communities which deal, on the one hand, with ex-ante assessment and, on the other hand, with ex-post evaluation.

This gap between ex-ante and ex-post perspectives determines for instance that data from ex-post evaluation do not systematically feedback into ex-ante planning. There is neither a systematic cyclical process, nor a broad availability of ex-post data that could feed into new initiatives (Smismans, 2015).

Figure 2.14 shows the DC results chain as a cycle in which the 'Policy' step corresponds to the strategic reflection that brings to the funding guidelines of DC and to all the calls for programmes or projects in DC.

Ex-ante evaluation enters the cycle in few points and namely during the 'Policy' step when strategic reasoning occurs and DC funding institutions allocate resources on different targets that in turns are inspired by the global sustainability framework (nowadays the SDGs).

Another ex-ante evaluation moment is when DC practitioners start the planning or the implementation of programmes or projects. In this phase it is common to implement needs assessments and to draw sound theories of change linking estimated inputs to forecasted outcomes in a new project proposal to be submitted for funding.

Conversely, ex-post evaluation takes place at the end of each step in the cycle and target different aspects of the DC implementation process. As it is shown by the inner arrow-circle, the five OECD-DAC criteria for evaluation of DC cover the different phases of the cycle.

As we already mentioned, ex-ante and ex-post evaluation teams rarely overlaps in timing and purposes so that evaluation cycle is not closed (OECD-DAC, 2013; Smismans, 2015).

As the DC evolves towards a new evaluation approach that encompasses both planning/funding and implementing steps of the DC system, the aim becomes to link ex-ante and ex-post through an improved collaboration between the two scholarly communities (Smismans, 2015) and through the use of sound and shared methodologies among levels.

This means in particular that ex-ante evaluations should gain a new role and be implemented side-by-side with ex-post evaluations. In other words, each time a new policy, programme or project is planned, ex-ante assessment should include the ex-post evaluation and vice versa (an ex-post evaluation should forecast the effect of its own recommendations in a kind of ex-ante assessment).
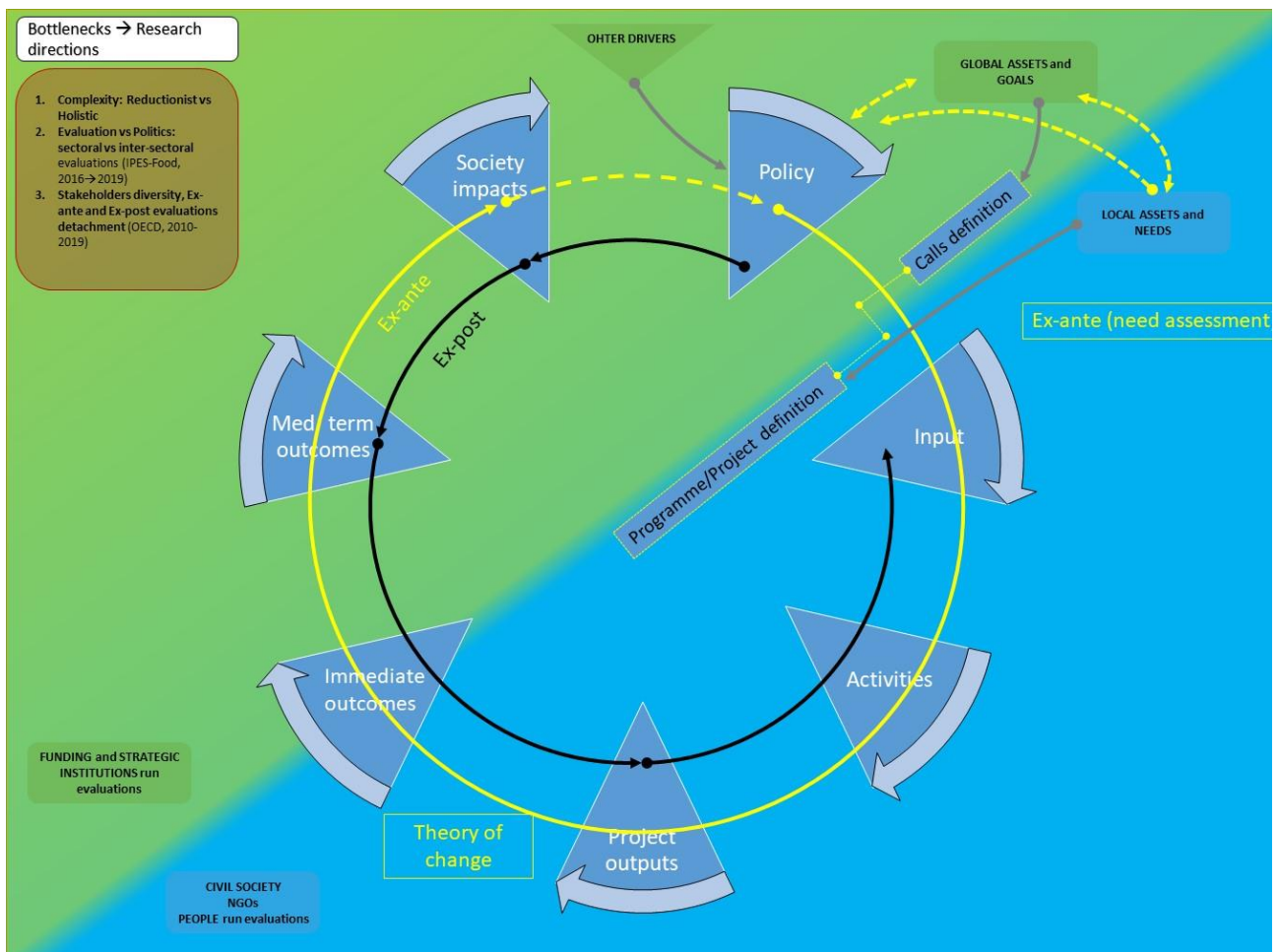
Ex-post evaluation should feed back into the DC ex-ante impact assessments, which actually take place only when the strategic and funding organizations define new funding frameworks or calls for projects.

This issue was clearly resumed by EC in a 2013 Evaluation Communication stating: 'There can be a tendency to look forward and focus on new initiatives. But changes are costly and take time to implement – so they need to be justified and greater attention needs to be paid to look back before moving forward' (EC, 2013).

As we illustrate in Figure 2.15, this can be achieved by locating evaluation firmly within the DC cycle, which, in other words, corresponds to spreading the evaluation culture among DC actors.

Figure 2.16 – Representation of the DC cycle as a spiral in which no round is equal to the previous one. Evaluation culture determines the location of ex-ante and ex-post evaluations between the subsequent DC programming and implementation cycles.
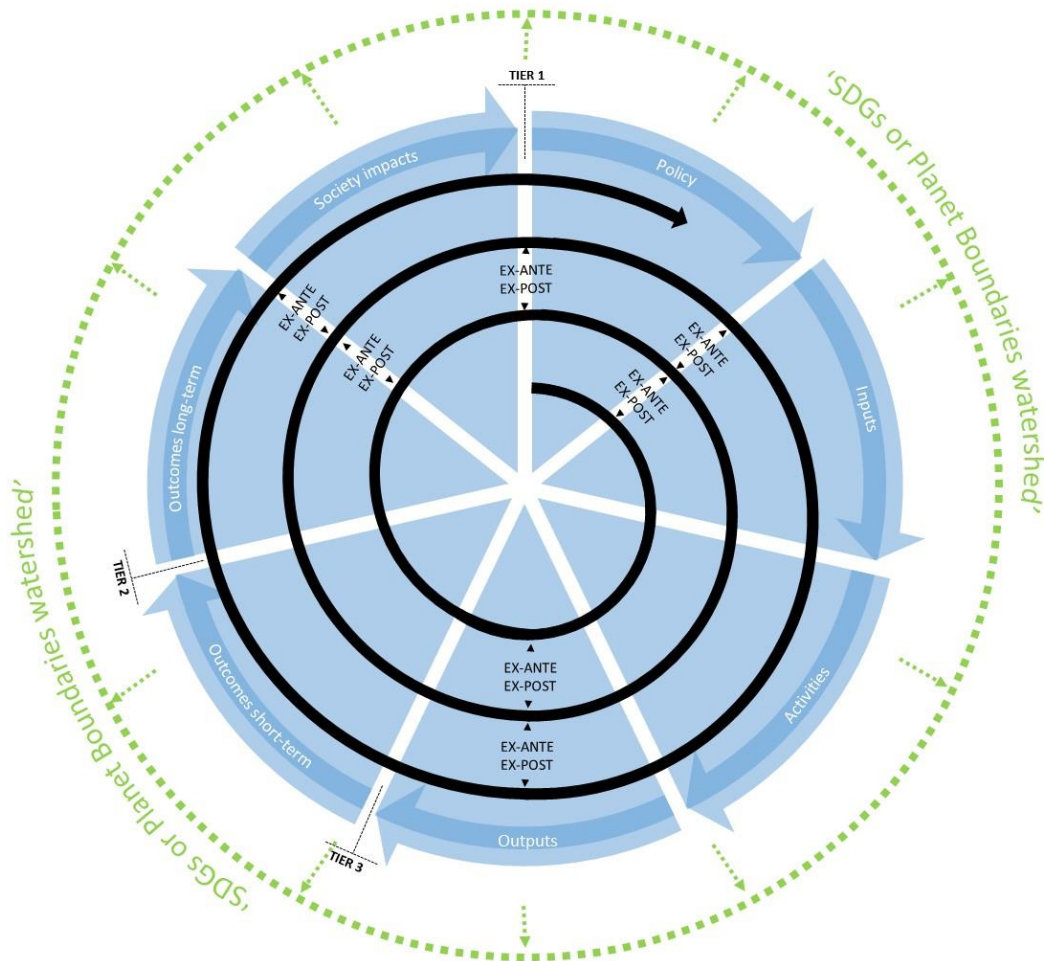
In Figure 2.15 we propose a visualization of the effect of a tight collaboration between ex-ante and ex-post perspectives in evaluating DC: the ex-ante / ex-post synergy would determine the possibility to guide mindfully the spiral trajectory towards the SDGs frontier.

On tiers 2 and 3 (Figure 2.12) the spreading of an evaluation culture could tackle these challenges because:

- it implies the networking and the participation among DC actors, paving the way for the existence of a truly shared and standardized communication protocol for evaluation planning and valorization;
- it implies a reviewed role of DC practitioners with increased consciousness about the effect of their evaluative action

72

- it implies new methodologies that are shared across the different levels and the different perspectives (ex-ante, ad-interim, final, ex-post).

Concluding, in the next chapters we focus on a specific case study where we tackle ex-ante evaluation, but the idea behind our experience is that the use of the proposed mixed-method approach and the applied participatory methodologies could be experimented at different levels of the DC system and with different 'timing' perspectives in order to create new knowledge concerning the third research direction identified so far, namely, 'DC system infrastructure: implementation methodologies and learning systems (focus on diversity of stakeholders and evaluation methods)'.

2.7   Decision aiding in agricultural development

Agricultural sector development has important connections with overall development of regions, and even more if the specific economic context is dependent from the primary sector. The reason behind the importance of agricultural development resides in multi-functionality of agriculture, here understood as the human activity of producing food through a sound relationships with the environment. The dimensions of multi-functionality are interpreted in literature in many different ways, but they always deal with environmental, social and economic aspects (Fagioli et al., 2017). Multi-functionality of agriculture has an impact not only on the economy, the environment and nature, but also on societal and cultural development (Cairol et al., 2009), and constitutes a path towards sustainable development (Caron et al., 2008). The measurement of the multidimensional functionality of the agri-food system is an important issue, as it allows the agricultural value chains to be interpreted under the lens of sustainable development pillars. This stands true both at farm management level and at global policy level.

At local farm or value chain level, literature defines Decision Support Systems (DSS) as mainly computerized systems, which include models and databases and are used in the decision-making process. They are 'tools' that help decision makers in the procedure of decision-making and in choosing the best (economic, social, or environmental) alternative solution. Several scientific sectors supported the development and constituted the necessary background for the effective planning of decision support systems. The Science of Informatics has contributed in the planning and applications of decision support systems with the supply of tools, material, and software. The sciences of Operational Research and Management and Business Administration provide the theoretical frame for the analysis of various decisions. The sciences of Behaviour, Sociology, and the Management of

Human Resources constitute sources of information that concern the manner in which human's potential behave at the handling of information and the decision-making process (Manos et al., 2010).

Scaling up from farm level to global level, the appropriate instrument for a multidimensional representation is a suitable set of indicators, which must be an integral part of an assessment methodology. Multiple criteria decision aiding (MCDA) plays a central role in this multidimensional evaluation process. MCDA is used to solve complex problems by assessing all the criteria (indicators), both individually and collectively, assigning specific importance to each criterion, and aggregating them in order to rank alternative policies for ruling the agri-food value chain (Fagioli et al., 2017).

Indicators, already introduced in paragraph 2.3.5, can pertain to different classes and the methodologies and technologies that are useful for measuring them are increasing rapidly (Brown, 2009).

In the next two paragraph, in order to introduce our contribution to the development of new methodological frameworks for evaluation in DC, we briefly introduce the three main methodologies that will be further described in chapters 4, 5 and 6 and that in our opinion may constitute tool for the spreading of evaluation culture in the DC system.

2.7.1   Remote sensing as a monitoring tool

In the quest for SMART indicators, the European Commission counts satellite earth-observation data and remote sensing as a cost-efficient way of collecting information about agriculture.

Since 1993 the EC DG Agriculture has promoted the use of 'Controls with Remote Sensing' (CwRS) as appropriate control system within the Common Agricultural Policy (CAP). The DG Agriculture ask yearly to the Joint Research Center of the EU (JRC) to check if agricultural area-based subsidies, which amount to more than 25 billion euro yearly, are correctly granted or to monitor crop production by distinguishing, identifying, measuring and forecasting yields for the main crop production areas in Europe.

This kind of technological tools is more and more interesting in the framework of multicriteria decision aiding methodologies, which are eager to build their argumentations and weightings on solid and costs-effective indicators.

Chapter 4 contains our own interpretation of how RS can be used in a mixed-method evaluation in order to collect and analyse value-added information on the agricultural sector of a given region.

2.7.2 Participatory Analytic hierarchy process

Historically, first MCDM methodologies (and MODM – Multi-objective decision making for not-discrete problems) have gained the attention of academia and decision makers because of their ability to inform evidence-based decisions (Mardani et al., 2015). We may define this kind of interest in MCDM as 'top-down' because the main reason for the use of the methodologies is to prove and witness the robustness of the decision maker's choice.

Decision making is tightly related to the rational and irrational Choice Theory (Scott, 2000) which in turns binds the analysis of the decision criteria to the evaluation theories and approaches (Reynolds et al., 2010).

As it happened in the wider sector of Evaluation, during the evolution of decision-making both researchers/developers and users switched their focus from accountability to effectiveness: it now appears clear that the implementation process of both evaluations and decision making have direct consequences on the degree of acceptance and therefore on the potential impact of final results or choices (Kazi, 2003; Pawson, 2002; Pawson et al., 2004; Weiss, 1998, 1982). So, in a time where participation has been proven to improve the implication of stakeholders (FAO, 2003; WFP, 2000), this is the reason why the operational tools that grant the possibility to widen participation in decision making get more and more attention.

A good example can be made on Analytic Hierarchy Process (AHP), Pairwise Ranking (PWR) and on their use as participatory methodologies.

AHP (T.L. Saaty, 1980; 1987; 2008) is one of the leading techniques in the wide family of Multicriteria Decision Making or Aiding (MCDM or MCDA) methodologies (Emrouznejad et al., 2017; Ho et al., 2018; Ishizaka et al., 2011). AHP have been applied widely in several sectors in the last 3 decades. First references are related to Zahedi, F. (1986) and Mohanty, R.P. (1993) works, showing AHP being used predominantly in the manufacturing function within organizations. Contemporary studies by Apostolou, B. and Hassell, J.M. (1993) investigates AHP within the accounting function showing increasing usage overtime. The healthcare industry is another target for AHP application and

Liberatore, M.J. and Nydick, R.L. (2008) reviewed over 50 articles relating AHP to decision making within the medical and healthcare profession.

Despite these examples show how AHP related literature is often very specifically oriented towards some sectors, existing reviews help to summarize main fields of AHP application: Seyhan, S. and Mehpare, T. (2010) shows that AHP has been applied heavily in manufacturing, environmental management and agriculture field, power and energy industry, transportation industry, construction industry and healthcare. To some lower degree AHP has also been applied diffusely to education, logistics, e-business, IT, R&D, telecommunication industry, finance and banking, urban management, defense industry and military, politics, government, marketing, tourism and leisure, sport, archaeology, auditing, and the mining industry.

In all these fields and sectors, AHP has been applied to help solving different kind of problems : (1) project selection (Cheng et al., 2005), (2) location selection (Yang et al., 1997; Eddie et al., 2005), (3) resource allocation (Ramanathan et al., 1995; Braunschweig., 2004; Cheng et al., 2005), (4) risk management (Muhammad A. Mustafa et al., 1991), (5) technology selection, (6) conflict management (Lam et al., 2005), (7) project evaluation and (8) benchmarking (Dey, 2002).

Among these AHP targets, literature witness the diffused use of AHP in resource allocation problems and the wealth of journals available on the subject indicates AHP's popularity as well as its ease of use (Ramanathan et al., 1995; Cheng et al., 2005). Despite the existing high pressure on international organizations to get the best out of the resources they have available and to account for the selection of work axes in several countries, we found no evidence of AHP being applied to resource allocation in the field of development cooperation.

In fact, despite many applications of AHP to real-world cases exist, development cooperation organizations tend to neglect the use of more complex MCDM methods such as AHP (De Marinis et al, 2019) and prefer more operational tools such as simplified pairwise ranking (SPWR) when dealing with the planning and implementation of development projects (DFID, 2002; FAO, 2006; IFAD, 2000; Russel, 1997; UNESCO, 2008; World Bank, 2018).

Chapter 6 contains our own interpretation of how AHP can be modified in order to work as a transformative techniques in ex-ante evaluation of priority interventions on the agricultural sectors.

2.7.3 Simplified Pairwise Ranking

Simplified Pairwise Ranking (SPWR) is a simplified form of the Pair-wise Ranking technique (PWR) (Narayanasamy, 2009): it helps a decision maker to binary pair-compare alternatives and to rank them according to their relative frequency of preference. Also called Priority Grid or Matrix, PWR provides a rational and structured approach to derive the group's priorities, whilst facilitating much negotiation and consensus building by the group (FAO, 2006a).

SPWR therefore consists in the use of binary pair-wise comparisons as stand-alone tools. While PWR is normally used in the framework of AHP by asking to the Decision Makers (DMs) 'how many times A is more important than B?', SPWR simply asks 'What is better, A or B?'. Moreover, while AHP deconstructs the decision into a hierarchy and builds PCMs for each level and node, SPWR simply compare alternatives against a global goal. This is the reason why it is also called binary PWR or simplified PWR.

SPWR is widely used by Development cooperation practitioners in order to evaluate and rank local resources, problems, solutions and impacts of development interventions worldwide in the sector of agricultural development (DFID, 2002; FAO, 2006a; IFAD, 2000; Russel, 1997; UNESCO, 2008; World Bank, 2018). Literature witnesses that SPWR, and also its 'grandfather' AHP among others techniques and tools (Aznar et al.,2011; Gupta et al., 2016; Ishizaka et al., 2011; Keeney et al., 2011), can be used with a bottom-up and participative approach, working as a consensus-building tool, bringing the group of stakeholders to discuss on single criteria and weights and to understand each other point of view (Aznar et al., 2011; Russo et al., 2015a). Practitioners of development project in agriculture are interested in practical and rapid assessments of needs which might be shared and accepted by the local community. In this context, a good degree of consensus is an important outcome as much as the consistency of the resulting list of items. Here is to be found the reason why the SPWR is preferred: the technique grants the possibility to run rapid brainstorming and focus groups with all the interested stakeholders and to produce participatory outcomes in terms of a ranked lists of preferences, might they be needs or priorities.

To our knowledge, few previous studies explicitly deal with SPWR applied to agricultural development, with participatory approach within the framework of international development initiatives. These scientific contributions deal with three topics, namely the participatory selection of crop varieties (Rajula Shanthy et al., 2010), the use of participatory rural appraisal technique

(Koralagama et al., 2007) and the livestock/veterinary sector assessment (Catley et al., 1996; Kebede et al., 2003; Stroebel et al., 2008).

Chapter 5 contains the description of the implementation of SPWR in a specific case study.

2.8   References

Adelle, C., Jordan, A., & Turnpenny, J. (2012). Proceeding in Parallel or Drifting Apart? A Systematic Review of Policy Appraisal Research and Practices. *Environment and Planning C: Government and Policy*, *30*(3), 401–415. https://doi.org/10.1068/c11104

American Evaluation Association. (2007). *Systems concepts in evaluation*. (B. Williams & I. Imam, Eds.), *Systems concepts in evaluation: An expert anthology*. Inverness: EdgePress.

Apostolou, B. and Hassell, J.M. (1993), An overview of the analytic hierarchy process and its use in accounting research, *Journal of Accounting Literature*, 12(1), 1–28.

Arnold, R. D., & Wade, J. P. (2015). A definition of systems thinking: A systems approach. In *Procedia Computer Science* (Vol. 44, pp. 669–678). https://doi.org/10.1016/j.procs.2015.03.050

Aznar, J., Guijarro, F., & Moreno-Jiménez, J. M. (2011). Mixed valuation methods: A combined AHP-GP procedure for individual and group multicriteria agricultural valuation. *Annals of Operations Research*, *190*(1), 221–238. https://doi.org/10.1007/s10479-009-0527-2

Bacon, F. (1620). *Novum organum*. London: John Bill.

Behrman, J. R. (2010). The International Food Policy Research Institute (IFPRI) and the Mexican PROGRESA Anti-Poverty and Human Resource Investment Conditional Cash. *World Development*, *38*(10), 1473–1485. https://doi.org/10.1016/j.worlddev.2010.06.007

Best, A., & Holmes, B. (2010). Systems thinking, knowledge and action: Towards better models and methods. *Evidence and Policy*, *6*(2), 145–159. https://doi.org/10.1332/174426410X502284

Better Evaluation. (2014). Rainbow Framework. Retrieved from https://www.betterevaluation.org/en/rainbow_framework

Bezzi, C. (2010). *Il nuovo disegno della ricerca valutativa*. Milan: Franco Angeli.

Bhatkal, T., Samman, E., & Stuart, E. (2015). *Leave no one behind - the real bottom billion*. London: Overseas Development Institute.

BOND. (2019). BOND evaluation methods choice tool. Retrieved August 15, 2019, from https://www.bond.org.uk/resources/evaluation-methods-tool

Braunschweig, T., & Becker, B. (2004). Choosing research priorities by using the analytic hierarchy process: an application to international agriculture. *R&D Management*, *34*(1), 77–86. https://doi.org/10.1111/j.1467-9310.2004.00324.x

Brown, D. (2009). Good Practice Guidelines for Indicator Development and Reporting. In *Third World Forum on 'Statistics, Knowledge and Policy' Charting Progress, Building Visions, Improving Life* (p. 8). Busan.

Butland, B., Jebb, S., Kopelman, P., McPherson, K., Thomas, S., Mardell, J., & Parry, V. (2007). *Foresight. Tackling Obesities: Future Choices – Project report*. *Government Office for Science*. London: Department

of Innovation Universities and Skills. https://doi.org/https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/287937/07-1184x-tackling-obesities-future-choices-report.pdf

Cabrera, D., & Colosi, L. (2008). Distinctions, systems, relationships, and perspectives (DSRP): A theory of thinking and of things. *Evaluation and Program Planning*, *31*(3), 311–317. https://doi.org/10.1016/j.evalprogplan.2008.04.001

Cairol, D., Coudel, E., Knickel, K., Caron, P., & Kröger, M. (2009). Multifunctionality of Agriculture and Rural Areas as Reflected in Policies: The Importance and Relevance of the Territorial View. *Journal of Environmental Policy & Planning*, *11*(4), 269–289. https://doi.org/10.1080/15239080903033846

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. *Information Processing* (Vol. 20). Boston: Houghton Mifflin Company. https://doi.org/10.1016/0306-4573(84)90053-0

Campbell, D. T., & Stanley, J. C.(1966). *Experimental and quasi-experimental designs for research.* Skokie: Rand McNally.

Campbell, D., & Cook, T. (1979). *Quasi-experimentation: Design and analysis for field settings*. Chicago: Rand McNally. https://doi.org/10.1207/s15327752jpa4601_16

Caron, P., Reig, E., Roep, D., Hediger, W., Le Cotty, T., Barthélémy, D., Sabourin, E. (2008). Multifunctionality: refocusing a spreading, loose and fashionable concept for looking at sustainability? *International Journal of Agricultural Resources, Governance and Ecology*, *7*(4–5), 301–318.

Cassen, R., & Associates (1986). *Does aid work?* (1st ed.)*.* Oxford: Oxford University Press.

Catley, A. P., & Mohammed, A. A. (1996). The use of livestock-disease scoring by a primary animal-health project in Somaliland. *Preventive Veterinary Medicine*, *28*(3), 175–186. https://doi.org/https://doi.org/10.1016/0167-5877(96)01038-0

CDC - Center for Disease Control and Prevention. (1999). Framework for program evaluation in public health. https://doi.org/10.1016/j.cardfail.2009.03.006

Chambers, R. (2005) *Ideas for Development*, London and Sterling: Earthscan.

Chambers, R., Pacey, A., & Thrupp, L. A. (1989). *Farmer First*. Rugby: Practical Action Publishing. https://doi.org/doi:10.3362/9781780440149

Chen, H. T. (2016). Interfacing theories of program with theories of evaluation for advancing evaluation practice: Reductionism, systems thinking, and pragmatic synthesis. *Evaluation and Program Planning*, *59*, 109–118. https://doi.org/10.1016/j.evalprogplan.2016.05.012

Chen, H., & Rossi, P. H. (1983). Theory-Driven Approach ~ methodology. *Evaluation Review*, *7*(3), 283–302. https://doi.org/10.1177/0193841X8300700301

Cheng, E. W. L., & Li, H. (2005). Analytic Network Process Applied to Project Selection. *Journal of Construction Engineering and Management*, *131*. https://doi.org/10.1061/(ASCE)0733-9364(2005)131:4(459)

Cheng, E.W.L. (2005). The analytic network process (ANP) approach to location selection: a shopping mall illustration. *Construction Innovation*, *5*(2), 83–97. https://doi.org/10.1108/14714170510815195

Clements, P. (1996). *Development as if impact mattered: A comparative organizational analysis of USAID, The World Bank and Care based on case studies of projects in Africa.* Ann Arbor: A Bell & Howell Company.

Conover, P. J. (1985). The Impact of Group Economic Interests On Political Evaluations. *American Politics Quarterly*, *13*(2), 139–166. https://doi.org/10.1177/1532673X8501300201

Cracknell, B. E. (2000). *Evaluating development aid.* Thousand Oaks: Sage.

DFID. (2002). *Monitoring and Evaluation Series - Field Guide for Conducting Participatory Impact Assessment*. (L. Oruko, Ed.). London: Client-oriented Agricultural Research and Dissemination Project.

DFID. (2013). *International Development Evaluation Policy*. London: Department for International Development.

Dhaouadi, M. (1990). Ibn Khaldun: the founding father of eastern sociology. *International Sociology*, *5*(3), 319–335. https://doi.org/10.1177/026858090005003007

Dorfman, R. (Ed.) (1963). *Measuring benefits of government investments*. Washington: The Brookings Institution.

EC - European Commission. (2013). *Evalsed—The Resource for the Evaluation of Socio-Economic Development*. Brussels: European Commission.

EC - European Commission. (2014). *Evaluation matters : The evaluation policy for European Union development co-operation*. Brussels, EU: European Commission.

EC. (2013). *Communication from the Commission to the European parliament, the council, the European economic and social committee and the committee of the regions. Strengthening the foundations of Smart Regulation – improving evaluation*. Brussels: European Commission.

Eggers, H. W. (2002). Project Cycle Management: A personal reflection. *Evaluation 8(4),496-504.*

Emrouznejad, A., & Marra, M. (2017). The state of the art development of AHP (1979–2017): A literature review with a social network analysis. *International Journal of Production Research*, *55*(22), 6653–6675. https://doi.org/10.1080/00207543.2017.1334976

EUROPAID. (2019). Results and indicators. Retrieved December 20, 2019 from https://ec.europa.eu/europeaid/sites/devco/files/19. results-and-indicators-10122013_en.pdf

European Commission. (2015). *The Monitoring and Evaluation Framework for the Common Agricultural Policy*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2762/5243

European Commission. (2017). *A Social Multi-Criteria Framework for ex-ante Impact Assessment*. (G. Munda, Ed.). Luxembourg: Publications Office of the European Union. https://doi.org/10.2760/909528

EvalPartners. (2016). *No one left behind. Evaluating SDGs with an equity-focused and gender-responsive lens. Report of the 2016 New York High-Level Panel and Technical Seminar*. (A. Kalugampitiya, N. Raaber, A. Ocampo, & F. Tateossian, Eds.). New York: UN WOMAN Independent Evaluation Office.

Fagioli, F., Rocchi, L., Paolotti, L., Słowiński, R., & Boggia, A. (2017). From the farm to the agri-food system: A multiple criteria framework to evaluate extended multi-functional value. *Ecological Indicators*, *79*, 91–102. https://doi.org/https://doi.org/10.1016/j.ecolind.2017.04.009

FAO. (2003). *Guidelines on Beneficiary Participation in Agricultural and Rural Development*. (B. van Heck, Ed.) (2nd ed.). Rome: Rural Institutions and Participation Service, Rural Development Division Food and Agriculture Organization of the United Nations.

FAO. (2006). *Participatory Rural Appraisal (PRA) Manual*. (R. Paul, Ed.). Rome: Food and Agriculture Organization of the United Nations.

FAO. (2006a). *Toolkit for the Application of Participatory and Negociated Territorial Development*. Rome: Food and Agriculture Organization of the United Nations.

Fitz-Gibbon, C. T., & Morris, L. L. (1996). Theory-based evaluation. *Evaluation Practice*, *17*(2), 177–184. https://doi.org/10.1016/S0886-1633(96)90024-0

Fitzpatrick, J. L. (2012). An introduction to context and its role in evaluation practice. *New Directions for Evaluation*, *2012*(135), 7–24. https://doi.org/10.1002/ev.20024

Fitzpatrick, J., Sanders, J., & Worthen, B. (2004). Setting boundaries and analyzing the evaluation context. In *Program evaluation: Alternative approaches and practical guidelines* (pp. 199–229).

Flygare, E., Gill, P. E., & Johansson, B. (2013). Lessons from a concurrent evaluation of eight antibullying programs used in Sweden. *American Journal of Evaluation*, *34*(2), 170–189. https://doi.org/10.1177/1098214012471886

Forrester, J. W. (1961). *Industrial dynamics*. Cambridge: MIT Press.

Frandberg, T. (2003). Systems Thinking - A Studie of Alternatives of R. Flood, M. Jackson, W. Ulrich, G. Midgley. *International Journal of Computers, Systems and Signals*, *4*(1), 33–41.

Furubo, J.-E., Rist, R. C., & Sandahl, R. (2002). *International atlas of evaluation Jan-Eric Furubu, Ray C. Rist, Rolf Sandahl, editors.* New Brunswick; London: Transaction Publishers.

Geoffrey, D. M. (1970). *A new dictionary of sociology*. London: Routledge & Kegan Paul Limited.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2016). *Impact evaluation in practice - second edition*. Washington DC: International Bank for Reconstruction and Development / The World Bank. https://doi.org/10.1596/978-1-4648-0779-4

Greenberg, D. H., & Morris, S. (2005). Large-scale social experimentation in Britain: What can and cannot be learnt from the Employment Retention and Advancement Demonstration? *Evaluation*, *11*(2), 223–242. https://doi.org/10.1177/1356389005055529

Gupta, M., & Mohanty, B. K. (2016). An algorithmic approach to group decision making problems under fuzzy and dynamic environment. *Expert Systems with Applications*, *55*, 118–132. https://doi.org/10.1016/j.eswa.2016.02.002

Guyot, W. M. (1978). Summative and Formative Evaluation. *The Journal of Business Education*, *54*(3), 127–129. https://doi.org/10.1080/00219444.1978.10534702

Hayes, S. P. (1959). *Evaluating development projects.* Belgium: UNESCO.

Hirschman, A. O. (1967). *Development projects observed.* Washington: Brookings Institution.

Ho, W., & Xin, M. (2018). The state-of-the-art integrations and applications of the analytic hierarchy process. *European Journal of Operational Research*, *267*, 399–414. https://doi.org/10.1016/j.ejor.2017.09.007

IEG-WB. (2009). *2009 Annual Review of Development Effectiveness*. Washington: The World Bank. https://doi.org/10.1596/978-0-8213-8135-9

IFAD. (2000). *Promoting Participation in Development Projects: IFAD's Experience in Promoting Participatory Approaches in Development Projects in Asia*. (M. T. L. Debuque, Ed.). Somewhere: IFAD publishing services.

IFAD. (2015). *Evaluation Manual - second edition*. (Independent Office of Evaluation of IFAD (IOE), Ed.). Rome: Palombi & Lanci Tipografia srl.

Ishizaka, A., & Labib, A. (2011). Review of the main developments in the analytic hierarchy process. *Expert Systems with Applications*, *38*(11), 14336–14345. https://doi.org/10.1016/j.eswa.2011.04.143

Jiaqin, Y. (1997). An AHP decision model for facility location selection. *Facilities*, *15*(9/10), 241–254. https://doi.org/10.1108/02632779710178785

Keeney, R. L., & Nau, R. (2011). A theorem for Bayesian group decisions. *Journal of Risk and Uncertainty*, 43 (1), 1–17. https://doi.org/10.1007/s11166-011-9121-5

Kellog Foundation. (2017). *Step-by-step guide to evaluation*. Retrieved from https://www.wkkf.org/resource-directory/resource/2017/11/wk-kellogg-foundation-step-by-step-guide-to-evaluation

Kemeny, J. G., & Oppenheim, P. (1956). On reduction. *Philosophical Studies*, *7*(1), 6–19. https://doi.org/10.1007/BF02333288

Kim, P. S. (2002). Cultural Change in Government: Promoting a High-Performance Culture. A Review of Ten Years of Modernisation. In *The HRM Perspective Human Resources Management (HRM) Working Party Meeting*. Paris: OECD Publishing.

Koralagama, D. N., Wijeratne, M., & De Silva, W. N. (2007). Emergence of participatory rural appraisal (PRA) technique as a strategy towards sustainable development: A Sri Lankan experience. *Journal of Agriculture and Rural Development in the Tropics and Subtropics*, 108(2), 149–160.

Krutilla, J. V., & Eckstein, O. T. (1964). M*ultiple purpose river development; studies in applied economic analysis.* Baltimore: The John Hopkins Press.

Kumar, D. P. (2002). Benchmarking project management practices of Caribbean organizations using analytic hierarchy process. *Benchmarking: An International Journal*, *9*(4), 326–356. https://doi.org/10.1108/14635770210442680

Kumar, K. (1987). *Rapid, low-cost data collection methods for A.I.D.* Washington: USAID.

Kumar, K. (1989). *Conducting key informant interview in developing countries.* Washington: USAID.

Lam, P.-K., & Chin, K.-S. (2005). Identifying and prioritizing critical success factors for conflict management in collaborative new product development. *Industrial Marketing Management*, *34*(8), 761–772. https://doi.org/https://doi.org/10.1016/j.indmarman.2004.12.006

Leischow, S. J., Best, A., Trochim, W. M., Clark, P. I., Gallagher, R. S., Marcus, S. E., & Matthews, E. (2008). Systems Thinking to Improve the Public's Health. *American Journal of Preventive Medicine*. https://doi.org/10.1016/j.amepre.2008.05.014

Lewis, A. (1951). *Measures for the economic development of under-developed countries.* New York: The United Nations.

Liberatore, M. J., & Nydick, R. L. (2008). The analytic hierarchy process in medical and health care decision making: A literature review. *European Journal of Operational Research*, *189*(1), 194–207. https://doi.org/https://doi.org/10.1016/j.ejor.2007.05.001

LINK, & Social Value Italia. (2007). The impact of international development cooperation. A methodological proposal. Retrieved December 9, 2019, from https://www.cosv.org/wp-content/uploads/2019/04/EVALUATING-THE-IMPACT-OF-INTERNATIONAL-DEVELOPMENT-COOPERATION.pdf

Manos, B. D., Papathanasiou, J., Bournaris, T., & Voudouris, K. (2010). A DSS for sustainable development and environmental protection of agricultural regions. *Environmental Monitoring and Assessment*, *164*(1–4), 43–52. https://doi.org/10.1007/s10661-009-0873-1

Mardani, A., Jusoh, A., Nor, K. M. D., Khalifah, Z., Zakwan, N., & Valipour, A. (2015). Multiple criteria decision-making techniques and their applications - A review of the literature from 2000 to 2014. *Economic Research-Ekonomska Istrazivanja*, *28*(1), 516–571. https://doi.org/10.1080/1331677X.2015.1075139

Martini, A., & Sisti, M. (2009). *Valutare il successo delle politiche pubbliche*. Bologna: Il Mulino.

Maynard, R., Randall, B., Schore, J., Jackson, R., Leach, V., & Wooldridge, J. (1979). *Supported Work Demonstration: Effects During the First 18 Months after Enrollment*. (M. D. R. CORPORATION, Ed.). Princeton: Mathematica Policy Research.

Mayne, J. (2008). *Building an Evaluative Culture for Effective Evaluation and Results Management* (ILAC Worki). Rome: Institutional Learning and Change Initiative.

McIntosh, C. (2008). Estimating Treatment Effects from Spatial Policy Experiments: An Application to Ugandan Microfinance. *Review of Economics and Statistics*, *90*(1), 15–28. https://doi.org/10.1162/rest.90.1.15

McKean, R. N. (1966). *Efficiency in government through systems analysis.* Hoboken: Wiley Publications

Mebrahtu, E. (2002). Perceptions and Practices of Monitoring and Evaluation: International NGO Experiences in Ethiopia. *Development in Practice*, *12*(3/4), 501–517.

Mertens, D. (2005). The Inauguration of the International Organization for Cooperation in Evaluation. *American Journal of Evaluation*, *26*, 124–130. https://doi.org/10.1177/1098214004273186

Mertens, D. M. (2012). What Comes First? The Paradigm or the Approach? *Journal of Mixed Methods Research*, *6*(4), 255–257. https://doi.org/10.1177/1558689812461574

Mertens, D. M., & Wilson, A. T. (2019). *Program evaluation theory and practice*. New York: THE GUILFORD PRESS.

Mingers, J., & White, L. (2010). A review of the recent contribution of systems thinking to operational research and management science. *European Journal of Operational Research*, *207*(3), 1147–1161. https://doi.org/10.1016/j.ejor.2009.12.019

Mishan, E. J. (1971). *Cost-benefit analysis,* London: Unwin University Books.

Mitchell, G. D. G. D. (1970). *Sociology: the study of social systems*. London: University Tutorial Press.

Mohanty, R., & Deshmukh, S. G. (1993). Use of AHP for evaluation of Sources. *International Journal of Physical Distribution & Logistics Management*, *23*, 22–28.

Morris, L. L., & Fitz-Gibbon, C. T. (1978). *How to Measure Program Implementation. Program Evaluation Kit, 4.* (Center for the Study of Evaluation, Ed.). Beverly Hills: SAGE Publications Inc.

Muhammad A.Mustafa, & Jamal F.Al-Bahar. (1991). Project risk assessment using the analytic hierarchy process. *IEEE Transactions on Engineering Management*, *38*(1), 46–52. https://doi.org/10.1109/17.65759

Narayanasamy, N. (2009). *Participatory Rural Appraisal: Principles, Methods and Application*. New Delhi: SAGE Publications India Pvt Ltd. https://doi.org/10.4135/9788132108382

Newman, I., & Ridenour, C. (1998). Qualitative Quantitative Research Methodology Exploring the Interactive Continuum. *Educational Leadership Faculty Publications*, *122*.

OECD. (1986). *Methods and procedures in aid evaluation.* Paris: OECD.

OECD. (1996). *The story of official development assistance. A history of the development assistance committee and the development co-operation directorate in dates, names and figures*. (H. Fuhrer, Ed.). Paris: OECD Pubblications.

OECD. (2005). Paris Declaration and Accra Agenda for Action. Retrieved July 25, 2019, from https://www.oecd.org/dac/effectiveness/parisdeclarationandaccraagendaforaction.htm

OECD. (2011). The Busan Partnership for Effective Development Co-operation. Retrieved July 20, 2019, from https://www.oecd.org/development/effectiveness/busanpartnership.htm

OECD. (2014). *Measuring and managing results in development co-operation A review of challenges and practices among DAC members and observers*. Paris: OECD Pubblications.

OECD. (2016). *Evaluation Systems in Development Co-operation*. Paris: OECD Publishing. https://doi.org/https://doi.org/https://doi.org/10.1787/9789264262065-en

OECD. (2017). *Strengthening the results chain: Synthesis of case studies of results-based management by providers*. (R. Zwart, Ed.), *OECD Development Policy Papers*. Paris: OECD Publishing. https://doi.org/doi:10.1787/544032a1-en

OECD. (2017). *Strengthening the results chain: Synthesis of case studies of results-based management by providers*. (R. Zwart, Ed.), *OECD Development Policy Papers*. Paris: OECD Publishing. https://doi.org/doi:10.1787/544032a1-en

OECD. (2019). *Measuring Distance to the SDG Targets 2019 An Assessment of Where OECD Countries Stand*. *Multilingual Summaries*. Paris: OECD Publishing. https://doi.org/10.1787/e0f4d2ac-en

OECD. (2019). Whati is Impact Assessment? Retrieved December 14, 1BC, from https://www.oecd.org/sti/inno/What-is-impact-assessment-OECDImpact.pdf

OECD-DAC. (2000). *RESULTS BASED MANAGEMENT IN THE DEVELOPMENT CO-OPERATION AGENCIES: A REVIEW OF EXPERIENCE*. (A. Binnendijk, Ed.). Paris: DAC Working Party on Aid Evaluation.

OECD-DAC. (2010). *Evaluating development co-operation, summary of key norms and standards*. (OECD-DAC network on development evaluation, Ed.) (Second Ed.). Paris: OECD Publishing. https://doi.org/10.1007/BF02927277

OECD-DAC. (2010). *Glossary of Evaluation and Results Based Management (RBM) Terms*. *Working Party on Aid Evaluation, OECD*. Paris Cedex: OECD Publishing.

OECD-DAC. (2013). *Evaluating Development Activities: 12 lessons from the OECD-DAC*. Paris: OECD Publishing.

Osborne, D., & Gaebler, T. (1993). *Reinventing government: How the entrepreneurial spirit is transforming the public sector.* New York: Perseus Books.

Owen, J. M., & McDonald, D. E. (1999). Creating an evaluation culture in international development cooperation agencies. *Journal of International Cooperation in Education*, *2*(2), 41–53.

Patton, M. Q. (2010). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. London: Guilford Publications. https://doi.org/10.1002/casp

Pawson, R. (2002). Evidence-based Policy: In Search of a Method. *Evaluation*, *8*(2), 157–181. https://doi.org/10.1177/1358902002008002512

Pawson, R., & Tilley, N. (1994). What works in evaluation research? *British Journal of Criminology*, *34*(3), 291–306. https://doi.org/10.1525/sp.2007.54.1.23.

Pawson, R., & Tilley, N. (1997). *Realist Evaluation*. London: SAGE Publications Ltd.

Pawson, R., & Tilley, N. (2004). Realist evaluation. In *Pawson, R., & Tilley, N. (1997). Realist Evaluation.* (pp. 1–36). London: SAGE Publications.

Peter Rossi, Howard E. Freeman e Mark W. Lipsey, Evaluation. A systematic approach, 6th ed., Sage, Thousand Oaks, CA. 1999, pp. 67-70.

Picciotto, R. (2009). Evaluating development: Is the country the right unit of account? In M. Segone (Ed.), *Country-led monitoring and evaluation systems* (pp. 32–55). Geneve: UNICEF.

Rajula Shanthy, T. (2010). Participatory varietal selection in sugarcane. *Sugar Tech*, *12*(1), 1–4. https://doi.org/10.1007/s12355-010-0001-0

Ramanathan, R., & Ganesh, L. S. (1995). Using AHP for resource allocation problems. *European Journal of Operational Research*, *80*(2), 410–417. https://doi.org/https://doi.org/10.1016/0377-2217(93)E0240-X

Rawlings, L. (2003). *Place randomized trials in Latin America.* Campbell Collaboration.

Reynolds, M., & Sue, H. (2010). *Systems Approaches to Managing Change: A Practical Guide*. (M. Reynolds, Ed.). Springer London. https://doi.org/10.1007/978-1-84882-809-4

Rog, D. J., Fitzpatrick, J. L., & Conner, R. F. (2012). Editors' notes. *New Directions for Evaluation*, *2012*(135), 1–6. https://doi.org/10.1002/ev.20023

Rossi, P. H., Freeman. H. E., & Wright, J. D. (1980) *Doing evaluation.* Paris: UNESCO.

Rossi, P. H., Freeman. H. E., & Wright. J. D. (1978). *Evaluation: A systematic approach* (1st ed.). Thousand Oaks: Sage.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Chapters 1-5*. In *Evaluation: A Systematic Approach* (7th ed.). Thousand Oaks: SAGE Publications.

Roxana Melenciuc, I. (2015). The emergence of an evaluation culture in the EU: between international practices and Europeanization pressures. *Europolity-Continuity and Change in European Governance*, *9*(1), 137–160.

Russel, T. (1997). Pair Wise Ranking Made Easy. In *PLA Notes* (28, pp. 25–26). London: International Institute for Environment and Development.

Russo, R. D. F. S. M., & Camanho, R. (2015). Criteria in AHP: A systematic review of literature. *Procedia Computer Science*, *55*(July), 1123–1132. https://doi.org/10.1016/j.procs.2015.07.081

Saaty, T. L. (1980). The Analytic Hierarchy Process. *Education*, 1–11. https://doi.org/10.3414/ME10-01-0028

Saaty, T. L. (1987). The analytic hierarchy process-what it is and how it is used. *Mathematical Modelling*, *9*(3–5), 161–176. https://doi.org/10.1016/0270-0255(87)90473-8

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83. https://doi.org/10.1504/IJSSCI.2008.017590

Save the Children. (2019). SMART indicators. Retrieved December 20, 2019, from https://sites.google.com/site/savethechildrendme/Home/smart-indicators

Sasaki, R. (2006). A Review of the History and the Current Practice of Aid Evaluation. *Journal of MultiDisciplinary Evaluation*, *3*(5), 55–88.

Scott, J. (2000). Rational choiche theory. In G. Browning, A. Halcli, & F. Webster (Eds.), *Understanding Contemporary Society: Theories of the Present*. London: SAGE Publications Ltd. https://doi.org/10.4135/9781446218310

Scriven, M. (1966). *The methodology of evaluation*. Lafayette: Purdue University.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental & Quasi-Experimental Designs for Generalised Causal Inference*. Boston: Houghton Mifflin Company. https://doi.org/10.1016/j.evalprogplan.2004.01.006

Sipahi, S., & Timor, M. (2010). The analytic hierarchy process and analytic network process: an overview of applications. *Management Decision*, *48*(5–6), 775–808. https://doi.org/10.1108/00251741011043920

Smismans, S. (2015). Policy Evaluation in the EU: The Challenges of Linking Ex-ante and Ex-post Appraisal. *European Journal of Risk Regulation*, *6*(1), 6–26. https://doi.org/DOI: 10.1017/S1867299X00004244

Stroebel, A., Swanepoel, F. J. C., Nthakheni, N. D., Nesamvuni, A. E., & Taylor, G. (2008). Benefits obtained from cattle by smallholder farmers: a case study of Limpopo Province, South Africa. *Australian Journal of Experimental Agriculture*, *48*(6–7), 825–828. https://doi.org/10.1071/EA08058

Stufflebeam, D. L. (2002). *Evaluation Models*. (D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan, Eds.) (Vol. 49). Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47559-6

Tilley, N., & Pawson, R. (2000). Realistic Evaluation: An Overview. *The British Journal of Sociology*, *49*(September), 331. https://doi.org/10.2307/591330

Tyler, R. W. (1942). General statement on evaluation. *Journal of Educational Research*, *35*(7), 492–501. https://doi.org/10.1080/00220671.1942.10881106

Ulrich, W. (1987). Critical heuristics of social systems design. *European Journal of Operational Research*, *31*(3), 276–283. https://doi.org/10.1016/0377-2217(87)90036-1

UNDP-IEO. (2019). *UNDP Evaluation Guidelines*. New York: Independent Evaluation Office of UNDP.

UNEG. (2016). *Norms and Standards for Evaluation*. (UNEG - United Nations Evaluation Group, Ed.), *UNEG*. New York: United Nations Evaluation Group. https://doi.org/10.1016/j.scitotenv.2016.11.020

UNESCO. (2008). *Data Requirements for Integrated Urban Water Management*. (T. Fletcher & A. Deletic, Eds.). London: CRC Press.

UNICEF. (2014). *Overview: Data Collection and Analysis Methods in Impact Evaluation*. (G. Peersman, Ed.). Florence: United Nations Children's Fund.

USAID. (2013). *Conducting Mixed-Method Evaluations*. Washington: Bureau of Policy, Planning and Learning of USAID.

Valadez, J., & Bamberger, M. (1994) *Monitoring and evaluating social programs in developing countries: A handbook for policymakers, managers, and researchers. W*ashington: The World Bank.

Weiss, C. H. (1982). Policy Research in the Context of Diffuse Decision Making. *The Journal of Higher Education*, *53*(6), 619–639. https://doi.org/10.1080/00221546.1982.11780501

Weiss, C. H. (1993). Where politics and evaluation research meet. *Evaluation Practice*, *14*(1), 93–106. https://doi.org/https://doi.org/10.1016/0886-1633(93)90046-R

Weiss, C. H. (1998). Have We Learned Anything New About the Use of Evaluation? *American Journal of Evaluation*, *19*(1), 21–33. https://doi.org/10.1177/109821409801900103

WFP. (2000). PARTICIPATORY APPROACHES. In D. Spearman (Ed.), *Executive Board Third Regular Session* (p. 18). Rome.

WHO. (2008). Choosing evaluation methods. In *Evaluating household energy and health interventions. A catalogue of methods* (pp. 35–44). Geneva: WHO Press.

WHO. (2009). *Systems thinking for health systems strengthening*. (D. De Savigny & T. Adam, Eds.), *Alliance for Health Policy and Systems Research*. Geneve: WHO Press. https://doi.org/10.1155/2010/268925

Wholey, J. S. (2010). *Handbook of practical program evaluation*. (J. S. Wholey, H. P. Hatry, & K. E. Newcomer, Eds.), *Essential texts for nonprofit and public leadership and management*. San Francisco: Jossy-Bass.

World Bank. (1992). *Effectiveness implementation: Key to development impact* (The Wapenhans Report). Washington: The World Bank.

World Bank. (2018). World Bank - Participatory Rural Appraisal (PRA) Techniques. Retrieved May 24, 2019, from http://go.worldbank.org/EU99NNIVS0

Zahedi, F. (1986). The Analytic Hierarchy Process: A Survey of the Method and Its Applications. *Interfaces*, *16*(4), 96–108.

3 Case study

3.1 Evaluation of agricultural potential in the Diocese of Goma, Nord Kivu, DRC

The framework for this case study is provided by a 3-years EU funded project called ARDST 'Appui au retour de réfugiés et déplacés par le biais de la sécurisation de terres en Diocèse de Goma', led by Caritas Development Goma NGO. The project started in February 2016 and ended in June 2019. Another partner of the project is the Rwandan NGO 'Les héritiers de terre'. The ARDST project is part of the EU strategy for peace seeking and keeping operations in the African Great Lakes region and contributes to the implementation of the Protocol on the Property Rights of Returnee signed at the International Conference of Great Lakes Region (ICGLR) on November 30, 2006. ARDST project's core is the sensitization and negotiation process aiming at fostering refugees return in DRC but, given the important role of agricultural sector for the local economy, agricultural development is mentioned as the main driver for fostering sustainable return and settlement of refugees in a socioeconomic system already hosting settled population.

In this context, in 2017, our DISAA research group was asked to lead a strategic sectoral evaluation in the framework of the ARDST project. The assessment of potentials is considered necessary in the first place to describe the 'state of the art' and to proceed with the identification of potential agri-food development interventions in the diocese of Goma (Goma Diocese or GD).

The aim of the evaluation was to inform the choice of interventions meant to trigger agricultural sustainable development in the Diocese of Goma. Therefore, our case study is about ex-ante evaluation and decision-making, as it appears in the workflow scheme in Figure 3.3 in paragraph 3.1.2. During the phase of evaluation design, a team was formed in order to carry out the evaluation process. The team was made by the 14 multidisciplinary experts working on the ARDST project for Caritas Development Goma NGO.

3.1.1 Study area

The study area lays in the eastern Democratic Republic of the Congo (DRC, Fig. 1A). In DRC the administrative hierarchy is as follows: the national territory is divided, since 2006, into 26 Provinces. Each Province is divided in Municipalities (urban areas) and Territories (rural areas). Municipalities are furtherly divided in cities and cities in communes and districts or neighbourhoods. Territories are

furtherly divided in Local Communities (or the Collectivities), the Collectivity in Groups, and the Group in Villages (Zongwe, 2019).

Figure 3.1 - Geographical framework of the study area. The Democratic Republic of the Congo (A); The Diocese of Goma, located at the extreme east of the DRC (B).
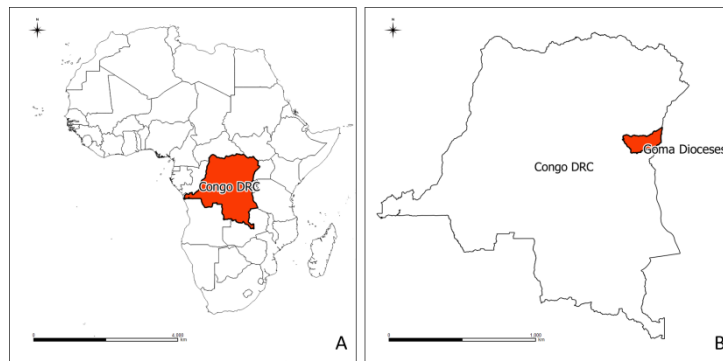
The study area is precisely defined by this ecclesiastical administrative hierarchy and corresponds to the Diocese of Goma (GD, surface: 26 223 km2). The territory of the GD is also known as 'small North' ('le petit Nord') and is located in the North Kivu province and in a small strip of the Kalehe Territory, from Minova to Nyabibwe (Figure 3.1B), laying in the South Kivu province. Within the diocese there are 11 Collectivities belonging to 5 territories (Goma, Rutshuru, Masisi, Walikale, Kalehe). The GD is part of the ecclesiastical province of Bukavu and consists of 27 parishes. Each parish is made up of different branches. These branches, called in Swahili '*Wigigi*' ('*Kigigi*' is the singular), are not exclusively ecclesial administrative units, because they also correspond to a pre-existing territorial division based on customary traditions. The *Wigigi* in the diocese of Goma are 105. In the territories belonging to each branch, the people are grouped into local ecclesial communities called *Shirika*. Within each Collectivity, it is estimated an average number of *Shirika* equal to 305 (CARITAS Development Goma, 2015), but this number varies as the *Shirika* are dynamic groups of people and not territorial units defined once and for all. The study area is characterized by one of the highest population densities in Africa (2 211 000 inhabitants on 26 223 km2 of territory equivalent to a mean population density of 88.4 ab./km2). At the same time the population of the area is characterized by a very high poverty rate (72.9% of the population in Nord Kivu live below the poverty line, against 71.2% in the whole DRC (UNDP, 2019; World Poverty Clock, 2019).

The main causes of this chronic situation are to be found in the historical events that have marked the socio-demographic evolution of the region.

As a result of these events, today the study area is inhabited by a very large and young population (50% of the population is less than 15 years old). Unemployment is relatively high (6.1% of the population). Child labour constitutes a diffused problem with a rate of 19.2%. The agricultural sector provides more than 7 jobs in 10. North Kivu has a primary school enrolment rate of 53.2%. The infant mortality rate is relatively low (57 ‰) compared to the national average (92 ‰). The average family size is 5.5 individuals per nucleus. Regarding sanitation, very few families are connected to drinking water (16.6%) and electricity (4.3%). 99.8% of households do not benefit from waste disposal services and 7% do not have access to toilets. Health services are few and poor: there are 47 hospitals throughout the province of North Kivu (12 beds per 100,000 inhabitants and 1 doctor per 24,030 inhabitants, figures very far from the WHO standard of 1 doctor per 10,000 inhabitants).

From an ethnic point of view the region of the two Kivus, includes three cultural areas: the north-east is populated by ethnic groups of probable Nilotic origin (Alur, Tutsi) combined with ethnic groups of Bantu origin (Hutu), the great agricultural region of Kivu is inhabited by Nande, Shi, Havu and Hunde, Twa and finally the forest region is occupied mainly by Bembe, Lega and other smaller groups (Vansina, 1966).

The economy of North Kivu is based on the primary sector (agriculture, livestock, fishing, forestry, mining) which represents about 49.7% of the provincial GDP, and that employs about 80% of the active population (UNDP, 2009). The agricultural, livestock and mining products, as well as the few processed products, are conveyed on local markets or exported following the traditional trade routes toward the ports of Mombasa (Kenya) and Dar-El-Salam (Tanzania). It should be underlined that a very significant proportion, at least 50%, of exports from the DRC is not registered by government officials, partly because traders underestimate declared exports to avoid paying taxes, partly because of the weak and corrupt bureaucracy. From the point of view of external access, the study area shows the airports of Goma and Kigali (Rwanda).

The study area is crossed by major road axes, which take the name of the crossed territories. The main road axes are part of the national road network and connect the study area to the capital cities of the neighbouring provinces (no road connection with the 1500 km-away national capital Kinshasa). The roads outside the capital and the main cities of the region, with the exception of the roads heading to Kigali (Rwanda), are unpaved clay roads, often interrupted due to the varied orography

(with altitudes ranging from about 1300 to the tops of the volcanoes that reach 3500-4000 m a.s.l), the heavy rain seasons and the 'lazy' upkeep provided by state authorities.

The study area is part of an agroecological area characterized by a long productive season (210-365 days of growth per year according to FAO, 2000). Local agricultural system can be framed in the agroecological zones as defined by FAO (2001): moving westward from the higher altitudes that lie along the borders with the neighbouring Uganda, Rwanda and Burundi and heading towards the forest, the farming system called 'Highland Perennial' gives the way to a 'Forest Based' system. (FAO, 2001). In this setting two production systems exist: the small farmers, practicing subsistence family agriculture on small plots, and the large landowners who instead mainly grow cash crop plantations in mono-cultural cropping systems for business[12] or breed cattle in extensive systems.

3.1.2 Workflow

Overall workflow was inspired by existing guidelines and namely by the evaluation design process proposed by Bezzi (2010) and illustrated in figure 3.2.

Figure 3.2 – Evaluation design process by Bezzi C. (2010, «Il nuovo disegno della ricerca valutativa»).
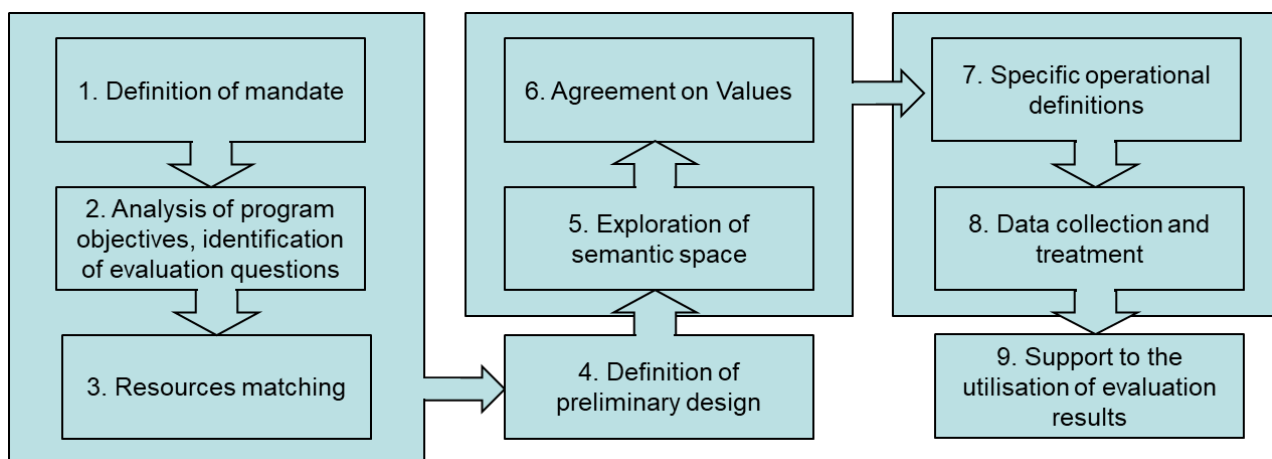
3.1.2.1 Mandate, evaluation design and questions

As Bezzi (2010), among other authors and organizations, explains, the evaluation process starts with the clear definition of a mandate, which is very important for defining and sharing the role of the evaluator and the expected results of his work. The impact of the evaluation itself is correlated to the

---

[12] For a complete description of the study area, namely from a technical agricultural point of view, the "Rapport de l'évaluation des potentialité agricoles en Diocèse de Goma", is available on demand.

degree of clearance of the evaluator: no matter the type of evaluation, if the evaluator is perceived as an external actor snooping in a pre-existing system, his ability to access information will be reduced, he will not be able to establish a constructive relationship with stakeholders and finally the results of his evaluation will be useless. For this reason, we had several meetings in order to clarify our approach and the evaluation questions. The agreed mandate was to lead a mixed-method evaluation of agricultural potentials in GD and a participatory decision-making process about triggering interventions. Four main evaluation questions were validated, as shown in Table 3.1.

Table 3.1 - Main research questions of the strategic sectoral evaluation carried out in the GD under the ARDST project.

	Evaluation question	Scale	Sources / methods for information gathering
A	What is happening today in rural areas?	Goma Diocese	Literature, Census, Remote sensing, Modified Delphi Method
B	What are the elements to be considered in describing agricultural potential?		Literature, Census, Modified Delphi Method
C	What is the land suitability in the territory of the Goma Diocese?		Participatory land suitability estimation, Remote sensing, photo interpretation, ad hoc surveys
D	What are the priority interventions in order to trigger agricultural sustainable development?		Modified Delphi Method, SPWR, AHP

3.1.2.2 Meta-evaluation and identification of sources

During the preparation phase we also worked out a meta-evaluation in order to understand where to look for data and how to access them. The goal of evaluating agricultural sector in the Diocese of Goma brought us to tackle many constrains.

The more general and expectable constrain is the lack of reliable and updated secondary data. Firstly, the history of the region explains why state institutional archives do not have updated nor complete secondary data. Secondly, even if FAO, WB, IFAD, CIAT and other international organizations do have repositories of publications on agricultural development in the region, no one of them is sufficiently focused on our study area.

The second, consequent and somehow more specific constrain is related to difficult physical accessibility to primary data, meaning that even if information could be plausibly available 'on the ground' by direct survey, the local orography, the lack of infrastructure connectivity and mostly the

physical insecurity due to armed groups presence prevent the access to this information, making dangerous for evaluators seeking primary data in the study area (Vlassenroot, 2006).

The third constrain is related to the fact that, given the historical background and the present humanitarian situation of the study area, the region is overridden by development cooperation initiatives whose praiseworthy work, on one hand, lifted the population during decades of security, health and famine emergency, while on the other hand created a diffused feeling of dependency from the external aid, preventing people spirit of initiative and implication in local development (Kabonga, 2017).

Figure 3.3 - Case study workflow: detailed evaluation and decision-making design. The evaluation design described here by the 1 to 9 bold elements is inspired from the more general scheme by Bezzi C. (2010, «Il nuovo disegno della ricerca valutativa»).
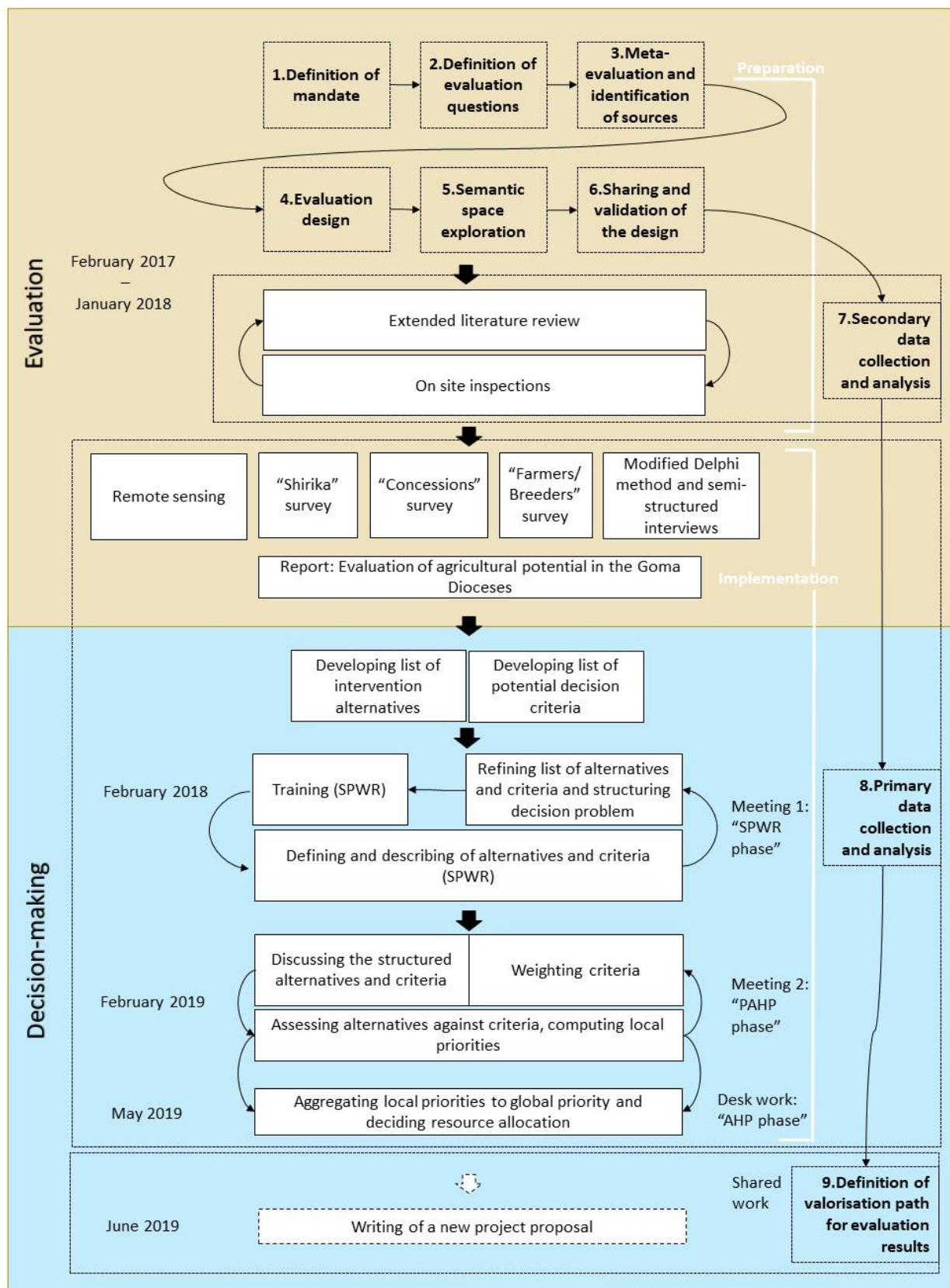
3.1.2.3 Evaluation design, semantic space exploration and design validation

In order to 'break this wheel' of externally planned local development and in order to tackle the other constrains, we created a specific evaluation design which responds to the needs of fostering local stakeholders active implication, accessing multiple sources of data and producing updated and locally-focused primary data. In fact, even if regional and sub-regional studies may constitute useful background documents, both the need for updated information and the need for involving local stakeholders and decision-makers brought us to identify a mixed-method participatory approach to evaluation and decision-making. Semantic space exploration phase includes all the exchanges with the evaluation team whose aim was to understand and share the meaning of design choices and expected implications of the evaluation outcomes. The validated data collection strategy is illustrated in Figure 3.4 and further described in Figure 3.6 at the end of this chapter. Three surveys were planned, and two other sources of information were proposed for triangulation and for enriching the data collection. Namely remote sensing and a modified Delphi method were inserted in the plan of the evaluation.

The validated overall evaluation and decision-making design is sketched in Figure 3.3 in previous paragraphs.

Figure 3.4 . Showing how the evaluation questions determine the need for multicriteria evaluation and multi-source identification.
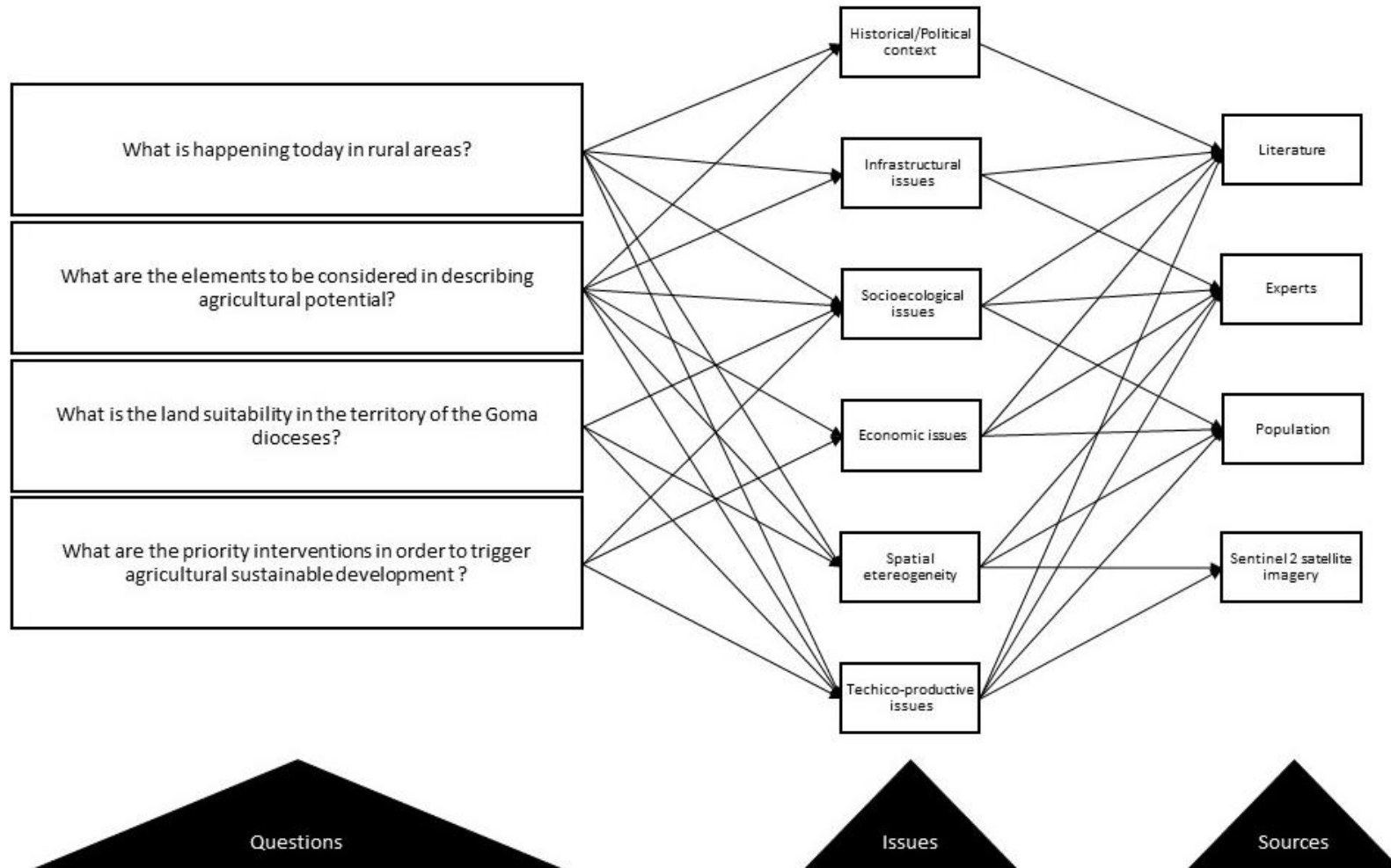
3.1.2.4 Secondary data collection and treatment

First, during the completion of the design phase, an extended literature review was conducted about agricultural development in Nord Kivu. Results were compared with personal observations during nine on-sites inspections in different zones of the Goma Diocese[13]. All the collected information served to inform the further design of data collection strategies and tools through periodic sharing, discussion and validation of following scheduled steps.

3.1.2.5 Primary data collection and treatment

In this paragraph we provide an introduction to the methodologies and techniques that we used along the evaluation and decision making paths. Further details about methodologies are provided in chapters 4, 5 and 6.

3.1.2.5.1 Surveys

Once the design had been validated by the team and the analysis of secondary data was in its final stage, we were able to start the implementation of the first two field surveys[14], 'Shirika' and 'Concessions', whose aim was to capture information from the representative of local communities (Shirika[15]) and from landowners (Concessions' owners). The third survey was aimed at collecting detailed technical information about agricultural activities and it was implemented in February – May 2019 by deploying eight surveyors and producing 40 valid observations of 336 variables.

A survey is an observation tool which quantifies and compares information. Data are collected from a sample of the population through a survey form or questionnaire, and therefore they are expression of the knowledge detained by the population. In developing countries, face-to-face questionnaires remain the surest tool to obtain the information required (EUROPAID, 2006).

---

[13] 9 site inspections were carried out during three 15-days missions in August 2017, February 2018 and August 2018. Site inspections were carried out in nine different localities: Bunangana, Katale and Rutshuru (in Rutshuru territory, Nord Kivu), Masisi, Minova, Bibwe, Katoyi and Lushebere (in Masisi territory, Nord Kivu), in Kiniezire (in Kalehe territory, South Kivu). During site inspections participatory observation and semi-structured interviews were the main methodologies used for collecting information.

[14] "Shirika" survey was implemented between February 2017 and February 2018 by deploying 23 surveyors, producing 1,008 valid observations of 128 variables and covering 59,3% of existing Shirika; "Concessions" survey was implemented between March 2017 and March 2018 by deploying 23 surveyors and producing 2,558 valid observations of 31 variables.

[15] Shirika is the local name in Swahili for the smallest administrative unit of the ecclesiastical structure in the region. The Goma Diocese is divided in

More information about the three surveys is presented in Table 3.2.

Table 3.2 – Detailed information about the three surveys implemented in the framework of the ARDST project.

Survey title	Survey ors	N° of valid questionn aires	N° of questions x theme	Info about sample representativeness
Shirika	23	1008	Total: 39 questions ✓ 8 x general info ✓ 4 x info on *Shirika* ✓ 2 x info on occupied land ✓ 3 x info on available land ✓ 6 x info on water availability and soil erosion ✓ 2 x info on land use ✓ 7 x info on production and marketing organisation ✓ 6 x info n other existing activities	We estimate that our sample cover 59,3% of existing *Shirika* in the study area (GD).
Concessio ns	23	2558	Total: 16 questions ✓ 5 x info on general info on the concession ✓ 5 x info on land tenure rights and land use ✓ 6 x info on existing conflict over land use or property	No information exists on the total number of existing concessions. We estimate that our sample cover 173,526.3 ha (1,735.26 km2) which corresponds to about 8% of the total area of the GD.
Agricultur e and animal husbandr y	8	40	Total: 209 questions ✓ 14 x general info ✓ 4 x general info on agropastoral activities ✓ 138 x technical info on agricultural production ✓ 39 x technical inf on animal husbandry ✓ 5 x info on social and work organisation ✓ 9 x info on other existing local resources	The sample was built by asking each of the 8 surveyors to select 5 literate and dynamic farmers or animal breeders on each zone of implementation of the project (8 axes).

Results from the surveys were collected in three separate databases and treated in order to produce descriptive statistic. Pearson correlation was computed between all the relevant variables in order to identify existing relationships. Results from the analysis were used to inform, through triangulation with other methodologies, the final ARDST evaluation report entitled 'Evaluation du potentiel agricole dans le Diocèse de Goma' and the following steps of the decision-making process.

Direct surveys were implemented in parallel with remote sensing and a modified version of Delphi method.

3.1.2.5.2 Modified Delphi method

The Delphi method is a structured technique or method, originally developed as a systematic, interactive forecasting method which relies on a panel of experts. (de Loë et al., 2016) The experts are individually interviewed in two or more rounds. After each round, a facilitator provides an anonymised summary of the experts' opinions from the previous round, as well as the reasons they provided for their arguments. Thus, experts are encouraged to revise their earlier answers considering the replies of other members of the panel. It is believed that during this process the range of the answers will decrease, and the group will converge towards one shared answer. Finally, the process is stopped following a predefined stop criterion (e.g., number of rounds, achievement of consensus, stability of results).

We used the project team experience and information from the literature review to identify individuals knowledgeable about agricultural systems in the diocese of Goma, seeking both depth of knowledge and diversity of orientation and background to cover many perspectives and components of the system.

We implemented a modified Delphi method described hereafter: 24 individuals were identified by the evaluation team and interviewed in February 2018. Additional 16 individuals were obtained from the first round of interviews through 'snowball' sampling (Naderifar et al. 2017) and were interrogated during a second turn of semi structured interviews in August 2018. This procedure provided extensive information about both the agricultural sector state-of-art and its potentials, granting several different perspectives according to subjective role in the network of organisations working in the local agricultural system. The procedure allowed to identify five individuals among the total 40 respondents who were considered to be experts by most of the respondents. These five 'agreed-upon' experts[16] were questioned both during the first and second turn of interviews in order to refine and sum up the information produced during the implementation of the modified Delphi methodology. Full list of interviews is provided in Table 3.3 below. The data collected with this modified Delphi method were useful also for the following decision-making process that will be better described in chapters 5 and 6.

---

[16] IDs 11, 14, 15, 16, 23 in Table 3.3.

*Table 3.3 - Full list of semi structured interviews held in the framework of the modified-Delphi method. Rows highlighted in grey contains the five agreed-upon experts who were interviewed twice in order to implement the modified Delphi method. *ID marked with the asterisk are the "agreed-upon" experts who participated to both the rounds of the modified Delphi in order to sum-up and refine the results.*

Turn 1				Turn 2			
ID	Name	Structure / Profession	Place and date of interview	ID	Name	Structure / Profession	Place and date of interview
Key informants – experts of local agricultural development							
1	M. Gabo Musogo	CHEF COUTUMIER de Localité Mubeko en Groupement de Nyalipe	Katoyi, 15/02/18	26	Divers	GROUPE D'AGRICULTEURS/ELEVEURS qui habite à la frontière du Domain de Katale	Katale, 03/08/2018
2	M. Bapiasi Ngatsimbikyi	CHEF COUTUMIER de Chefferie, en territoire de Rutshuru, Paroisse de Jomba, Groupement de Jomba.	Jomba, 13/02/2018	27	M. Amisi André	AGRICULTEUR	Nyakariba, 01/08/2018
3	M. Ngumba Bwira Elias	CHEF COUTUMIER de Groupement de Nyamaboko 2 à Kikoma, Secteur de Katoyi	Katoyi, 15/02/18	28	M. Ezmawa Michel	ELEVEUR	Kitchanga, 02/08/2018
4	M. Ndaysaba Bakamira	CHEF COUTUMIER de Groupement de Kibabi 2 en Secteur de Katoyi	Katoyi, 15/02/18	29	M. Michel	SOTRAKI (transformation de café) / DIRECTEUR DE L'OUSINE	Goma, 03/08/2018
5	M. Bazamanza Ruesa	RECENSEUR de Groupement de Kibabi 2	Katoyi, 15/02/18	30	M. Jeredi Sabacungu	ELEVEUR	Nyakariba, 01/08/2018
6	M. Turikunkiko Bigembe Matabaro	CHEF COUTUMIER de Secteur et CHEF DE SECTEUR-COLLECTIVITE' de Katoyi	Katoyi, 15/02/18	31	M. Ngumije Niyonzima Laurien	DOMAIN DE KATALE (plantation et transformation de café) / DIRECTEUR DES PLANTATIONS	Katale, 03/08/2018
7	Abbé Janvier Aushime	CURE' DE LA PARROISE DE RUGHERI en Rutshuru	Rugheri, 13/02/2018	32	M. Nibamue Albert	AGRICULTEUR	Kitchanga, 02/08/2018
8	M. Bolhaza Rukoriki	GRAND CONCSSIONAIRE, Concession de Kaghara/Builoma en Territoire de Rutshuru, Paroisse de Jomba	Rutshuru, 13/02/2018	33	M. Oswald Sangana	AGRICULTEUR	Nyakariba, 01/08/2018
9	M. Muhire J. Bosco	GROUPE D'AGRICULTEURS, ELEVEURS AVEC AGRONOME du Secteur Katoyi	Katoyi, 15/02/18	34	M.me Seraphine Muhima	AGRICULTRICE	Kitchanga, 02/08/2018
10	M. Mapendo Fokasi	SECRETAIRE de Collectivité-Secteur de Katoyi	Katoyi, 15/02/18	35			
11*	M. Guido Semahane*	CARITAS GOMA / SUPERVISEUR DE TERRAIN de l'AXE	Katoyi, 15/02/18	11	M. Guido Semahane*	CARITAS GOMA / SUPERVISEUR DE TERRAIN de l'AXE Matanda-Masisi-Nyakariba	Goma, 10/08/2018

Turn 1				Turn 2			
ID	Name	Structure / Profession	Place and date of interview	ID	Name	Structure / Profession	Place and date of interview
		Matanda-Masisi-Nyakariba					
Farmers organisations							
12	M. Pierrot Mandefu Bwansiki	SYDIP / Responsable du bureau de Liaison de Goma	Goma, 12/02/2018	36	M. Murairi Janvier	ASSODIP asbl – Association pour le developpement des initiatives paysannes / PRESIDENT	Goma, 31/07/2018
13	M.me Kavira Bora	COOCENKI – Assistente Administrative	Goma, 14/02/2018	37	Jean Baptiste M.	FOPAC NORD KIVU/AJAC RDCongo (Association des Journalistes Agricoles) / JURNALIST	Goma, 09/08/2018
14*	M. Tharcis Bali*	FOPAC-NK / Responsable bureau de Goma	Goma, 14/02/2018	14	M. Tharcis Bali*	FOPAC-NK – Responsable bureau de Goma	Goma, 31/07/2018
15*	M.me Masika Marielaine*	UWAKI-NK (Umoja Wamama Wakulima wa Kivu – Nord Kivu - Union de femmes paysannes du Nord Kivu) / SECRETARIE EXECUTIVE	Goma, 21/02/18	15	M.me Masika Marielaine*	UWAKI-NK (Umoja Wamama Wakulima wa Kivu – Nord Kivu - Union de femmes paysannes du Nord Kivu) / SECRETARIE EXECUTIVE	Goma, 31/07/2018
16*	M. Theogene Sikiryamuva*	UWAKI-NK (Umoja Wamama Wakulima wa Kivu – Nord Kivu - Union de femmes paysannes du Nord Kivu) / CHARGE DE PROGRAMME	Goma, 21/02/18	16	M. Theogene Sikiryamuva*	UWAKI-NK (Umoja Wamama Wakulima wa Kivu – Nord Kivu - Union de femmes paysannes du Nord Kivu) / CHARGE DE PROGRAMME	Goma, 31/07/2018
17	M. Musubao Anuarite	LOFEPACO (Ligue Organisation de Femmes Paysannes du Congo) / AGRONOME	Goma, 21/02/18				
18	M.me Denise Kombi	LOFEPACO (Ligue Organisation de Femmes Paysannes du Congo) / CHARGE DE PROGRAMME	Goma, 21/02/18				
State offices – province							
19	M. Vulambo Kalissa	DPDD-NK (Division du Plan de Développement Durable du Comité Provincial du Nord Kivu) / CHEF DE DIVISION	Goma, 21/02/18	38	Dr. Nluwa Ephrem	Mairie de Goma – Service de l'agriculture, pèche et élevage / DIRECTEUR	Goma, 10/08/2018
20				39	Dr. Stella Sowhungo	Mairie de Goma – Service de l'agriculture, pèche et élevage / INGEGNEUR AGRONOME	Goma, 10/08/2018
21				40	Dr. Vincent Njabimana	Inspection provinciale de l'agriculture, peche et élevage du Nord Kivu / SERVICE CHIEF	Goma, 10/08/2018
Research institutes							

Turn 1				Turn 2			
ID	Name	Structure / Profession	Place and date of interview	ID	Name	Structure / Profession	Place and date of interview
22	Prof. Kennedy Kihangi Bindu	ULPGL-CREDDA (Université Libre des Pays des Grands Lacs - Centre de Recherche pour la Democratie et le Developement) / Secrétaire administratif AND Vice Recteur	Goma, 20/02/18				
23*	M. Audry Muke Mangekele*	INERA-GOMA (Institut National de Recherche Agricole, antenne de Goma) / LOCAL REPRESENTER IN GOMA	Goma, 21/02/18	23	M. Audry Muke Mangekele*	INERA-GOMA (Institut National de Recherche Agricole) / ANTENNE DE GOMA	Goma, 31/07/2018
International organisations							
24	Dr. Francesco Minisini	AGENCE ITALIENNE DE COOPERATION AU DEVELOPEMENT (AICS) / Chef de programme, Bureau de Goma	Goma, 20/02/18				
25	Dr. Martina Tenko	DELEGATION DE L'UNION EUROPEENNE en RDC / Conseillere Politique, Section Politique, Communication, Presse et Information	Goma, 20/02/18				

3.1.2.5.3 Remote sensing

In current usage, the term 'remote sensing' generally refers to the use of satellite- or aircraft-based sensor technologies to detect and classify objects on Earth, including on the surface and in the atmosphere and oceans, based on electromagnetic radiation. It may be split into 'active' remote sensing (such as when a signal is emitted by a satellite or aircraft and its reflection by the object is detected by the sensor) and 'passive' remote sensing (such as when the reflection of sunlight is detected by the sensor). In the present experience we used passive remote sensing.

Goals of the remote sensing technique were the production of maps concerning land cover in the project area, in particular:

1. A land cover map (4 classes) pertaining to the entire GD (indicative scale 1: 50,000, geometric resolution raster 10 m);

2. A land cover map (4 classes, indicative scale 1: 10,000, geometric resolution raster 10 m) pertaining to the Katoyi community;

3. A map of the prevailing (2) agricultural production systems spread throughout the territory for each of the collectivities designated within the GD (based on the characterization of the level of fragmentation).

Remote sensing workflow is shown in Figure 3.5. Further details are provided in chapter 4.

Figure 3.5 - Showing the remote sensing technique workflow.
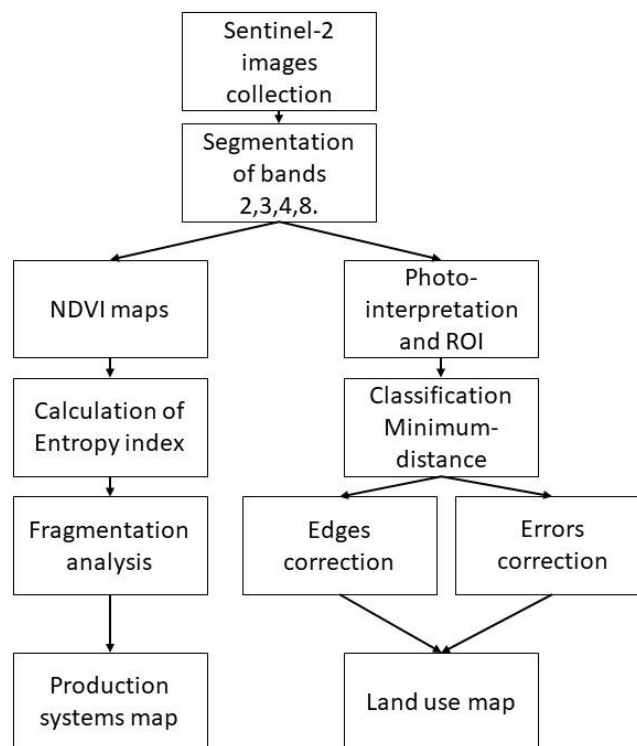
3.1.2.5.4 Simplified pairwise ranking

Simplified PWR (SPWR) is a simplified form of the Pair-wise Ranking technique (PWR) (Narayanasamy, 2009): it helps a decision maker to binary pair-compare alternatives and to rank them according to their relative frequency of preference. Also called Priority Grid or Matrix, PWR provides a rational and structured approach to derive the group's priorities, whilst facilitating much negotiation and consensus building by the group (FAO, 2003 and 2006a).

Even if many applications of structured Multicriteria Decision Making or Aiding (MCDM or MCDA) methodologies to real-world cases exist, Development cooperation organizations tend to neglect the use of more complex MCDM methods such as AHP and AHP-nested PWR. These organizations prefer a simplified form of PWR (SPWR) when dealing with the rapid appraisal of the context that normally precedes the planning and implementation of development projects (FAO, 2006a; Russel, 1997; World Bank, 2018).

Considering the overall evaluation and decision making process presented in Figure 3.3, in order to appropriately approach the identification of triggering interventions for agricultural development in a specific collectivity of the GD, the whole technical team of the ARDST project was asked to join a participative decision process animated though SPWR (Narayanasamy, 2009; Russel, 1997). This participatory path started in February 2018, focusing the results of the previous evaluation phase and aiming at identifying priorities for intervention in the Katoyi collectivity, GD, DRC. The whole ARDST project team was involved. 14 persons with different background and expertise participated to the SPWR meeting: four are agronomists, four are jurists, three are expert in sustainable rural development, one is a sociology professor, one is a civil engineer, one is a cartographer.

Alternative interventions were previously extrapolated from the results of a 'preparation' phase in which the agricultural sector of the GD was analyzed through mixed methods. Eight alternatives were then validated and further described in order to build common understanding of the theory of change (Brest, 2010) connecting interventions to the overall objective. Full description of the SPWR technique is provided in chapter 5.

3.1.2.5.5 Analytic hierarchy process

Analytic Hierarchy Process (AHP) (T.L. Saaty, 1980; 1987; 2008) is one of the leading techniques in the wide family of Multicriteria Decision Making or Aiding (MCDM or MCDA) methodologies (Emrouznejad et al., 2017; Ho et al., 2018; Ishizaka et al., 2011). In our opinion, if such a methodology is turned participatory, it becomes an important tool for optimising project multi-ownership which in turn, may lead to better performance from project managers, management and participants in general.

In the 6th chapter our application of participatory AHP in a real-world case study, related to agricultural development in the GD, Democratic Republic of Congo, is discussed. The practical

objective of using AHP in the case study was to elicit discussion on priority criteria and to have new useful elements to determine resource allocation in the framework of the planning of new interventions in the agricultural sector. Main research questions of this phase of the research were: (1) What are the most important criteria behind the prioritization of interventions to achieve sustainable development of the agricultural sector in the GD, DRC? (2) How should financial resources be allocated on identified priority interventions?

3.1.2.6   *Use of evaluation results*

As we anticipated at the beginning of this chapter, the aim of the evaluation was to inform the choice of interventions meant to trigger agricultural sustainable development in the GD. As it is shown in Figure 3.4, our evaluation design relied on multiple sources of information, namely literature, sectoral experts, population (modified Delphi method, participatory observation and focus groups) and remote sensing repositories. From each source, with tailored methodologies, information was collected and organized in separate reports which finally have been integrated in order to synthesize knowledge and produce both the technical evaluation report (available on demand) and the scientific papers collected in the following chapters of the present thesis.

In Figure 3.6 the procedure for summing-up information is sketched showing how different sources and methodologies contributed to the drafting of the final report on agricultural potential and triggering interventions for sustainable agricultural development in the GD. In the same Figure 3.6 the specific steps or parts that allowed us to write scientific papers are highlighted and will be better described in the following chapters.

Figure 3.6 - Showing the synthesis procedure used to integrate different sources and methodologies into deliverable results. This Figure can be compared to Figure 3.4 in order to link the main research questions to final deliverables.
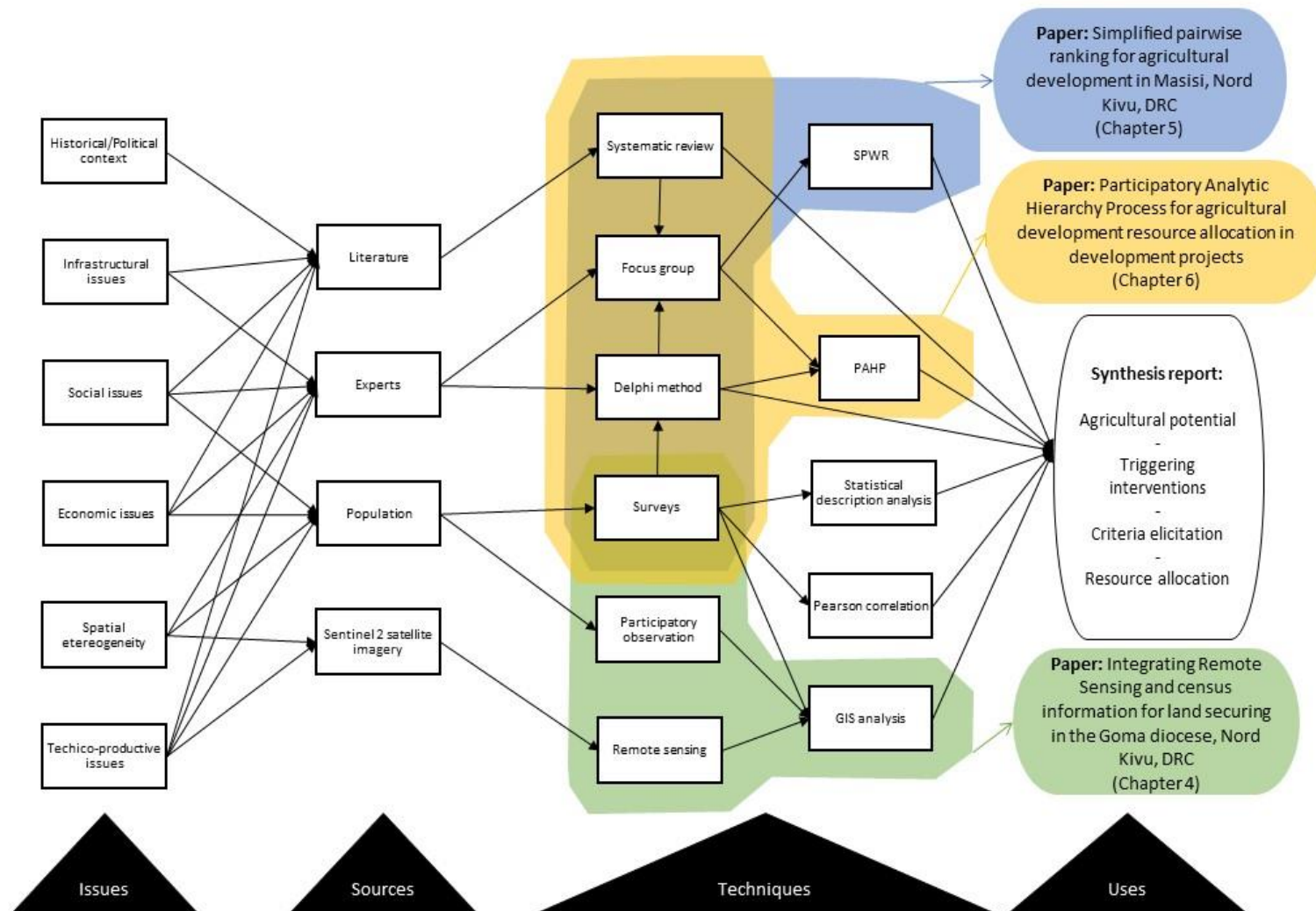
3.2 References

Bezzi, C. (2010). *Il nuovo disegno della ricerca valutativa*. Milan: Franco Angeli.

Brest, P. (2010). *The Power of Theories of Change*. (Leland Stanford Jr. University, Ed.). Standford: Stanford Social Innovation Review.

CARITAS Development Goma. (2015). Document de projet ARDST 'Contribuer à apporter des pistes de solution aux causes profondes des conflits à l'Est de la République démocratique du Congo, par le biais d'un appui à la sécurisation foncière.' Goma, RDC.

de Loë, R. C., Melnychuk, N., Murray, D., & Plummer, R. (2016). Advancing the State of Policy Delphi Practice: A Systematic Review Evaluating Methodological Evolution, Innovation, and Opportunities. *Technological Forecasting and Social Change*, *104*, 78–88. https://doi.org/https://doi.org/10.1016/j.techfore.2015.12.009

Emrouznejad, A., & Marra, M. (2017). The state of the art development of AHP (1979–2017): A literature review with a social network analysis. *International Journal of Production Research*, *55*(22), 6653–6675. https://doi.org/10.1080/00207543.2017.1334976

EUROPAID. (2006). *Méthodologie d'évaluation de l'aide extérieure de la commission européenne. Volume 4. Outils d'évaluation.* Luxembourg: European Commission.

FAO, 2000. Derived from the Global Agro-Ecological Zones Study, Food and Agriculture Organization of the United Nations (FAO), Land and Water Development Division (AGL) with the collaboration of the International Institute for Applied Systems Analysis (IIASA), 2000. Data averaged over a period of 37 years. Raster dataset has been exported as ASCII raster file type. Retrieved July 10, 2019 from: http://www.fao.org/nr/gaez/en/

FAO. (2001). *Farming Systems and Poverty*. (J. Dixon, A. Gulliver, & D. Gibbon, Eds.). Rome: FAO Publishing and Multimedia Service. https://doi.org/10.1017/S0014479702211059

FAO. (2003). *Participatory Development: guidelines on beneficiary participation in agricultural and rural development*. (B. van Heck, Ed.) (2nd ed.). Rome: Rural Institutions and Participation Service, Rural Development Division Food and Agriculture Organization of the United Nations.

FAO. (2006a). *Toolkit for the Application of Participatory and Negociated Territorial Development*. Rome: FAO Publishing and Multimedia Service.

Ho, W., & Xin, M. (2018). The state-of-the-art integrations and applications of the analytic hierarchy process. European Journal of Operational Research, 267, 399–414. https://doi.org/10.1016/j.ejor.2017.09.007

Ho, W., & Xin, M. (2018). The state-of-the-art integrations and applications of the analytic hierarchy process. *European Journal of Operational Research*, *267*, 399–414. https://doi.org/10.1016/j.ejor.2017.09.007

Kabonga, I. (2017). Dependency Theory and Donor Aid: A Critical Analysis. *Journal of Development Studies*, *46*. https://doi.org/10.25159/0304-615X/1096

Naderifar, M., Goli, H., & Ghaljaei, F. (2017). Snowball Sampling: A Purposeful Method of Sampling in Qualitative Research. *Strides in Development of Medical Education*, *In Press*. https://doi.org/10.5812/sdme.67670

Narayanasamy, N. (2009). *Participatory Rural Appraisal: Principles, Methods and Application*. New Delhi: SAGE Publications India Pvt Ltd. https://doi.org/10.4135/9788132108382

Russel, T. (1997). Pair Wise Ranking Made Easy. In *PLA Notes* (pp. 25–26). London: International Institute for Environment and Development.

Saaty, T. L. (1980). The Analytic Hierarchy Process. *Education*, 1–11. https://doi.org/10.3414/ME10-01-0028

Saaty, T. L. (1987). The analytic hierarchy process-what it is and how it is used. *Mathematical Modelling*, *9*(3–5), 161–176. https://doi.org/10.1016/0270-0255(87)90473-8

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83. https://doi.org/10.1504/IJSSCI.2008.017590

UNDP. (2009). *Nord Kivu Province profile*. (G. Aho, J. R. Ravelosoa, & A. Lukaku Zinga, Eds.). Kinshasa: UNDP Publishing.

UNDP. (2019). *Poverty and household living conditions in the North-Kivu Province*. Retrieved December 27, 2019, from https://www.cd.undp.org/content/rdc/en/home2/library/poverty/pauvrete-et-condition-de-vie-des-menages1.html

Vansina, J. (1966). *Introduction à l'ethnographie du Congo*. Kinshasa: Université Lovanium.

Vlassenroot, K. (2006). War and Social Research. *Civilisations*, *54*(1–2), 191–198. https://doi.org/10.4000/civilisations.442

World Bank. (2018). World Bank - Participatory Rural Appraisal (PRA) Techniques. Retrieved May 24, 2019, from http://go.worldbank.org/EU99NNIVS0

World Poverty Clock. (2019). World poverty clock. Retrieved May 24, 2019, from https://worldpoverty.io/blog/index.php?r=18

Zongwe, D. P. (2019). Decentralization in the Democratic Republic of the Congo. *Autonomy Arrangements in the World*. https://doi.org/10.13140/RG.2.2.20028.08321

4 Integrating Remote Sensing and census information for land securing in Nord Kivu, DRC[17]

4.1 Introduction

4.1.1 East-DRC: Historical background

Since the Democratic Republic of Congo (DRC) is one of the richest countries in the world in terms of natural resources (ADB, 2009; Bwana, 2011), the control of lands in the region has always been at the core of profound disputes (Naidoo, 2003).

In such a context of internal and external economic interests, land disputes can be considered obviously both as sources and perpetuating factors of conflict in the eastern DRC (Verweijen & Marijnen, 2016). The most visible land-related conflicts, contributing to large-scale violence, are those that put ethnic communities against each other. However, other forms of land-related conflict are very widespread, including community-level conflicts between farmers and large-scale concessionaires, between rural communities and mining companies, between pastoralists and farmers, and between national parks institutions and surrounding populations. While generally accompanied by low levels of violence, the grievances related to such conflicts often impact on the security and stability at the local level (Mathys & Vlassenroot, 2016).

The weakness of land legislation, associated with widespread corruption, led to a massive alienation of lands under the two existing land tenure systems, namely customary and statutory tenures. On the other hand, the Government has already signed a huge number of international agreements with multinational companies for large-scale land exploitation projects, mostly for mining and biofuel production (Landmatrix, 2017). These two causes brought to an incredibly difficult situation for the increasing population of the region, who now is cornered in the less productive areas (Frankema & Buelens, 2013).

The situation is even more complicated by the huge number of refugees in the region. According to official data (UNHCR, 2019), some 626,000 refugees (Externally Displaced Persons – EDP) from the

---

[17] This chapter contains the pre-print version of the paper by De Marinis P., Manfron G., Sali G., Facchi A., 2019. Remote Sensing and land stabilization in Nord Kivu, DRC, CUCS2017 Congress paper on Journal of Universities and international development cooperation, ISSN: 2531-8772, published.

DRC are still living in neighbouring countries, including Burundi, Uganda, the United Republic of Tanzania and Rwanda. The humanitarian profile of North Kivu, in particular, is alarming, with an estimated 863,400 internally displaced persons (IDP) between January 2009 and November 2014. All these people have the right to return to their home, as declared by several International Treaties concerning human rights in general and the refugee condition in particular (CIRGL, 2006).

For all these reasons, the agenda of international organizations dealing with peace seeking and keeping in the Eastern Congo are now focusing, among other issues, on drawing a clear sketch of land tenure and land ownership distribution in the region.

4.1.2 The project 'Appui au retour de réfugiés et déplacés par le biais de la sécurisation de terres en Diocèse de Goma'

The Department of Agricultural and Environmental Sciences (DISAA) of the University of Milan, is involved with Caritas Goma NGO in a three-year project called ARDST 'Appui au retour de réfugiés et déplacés par le biais de la sécurisation de terres en Diocèse de Goma', funded by the European Commission. The ARDST project aims, among other goals, at identifying reference pilot areas suitable for the re-settlement of Congolese refugees actually residing in Rwanda.

Identification of settlement sites is always a complex matter, and, in such a region, it needs increased awareness of complex interactions among humans, environment, and potential land management decisions. This kind of complex decision making creates a demand for integrated multidisciplinary decision support tools within a spatial framework (Baker, Miller, & Paige, 2009).

On this assumption, our team experimented innovative methodologies, such as satellite remote sensing (RS) techniques, to provide new cartographic and thematic land use informative layers. In addition, the ARDST project is conducting an important data gathering effort over the target area (Diocese of Goma, 26,223 km^2) in order to collect detailed information about existing agricultural resources, conflicts, and potentialities. Therefore, in the framework of the ARDST project the DISAA proposes the setup of a Spatial Decision Support System (SDSS) aimed at Multi Criteria Evaluation (MCE) for the identification of pilot zones.

4.1.3   Spatial Decision Support Systems and Remote Sensing

SDSSs are GIS-based models that 'enhances a person or group's ability to make decisions' (Power et al., 2014). As tools for improving evidence-based choices, SDSS are widely used in several sectors such as policy making, enterprise management, environmental studies and more recently sustainable development studies. As GIS-based systems, SDSSs are made of a geo-referenced database that is merged with a DSS which in turn is often related to some kind of Multi Criteria Analysis (MCA) (Chakhar et al., 2008).

Generally speaking, SDSS are widely used in order to tackle land development choices (Agatsiva & Oroda, 2002). Experiences in the East Africa region are reported by several authors (Baldyga, Miller, Driese, & Gichaba, 2008).

In the framework of International Cooperation for Development, and namely among the programmes and projects funded by the EU, other experiences of SDSS development also exist (ITC, 2009; Janusz, 2016; Manakos & Braun, 2014; Refsgaard et al., 1998; Uyan, Cay, & Akcakaya, 2013). Satellite image has been used so far to classify and map land cover and land use changes with different techniques and data sets. Unsupervised and supervised approaches are the most commonly adopted for satellite images classification (Butt, Shabbir, Ahmad, & Aziz, 2015; Lu, Mausel, Brondizio, & Moran, 2004; Rundquist, Narumalani, & Narayanan, 2001; Zhang, Zhang, & Zhang, 2000).

The application of supervised approaches implies that the user or image analyst 'supervises' the pixel classification process by specifying which pixels values or spectral signatures should be associated with each class according to known representative sample sites, called Training Sites. The 'trained' algorithm is then able to classify the whole image. Precision of supervised classification depends heavily on the training sites, the skills of the expert who process the image, and the spectral distinctness of the classes.

A variety of supervised classification methods have been developed and applied to satellite images for land use mapping. Among these, some applications were specifically designed for land use monitoring in the DRC. Important examples are the contribution of Hansen et al. (2008) which focused on the use of Landsat images to map forest land-cover typologies exploiting unsupervised decision tree approach. They succeed in validating maps having 75% of overall accuracy. Duveiller et al., (2008) instead, focused on deforestation mapping in central Africa in the period 2000-2010

applying unsupervised object-based classification procedures to Landsat images. The overall accuracy estimated for their thematic products was 91%. Vancustem et al. (2009) proposed a semi-automatic classification of SPOT (Satellite Pour l' Observation de la Terre) temporal-spectral information, to produce a map discerning 18 vegetation types in the DRC.

An additional source of land-use information is census data, which has recently been explored for its relationship to remotely sensed imagery (Cardille & Foley, 2003). Land use/cover changes have been identified as a useful tool to aid the process of understanding human-environment interaction (Dale, Oneill, Pedlowski, & Southworth, 1993). Thematic and census land use descriptors can be integrated in a geographical information system (GIS) on the base of a given administrative level.

4.1.4   Scope of the paper

This paper reports the implementation and results of a preliminary methodological test focused on the use of RS to collect data on the land cover/use, on the collection of the Census data on the ground, and on the potential added value of merging these two sources of information in order to support the multicriteria choice of pilot zones. Therefore, even if the final scope of the research is the set up a SDSS for the identification of pilot zones, based on the multicriteria evaluation (MCE) of land hosting capability, this paper only reports preliminary steps. In order to identify zones with 'good' hosting capability, the ARDST project has defined several criteria. The principal criteria considered are presented in Figure 4.1, where the scheme of the proposed SDSS is shown.

Figure 4.1 - Scheme of the SDSS draft for the evidence based identification of ARDST project pilot zones, containing a list of the principal choice criteria which have been considered in this work.
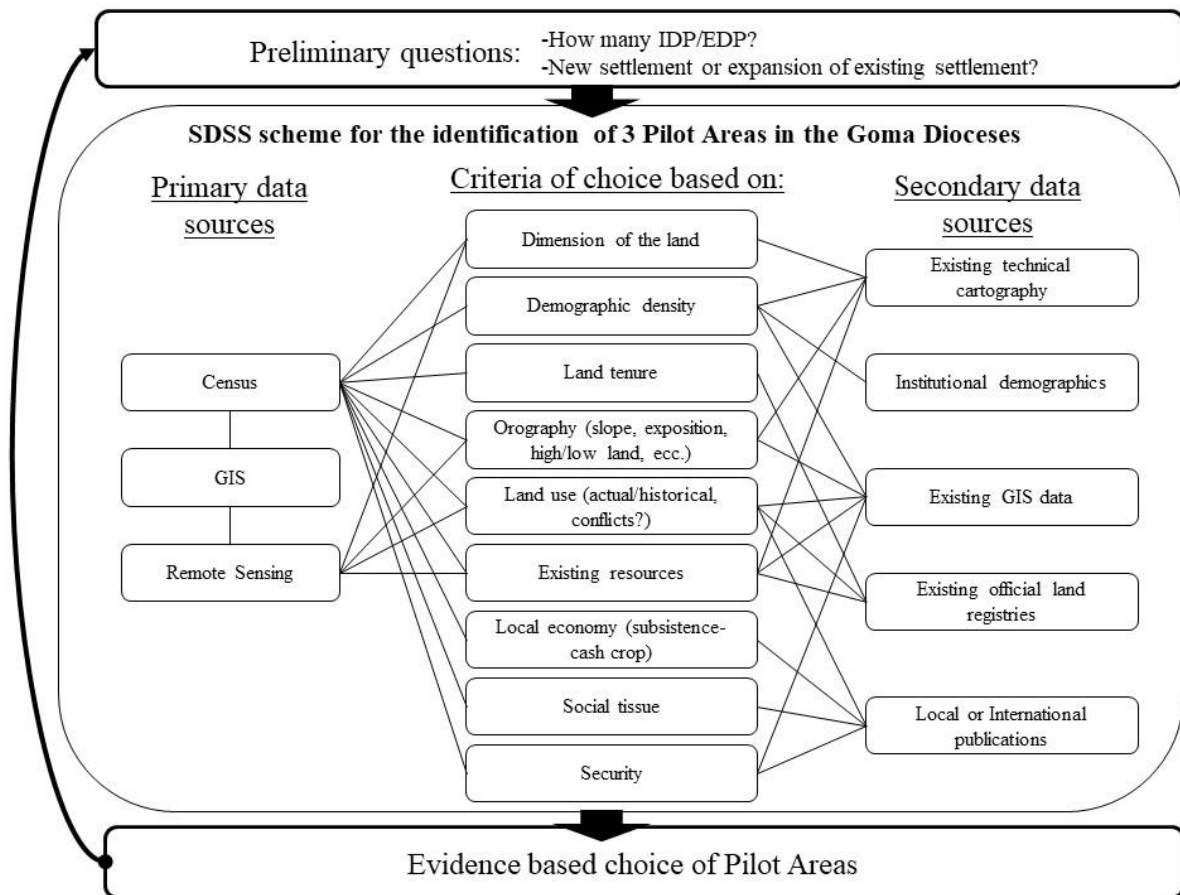
It clearly appears how the choice of pilot areas depends on several criteria whose measurement relies on different data sources, both primary and secondary.

Three administrative units, namely the collectivities of Wanyanga, Bakano and Bashali, placed in the study area of the GD, have been selected on the base of the available Census results for this preliminary methodological test. The methodological approach concerns the integration in GIS environment of land use products derived by RS and detailed census data about existing agricultural resources, potentialities and existing conflicts. The generated land-use-knowledge is therefore used to build useful evidences for the identification of pilot zones and, generally speaking, to foster peace-seeking operations in the region.

Preliminary results highlight how the integration of different information sources lead to: (i) cross-validation between sources, (ii) identification of conflicting information between sources, and (iii) creation of emergent information stemming from the joining of the two sources.

Results, proposed in form of thematic maps and aggregations of census survey data, are discussed formulating new research perspectives in view of more detailed land hosting capability evaluation to be conducted in the project framework.

4.2 Materials

4.2.1 Study area

Our attention focused on the DRC (fig. 2A). Here, the administrative subdivision is as follows: the national territory is divided into Provinces, the province into Districts and Cities, the district into Territories, the territory into Local Communities (or collectives or collectivities), the local community into Groups, and the group into Villages (Comité Provincial-SRP du Nord-Kivu, 2005).

In addition we have to take in consideration a supplementary hierarchical subdivision of the national territory, namely the ecclesiastical administrative subdivision. In fact, in this paper, the administrative level of collectives is used in order to merge the data from the RS and the data from the Census which are instead collected at lower levels of the ecclesiastical subdivision (*Shirika*).

The study area corresponds to the rural area of the GD, which is part of the Ecclesiastical province of Bukavu and which is composed of 23 parishes. Each parish is composed of several branches. These branches, also called in Swahili '*Wigigi*' ('*Kigigi*' is the singular), are not an exclusively ecclesial administrative units, because they correspond also to a pre-existing territorial division based on customary traditions. In each branch's territory, people are grouped in local ecclesial communities called *Shirika*. *Wigigi* in the GD are numbered at 105. Inside each collectivity, the project estimates an average number of *Shirika* equal to 305, but this number varies according to the fact that *Shirika* are dynamic groups of people and not once-for-all-defined territorial units.

Specifically, the target area of ARDST project, namely the GD (surface: 26 223 km^2), lays in the North Kivu Region and in a tiny strip of South Kivu Region, from Minova to Nyabibwe (fig. 2B). Inside the Diocese there are 11 collectives appertaining to 5 territories (Goma, Rutshuru, Masisi, Walikale, Kalehe).

The present paper focuses on three collectives belonging to the GD: Bakano, Bashali and Wanyanga. These administrations cover respectively the 16.2% (4 251 km^2), the 4.4% (1 144 km^2) and the 39.9% (10 487 km^2) of the Diocese (fig. 2C).

114

Figure 4.2 – Geographic positioning of the three collectives representing the investigated study sites. The DRC, Africa (A); the Goma Diocese, in the Congo DRC' North Kivu region (B); the collectives of Wanyanga, Bakano and Bashali.
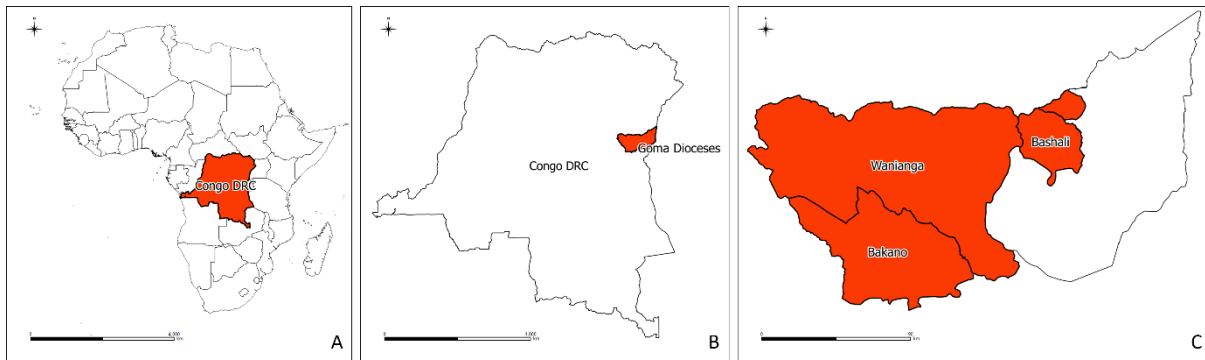
4.2.2 Ground reference data

In the Nord Kivu region, different land tenures correspond to different land use/cover patterns: while large concessions result in large, uniform territorial units typically covered by monoculture plantations or pastures, smallholders typically own small plots which result in fragmented, heterogeneous territorial units. This diversity in land use and ownership determines a meaningful land cover pattern in the region.

A data collection campaign was conducted in May 2017 in the GD with the purpose of collecting GPS points concerning different land use classes. A total of 110 polygons of surveyed areas were drown from the GPS point sample collection and successively used to train and test supervised classification algorithms. The polygons are representative for a total area of 904.3 ha. Table 4.1 summarizes the main features of the database by listing the four land use classes as well as the number of polygons drown for each class and the respective area of concern.

Table 4.1 - Summary of the information collected in the geo-database produced through a field survey activity carried out in May 2017.

Observed land use	Available polygons	Total area [ha]
Agriculture	39	343.4
Anthropic	25	131
Forest	12	20.6
Pasture	34	409.3
Total	110	904.3

4.2.3 Satellite data

Eight Sentinel-2A satellite images, covering together the entire surface of GD were downloaded from the European Space agency (ESA) Copernicus open access hub archive

(https://scihub.copernicus.eu/). In particular, we exploited level 1C (L1C) Sentinel images, provided as a set of tiles of 100 square kilometres. These products contain applied radiometric and geometric corrections (including orto-rectification and spatial registration). We selected the less cloud-contaminated acquisitions for each tile during the reference year 2017. Details on spatial resolution and the different spectral acquisition bands available in Sentinel-2A L1C images are shown in Figure 4.3.

Figure 4.3 - Spatial resolution versus wavelength intervals of Sentinel-2A images. Sensor span of 13 spectral bands from the visible and the near-infrared to the short wave infrared at different spatial resolutions (pixel size) ranging from 10 to 60 m. (source http://esamultimedia.esa.int/docs/Earthobservation/Sentinel-2_ESA_Bulletin161.pdf)
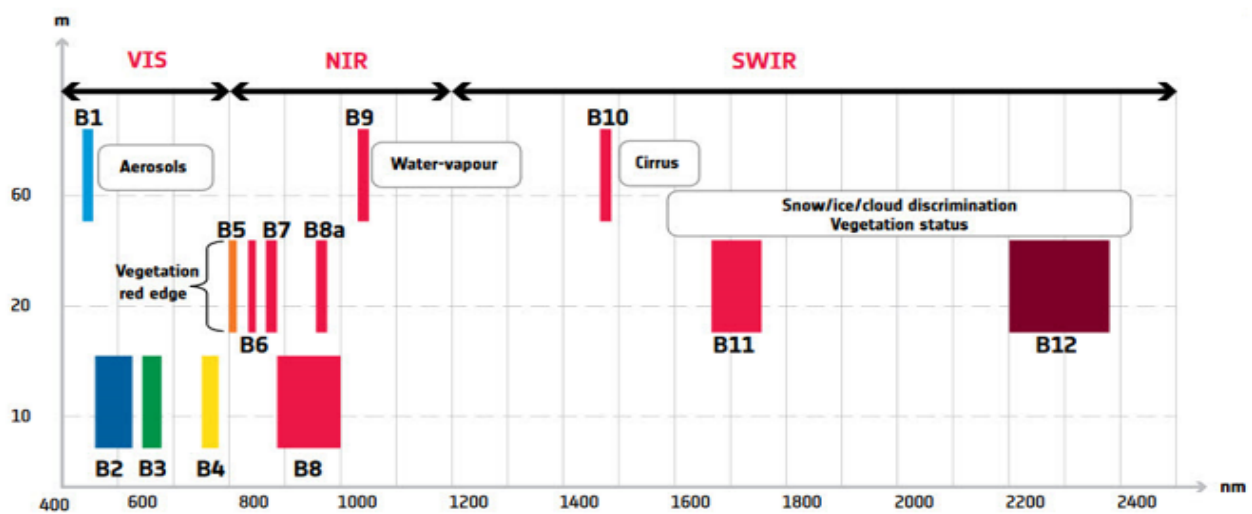
A Digital Elevation Model (DEM) provided by the NASA Shuttle Radar Topographic Mission (SRTM - http://srtm.csi.cgiar.org) was used to gather information about altitude and slopes characterizing the area of study (Jarvis, H.I., Reuter, Nelson, & Guevara, 2016).

4.2.4   Census information

The ARDST project has a strong preliminary component related to sensitization about land tenure issues in front of EDP/IDP return. In the framework of these activities aimed at building a positive environment for the implementation the whole project, a survey in the GD was carried out by 8 field supervisors and 23 field surveyors of Caritas Goma NGO. The structure of the survey form is resumed in Table 4.2. The survey aimed at collecting at least one interview per each *Shirika* in the GD, which are estimated to be about 305 per collectivity. The survey was scheduled between May 2016 and December 2017. This paper reports the results of the first 339 interviews that were dispensed between May and August 2016 in 339 *Shirika* laying in three collectives, as reported in Table 4.3.

116

Table 4.2 - Structure and contents of the Survey form used for gathering information about existing agricultural resources, conflicts, and potentialities

Topics	Questions	Description
General information on the *Shirika*	4	Features of the *Shirika* (with a focus on co-existing ethnic groups).
Land ownership and land use	2	Land uses and land owners.
Soil erosion	9	Soil erosion: its presence and its intensity
Agricultural land use	2	Agriculture: crops, depending on the type of land owner.
Information about land tenure in owned/vacant lands	3	Land tenure on owned/vacant lands.
Existing value chains (production-marketing)	5	Local production/consumption chain.
Other activities	5	Other peculiar activities eventually existing in the *Shirika*
Externally displaced people and land	8	Willingness of resident population to manage the re-settlement of externally displaced people.
Internally displaced people and land	8	Willingness of resident population to manage the re-settlement of internally displaced people.

Table 4.3 - Number of interviews per Collectivity.

Province	Territory	Collectivity	Interviewed chiefs of *Shirika*
Nord Kivu	Walikale	Bakano	29
Nord Kivu	Masisi	Bashali	167
Nord Kivu	Walikale	Wanyanga	143
Total			339

4.3 Method

The applied methodology involved an independent processing of two different data sources and their successive integration in GIS environment. On one hand, remote sensed satellite images were analyzed with the aim of producing land use maps of the three collectivities of interest. On the other hand, census data were gathered, organized in a database and finally summarized at collectivity level. Once retrieved at the same spatial level, data sources were integrated by overlapping them in a GIS environment and were used to build a SDSS based on MCE of land hosting capability. Figure 4.4 proposes a flow chart of the main methodological steps followed in the proposed methodology.

Figure 4.4 - Flow-chart summarizing the main phases of the designed methodology.
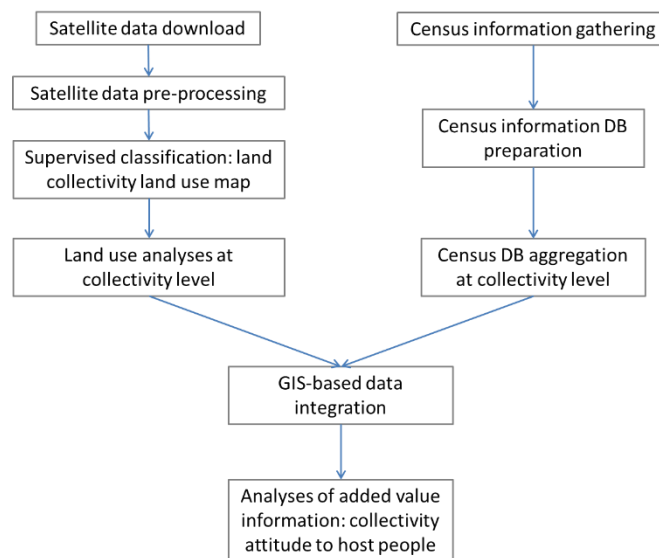
4.3.1 Analyses of Satellite data

A selection and download of the best available cloud free Sentinel-2 images were conducted on the ESA web catalog (January-June 2017 time window). Eight images entirely covering the GD were chosen.

A preliminary image pre-processing phase was carried out on selected images by resampling all the images spectral bands components at a 10-m resolution and subsequently by creating a composite mosaic as cloud-free as possible of the GD.

The mosaic was then classified using a Spectral Angle Mapper (SAM) supervised classifier (Kruse et al., 1993). All the Sentinel-2A bands with native spatial resolution of 10-m and 20-m were used as independent input variables in the classification procedure.

We used 50% of the field data to train the algorithm (calibration) and the other 50% for testing the land use map (validation) produced by the supervised classification. Four target classes were formalized to be automatically mapped by the SAM algorithm. The classes were: (a) pastoral land, (b) forest, (c) agricultural land and (d) anthropic (built) land areas.

A thematic land use map of the GD represented the output of the classification phase. A pixel-to-pixel comparison between the validation dataset (reference data) and the land use map (estimated data) was then conducted using the error matrix methodology (Brivio, Lechi, & Zilioli, 2006; Congalton, 1991). After the validation, three map resizes were made in correspondence of Wanyanga, Bakano

118

and Bashali communities and statistics related to the percentage distribution of the estimated land use classes were compiled.

4.3.2   Analyses of Census data

Data from the first 339 paper questionnaires were preliminarily analyzed through participative approach: one *ad hoc* focus group was organized in collaboration with the whole survey staff in Caritas Goma headquarter during May 2017. This first participatory analysis of the data served as a first-glance assessment of the integrity and internal consistency of the survey data. This step made it possible to highlight some critical issue in the survey form and in the data flow toward sources integration. The 339 questionnaires were manually digitalized and recorded in an electronic database. Among the possible aggregation levels, determined by the existing administrative units and related GIS reference data, the collectivity (or collective) level was chosen as a good compromise between clarity of the exposition and ease in the overlay of the data from satellite and census. After the creation of this preliminary database and the choice of a consistent aggregation level, data were analyzed with the software Stata (StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.).

4.3.3   Integration of remote sensed and census data

The integration of data form Census and RS was approached in two steps:

Census information were gathered at collectivity level, as it is shown in Table 4.5 (Results section). Aggregated data were mapped on GIS and used to build a simple synoptic table, as show in Figure 4.7 (Results section). The interpretation of the emerging relations between the data from the two sources was analysed in ad hoc participatory sessions with the aim of systematically explaining the relations.

4.4   Results

4.4.1   Collective-level land use maps

Figure 4.5 shows Sentinel-2A reflectance average response to four land use classes of interest in Goma as depicted in the calibration database. Pastoral land use (Figure 4.5A) showed the higher spectral signature response (i.e. the most reflective land use typology, Figure 4.5E), the signature was similar to the forest and agricultural land uses, but with higher magnitude. Forest spectral signature

(Figure 4.5B) behalf similarly to the agriculture land use (Figure 4.5C), with differences mainly marked in the Short Wave Infra-Red region (SWIR) between 1200 nm to 2200 nm and corresponding to Sentinel-2 bands b11 and b12. For this portion of the spectra, forest land covers were more reflective than the agricultural one (Figure 4.5E). Anthropic areas showed the lower spectral signature (i.e. the less reflective, Figure 4.5E), with the lover differences between the Red (around 650 nm, Sentinel-2 band b4) and NIR (around 850 nm, Sentinel-2 band b8) regions due to the absence of photosynthetic activity and having almost the same reflectance values from 800 nm to 2200 nm (Figure 4.5D). All the spectral signatures presented reflectance maximum values in the NIR region (Sentinel-2 band b8).

Figure 4.5 – Training dataset. Average Sentinel-2 reflectance responses and standard deviations for four land use classes. (a) Pastoral land use, (b) forest land use, (c) agricultural land use, (d) anthropic land use, (e) all the target classes together.
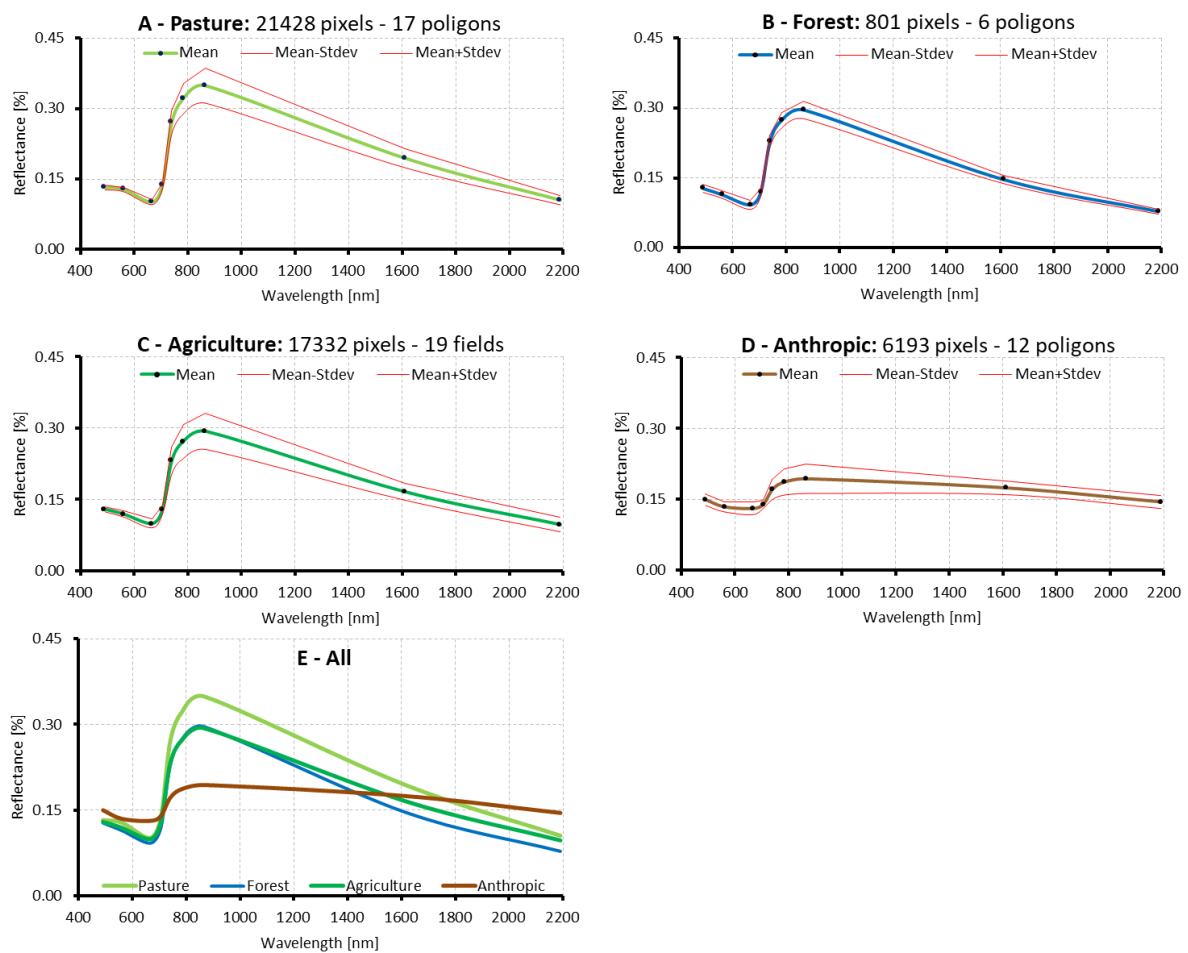

Table 4.4 shows the validation results of the retrieved land use map for the study area of the Goma. This classification result has to be considered a preliminary methodological exercise, mainly devoted to set-up the broad project's methodological framework rather than maximize its single component

performances. Given this, the classification reached an overall accuracy of 66% meaning 66/100 cases of agreement between field data and RS classification. The anthropic and pastureland use classes were mapped with acceptable levels of omission and commission errors, lower or equal to 29%. Further improvements are needed with reference to both agricultural and forest land use estimations, which was conservatively classified in view of decreasing mutual misclassification.

Table 4.4 – Error matrix showing a pixel-by-pixel comparison between the validation dataset (reference dada) and the thematic classification map (estimated data). CE: commission error, OE: omission error, OA: overall accuracy.

	Validation data				Accuracy indicators		
	Agriculture	Anthropic	Forest	Pasture	CE	OE	OA
Land use map:	[px]	[px]	[px]	[px]	[%]	[%]	[%]
Agriculture	17802	2079	528	4337	28	42	66
Anthropic	51	7565	0	0	1	29	-
Forest	9615	652	1350	551	89	28	-
Pasture	3472	351	8	14647	21	25	-

The SAM classifier allowed the production of a land use map of the study area. This map was resized for target collectives: Wanyanga (Figure 4.6A), Bashali (Figure 4.6B) and Bakano (Figure 4.6C).

With reference to the whole GD (map not reported), the classification mapped a relevant presence of forest that counts for 74% of the study area (around 19 405 km^2). The agricultural land uses were instead identified on 20% of the map surface (5 244 km^2) and were mainly concentrated near rivers and roads. Pastoral and anthropic land uses were mapped in low percentages, with 5% (1 311 Km^2) and 1% (262 km^2) respectively. For the collectivity of Wanyanga (Figure 4.6A) a higher presence of forest (+13%) and a lower presence of agricultural land use (-8%) were found in comparison with the entire Diocese. In the Eastern part of the collectivity, a higher presence of agricultural land use was mapped. In the central part instead, it is possible to note some cloud and cloud shadow noises. These represent for our purposes an unavoidable loss of land use information. Figure 4.6B shows the land use map of Bashali. This collectivity has land use proportions in line with the Goma study area. Although a marked presence of cloud contamination, it is possible to note a predominant presence of forests in the Western side as well as a predominant presence of agricultural land use on the Eastern side. In the Northeastern part of Bashali the algorithm identified the presence of pastoral land use and in the Southeastern the presence of three main water bodies. These latter, are presumably volcanic lakes. The land use map of Bakano (Figure 4.6C) reported a strong presence of forest (15% more than the entire GD) and a minor presence (-9%) of agricultural land use. This was mainly concentrated in the central part of the map, along the East to West direction, and in the South.

In addition to the land use information, we also obtained data concerning average (AVG) elevation and slope for the study area of Goma and for the three collectives we focused on. The SRTM dataset showed for Goma AVG elevation of 1299 m and 3.33% AVG slope. Weather for the three collectivities: 1026 with 2.59% (Wanyanga), 1626 with 4,53% (Bashali) and 1056 with 2.44% (Bakano) for the AVG elevation and the AVG slope respectively.

Figure 4.6 – Resize of the land use map of GD for the three collectives object of this study: (A) Wanyanga, (B) Bakano and (C) Bashali. The legend reports four target land uses (agricultural, anthropic, forest and pastoral), map sources of noise (cloud and cloud shadow) and water bodies.
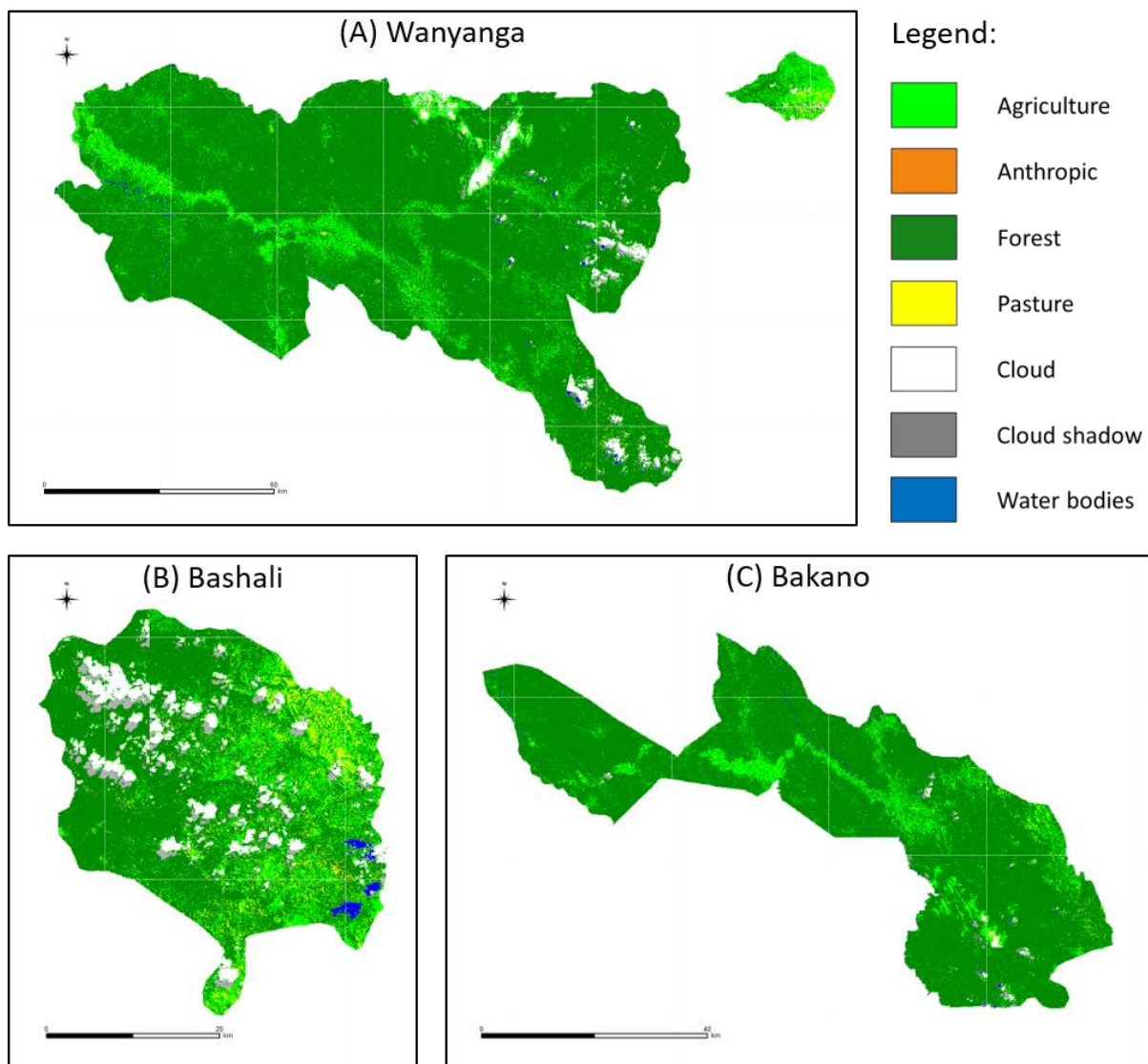
4.4.2   Aggregation of census data at the collectivity level

Table 4.5 shows some results of the census for the three target collectivities.

In relation to its surface, Bashali is the most densely populated community and respondents mention few erosion problems, seasonal water scarcity, few conflicts over the land and apparently no land appertaining to EDP/IDP. Moreover, according to respondents, Bashali shows the lowest propensity to the return of EDP/IDP even if vacant lands seem to be available.

The Collectivity of Bakano is characterized by the lowest population density, low level of conflict even if vacant lands are not available, seasonal water scarcity, no presence of lands appertaining to EDP/IDP and no will for them to settle back.

The collectivity of Wanyanga shows 13,8 inhabitants for square kilometre, diffused erosion phenomenon, seasonal water scarcity, presence of conflicts over the lands even if vacant lands seem to be available. Similarly to the other two other collectivities Wanyanga's respondents are not willing to accept the return of EDP/IDP.

Table 4.5 - Census results aggregated for the three target collectivities.

	Bakano		Bashali		Wanyanga		Total	
Surface (Sq.km)	4251		1144		10487		15882	
Population (Inhab.)	16584		17292		145024		178900	
Pop. density (Inhab./Sq.km)	3,9		15,1		13,8		11,3	
	Bakano		Bashali		Wanyanga		Totale	
Number of interviews	Count	%	Count	%	Count	%	Count	%
	29	9%	167	49%	143	42%	339	100%
Question	Answers							
1. Is soil erosion a problem in the area?	Count	%	Count	%	Count	%	Count	%
N	13	45%	58	35%	108	76%	179	53%
Y	16	55%	109	65%	35	24%	160	47%
Tot.	29	100%	167	100%	143	100%	339	100%
2. Is there water yearly-round (as flow or well) ?	Count	%	Count	%	Count	%	Count	%
N	1	3%	6	4%	4	3%	11	3%
Y	28	97%	158	96%	139	97%	325	97%
Tot.	29	100%	164	100%	143	100%	336	100%
3. Are there any vacant lands?	Count	%	Count	%	Count	%	Count	%
N	10	34%	153	92%	121	85%	284	84%
Y	19	66%	14	8%	22	15%	55	16%
Tot.	29	100%	167	100%	143	100%	339	100%

	Bakano		Bashali		Wanyanga		Total	
4. Do conflicts over land exist in the area?	Count	%	Count	%	Count	%	Count	%
N	5	17%	76	46%	97	68%	178	53%
Y	24	83%	89	54%	45	32%	158	47%
Tot.	29	100%	165	100%	142	100%	336	100%
5. Are there any land appertaining to EDP/IDP in the area?	Count	%	Count	%	Count	%	Count	%
N	0	0%	8	6%	0	0%	8	4%
Y	15	100%	134	94%	70	100%	219	96%
Tot.	15	100%	142	100%	70	100%	227	100%
6. Is the resident population willing to accept the return of EDP/IDP?	Count	%	Count	%	Count	%	Count	%
N	9	38%	16	11%	57	42%	82	27%
Y	15	63%	133	89%	79	58%	227	73%
Tot.	24	100%	149	100%	136	100%	309	100%

4.4.3 Data integration

The data coming from the two sources (Census and Remote Sensing) were integrated in a synoptic table (Figure 4.7), also through the ARCGIS software by using the 'collectivity' administrative unit on GIS (MONUSCO, 2016). As anticipated, an ad hoc focus group was organised in Goma in order to share the synoptic table and to collectively interpret the crossing of the two sources of information.

Figure 4.7 - Data integration for the two sources, presented as histograms of different features and corresponding coloured layers of the three target collectives (Bakano, Bashali and Wanyanga).
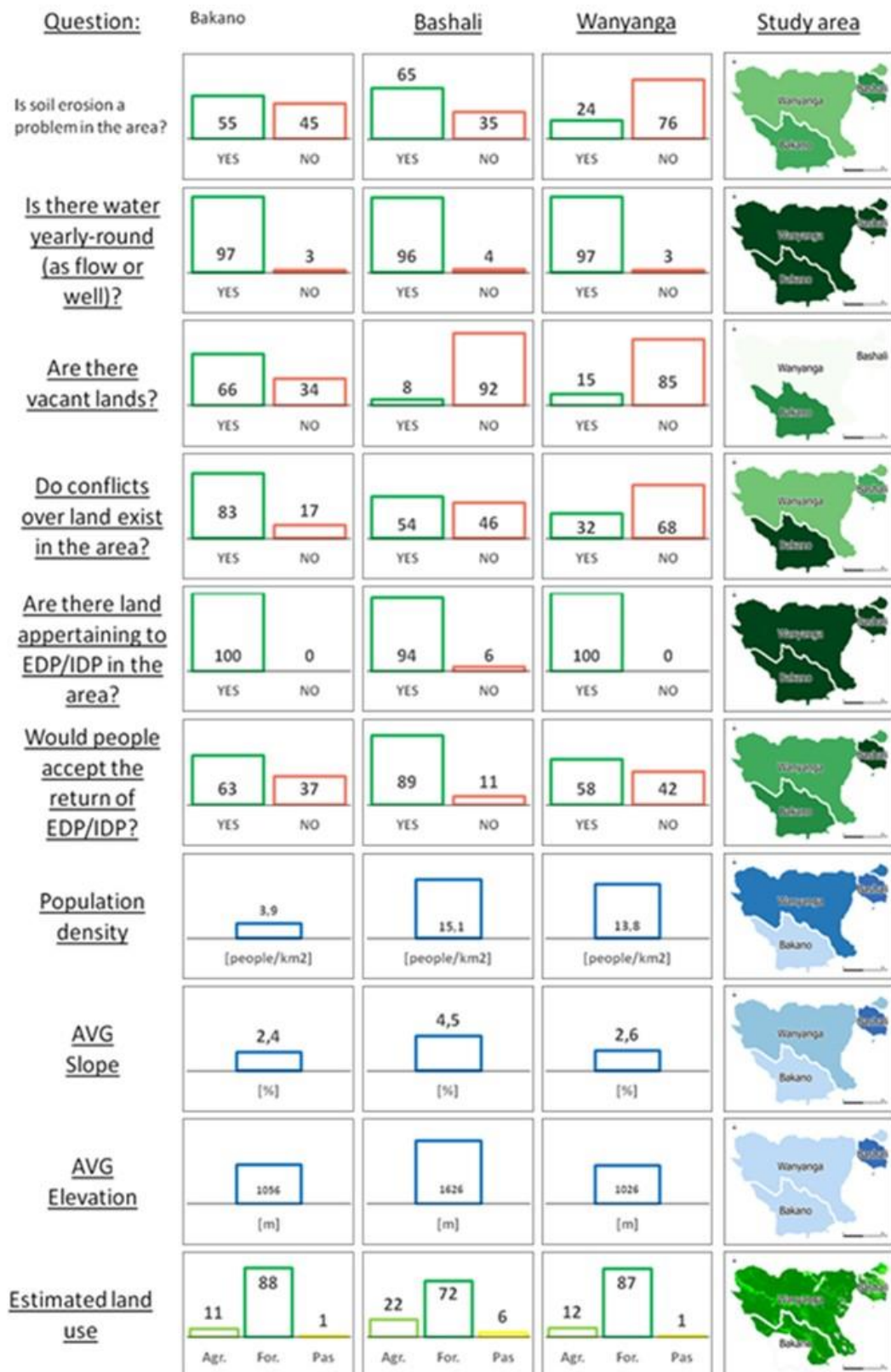
4.5 Discussions and conclusions

Through the implementation of the proposed methodology, we obtained the integration of different sources of information, respectively based on remotely-sensed images and on the field survey.

The proposed methodology, even in this preliminary phase, is able to overlay different data sources and therefore allows multicriteria evidence-based considerations.

In fact, the overlay of information provided by the two sources produced 'added' information that is useful to support the MCE of land hosting capacities and the identification of potentially suitable pilot areas.

It is interesting to remark that the coupling of the two information sources is able to support the process, as a SDSS, in different ways and at different degrees. In fact, the overlay of the two sources produces results that can be classified according to the following three cases.

1. Case of mutual confirmation: the two sources confirm each other and therefore bring to a stronger knowledge base.

2. Case of contradiction: the two sources contradict each other. This case underlines critical issues in the project' knowledge base or in the proposed methodology for data integration. These cases hints at the need of further research.

3. Case of 'added value': it represents the main goal of the methodological approach. The two sources of information allow the creation of additional information 'over the simple sum of the parts', which leads to more complex evaluations.

As an example of this third case, which is strongly desirable in the study, we can make the case of Bashali collectivity. The proposed methodology works as a SDSS tool by integrating the data related to all the proposed criteria (Figure 4.1). In particular, if we only consider the knowledge collected during field visits, Bashali is the richest collectivity among the three target collectivities and therefore is one of the sound zones where to propose a pilot and where the 'suitable conditions for the refugees' return may already exist'. If we take in consideration the results of RS analyses, at first glance, the Bashali collectivity has a higher presence of agricultural and pastoral land and could therefore be confirmed as the best collectivity where to propose a pilot. On the contrary, coupling the information from RS with census data, it clearly appears that Bashali is also the most densely

126

populated collectivity, the one with more lands appertaining to missing EDP/IDP and the one that shows the lowest presence of vacant lands. Even if most of the respondents in Bashali show a positive attitude toward the return of refugees, the composite information we can draw from the overlay of sources may explain the medium-high presence of conflicts over land use. Then, if we also take in consideration the average slope of the collectivity, we can add another item to the reflexion and therefore resize our interest on this collectivity to be the host of one of the pilots.

This evidence based, 'added knowledge' about Bashali is possible only if projects like the ARDST adopts a SDSS based on MCE, as the one drafted in the present paper.

Further research will be conducted in the framework of the ARDST project, and more detailed information to support the choice of pilot zones will be produced. A next step will be the use of a multi-temporal analysis approach in the analysis of satellite images. As a matter of fact, agricultural and pastoral land extensions in Bashali are higher than in other collectives (as can be observed by RS). As we know from literature and as we confirmed with our census, Bashali collectivity, and Masisi territory as a whole, suffers from chronic disputes over land tenure, because all the lands are already exploited. Land conflict in Masisi are often generated by the dichotomy farmers-pastoralists where farmers are in majority smallholders while pastoralist areas are dominated by huge, under-exploited, concessions. The RS methodology implemented in this first step-methodological approach is still not able to produce land cover classes on the base of the difference between landscape patterns, but this result will be obtained with multi-temporal analysis in the second phase of the project. The information obtained, joined with information collected by census about vacant lands and existence/features of large concessions will allow to strengthen the proposed SDSS.

However, it is crucial to stress that, in a complex situation as in Nord Kivu, any decision about the re-settlement of IDP/EDP should be preceded by a strong consultation process with local people and institutions, and in this sense the SDSS results can constitute only a discussion base for supporting these operations.

4.6 References

ADB. (2009). *Democratic Republic of Congo. Economic and sector work. Regional economic development in Bas Congo in the context of decentralization in the Democratic Republic of Congo (DRC)*. Kinshasa: African Development Bank Country operations department Centre Region (ORCE).

Agatsiva, J., & Oroda, a. (2002). Remote Sensing and Gis in the Development of a Decision Support System for Sustainable Management of the Drylands of Eastern Africa : a Case of the Kenyan Drylands Abstract : *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *34*, 42–49. https://doi.org/ISSN/ISBN:1682-1750

Baker, T., Miller, S., & Paige, G. (2009). The Spatial Environmental and Agricultural Decision Support System (SEADS). In *SUMAWA Mau Forest Complex Conference*. Njoro.

Baldyga, T. J., Miller, S. N., Driese, K. L., & Gichaba, C. M. (2008). Assessing land cover change in Kenya's Mau Forest region using remotely sensed data. *African Journal of Ecology*, *46*(1), 46–54. https://doi.org/10.1111/j.1365-2028.2007.00806.x

Brivio, A., Lechi, G., & Zilioli, E. (2006). *Principe e metodi di Telerilevamento*. Milan: CittàStudi Edizioni.

Butt, A., Shabbir, R., Ahmad, S. S., & Aziz, N. (2015). Land use change mapping and analysis using Remote Sensing and GIS: A case study of Simly watershed, Islamabad, Pakistan. *The Egyptian Journal of Remote Sensing and Space Science*, *18*(2), 251–259. https://doi.org/10.1016/j.ejrs.2015.07.003

Bwana, S. M. (2011). *Democratic Transition in the Democratic Republic of Congo After The Cold War. Mobutu Sese Seko and Laurent-Désiré Kabila's Leadership Scrutinized*. (C. Fernandez, Ed.). Lund: Lund University.

Cardille, J. A., & Foley, J. A. (2003). Agricultural land-use change in Brazilian Amazônia between 1980 and 1995: Evidence from integrated satellite and census data. *Remote Sensing of Environment*, *87*(4), 551–562. https://doi.org/10.1016/j.rse.2002.09.001

Chakhar, S., & Mousseau, V. (2008). Multicriteria Decision Making, Spatial. In S. Shekhar & H. Xiong (Eds.), *Encyclopedia of GIS* (pp. 747–753). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-35973-1_839

CIRGL. (2006). *Déclaration de Dar-es-Salaam sur la paix, la sécurité, la démocratie et le développement dans la région des Grands Lacs*. Dar-es- Salaam: Conférence Internationale sur la Paix , la Sécurité , la Démocratie et le Développement dans la Région des Grands Lacs.

Comité Provincial-SRP du Nord-Kivu. (2005). *Document de stratégie de réduction de la pauvreté*. Goma: Comité Provincial-SRP Nord-Kivu.

Congalton, R. G. (1991). A Review of Assessing the Accuracy of Classification of Remotely Sensed Data A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment*, *4257*(September), 34–46. https://doi.org/10.1016/0034-4257(91)90048-B

Dale, V. H., Oneill, R. V, Pedlowski, M., & Southworth, F. (1993). Causes and Effects of Land-Use Change in Central Rondonia, Brazil. *Photogrammetric Engineering and Remote Sensing*, *59*(6), 997–1005.

Duveiller, G., Defourny, P., Desclée, B., & Mayaux, P. (2008). Deforestation in Central Africa: Estimates at regional, national and landscape levels by advanced processing of systematically-distributed Landsat extracts. *Remote Sensing of Environment*, *112*(5), 1969–1981. https://doi.org/10.1016/j.rse.2007.07.026

Frankema, E., & Buelens, F. (2013). *Colonial exploitation and economic development: The Belgian Congo and the Netherlands indies compared*. (E. Frankema & F. Buelens, Eds.). London: Routledge. https://doi.org/10.4324/9780203559406

Hansen, M. C., Roy, D. P., Lindquist, E., Adusei, B., Justice, C. O., & Altstatt, A. (2008). A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sensing of Environment*, *112*(5), 2495–2513. https://doi.org/10.1016/j.rse.2007.11.012

ITC. (2009). Final Report: Changing Hydro-meteorological Risks – as Analyzed by a New Generation of European Scientists. In *Applications Guide for Determining the Yield Strength of In-Service Pipe by Hardness Evaluation: Final Report* (pp. 1–66). Published in Europe: ASME Press. https://doi.org/10.1115/1.802915.ch1

Janusz, J. (2016). Fast simulation tool: SDSS for housing development and more. *International Journal of Sustainable Development and Planning*, *11*(1), 39–48. https://doi.org/10.2495/SDP-V11-N1-39-48

Jarvis, A., H.I., Reuter, A., Nelson, A., & Guevara, E. (2016). Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database. *CGIAR CSI Consortium for Spatial Information*, 1–9. https://doi.org/http://srtm.csi.cgiar.org

Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., & Goetz, A. F. H. (1993). The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, *44*(2–3), 145–163. https://doi.org/10.1016/0034-4257(93)90013-N

Landmatrix. (2017). Land deals in the DRC. Retrieved June 18, 2019, from www.landmatrix.org/

Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2004). Change detection techniques. *International Journal of Remote Sensing*, *25*(12), 2365–2407. https://doi.org/10.1080/0143116031000139863

Manakos, I., & Braun, M. (2014). *Land Use and Land Cover Mapping in Europe*. *Land Use and Land Cover Mapping in Europe* (Vol. 18). https://doi.org/10.1007/978-94-007-7969-3

Mathys, G., & Vlassenroot, K. (2016). It's not all about the land. Land disputes and conflict in the eastern Congo. *Rift Valley Institute PSRP Briefing Papers*, *14*.

MONUSCO. (2016). MONUSCO GIS database. Goma.

Naidoo, S. (2003). *The War Economy in the Democratic Republic of Congo*. *IGD Occasional Papers*. Braamfontein: Institute for Global Dialogue.

Power, D. J., Sharda, R., & Burstein, F. (2014). Decision Support Systems. In *Wiley Encyclopedia of Management* (pp. 1–4). Hoboken: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118785317.weom070211

Refsgaard, J. C., Ramaekers, D., Heuvelink, G. B. M., Schreurs, V., Kros, H., Rosén, L., & Hansen, S. (1998). Assessment of 'cumulative'uncertainty in spatial decision support systems: Application to examine the contamination of groundwater from diffuse sources (UNCERSDSS). In *European Climate Science Conference, Vienna* (pp. 19–23). Vienna: European Commission.

Rundquist, D. C., Narumalani, S., & Narayanan, R. M. (2001). A review of wetlands remote sensing and defining new considerations. *Remote Sensing Reviews*, *20*(3), 207–226. https://doi.org/10.1080/02757250109532435

Uyan, M., Cay, T., & Akcakaya, O. (2013). A Spatial Decision Support System design for land reallocation: A case study in Turkey. *Computers and Electronics in Agriculture*, *98*, 8–16. https://doi.org/10.1016/j.compag.2013.07.010

Vancutsem, C., Pekel, J. F., Evrard, C., Malaisse, F., & Defourny, P. (2009). Mapping and characterizing the vegetation types of the Democratic Republic of Congo using SPOT VEGETATION time series. *International Journal of Applied Earth Observation and Geoinformation*, *11*(1), 62–76. https://doi.org/10.1016/j.jag.2008.08.001

Verweijen, J., & Marijnen, E. (2016). The counterinsurgency/conservation nexus: guerrilla livelihoods and the dynamics of conflict and violence in the Virunga National Park, Democratic Republic of the Congo. *The Journal of Peasant Studies*, 1–21. https://doi.org/10.1080/03066150.2016.1203307

Zhang, S. Q., Zhang, S. K., & Zhang, J. Y. (2000). A study on wetland classification model of remote sensing in the Sangjiang Plain. *Chinese Geographical Science*, *10*(1), 68–73. https://doi.org/10.1007/s11769-000-0038-1

5 Simplified pairwise ranking for agricultural development in Masisi, Nord Kivu, Democratic Republic of Congo[18]

5.1 Introduction

Development Cooperation initiatives, in all their relevant phases from planning through implementation, monitoring and final or ex-post evaluation, often involve complex real-world decision-making problems. Agri-environmental and generally landscape-scale issues are the typical target of participative decision-making procedures as they involve community resources planning (Allain et al., 2017). To successfully solve the complex real-world problems multi-criteria group decision-making approaches are recognized as reliable and effective (Guerra et al., 2015; Baffoe, 2019).

Building on several background well-known theories such as the Social Innovation Theory (Chambon et al., 1982), the Socio-ecological System Theory (Ostrom, 2009), and keeping in mind that Multicriteria Decision-making methodologies (MCDM) are bound to evaluation processes (Reynolds et al., 2010), the present paper investigates the use of MCDM in the framework of a formative evaluation and discusses the application of simplified form of pairwise ranking (SPWR) in a real-world case study related to agricultural development in the Collectivity of Katoyi, Nord Kivu, Democratic Republic of Congo (DRC). Attention is focused on assessing SPWR as a useful tool for building shared priority ranking and for widening consensus (Author et al., 2019).

Main research questions are: (1) What are the priorities for intervention in order to achieve sustainable development of the agricultural sector in the Collectivity of Katoyi, Nord Kivu, DRC? (2) Is SPWR a sound technique to prioritize different alternatives for intervention while improving group consensus in real-world complex cases?

Present section (paragraph 5.1) introduces the topic of SPWR in the framework of MCDM, existing relevant applications and existing methodological criticisms to its application. Section 2 (paragraph 5.2) describes the case study and describes the SPWR methodology. Section 3 (paragraph 5.3)

---

[18] This chapter contains the pre-print version of the paper by De Marinis P., Sali G., 2009. Simplified pair-wise ranking for agricultural development in Masisi, Nord Kivu, Democratic Republic of Congo, Submitted to Evaluation and Programme Planning, ISSN: 0149-7189, currently under review.

includes results from the case study. Finally, Section 4 (paragraph 5.4) discusses the results by commenting them in front existing critics to SPWR and hints at research perspectives on the topic.

5.1.1 MCDM and participation

Historically, first MCDM methodologies have gained the attention of academia and decision makers because of their ability to inform evidence-based decisions (Mardani et al., 2015).

Thanks to the stratification of knowledge on MCDM and to the changes occurred in the understanding of complex systems, we now understand that decision-making is positively influenced by participation (Dolan et al., 2007) especially in dealing with agricultural development (Julio Berbel, 2018). From this perspective the techniques to be used in the process of identification and evaluation of decision alternatives must be at the same time rigorous and fully opened to participation in order to inform evidence-based choices while building a consensus (Groselj, 2018).

5.1.2 Simplified pair-wise ranking, definition and applications

Simplified PWR (SPWR) is a simplified form of the pairwise ranking technique (PWR) (Narayanasamy, 2009): it helps a decision maker to binary pair-compare alternatives and to rank them according to their relative frequency of preference. Also called Priority Grid or Matrix, PWR provides a rational and structured approach to derive the group's priorities, whilst facilitating much negotiation and consensus building by the group (FAO, 2003 and 2006a).

SPWR therefore consists in the use of binary pairwise comparisons as stand-alone tools. While PWR is normally used in the framework of Analytic Hierarchy Process (AHP) by asking DMs 'how many times A is more important than B?', SPWR simply asks, 'What is better, A or B?'

Moreover, while AHP deconstructs the decision into a hierarchy and builds Pariwise Comparison Matrixes (PCMs) for each level and node, SPWR simply compare alternatives against a global goal. This is the reason why it is also called binary PWR or simplified PWR.

SPWR is widely used by DC practitioners in order to evaluate and rank local resources, problems, solutions and impacts of development interventions worldwide in the sector of agricultural development (IFAD, 2000; DFID, 2002; FAO, 2006a; UNESCO, 2008; World Bank, 2018). SPWR is a simplified version of the pairwise ranking (PWR) technique that is used in the framework of Analytic Hierarchy Process (AHP) (T.L. Saaty, 1980; 2008) and namely it is often referred to as 'binary AHP'

(Jensen, 1986; Nishizawa, 1995; Iida, 2009) . AHP is in turn one of the leading techniques in the wide family of Multicriteria Decision Making or Aiding (MCDM or MCDA) methodologies (Ho et al., 2018; Ishizaka et al., 2011). MCDM methods and tools regroup a long and increasing list of methodologies whose aim is to help decision makers in identifying suitable solutions to a given question in a complex context (Koksalan et al., 2013; Mardani et al., 2015; Steuer et al., 2016).

Literature witnesses that SPWR and AHP can be used with a bottom-up and participative approach (Ishizaka et al., 2011; Keeney et al., 2011), working as a consensus-building tool, bringing the group of stakeholders to discuss on single criteria and weights and to understand each other point of view (Aznar et al., 2011; Russo et al., 2015).

Despite many applications of AHP to real-world cases exist, DC organizations tend to neglect the use of more complex MCDM methods such as AHP and AHP-nested PWR. These organizations prefer a simplified form of PWR (SPWR) when dealing with the planning and implementation of development projects (FAO, 2006a; Russel, 1997; World Bank, 2018).

Practitioners of development projects in agriculture are interested in practical and rapid appraisals of needs which might be shared and accepted by the local community. In this context, a good degree of consensus is an important outcome as much as the consistency of the resulting list of items. Here is to be found the reason why the SPWR is preferred: the technique grants the possibility to run rapid brainstorming and focus groups with all the interested stakeholders and to produce participatory outcomes in terms of a ranked list of preferences, might they be needs or priorities.

To our knowledge few previous studies explicitly deal with SPWR applied to agricultural development, with participatory approach within the framework of DC initiatives. These scientific contributions deal with three topics, namely the participatory selection of crop varieties (Shanthy et al., 2010), the use of participatory rural appraisal technique (Koralagama et al., 2007) and the livestock/veterinary sector assessment (Catley et al., 1996; Kebede et al., 2003; Stroebel et al., 2008).

5.2 Materials and methods

5.2.1 Case study: Evaluation of agricultural potential and identification of triggering interventions in the Diocese of Goma, Nord Kivu, Democratic Republic of Congo

The study area, in the framework of what is called the 'geologic scandal' of the DRC (International Monetary Fund, 2014; RAID, 2007), is the Diocese of Goma (GD) and lies in the eastern region of the DRC, at the border with Uganda and Rwanda, in the Nord Kivu province. It covers some 25000 sq.km, it is inhabited by some 1,443,000-total people and it is interested by chronic insecurity due to the presence of several active armed groups (MONUSCO, 2018).

Among specific objectives of the ARDST project, our research group was asked to contribute to the strategic evaluation of the agricultural sector aimed at identifying triggering interventions for the sustainable development of the agricultural sector in the GD.

Agriculture is the leading economic activity in the region even if yields are low and the lack of security both in terms of land tenure and social security determine huge constrains to local development. Main other critical issues that prevent endogenous development are the insufficient road network, the low technical level of farmers and advisory services, the increasing loss of fertile lands due to soil erosion, lack of food processing, unfair value chains, scarce performance of farmers' associations, conflicts for land use. These issues strike in different ways the two existing productive systems within the regional agricultural sector: a business-oriented system based essentially on tea cropping and cattle husbandry installed on huge surfaces and a more diversified and fragmented subsistence agriculture system mainly focused on cassava, beans, corn, sugar cane, legumes, bananas, potatoes, sweet potatoes and yam (De Marinis et al., 2019a).

Reaching sustainable agricultural development in such a complex context demands specific approaches in order to achieve both the needed community involvement and robust results. Following numerous international organization guidelines (IFAD, 2000; DFID, 2002; FAO, 2006a; UNESCO, 2008; World Bank, 2018), we selected the SPWR technique in order to implement a participative and operational decision-making process about priorities for interventions.

In order to appropriately approach the identification of triggering interventions, the whole technical team of the ARDST project was asked to join a participative decision process animated through SPWR (Narayanasamy, 2009; Russel, 1997). This participatory path started February 2018 focusing priorities

for intervention in the area of Katoyi, Masisi, DRC. Fourteen development practitioners with different backgrounds and expertise participated to the SPWR meeting: four are agronomists, four are jurists, three are experts in sustainable rural development, one is a sociologist professor, one is a civil engineer, one is a cartographer

Alternative interventions were previously extrapolated from the results of a 'preparation' phase in which the agricultural sector of the GD was analysed through mixed methods. Eight alternatives were then validated and further described in order to build common understanding of the theory of change (Brest, 2010) connecting intervention to the overall objective. Workflow is resumed in figure 5.1.

Figure 5.1 - Workflow of SPWR application on the case study.

135
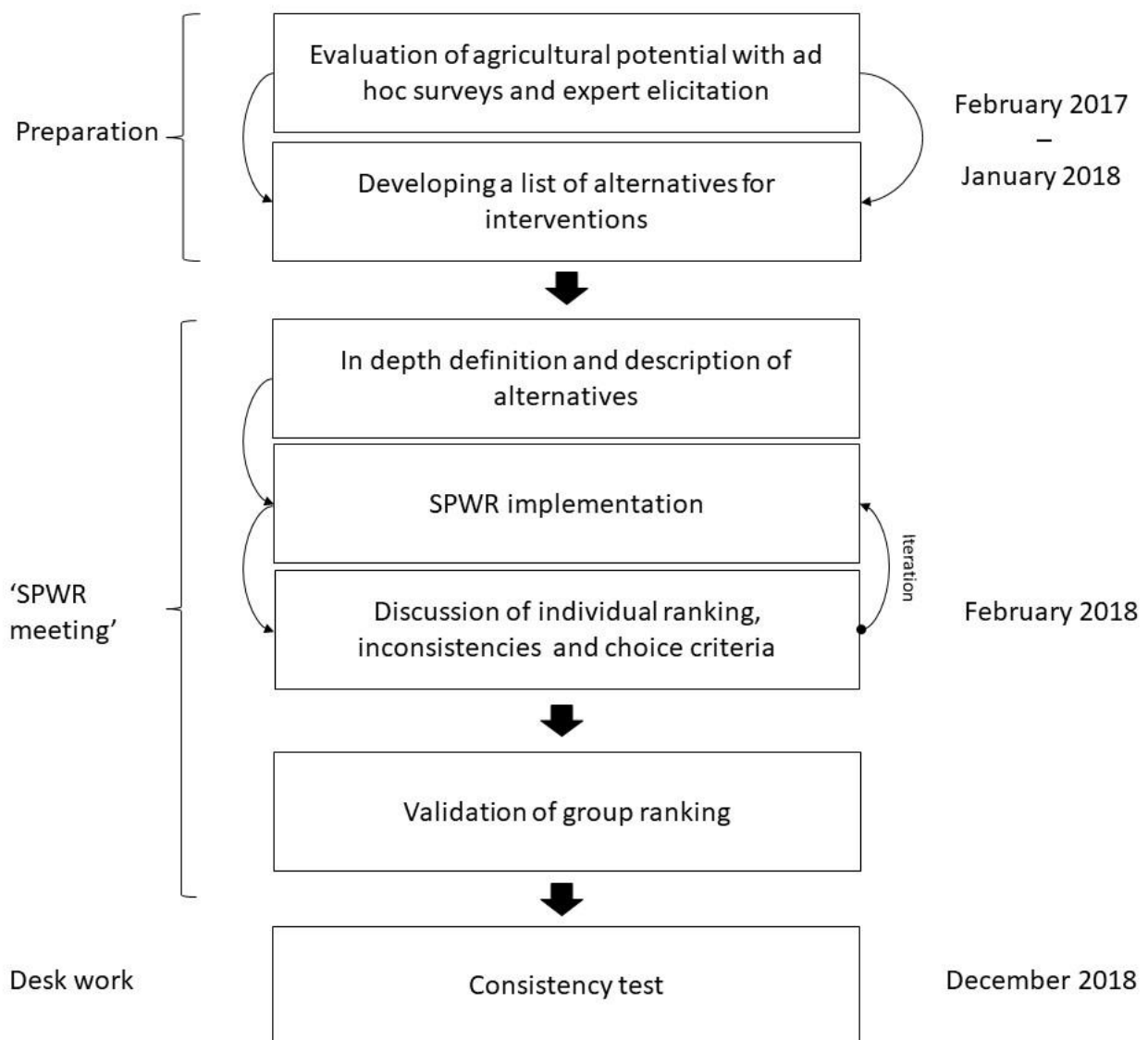
5.2.2 Simplified Pairwise ranking

Pairwise Ranking or Preference Ranking is a tool to set priorities between different options. In Pairwise Ranking, each individual item is compared directly against the others so as to emerge with a ranking from highest (best) to lowest (least) (FAO, 2006a). Table 5.1 resumes the main objectives and key questions targeted by PWR according to FAO.

Table 5.1 – Objectives and key questions in SPWR, from FAO (FAO, 2006a).

Objectives	• to determine the main preferences/ priorities/ needs/ constraints or problems of individuals or groups for a set of items/ subjects/ topics • to compare the priorities of different groups against one another • to support a group of people to discuss, analyse and prioritize their problems, needs and possible solutions ñ to find out about criteria for decision-making.
Key Questions	• Which item out of several ones is looked upon as most important/ favourable/ necessary/ pressing bay a certain group? • Which are the criteria for preferring one item to another? • How different are the preferences between different groups within the local population?

SPWR methodology was implemented during a specific focus group in order to build a priority ranking. Proceedings were noted carefully especially concerning the personal rankings, personal criteria behind the scoring and global ranking. During the final discussion phase, apparent inconsistencies in Decision Makers' (DMs) PCMs were pointed out and used in order to ask the participants to refine their judgement and to further discuss different points of view.

To construct a PCM, each activity is compared in turn with each of the others. Participants ask themselves: 'What's more important to trigger agricultural development in Katoyi, A or B?' Each participant to the working group expresses his/her own preference by choosing one of the two activities and marking it on the PCM. This is repeated until all problems have been compared with each other and the upper-right half-part of the PCM is full. For each preference participants are asked to mention the criteria upon which their decision has been made. The remaining lower-left part of the PCM is filled with the reciprocal values of the upper-right part. Scores are computed on the total number of times each activity is selected as 'the best' in the upper-right half of the PCM. Personal ranks are then computed according to the scores and reveal personal preferences or priorities. Individual scores (an example is provided in Table 5.2) were aggregated in order to build a group ranking of the alternatives.

Table 5.2 - Example of one PCM (one participant).

Activity	Activity Number								Score	Rank
	1	2	3	4	5	6	7	8		
1.Technical training and follow-up on production	1	1	1	1	5	6	1	1	5	**3°**
2.Producers aggregation	1	2	3	4	5	6	7	2	1	7°
3.Support to local R&D institutes in agriculture	1	3	3	3	5	6	3	3	4	4°
4.Improved varieties (supply and training/follow-up)	1	4	3	4	5	6	4	4	3	5°
5.Roads building	5	5	5	5	5	6	5	5	6	**2°**
6.Support in animal husbandry management	6	6	6	6	6	6	6	6	7	**1°**
7.Support in value chains improvement	1	7	3	4	5	6	7	7	2	6°
8.Markets (assessment and structural improvement)	1	2	3	4	5	6	7	8	0	8°

5.2.3   SPWR consistency treatment

Consistency of PCMs was treated in two separate moments of the workflow (see Figure 5.1). Firstly, the inconsistencies emerging during the final discussion held in the SPWR meeting were used to elicit more in-depth discussion and to improve the group consensus. Secondly, acknowledging the major critics[19] that PWR and SPWR techniques receive within the scientific debate, we used the approach of Graph Theory (Bang-Jensen et al., 2010) to run a consistency check at individual participant level. It is important to state here that we performed this consistency test only after the SPWR meeting held in Goma in February 2018. In fact, if the aim of the test is to select only consistent PCMs (and DMs), during the SPWR meeting we were not interested in eliminating participants but instead in valorising emergent inconsistencies in the discussion. Then, when this first goal was achieved during the meeting, results of the consistency test were useful to improve our comprehension of conflicting values among the participants, to measure the quality of participants' contributions along the decision process and to evaluate the relevance of identified inconsistencies in the final group ranking.

We used the consistency test proposed by Iida (2009). Iida (2009), adapting a more general theorem by Kendall et al. (1940), demonstrates a specific theorem for the analysis of internal consistency of binary AHP and namely of nested pairwise comparison matrixes.

The consistency test follows two steps:

(1) Calculation of d (number of cyclic triads in the preference matrix digraph) and comparison of d with standard thresholds

---

[19] See Section 4 (paragraph 5.4) for the discussion of SPWR against its major methodological critics'.

(equation 1) $\qquad d = \frac{n(n-1)(n-2)}{6} - \frac{1}{2}\sum_{i=1}^{n} a_i^2$

Where n is the number of pairwise-compared elements or the number of rows/columns in the square preference matrix.

(2) Calculation of a consistency coefficient ζ that is previously defined (Kendall et al., 1940) as

(equation 2) $\qquad \zeta = 1 - \frac{24d}{n^3-n}$ $\;$ if n is odd, $\;\zeta = 1 - \frac{24d}{n^3-4n}$ $\;$ if n is even,

Where d is the number of cyclic triads in the digraph produced by the preference matrix under exam.

Iida (2009) explains that '*It is easy to see that 0 ≤ ζ ≤ 1 and ζ = 1 if and only if there is no circular triad in the digraph. As ζ decreases to zero, the inconsistency, which is measured by the number of circular triads, increases. So ζ can be used as a measure of consistency in a preference matrix. Clearly the objects may be completely ranked linearly by the matrix if and only if ζ = 1.*'

Following (Kendall et al., 1940), Iida (2009) produces a table containing the threshold number of cyclic triads in a digraph to be used in order to state the consistency of a PCM. The threshold varies according to the number of alternatives in the PCM or nodes in the digraph and for our specific situation it is equal to seven.

Under these circumstances, we calculated d and ζ for our preference matrixes by using equations 1 and 2.

5.3 Results

5.3.1 Ranking results

An example of the PCMs produced during the SPWR meeting is given in the previously presented Table 5.2.

Preliminary global group ranking (Table 5.3) is then computed by adding the score obtained by each activity from each participant. Final group ranking is produced through iterative discussion of recorded global preferences. At each round of iterative discussion, a higher degree of consensus was facilitated. The validated group ranking is shown in Table 5.4.

Validated global results (Table 5.4 'whole group') show that intervention priority, according to the whole working group, is (1) Roads building, (2) Technical training and follow-up on production, (3)

Improved varieties (supply and training/follow-up), (4) Producers aggregation, (5) Markets (assessment and structural improvement), (6) Support in value chains improvement, (7) Support in animal husbandry management, (8) Support to local R&D institutes in agriculture. These results were discussed within the working group and validated. Criteria behind the resulting ranking were asked at the stage of personal ranking presentation, noted and then discussed collectively.

During the final desk phase and namely at the end of the consistency test, one of the PCM went lost, therefore was discarded (participant 13). Producing the final ranking after the consistency test and therefore by considering only consistent PCMs reveals a rank reversal for alternatives ranked 4°,5°,6°,7°. It means that alternatives in the first three and in the last ranks remain the same. Reversals occur between Intervention 8 « Markets » and Intervention 2 « Farmers organizations », and between Intervention 6 « Animal husbandry management » and Intervention 7 « Value chains improvement». The reversal is significant, but it doesn't prevent from considering the ranking validated by the group at the end of the SPWR session inferior in terms of suitable guide for the decision, because even if a ranking is produced with not-completely-consistent judgements, it may still be used as a base for a new discussion and decision (Iida, 2009).

Table 5.3 - Preliminary ranking computed on individual rankings through simple sum of ratings.

	P. 1	P. 2	P. 3	P. 4	P. 5	P. 6	P. 7	P. 8	P. 9	P. 10	P. 11	P. 12	P. 13	P. 14	P. 15	Score TOT	Rank TOT
1.Technical training and follow-up on production	5	6	6	4	4	3	3	6	3	2	6	3		3	2	56	2°
2.Producers aggregation	1	0	4	2	4	7	5	2	4	2	5	6		3	6	51	4°
3.Support to local R&D institutes in agriculture	4	3	1	2	5	0	2	0	0	4	0	1		3	0	25	8°
4.Improved varieties (supply and	3	5	2	5	3	4	4	5	2	3	6	5		4	2	53	3°

	P. 1	P. 2	P. 3	P. 4	P. 5	P. 6	P. 7	P. 8	P. 9	P. 10	P. 11	P. 12	P. 13	P. 14	P. 15	Score TOT	Rank TOT
training/fol low-up)																	
5.Roads building	6	5	6	7	5	4	2	6	7	5	5	5		7	4	74	1°
6.Support in animal husbandry manageme nt	7	3	4	2	1	2	4	5	2	2	2	1		2	3	40	7°
7.Support in value chains improvem ent	2	1	3	4	1	3	5	1	5	5	3	4		5	4	46	6°
8.Markets (assessme nt and structural improvem ent)	0	5	2	2	5	5	3	3	5	5	1	3		1	7	47	5°

Table 5.4 - Group final ranking (validated through participatory discussion) compared with the ranking obtained (a-posteriori) by eliminating inconsistent PCMs.

RANK	Final 'whole group' ranking validated through discussion		Final '7 consistent DMs' ranking	
1°	Interv. 5	Roads building	Interv.5	Roads building
2°	Interv. 1	Technical training and follow-up on production	Interv.1	Technical training and follow-up on production
3°	Interv. 4	Improved varieties (supply and training/follow-up)	Interv.4	Improved varieties (supply and training/follow-up)
4°	Interv. 2	Producers aggregation	Interv.8	Markets (assessment and structural improvement)
5°	Interv. 8	Markets (assessment and structural improvement)	Interv.2	Producers aggregation
6°	Interv. 7	Support in value chains improvement	Interv.6	Support in animal husbandry management
7°	Interv. 6	Support in animal husbandry management	Interv.7	Support in value chains improvement
8°	Interv. 3	Support to local R&D institutes in agriculture	Interv.3	Support to local R&D institutes in agriculture

140

5.3.2 Consistency results and treatment

Our objective in applying SPWR was not to extract data from the group but instead to run SPWR as a consensus-building technique which is capable, at the same time, to produce quantitative data about the discussion. For this reason, the consistency test was run after the participatory session and its results were not used during the session itself. Iida (2009) method for consistency check allowed us to compute values of d and ζ (number of cyclic triads in a preference matrix and significance of the test) for each participant and to decide -a posteriori- to what extent participants to the SPWR session were able to linearly prioritize the alternatives.

Table 5.5 shows each participant performance.

Table 5.5 - Performance of all the participants in terms of consistency of their expressed preferences. **d value over 7 indicates participants were not able to linearly prioritize the alternatives (Iida, 2009).*

Participant	d	ζ
P1	0	1
P8	2	0.9
P11	2	0.9
P15	3	0.85
P9	4	0.8
P2	5	0.75
P6	6	0.7
P4	9*	0.55
P12	9*	0.55
P14	9*	0.55
P5	11*	0.45
P3	12*	0.4
P10	14*	0.3
P7	16*	0.2

5.4 Discussion

According to the working group, roads are the 'most triggering' intervention due to the fact that the actual condition of the road network *'already prevents the marketing of the existing production'*[20]. Road status constitutes one of the main bottlenecks of the whole development issue in the region (Hesselbein, 2007; Roberts, 2018). Considering the transversal impact of roads network on all the

---

[20] This and all the following quotations in italic have been collected during the SPWR session by Pietro De Marinis.

other issues tackled by the other ranked interventions, the commentary on this intervention may be understood as a background for the whole dissertation in this paragraph.

In fact, main criterion for choosing 'roads building' as absolute priority is repeated by participants: road network bad status is the first bottleneck for the rural community in terms of displacing and transport. Both people and goods must travel at very low speed and high costs, mainly due to the incidence of accidents, breakdowns and illegal tolls. This result match existing literature about the state of roads in Nord Kivu province (OCHA - ReliefWeb, 2018;) and can be confirmed by our personal inspection in February 2018.

The need for 'roads building' is a symptomatic example of '*the frustrating situation of the DRC*'[21] in which 'the *absence in the best case and the antagonism in the worst*'[22], of the state institutions, due to the 'kleptocratic state class'(Elsenhans, 1977, 1981), is mentioned continuously.

The following ranked interventions share at least one important criterion: all the most quoted interventions share the importance of human capital building as an important feature. According to the working group, this kind of intervention is preferred in respect of more material aid.

The second-ranked intervention is about technical training to farmers and local technicians which was described as a composite intervention including best new practices, supply of technical inputs for practical training and the creation of at least one permanent local training centre as reference for follow-up stages. Main reason behind the weight attributed to this intervention may be found in the relatively low yields obtained by local farmers and breeders. According to the working group the main reason for the need for external support in this context is not the lack of good local technicians but their ineffectiveness due to chronic insufficient funding. The second criterion is the unsustainable trend in soil degradation that follows a short-term oriented use of the land. Among other cited criteria it is important to underline the feeling about the plausible impact of technical training to farmers which may directly trigger community empowerment and consequently improve its direct implication

---

[21] Statement by M. Vulambo Kalissa, Chef de division DPDD-NK (Division du Plan de Développement Durable du Comité Provincial du Nord Kivu), vulkasoyi@gmail.com – extracted from the interview held in Goma on the 21th of February 2018 and conducted by De Marinis Pietro in the framework of the ARDST project.
[22] Statement by M. Murairi Janvier, president of ASSODIP asbl (Association pour le developpement des initiatives paysannes), assodipkivu@yahoo.fr – extracted from the interview held in Goma on the 31th of July 2018 and conducted by De Marinis Pietro and Gatti in the framework of the ARDST project.

in local development. This sense of 'autarchic local development' is once more due to the broadly perceived absence of state institutions in tackling local development challenges.

The third mentioned intervention is 'improved varieties' and include the supply of improved seeds and the training about on-site and participatory varieties improvement. The inclusion of supply and training in the intervention has been the result of a passionate discussion about the low long-term impact of improved seeds delivery as consequence of both hybrid varieties sterility in second generations and inbreeding yield depression phenomenon. So, training is cited once more as a tool for triggering long-term development of the region, while the existing quest for a short-term impact is behind the importance attributed to improved seeds delivery.

According to the working group the main criterion behind the importance of intervening on improved varieties is the belief in the triggering potential of a tool which is provided directly to the community, with only a 'satellite' participation of local research institutes and authorities, always seen as actors of an institutional system with chronic attitude towards corruption, diversions and misappropriation of public and humanitarian funds (Hidalgo, 2015; United States Department of State, 2011).

The fourth-ranked intervention is related to 'producers aggregation', here understood as a set of activities aimed at (1) building the sense of community which is the root of collaboration among actors of the same productive community, expected to tackle sector internal competitiveness and to trigger mutualism and cooperation towards common goals (2) training producers on the bureaucratic process for the creation and administration of associations and cooperatives, (3) following up and animating associations and cooperatives over time. Main criteria behind the importance accorded to this intervention may be resumed as '*unity is strength*' and are related to the belief that '*aggregating the producers has multiple emergent impacts such as social unity improvement, cooperation and conflict reduction*' which is indeed a very important impact in such a region where conflicts are common and in a sector that sees the continuous fight between farmers and animal breeders for the use of the land (Mathys et al., 2016). This is also clearly stated by multiple participants when comparing this intervention with the one related directly to animal husbandry management: '*Because, given that animal husbandry is a matter of large surfaces which needs community collaboration, producers' organizations are seen as a possible solution in order to build social cohesion and collaboration*' [3]. Moreover, an aggregated productive base is believed to have more power in terms of self-development, here understood as the ability of a community to maintain local

infrastructure such as roads and markets, to approach the improvement of value chains, to be targeted by further research and development initiatives and so on.

The importance deemed by participants to farmers' organizations in agricultural development matches scientific literature and reports from development organizations (Abaru et al., 2006; Abebaw et al., 2013; Bachke, n.d.; Bernard et al., 2009; SOS Faim, 2014), but the rank obtained by this intervention also seems to match the specific situation of farmers' organizations in DRC, '*which is not as developed as in other African countries*'[23].

The following triplet shares the need of improving community planning and collaboration for common goals such as 'value chain improvement through food conditioning/transformation' and 'better market structures and strategies'. This goal may stem spontaneously from the empowerment of the community and mostly from the strengthening of social cohesion which may be reached through more branched producers' organizations.

The fifth and sixth-ranked interventions share both several criteria and several interlinks in the description of related activities. Indeed, the two interventions were described as complementary: intervention 7 ('value chains') included the topic of 'transformation' and of 'collaboration between actors along the value chain', while intervention 8 ('markets') was more concerned with marketing strategies and market infrastructure building or maintenance. These interventions were scarcely prioritized but when they were, it has been on the base of a '*let's valorise what we already have*' imperative criterion, meaning that improvement in marketing, in a broad sense, is perceived as secondary to the augmentation of production. Anyway, the average ranking position may be explained by the fact that, on the one hand, the working group shared the belief that agricultural development may be triggered by 'excellence farmers or breeders', interested in and already capable of producing and marketing for cash on large scale, on the other hand, participants were aware of the importance of the wide base of smallholders whose economy may be hugely increased by working on process innovation such as the value chain's performance improvement and agricultural products processing.

---

[23] Statement by Jean Baptiste Musantamana, FOPAC Nord Kivu/AJAC RDCongo (Association des Journalistes Agricoles) – extracted from the interview held in Goma on the 9th of August 2018 and conducted by De Marinis and Gatti in the framework of the ARDST project.

The seventh-ranked intervention is about 'animal husbandry management external support' meaning direct intervention in aiding animal breeders not to fall in the struggle with farmers. This intervention was expected to reach higher ranks in such a context because the Nord Kivu province and namely the collectivity of Katoyi, which lies in the Masisi territory, are well known for huge cattle pastures and poor smallholder farmers.

The positioning of this intervention has been explained as a matter of self-organization attitude at community level concerning this issue: *'animal husbandry issue should not be targeted from the external because if it is, it would translate in subtracting the community from its own responsibility and finally prevent the only long-term solution to the issue, which is self-organization'*. It was also mentioned that *'the resolution of the chronic struggle between pastoralists and farmers is not that simple as it may appear because it is a multi-scale conflict: small breeders may be in conflict against smallholder neighbour farmers because their own few animals, usually managed by contracted shepherds in huge rented pastures owned by large land owners, are not managed properly'*. The lack of collaboration between large pastures owners and smallholders, both farmers and breeders, translates in a multiscale conflict in which no one has the resources nor the will to intervene. In such a context, one participant said, *'the intervention of external actors in financing or implementing fences or trenches would result only in a short-term impact without solving the lack of collaboration among the users of the land'*.

The last ranked intervention is the support to local research. The intervention was meant to build local research capacities that should operate as engines of innovation. Despite this description, local institutions are perceived as '*lost in corruption*' and so not worthy external support.

5.4.1   Consistency test results and role of inconsistencies during the participatory SPWR

Results show that 7 out of 14 participants (50%) were not able to linearly prioritize the alternatives. Kendall et al. (1940) considered the following possible explanation for not-linear ranking in binary PWR: *'(a) The observer may be a bad judge. (b) Some of the objects may diverge by amounts which fall below the threshold of distinguishability for the observer. (c) The property under judgement may not be a linear variate at all. (d) Several of the effects may be operating simultaneously.'*

Taking into consideration the level of strong expertise of participants, recorded levels of intransitivity may be due the fact that some of the objects in comparison differ by amounts which fall below the

threshold of distinguishability for the observer or to the fact that compared factors have properties which are not linear at all. Both the arguments are indeed realistic and may coexist because, even though holistic judgement is used, the difference between two 'incomparables' alternatives such as 'roads building' and 'technical training' may not be perceived nor used to inform a consistent preference. Moreover, 'incomparable alternatives' also show features that are not linearly related and therefore can't be ranked linearly at all.

It is important to remind that the seven 'bad judges' were an important asset for eliciting the discussion because they were the object of several criticisms during the discussion towards the validation of a shared final ranking. Their 'mistakes' fuelled the discussion and, given the constructive environment in which the discussion took place, we can't really blame the existence of these inconsistencies, nor of the participants.

5.4.2   Weaknesses and arguments in favour of SPWR

Given the important role of inconsistencies in our study, both in terms of enriching value and biasing effect on some of the DMs judgements, from the methodological point of view, we have to take in consideration the discussion existing in scientific literature about the lack of robustness concerning PWR, and SPWR to a higher degree. The discussion focuses on the inconsistency of the results of PWR/SPWR (Tran, 2013; IFAD, 2015; Kou et al., 2016; Krejčí et al., 2018; Kułakowski, 2018;) and criticisms can be resumed in the following 4 issues:

Firstly, PWR often brings people to compare elements that are not comparable from a cognitive point of view because it does not implement a strict structuration of the decision. For example, how can we compare a 'hard' intervention such as 'roads paving' with a 'soft' intervention such as 'technical training for farmers'?
Even if this observation has been found relevant in several fields of study (Kuester et al., 2015; Orsi, 2015) a supportive answer comes from Hiesh (2007), who explains that non-comparative, but still useful to choose, attitudes may be used when comparing two incomparable items in order to select the better one. In other words, incomparable things can be compared upon indirect effect of their incomparable features.
Secondly, SPWR elicit holistic perception, without any hierarchical conceptualization of the criteria against which the elements are pairwise compared.

Thirdly, SPWR is also criticized because it does not give people the possibility to fade their preference and simply put people in front of a black-or-white choice. Of course, it is a simplification, but we may say, a simplification that is good for the sake of obtaining rapid results and collecting multidisciplinary or multi-stakeholders' opinions. In fact, the sensitivity lost in avoiding internal weighting is regained through the possibility of running the whole technique in a participatory meeting.

Finally, SPWR is criticized because it does not include a consistency test procedure and therefore it does not take in consideration bias generated by intransitivity in pairwise comparisons, by decision agent's indifference or preference incompleteness about alternatives.

Anyway, strict transitivity, often called linear transitivity, is considered an ideal outcome of preference expression even by the first theorist of AHP (Saaty, 1984). Other authors (Wedley, 1993; Nurmi, 2014) state in real-life cases it is misleading to associate complete and transitive preference relations with rationality and robustness of the technique, especially if one assumes that AHP and other techniques such as PWR are meant to aid the decision and not to substitute the decision maker. All these critics are realistic and were met in the implementation of SPWR, but eventually the experiment was successful both in terms of building new consensus about the decision and in terms of final ranking.

From a methodological point of view all mentioned critics may be answered thanks to the existing debate about performances of choice's decomposition models and holistic judgements. Even if some authors demonstrated that choice decomposition models and robust weight elicitation models outperform holistic judgements in ad hoc experiments (Larrick, 2008, Fischer et al., 1995), holistic ability varies among people and can be improved with training. It means that every judgement has a systematic error, and this stands true both for models and holistic judgements. Our study both confirms these critics and shows that human brain can bypass inconsistencies and go for a sound results in group decision-making. In fact, if the intrinsic non-linearity of the alternatives is the cause determining the need to discard 50% of participants' opinions, on the other hand, results show that 50% of involved DMs were able to 'compare incomparables' and were able to produce consistent judgements without decomposing the choice. They probably acted on the base of what Hsieh (2007) wrote about human brains being normally able to integrate, even if with occasional errors, considerations that do not pertain to the strictly comparative value of two alternatives, granting the possibility to choose between 'strictly incomparable things' and 'across multiple criteria'.

According to Iida (2009), the whole group priority result can be used because *'even if a preference matrix is not completely logically consistent according to tests, and therefore we think that*

147

alternatives cannot be ranked linearly basically or because judges weren't able enough, we still don't need to completely reject the ranking by the binary AHP.' In fact, we can use it as a reference ranking in situations where complexity plays the king's role. Such a ranking may be difficult to use blindly as a solid stategical reference but still it has been produced with the participation of multiple stakeholders and disciplines. For this reason, it does have the ability to compose the elements of the decision and also the group of decision makers itself by building consensus.

5.5   Conclusions

In general terms, the ranking obtained with SPWR and validated by the working group, matches the priority list of several international organizations which point out the importance of improving agricultural productivity by the means of structural interventions based on farmers' organizations, technical training and market valorisation (FAO, 2017).

Despite poor scientific background of SPWR, the technique was found able to elicit the multidisciplinary group, to analyse the composite issue of agricultural development in the Collectivity of Katoyi and to build a consensus ranking of interventions that matches with existing recognized guidelines.

5.5.1   Research perspectives

In general terms our results show that further research is needed in order to adapt MCDM to the diversity of complex situations typical of primary sector development. Given both the scarce scientific background of SPWR and its diffused use in the framework DC, further studies should focus on extensive application of simplified and participatory MCDM methodologies trying to blend robust methodologies with consensus-oriented techniques.

5.5.2   Lessons learned

Qualitative methodologies such as SPWR are very useful to elicit participation and consensus building but the implementation process takes large amounts of time. Constructive and formative discussion happens when there are no time constrains and the atmosphere inside the group is relaxed. In order to deploy the best conditions for implementation, a careful preparation should be carried out. Present experience was conducted by one researcher who animated the group decision-making

process but huge improvements in the overall workflow could be achieved with a team of animators and communicators.

5.6 References

Abaru, M. B., Nyakuni, A., & Shone, G. (2006). Strengthening farmers' organizations: the experience of RELMA and ULAMP. *ICRAF Working Paper*, *23*.

Abebaw, D., & Haile, M. G. (2013). The impact of cooperatives on agricultural technology adoption: Empirical evidence from Ethiopia. *Food Policy*, *38*(1), 82–91. https://doi.org/10.1016/j.foodpol.2012.10.003

Allain, S., Plumecocq, G., & Leenhardt, D. (2017). How Do Multi-criteria Assessments Address Landscape-level Problems? A Review of Studies and Practices. *Ecological Economics*, *136*, 282–295. https://doi.org/10.1016/j.ecolecon.2017.02.011

Aznar, J., Guijarro, F., & Moreno-Jiménez, J. M. (2011). Mixed valuation methods: A combined AHP-GP procedure for individual and group multicriteria agricultural valuation. *Annals of Operations Research*, *190*(1), 221–238. https://doi.org/10.1007/s10479-009-0527-2

Bachke, M. E. (2009). *Are farmers' organizations a good tool to improve small-scale farmers' welfare?* Ås: Norwegian University of Life Sciences.

Baffoe, G. (2019). Exploring the utility of Analytic Hierarchy Process (AHP) in ranking livelihood activities for effective and sustainable rural development interventions in developing countries. *Evaluation and Program Planning*, *72*, 197–204. https://doi.org/https://doi.org/10.1016/j.evalprogplan.2018.10.017

Bang-Jensen, J., & Gutin, G. Z. (2009). *Digraphs. Theory, Algorithms and Applications*. London: Springer London. https://doi.org/10.1007/978-1-84800-998-1

Bernard, T., & Spielman, D. J. (2009). Reaching the rural poor through rural producer organizations? A study of agricultural marketing cooperatives in Ethiopia. *Food Policy*, *34*(1), 60–69. https://doi.org/10.1016/j.foodpol.2008.08.001

Brest, P. (2010). *The Power of Theories of Change*. (Leland Stanford Jr. University, Ed.). Standford: Stanford Social Innovation Review.

Catley, A. P., & Mohammed, A. A. (1996). The use of livestock-disease scoring by a primary animal-health project in Somaliland. *Preventive Veterinary Medicine*, *28*(3), 175–186. https://doi.org/https://doi.org/10.1016/0167-5877(96)01038-0

Chambon, J.-L., Devevey, J.-M., & David, A. (1982). *Les innovations sociales* (1e édition). Paris: Presses universitaires de France.

De Marinis, P., & Sali, G. (2019a). *Rapport conclusif de l'étude des potentialités agricoles en Diocese de Goma. Projet ARDST*. Goma: Caritas Development Goma.

De Marinis, P., & Sali, G. (2019). Simplified pairwise ranking for identification of triggering interventions on agricultural development in Masisi, Nord Kivu, Democratic Republic of Congo. *Evaluation and Program Planning*, *under revi*ew.

Dolan, P., Edlin, R., Tsuchiya, A., & Wailoo, A. (2007). It ain't what you do, it's the way that you do it: Characteristics of procedural justice and their importance in social decision-making. *Journal of Economic Behavior & Organization*, *64*(1), 157–170. https://doi.org/https://doi.org/10.1016/j.jebo.2006.07.004

Elinor Ostrom. (2009). A General Framework for Analyzing Sustainability of Social-Ecological Systems. *Science*, *325*(July), 1962–1980. https://doi.org/10.1002/2016JE005088

Elsenhans, H. (1977). Die staatsklasse/staatsbourgeoisie in den unterentwickelten ländern zwischen privilegierung und legitimationszwang. *Verfassung Und Recht in Übersee / Law and Politics in Africa, Asia and Latin America*, *10*(1), 29–42.

Elsenhans, H. (1981). *Abhäaengiger Kapitalismus oder büuerokratische Entwicklungsgesellschaft. Versuche üueber den Staat in der Dritten Welt*. Frankfurt: Campus-Verl.

FAO. (2003). *Participatory Development: guidelines on beneficiary participation in agricultural and rural development*. (B. van Heck, Ed.) (2nd ed.). Rome: Rural Institutions and Participation Service, Rural Development Division Food and Agriculture Organization of the United Nations.

FAO. (2006). *Participatory Rural Appraisal (PRA) Manual*. (R. Paul, Ed.). Rome: Food and Agriculture Organization of the United Nations.

FAO. (2006). *Toolkit for the Application of Participatory and Negociated Territorial Development*. Rome: FAO Publishing and Multimedia Service.

FAO. (2017). *The State of Food and Agriculture - Leveraging Food Systems for Inclusive Rural Transformation*. Rome: FAO Publishing and Multimedia Service.

Fischer, G. W. (1995). Range sensitivity of attribute weights in multiattribute value models. *Organizational Behavior and Human Decision Processes*, *62*, 252–266.

Groselj, P. (2018). Consensus model in analytic hierarchy process. In Z. Zivkovic (Ed.), *XIV international conference on strategic management* (Vol. 14, pp. 24–35). BOR: Univ Belgrade.

Guerra-López, I., & Hicks, K. (2015). The participatory design of a performance oriented monitoring and evaluation system in an international development environment. *Evaluation and Program Planning*, *48*, 21–30. https://doi.org/10.1016/j.evalprogplan.2014.09.003

Hesselbein, G. (2007). The rise and decline of the Congolese State : an analytical narrative on state-making. *Crisis States Research Centre Working Papers*, *2*(21).

Hidalgo, S. (2015). *Country Report: DRC. Evaluation of the Common Humanitarian Fund.* Geneve: OCHA.

Ho, W., & Xin, M. (2018). The state-of-the-art integrations and applications of the analytic hierarchy process. *European Journal of Operational Research*, *267*, 399–414. https://doi.org/10.1016/j.ejor.2017.09.007

Hsieh, N.-H. (2007). Is Incomparability a Problem for Anyone? *Economics and Philosophy*, *23*(1), 65–80. https://doi.org/10.1017/s0266267107001228

IFAD. (2000). *Promoting Participation in Development Projects: IFAD's Experience in Promoting Participatory Approaches in Development Projects in Asia*. (M. T. L. Debuque, Ed.). Somewhere: IFAD publishing services.

IFAD. (2015). *Projet d ' Appui au Secteur Agricole dans la Province du Nord Kivu (PASA-NK)*. Rome: IFAD publishing services.

Iida, Y. (2009). The number of circular triads in a pairwise comparison matrix and a consistency test in the AHP. *Journal of the Operations Research Society of Japan*, *52*(2), 174–185. https://doi.org/10.15807/jorsj.52.174

International Monetary Fund. (2014). Democratic Republic of the Congo: Selected Issues. *Country Report*, (15). https://doi.org/10.1111/j.1365-3156.2008.02150.x.dhfr

Ishizaka, A., & Labib, A. (2011). Review of the main developments in the analytic hierarchy process. *Expert Systems with Applications*, *38*(11), 14336–14345. https://doi.org/10.1016/j.eswa.2011.04.143

JENSEN, R. E. (1986). Comparison of consensus methods for priority ranking problems. *Decision Sciences*, *17*(2), 195–211. https://doi.org/10.1111/j.1540-5915.1986.tb00221.x

Julio Berbel, Thomas Bournaris, Basil Manos, Nikolaos Matsatsinis, & Davide Viaggi. (2018). *Multicriteria Analysis in Agriculture*. New York: Springer International Publishing. https://doi.org/10.1007/978-3-319-76929-5

Kebede, F., Shiferaw, G., Admassu, B., & Berhanu, A. (2003). Participatory study on Afar traditional Camel production management and disease. *AGRIS FAO - EARO/Kombolcha Regional Veterinaly Laboratory*, 22.

Keeney, R. L., & Nau, R. (2011). A theorem for Bayesian group decisions. *Journal of Risk and Uncertainty*, *43*(1), 1–17. https://doi.org/10.1007/s11166-011-9121-5

Kendall, M. G., & Smith, B. B. (1940). On the method of paired comparisons. *Biometrika*, *31*(3–4), 324–345. https://doi.org/10.1093/biomet/31.3-4.324

Koksalan, M., Wallenius, J., & Zionts, S. (2013). An Early History of Multiple Criteria Decision Making. *Journal of Multi-Criteria Decision Making*, *77*(1), 209–223. https://doi.org/10.1002/mcda

Koralagama, D. N., Wijeratne, M., & De Silva, W. N. (2007). Emergence of participatory rural appraisal (PRA) technique as a strategy towards sustainable development: A Sri Lankan experience. *Journal of Agriculture and Rural Development in the Tropics and Subtropics*, *108*(2), 149–160.

Kou, G., Ergu, D., Lin, C., & Chen, Y. (2016). Pairwise comparison matrix in multiple criteria decision making. *Technological and Economic Development of Economy*, *22*(5), 738–765. https://doi.org/10.3846/20294913.2016.1210694

Krejčí, J., & Ishizaka, A. (2018). FAHPSort: A Fuzzy Extension of the AHPSort Method. *International Journal of Information Technology and Decision Making*, *17*(4), 1119–1145. https://doi.org/10.1142/S0219622018400011

Kuester, S., Feurer, S., Schuhmacher, M. C., & Reinartz, D. (2015). Comparing the incomparable? How consumers judge the price fairness of new products. *International Journal of Research in Marketing*, *32*(3), 272–283. https://doi.org/10.1016/j.ijresmar.2014.09.006

Kułakowski, K. (2018). Inconsistency in the ordinal pairwise comparisons method with and without ties. *European Journal of Operational Research*, *270*(1), 314–327. https://doi.org/10.1016/j.ejor.2018.03.024

Larrick, R. P. (2008). *Blackwell Handbook of Judgment and Decision Making*. Hoboken: Blackwell Publishing Ltd. https://doi.org/10.1002/9780470752937.ch16

Mardani, A., Jusoh, A., Nor, K. M. D., Khalifah, Z., Zakwan, N., & Valipour, A. (2015). Multiple criteria decision-making techniques and their applications - A review of the literature from 2000 to 2014. *Economic Research-Ekonomska Istrazivanja*, *28*(1), 516–571. https://doi.org/10.1080/1331677X.2015.1075139

Mathys, G., & Vlassenroot, K. (2016). It's not all about the land. Land disputes and conflict in the eastern Congo. *Rift Valley Institute PSRP Briefing Papers*, *14*.

MONUSCO. (2018). UN reports hundreds of human rights violations as security situation in North Kivu deteriorates. Retrieved April 22, 2019, from https://monusco.unmissions.org/en/un-reports-hundreds-human-rights-violations-security-situation-north-kivu-deteriorates

Narayanasamy, N. (2009). *Participatory Rural Appraisal: Principles, Methods and Application*. New Delhi: SAGE Publications India Pvt Ltd. https://doi.org/10.4135/9788132108382

Nishizawa, K. (1995). A consistency improving method in binary AHP. *Journal of the Operations Research Society of Japan*, *38*(1), 21–33. https://doi.org/10.15807/jorsj.38.21

Nurmi, H. (2014). Are We Done With Preference Rankings ? If We Are , Then What ? *Operations Research and Decisions*, *24*(4), 63–74. https://doi.org/10.5277/ord140405

OCHA - ReliefWeb. (2018). Roads snapshot in DRC. Retrieved November 18, 2019, from https://reliefweb.int/sites/reliefweb.int/files/resources/296FAADD2CD1C4EF852571B5005CD59C-unjlc-cod-24jul.pdf

Orsi, F. (2015). *Value Theory*. London: Bloomsbury.

RAID. (2007). *Democratic Republic of the Congo: Key Mining Contracts in Katanga: The economic argument for renegotiation*. Oxford: Rights and Accountability in Development.

Reynolds, M., & Holwell, S. (2010). *Systems Approaches to Managing Change: A Practical Guide*. London: Springer International Publishing.

Roberts, K. (2018). Understanding the Causes of Poverty in the DRC. *Borgen Magazine*.

Russel, T. (1997). Pair Wise Ranking Made Easy. In *PLA Notes* (pp. 25–26). London: International Institute for Environment and Development.

Russo, R. D. F. S. M., & Camanho, R. (2015). Criteria in AHP: A systematic review of literature. *Procedia Computer Science*, *55*(July), 1123–1132. https://doi.org/10.1016/j.procs.2015.07.081

Saaty, T. L. (1980). The Analytic Hierarchy Process. *Education*, 1–11. https://doi.org/10.3414/ME10-01-0028

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83. https://doi.org/10.1504/IJSSCI.2008.017590

Saaty, T. L., & Vargas, L. G. (1984). Inconsistency and rank preservation. *Journal of Mathematical Psychology*, *28*(2), 205–214. https://doi.org/10.1016/0022-2496(84)90027-0

SOS Faim. (2014). The need for a strong farmer movement in the DRC. *SOS FAIM - Farming Dynamics*, (34).

Steuer, R. E., & Zionts, S. (2016). Short MCDM History. *International Society on MCDM*.

Stroebel, A., Swanepoel, F. J. C., Nthakheni, N. D., Nesamvuni, A. E., & Taylor, G. (2008). Benefits obtained from cattle by smallholder farmers: a case study of Limpopo Province, South Africa. *Australian Journal of Experimental Agriculture*, *48*(6–7), 825–828. https://doi.org/10.1071/EA08058

Tran, N. M. (2013). Pairwise ranking: Choice of method can produce arbitrarily different rank order. *Linear Algebra and Its Applications*, *438*(3), 1012–1024. https://doi.org/10.1016/j.laa.2012.08.028

UNESCO. (2008). *Data Requirements for Integrated Urban Water Management*. (T. Fletcher & A. Deletic, Eds.). London: CRC Press.

UNHCR. (2019). DRC Situation. Retrieved November 20, 2019, from https://data2.unhcr.org/en/situations/drc

United States Department of State. (2011). *Bureau of Democracy, Human Rights and Labor Country Reports on Human Rights Practices for 2011: DRC*. Washington: United States Department of State. Bureau of Democracy, Human Rights and Labor.

Wedley, W. C. (1993). Consistency prediction for incomplete AHP matrices. *Mathematical and Computer Modelling*, *17*(4), 151–161. https://doi.org/https://doi.org/10.1016/0895-7177(93)90183-Y

World Bank. (2018). World Bank - Participatory Rural Appraisal (PRA) Techniques. Retrieved May 24, 2019, from http://go.worldbank.org/EU99NNIVS0

6 Participatory Analytic Hierarchy Process for agricultural development resource allocation in development projects[24]

6.1 Introduction

In the 2030 Agenda for sustainable development, agriculture plays a central role (Food and Agriculture Organization (FAO), 2017). The transition towards a more sustainable agricultural sector is a transversal challenge worldwide and implies the shift towards renewed approaches to the planning and evaluation of policies and specific interventions (International Panel of Experts on Sustainable Food Systems (IPES-Food), 2016; 2017). The diffusion of a new culture of evaluation in development agendas is believed to be a part of the leverage tool for this much-needed shift (Barbier et al., 2012).

In this perspective, agricultural development is an important area because of its overall societal impact and fundamental role in reducing poverty and boosting economic growth. Nearly 90% of the population in many developing countries depend on agriculture and farming for their livelihood, as agricultural production provides income, employment, and food, as well as raw materials for the processing industry and exports (European Union (EU), 2019). Consequently, the problem of choosing the best path towards sustainable development in the agricultural sector is currently central to development cooperation (DC). Agri-environmental and landscape-scale issues are the typical targets of participative and multicriteria decision-making procedures as they involve community resource planning and multidisciplinary knowledge (Dunnett et al., 2018). Consistent with what ensues in the field of impact assessment (Cornwall et al., 2017), some authors report that in order to successfully solve the complex real-world problems, multi-criteria group decision-making approaches and multidisciplinary analysis are recognized as reliable and effective (Groselj, 2018). Such tools are required to provide more evidence-based and shared choices, possibly linking interventions from the basic technical level to the top policy level. Shifting from local scale perceptions to a global perspective, stakeholders' interaction may give rise to misunderstanding and conflict over strategic decisions. While several international development organizations emphasize the need to conceive

---

[24] This chapter contains the pre-print version of the paper by De Marinis P., Sali G., 2009. Participatory Analytic Hierarchy Process for agricultural development resource allocation in development projects, *Evaluation and Programme Planning*, in press.

and implement multi-stakeholder development governance systems and spread the practice of evaluation (FAO-CFS, 2015; United Nations Development Programme (UNDP), 2012; Mayne, 2008), we focus on effective methodologies for facilitating the task.

According to several authors, the requested methodologies share features such as the importance of participatory, multi-stakeholder, and multidisciplinary analysis approaches (IPES-Food, 2017).

Participation is a broad concept with multiple definitions. However, a commonality to all the definitions is the role of the community in decision-making (Claridge, 2004). For the scope of this study we adopt the definition proposed by Lane (1995 pp. 181-191): '*Participation means meaningful implication of individuals and groups at all stages of the development process including that of initiating action*.'

A participatory, multi-stakeholder approach approximates the diversity of interests and positions in the target system, attempting not only to spread democratic principles but also to increase the practical likelihood that the proposed actions and plans will be accepted and effectively implemented. In this perspective, and to acknowledge and understand the diversity of values among the different stakeholders, several authors agree that truly participatory and multi-stakeholder decision support systems must include a multidisciplinary assessment of alternatives and choice criteria (UNDP, 2012).

Therefore, interest in participatory and multi-criterial decision-making methodologies (MCDM) in agriculture has been growing rapidly, primarily because these methods are proficient in informing community decision-making (Pazek et al., 2005; Tiwari et al., 1999). Literature shows that the analytic hierarchy process (AHP) (Saaty, 1980), among other techniques and tools, can be used as a 'bottom-up' participatory method, working as a consensus-building tool to enable stakeholders to discuss single criteria and weights to understand each other's viewpoint in iterative decisions processes. Notably, the analytic hierarchy process is recognized as one of the leading techniques among the family of MCDM or multi-criteria decision aiding (MCDA) methodologies (Emrouznejad et al., 2017; Ho et al., 2018;). One of the main benefits of AHP, especially if implemented as participatory AHP (PAHP), is that it optimizes project multi-ownership which, in turn, leads to better performance from project managers, management, and participants in general (Israel et al., 2002).

The analytical hierarchy process allows to consider both objective and subjective factors in the decision-making process (Aznar et al., 2011; Russo et al., 2015). Thus, AHP may be relevant in operational fields where community planning and participatory decision-making processes are important for achieving the desired results of development initiatives. The reason for this relevance was assumed a long time ago with Shumway (1981, p. 171) declaring, '... *any ex-ante evaluation procedures are inherently subjective. The only difference is where subjectivity enters and how it is processed. Consequently, the legitimate role of subjectivity in ex-ante evaluation needs to be recognized clearly and respected.'* More recently Palcic et al. (2009) suggested that AHP can enhance methods for evaluating project proposals; however, the authors refer to the business sector and we could not find any literature specifically relevant to non-business-oriented initiatives.

In the field of DC, participatory approaches to assessment, research, management, and budgeting have been widely studied and applied over the last few decades (FAO, 2006a; World Bank, 2018). In fact, DC initiatives, in all their relevant phases from planning through implementation, monitoring, and evaluation, face the problem of identifying initiatives that can be successfully and sustainably implemented.

The present study investigates the application of PAHP as a tool for choice criteria elicitation and resource allocation in the framework of operational planning for agricultural development projects in the Dioceses of Goma (or Goma Diocese or GD), Nord Kivu, Democratic Republic of Congo (DRC). We attempt to present AHP as a participatory and multidisciplinary technique. In fact, the practical objective of using AHP in the specific framework of the case study is to elicit a group discussion on priority criteria and to have useful new elements to determine and argue a shared resource allocation pattern. The main research questions of this study are: (1) What are the most important criteria behind the prioritization of interventions aimed at achieving sustainable agricultural development in the Goma Diocese (GD) Nord Kivu, DRC? (2) How should financial resources be allocated on identified priority interventions? (3) What is (if any) the added value of using PAHP?

The rest of the study is structured as follows. Section 2 outlines the research workflow, it describes the case setting of the study and reports about the present and past PAHP/AHP application in agricultural development and DC. Section 3 presents the results and discussion. Section 4 outlines possible conclusions and suggests future research perspectives on the subject.

6.2   Materials and methods

6.2.1   Research workflow

As shown in Figure 6.1 and detailed in Figure 6.2, research workflow was implemented over two years and comprised a preparation phase, participatory work phase (PAHP phase), and desk work phase (AHP phase). Being in the framework of the 'Appui au retour de réfugiés et déplacés par le biais de la sécurisation de terres en Diocèse de Goma' (ARDST) project (see Section 2.4 for case study description), we were able to build the decision path by starting from in-depth data collection aimed at evaluating the agricultural potential of the study area. This was achieved through three field surveys, elicitation of experts through semi-structured interviews,[25] and land use analysis through remote sensing (De Marinis et al., 2019a). The whole decision process was animated through a mixed-methods approach and with the declared goal of widening participation and building a transformative environment where decision makers (DMs) could 'learn by doing'.

Figure 6.1 - Study Workflow

---
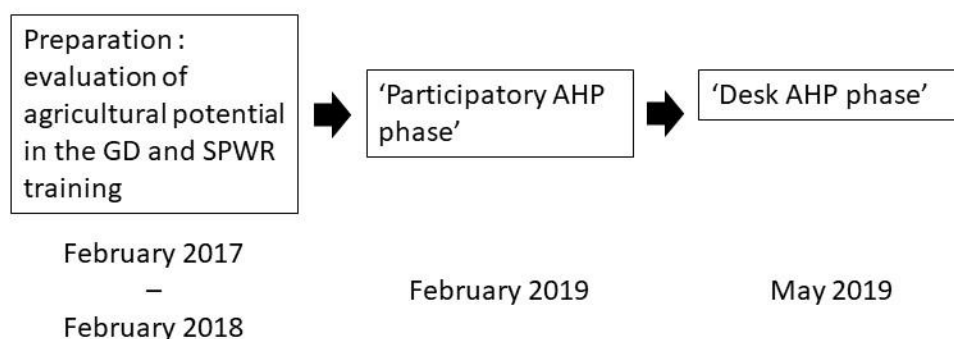
[25] Eight field supervisors managed the work of 24 surveyors in the GD: 'Shirika' survey produced 1005 observations for 131 variables concerning general local agricultural features and land tenure. A 'Concessions' survey produced 2554 observations of 25 variables concerning specific farm issues, 'Agriculture/élevage' survey produced 33 observations for 334 variables concerning agricultural and animal breeding practices. Nearly 35 experts from the local agricultural sector were interviewed during two rounds of semi-structured interviews in the Goma Diocese (GD).

Figure 6.2 - Decision process structure

Figure 6.2 - Decision process structure
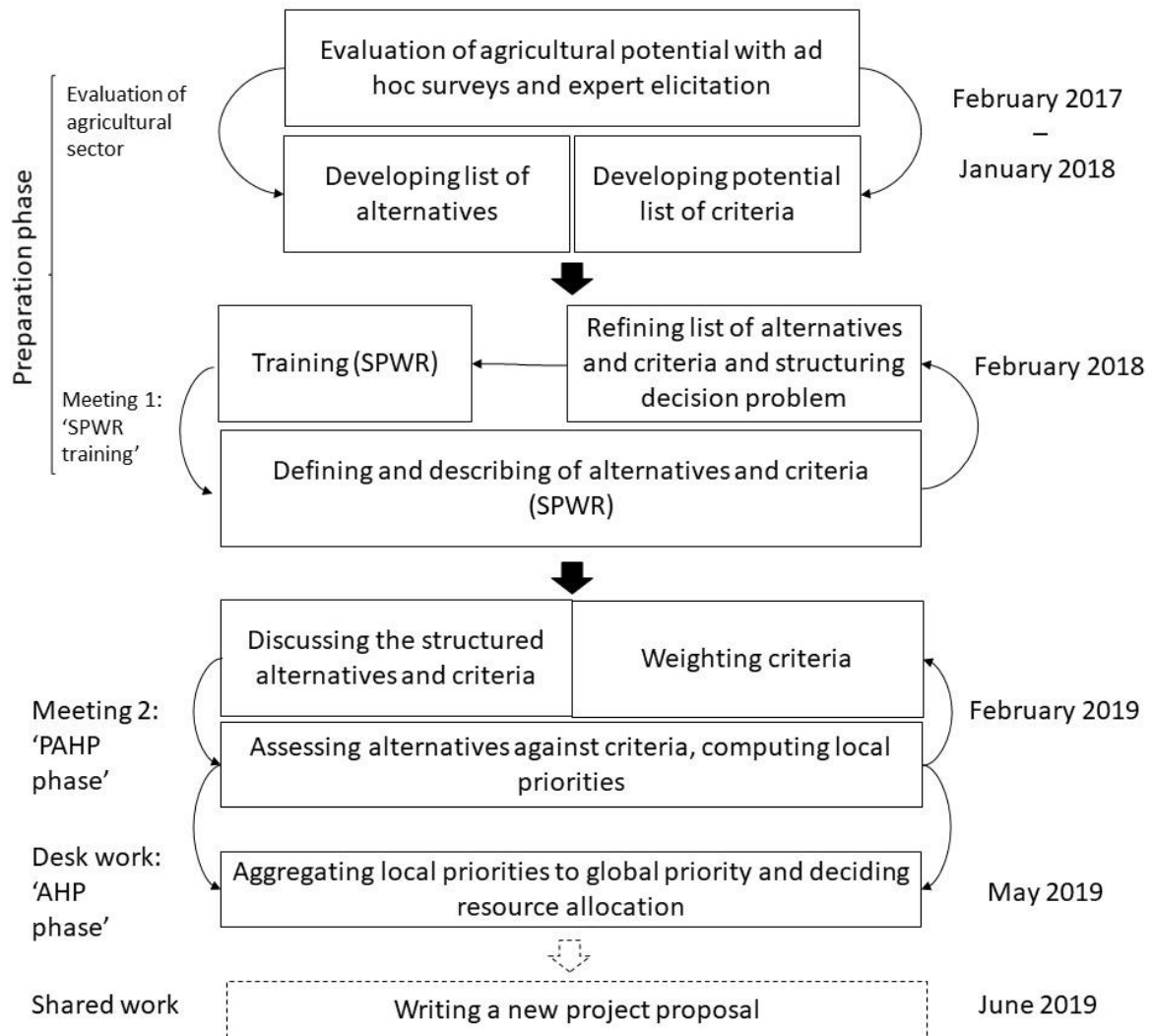
The preparation phase produced detailed information background on local agricultural system that resulted in a preliminary list of interventions expected to improve the sustainability of the sector. This list fed the following path. In order to appropriately approach the identification of triggering interventions, the same multidisciplinary team conducted the entire process applying a transformative participatory approach (Groselj, 2018). The participatory part of the path started during the preparatory phase, that is, during a mission in North Kivu (NK) (Goma, February 2018) and by organising a preliminary training session dedicated to the use of the Simplified Pairwise Ranking (SPWR) (FAO, 2006a; Narayanasamy, 2009; World Bank, 2018) focusing on priorities for intervention in a specific zone of the GD ('SPWR phase' in Figure 6.2). The working group comprised fourteen people with different backgrounds and expertise: four agronomists, four jurists, three experts in sustainable rural development, one sociologist (professor), one civil engineer, and one cartographer. The participants were selected according to the availability of experts already working in the target

zone in the framework of the ARDST project. They have different socioeconomic backgrounds and are not representative of stakeholders' diversity in the region; however, due to their practical experience in local agricultural development they can be considered as approximating all the different perspectives in agricultural development in the target region. The three main objectives of the SPWR session held in February 2018 are: (1) training the team in the use of MCDM methods in a simplified form (SPWR methodology), (2) defining and describing of interventions (alternatives) and criteria/sub-criteria, and (3) refining the preliminary list of interventions and criteria by limiting the scope of the following analysis to the five most relevant alternatives for agricultural development in the study area. This led to a preliminary decomposition of the problem followed by developing a preliminary decision hierarchy.

After the preparation phase, we first implemented AHP as PAHP for eliciting discussion on the relative importance of criteria and sub-criteria versus the global goal. During the third phase, we used AHP as a desk tool to allocate resources according to the weighting of alternatives against the global goal. The different phases are shown in Figure 6.2 and the relevant phases are described hereafter.

The PAHP phase allowed the working group to implement a participatory weighting of criteria and sub-criteria behind the target choice and to identify and valorise inconsistent judgements across the group. The AHP final phase allowed for aggregating individual refined priorities into a group consensus priority vector for resource allocation.

For implementing PAHP, the Priority Estimation Tool (PriEsT; Siraj et al., 2015) was used for calculating individual and local preferences through live calculation and refinement during a meeting held in February 2019 in Goma. The PriEsT is an open-license platform-independent Java-based tool that implements several prioritization methods and consistency measures.

The PAHP phase was aimed at weighting criteria and sub-criteria. Utilization of PAHP was intended to provide a solid methodological structure to the weighting procedure, namely to overcome the problem of potential bias often attributed to focus groups and other participatory techniques when anonymity is not granted along the process. One such type is the so-called 'dominance bias' (Crawford, 1997). This type of bias is produced by the existing differences in terms of communication ability, self-confidence, and authoritativeness among the participants (Sutton et al., 2013).

Among the several methods offered by the PriEsT app for local priorities calculation the geometric mean method (GMM) was selected (Dijkstra, 2013). The local preferences for each participant were recorded during the participatory work session, analysed, and the results were presented in front of the working team. The app helped to identify inconsistencies and adjusted weighting was produced in a two-turn iterative work session.

The desk AHP analysis was carried out using R (R Foundation for Statistical Computing, version 3.5.0) and Microsoft Excel (Microsoft Office 365) software with the aim of aggregating local priorities to a final and global 'aggregated group priority ranking'. The 'AHP SURVEY' package (CRAN Repository, version 0.4.0) was used to extract local priorities in a structured and automatic form. Microsoft Excel (Microsoft Office 356 version) was then used to aggregate priority vectors and matrixes for each DM and to produce a final aggregated group ranking of weighted alternatives to be used as a resource allocation guide. The GMM was used in computing both the local and global priorities (Dijkstra, 2013).

6.2.2   Research setting

6.2.2.1   Case study: Agricultural potential and triggering interventions in the Diocese of Goma, Nord Kivu, Democratic Republic of Congo

The study area lies in the eastern region of the DRC, bordering Uganda and Rwanda, in the Nord Kivu province and it is formalized as the Dioceses of Goma, covering approximately 25000 sq.km. The whole region is concerned with chronic insecurity due to the presence of several active armed groups (MONUSCO, 2018) in the framework of what is called the 'geologic scandal' of the DRC (International Monetary Fund, 2014).

The study area is characterized by one of the highest population density in Africa (2,211,000 inhabitants, population density of 88.4 inhabitants/km2) with the youth population (aged 15 years) accounting for 50%. The average family size is 5.5 individuals per nucleus. North Kivu has a primary school enrolment rate of 53.2%. From an ethnic point of view, the region of the two Kivus includes three cultural areas: the north-east is populated by ethnic groups of probable Nilotic origin (Alur, Tutsi) combined with ethnic groups of Bantu origin (Hutu), the great agricultural region of Kivu is inhabited by Nande, Shi, Havu and Hunde, Twa, and the forest region is occupied mainly by Bembe, Lega, and other smaller groups (Vansina, 1966).

Very few families have access to drinking water (16.6%) and electricity (4.3%), and 99.8% of the households do not benefit from waste disposal services. Compared with international WHO standards, infrastructure for delivering health services are few and poor: 12 hospital beds per 100,000 inhabitants and 1 doctor per 24,030 inhabitants.

Most of the roads outside the capital and the main cities of the region, with the exception of roads heading towards Kigali (Rwanda), are unpaved clay roads, often interrupted by the varied orography (with altitudes ranging from 1300 m to the tops of volcanoes that reach 3500–4000 m a.s.l.), combined with heavy rain seasons, and the poor maintenance by state authorities.

The economy of North Kivu is based on the primary sector (agriculture, livestock, fishing, forestry, mining), which represents about 49.7% of the provincial GDP and employs about 80% of the active population (UNDP, 2009).

According to the FAO (2001), the study area is part of an agroecological zone characterized by a long growing season of nearly 210–365 days per year. The local agricultural system can be framed in the agroecological zones: moving westward from the higher altitudes that lie along the borders with neighbouring Uganda, Rwanda, and Burundi and heading towards the forest, the farming system called 'Highland Perennial' gives way to a 'Forest Based' system. Two production systems exist in this setting: (1) small farmers, practicing subsistence family agriculture on small plots, and (2) large landowners who mainly grow cash crop plantations in monocultural cropping systems for business, or breed cattle in extensive systems.

It follows that the area is characterized by a high poverty rate, and 61 million people are currently living below the US$1.90 per day international poverty line. Overall, this amounts to 60.5% of the entire population in 2018 (World Poverty Clock, 2019). The operational context for this study is provided by an EU-funded project called ARDST 'Appui au retour de réfugiés et déplacés par le biais de la sécurisation de terres en Diocèse de Goma', led by Caritas Development Goma NGO. The ARDST project is framed in the European strategy for peace seeking and keeping operations in the African Great Lakes region and contributes to the implementation of the Protocol on the Property Rights of Returnees signed at the International Conference of Great Lakes Region (ICGLR) on November 30, 2006. Moreover, the ARDST project follows the guidelines of the European Development Fund National Strategy for the DRC, focusing on the increased diffusion of the state-of-law and augmented sustainability of the agricultural sector (EC, 2013). Among the specific objectives of the ARDST project,

our research group was asked to contribute to the strategic evaluation of the agricultural sector aimed at identifying interventions that trigger agriculture and at allocating financial resources within the framework of a new project proposal on the subject. Taking into consideration the need to innovate planning and evaluation methodologies in DC, as expressed by the Congolese government during the official EU-DRC negotiations (EC, 2013) and confirmed by the more general framework of international DC providers (IPES-Food, 2016; Barbier et al., 2012), we proposed a mixed method evaluation approach and the AHP technique as it has the ability to build more robust decisions and foster participation among the DMs.

6.2.2.2   The decision structure[26]

The study's target decision is about identifying the most important interventions to be implemented in the agricultural sector. These constitute the alternatives for the decision. As shown in Figure 6.3, the decision tree comprises four levels and seventeen factors that are described in Table 6.1. The global goal is to improve the sustainability of the agricultural sector (DSA) in the GD. In order to better describe the possible contributions to DSA, the concept of sustainability has been divided into three conventional components (Barbier, 1987), which constitute the sub-goals or criteria at level 2. Their comparison against the global goal produces the first weighting level in the decision tree (weights of DEC, DEN, DES). Descending towards the alternatives and to become practical in terms of effects that would impact the mentioned dimensions of sustainability, we were able to define eight sub-criteria (or desirable effects) whose comparison against the criteria produce the second weighting level in the decision tree (weights of RDI, DP, APC, RUAC, LE, AFS, ACP, ARP [27]). Finally, five alternatives for backward intervention constitute the causes of the identified desirable effects towards the

---

[26] In this section, we use acronyms for the goal, criteria, sub-criteria, and alternatives: DEC - Durabilité économique or economic sustainability; DEN - Durabilité environnementale or environmental sustainability; DES - Durabilité sociale or social sustainability; RDI - Réduire la dépendance de l'extérieur or reducing external dependencies; DP - Diversifier la production or diversifying production; APC - Augmenter le rendement des cultures or improving crops yield; RUAC - Réduire l'utilisation des agrochimiques or reducing the use of agrochemicals; LE - Lutter contre l'érosion des sols or preventing soil erosion; AFS - Augmenter la fertilité des sols or improving soil fertility; ACP - Augmenter la conservabilité des produits agricoles or improving the shelf life of agricultural products; ARP - Augmenter la rentabilité des produits agricoles or improving the rentability of agricultural products; : FTA - Formation et appui Technique Agricole or technical agricultural training ; ISA - Introduction des Semences Ameliorées (et connaissances) or introduction of improved seeds (and knowledge); TPA - Transformation des Produits Agricoles or processing of agricultural products; ISBP - Intervention de Structuration de la Base Productive or interventions in farmers' organization building ; AM - Accès au marché or access to market.

[27] Please refer to foot note 3 in previous pages.

achievement of the global goal (FTA, ISA, TPA, ISBP, AM[28]). All these levels of the decision structure were discussed in detail during the preparation phase and first meeting in order to bring the working group to share knowledge and interpretation of the elements of the decision tree.

---

[28] Please refer to foot note 3 in previous pages.

Figure 6.3 - Decision tree: The goal of the decision is to foster sustainability of the agricultural sector in the Goma Diocese (GD). Sustainability is divided into three acknowledged dimensions that constitute the main sub-goals or criteria for the decision. Eight desirable effects for the achievement of sub-goals are identified as the sub-criteria for the decision. Five alternatives for intervention constitute the alternatives in the target decision. All the elements appearing in the decision tree are described in Table 6.1.

Acronym	Complete name	Description
Global goal		
DSA	*Durabilité du Secteur Agricole* or Sustainability of the agricultural system sustainability	The FAO definition for sustainable agricultural development is 'the management and conservation of the natural resource base, and the orientation of technological and institutional change in such a manner as to ensure the attainment and continued satisfaction of human needs for present and future generations. Such development ... conserves land, water, plant and animal genetic resources, is environmentally non-degrading, technically appropriate, economically viable and socially acceptable'. (FAO, 1995)
Sub-goals or Criteria		
DEC	*Durabilité économique* or economic sustainability	Economic sustainability refers to practices that support long-term economic growth without negatively impacting environmental, social, and cultural aspects of the community.
DEN	*Durabilité environnementale* or environmental sustainability	Environmental sustainability is the management of our physical environment in a way that supports living within ecological limits, protects natural resources, and meets the needs of communities without compromising the ability of future generations to meet their own needs.
DES	*Durabilité sociale* or social sustainability.	Social sustainability encourages communities to promote social interaction and fosters community investment while respecting social diversity. It includes cultural sustainability, which is the idea that fostered practices honour traditional values, customs, spaces, and way of life.
Desirable effects or Sub-criteria		
RDI	*Réduire la dépendance de l'extérieur* or reducing external dependencies	Reducing external dependencies is an important goal because people feel frustrated by the incapacity of the local and national governance to valorise the huge potential of the area in terms of available resources and possibility to foster local well-being. The potential to gain self-sufficiency is experienced on a large scale as an urgent need when confronting the prospect of foreign markets rapidly entering the region. On a smaller scale, self-sufficiency is perceived by people as the possibility to profit from local resources that are perceived as abundant but inaccessible.
DP	*Diversifier la production* or diversifying agricultural production	Diversifying agricultural production is perceived as a chance to increase resiliency due to climate change, unpredictable agricultural seasons, unstable market prices, and unstable socio-political local conditions.
APC	*Augmenter le rendement des cultures* or improving crops yield	The increase in crop yield is perceived as one of the key desirable effect in agriculture.
RUAC	*Réduire l'utilisation des agrochimiques* or reducing agrochemicals use	The reduction in agrochemicals use is perceived as a means of decreasing the environmental and health impact of agriculture. Moreover, it is considered to be the best approach to decrease

Acronym	Complete name	Description
		production expenses (if other means are provided to increase the yield).
LE	*Lutter contre l'érosion des sols* or preventing soil erosion	Preventing soil erosion is perceived as a priority by the population living in such sloping regions. Considering that agriculture is traditionally implemented on hard slopes and that rainy seasons are quite aggressive in the entire region, the phenomenon of massive soil erosion is observed and is rapidly reducing agricultural potential.
AFS	*Augmenter la fertilité des sols* or improving soil fertility	Improving soil fertility is perceived as a prerequisite for expanding yields and reducing the need for fertilisers. In fact, soil fertility is the key asset for agricultural production.
ACP	*Augmenter la conservabilité des produits agricoles* or improving the shelf life of agricultural products	Agricultural products are prone to seasonal price fluctuations due to the availability of products during a specific period in the year and proximity to the agricultural market in the region. The possibility to increase the shelf life of products is perceived as a huge potential for reducing production waste and increasing and stabilising incomes.
ARP	*Augmenter la rentabilité des produits agricoles* or Improving rentability of agricultural products	Rentability of products in such a closed market context is the key triggering factor for improving well-being. Improved rentability is perceived as the sum of many factors such as the opportunity to sell during more profitable periods of the year, to improve and differentiate product quality in order to satisfy the needs of the elite and more profitable markets in the region.
Interventions or Alternatives		
FTA	*Formation et appui Technique Agricole* or agricultural technical training	This is a training and technical demonstration intervention including capacity building for producers as well as local technicians and training on new techniques. The training is always accompanied by a basic supply of inputs and implementation of one or more centres for technical assistance and sale of inputs and tools.
ISA	*Introduction des Semences Ameliorées* (et connaissances) or introduction of improved seeds (and knowledge)	It is a triple intervention: distribution of improved seeds belonging to improved local varieties (no unsuitable and infertile hybrids), training for farmers interested in the conservation of agro-biodiversity and its valorisation by improving local varieties and launching a local seed company.
TPA	*Transformation des Produits Agricoles* or processing of agricultural products	This is a composite intervention aimed at improving the performance of the target value chains. The intervention was described as a demonstration, within ad hoc demonstrative farms, of the profitability of agriculture through the introduction of profitable crops in functional rotation, integration of livestock and agriculture in order to value the existing synergies, creation of small decentralised processing units, artisanal but improved, and the establishment of local committees for maintaining structures and technical means.
ISBP	*Intervention de Structuration de la Base Productive* or intervention on farmers' organization building	This is a training intervention focused on structuring agricultural and agri-food associations and cooperatives, including economic support to meet the costs of formalization, post-creation animation to promote an associative spirit and good administrative functioning, and support for access to micro-credit or for the creation of local safety networks.

167

Acronym	Complete name	Description
AM	*Accès au marché* or access to markets	This is an action that supports market studies focused on target value chains and for building, according to the results of the studies, new structures well located for marketing (covered markets with storage space and means for facilitating the flow of foodstuffs). The action is complemented by the establishment of local cells for the maintenance of structures and technical means.

6.2.2.3   Criteria and sub-criteria interconnections

During the PAHP phase, that is, during the drafting of the decision structure, an ad hoc focus group was organized to evaluate the relationship existing between the identified criteria and sub-criteria.

Directed graphs were used (Biggs et al., 1986) to assess and visualize the existence of directed interconnection between the criteria and between the sub-criteria. Obviously, at the criteria level, three factors were found to be completely interconnected as the same definition of sustainability applies to the interconnection between the three dimensions. Concerning sub-criteria, the focus group produced the directed graph shown in Figure 6.4.

Moreover, sub-criteria were classified according to a short list of features that were argued to be useful to frame the existing interconnection. NodeXL Basic Template 2014 for Microsoft Excel (version 1.0.1.381, Social Media Research Foundation) was used to create the directed graph and to calculate graph metrics such as the number of inbound, outbound, and total connections for each vertex. The results are presented in Table 6.2.

All the effects composing the set of sub-criteria were found to be related to human capital building as capacity building affects local stakeholders' ability to handle all possible improvements in the actual situation. In contrast, only RDI and RUAC were argued to be related to other policy effects: reduction of external dependencies was linked to the market and commercial policies, and the decrease in agrochemicals use was related to the environmental policy. In the context of relationships involved during the production and processing phases of agricultural value chains, the sub-criteria are divided between the targeted production and processing effects, with only one case related to both phases. The RDI was argued to depend on both types of intervention. In terms of interconnections, Figure 6.4 shows the connected sub-criteria. The only two sub-criteria having only inbound and no outbound connections are ACP and ARP, indicating that these effects do not influence others.

Table 6.2 - *Showing the classification of sub-criteria according to some useful features for describing factor interconnectivity. Acronyms: RDI - Réduire la dépendance de l'extérieur or reducing external dependencies; DP - Diversifier la production or diversifying production; APC - Augmenter le rendement des cultures or improving crop yields; RUAC - Réduire l'utilisation des agrochimiques or reducing the use of agrochemicals; LE - Lutter contre l'érosion des sols or preventing soil erosion; AFS - Augmenter la fertilité des sols or improving soil fertility; ACP - Augmenter la conservabilité des produits agricoles or improving shelf life of agricultural products ; ARP - Augmenter la rentabilité des produits agricoles or improving rentability of agricultural products.*

	Max/min	Human capital related	Other policies related	Production related	Processing related	N° of inbound connections	N° of outbound connections	N° of total connections
ACP	MAX	Y	N	N	Y	0	2	2
AFS	MAX	Y	N	Y	N	4	4	8
APC	MAX	Y	N	Y	N	01	0	1
ARP	MAX	Y	N	N	Y	2	0	2
DP	MAX	Y	N	Y	N	1	4	5
LE	MAX	Y	N	Y	N	4	2	6
RDI	MIN	Y	Y	Y	Y	5	3	8
RUAC	MIN	Y	Y	Y	N	3	5	8

Figure 6.4 - *Directed graph showing existing interconnection between sub-criteria. Acronyms: RDI - Réduire la dépendance de l'extérieur or reducing external dependencies; DP - Diversifier la production or diversifying production; APC - Augmenter le rendement des cultures or improving crop yields; RUAC - Réduire l'utilisation des agrochimiques or reducing the use of agrochemicals; LE - Lutter contre l'érosion des sols or preventing soil erosion; AFS - Augmenter la fertilité des sols or improving soil fertility; ACP - Augmenter la conservabilité des produits agricoles or improving shelf life of agricultural products; ARP - Augmenter la rentabilité des produits agricoles or improving rentability of agricultural products.*

Created with NodeXL (http://nodexl.codeplex.com)
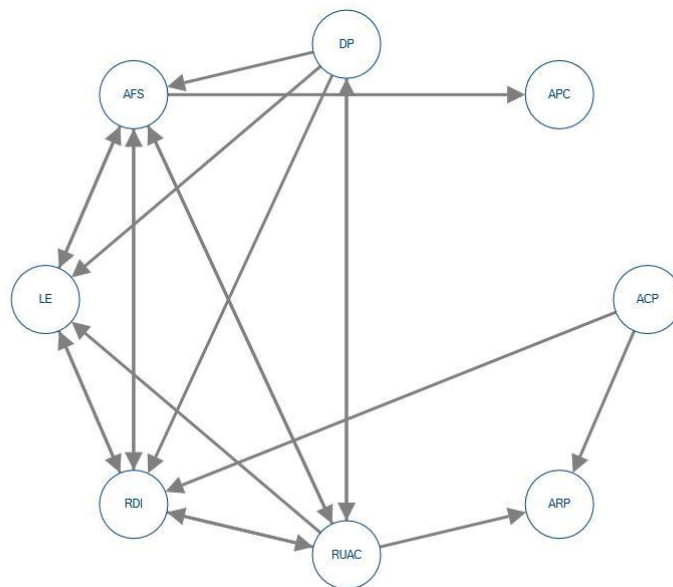
6.2.3 AHP: past and present applications

The AHP has been applied widely in several sectors in the last three decades (Golden et al., 1989). The first references are related to Zahedi's (1986) work showing organizations deploying AHP predominantly in the manufacturing process. Contemporary studies by Apostolou et al. (1993) investigates AHP within the accounting function and shows increasing usage overtime. The healthcare industry is another target for AHP applications, and Liberatore et al. (2008) reviewed over 50 studies relating AHP to decision-making in the medical and healthcare profession.

Despite these examples showing how AHP related literature is often specifically oriented towards some sector, existing reviews (Sipahi et al., 2010; Ishizaka, 2011) help to summarize the main areas of application of AHP: the authors show that AHP is widely applied in industries such as manufacturing, agriculture, environment, power and energy, transportation, construction, and healthcare. Moreover, AHP has also been applied diffusely to education, logistics, e-business, information technology (IT), research and development (R&D), telecommunications, finance and banking, urban management, defence and military, politics and government, marketing, tourism and leisure, sports, archaeology, auditing, and mining.

In all these fields and sectors, AHP has been applied to help solve a host of problems (Golden et al., 1989): (1) project selection and (2) location selection (Cheng et al., 2005), (3) resource allocation (Ramanathan et al., 1995; Braunschweig et al., 2004), (4) risk management (Mustafa et al., 1991), (5) technology selection, (6) conflict management (Lam et al., 2005), (7) project evaluation, and (8) benchmarking (Dey, 2002).

Despite the existing high pressure on international organizations to optimize available resources and account for the selection of intervention sectors in several countries, we did not find evidence of AHP being applied to resource allocation in the field of DC.

In fact, although AHP has been applied to real-world cases, DC organizations tend to neglect the use of more complex MCDM methods such as AHP (De Marinis et al., 2019) and prefer more operational tools such as SPWR when planning and implementing development projects (FAO, 2006a; World Bank, 2018).

The AHP methodology helps DMs to apportion a decision into a hierarchical structure (Saaty, 1987; 2008). This decision structure comprises L {L1, L2, … Lx} levels that normally correspond to at least

one global goal at L1 level, two or more criteria at L2 level, and two or more alternatives on the last Lx level. More than three levels can exist. The decision hierarchy or structure for our case study is presented in Figure 6.3, and it can be used as an example. Each level includes F {F1.1, F2.1, …Fx.y} decision factors. In our study (Figure 6.3), the decision has been structured into four levels containing: one overall goal on L1 (decision factor F1.1), three criteria on L2 (factors F2.1 to F2.3), eight sub-criteria on L3 (factors F3.1 to F3.8), and five alternatives on L4 (factors F4.1 to F4.5). The AHP method uses pairwise comparison matrixes (PCMs) to compare the relative importance of factors on one level against each criterion on the upper level of the decision structure. In case of multiple DMs, each DM must complete a pairwise comparison matrix for each level and for each factor at the mother-level (upper level in the decision structure). A PCM is a two-dimensional array where decision factors are compared pairwise by attributing a score on a 1–9 scale (or reciprocal values 1/9−1) by asking the question: 'how many times F2.1 is more important than F2.2 in front of F1.1?' Priority vectors are then calculated for each PCM (local priority vectors) and aggregated into a priority matrix for each level. In the case of our study, we concluded with eight priority vectors at level 4, three priority vectors at level 3, and one priority vector at level 2. Single priority vectors are aggregated into matrixes at each level and these matrixes are then aggregated into a final priority vector that contains the relative weight of alternatives against the global goal. When the AHP is applied in group sessions, individual final priority vectors can be aggregated into a group priority matrix, and a group or consensus priority vector can be computed.

The AHP can be implemented in four simple consecutive steps:

1) Computing the vector of criteria weights
2) Computing the matrix of sub-criteria weights
3) Computing the matrix of option scores
4) Ranking the options

Each step will be described in detail hereafter. It is assumed that m evaluation criteria and s sub-criteria are considered, and n options are to be evaluated.

In order to compute the weights for different criteria (F2.1 to F2.3 in Figure 6.3) against the overall goal, the AHP creates a pairwise comparison matrix A. Matrix A is a m×m real matrix (see Formula 2), where m is the number of evaluation criteria considered. Each entry a_{jk} of matrix A represents the importance of the jth criterion relative to the kth criterion. If $a_{jk} > 1$, then the jth criterion is more

important than the kth criterion; similarly, if $a_{jk} < 1$, then the jth criterion is less important than the kth criterion. If two criteria have the same importance, then the entry a_{jk} is 1. The relative importance between the two criteria is measured according to a numerical scale from 1 to 9. The values in matrix A are pairwise consistent by construction because the entries a_{jk} and a_{kj} satisfy the following constraint:

(1) $a_{jk} * a_{kj} = 1$

Obviously, $a_{jj} = 1$ for all j.

(2) $A(m * m) = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{j1} & \cdots & a_{jk} \end{bmatrix}$

Once matrix **A** is built, it is possible to derive from **A** the *normalized pairwise comparison matrix* **A**norm (3) by making the sum of entries on each column equal to 1;

(3) $A_{norm} = \begin{bmatrix} \bar{a}_{11} & \cdots & \bar{a}_{1k} \\ \vdots & \ddots & \vdots \\ \bar{a}_{j1} & \cdots & \bar{a}_{jk} \end{bmatrix}$

Each entry $\bar{a}_{jk}$ of matrix **A**norm is computed as

(4) $\bar{a}_{jk} = a_{jk} \Big/ \sum_{j=1}^{m} a_{jk}$

Finally, the criteria weight vector W_c (an m-dimensional column vector) is built by computing the geometric mean of the entries on each row of A_{norm}, that is,

$$W_c = \begin{vmatrix} w_1 \\ \vdots \\ w_j \end{vmatrix}$$

W_c contains the weights of m criterion against the global goal and its entries are computed using the GMM with Formula (5).

(5) $w_j = \sqrt[m]{\bar{a}_{j1} * \bar{a}_{j2} * \ldots * \bar{a}_{jk}}$

In order to compute the weights for different sub-criteria (F3.1 to F3.8 in Figure 6.3), the AHP starts creating PCMs B1, B2, and B3 in order to compare the eight sub-criteria against each one of the criteria at the above level. Each matrix B is a *s×s* real matrix, where *s* is the number of evaluation sub-

criteria considered. Formulas (1) to (5) are used again, as shown for weight computation for criteria, by replacing matrix A with matrixes B1, B2, and B3. In this case the aim is to obtain the three sub-criteria weight vectors w_1, w_2, and w_3, which are s-dimensional column vectors containing the relative weight of sub-criteria against each of the three criteria. These column vectors are built by averaging the entries on each row of $B1_{norm}$, $B2_{norm}$, and $B3_{norm}$. Finally, the three sub-criteria weight vectors are reunited in a *s×m* sub-criteria weight matrix W_s.

W_s is obtained as

(6) $W_s = |w_1 \quad w_2 \quad w_3|$

In order to compute the matrix containing the scores of options (from F4.1 to F4.5 in Figure 6.3), a PCM Ox is first built for each of the *s* sub-criteria. This means that we first build PCMs O1 to O8. Each matrix Ox is a n×n real matrix, where n is the number of options evaluated.

The AHP then applies to each matrix Ox, which is the same two-step procedure described for PCMs A, B1, B2, and B3 and Formulas (1) to (5), that is, it divides each entry by the sum of entries in the same column and averages the entries on each row, thus obtaining the score vectors o_x, where *x=1,…,s*. Each of the eight vectors contains the scores of the evaluated options with respect to the sth sub-criterion.

Finally, the score matrix $\mathbf{O_w}$ is obtained. It is a *n×s* real matrix where each entry o_{jk} represents the score of the nth option (F4.1 to F4.5 in Figure 6.3) with respect to the sth sub-criterion.

(7) $O_w = [o_1 \dots o_8]$

namely, the sth column of $\mathbf{O_w}$ corresponds to o_x vector.

Once the weight vector w_c, sub-criteria weight matrix W_s, and score matrix O_w have been computed, the AHP obtains a vector **v** of global scores by multiplying **O, Ws,** and **w_c,** that is,

(8) $v = (O_w * W_s) * w_c$

Where v is a n-dimensional column vector containing the aggregated (or global) weights of the n options against the global goal. In other words, the nth entry v_n of **v** represents the global score assigned by the AHP to the nth option.

If multiple DMs exist, the global priority vectors of alternatives against the global goal are aggregated using GMM in order to build the final group global priority vector.

As the final step, options ranking is accomplished by ordering the global scores in decreasing order, thus building the final ranking of the options.

Given that the choice of the priority vectors computation method and the aggregation mode can change the results of the analysis, we prefer to clearly state that our case does not concern a unique choice among exclusive alternatives, rather it refers to a distributive problem such as resource allocation. According to Saaty et al. (1993), the distributive mode of aggregation is recommended in distributive decision problems, and it implies normalization of weights as specified by the Formula (4).

Moreover, GMM (Formula 5) was picked among the possible methods for computing local priorities because it is insensitive to low consistency rates in PCMs (Dijkstra, 2013). For aggregating individual local priorities into group global priority, the GMM was held as a method in order to avoid the rank reversal in global ranking (Stoklasa et al., 2018).

In fact, we emphasize here the need for a computing method that grants both mathematical robustness related to reducing rank reversals and the possibility of accepting high local inconsistency degrees, which are typical of complex real contexts.

Given the participatory framework that characterizes the present work and the operational aim of the decision process, which constitutes the case study for our work, we treated inconsistency in two separate ways: during the PAHP phase, inconsistency of local priorities was only used in terms of highlighting the most inconsistent comparisons and to guide discussions on them. For the specific purpose of our work, the first phase did not have 'too inconsistent PCMs' and, therefore, there was no need for evaluating the PCMs' CR against a threshold. Conversely, during the desk-AHP phase, inconsistency analysis was run and used to extract only consistent decision makers whose priorities were used to allocate resources.

Concerning the inconsistency threshold adopted, it is widely documented in literature that the consistency ratio (CR) threshold for judging a PCM as consistent depends mainly on the matrix size, with higher thresholds when the number of alternatives in PCMs becomes higher (Wedley, 1993). In addition, it depends on the sample characteristics and the analysis (group and/or individual) type. For individual experts, the acceptable CRs are restricted to 0.10 or 0.15, while the CR consistency threshold for groups could be relaxed to 0.20 or 0.25 to allow for non-expert contributions (Ho, 2005). Building on these assumptions, we decided to use a threshold equal to 0.25 and to accept and

use scarcely consistent PCMs, as they reflect the perceptions of experts in a complex context, whose opinion, even if scarcely consistent, is the only tool available for informing decisions.

The CRs of all the PCMs in the decision structure have been computed according to Saaty (1980). The author has proved that for a consistent reciprocal matrix, a consistency ratio (CR) can be calculated by comparing the consistency index (CI) of each PCM with an appropriate random consistency index (RI). While RIs are tabled values that vary according to the amount of comparison elements (Saaty, 1980), CI can be calculated on the basis of a consistency measure λ related to each row of a PCM. λ is obtained for each row as the matrix product between the priority vector of the PCM and each row of the PCM, divided by the weight (or priority value) of the same row. Therefore, CI, which is a global measure of consistency for each PCM, is computed for each PCM as a deviation from mean λ using the following formula

$$(9)\ CI = \frac{\lambda_{mean} - n}{n - 1}$$

where λ_{mean} is computed by averaging the λ values of the PCM rows, and n is equal to the number of elements compared in the PCM.

CRs are computed by Formula (10) $CR = \frac{CI}{RI}$.

6.3 Results and discussion

6.3.1 PAHP for individual local priorities: criteria, sub-criteria ranking, and consensus building [29]

During the PAHP session, criteria and sub-criteria were analysed by pairwise comparison and AHP analysis was implemented using PriEsT software.

The PriEsT app calculates and shows local priority vectors and CRs for each PCM in the decision structure. Moreover, the software graphically shows the three most inconsistent comparisons for each PCM as a tool for eliciting discussions and for adjusting local preferences. This was done during two rounds of in-depth plenary discussion about local priorities and two individual work sessions. During each round, the participants modified their preferences according to the most inconsistent comparisons hinted at by the software, and the adjusted preferences were recorded. Local priority

---

[29] In this section, we use acronyms for goals, criteria, sub-criteria, and alternatives. Please refer to the captions of Tables 6.1, 6.2, 6.3, 6.4, 6.5, and 6.6 for explanation of acronyms.

vectors were, therefore, computed and discussed again during each round, bringing improvements for each participant's consistency and consensus about the relative importance of the selected criteria/sub-criteria. It is important to highlight here that the PAHP phase focused on the identification of the most important criteria/sub-criteria with a participative approach. This indicates that the AHP methodology was mostly used in this phase as a discussion-eliciting tool, and that the consistency improvement process was judged satisfactorily when the consistency ratio (CR) improved by approximately 0.2, despite the final CR local value that remained above the 0.24 threshold in many cases.

As mentioned earlier, we were able to identify and discuss most frequent inconsistencies in order to improve global performance. Examining the local priority vectors separately for the two levels of criteria and sub-criteria, we were able to identify where most frequent inconsistencies are generated. For comparing the criteria against the goal, inconsistencies were homogeneously distributed among the DMs and the criteria comparisons showing light prevalence of DEC_DES and DEN_DES inconsistent judgements. Examining the frequency of single criteria in inconsistent comparisons, we find that DES is the most frequent criteria in inconsistent choices.

In comparing the sub-criteria against the three criteria, the top-three inconsistent comparisons were RDI_APC, APC_RUAC, and RUAC_ARP, and even after the adjustment rounds, the APC, RUAC, RDI, and ARP sub-criteria remained the most frequent in inconsistent judgements.

The PAHP phase, after the two rounds of discussion on consistency improvement mentioned above, yielded individual aggregate priority vectors for the eight sub-criteria against the global goal. The results are shown in Table 6.3. Local priority vectors for criteria show high diversity among the DMs' opinion. In Figure 6.5 and Table 6.4, the local priority vectors concerning first level comparison for the eleven DMs can be visualized. Only two DMs attribute equal weight to the three criteria. Notably, seven DMs among the eleven attributes greater importance to environmental sustainability, while only two DMs prioritize economic sustainability and one DM attributes equal weight to environmental and economic sustainability. In the opinion of six out of eleven DMs, social sustainability is the least important and never gets the first rank. Following Wedley (1993) thresholds, CRs witness acceptable consistency levels for only six DMs out of eleven (only six DMs show CR < 0.24).

Table 6.3 - Aggregated local priority vectors of sub-criteria against the global goal for eleven DMs. Acronyms: RDI - Réduire la dépendance de l'extérieur or reducing external dependencies; DP - Diversifier la production or diversifying production; APC - Augmenter le rendement des cultures or improving crop yields; RUAC - Réduire l'utilisation des agrochimiques or reducing the use of agrochemicals; LE - Lutter contre l'érosion des sols or preventing soil erosion; AFS - Augmenter la fertilité des sols or improving soil fertility; ACP - Augmenter la conservabilité des produits agricoles or improving shelf life of agricultural products; ARP - Augmenter la rentabilité des produits agricoles or improving rentability of agricultural products.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
RDI	0.12	0.12	0.05	0.05	0.14	0.16	0.03	0.04	0.12	0.12	0.06
DP	0.14	0.09	0.09	0.17	0.30	0.31	0.08	0.18	0.12	0.14	0.06
APC	0.13	0.28	0.12	0.14	0.12	0.17	0.10	0.21	0.08	0.14	0.18
RUAC	0.13	0.16	0.06	0.07	0.12	0.14	0.16	0.07	0.10	0.20	0.05
LE	0.13	0.07	0.48	0.13	0.07	0.09	0.13	0.15	0.08	0.13	0.05
AFS	0.13	0.12	0.07	0.15	0.15	0.06	0.23	0.09	0.12	0.12	0.23
ACP	0.08	0.06	0.06	0.13	0.06	0.04	0.06	0.12	0.21	0.09	0.04
ARP	0.14	0.10	0.07	0.16	0.04	0.03	0.20	0.14	0.18	0.08	0.32
CR	0.18	0.42	0.68	0.08	0.51	0.39	0.16	0.33	0.14	0.21	0.30

Figure 6.5 - Graph showing the priority vectors for eleven DMs for level 1 (criteria against goal).
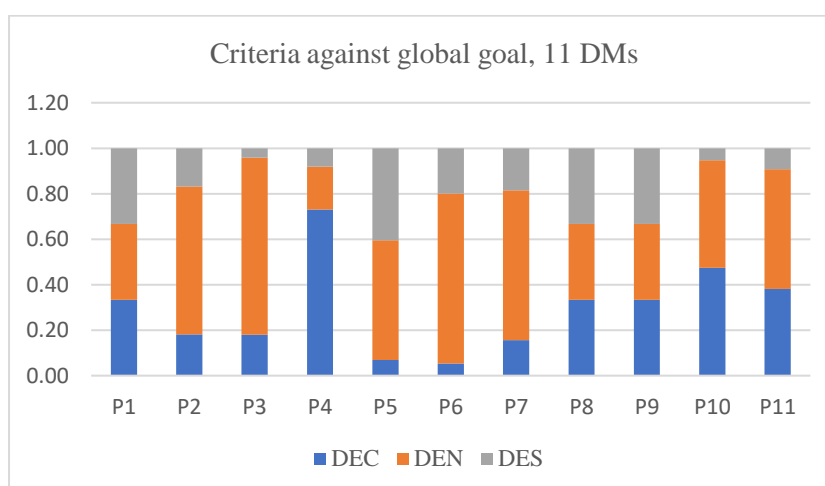

Table 6.4 - Local Priority Vectors for eleven DMs comparing criteria against the goal (level 1 comparison). Acronyms: DEC - Durabilité économique or economic sustainability; DEN - Durabilité environnementale or environmental sustainability; DES - Durabilité sociale or social sustainability.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
X DEC	0.33	0.18	0.18	0.73	0.07	0.05	0.16	0.44	0.33	0.47	0.38
X DEN	0.33	0.65	0.78	0.19	0.53	0.75	0.66	0.39	0.33	0.47	0.53
X DES	0.33	0.17	0.04	0.08	0.40	0.20	0.19	0.17	0.33	0.05	0.09
CR	*0.43*	*0.31*	*0.48*	*0.06*	*0.16*	*0.70*	*0.24*	*0.02*	*0.00*	*0.00*	*0.12*

Local priority vectors for sub-criteria show even higher diversity among the DMs reflecting different perspectives and perceptions regarding the relative importance of target desirable effects on the

agricultural sector. The CRs show high variability and according to Wedley (1993), only five out of eleven DMs produced consistent judgements despite their indisputable expertise in rural development (only five DMs show CR < 0.24).

Inconsistencies identified during the PAHP phase were useful to elicit discussions. In general, inconsistencies related to the comparison of L2 criteria against the L1 global goal may be attributed to the intrinsic interconnection between the three dimensions of sustainability. Specifically, L2 vs. L1 comparison showed that DES caused difficulties in individual preference expression. This may be attributed to the definition and operationalization of the concept of social sustainability with respect to the most acknowledged concepts of economic and environmental sustainability.

Inconsistencies related to the comparison of L3 sub-criteria against each of the three L2 criteria may be generally attributed to the complexity of operational interventions in complex contexts and the intrinsic interconnections that practical interventions on agricultural development have. Specifically, L3 vs. L2 comparisons showed ARP, APC, RDI, and RUAC being the main source of inconsistent judgement. Among these sub-criteria, which we remind being desirable effects of interventions, we find only two 'minimizing' effects and the two 'maximizing' effects with no outbound relations with other criteria, as shown in Figure 6.4. Inconsistent judgements about the two 'minimizing' effects could be attributed to the difficulty in comparing the impact of minimizing factors in a set of maximizing factors. In other words, it means that these inconsistencies may be related to some form of inertia in DMs' reasoning. Inconsistent judgements about APC and ARP could be because these factors were perceived as dominated by the others in terms of being completely dependent on others, and this in turn determines inconsistencies when evaluating the relative importance of these two sub-criteria in respect of other factors. It is important to underline that inconsistent DMs were fundamental for our participatory implementation of AHP because of their role in indicating the most difficult comparisons and, therefore, boosting the discussion meant to build consensus among the participants. For this reason, we find it useful to report individual local priority vectors with CRs above the threshold in Tables 6.5 and 6.6.

The group global priority vector for the five alternatives was computed using R and Excel software. R was used to reproduce AHP analysis for individual global priorities. The results were compared with those derived from the PAHP phase generated by the PriEsT app in order to triangulate the results and to have consolidated priority matrixes for each level of the decision hierarchy and for global priorities. The two procedures, following the same GGM aggregation methodology, produced

matching results in terms of priority vectors and CRs. The priority matrixes produced by R software for each level and DM were aggregated using GMM on an Excel spreadsheet. The results are presented in Table 6.5.

Table 6.5 – Eleven DMs' individual priority vectors and aggregated group priority vector for the alternatives against the goal. Acronyms: FTA - Formation et appui Technique Agricole or technical agricultural training; ISA - Introduction des Semences Ameliorées (et connaissances) or introduction of improved seeds (and knowledge); TPA - Transformation des Produits Agricoles or processing of agricultural products; ISBP - Intervention de Structuration de la Base Productive or intervention on farmers' organizations building; AM - Accès au marché or access to market.

Rank	Alternatives (interventions)	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	Group Global priority vector
1	FTA	0.42	0.22	0.42	0.40	0.32	0.38	0.38	0.39	0.33	0.23	0.31	0.35
2	ISA	0.12	0.18	0.25	0.17	0.16	0.22	0.25	0.29	0.36	0.23	0.26	0.23
3	ISBP	0.21	0.31	0.16	0.18	0.22	0.13	0.22	0.12	0.08	0.21	0.12	0.17
4	ITPA	0.14	0.17	0.10	0.14	0.16	0.14	0.09	0.15	0.15	0.14	0.10	0.14
5	AM	0.10	0.13	0.07	0.10	0.14	0.12	0.07	0.05	0.08	0.19	0.21	0.11
	Sum	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	CR	0.12	0.35	0.50	0.08	0.76	0.37	0.09	0.29	0.26	0.25	0.28	

On the one hand, Table 6.6 shows the individual global priority vectors for the whole team because, as already mentioned, inconsistent DMs were valuable for the scope of the PAHP application, that is, eliciting discussion and building consensus.

According to the eleven DMs group, the group's global priority vector shows the ranking of alternatives (and resource allocation on the interventions) as follows:

1. Technical training on cropping techniques (FTA – Intervention de formation technique agricole): 35%
2. Supply of improved seeds and training on improved seed production (ISA - Introduction de semences améliorées): 23%
3. Farmers' organization building (ISBP – Intervention de structuration de la base productive): 17%
4. Agricultural products processing (ITPA – Introduction de la transformation des produits agricole): 14%
5. Access to market (AM – Accès au marché): 11%

On the other hand, in order to build a resource allocation pattern on solid methodological basis, we ran the mentioned consistency test and retained only the consistent DMs, whose aggregated CR is below the 0.24 threshold, in order to compute the resource allocation pattern. Table 6.6 shows the results. Comparing left and right parts of Table 6.6, it is clear that rank reversal occurs only for the alternatives ranked 2nd and 3rd (ISA and ISBP), and the rank reversal effect in terms of changing

weighting for resource allocation is limited to +/- 5% for both factors. The final weights of alternatives determine the following resource allocation pattern:

1. Technical training on cropping techniques (FTA – Intervention de formation technique agricole): 40%
2. Farmers' organization building (ISBP – Intervention de structuration de la base productive): 20%
3. Improved seed supply and training on improved seed production (ISA - Introduction de semences améliorées): 18%
4. Processing of agricultural products (ITPA – Introduction de la transformation des produits agricole): 12%
5. Access to market (AM – Accès au marché): 10%

Table 6.6 – Three 'consistent' DMs' individual priority vectors and aggregated group priority vector for the alternatives against the global goal (resource allocation vectors). Acronyms: FTA - Formation et appui Technique Agricole or technical agricultural training; ISA - Introduction des Semences Ameliorées (et connaissances) or introducing improved seeds (and knowledge); TPA - Transformation des Produits Agricoles or processing of agricultural products; ISBP - Intervention de Structuration de la Base Productive or intervention on farmers' organization building ; AM - Accès au marché or access to market.

Ranking	Alternatives (interventions)	P1	P4	P7	Group Global priority vector
1.00	FTA	0.42	0.40	0.38	0.40
2.00	ISBP	0.21	0.18	0.22	0.20
3.00	ISA	0.12	0.17	0.25	0.18
4.00	ITPA	0.14	0.14	0.09	0.12
5.00	AM	0.10	0.10	0.07	0.10
Sum		1.00	1.00	1.00	1.00
CR		0.12	0.08	0.09	

6.4 Conclusion and research perspectives

The ARDST project constituted the framework for the implementation of a mixed methods evaluation aimed at decision-making. The whole evaluation and the decision path have been animated through participatory methodologies that informed an improved consensus and a shared decision about both the factors guiding the choice and suitable priority interventions in the agricultural sector of Goma Diocese (GD). During the complete decision path, the case study presented here is a perfect setting for a two-step AHP implementation, namely for experimenting PAHP and producing new insights both in terms of methodological implementation and operational results.

From a methodological point of view, our AHP implementation in group decision-making shows all the criticality found in the existing academic debate on robustness of this methodology (Krejčí et al., 2018; Kułakowski, 2018; Pan, et al., 2014; Russo et al., 2015; Tran, 2013). Nonetheless, AHP was

found to turn subjectivity and inconsistency into valuable assets and elicit discussion to improve consensus within the working team. The implementation of PAHP was useful both to structure experts' contributions and to improve DMs' ability to perform multi-criteria weighting and priority allocation. In other words, we were able to use inconsistent pairwise comparisons in order to foster participation and to train DMs to perform holistic judgements. A holistic perception is an important skill when dealing with the complexity of sustainable agricultural development because when incomparable things must be compared this is the only human ability that allows to support choices (Hsieh, 2007; Larrick, 2008). Considering the high level of data inconsistency produced by the application of AHP in our case study, we experienced how strict transitivity, often called linear transitivity, must be considered an ideal outcome of preference expression, exactly as proposed by the first theorist of AHP (Saaty, 1984). Several other authors (Wedley, 1993; Nurmi, 2014) state that in real-life situations it is misleading to associate complete and transitive preference relations with rationality and robustness of the technique, especially if it is assumed that AHP and other techniques, such as pairwise ranking, are meant to aid the decision and not to substitute the decision maker.

Nurmi (2014) reports that *'It is generally agreed that intransitivities can occur, especially when the number of items being compared under a multi-criteria framework gets to be greater than five. It is also generally agreed that, if intransitivities are found, they should be analysed and changed, if deemed appropriate. That is, there is no inherent rule that says a set of comparisons should not contain any intransitivities, but they should be made explicit'* (Gass, 1998 pp. 616-624).

In other words, DMs may be good reasons for having intransitive individual preference relations and incomplete preferences. These good reasons were pointed out by the application of PAHP and valorized in terms of team building and in-depth comprehension of different perspectives.

In fact, from an operational point of view, the PAHP methodology is suited for training the project team in MCDM methodologies, for eliciting discussions and for identifying a shared resource allocation pattern that matches the existing international guidelines for agricultural development in the region (International Fund for Agricultural Development (IFAD, 2015; FAO, 2017).

Moreover, participants in the decision process appreciated the use of PAHP methodology because it provides the opportunity to structure their own participation and self-improvement work. This has two-fold implications: first, assessing how the proposed methodology was able to overcome the possible bias recorded for the participatory group methodologies; second, spreading participation and the culture of evaluation.

First, asking the participants to work individually on PCMs, and the fact that PCMs drive participants to focus separately on the different levels of the decision structure, helps to reduce the impact of respondent-induced bias, exactly as these biases are managed and reduced at different degrees in the interactive, nominal, and Delphi groups (Sutton et al., 2013). In our experience, the values established by the participatory approach overcome the negative impact of the lack of anonymity during the process, both in terms of the resulting data quality and consensus building.

Second, in our experience, the participatory approach to evaluation and decision-making we used, denotes the choice of data collection and treatment methodologies in which plurality of disciplines and stakeholders were actively sought and equally considered. Working with a constructivist and transformative approach to evaluation, and focusing on decision-making, participation implies the physical contribution of a plurality of disciplines, and the use of a procedure that grants participants equal contribution potential and the opportunity to improve each participant's ability to better contribute in further decision-making processes.

Participation also implies uncovering biases that become a shared framework for the decision process.

According to Fals-Borda (1991), participation is a real and endogenous experience of and for the common people, as it reduces the differences between both experts and communities and mental and manual labour. O'Neill and Colebatch (1989) identified that participation is real when participants can determine their outcomes, and this becomes relevant in the field of DC and the discussion on evaluation.

Concerning the comparisons of L3 vs. L2 vs. L1 factors in the decision tree (Fig.3), the PAHP path was useful in improving team consensus and identifying the most important factors underpinning the choice of forecasted interventions. These factors (criteria and sub-criteria) are important to build a new solid project proposal that incorporates both the ambition of drafting a shared proposal and focusing on the most deeply felt desirable effects being sought. In fact, in the framework of a project proposal, the choice of proposed interventions is normally the result of an ex-ante evaluation. More precisely, the relative importance of the arguments underlying the choice is subjective and is rarely investigated with advanced MCDMs such as AHP. The reasons behind the choice are often stated by referring to international guidelines issued by developmental organizations such as the FAO, World Bank or others. Nonetheless, even if these references completely cover the statements, it is difficult

182

to deny the 1981 statement made by Shumway (1981, p. 171) declaring, ' *... any ex-ante evaluation procedures are inherently subjective. The only difference is where subjectivity enters and how it is processed. Consequently, the legitimate role of subjectivity in ex-ante evaluation needs to be recognized clearly and respected.'* Subjectivity has to be valorized during the decision-making process.

During our research, we were able to use the most frequent inconsistencies in PCMs in order to elicit the discussion and adjust local preferences. Thus, subjectivity and inconsistent subjective judgements were pointed out and approached as resources for improving both the entire decision process and the participants' ability to 'take action' in project cycles by becoming actors of an evaluation culture. Our experience, and namely the work carried out by comparing L3, L2, and L1 factors in the decision tree (Fig.3), shows that PAHP can foster the valorization of different opinions in group decision-making and, therefore, strengthens the existing idea that use of PAHP allows diverse subjective and objective factors to be considered in the decision process, as evidenced by other authors (Aznar et al., 2011; Russo et al., 2015). In other words, PAHP has been found to be useful in structuring and giving pace to the participation of experts, allowing for building synergies between research methodologies and practical decisions.

From the resource allocation perspective, our study confirms once more (Ramanathan et al., 1995; Cheng et al., 2005) that AHP can be used as a resource allocation tool and provides first insights about the possibility to adapt it in the field of DC. Despite the existing high pressure on international organizations to obtain the best out of their limited resources and to account for the selection of work axes in several countries, few authors investigate the effect of structuring participatory decision processes in the field of DC. If an evaluative culture must be spread along the results chain in DC, the use of participatory MCDM, such as the proposed PAHP, must become the standard in all the phases of the policy-to-project cycle. For this reason, even if some simplified forms of MCDM tools are currently appropriate (DFID, 2002; FAO, 2006; 2006a; IFAD, 2000; UNESCO, 2008; World Bank, 2018), our experience suggests that PAHP deserve to become a relevant operational tool in the field of DC in order to spread the frequently summoned evaluation culture.

6.4.1   Lessons learned

If the transition towards more sustainable agricultural development implies the shift towards renewed approaches to the planning of policies and specific interventions (IPES-Food, 2017), approaching development with the lens of socio-ecological systems framework (Ostrom, 2009) forces

us to implement new methodologies for community planning and decision aiding in a participatory situation. In such a context, where the degree of sharing and local acceptance of proposed interventions are important for achieving the expected results, PAHP should be used to inform more shared decisions about intervention.

On the one hand, the present experience perceives the need to continue operational research on MCDM tools that are appropriate for both higher policy and ground-level interventions. Such type of research should be kept operational in order to not lose contact with reality and to implement 'live' training to DC practitioners. On the other hand, the registered levels of inconsistency among the preferences expressed by the DMs show that the creation of a hierarchical structure of an operational decision in a complex context is indeed a subjective action that is absolutely prone to overlapping of composing factors which, in turn, renders complete consistency of comparison utopian. Training international development practitioners in using their holistic judgement is often neglected and this could profit from the use of methodologies such as PAHP. The high appreciation level participants expressed during the experience enables us to affirm that further research could profitably focus on the impact of using PAHP on the ability of a DM team to remain consistent when evaluating alternatives and criteria in complex contexts.

6.5 References

Apostolou, B., & Hassell, J. M. (1993). An empirical examination of the sensitivity of the analytic hierarchy process to departures from recommended consistency ratios. *Mathematical and Computer Modelling*, *17*(4), 163–170. https://doi.org/https://doi.org/10.1016/0895-7177(93)90184-Z

De Marinis et al. (2019). Simplified pairwise ranking for identification of triggering interventions on agricultural development in Masisi, Nord Kivu, Democratic Republic of Congo. *Evaluation and Program Planning*, *under revi*ew.

De Marinis et al. (2019a). Integrating remote sensing and census information for land securing in Nord Kivu, DRC. *Journal Of Universities and Internaional Development Cooperation*, *1*(1).

Aznar, J., Guijarro, F., & Moreno-Jiménez, J. M. (2011). Mixed valuation methods: A combined AHP-GP procedure for individual and group multicriteria agricultural valuation. *Annals of Operations Research*, *190*(1), 221–238. https://doi.org/10.1007/s10479-009-0527-2

Barbier, E. B. (1987). The Concept of Sustainable Economic Development. *Environmental Conservation*, *14*(2), 101–110. https://doi.org/10.1017/S0376892900011449

Barbier, J.-C., & Hawkins, P. (2012). *Evaluation Cultures, Sense-Making in Complex Times*. London: Routledge.

Biggs, L., Lloyd, E. K., & Wilson, R. J. (1986). *Graph theory* (2016/08/01). Oxford: Oxford University Press.

Braunschweig, T., & Becker, B. (2004). Choosing research priorities by using the analytic hierarchy process: an application to international agriculture. *R&D Management*, *34*(1), 77–86. https://doi.org/10.1111/j.1467-9310.2004.00324.x

Cheng, E. W. L., & Li, H. (2005). Analytic Network Process Applied to Project Selection. *Journal of Construction Engineering and Management*, *131*. https://doi.org/10.1061/(ASCE)0733-9364(2005)131:4(459)

Claridge, T., 2004. Designing social capital sensitive participation methodologies. Report, Social Capital Research, Brisbane, Australia.

Cornwall, A., & Aghajanian, A. (2017). How to Find out What's Really Going On: Understanding Impact through Participatory Process Evaluation. *World Development*, *99*, 173–185. https://doi.org/10.1016/j.worlddev.2017.07.010

Crawford, I. (1997). Marketing Research and Information Systems. Rome: Food and Agriculture Organization of the United Nations. Retrieved December 24, 2019 from https://books.google.it/books?id=ns8_OpkPUi8C.

Dey, P. K. (2002). Benchmarking project management practices of Caribbean organizations using analytic hierarchy process. *Benchmarking: An International Journal*, *9*(4), 326–356. https://doi.org/10.1108/14635770210442680

DFID. (2002). *Monitoring and Evaluation Series - Field Guide for Conducting Participatory Impact Assessment*. (L. Oruko, Ed.). London: Client-oriented Agricultural Research and Dissemination Project.

Dijkstra, T. K. (2013). On the extraction of weights from pairwise comparison matrices, *21*, 103–123. https://doi.org/10.1007/s10100-011-0212-9

Dunnett, A., Shirsath, P. B., Aggarwal, P. K., Thornton, P., Joshi, P. K., Pal, B. D., … Ghosh, J. (2018). Multi-objective land use allocation modelling for prioritizing climate-smart agricultural interventions. *Ecological Modelling*, *381*(December 2017), 23–35. https://doi.org/10.1016/j.ecolmodel.2018.04.008

EC. (2013). Programme indicatif national 11eme fonds europeen de developpement 2014-2020. Retrieved from https://ec.europa.eu/europeaid/countries/congo-democratic-republic_en

Emrouznejad, A., & Marra, M. (2017). The state of the art development of AHP (1979–2017): A literature review with a social network analysis. *International Journal of Production Research*, *55*(22), 6653–6675. https://doi.org/10.1080/00207543.2017.1334976

EU. (2019). Food and agriculture sector overview. Retrieved December 5, 2019, from https://ec.europa.eu/europeaid/sectors/food-and-agriculture_en

Fals-Borda, O., & Rahman, M. A. (1991). *Action and Knowledge -Breaking the Monopoly with Participatory Action*. New York: Apex Press.

FAO. (1995). *Dimensions of need - An atlas of food and agriculture*. (T. Loftas, J. Lyons, H. de Mattos-Shipley, F. Greenland, J. Holgate, & A. Sanchez, Eds.). Rome: Food and Agriculture Organization of the United Nations.

FAO. (2001). *Farming systems and poverty. Improving farmers' livelihoods in a changing world.* (M. Hall, J. Dixon, A. Gulliver, & D. Gibbon, Eds.). Rome: FAO Publishing and Multimedia Service.

FAO. (2006). *Participatory Rural Appraisal (PRA) Manual*. (R. Paul, Ed.). Rome: Food and Agriculture Organization of the United Nations.

FAO. (2006a). *Toolkit for the Application of Participatory and Negociated Territorial Development*. Rome: FAO Publishing and Multimedia Service.

FAO. (2017). *The State of Food and Agriculture - Leveraging Food Systems for Inclusive Rural Transformation*. Rome: FAO Publishing and Multimedia Service.

FAO-CFS. (2015). *Committee on World Food Security: A Multi-Stakeholder, Evidence-Based Approach to Policy Making*. Rome: FAO Publishing and Multimedia Service.

Gass, S. I. (1998). Tournaments, transitivity and pairwise comparison matrices. *Journal of the Operational Research Society*, *49*(6), 616–624. https://doi.org/10.1057/palgrave.jors.2600572

Golden, B. L., Wasil, E. A., Daniel, W. D., Alexander, J. M., Dolan, J. G., Harker, P. T., … Others. (1989). *The Analytic Hierarchy Process*. (Bruce L. Golden, E. A. Wasil, & P. T. Harker, Eds.). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-50244-6

Ho, D.. (2005). The importance of property-specific attributes in assessing CBD office building quality. *Journal of Property Investment & Finance*, *23*(5), 424–444. https://doi.org/10.1108/14635780510616025

Ho, W., & Xin, M. (2018). The state-of-the-art integrations and applications of the analytic hierarchy process. *European Journal of Operational Research*, *267*, 399–414. https://doi.org/10.1016/j.ejor.2017.09.007

Hsieh, N.-H. (2007). Is Incomparability a Problem for Anyone? *Economics and Philosophy*, *23*(1), 65–80. https://doi.org/10.1017/s0266267107001228

IFAD. (2000). *Promoting Participation in Development Projects: IFAD's Experience in Promoting Participatory Approaches in Development Projects in Asia*. (M. T. L. Debuque, Ed.). Somewhere: IFAD publishing services.

IFAD. (2015). *Projet d ' Appui au Secteur Agricole dans la Province du Nord Kivu (PASA-NK)*. Rome: IFAD publishing services.

International Monetary Fund. (2014). Democratic Republic of the Congo: Selected Issues. *Country Report*, (15). https://doi.org/10.1111/j.1365-3156.2008.02150.x.dhfr

IPES-Food. (2016). *From uniformity to diversity: a paradigm shift from industrial agriculture to diversified agroecological systems*. International Panel of Experts on Sustainable Food systems.

IPES-Food. (2017). *What Makes Urban Food Policy Happen? Insights from Five Case Studies*. International Panel of Experts on Sustainable Food systems.

Ishizaka, A., & Labib, A. (2011). Review of the main developments in the analytic hierarchy process. Expert Systems with Applications, 38(11), 14336–14345. https://doi.org/10.1016/j.eswa.2011.04.143

Israel, B. A., Schulz, A. J., Parker, E. A., & Becker, A. B. (2002). Review of community-based research: Assessing Partnership Approaches to Improve Public Health. *Annual Review of Public Health*, *19*(1), 173–202. https://doi.org/10.1146/annurev.publhealth.19.1.173

Krejčí, J., & Ishizaka, A. (2018). FAHPSort: A Fuzzy Extension of the AHPSort Method. *International Journal of Information Technology and Decision Making*, *17*(4), 1119–1145. https://doi.org/10.1142/S0219622018400011

Kułakowski, K. (2018). Inconsistency in the ordinal pairwise comparisons method with and without ties. *European Journal of Operational Research*, *270*(1), 314–327. https://doi.org/10.1016/j.ejor.2018.03.024

Lam, P.-K., & Chin, K.-S. (2005). Identifying and prioritizing critical success factors for conflict management in collaborative new product development. *Industrial Marketing Management*, *34*(8), 761–772. https://doi.org/https://doi.org/10.1016/j.indmarman.2004.12.006

Lane, J. (1995) Non-governmental organizations and participatory development: the concept in theory versus the concept in practice. In '*Power and Participatory Development*'. (Ed. S Wright). London: Intermediate Technology Publications.

Larrick, R. P. (2008). *Blackwell Handbook of Judgment and Decision Making*. Hoboken: Blackwell Publishing Ltd. https://doi.org/10.1002/9780470752937.ch16

Liberatore, M. J., & Nydick, R. L. (2008). The analytic hierarchy process in medical and healthcare decision-making: A literature review. *European Journal of Operational Research*, *189*(1), 194–207. https://doi.org/https://doi.org/10.1016/j.ejor.2007.05.001

Mayne, J. (2008). *Building an Evaluative Culture for Effective Evaluation and Results Management* (ILAC Worki). Rome: Institutional Learning and Change Initiative.

MONUSCO. (2018). UN reports hundreds of human rights violations as security situation in North Kivu deteriorates. Retrieved April 22, 2019, from https://monusco.unmissions.org/en/un-reports-hundreds-human-rights-violations-security-situation-north-kivu-deteriorates

Mustafa, Muhammad A. & Al-Bahar, Jamal F. (1991). Project risk assessment using the analytic hierarchy process. *IEEE Transactions on Engineering Management*, *38*(1), 46–52. https://doi.org/10.1109/17.65759

Narayanasamy, N. (2009). *Participatory Rural Appraisal: Principles, Methods and Application*. New Delhi: SAGE Publications India Pvt Ltd. https://doi.org/10.4135/9788132108382

Nurmi, H. (2014). Are We Done With Preference Rankings ? If We Are , Then What ? *Operations Research and Decisions*, *24*(4), 63–74. https://doi.org/10.5277/ord140405

O'Neill, N. K. F. (1989). *Public participation in local government : summary : a report to the Hawkesbury Shire Council*. (N. O'Neill & H. K. Colebatch, Eds.). Windsor: Hawkesbury Shire Council.

Ostrom, E. (2009). A General Framework for Analyzing Sustainability of Social-Ecological Systems. *Science*, *325*(5939), 419 LP – 422. https://doi.org/10.1126/science.1172133

Palcic, I., & Lalic, B. (2009). Analytical hierarchy process as a tool for selecting and evaluating projects. *International Journal of Simulation Modelling*, *8*(1), 16–26. https://doi.org/10.2507/IJSIMM08(1)2.112

Pan, D., Lu, X., Liu, J., & Deng, Y. (2014). A Ranking Procedure by Incomplete Pairwise Comparisons Using Information Entropy and Dempster-Shafer Evidence Theory. *Scientific World Journal*, *2014*. https://doi.org/10.1155/2014/904596

Pazek, K., & Rozman, Č. (2005). Application of analytical hierarchy process in agriculture. *Poljoprivreda (Osijek)*, *11*.

Ramanathan, R., & Ganesh, L. S. (1995). Using AHP for resource allocation problems. *European Journal of Operational Research*, *80*(2), 410–417. https://doi.org/https://doi.org/10.1016/0377-2217(93)E0240-X

Russo, R. D. F. S. M., & Camanho, R. (2015). Criteria in AHP: A systematic review of literature. *Procedia Computer Science*, *55*(July), 1123–1132. https://doi.org/10.1016/j.procs.2015.07.081

Saaty, T. L. (1980). The Analytic Hierarchy Process. *Education*, 1–11. https://doi.org/10.3414/ME10-01-0028

Saaty, T. L. (1987). The analytic hierarchy process-what it is and how it is used. *Mathematical Modelling*, *9*(3–5), 161–176. https://doi.org/10.1016/0270-0255(87)90473-8

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83. https://doi.org/10.1504/IJSSCI.2008.017590

Saaty, T. L., & Vargas, L. G. (1984). Inconsistency and rank preservation. *Journal of Mathematical Psychology*, *28*(2), 205–214. https://doi.org/10.1016/0022-2496(84)90027-0

Saaty, T. L., & Vargas, L. G. (1993). Experiments on rank preservation and reversal in relative measurement. *Mathematical and Computer Modelling*, *17*(4), 13–18. https://doi.org/https://doi.org/10.1016/0895-7177(93)90171-T

Shumway, C. R. (1981). Subjectivity in Ex-ante Research Evaluation. *American Journal of Agricultural Economics*, *63*(1), 169–173. https://doi.org/10.2307/1239826

Sipahi, S., & Timor, M. (2010). The analytic hierarchy process and analytic network process: an overview of applications. *Management Decision*, *48*(5–6), 775–808. https://doi.org/10.1108/00251741011043920

Siraj, S., Mikhailov, L., & Keane, J. A. (2015). PriEsT: an interactive decision support tool to estimate priorities from pairwise comparison judgments. *International Transactions in Operational Research*, *22*(2, SI), 217–235. https://doi.org/10.1111/itor.12054

Stoklasa, J., & Krejci, J. (2018). Aggregation in the analytic hierarchy process : Why weighted geometric mean should be used instead of weighted arithmetic mean. *Expert Systems With Applications*, *114*, 97–106. https://doi.org/10.1016/j.eswa.2018.06.060

Sutton, S. G., & Arnold, V. (2013). Focus group methods: Using interactive and nominal groups to explore emerging technology-driven phenomena in accounting and information systems. *International Journal of Accounting Information Systems*, *14*(2), 81–88. https://doi.org/https://doi.org/10.1016/j.accinf.2011.10.001

Tiwari, D. N., Loof, R., & Paudyal, G. N. (1999). Environmental–economic decision-making in lowland irrigated agriculture using multi-criteria analysis techniques. *Agricultural Systems*, *60*(2), 99–112. https://doi.org/https://doi.org/10.1016/S0308-521X(99)00021-9

Tran, N. M. (2013). Pairwise ranking: Choice of method can produce arbitrarily different rank order. *Linear Algebra and Its Applications*, *438*(3), 1012–1024. https://doi.org/10.1016/j.laa.2012.08.028

UNDP. (2009). *Nord Kivu Province profile*. (G. Aho, J. R. Ravelosoa, & A. Lukaku Zinga, Eds.). Kinshasa: UNDP Publishing.

UNDP. (2012). *Multi-Stakeholder Decision-Making: A Guidebook for Establishing a Multi-Stakeholder Decision-Making Process*.

UNESCO. (2008). *Data Requirements for Integrated Urban Water Management*. (T. Fletcher & A. Deletic, Eds.). London: CRC Press. https://www.taylorfrancis.com/books/9781482266191

Vansina, J. (1966). *Introduction à l'ethnographie du Congo*. Kinshasa: Université Lovanium.

Wedley, W. C. (1993). Consistency prediction for incomplete AHP matrices. *Mathematical and Computer Modelling*, *17*(4–5), 151–161. https://doi.org/10.1016/0895-7177(93)90183-Y

World Bank. (2018). Participatory Rural Appraisal (PRA) Techniques. Retrieved May 24, 2019, from http://go.worldbank.org/EU99NNIVS0

World Poverty Clock. (2019). World poverty clock. Retrieved May 24, 2019, from https://worldpoverty.io/blog/index.php?r=18

Zahedi, F. (1986). The Analytic Hierarchy Process: A Survey of the Method and Its Applications. *Interfaces*, *16*(4), 96–108.

7   Conclusions

In Development Cooperation, the dissemination at all levels of a culture of evaluation and the construction of solid, adaptive and inclusive feedback and decision-making systems are seen as key objectives to foster the impact of interventions aiming sustainability (OECD, 2013). These goals determine the need to frame all development interventions, from global policies to local projects, in a whole interconnected chain, able to link the results on different scales and to evaluate, both ex-ante and ex-post, the overall contribution towards the SDGs (OECD, 2017).

With regard to agricultural development, an intensification of research in the field of evaluation can be identified during the '60, in parallel with the spreading of sustainability targets. However, it is only since the 1990s that policies have neatly set the goal of converging on the sustainability of human development on a global scale. Therefore, since then, and in a more decisive manner in the last decades, political decision makers and large international organizations have faced the problem of assessing the real impact of policies and interventions in achieving the 'sustainability' objective (United Nations Conference on Environment and Development, 1992; OECD, 2016; 2019).

Given the different points of view of the multiple actors in DC and the complexity and diversity of the socio-ecological and economic contexts in which they act, the application of the result-based management approach determines methodological and operational difficulties that are currently the subject of debate both in the academic sphere and among DC providers (Sasaki , 2006; OECD, 2014). In order to tackle these constraints, development should be managed as a dynamic concept (Emas, 2015; Sauvé et al., 2016; Fuuda-Parr, 2016) and therefore, responsive and adaptable evaluation frameworks should be created.

The evaluation process has to be embedded in the development policy process: study, design, implementation and evaluation phases have to be dynamically interconnected to form such a system (Best et al., 2010). New methodologies, shared across the system, should be designed in order to let the different evaluation actors link their results across the scales in the DC system.

Our research allowed, first of all, to retrace the existing debate on the issue of the evaluation in DC. Secondly, the research focused on a specific case study located in the GD, NK, DRC, in order to experiment a constructivist and transformative approach and new evaluation methodologies.

We were confronted with the evaluation of the agricultural sector's development potential. Given that agricultural potential is composed of the whole diversity of services that specific socioecological systems can provide, we focused on a given study area and on the identification of priority interventions meant to trigger an improved sustainability of the agricultural sector. We experimented a participatory approach to the identification of the priority of possible sustainability-oriented interventions and namely we experienced how a constructivist/transformative approach can serve this purpose.

In particular, our literature review has allowed to reconstruct the state of the art and to focus, among the existing evaluation approaches and methods, those most suitable in order to face a specific case of agroecological planning in a complex context, such as the GD.

Literature (Chambers, 2017; Mertens, 2019) suggests that, especially in complex contexts, the use of a constructivist (or even transformative) approach (Denzin et al., 2005) and of mixed methods, including the participatory approach (Chambers, 1994), can contribute to the setting of a 'local' DC system which is able to identify and implement development interventions really contributing to sustainability (Guba, 1981; 1989; 1990). Our research has allowed us to experiment with the construction of a multi-method strategy for assessing the potential for agricultural development of the GD. From a methodological point of view, the results obtained with our constructivist, participatory, systematic (Rossi et al., 2004) approach and thanks to the bundle of mixed methods we used, confirm that complex systems demand for adaptive evaluation and decision-making models of action. In this perspective, the ad hoc bibliographic research, the three surveys, the participatory path for information collection and analysis and the study of the area through remote sensing fitted for the purpose. The results were integrated into a single corpus allowing the triangulation of the sources and the production of an in-depth technical evaluation report (not published, property of Caritas Development Goma NGO, available on request). The main added value produced by this phase of work was the transparency of the evaluation process that in turns facilitated appropriation of results by the participants. This was due essentially to the participatory methodology (OECD, 2008).

Moreover, the information produced was used to develop a decision-support model, based on SPWR and AHP, regarding the identification of priority interventions. The DSS allowed to analyse in a participatory manner the reason for accorded priority and so to outline the TOC behind each

activity. In particular we implemented a modified version of AHP (Saaty, 1980; 1987; 2008), namely participatory AHP, first in its simplified form of SPWR and then in its most complete methodological form (Forman, 1998; Narayanasamy, 2009; Boukherroub et al., 2017). The process animated during the research allowed for the creation of a good degree of consensus within the working group in charge of planning and writing the new project proposal that, building on the research results, matches the existing international guidelines for agricultural development in the region (IFAD, 2015; FAO, 2017).

The main added value of this priority identification phase has been the demonstration to the participants of the possibility of implementing rigorous albeit participatory, methodologies for structuring problems and identifying priority solutions. This was achieved by implanting an action research path which was able to 'train-by-doing'.

During our research, we were able to foster the valorization of different opinions in group decision-making, so strengthening the existing idea that the use of PAHP allows diverse subjective and objective factors to be considered in the decision process as other authors already proved (Aznar et al., 2011; Russo et al., 2015).

Furthermore, the DSS was able to solve the resources allocation problem in the context of a new project proposal containing the previously identified interventions. Once again, the participatory approach stands for the main added value of these results, because it was found able to determine an immediate appropriation of the TOC and to point out the relative importance underlying each intervention, thus going beyond the boundaries of the single design phase and determining a positive influence on the subsequent implementation phases of the activities themselves.

This is what an evaluation culture should determine in the first place: the possibility to run evaluative actions that go beyond the boundaries of the single DC phase (in the same programme cycle) or a single DC cycle (across subsequent programme cycles, in long term perspective), determining a positive influence on the subsequent phases or cycles, and therefore improving the management towards sustainability.

In conclusion, it seems important to stress that the shared value of the results obtained from the three working phases we described in chapters 4, 5 and 6 depends largely on the choice of the global approach and the tools / methodologies used. This choice was guided by the desire to

enhance the transformative effect of the evaluation process itself through the involvement of various actors. In fact, the use of a specifically participatory research method has allowed not only to collect primary data and to apply a structured evaluation and decision-making process, but it has also allowed to train the project team and actively involve many stakeholders of the local agricultural sector. According to several recent studies, these intrinsic elements of a constructivist and transformative evaluation process determine a real increase in the ability of an intervention to impact on social, economic and environmental sustainability (Mertens, 2019).

The results obtained during the present research project are therefore positive, above all for the practical interest in the local project cycle management system of the NGO Caritas Development Goma which should be able to take advantage, in a short and medium term, of the methodologies described so far. Secondly, the results are interesting for the academic community and in particular for the supporters of a constructivist and social-justice oriented approach in the evaluation of development programmes. The AHP methodology, although it is no longer an innovative methodology in itself and it has been widely used to inform decision-making, has rarely been used in a transformative perspective as a method of seeking consensus in multi-stakeholder decisions (Sipahi et al., 2010; Dong et al., 2016). The emphasis in our research on the participatory approach made it possible to experiment with the combination of mixed data collection methods and with a consolidated decision-aiding methodology, originally implemented through the animation of deliberative participatory discussions. The way we implemented AHP constitutes a novelty both in regard of its most common applications and of the more usual body of consensus-building techniques, therefore generating new knowledge about its adaptability and thus proposing to spread its use in DC.

The results obtained are, however, to be considered preliminary. In fact, they can be considered incomplete in two directions: firstly towards a scale of greater detail, since the results of our research could be valorized by studying further models to tackle spatial heterogeneity through the planning of technical activities on the ground, with particular reference to land suitability to different uses. Secondly, in the direction of a more global scale, since our results emerge from a single case study and don't allow for a generalization. Further research should be implemented in order to identify a new validated and easily replicable operational model based on our approach and methodologies.

The addressed issues therefore require further research which, according to our experience, should focus on the following two lines of work.

1. The creation of a detailed SDSS that allows to further detail the interventions identified so far.

   In fact, although the proposed framework was able to accompany the activities in the different phases of, planning, design and implementation of the evaluation and during the elaboration of a new project proposal, it was not possible to investigate more technical aspects such as the production potential and the sustainability of the different agricultural crops on the target territory.

   In other words, our mixed method approach to information gathering and analysis, allowed for the confirmation that the increase in crop yields is one of the priorities, even if it is not the only neither the first one. Further research is needed to create specific SDSS able to guide the production choice in the different zones. Although during this research a specific SDSS focused on the land capability / suitability methodology (FAO, 1993; Kutter et al., 1997; FAO, 2007) was prompted, time constrains made it not possible to complete the path.

   In particular, during our research it was possible to carry out three focus groups aimed at defining the descriptive criteria relating to the territory (they are the parameters with which the land capability is described: 'measuring the vocation *of what* to what') and the comparative criteria related to the potential uses of the territory itself (they constitute the subject in relation to which the vocation is measured: 'measuring the vocation of what *to what*'). It was also possible to build, starting from the cartographic bases collected during the study of the area by remote sensing, a GIS of the Diocese of Goma containing the information layers related to the descriptive and comparative parameters mentioned. Further time and resources are needed to complete the production of the cartography of the land capability of the study area. In fact, according to the best practices for the construction of a SDSS, such interactive cartography would be potentially very useful if used to further detail the results obtained so far (Van Ittersum et al., 2001).

   These further research lines should be developed with a participatory approach and through an implementation process that includes design, integration, calibration and validation, because the efficiency with which the SDSS manage to facilitate reasoning and choice is in fact strongly correlated to the methodology for their own construction (OECD, 2008).

Although the debate is still open on the potential of the different SDSS construction methodologies, numerous researches (Mac Kintosh et al., 2001; Keenan, 2002; Sugumaran et al., 2007; Raudsepp-Hearbe et al., 2010) show that their use may be useful both in the participatory processes of territorial governance (Worrall et al., 1997; Keenan et al., 2019) and in multiscale evaluations (Smith et al., 1993; Lopez-Ridaura et al., 2005).

Moreover, our research strategy included remote sensing techniques and produced a preliminary land use map of the Diocese of Goma, with special focus on the automated identification of agricultural production systems through the analysis of the entropy of satellite images. In this perspective, further researches are needed to improve the procedure and to deliver a validated remote sensing technique for land use classification.

2. Further research is needed in order to formalize a new framework for ex-ante evaluation of agricultural development programmes which is easily replicable and that allows to respect the currently accepted criteria and to use inclusive and ecological methodologies such as the ones we proposed.

Too many models and frameworks, more or less structured, formalized and branded already exist, therefore we do not suggest lengthening the list. Rather we think that the ongoing synthesis of basic values already existing in the pool of the different evaluation paradigms should be accelerated.

In other words, basic values that are fundamental for having a sustainable 'development system' should be clearly pointed out in order to have a generalized framework for evaluation and decision-making in agricultural development.

As a matter of fact, the procedure we used was strongly influenced by the peculiarities of the case study in question. In order to satisfy the needs, it has been conceived upon numerous instances that are typical of the different existing evaluation paradigms.

This is the case of the focus on methods that is typical of the post-positivist evaluation flow, which we integrated in building a responsive mixed method including quantitative techniques. It is also the case of the focus on uses, typical of the pragmatic paradigm, and the case of the attention to values diversity and to qualitative information emerging from relationships configuration, typical of the constructivist.

Finally, it is the case of empowering role attributed, through widespread and fair participation, to the evaluation and decision-making process itself, which is typical of the

transformative paradigm. By integrating these values, we were able to ensure scientific rigour while maintaining the necessary methodological adaptability in dealing with, and valorizing, the complexity. The key and generalizable assets of our work can be summarized as the following:

- Openness to complexity.

  In order to increase the sustainability of the interventions and therefore significantly contribute to the sustainable evolution of the target system, the multi-stakeholders' approach and the multidisciplinary opening remain fundamental characteristics of the development governance systems (WFP, 2000; FAO, 2003). The attitude towards the complexity of the system in question is important to improve the development system in every phase (Janssen, 2002). The way local stakeholders perceive the development system is fundamental: preliminary studies, evaluations (ex-ante, ad-interim, final or ex-post), selection of alternative interventions and the implementation processes of agricultural development programmes depend largely on the way these activities are proposed (Milan Food Policy Office, 2016; UNEP, 2019).

- Methodological adaptability, governance responsiveness and learning ability of the development evaluation system.

  Stemming from the attitude towards complexity, flexibility remains a key feature of sustainability-oriented development systems. But flexibility implies, like in a matryoshka, the intelligence to choose between alternatives, where intelligence means both the data and the ability to discern information from them. DC systems are nowadays striving for effective results management systems that anyway are still lacking (OECD-DAC, 2000, 2013; 2014; OECD, 2016) and that need the contribution of ad hoc research.

- Eco-engagement and agroecology.

  It means engagement for the sake of 'eco' or 'oikos' which means 'our common home'. This engagement, formerly part of the transformative evaluation paradigm portfolio of values, becomes fundamental for the DC governance because social fairness and equality go together with the sustainable management of the environment through wise economical administration of resources (IPES-Food, 2017). If sustainability must be achieved, the development system should reason

196

eco-logically in a way that includes 'eco-nomically', bringing back the nowadays concepts of economy to something less financial and much more human or natural. Agroecology is nowadays spreading from grassroots realities towards higher governance level as a possible new configuration of values and operational methodologies fundamental to shift the overall development paradigm (IPES-Food, 2016; 2018).

In order to improve the actual assets of DC (and evaluation) systems, further research is needed to fully integrate more reductionist approaches within a development management system that has to be holistic in front of complexity (Solesbury, 2001). In fact, reductionist approach is mainly viable for detailed and specific queries in confined subsystems, aimed at the formalization of specific assessment methods and models capable of managing and comparing more and more information in shared knowledge management systems. The holistic approach is mandatory to profitably react towards global complexity and interconnection of subsystems. Consequently, further research is needed on adaptive knowledge management methodologies, such as PAHP among many others, that are able to let DC providers speak a common language between countries and scales in the quest for development sustainability (OECD, 2014; 2016; 2019; UNEP, 2019).

Research could take the direction towards the ability of such a system to adapt as the time passes by, for instance by spreading the implementation of widespread action-research and learning-by-doing training in evaluation.

Otherwise, more insights are needed on the practical effect of shifting governance structures according to the agroecological set of values, by studying the effects both on more technical aspect such as agronomical and ecological aspects, and on higher scale issues such as food, economic, social and environmental policies.

7.1 References

Aznar, J., Guijarro, F., & Moreno-Jiménez, J. M. (2011). Mixed valuation methods: A combined AHP-GP procedure for individual and group multicriteria agricultural valuation. *Annals of Operations Research*, *190*(1), 221–238. https://doi.org/10.1007/s10479-009-0527-2

Best, A., & Holmes, B. (2010). Systems thinking, knowledge and action: Towards better models and methods. *Evidence and Policy*, *6*(2), 145–159. https://doi.org/10.1332/174426410X502284

Boukherroub, T., LeBel, L., & Ruiz, A. (2017). A framework for sustainable forest resource allocation: A Canadian case study. *Omega*, *66*, 224–235. https://doi.org/https://doi.org/10.1016/j.omega.2015.10.011

Chambers, R. (1994). The origins and practice of participatory rural appraisal. *World Development*, *22*(7), 953–969. https://doi.org/https://doi.org/10.1016/0305-750X(94)90141-4

Chambers, R. (2017). *Can We Know Better? Reflection for development.* Bourton on Dunsmore: Practical Action Publishing Ltd. https://doi.org/10.3362/9781780449449

Denzin, N. K., & Lincoln, Y. S. (2005). *The SAGE handbook of qualitative research*. Thousand Oaks: Sage Publications.

Dong, Q., & Cooper, O. (2016). A peer-to-peer dynamic adaptive consensus reaching model for the group AHP decision making. *European Journal of Operational Research*, *250*(2), 521–530. https://doi.org/10.1016/j.ejor.2015.09.016

Dumanski, A. J. S. J. (1993). *FESLM: An international framework for evaluating sustainable land management*. Rome: Food and Agriculture Organization of the United Nations.

Emas, R. (2015). *The Concept of Sustainable Development: Definition and Defining Principles* (Brief for GDSR 2015). Florida International University.

Fao. (1993). *Guidelines for land-use planning*. *FAO Development Series* (Fao Develo). Rome: FAO Publishing and Multimedia Service.

FAO. (2003). *Participatory Development: guidelines on beneficiary participation in agricultural and rural development*. (B. van Heck, Ed.) (2nd ed.). Rome: FAO Publishing and Multimedia Service.

FAO. (2007). *Land evaluation: towards a revised framework*. Rome: Electronic publishing policy and support branch. https://doi.org/10.1016/j.geoderma.2008.11.001

FAO. (2017). *The State of Food and Agriculture - Leveraging Food Systems for Inclusive Rural Transformation*. Rome: FAO Publishing and Multimedia Service. http://www.fao.org/3/a-i7658e.pdf

Forman, E., & Peniwati, K. (1998). Aggregating individual judgments and priorities with the analytic hierarchy process. *European Journal of Operational Research*, *108*(1), 165–169. https://doi.org/https://doi.org/10.1016/S0377-2217(97)00244-0

Guba, E. C. (1990). The alternative paradigm dialog. In E. G. Guba & A. P. Conference (Eds.), *The paradigm dialog* (pp. 17–27). Newbury Park: Sage Publications.

Guba, E. G. (1981). *Effective evaluation: Improving the usefulness of evaluation*. San Francisco: Jossy-Bass.

Guba, E. G. (1989). *Fourth generation evaluation*. (Y. S. Lincoln, Ed.). Newbury Park: Sage Publications.

IFAD. (2015). *Projet d ' Appui au Secteur Agricole dans la Province du Nord Kivu (PASA-NK)*. Rome: IFAD publishing services.

IPES-Food. (2016). *From uniformity to diversity: a paradigm shift from industrial agriculture to diversified agroecological systems*. International Panel of Experts on Sustainable Food systems. www.ipes-food.org

IPES-Food. (2017). *What Makes Urban Food Policy Happen? Insights from Five Case Studies*. International Panel of Experts on Sustainable Food systems. www.ipes-food.org

IPES-Food. (2018). *Breaking away from industrial food and farming systems: Seven case studies of agroecological transition*. International Panel of Experts on Sustainable Food systems. www.ipes-food.org

Janssen, M., & Van Peach, H. (2002). *Complexity and Ecosystem Management: The Theory and Practice of Multi-agent Systems*. Cheltenham: Edward Elgar Pub.

Kee, P. G. W., & Mackintosh, R. (2001). *he freedom economy : gaining the M-commerce edge in the era of the wireless Internet*. New York: Osborne/McGraw-Hill.

Keenan, P. B., & Jankowski, P. (2019). Spatial Decision Support Systems: Three decades on. *Decision Support Systems*, *116*, 64–76. https://doi.org/https://doi.org/10.1016/j.dss.2018.10.010

Keenan, P.B. (2002). Geographic Information Systems. In H. Bidgoli (Ed.), *Encyclopedia of Information Systems*. San Diego: Academic Press.

Kutter, A., Nachtergaele, F. O., & Verheye, W. H. (1997). The new FAO approach to land use planning and management , and its application in Sierra Leone. *Itc Journal*, *3*(4).

López-Ridaura, S., van Keulen, H., van Ittersum, M. K., & Leffelaar, P. a. (2005). Multi-scale sustainability evaluation of natural resource management systems: Quantifying indicators for different scales of analysis and their trade-offs using linear programming. *International Journal of Sustainable Development & World Ecology*, *12*(2), 81–97. https://doi.org/10.1080/13504500509469621

Milan Food Policy Office. (2016). Open research needs & most interesting innovation. Retrieved December 6, 2019, from http://www.milanurbanfoodpolicypact.org/wp-content/uploads/2018/06/study.pdf

Narayanasamy, N. (2009). *Participatory Rural Appraisal: Principles, Methods and Application*. New Delhi: SAGE Publications India Pvt Ltd. https://doi.org/10.4135/9788132108382

OECD (Organisation for Economic Co-operation and Development). (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. *Methodology* (Vol. 3). https://doi.org/10.1787/9789264043466-en

OECD. (2005). *Paris Declaration on Aid Effectiveness*. Paris: OECD. http://dx.doi.org/10.1787/9789264098084-en.

OECD. (2014). *Measuring and managing results in development co-operation A review of challenges and practices among DAC members and observers*. Paris: OECD Pubblications. https://www.oecd.org/dac/peer-reviews/Measuring-and-managing-results.pdf

OECD. (2016). *Evaluation Systems in Development Co-operation*. Paris: OECD Publishing. https://doi.org/https://doi.org/https://doi.org/10.1787/9789264262065-en

OECD. (2017). *Strengthening the results chain: Synthesis of case studies of results-based management by providers*. (R. Zwart, Ed.), *OECD Development Policy Papers*. Paris: OECD Publishing. https://doi.org/doi: 10.1787/544032a1-en

OECD. (2019). *Measuring Distance to the SDG Targets 2019 An Assessment of Where OECD Countries Stand*. *Multilingual Summaries*. Paris: OECD Publishing. https://doi.org/10.1787/e0f4d2ac-en

OECD-DAC. (2000). *Results based management in the development co-operation agencies: a review of experience*. (A. Binnendijk, Ed.). Paris: DAC Working Party on Aid Evaluation.

OECD-DAC. (2013). *Evaluating Development Activities: 12 lessons from the OECD-DAC*. Paris: OECD Publishing.

Raudsepp-Hearne, C., Peterson, G. D., & Bennett, E. M. (2010). Ecosystem service bundles for analyzing tradeoffs in diverse landscapes. *Proceedings of the National Academy of Sciences*, *107*(11), 5242 LP – 5247. https://doi.org/10.1073/pnas.0907284107

Rossi, P. H. (2004). *Evaluation : a systematic approach*. (P. H. Rossi, M. W. Lipsey, & H. E. Freeman., Eds.). Thousand Oaks: SAGE Publications.

Russo, R. D. F. S. M., & Camanho, R. (2015). Criteria in AHP: A systematic review of literature. *Procedia Computer Science*, *55*(July), 1123–1132. https://doi.org/10.1016/j.procs.2015.07.081

Saaty, T. L. (1980). The Analytic Hierarchy Process. *Education*, 1–11. https://doi.org/10.3414/ME10-01-0028

Saaty, T. L. (1987). The analytic hierarchy process-what it is and how it is used. *Mathematical Modelling*, *9*(3–5), 161–176. https://doi.org/10.1016/0270-0255(87)90473-8

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83. https://doi.org/10.1504/IJSSCI.2008.017590

Sasaki, R. (2006). A Review of the History and the Current Practice of Aid Evaluation. *Journal of MultiDisciplinary Evaluation*, *3*(5), 55–88.

Sauvé, S., Bernard, S., & Sloan, P. (2016). Environmental sciences, sustainable development and circular economy: Alternative concepts for trans-disciplinary research. *Environmental Development*, *17*, 48–56. https://doi.org/https://doi.org/10.1016/j.envdev.2015.09.002

Sipahi, S., & Timor, M. (2010). The analytic hierarchy process and analytic network process: an overview of applications. *Management Decision*, *48*(5–6), 775–808. https://doi.org/10.1108/00251741011043920

Solesbury, W. (2001). *Evidence Based Policy: Whence it Came and Where it's Going*. London: ESRC UK Centre for Evidence Based Policy and Practice.

Sugumaran, V., & Sugumaran, R. (2007). Web-based Spatial Decision Support Systems (WebSDSS): Evolution, Architecture, Examples and Challenges. *Communications of the Association for Information Systems*, *19*. https://doi.org/10.17705/1CAIS.01940

UNEP. (1992). *The Rio Declaration on environment and development*. Rio de Janeiro: UNEP.

UNEP. (2019). *Collaborative Framework for Food Systems Transformation for sustainable food systems*. (A. Sweeting, Ed.). Nairobi: UNEP. https://www.oneplanetnetwork.org/resource/collaborative-framework-food-systems-transformation-multi-stakeholder-pathway-sustainable

van Ittersum, M. K., Roetter, R. P., van Keulen, H., de Ridder, N., Hoanh, C. T., Laborte, A. G., … Tawang, A. (2004). A systems network (SysNet) approach for interactively evaluating strategic land use options at sub-national scale in South and South-east Asia. *Land Use Policy*, *21*(2), 101–113. https://doi.org/https://doi.org/10.1016/j.landusepol.2004.02.001

WFP. (2000). *Participatory approaches* (Policy issue agenda). Rome: World Food Program.

Worrall, L., & Bond, D. (1997). Geographical Information Systems, Spatial Analysis and Public Policy: The British Experience. *International Statistical Review / Revue Internationale de Statistique*, *65*(3), 365–379. https://doi.org/10.2307/1403377

8   Acknowledgements