Institut für Lebensmittel- und Ressourcenökonomik der
Rheinischen Friedrich-Wilhelms-Universität Bonn

Evidence on the role of context for agricultural innovation and conservation

I n a u g u r a l - D i s s e r t a t i o n

zur
Erlangung des Grades

Doktor der Agrarwissenschaften
(Dr.agr.)

der
Landwirtschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

von

Dario Schulz

aus

Bonn

Bonn, 2024

Referent:	Prof. Dr. Jan Börner
Koreferent	Prof. Dr. Joachim von Braun
Externer Gutachter:	Prof. Dr. Robert Finger
Tag der mündlichen Prüfung:	03.07.2024

Danksagung

Diese Forschung war nur möglich durch die Unterstützung vieler anderer Personen, denen ich meinen Dank aussprechen möchte. Insbesondere möchte ich Taynã für ihre unerschütterliche Unterstützung und Geduld während dieses Unternehmens danken. Ohne sie hätte ich weder den akademischen Karriereweg in Betracht gezogen, noch hätte ich die für wissenschaftliche Untersuchungen notwendige Ausdauer aufbringen können.

Ich danke vor allem meinem Doktorvater Prof. Dr. Jan Börner für seine kontinuierliche Unterstützung seit meinem Masterstudium, unter anderem in Form von konstruktivem Feedback und seinem stets offenen Ohr. Außerdem schätze Ich die strategischen Ratschläge von Prof. Dr. Joachim von Braun sehr, und bin dankbar für seine Übernahme des Koreferats. Ein besonderer Dank gilt auch Prof. Dr. Robert Finger, der mir einen Forschungsaufenthalt an der ETH ermöglichte und maßgeblich an der Erstellung von Kapitel 6 beteiligt war. Bei meinen Kollegen am ILR, in der ESLU-Gruppe, und bei PhenoRob möchte Ich für die unzähligen spannenden Diskussionen, Ratschläge, aber vor allem die angenehme und bereichernde Zeit bedanken. Insbesondere danke Ich Dr. Gustavo Oliveira de Magalhães, Dr. Javier Muro, Dr. Christian Stetter, Philipp Feisthauer, Lennart Kokemohr, Rienne Wilts, Marco Ferro, Helena Engemann, Yannic Damm, Anna Massfeller, Johannes Kopton, und Dominik Suri, die einen direkten oder indirekten Beitrag zur Fertigstellung dieser Arbeit geleistet haben. Für die Bereitstellung von Daten danke Ich Santiago Mesa, Ashley Monsival, Robert Johnson, Robert Strey, und ganz besonders den Landwirt:Innen, die an den Umfragen und Versuchen teilgenommen haben.

Zudem möchte Ich mich bei Rodrigo, Luciano, Paul, und Claus bedanken, die mir stets halfen die Balance zu bewahren, sowie Frederik, Pascal, für die heiteren Zeiten und langjährigen Beistand. Schließlich möchte Ich meinen Eltern und meiner Familie danken, die mich mein ganzes Leben lang unterstützt haben.

Kurzfassung

Ziel dieser Dissertation ist die Bewertung landwirtschaftlicher Innovationen und politischer Maßnahmen zur Verbesserung der Nachhaltigkeit von Agrar- und Lebensmittelsystemen, indem das Potential und die Auswirkungen von Innovationen und politischer Maßnahmen kontextuell bewertet werden. Das erste Thema untersucht die Wirksamkeit von digitalen landwirtschaftlichen Beratungsdiensten bei der Verbesserung des Wissensstandes, der Praxisübernahme und der landwirtschaftlichen Produktivität von Kleinbauern. Die Ergebnisse von zwei randomisierten Kontrollstudien zeigen, dass das Format der digitalen Informationsübermittlung und der Studienkontext einen erheblichen Einfluss auf ihre Wirksamkeit haben. Während SMS-Nachrichten möglicherweise nicht ausreichen, um komplexes Wissen zu vermitteln, können multimediale Inhalte über Plattformen wie WhatsApp bei gut ausgebildeten Bauern erfolgreicher sein. Daher sollten digitale Maßnahmen auf den spezifischen Kontext und die vorhandenen Fähigkeiten der Landwirte abgestimmt sein.

Das zweite Forschungsthema geht über die Auswirkungen auf Betriebsebene hinaus und untersucht die Verbreitung von landwirtschaftlichen Innovationen. Die Untersuchung der Verbreitung einer Pflanzenschutz-App in Indien zeigt die entscheidende Rolle der Netzwerkinfrastruktur bei der Überwindung der digitalen Kluft und dass lokale Lernnetzwerke eine frühe und intensive Nutzung der Technologie fördern. Eine globale Meta-Analyse zeigt, dass Faktoren wie Boden und Kapital im Allgemeinen kritischer für Innovationen sind, die stark auf diese Faktoren angewiesen sind, dieser Effekt jedoch kontextuell systematisch variiert. Dies unterstreicht die Relevanz, bei der Entwicklung und Förderung landwirtschaftlicher Innovationen den lokalen Kontext zu berücksichtigen.

Das dritte Thema dieser Forschungsarbeit verlagert den Schwerpunkt von digitalen Innovationen auf politische Ansätze für eine nachhaltige ländliche Entwicklung, indem es die Umweltauswirkungen von Landbewirtschaftungspraktiken bewertet. Konkret wird der Zielkonflikt zwischen landwirtschaftlicher Produktivität und biologischer Vielfalt auf räumlich explizite Weise am Beispiel der Schnitthäufigkeit im Grünland untersucht. Darüber hinaus wird die Wirksamkeit von REDD+-Initiativen bei der Eindämmung des Klimawandels untersucht. Die Studien verdeutlichen die Notwendigkeit einer kontextsensitiven Umweltpolitik und liefern entsprechende Instrumente. Insgesamt unterstreicht die Forschung die Relevanz, digitale Interventionen maßzuschneidern, Netzwerkeffekte zu fördern und den lokalen Kontext zu berücksichtigen. Die Forschung unterstreicht die Notwendigkeit einer räumlich differenzierten Umweltpolitik, die Zielkonflikte berücksichtigt.

Abstract

This dissertation aims to explore how agricultural innovation and policies can make agri-food systems more sustainable by assessing impacts and diffusion patterns of innovations, as well as policy impacts across geographic contexts. The first theme investigates the effectiveness of digital farm advisory services on farm level in improving knowledge, practice adoption, and agricultural productivity among smallholder farmers. Results from two randomized control trials show that the format of digital information delivery and study context significantly impacts its effectiveness. While SMS messages may be insufficient for conveying complex agricultural knowledge, multimedia content delivered through platforms like WhatsApp can be more successful among well-trained farmers. This highlights the importance of tailoring digital interventions to the specific context and the existing skillset of farmers.

The second theme of the research goes beyond farm-level impacts and explores the diffusion of agricultural innovations. Studying the spread of a plant health app in India reveals the critical role of network infrastructure in overcoming the digital divide, and that extension agents and peer-to-peer learning networks promote early and intensive use of the technology. Widening the scope, a global meta-analysis reveals that factors like land availability, access to capital, and existing knowledge are generally more critical for innovations that heavily rely on these factors. However, this effect diminishes in regions where these factors are already abundant. This underscores the importance of considering the local context when designing and promoting agricultural innovations.

The third topic of this research moves the focus from digital innovations towards broader policy considerations for sustainable rural development by assessing environmental implications of land management practices. Specifically, the trade-off between agricultural productivity and biodiversity are explored in a spatially explicit manner using mowing frequency in grasslands as an example. In addition, the effectiveness of REDD+ initiatives in mitigating climate change is examined. The studies highlight the need and provide tools for context-sensitive environmental policies.

Overall, the research emphasizes the importance of tailoring digital interventions, fostering network effects, and considering the local context for maximizing the impact of these technologies. The research also underscores the need for environmental policies that account for trade-offs and the importance of spatial policy targeting sustainable rural development.

CONTENTS

List of tables

List of figures

Abbreviations

CHE	Correlated and Hierarchical Effects Model
GAP	Good Agricultural Practices
GPS	Global Positioning System
ICT	Information- and Communication Technology
IIH	Induced Innovation Hypothesis
ITT	Intention to Treat Effect
LATE	Local Average Treatment Effect
NGO	Non-Governmental Organization
PES	Payments for Environmental Services
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT	Randomized Control Trial
REDD+	Reducing emissions from deforestation and forest degradation, and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries
SDG	Sustainable Development Goals
SMS	Short Message Service
UN	United Nations
UNFCCC	United Nations Framework Convention on Climate Change

Chapter 1
Research Context[1]

1.1 Background

The agricultural sector faces the challenge of responding to the increasing demand for food, feed, and fiber in an environmentally sustainable manner. Agri-food systems are responsible for over 30% of global greenhouse gas emissions, including emissions from crop and livestock production and land use change (Tubiello et al. 2022). Agriculture is the main driver of land-use change and its expansion and intensification continue to degrade natural habitats (Kehoe et al., 2017; Pendrill et al., 2022). This degradation of natural ecosystems such as forests and grasslands drives biodiversity loss on a global scale (IPCC 2022; Newbold et al. 2015). In addition to effects of land use change, the application of chemical inputs such as nitrogen fertilizer and pesticides creates risks for surrounding ecosystems, and human health (Robertson and Vitousek 2009; Tang et al. 2021). In consequence, the agricultural sector contributes to the transgression of multiple planetary boundaries (Campbell et al. 2017). However, the ability to address these challenges is stressed by intersecting crises including climate change, conflicts, and shocks such as the Covid-19 pandemic (Hendriks et al. 2022).

[1] The present research has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2070—390732324 as part of the cluster of excellence "*Robotics and phenotyping for sustainable crop production – PhenoRob*".

Regarding the Sustainable Development Goals (SDGs) as an overarching policy agenda for 193 United Nations member states (United Nations 2015), solutions require alignment of multiple social, environmental, and economic outcomes. While there are many synergies among the SDGs, there are also trade-offs resulting from competing objectives (Pradhan et al. 2017). In the context of agriculture, specific land use decisions may imply trade-offs between ecosystem services. For example, they can affect levels of provisioning services such as yields, regulating services such as carbon sequestration, and services supporting biodiversity. Failure to address and minimize these trade-offs risks surpassing irreversible biophysical tipping points (Rockström et al. 2009) or violating the second principle of the Sustainable Development Goals of leaving no on behind (United Nations 2015). Thus, solutions to navigate these trade-offs are required for safeguarding the intertwined goals of food security and sustainability.

Agricultural innovation has been identified as a key factor to mitigate some of the negative consequences of modern agriculture (Costa et al. 2022; Poore and Nemecek 2018). Innovations, encompassing technologies and practices impacting productivity, are typically adopted and impact directly at the farm level. In particular, digital innovations are expected to contribute to more sustainable production (Herrero et al. 2020; von Braun et al. 2021), i.e., reduce tradeoffs between production and adverse environmental effects. For example, remote sensing tools enable precision agriculture, leveraging variability in site-specific conditions to improve resource use efficiency. Furthermore, information and communication technologies (ICTs), have become increasingly available across the world and can reduce existing barriers of access to and transmission of information (Nakasone and Torero 2016; Spielman et al. 2021). The use of ICTs has been linked to improved farm-level decision making resulting in higher land- and labor productivity (Fabregas et al. 2019; Rajkhowa and Baumüller 2024). However, although many innovations are seemingly economically and environmentally beneficial, uptake has often been limited in scale.

In consequence, upscaling innovation adoption to generate the desired impacts requires the development and implementation of different policies (Finger 2023). Agricultural policies can stimulate innovation adoption via market based or regulatory measures. Digital innovations such as remote sensing tools can also enable new types of policies and programs based on the scalability of information flows and analytical tools for planning, intervening, and monitoring (Ehlers et al. 2021; Jain et al. 2019).

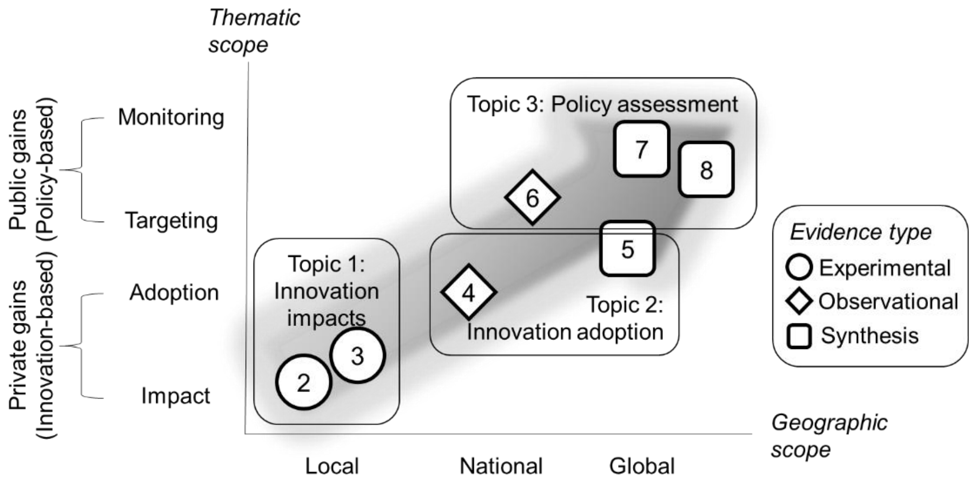
A key function of policies is to reduce the negative consequences of market failures. In the absence of markets as a price-setting institution, one refers to externalities, which are situations in which individual decisions do not lead to socially optimal outcomes, for example when private and social costs of pollution diverge. Given the many market failures around regulating and supporting ecosystem services, conservation policies are particularly important for steering and adjusting their provision. Environmental policies may target different environmental objectives, such as climate change mitigation or biodiversity preservation, and employ different mechanisms including incentives such as subsidies and conditional payments, or disincentives such as command and control regulation of alternative technologies. Such environmental policies directly affect land use decisions and are often tightly linked to or embedded in agricultural policy. Many agricultural policies mitigate some negative externalities, for example by regulating environmental pollutants including nitrate and pesticides (Fermeglia 2023; Wuepper et al. 2023). Yet, many negative environmental externalities remain inadequately or incompletely addressed by current policies, partially because of trade-offs between goals.

The potential for digital innovation to deliver expected improvements to the agri-food system rests on their adoption and actual impact, but these adoption patterns and impacts are often heterogeneous (Ogundari and Bolarinwa 2018; Ruzzante et al. 2021). That is, they may vary across geographic regions, socioeconomic subgroups, biophysical gradients - in short: the context. Similarly, environmental policies have been found to vary in their effectiveness and cost-efficiency across institutional contexts (Börner et al. 2020; Wuepper et al. 2024, 2019). Therefore, this research's overall goal is to better understand the role of context for agricultural innovation and to inform context-aware policies.

In general terms, profitability is a necessary condition for innovation at the farm level, primarily incentivized by private gains. While societal benefits may arise as positive externalities, they are not essential for farm-level innovations. This contrasts agricultural and environmental policies, which often target public goods like intact nutrient cycles and climate regulation. These policies sometimes require mitigating private losses through mechanisms like compensating forgone profits or providing subsidies. Policies also differ regarding the required scope of analysis. Since they are developed and implemented at the national level, they necessitate a broader assessment scope, often at the national or global level. Given these considerations, the structure of this dissertation is organized along these spatial scopes of analysis, ranging from the farm level to the global level, and from the focus on private gains from innovations to the consideration of public gains from policies. The benefits of private innovation can be gauged in dimensions of impact and adoption, while implementation effectiveness of policies requires monitoring and targeting for efficient resource allocation.

The remainder of this thesis is organized into three topics: (1) the impacts of farm-level innovations; (2) patterns of spatial innovation adoption; and (3) the broader impacts of policy interventions. This thematic and geographic scope is illustrated in Figure 1.1.

Figure 1.1: Schematic framework of chapters



***Note**: Numbers refer to chapters and are positioned along the thematic (Y-axis) and geographic (X-axis) scope. Thematic scope is divided into innovations and policies, based on their primary function of providing private and public benefits, respectively. Geographic scope ranges from local studies covering a specific subnational region, via national to global analyses. Experimental evidence is indicated by circle, observational evidence by diamond, evidence synthesis by rounded square. Black boxes indicate the three overarching research topics, and which individual chapters contribute to it. A grey-shaded arrow indicates empirical shift from internal to external validity. Reading example: Chapter 4 provides observational evidence on the adoption of agricultural innovation that provides private benefits covering an entire country; Chapter 7 assesses policies via a global evidence synthesis dealing with monitoring and targeting of these policies.*

1.2 Motivation and general research question

The uptake of information and communication technologies has been associated with higher agricultural productivity in low- and middle-income countries (Rajkhowa and Baumüller 2024). Digital farm advisory services can be a scalable and effective way to provide relevant information to farmers with low transaction costs. Previous studies identified overall positive impacts of digital advisory tools on farmer’s knowledge, adoption of recommended practices and production related outcomes (Fabregas et al. 2019; Rajkhowa and Qaim 2021; van Campenhout et al. 2021). Yet, evidence on farm-level impacts of digital innovations in low- and middle-income countries remains limited to specific digital information interventions and is context-specific (Baumüller 2018; Spielman et al. 2021). Previously evaluated interventions focused on specific topics such as weather and market prices (Camacho and Conover 2019) or how to tackle a specific disease (Tambo et al. 2019). The contextual differences among low- and middle-income countries and the diversity of digital advisory services warrant additional evidence to generalize and learn from previous findings. Therefore, the guiding research question for the first thematic block is:

- Research Question 1: *What impact do digital agricultural advisory services have at farm level in different contexts?*

This first research question and thematic block motivates chapters 2 and 3, both of which evaluate the causal impact of a digital farm advisory service at farm level. They differ with respect to the concrete information intervention technology and geographic study context, namely short text messages in Haiti (chapter 2) and WhatsApp chats in El Salvador (chapter 3), respectively.

The second thematic block evolves around spatial factors of innovation adoption. The profitability of innovations often depends on agro-ecologic conditions as well as input- and output markets and prices. The same profit maximizing farmer could thus take different rational adoption decisions depending on external conditions. Understanding the innovation adoption *vis-à-vis* variation in contextual conditions is therefore the main motivation for chapters 4 and 6. Therefore, the guiding research question for the second topic question is:

- Research Question 2: *How does context shape the adoption and diffusion of agricultural innovation across scales?*

Individuals access markets and information through infrastructure which in turn affects their decision to adopt innovations. Specific to digital innovations, the relevance of internet coverage and remoteness for innovation diffusion have often been highlighted (Mehrabi et al. 2021), but so far not been empirically assessed across large spatiotemporal scales. Moreover, diffusion studies often rely on spatiotemporally aggregated administrative records such as census data (e.g. Assunção et al. 2019). Such aggregation may mask relevant environmental and socioeconomic processes that influence innovation adoption but occur at spatial scales smaller than an administrative region but larger than a farm. Phenomena like social learning, or spatial neighborhood effects have received increasing attention (Abdulai 2023; Genius et al. 2014; Maertens and Barrett 2013), but previous studies have been limited to short time frames, small study regions and a limited subset of innovation adopters. Expanding the evidence of contextual innovation diffusion determinants and spatial neighborhood effects based on a comprehensive set of digital innovation adopters across an entire country (here India) is the main motivation for Chapter 4.

In addition to contextual drivers and barriers at higher spatial scales, there are farm-level determinants of innovation adoption. A large body of literature has been concerned with identifying and quantifying determinants of innovation adoption at farm level, resulting in several qualitative and quantitative reviews (Feder & Umali, 1993; Knowler & Bradshaw, 2007; Ruzzante et al., 2021; Shang et al., 2021). However, most primary studies of innovation adoption focused on small samples of adopters using cross-sectional or short panel data, often limited to specific regions. Consequently, there are substantial knowledge gaps with respect to how findings from individual primary studies can be generalized across agricultural contexts and innovation types. Specifically, the theoretical linkage between micro-level adoption determinants and macro-level structural factors has not been empirically assessed. This missing link motivates chapter 5, in which the global variation of farm-level innovation adoption determinants is explored using the induced innovation hypothesis (Binswanger 1974; Hicks 1932). The induced innovation hypothesis postulates that innovation occurs to make more efficient use of the relatively more expensive production factor, implying that global differences in factor abundances influence the adoption of innovations that rely on these factors (Hayami and Ruttan 1971).

Research topics 1 and 2 focus on farm-scale innovations, that are driven by expected private gains for the adopters. Although chapter 5 initializes a broad scale understanding of driving factors relevant to policy makers that want to target innovation adoption, it still focuses on farm-level innovations that may or may not provide positive externalities. In contrast, a central goal of many environmental policies is to provide public benefits, and incentivizing the adoption of innovations and practices with positive externalities is one means for achieving it. Hence, policy makers must be aware of the potential and realized effectiveness and spatial distribution of public gains and tradeoffs from such policies (Cord et al. 2017).

Understanding impacts of environmental policies and its contextual heterogeneity beyond individual farms and farmers is therefore the main motivation for research topic 3, accompanied by the research question:

- Research Question 3: *How can geographic context inform the design of environmental policies?*

To make an informed decision about the trade-offs between competing objectives such as food security and biodiversity conservation, policy makers need to assess cost and benefits of the counterfactual scenario, i.e., the hypothetical scenario where a policy was (not) implemented (depending on whether it is an *ex-ante* or *ex-post* assessment). The availability of high-resolution remote sensing products in combination with spatial models provides new opportunities to assess such costs and benefits across policy-relevant scopes (Ehlers et al. 2021). The complexity of many socioecological systems requires the combination of empirical and modelling approaches for policy analysis (Schlüter et al. 2023) and thereby motivates the development of a novel approach described in Chapter 6. For the case of grasslands, which provide a wide range of provisioning, regulating, and supporting ecosystem services, the chapter illustrates the integration of remote sensing, causal machine learning, and biophysical modelling to assess cost and benefits of a hypothetical conservation policy. Previous experimental and correlational studies have established a negative relationship between grassland mowing frequency and species richness (Weber et al. 2023), but a positive relationship between mowing frequency and yields (Isselstein et al. 2005). A hypothetical incentive-based biodiversity conservation approach could therefore compensate farmers for lower mowing frequencies by the value forgone yields, but this requires knowledge about the counterfactual biodiversity gains and forgone yields. Spatially explicit information about the impact of extensification on species richness and yields are not yet available, but needed to assess the potential trade-off in ecosystem services (Cord et al. 2017). The goal of chapter 6 was therefore to provide such context-specific impact estimates at a national scale and to assess spatial targeting scenarios in terms of their effectiveness and cost-efficiency.

The effectiveness of environmental policies not only differs within countries. A rapidly increasing body of literature evaluates the effectiveness of environmental policies, pointing towards variation across policy tools and context specific environmental pressure (Börner et al. 2020). Reduced Emissions from Deforestation and Degradation (REDD+) is the umbrella term for a range of carbon market financed policies and programs that aim to mitigate carbon emissions by means of forest protection (Turnhout et al. 2017). Over the past fifteen years, there have been multiple REDD+ projects around the world, ranging from regional pilot projects to national level initiatives (Simonet et al. 2018). Some of these initiatives have been evaluated with respect to their effectiveness, but so far, a systematic synthesis of the available evidence is missing. Furthermore, an assessment of impact variability across settings and comparison relative to other conservation policies and programs is missing. Therefore, the goal of chapter 7 is to collect and analyze the available evidence on REDD+ effectiveness regarding its environmental and socioeconomic outcomes. Aligned with the central topic of this dissertation, a central question is how deforestation pressure as a contextual factor moderates REDD+ effectiveness.

Given the context-dependent effectiveness of environmental policies, a central question concerns the underlying mechanisms at different scales. Farm-level observations can relate to very different geographic scales, depending on farm size. Especially in the context of farm structural change it is important to understand how these scales relate to aggregate policy outcomes. The last chapter 8 therefore revolves around the argument that smaller farms are globally better able to mitigate the trade-off between biodiversity and yields, a claim put forward by Ricciardi et al. (2021). They conducted a meta-analysis on the association between landscape heterogeneity and levels of yields and biodiversity, finding a negative association between landscape heterogeneity and both crop yields and biodiversity indicators. Based on these results, they suggest that smaller farms provide higher area-yields and are better for the environment. The goal of chapter 8 is to verify the soundness of the empirical evidence and the validity of the implied argument.

To answer these questions, a range of empirical methods is introduced in the following section before summarizing the contribution of each chapter to answer the research questions.

1.3 Overview of Methods

Reliable evidence is crucial to inform public policy making, justify public spending or to scale up beneficial innovations. *Ex-ante* studies anticipate potential impacts and contribute to the design of policies or upscaling programs, while *ex-post* studies monitor performance. Rigorous impact evaluations have been conducted for decades in the field of development, in particular agricultural research (Banerjee and Duflo 2009) and key methodological advancements occurred in the domains of causal inference and statistical representativeness (Stevenson et al. 2023). In comparison, the field of conservation has been lagging in terms of rigorous impact evaluations, albeit past calls for more systematic evidence generation (Baylis et al. 2016; Ferraro and Pattanayak 2006; Wauchope et al. 2021). The increasing body of evidence from both fields raises questions about the reliability, generalizability, and complementarity of methods, particularly if findings contrast. Burivalova et al. (2019) highlight that there are three types of studies that can – at least in principle – estimate the causal impact between intervention exposure and outcomes: First, experimental studies such as randomized control trials (RCTs), second quasi-experimental studies (also referred to as observational studies), and third systematic reviews including meta-analyses. However, biases and limitations can still arise from sample representativeness, study design, and empirical assumptions. This thesis illustrates methodological pluralism by integrating and considering different forms of rigorous evidence motivated by their complementing degrees of validity, as described in the following.

RCTs have become a widely used method in social science over the past 20 years and gained public attention after the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel was given to Esther Duflo-Banerjee, Abhijit Banerjee, and Michael Kremer for their contributions in experimental economics (The Prize in Economic Sciences 2023). The basic principle is that a certain intervention is randomly assigned to a treatment and control group, and its attributable impact can be measured as the difference in means between treatment and control group outcomes (Angrist and Pischke 2009). If implemented correctly, RCTs provide a high level of internal validity and credibility and are particularly well-suited to assess competing behavioral assumptions or intervention mechanisms. Chapter 2 and 3 of this dissertation present two applications of randomized control trials to evaluate the impact of digital information provision. The methodological choice to conduct these studies as RCTs was based on the low anticipated costs, ethically uncritical intervention setup and in one case the timeline of the intervention which did not allow for a detailed baseline survey. RCTs are often considered to be the gold standard of evaluation, but Bédécarrats et al. (2020) critically questions this stand. Experimental studies are often prone to sample bias and provide limited external validity because implementation constraints and ethical considerations may result in deliberate study samples not representative of the wider population. Both RCTs presented here were implemented among clients of the identical social enterprise providing access to inputs and markets to small-scale farmers, pointing towards limitations with respect to sample representativeness and limits to upscaling on the one hand, but also enhanced between-study comparability on the other hand. One critical aspect of RCTs is the pronounced information- and power asymmetry between study participants on the one side and implementer and evaluators on the other side, thereby often reflecting and potentially reproducing sensitive socioeconomic gradients (Bédécarrats et al. 2020). The chapters presented here addressed such concerns by establishing prior informed consent and institutional ethical assessment of the study plan. In sum, RCTs are useful to evaluate impact channels of certain types of interventions such as the digital advisory services assessed here with high level of credibility. Yet, the focus on the identification of impacts allows only limited knowledge gains about the process of intervention take up, which may be more relevant for the aggregate policy outcome than the impact on compliers. Since practical, epistemic and ethical hurdles often restrict their applicability and generalizability of findings, observational studies and quasi-experimental designs play a crucial role for understanding and modelling population-level processes for a wide range of interventions.

The limited scale and generalizability of many experimental studies can be overcome with methods based on observational data that may be more easily and cost-efficiently scaled (Bédécarrats et al. 2020; Heckman 2020). Such studies may be feasible for a wider array of research questions, including those related to interventions that would not be feasible to randomize. Also, they can often be based on already existing secondary data such as administrative data or remote-sensing products, which often have the advantage of being more representative. This increased generalizability to entire populations makes resulting findings more relevant for policy makers if the identification strategy is convincing. Causal identification of intervention impacts can be convincingly achieved by using quasi-experimental designs on observational data (Angrist and Pischke 2009). Especially for cross-sectional studies, causality is harder to establish because it typically relies unconfoundedness conditional on observed variables. Therefore, the main disadvantage of observational studies compared to RCTs is that internal validity relies on additional and stricter assumptions.

Chapters 4 and 6 use observational datasets and quasi-experimental methods to study innovation diffusion and environmental policy impacts, respectively. Both cases leverage secondary remote sensing data to enable the analysis at national scale. Chapter 4 focusses on spatiotemporal diffusion patterns of agricultural innovation across India, using more traditional econometric techniques such as duration analysis and fixed effects regression, which allow to assess associations between contextual factors and innovation diffusion. However, policy makers are increasingly interested in causal impacts on individuals or specific subpopulations, for example for targeting. Some machine learning tools allow us to go beyond average towards individual and heterogeneous treatment effects (Athey and Imbens 2016). Therefore, chapter 6 builds upon the causal forest approach to empirically model individual counterfactual impacts of a hypothetical management intervention across Germany. Causal forests follow the principle of weighting observations by an inverse probability score, like matching. A key advantage here is the non-parametric estimation of an augmented probability score, that leverages not only linear combinations of observed variables into account, but also all their linear and non-linear interactions (Stetter et al. 2022). In doing so, it controls also for non-observed confounders that are indirectly represented in the observed variable space, and thereby provides stronger support for the conditional unconfoundedness assumption.

In sum, key strengths of observational studies lie in the wide range of addressable research questions, the possibility to rely on less costly secondary data and thereby evaluate potentially causal processes of policy relevance with higher external validity than RCTs with a comparable budget. However, the variability in secondary data collection protocols and empirical methods used to deal with the specifics of the data generating process often make it hard to disentangle whether results are driven by the study design or actual differences in underlying populations, which are themselves dynamic. Therefore, it is useful to synthesize and analyze findings of (quasi-) experimental primary studies.

The central idea of meta-analysis is to collect the relevant body of literature on a given topic in a systematic way, ensure comparability of findings and then synthesize and analyze them by means of weighted regression. Chapters 5 and 7 comprise and leverage empirical results from globally distributed primary studies in combination with spatially explicit contextual data to understand the association between geographic context and variability in the outcomes of interest. Meta-analyses depend crucially on a sizeable body of literature, so especially for very newly emerging topics this approach is often not feasible. In the case of chapter 5, we identified hundreds of studies ranging back up to fifty years, so this was no concern, but in chapter 7 fewer empirical studies had implications for the generalizability of results. A key advantage of meta-analysis is the increased external validity and clear statement of confidence in the body of evidence, which can be used to inform future research. Research and publication bias are serious concerns that can to a certain degree be identified using meta-analytic techniques, as illustrated in chapters 5, 7, and 8. Thus, meta-analytic approaches help to overcome limited generalizability of individual study findings across populations and contexts, and to assess to what extent specificities of the data and method shape results. However, their reliability crucially depends on characteristics of included primary studies and credibility of the research framing, because aggregation any synthesis by itself does not ensure the quality thereof, nor the validity of the imposed research question.

Re-analysis is a means to validate previous findings, and potentially contribute a new perspective on them. Hence, re-analysis can either strengthen credibility if results are reproducible, or contribute to the scientific debate around consensus if results or their interpretation deviate. Thereby, it is a means to address research bias, i.e., validating that the right questions are asked, appropriate methods are applied and transparently communicated. This is particularly relevant in the context of the p-value debate and replication crisis (Heckelei et al. 2021). This motivates the concluding chapter 8, which is not original research, but a re-analysis of a published article by Ricciardi et al. (2021). Along these lines, all chapters presented here contain data available statements and readers are invited to engage in replication exercises and explore the “universe of uncertainty” (Breznau et al. 2022).

1.4 Contribution

This thesis contributes to the three overarching research topics by assessing impacts of digital innovations at farm level, diffusion drivers of agricultural innovation at national and global level and disentangling contextual factors that shape the effectiveness of different environmental policies within and across countries. To do that, it leverages different data types and empirical approaches with complementing degrees of internal and external validity.

The approach, key results and overall contributions of the chapters are summarized in the following.

Research topic 1: farm-level impacts of digital innovations

Chapter 2 describes a pre-registered and ethically approved experiment among smallholder peanut farmers in the central plateau of Haiti, a region characterized by a large yield gap and political instability. The chapter contributes by increasing the evidence base on overall impacts of a real-world[2] short text-message based information intervention on knowledge, practice adoption and yields, or rather the lack thereof in the Haitian context. Contrary to previous studies (Fafchamps and Minten 2012; Larochelle et al. 2019), our findings indicate that the intervention was not effective in increasing any of the measured outcome levels, not even within subgroups or after repeating the intervention in the following year, which few studies considered previously (Oyinbo et al. 2022). Thus, we contribute by discussing the role of contextual factors in shaping our findings and discussing lessons learned for similar studies being carried out in settings with implementation constraints. We argue that short text messages provide little incremental benefits in contexts of relatively well-trained farmers, and external shocks may have restricted trust in non-personal information sources. The trade-off between using more elaborate digital communication channels that may provide personalized advice such as mobile apps and reaching marginalized groups that need such advice the most, requires further investigation.

Taking up the identified need to evaluate more elaborate information channels, chapter 3 provides complementary experimental evidence on the farm level impact of a pilot digital farm advisory service among

[2] I thank Robert Finger for pointing out this important attribute.

smallholders in El Salvador. It was implemented by the identical social enterprise as in Chapter 2 to assess the impact of an interactive chat tool providing multimedia content about pest management, thereby extending previous studies in which farmers could not request contents and content complexity was limited (160 signs). A novelty and advancement compared to the SMS intervention was the availability of administrative data that allowed to monitor treatment status, i.e., whether a participant engaged with the chat-tool and visualized the contents. In contrast to chapter 2, the intervention was more personalized and resulted in significant knowledge gains among treated participants that engaged with the tool. These findings indicate that more elaborate multimedia content provided on-demand was effective in this setting, which is consistent with other studies (Arouna et al. 2021; Giulivi et al. 2022; Rajkhowa and Qaim 2021), but had not been documented in the Latin American context.

Regarding the first research question, digital advisory services can positively impact farmers' knowledge, but their effectiveness depends on the content complexity and usefulness. Contextual factors like pre-existing knowledge levels, trust in non-personal information sources and access to digital tools influence their effectiveness.

Research topic 2: contextual patterns of innovation adoption

This research topic shifts the focus from farm level impacts within specific regions to geographically broader innovation diffusion patterns (Figure 1.1). Establishing that some digital advisory tools are effective in generating farm-level impacts on knowledge or practices is a crucial step in understanding their overall contribution to solve challenges related to sustainable production and food security. But being implemented as experiments, chapter 2 and 3 allocated the innovation randomly within a specific sample of farmers. In reality, there is a complex interplay of contextual and behavioral factors governing the adoption and diffusion of innovations (Dessart et al. 2019; Rogers 2003). Unravelling these factors for larger populations requires observing not experimentally assigned, but endogenously driven adoption patterns, which previous studies on agricultural apps did not observe at large (Bounkham et al. 2022; Michels et al. 2020; Thar et al. 2021). Chapter 4 contributes in this respect by characterizing the spatiotemporal digital innovation diffusion at country scale along with the quantification of spatially explicit drivers and barriers of adoption. Using a unique large and comprehensive but proprietary dataset of GPS-referenced and time-stamped user requests in Plantix, a plant health monitoring app, we identify a clear digital divide; specifically, that more remote regions with lower network connectivity lag in terms of adoption. Furthermore, we confirm and extend evidence on peer effects on adoption by using a weekly panel dataset covering four years to model spatiotemporal spillover effects, whereas previous studies relied on less frequent and spatially more restricted observations (Genius et al. 2014; Maertens and Barrett 2013). This high resolution enabled to estimate quickly diminishing spillover effects in our case, i.e., within just four weeks and within a radius of 50 kilometers. Finally, we illustrate a novel approach to differentiate between stationary and mobile users based on the spatial radius of their requests, contributing to the literature on network structure and spatial effects (Abdulai 2023; Krishnan and Patnam 2014). The spillover effect of stationary users – likely peers – is larger in magnitude, but the effect of mobile users – potentially extension agents - play a bigger role for adoption at the extensive margin.

To bridge the micro- and macro level of innovation diffusion, chapter 5 contributes to the generalizability of agricultural innovation adoption determinants. We overcome limitations of previous reviews in terms of innovation and geographic focus (Feder and Umali 1993; Knowler and

Bradshaw 2007; Prokopy et al. 2019; Ruzzante et al. 2021; Shang et al. 2021) by generating a unique dataset of adoption determinants and applying a theoretical framework that allows us to derive some more generalizable insights regarding their variation. This dataset is unprecedented in size and detail, containing adoption determinants for a wide range of agricultural innovations originating from over 300 primary studies across the world, and we use it to formulate reporting guidelines that aim to enhance comparability of future adoption studies. Results from our meta-analysis underline the importance of extension and access to capital as most important drivers for farm level innovation, and quantify numerous other drivers. In a second step, we formulate and apply a theoretical framework motivated by the induced innovation hypothesis, which postulates that innovation occurs to make more efficient use of the relatively more expensive production factor and was previously used to link innovation adoption to macro level context variables (Hayami and Ruttan 1971). Using country-level secondary data on the abundance of land, labor, capital and knowhow, and characterizing innovations via traits related to these four factors, we quantify the interplay of context and traits using multi-level meta-regression. By formulating the theoretical foundation and widening the geographic scope, we extend the analysis by Ruzzante et al. (2021). Our results show that farm size and capital matter more for the adoption of land- and capital intensive innovations, respectively, if these factors are less abundant in the geographic study context. These findings are in line with the induced innovation hypothesis and emphasize the role of policy to align targeting strategies to regional production contexts and user characteristics.

Regarding the second research question, contextual factors such as network infrastructure had a significant influence on the spatiotemporal diffusion of a digital agricultural advisory services Plantix across India. Spatial spillover effects were substantial, but short-lived and limited in radius. Macro-level land- and capital endowments explain between-country variation in adoption determinants of agricultural innovations that use these factors intensively.

Research topic 3: contextual impacts of policy interventions

The last research topic shifts the focus from farm level innovations to broader level policy analysis covering national to global scopes (Figure 1.1). We contribute to better understanding context-driven heterogeneity in environmental policy effectiveness. Chapter 6 provides a new methodological approach to estimate causal, spatially explicit effects of mowing frequency as proxy of management intensity on plant species richness in grasslands across Germany. A unique, remote-sensing based big dataset in combination with generalized random forests (Athey et al. 2019) enables us to assess the causal effect of changes in mowing frequency on species richness at parcel level across Germany. In line with Weber et al. (2023), we find that increases in mowing frequency cause plant species richness to decrease, but we expand this previous assessment by quantifying the spatial heterogeneity of the effect along biophysical and socioeconomic contextual gradients. In a second step, we combine these estimates with those derived from a biophysical grass growth model to quantify the changes in yields under different mowing frequencies, illustrating an integration of empirical and modelling approaches, as recently called for by Schlüter et al. (2023). Using these counterfactual estimates of plant species richness and associated changes in yields, we assess effectiveness and cost-efficiency of different conservation strategies. Based on the foregone yields associated with reduced mowing frequency, we estimate that opportunity cost per additional plant species vary by an order of magnitude across Germany, implying considerable efficiency gains from spatial policy targeting in a scenario where farmers would be compensated for reducing their management intensity. Our flexible approach could help to design and implement novel model-based biodiversity payment schemes (Bartkowski et al. 2021).

Chapter 7 assesses the evidence on the effectiveness of forest conservation programs under the umbrella of Reduced Emissions from Deforestation and Degradation (REDD+), adding to previous qualitative reviews (Turnhout et al. 2017) and meta-studies on related, but conceptually wider payments for ecosystem services (Samii et al. 2014; Wunder et al. 2020). We here systematically review and quantitatively synthesize the available rigorous impact evaluations of REDD+ initiatives and carbon-related payment for ecosystem services schemes globally. Results from our meta-analysis show that REDD+ impacts on environmental and social outcomes are small and heterogeneous, yet similar in magnitude compared to other forest conservation mechanisms. We then try to disentangle this heterogeneity using program-specific REDD+ design characteristics and regional estimates of preceding deforestation pressure as a contextual variable in a moderation analysis. Adding to previous analysis on a wider range of forest conservation interventions (Börner et al. 2020), we find that REDD+ initiatives are more effective in contexts of high deforestation pressure and where initiatives applied spatial targeting approaches. This underscores the importance of considering context-driven impact heterogeneity for future policy design.

In the last chapter 8 we take a closer look at farm structure as a moderating element of the often implied trade-offs between regulating and provisioning ecosystem services, i.e., the biodiversity-yield relationship. We were able to reproduce results from Ricciardi et al. (2021) using their provided data and methods. However, we describe and document several limitations, including the data collection process, the effect size calculation, data representativeness, the empirical estimation technique and - most crucially -

the interpretation of the results. We raise doubts about the generalizability of the findings, and add a conceptual critique independent of the soundness of empirical findings. We sustain previous studies (Garzón Delvaux et al. 2020) and the long standing theoretical debate (Helfand and Taylor 2021) that land productivity (yields) alone is not a valid performance indicator because it neglects productivity of other factors such as labor. Along these lines, we caution to draw policy conclusions based on associations rather than causal effects, implying that policy should not aim to reduce farm sizes, but rather support the mechanisms that lead to better productivity and environmental performance – independent of farm size.

Regarding the third research question, the geographic (i.e., biophysical and socioeconomic) context explains substantial variation in environmental policy effectiveness, both at plot level and country level. Environmental policies may target areas that are a-priori more sensitive to policy-induced changes to improve effectiveness. However, such policies should be based on understanding the causal relationships that drive their effectiveness, rather than taking associations as a shortcut.

1.5 Limitations and future research avenues

The two experimental studies on the first research topic are limited in their sample representativeness. Both studies recruited participants from a pool of clients of a locally active social enterprise, and results should therefore be generalized with caution. The implementation within specific contexts and subpopulations is also a common critique RCTs in general (Bédécarrats et al. 2020). Future meta studies should therefore synthesize the available evidence and quantify to what extent contextual factors *vis-à-vis* intervention characteristics matter for effectiveness. On an epistemic level, RCTs lend themselves particularly well to study impact of goods and services that are rival and excludable because they can be individually targeted. Thereby, the use of RCTs may contribute to a research bias away from public goods, and - based on their often comparably high cost associated with randomization, intervention, and data collection – contribute to suboptimal research funding allocation (Bédécarrats et al. 2020).

A key limitation of Chapters 4 and 5 is the observational nature of the underlying data and corresponding stronger assumptions required to make causal claims. Therefore, we interpret the contextual drivers and barriers of Plantix app diffusion only as associations. Similarly, although we try to control for unobserved time-variant confounders at the district level in the neighborhood analysis, endogenously driven self-selection at levels below district could affect our results, for example if local app usage results in effective pest-management measures and thus reducing the need to use the app again. More detailed farm-level data on related decisions and impacts would be required. Based on the spatially heterogeneous adoption patterns, it is likely that benefits of usage are also not homogenously distributed. This also links to a limitation of Chapter 5, in that the available adoption literature may reflect context-dependent (expected) benefits of the studied innovation. This may imply a research bias in the sense that not all innovations are implemented and their adoption studied everywhere at the same rate. The implication of non-adoption in regions were benefits would be high therefore needs to be assessed and appropriate diffusion mechanisms should be implemented to improve overall impact. This, however, requires spatially explicit knowledge about the potential impact of changing management decisions caused by innovation adoption.

For the grassland study (Chapter 6), the availability and quality of remote-sensing products should be expanded to enable applications to other conceptually relevant variable constructs of environmental indicators, i.e., beyond plant species richness and yields. The adequacy of the employed proxies of use intensity and ecosystem services depends ultimately on the application and must be carefully considered. For the REDD+ meta-analysis (Chapter 7), we can in principle claim causal effects, but must be cautious regarding its external validity. We identified potential publication bias for environmental outcomes, indicating that negative or non-significant findings may have been published less frequently. Most of the included studies focused on small scale pilot projects with a regional bias towards Latin America, so it is hard to extrapolate findings to potentially national level schemes at the global level.

The integration of feasibility- and impact assessments are a promising way to identify priority areas and improving targeting efficacy. Future research should expand the range of impact dimensions, which requires innovative data products and a closer linkage of modelling- and causal analysis (Schlüter et al. 2023). Harmonization of research protocols wherever possible is useful to compare findings across contexts and methods to increase learning potential from them (Slough and Tyson 2022). Practices like pre-registration and open science principles such as providing FAIR data access are important pillars of a transparent, replicable, and thereby credible and trustworthy research (Schwab et al. 2022). Since the available empirical methods have their merits and limitations in terms of internal and external validity, they form the basis for methodological triangulation and evidence complementarity. Especially in times of big data, statistical power should not be used exclusively as a proxy of credibility, and new methods need to be developed to aggregate, weight, and (re-)valuate evidence on additional metrics of validity and soundness.

1.6 Conclusion and Policy Implications

The research findings have implications for various stakeholders including extension service providers, developers of digital agricultural innovations, and agricultural and environmental policymakers. These implications are relevant on both local and global scales, spanning low- to high-income countries. The findings shed light on the challenges and opportunities faced by digital advisory services in generating impacts at the farm level (Chapter 2 + 3). In the context of upscaling, the research identifies factors driving and hindering the adoption of digital innovations (Chapter 4). By providing generalizable insights into spatial diffusion patterns and adoption determinants of other agricultural innovations, this research aids in targeting them more effectively (Chapter 5). Additionally, the heterogeneous impacts of different environmental policies are quantified, enabling policymakers to improve policy targeting for enhanced effectiveness and cost-efficiency (Chapter 6 + 7).

To make their services impactful, digital extension service providers should prioritize making information sources trusted and providing added value to recipients; this includes timeliness, usefulness and actionability of advice. More elaborate communication vehicles such as apps, multimedia contents and chatbots have therefore a higher potential, because more complex contents can be made available. Developers of digital advisory services should optimize content for the respective target groups, considering information needs, digital literacy, and access to different communication vehicles. Policy makers should support the development and implementation of advisory services; this support can be in terms of collaborating in data collection to identify farmer’s needs and evaluate impacts of provided services, but also marketing- and awareness campaigns as well as structurally improve access through network infrastructure and education investment. We underscore the globally high importance of extension as an innovation adoption determinant (Chapter 5) and the role of mobile Plantix users in the case of India (Chapter 4). Providers of traditional extension services and developers of (digital) innovations may therefore collaborate closer to disseminate information about innovations, because both formal and informal information networks are important to create trust and adoption of innovations. This may include public-private or private-private partnerships to improve extension quality and reach. Digital extension may supplement traditional extension where it is sufficiently effective, while they may complement each other in situations where the latter is less effective, be it due to lacking digital infrastructure or other information barriers such as literacy or trust. Thereby, the cost-efficacy and reach of overall extension system can be improved to the benefit of marginalized communities.

An overarching implication of our research is the role of context in shaping both the development of innovations and the formulation of policies. By tailoring innovations to address context-specific challenges, higher innovation adoption rates may be achieved. As illustrated by the results of Chapter 5, initiatives promoting capital-intensive innovations like agroforestry or agricultural robotics may be more effective in regions with ample access to capital. However, in contexts where capital is scarce, innovation developers should explore ways to reduce capital intensity, or policymakers should focus on improving access to capital. Similarly, spatial targeting is essential for infrastructural development and context-sensitive policy design. This research demonstrates significant potential gains in policy effectiveness and cost-efficiency when environmental policies are targeted to specific geographic areas, such as those with the highest a-priori sensitivity to policy-induced changes. Similarly, digital innovations can benefit from targeted investments in network infrastructure and education to ensure maximum impact. Yet, the spatial targeting must also consider tradeoffs in outcomes, which may include socioeconomic versus environmental gains. In such cases, compensating actors' private costs can mitigate undesirable consequences in terms of welfare or equality resulting from environmental policies. Knowing these costs requires spatially explicit modelling and innovative trade-off assessment tools, such as those presented in Chapter 6.

The context specificity of research imposes challenges for transferability of findings, but this should not be seen a reason for not trying to identify generalizable patterns. Rather, context emerges on the meta-level and helps to disambiguate findings. This dissertation illustrates how the availability of case-based literature can serve as basis for meta-analysis and secondary data in combination with tools such as machine learning facilitates better understanding of the role of context. The chapter complement each other by their varying degrees of internal and external validity, by considering both agricultural innovation and environmental policy, and by considering both feasibility and impacts of different interventions. What emerges from the use of different methods, is that they all have their merits and should not be hierarchized *a-priori*. In conclusion, understanding the role of context in determining innovation and policy impacts by employing the range of empirical tools is essential for designing and implementing effective context-sensitive policies that contribute to solving the challenges the agri-food system is confronted with.

1.8 References

Abdulai, A. (2023). Information acquisition and the adoption of improved crop varieties. *American Journal of Agricultural Economics*, *105*(4), 1049–1062.

Angrist, J. D., and Pischke, J.-S. (2009). *Mostly harmless econometrics: an empiricist's companion*. Princeton: Princeton University Press.

Arouna, A., Michler, J. D., Yergo, W. G., and Saito, K. (2021). One Size Fits All? Experimental Evidence on the Digital Delivery of Personalized Extension Advice in Nigeria. *American Journal of Agricultural Economics*, *103*(2), 596–619.

Athey, S., and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(27), 7353–7360.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, *47*(2), 1148–1178.

Banerjee, A. V., and Duflo, E. (2009). The Experimental Approach to Development Economics. *Annual Review of Economics*, *1*(1), 151–178.

Bartkowski, B., Droste, N., Ließ, M., Sidemo-Holm, W., Weller, U., and Brady, M. V. (2021). Payments by modelled results: A novel design for agri-environmental schemes. *Land Use Policy*, *102*, 105230.

Baumüller, H. (2018). The Little We Know: An Exploratory Literature Review on the Utility of Mobile Phone-Enabled Services for Smallholder Farmers. *Journal of International Development*, *30*(1), 134–154.

Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P. J., … Wunder, S. (2016). Mainstreaming Impact

Evaluation in Nature Conservation: Mainstreaming impact evaluation. *Conservation Letters*, *9*(1), 58–64.

Bédécarrats, F., Guérin, I., and Roubaud, F. (Eds.). (2020). *Randomized control trials in the field of development: a critical perspective*. Oxford: Oxford University Press.

Binswanger, H. P. (1974). A microeconomic approach to induced innovation. *The Economic Journal*, *84*(336), 940–958.

Börner, J., Schulz, D., Wunder, S., and Pfaff, A. (2020). The Effectiveness of Forest Conservation Policies and Programs. *Annual Review of Resource Economics*, *12*(1), 45–64.

Bounkham, P., Ahmad, M. M., and Yaseen, M. (2022). Determinants and benefits of using smartphone-based information for vegetable production and marketing: a case of Lao farmers. *Development in Practice*, *32*(7), 981–994.

Breznau, N., Rinke, E. M., Wuttke, A., Nguyen, H. H. V., Adem, M., Adriaans, J., … Żółtak, T. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(44), e2203150119.

Burivalova, Z., Miteva, D., Salafsky, N., Butler, R. A., and Wilcove, D. S. (2019). Evidence Types and Trends in Tropical Forest Conservation Literature. *Trends in Ecology & Evolution*, *34*(7), 669–679.

Camacho, A., and Conover, E. (2019). The impact of receiving SMS price and weather information on small scale farmers in Colombia. *World Development*, *123*(3), 104596.

Campbell, B. M., Beare, D. J., Bennett, E. M., Hall-Spencer, J. M., Ingram, J. S. I., Jaramillo, F., … Shindell, D. (2017). Agriculture production as a major driver of the Earth system exceeding planetary boundaries. *Ecology and Society*, *22*(4), art8.

Cord, A. F., Bartkowski, B., Beckmann, M., Dittrich, A., Hermans-Neumann, K., Kaim, A., … Volk, M. (2017). Towards systematic analyses of ecosystem service trade-offs and synergies: Main concepts, methods and the road ahead. *Ecosystem Services*, *28*, 264–272.

Costa, C., Wollenberg, E., Benitez, M., Newman, R., Gardner, N., and Bellone, F. (2022). Roadmap for achieving net-zero emissions in global food systems by 2050. *Scientific Reports*, *12*(1), 15064.

Dessart, F. J., Barreiro-Hurlé, J., and van Bavel, R. (2019). Behavioural factors affecting the adoption of sustainable farming practices: a policy-oriented review. *European Review of Agricultural Economics*, *46*(3), 417–471.

Ehlers, M.-H., Huber, R., and Finger, R. (2021). Agricultural policy in the era of digitalisation. *Food Policy*, *100*, 102019.

Fabregas, R., Kremer, M., and Schilbach, F. (2019). Realizing the potential of digital development: The case of agricultural advice. *Science (New York, N.Y.)*, *366*(6471).

Fafchamps, M., and Minten, B. (2012). Impact of SMS-Based Agricultural Information on Indian Farmers. *The World Bank Economic Review*, *26*(3), 383–414.

Feder, G., and Umali, D. L. (1993). The Adoption of Agricultural Innovations: A Review. *Technological Forecasting and Social Change*, *43*, 215–239.

Fermeglia, M. (2023). The Nitrates Directive under the European Green Deal: time to deliver! *ERA Forum*, *24*(2), 169–182.

Ferraro, P. J., and Pattanayak, S. K. (2006). Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments. *PLoS Biology*, *4*(4), e105.

Finger, R. (2023). Digital innovations for sustainable and resilient agricultural systems. *European Review of Agricultural Economics*, *50*(4), 1277–1309.

Garzón Delvaux, P. A., Riesgo, L., and Gomez y Paloma, S. (2020). Are small farms more performant than larger ones in developing countries? *Science Advances*, *6*(41), eabb8235.

Genius, M., Koundouri, P., Nauges, C., and Tzouvelekas, V. (2014). Information Transmission in Irrigation Technology Adoption and Diffusion: Social Learning, Extension Services, and Spatial Effects. *American Journal of Agricultural Economics*, *96*(1), 328–344.

Giulivi, N., Harou, A. P., Gautam, S., and Guereña, D. (2022). Getting the message out: Information and communication technologies and agricultural extension. *American Journal of Agricultural Economics*, ajae.12348.

Hayami, Y., and Ruttan, V. W. (1971). *Agricultural development: An international perspective*. Baltimore: Johns Hopkins University Press.

Heckelei, T., Huettel, S., Odening, M., Rommel, J., Heckelei, T., Huettel, S., … Rommel, J. (2021). *The replicability crisis and the p-value debate – what are the consequences for the agricultural and food economics community?* (Discussion Paper No. 2021:2). Bonn: Institute for Food and Resource Economics. Retrieved from https://ageconsearch.umn.edu/record/316369

Heckman, J. J. (2020). Epilogue: Randomization and Social Policy Evaluation Revisited. In F. Bédécarrats, I. Guérin, and F. Roubaud, editors, *Randomized Control Trials in the Field of Development*, pages 304–330. Oxford University PressOxford.

Helfand, S. M., and Taylor, M. P. H. (2021). The inverse relationship between farm size and productivity: Refocusing the debate. *Food Policy*, *99*, 101977.

Hendriks, S. L., Montgomery, H., Benton, T., Badiane, O., Castro De La Mata, G., Fanzo, J., … Soussana, J.-F. (2022). Global environmental climate change, covid-19, and conflict threaten food security and nutrition. *BMJ*, e071534.

Herrero, M., Thornton, P. K., Mason-D'Croz, D., Palmer, J., Benton, T. G., Bodirsky, B. L., … West, P. C. (2020). Innovation can accelerate the transition towards a sustainable food system. *Nature Food*, *1*(5), 266–272.

Hicks, J. R. (1932). *The Theory of Wages*. London: Palgrave Macmillan UK.

IPCC. (2022). *Climate Change and Land: IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems*. Cambridge University Press.

Isselstein, J., Jeangros, B., and Pavlů, V. V. (2005). Agronomic aspects of biodiversity targeted management of temperate grasslands in Europe - A review. *Agronomy Research*, *3*, 139–151.

Jain, M., Balwinder-Singh, Rao, P., Srivastava, A. K., Poonia, S., Blesh, J., … Lobell, D. B. (2019). The impact of agricultural interventions can be doubled by using satellite data. *Nature Sustainability*, *2*(10), 931–934.

Kehoe, L., Romero-Muñoz, A., Polaina, E., Estes, L., Kreft, H., and Kuemmerle, T. (2017). Biodiversity at risk under future cropland expansion and intensification. *Nature Ecology & Evolution*, *1*(8), 1129–1135.

Knowler, D., and Bradshaw, B. (2007). Farmers’ adoption of conservation agriculture: A review and synthesis of recent research. *Food Policy*, *32*(1), 25–48.

Krishnan, P., and Patnam, M. (2014). Neighbors and Extension Agents in Ethiopia: Who Matters More for Technology Adoption? *American Journal of Agricultural Economics*, *96*(1), 308–327.

Larochelle, C., Alwang, J., Travis, E., Barrera, V. H., and Dominguez Andrade, J. M. (2019). Did You Really Get the Message? Using Text Reminders to Stimulate Adoption of Agricultural Technologies. *The Journal of Development Studies*, *55*(4), 548–564.

Maertens, A., and Barrett, C. B. (2013). Measuring Social Networks’ Effects on Agricultural Technology Adoption. *American Journal of Agricultural Economics*, *95*(2), 353–359.

Mehrabi, Z., McDowell, M. J., Ricciardi, V., Levers, C., Martinez, J. D., Mehrabi, N., … Jarvis, A. (2021). The global divide in data-driven farming. *Nature Sustainability*, *4*(2), 154–160.

Michels, M., Bonke, V., and Musshoff, O. (2020). Understanding the adoption of smartphone apps in crop protection. *Precision Agriculture*, *21*(6), 1209–1226.

Nakasone, E., and Torero, M. (2016). A text message away: ICTs as a tool to improve food security. *Agricultural Economics*, *47*(S1), 49–59.

Newbold, T., Hudson, L. N., Hill, S. L. L., Contu, S., Lysenko, I., Senior, R. A., … Purvis, A. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, *520*(7545), 45–50.

Ogundari, K., and Bolarinwa, O. D. (2018). Impact of agricultural innovation adoption: a meta-analysis. *Australian Journal of Agricultural and Resource Economics*, *62*(2), 217–236.

Oyinbo, O., Chamberlin, J., Abdoulaye, T., and Maertens, M. (2022). Digital extension, price risk, and farm performance: experimental evidence from Nigeria. *American Journal of Agricultural Economics*, *104*(2), 831–852.

Pendrill, F., Gardner, T. A., Meyfroidt, P., Persson, U. M., Adams, J., Azevedo, T., … West, C. (2022). Disentangling the numbers behind agriculture-driven tropical deforestation. *Science*, *377*(6611), eabm9267.

Poore, J., and Nemecek, T. (2018). Reducing food’s environmental impacts through producers and consumers. *Science*, *360*(6392), 987–992.

Pradhan, P., Costa, L., Rybski, D., Lucht, W., and Kropp, J. P. (2017). A Systematic Study of Sustainable Development Goal (SDG) Interactions. *Earth’s Future*, *5*(11), 1169–1179.

Prokopy, L. S., Floress, K. M., Arbuckle, J. G., Church, S. P., Eanes, F. R., Gao, Y., … Singh, A. S. (2019). Adoption of agricultural conservation practices in the United States: Evidence from 35 years of quantitative literature. *Journal of Soil and Water Conservation*, *74*(5), 520–534.

Rajkhowa, P., and Baumüller, H. (2024). Assessing the potential of ICT to increase land and labour productivity in agriculture: Global and regional perspectives. *Journal of Agricultural Economics*, 1477-9552.12566.

Rajkhowa, P., and Qaim, M. (2021). Personalized digital extension services and agricultural performance: Evidence from smallholder farmers in India. *PloS One*, *16*(10), e0259319.

Ricciardi, V., Mehrabi, Z., Wittman, H., James, D., and Ramankutty, N. (2021). Higher yields and more biodiversity on smaller farms. *Nature Sustainability*, *4*(7), 651–657.

Robertson, G. P., and Vitousek, P. M. (2009). Nitrogen in Agriculture: Balancing the Cost of an Essential Resource. *Annual Review of Environment and Resources*, *34*(1), 97–125.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., … Foley, J. A. (2009). A safe operating space for humanity. *Nature*, *461*(7263), 472–475.

Rogers, E. M. (2003). *Diffusion of innovations*. New York: Free Press.

Ruzzante, S. W., Labarta, R., and Bilton, A. (2021). Adoption of agricultural technology in the developing world: A meta-analysis of the empirical literature. *World Development*, *146*(1), 105599.

Samii, C., Lisiecki, M., Kulkarni, P., Paler, L., Chavis, L., Snilstveit, B., … Gallagher, E. (2014). Effects of Payment for Environmental Services (PES) on Deforestation and Poverty in Low and Middle Income Countries: A Systematic Review. *Campbell Systematic Reviews*, *10*(1), 1–95.

Schlüter, M., Brelsford, C., Ferraro, P. J., Orach, K., Qiu, M., and Smith, M. D. (2023). Unraveling complex causal processes that affect sustainability requires more integration between empirical and modeling approaches. *Proceedings of the National Academy of Sciences*, *120*(41), e2215676120.

Schwab, S., Janiaud, P., Dayan, M., Amrhein, V., Panczak, R., Palagi, P. M., … Held, L. (2022). Ten simple rules for good research practice. *PLOS Computational Biology*, *18*(6), e1010139.

Shang, L., Heckelei, T., Gerullis, M. K., Börner, J., and Rasch, S. (2021). Adoption and diffusion of digital farming technologies - integrating farm-level evidence and system interaction. *Agricultural Systems*, *190*(1), 103074.

Simonet, G., Agrawal, A., Bénédet, F., Cromberg, M., de Perthuis, C., Haggard, D., … Vaillant, G. (2018). ID-RECCO, International Database on REDD+ projects and programs, linking Economic, Carbon and Communities data (Version 3.0) [Data set].

Slough, T., and Tyson, S. A. (2022). External Validity and Meta-Analysis. *American Journal of Political Science*, ajps.12742.

Spielman, D., Lecoutere, E., Makhija, S., and van Campenhout, B. (2021). Information and Communications Technology (ICT) and Agricultural Extension in Developing Countries. *Annual Review of Resource Economics*, *13*(1).

Stetter, C., Mennig, P., and Sauer, J. (2022). Using Machine Learning to Identify Heterogeneous Impacts of Agri-Environment Schemes in the EU: A Case Study. *European Review of Agricultural Economics*, *49*(4), 723–759.

Stevenson, J. R., Macours, K., and Gollin, D. (2023). The Rigor Revolution: New Standards of Evidence for Impact Assessment of International Agricultural Research. *Annual Review of Resource Economics*, *15*(1), 495–515.

Tambo, J. A., Aliamo, C., Davis, T., Mugambi, I., Romney, D., Onyango, D. O., … Byantwale, S. T. (2019). The impact of ICT-enabled extension campaign on farmers' knowledge and management of fall armyworm in Uganda. *PloS One*, *14*(8), e0220844.

Tang, F. H. M., Lenzen, M., McBratney, A., and Maggi, F. (2021). Risk of pesticide pollution at the global scale. *Nature Geoscience*, *14*(4), 206–210.

Thar, S. P., Ramilan, T., Farquharson, R. J., Pang, A., and Chen, D. (2021). An empirical analysis of the use of agricultural mobile applications among smallholder farmers in Myanmar. *The Electronic Journal of Information Systems in Developing Countries*, *87*(2), e12159.

The Prize in Economic Sciences. (2023). *NobelPrize.org* (Nobel Prize Outreach AB). Retrieved from https://www.nobelprize.org/prizes/economic-sciences/2019/summary

Tubiello, F. N., Karl, K., Flammini, A., Gütschow, J., Obli-Laryea, G., Conchedda, G., … Torero, M. (2022). Pre- and post-production processes increasingly dominate greenhouse gas emissions from agri-food systems. *Earth System Science Data*, *14*(4), 1795–1809.

Turnhout, E., Gupta, A., Weatherley-Singh, J., Vijge, M. J., De Koning, J., Visseren-Hamakers, I. J., … Lederer, M. (2017). Envisioning REDD+ in a post-Paris era: between evolving expectations and current practice. *WIREs Climate Change*, *8*(1), e425.

United Nations. (2015). Transforming our World: The 2030 Agenda for Sustainable Development.

van Campenhout, B., Spielman, D. J., and Lecoutere, E. (2021). Information and Communication Technologies to Provide Agricultural Advice to Smallholder Farmers: Experimental Evidence from Uganda. *American Journal of Agricultural Economics*, *103*(1), 317–337.

von Braun, J., Afsana, K., Fresco, L. O., and Hassan, M. (2021). Food systems: seven priorities to end hunger and protect the planet. *Nature*, *597*(7874), 28–30.

Wauchope, H. S., Amano, T., Geldmann, J., Johnston, A., Simmons, B. I., Sutherland, W. J., and Jones, J. P. G. (2021). Evaluating Impact Using Time-Series Data. *Trends in Ecology & Evolution*, *36*(3), 196–205.

Weber, D., Schwieder, M., Ritter, L., Koch, T., Psomas, A., Huber, N., … Boch, S. (2023). Grassland-use intensity maps for Switzerland based on satellite time series: Challenges and opportunities for ecological applications. *Remote Sensing in Ecology and Conservation*, rse2.372.

Wuepper, D., Borrelli, P., and Finger, R. (2019). Countries and the global rate of soil erosion. *Nature Sustainability*, *3*(1), 51–55.

Wuepper, D., Crowther, T., Lauber, T., Routh, D., Le Clec'h, S., Garrett, R. D., and Börner, J. (2024). Public policies and global forest conservation: Empirical evidence from national borders. *Global Environmental Change*, *84*, 102770.

Wuepper, D., Tang, F. H. M., and Finger, R. (2023). National leverage points to reduce global pesticide pollution. *Global Environmental Change*, *78*, 102631.

Wunder, S., Börner, J., Ezzine-de-Blas, D., Feder, S., and Pagiola, S. (2020). Payments for Environmental Services: Past Performance and Pending Potentials. *Annual Review of Resource Economics*, 12(1), 209–234.

Chapter 2
No impact of repeated digital advisory service to Haitian peanut producers[3]

2.1 Introduction

Digital technologies are considered an important pillar of agricultural transformation towards food security and sustainability (von Braun et al. 2021). Smallholder farmers in developing countries often take production decisions based on limited information, which dampens productivity and farm-household incomes. Information and Communication Technologies (ICT) in particular have received increasing attention for their potential to disseminate information to farmers in developing countries (Baumüller 2018; Spielman et al. 2021). ICTs are expected to support farmers on different levels, from digital platforms for input procurement and capital access to digital advisory services and marketing channels. Agricultural advisory services for farmers rely on diverse communication channels

[3] This chapter is published as: Schulz, D., and Börner, J. (2023). No impact of repeated digital advisory service to Haitian peanut producers. *Q Open*, qoad023. DOI: 10.1093/qopen/qoad023

Author contribution: DS: conceptualization, data collection and curation, formal analysis, writing original draft; JB: conceptualization, funding acquisition, writing - review and editing the paper.

including videos, interactive voice recordings, smartphone apps for extension agents, and short text messages (SMS). The emerging empirical evidence across different contexts and ICTs indicate overall promising positive impacts on recommended practice adoption and yields (Fabregas et al. 2019).

Several studies have tried to identify mechanisms through which ICT-enabled information provision can lead to productivity- and livelihood improvements. According to the common theory of change, sub-optimal input use decisions result from a lack of knowledge or attention to employ it. Sending SMS to farmers has been shown to increase their knowledge and adoption of recommended practices (Camacho & Conover 2019; Carrión-Yaguana et al. 2020; Ding et al. 2022; Larochelle et al. 2019). Similarly, knowledge and adoption were shown to increase when exposing farmers to extension videos, or a combination of video and SMS (Maredia et al. 2018; Tambo et al. 2019; Vandevelde et al. 2021). Going beyond adoption of practices, van Campenhout et al. (2021) found a positive impact of a video-based training on yields, but no clear evidence that SMS-nudges or interactive voice recordings further added to that effect. Finally, several studies investigated the effectiveness of smartphone applications for personalized agricultural advice, such as in nutrient management or input procurement, finding sizable positive impacts on yields and profits (Arouna et al. 2021; Oyinbo et al. 2022; Rajkhowa & Qaim 2021). While such apps can provide tailored contents and enable bidirectional information flows, they are also less accessible than direct messaging and require prior training, a smartphone, and eventually also an internet connection. This makes such services costlier and less inclusive than unidirectional information provision, especially in areas with limited internet connectivity, smartphone adoption, and digital literacy. That is why app-based services are rarely provided to farmers directly, but instead used by extension agents or local group leaders as intermediaries.

The overall impacts of digital information provision are well-documented, but there is limited evidence on at least four related aspects. First, information spillovers received limited attention, with notable exceptions such as Vandevelde et al. (2021). Such spillovers may occur when digital information is shared or neighbouring peers learn about recommended practices by observation. Second, most previous studies looked at single intervention periods, but Oyinbo et al. (2022) reported that the impact decreased from the first to the second year. This relates to the question of permanence, i.e. the stability of the effect of an intervention over time. Furthermore, it remains important to understand whether a positive impact can be maintained or even increased by repeating an intervention in consecutive seasons. Third, little is known about whether ICT-enabled information delivery works in isolation or only in combination with other support mechanisms. In particular, whether digital advice can or should replace traditional extension service, or rather complement it (Maredia et al. 2018). Most previous studies were performed in a setting where farmers had homogeneous access to additional support, such as extension service or group training (e.g. Larochelle et al. 2019). Relatedly, information provision may not produce the desired changes in agricultural practices if the required inputs are not accessible (Jack 2011). Hence, it remains to be understood how reliably ICT works along a gradient of input- and market access. Fourth, some studies suggest substantial impact heterogeneity across socially sensitive gradients such as farm size (Ding et al. 2022; Mehrabi et al. 2021) and education (Carrión-Yaguana et al. 2020).

This research aimed at addressing these knowledge gaps by expanding the experimental evidence on SMS-based information provision to Haiti, a challenging environment for both agricultural production and extension. Our study contributes in two ways. First, we examine the relationship between digital information provision and changes in agricultural practices via knowledge gain and changes in farming income via increased yields. By harmonizing our intervention and outcome measures for comparability with existing studies, we also enhance the generalizability of research in this topic. Second, we aimed to enrich our understanding of the causal mechanisms behind ICT impacts by focusing on spatial and temporal spillovers and heterogeneous impacts by considering them in our research design (see Section 2.4). Several implementation constraints implied limitations in our ability to do so and we hope to contribute to future research by systematically highlighting both constraints and lessons learned throughout the manuscript. To conduct this study, we identified a local enterprise that was planning to launch an ICT-based extension service and was willing to do it in in a randomized fashion. Notably, the intervention was not initiated by researchers and thus reflects a real-world setup. Agronomic and organizational experience of local partner organization and especially existing survey infrastructure can be a great advantage; in our case it allowed this research to be planned and implemented in a few months. On the other hand, such existing organizational structures may be less flexible in certain aspects and require some compromises. In our case this implied, among other things, to accept a) the unavailability of baseline data, b) limited control over timing and content of messages, and c) limited sample representativeness of the organization's clients. While some shortcomings are known *ex-ante*, others may appear as violations of implicit assumptions *ex-post*. For example, in the absence of baseline data group imbalances and partial non-random treatment allocation are notoriously hard to detect at an early stage. Since this study is about ICTs, it is worth mentioning that this entire research was conducted during the COVID-19 pandemic and all communication occurred digitally. This made it particularly challenging in the context of regional and temporal connectivity gaps, because conducting an experiment requires substantial transfer of knowledge and information between researcher and implementing organization. Further constraints particular to politically unstable contexts such as Haiti included temporal unavailability of gasoline (needed by enumerators to visit farmers), and security concerns that impeded field visits by the investigators to conduct focus group discussions. More details and some implications of these challenges are discussed in section 2.4.

Our results showed that the SMS-based intervention was ineffective in improving any of the measured outcomes. Given these unexpected results, we repeated the intervention in a slightly modified manner, which enabled us to vary the treatment intensity and consider effect permanence. Again, we found that the intervention was overall not effective, with no evidence of subgroup differences. This paper is structured as follows. In Section 2.2 we describe the context of our study as well as the ICT-intervention design and empirical framework. Section 2.3 presents our results, which are discussed in the light of internal and external validity in Section 2.4. Finally, we conclude in Section 2.5.

2.2 Material and Methods

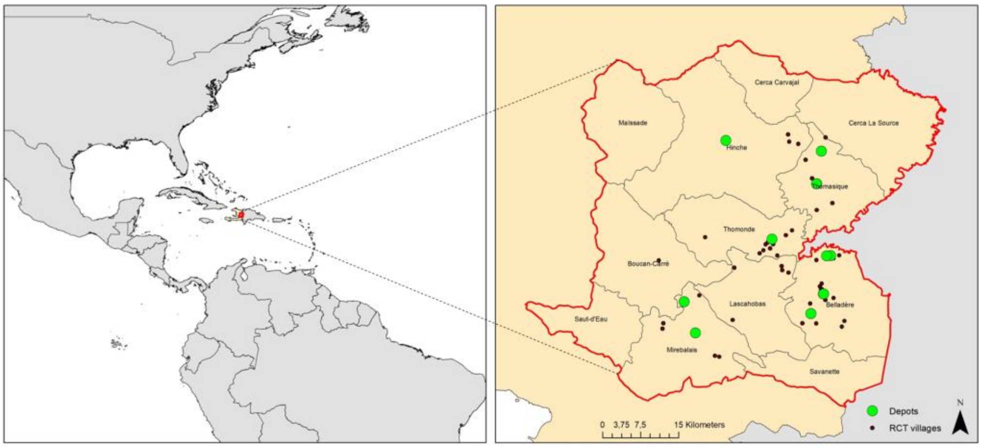
This study was pre-registered at the Open Science Foundation and approved by the Research Ethics Committee of the Center for Development Research (ZEF), University of Bonn[4]. All study participants orally gave prior informed consent to partake in the study.

[4] Pre-analysis plan for this study: https://osf.io/cazqw.

2.2.1 Study context

Figure 2.1 shows our study region: the central plateau, one of the main areas for peanut production in Haiti (Fulmer 2018). This region has two seasons, the first starting in March, the second around August. Typically, farmers plant peanuts only in the first season and another crop such as beans during the second, although two peanut seasons are also common. At an estimated 300-900 kg/ha, average yields remain far below the reported 3-4 tons/ha achieved during agronomic trials in the same region (Fulmer et al. 2020; Kostandini et al. 2021; Tyroler 2018). The main factors responsible for these low yields include soil- and seed quality, management practices related to planting, weed- and pest control, improper storage methods that do not protect from rodents and fungi, floods, droughts and theft (Fulmer et al. 2020; ICF International 2013). We conducted the experiment among smallholder peanut farmers that work with a local social enterprise (SE). The local partner has been active in Haiti since 2015 and provides agricultural extension services as well as input-, credit-, and market access to smallholders as part of a contract farming scheme. Our study aligns with efforts along the peanut value chain by the SE aiming to overcome several structural barriers by providing support along all production stages from field preparation to selling. In particular, the SE conducts physical farmer group trainings and offers a range of services such as provision of seeds, fertilizer, pesticide application, management advice and market access as part of a contract farming scheme. Such schemes have been shown to improve input usage and yields (Ruml & Qaim 2020). The location of depots where farmers can purchase inputs or sell their produce is indicated by the green dots in Figure 2.1 on the right side.

Figure 2.1: Study region in Haiti



Note: *Red line indicates central plateau region in Haiti. Black dots are villages participating in the experiment, green dots indicate location of depots of the social enterprise where farmers can get inputs and sell produce.*

Around 3.000 farmers attend the seasonal group trainings every year, while due to organizational constraints only about 300 are offered the farming contract scheme. The contract is conditional on stated interest by the farmers and observed farm characteristics that are conducive to peanut production. However, all participants of the initial group trainings could state their interest to receive information via SMS and thereby formed the pool of eligible farmers for receiving the GAP-intervention. Independent of the intervention, all interested farmers received weather warnings on their phones. The delivery status of these messages was used to verify that farmers in the control group would have been technically able to receive the content.

2.2.2 GAP intervention design

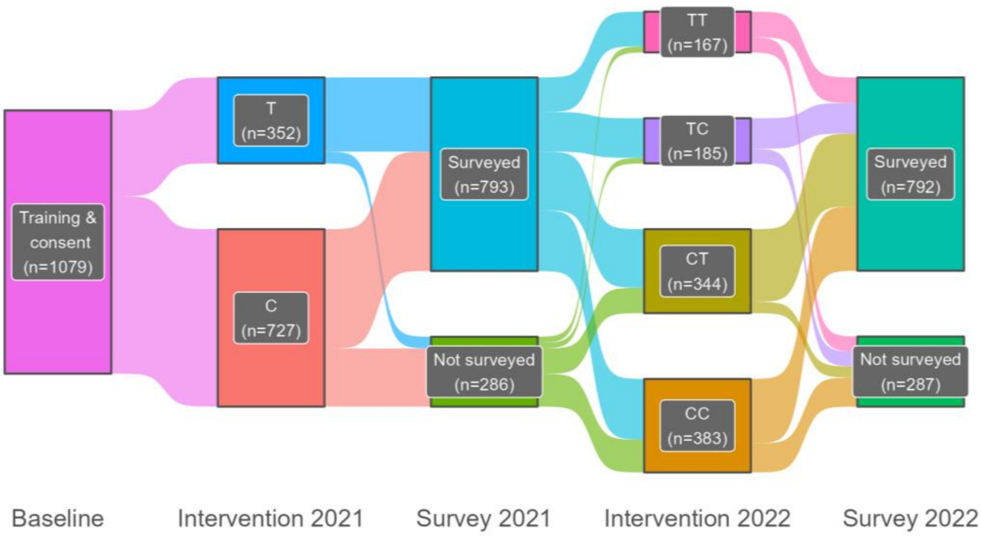
We used a random phase-in experimental design to study the impact of providing information about good agricultural practices (GAP) via SMS (Duflo et al. 2006). Our theory of change is presented in the supplementary material (see also Figure S7). All GAP-contents were developed and checked by local agronomists and then sent to a randomly selected treatment group in Haitian Creole (TechnoServe 2014). A complete overview of all messages with English translation is presented in Table S2. To select our sample, we used the database of farmers that participated in any group training on peanut cultivation conducted by the SE in 2021. Therefore, while only (self-)selected farmers became contract farmers, our ICT-based information intervention also targeted farmers not participating in the contract scheme and thus allowed us to study the complementarity of physical and digital information provision under varying sets of market constraints in the heterogeneity analysis. The intervention occurred in two separate waves, namely between April and August of the years 2021 and 2022, but with slight differences between the two years. Figure 2.2 summarizes the intervention and sampling design in a timely order with the respective group sizes in form of a Sankey diagram. In 2021, the treatment was randomized in two steps with the intention to identify spillover effects. In the first step, the total of 46 available villages were randomly assigned as treatment or control with equal probability. We chose to randomize at the village level in order to reduce information spillovers among farmers within the same village. However, due to the low number of available villages (46), we chose to employ non-bipartite covariate matching (Imai et al. 2009) to ensure that key village-level characteristics (such as distance to major roads) were balanced across both groups. In the second step, we randomly assigned treatment intensities among the treated villages. With equal probabilities, the share of known farmers in a village that would receive GAP-content could be 15%, 30%, 60% or 100%. The intention was to create a gradient in the probability of spillovers: the more farmers in a given neighborhood receive GAP-content, the higher the likelihood that one of them would share that information with a non-receiving farmer. We planned to detect spillover effects on the individual level among farmers in treatment villages that were randomly assigned to not receive the treatment using the village-level treatment share as a continuous treatment indicator. Therefore, we also surveyed farmers in treatment villages that did not receive the GAP-messages to measure potential spillovers. In the absence of a measurable impact, we decided in 2022, to randomize the treatment on the individual level but within strata of the 2021 treatment status. That is, we randomly assigned 50% of the treated farmers and 50% of the control farmers from 2021 to treatment and control, respectively. This resulted in four distinct treatment arms: 1) TT received treatment in both years; 2) TC received treatment in 2021 but not in 2022; 3) CT received treatment in 2022 but not in 2021; and 4) CC did not receive treatment in either year. The reason for randomization at the individual level in 2022 were concerns about lack of statistical power under a cluster randomization with four de-facto treatment arms.

2.2.3 Sampling design and sample characteristics

Before treatment assignment in 2021, we conducted power calculations to inform the required sample size (see pre-registration for details), and employed a truncated proportional random sampling approach to account for vastly different village sizes (min=1, max=433). Our intention was to limit an expected high intra-cluster correlation in proportional samples from large villages by imposing an upper limit. Eligible farmers must be reachable via phone, grow peanuts, and have attended at least one group event organized by the social enterprise. Based on this approach, we anticipated a total sample size of 1001 farmers, but only obtained consent and follow-up data from 933. A phone-based baseline survey was conducted in May 2021 to obtain prior informed consent by all participants and verify their eligibility. A follow-up survey was conducted during physical visits at the farms during August and September of the years 2021 and 2022, respectively. Both treatment and the follow-up survey were organized by the social enterprise.

Farmers were rewarded with a small token of gratitude after finishing the follow-up surveys, because initial tests showed that it would lead to a higher survey completion rate. Nevertheless, for several reasons, it was not possible to visit the exact same farmers in both years. Known reasons include death, migration or repeated non-presence of farmers, constrained mobility of local supervisors and enumerators due to temporal unavailability of gasoline, lack of internet connectivity in the field to supervise the sampling remotely and a short time-window imposed by budget constraints. Therefore, we have an unbalanced panel dataset, where 592 farmers were visited in both years, and an additional 199 and 198 farmers were interviewed in 2021 and 2022, respectively. Figure 2.2 also shows the number of surveyed farmers in each group and year.

Figure 2.2: SMS intervention and sampling overview



Note*: Sankey diagram showing group sizes and survey status for 2021 and 2022. T indicates treatment, i.e. receival of SMS with good agricultural practice recommendations and weather warnings, C indicates control, i.e., receival of weather warnings only. TC indicates treated in 2021 and control in 2022, whereas CT indicates Control in 2021 and Treatment in 2022. CC and TT indicate consecutive control and treatment, respectively. The 2021* *survey contained questions about household variables prior to the intervention, which we used to assess balance. For example: "What kind of roof material did your house have five years ago?".*

2.2.3.1 Independent variables

For this analysis, our treatment indicator is whether GAP content for each production stage was delivered to a farmer's phone number as indicated by administrative records. We collected control variables relating to farmer and farm characteristics, and show their summary statistics for both years and by treatment status in Table 2.1. However, control variables in the 2021 survey are based on five-year recall questions and thus refer to the pre-treatment status. This was done to assess covariate balance in the absence of a complete baseline survey. There are several missing control variables in both groups and years as indicated by the number of "Unknown". As specified in our pre-analysis plan, we excluded these observations from the analysis, but provide results with imputed (mean) values as a robustness check in the supplementary information (Figure S5). Columns 4 and 7 of Table 2.1 indicates whether significant differences between treatment and control group were detected via Wilcoxon rank sum test (for continuous variables) and Pearson's Chi-squared test (for binary variables). Notably, covariates are only balanced for the year 2022, but not prior to the intervention (data collected in 2021 referred to 2016). Respondents in the treatment group in 2021 were significantly closer to major roads and cities and had significantly more ICT-relevant assets (TV, electricity) than the control group. The fact that these imbalances exist in both time-variant and time-invariant variables implies that our randomization was not fully successful in the first year, which we attribute to deviations from our study design (Section 2.5). The variable "priority village" indicates that respondents were located in one of the villages in which the SE reported relatively intense activity and positive feedback during previous years.

Table 2.1: Control variable summary

Characteristic	**2021 Sample**			**2022 Sample**		
	Control N = 487[1]	**Treated N = 304**[1]	**p-value**[2]	**Control N = 388**[1]	**Treated N = 402**[1]	**p-value**[2]
Age (years)	47.41 (13.34)	46.32 (13.46)	0.18	47.33 (13.71)	48.21 (13.01)	0.28
Unknown	13	5		7	13	
Gender (1=female)	172 (35%)	112 (37%)	0.66	126 (32%)	135 (34%)	0.74
Literate (1=yes)	-	-		301 (78%)	316 (79%)	0.73
Unknown	487	304				
years of schooling	-	-		6.94 (4.95)	7.20 (4.89)	0.60
Unknown	487	304				
Number of rooms	2.78 (0.81)	2.86 (0.85)	0.42	3.20 (1.11)	3.13 (1.08)	0.33
Material roof	16 (3.3%)	13 (4.3%)	0.47	22 (5.7%)	14 (3.5%)	0.14
Material floor	1.67 (0.84)	1.95 (0.81)	<0.001	1.72 (0.83)	1.71 (0.85)	0.74
Material wall	2.28 (0.45)	2.46 (0.50)	<0.001	2.32 (0.47)	2.31 (0.46)	0.83
Draft animals (#)	1.37 (1.30)	1.24 (1.32)	0.12	1.46 (1.32)	1.53 (1.52)	0.96
Owns plough (1=yes)	110 (23%)	52 (17%)	0.063	113 (29%)	119 (30%)	0.88
Crops cultivated (#)	2.57 (1.41)	2.47 (1.14)	0.89	2.53 (1.36)	2.48 (0.97)	0.89
Sanitary facility quality	2.30 (0.97)	2.17 (0.85)	0.22	2.25 (0.80)	2.17 (0.79)	0.15
Water access	2.40 (0.50)	2.51 (0.51)	0.005	2.43 (0.53)	2.43 (0.52)	>0.99
Charcoal stove (1=yes)	163 (33%)	188 (62%)	<0.001	103 (27%)	115 (29%)	0.52
Has electricity (1=yes)	80 (16%)	110 (36%)	<0.001	91 (23%)	78 (19%)	0.17
Owns vehicle (1=yes)	84 (17%)	60 (20%)	0.38	56 (14%)	65 (16%)	0.50
Owns TV (1=yes)	39 (8.0%)	69 (23%)	<0.001	46 (12%)	48 (12%)	0.97
Has off-farm job (1=yes)	90 (18%)	47 (15%)	0.27	80 (21%)	82 (20%)	0.94
Income farming (HTG)	10,747.01 (16,017.35)	9,465.59 (9,586.17)	0.62	4,987.50 (6,316.12)	5,182.32 (5,861.99)	0.41
Unknown	290	189		85	82	

Characteristic	2021 Sample			2022 Sample		
	Control N = 487[1]	Treated N = 304[1]	p-value[2]	Control N = 388[1]	Treated N = 402[1]	p-value[2]
Income total (HTG)	20,110.30 (26,023.87)	14,253.68 (23,067.15)	0.072	10,040.98 (9,012.14)	10,429.34 (10,913.08)	0.86
Unknown	299	192		85	82	
Distance any road (km)	0.41 (0.66)	0.38 (0.50)	0.055	0.42 (0.64)	0.35 (0.60)	0.050
Unknown	13	9		51	110	
Travel time (min)	38.45 (22.61)	29.65 (26.45)	<0.001	35.03 (22.63)	37.17 (23.02)	0.29
Unknown	13	9		51	110	
Total farm size (ha)	1.62 (1.29)	1.70 (1.85)	0.73	1.18 (1.00)	1.33 (1.22)	0.26
Unknown	88	68				
Cultivated peanut area (ha)	0.87 (0.72)	0.75 (0.60)	0.002	0.68 (0.48)	0.71 (0.50)	0.34
Contract 2020 S1 (1=yes)	47 (9.7%)	33 (11%)	0.58	37 (9.5%)	42 (10%)	0.67
Contract 2020 S2 (1=yes)	43 (8.8%)	19 (6.3%)	0.19	29 (7.5%)	34 (8.5%)	0.61
Contract 2021 S1 (1=yes)	78 (16%)	49 (16%)	0.97	62 (16%)	68 (17%)	0.72

[1] *Mean (SD); n (%)*

[2] *Wilcoxon rank sum test; Pearson's Chi-squared test*

2.2.3.2 Outcome variables

We collected information on knowledge, self-reported practice adoption, and self-reported productivity. Knowledge questions relating to different stages of crop-production had three answer options; one correct answer, one wrong answer and an option “I don’t know”. They were recoded into binary variables taking the value of one if the correct answer was given and zero otherwise. Six knowledge questions were aggregated into a knowledge score by taking the simple mean[5]. Similarly, eight practice adoption questions

[5] The knowledge score includes three questions on disease identification and three on (post-) harvest practices. Following the pea-analysis plan, two questions regarding storage were excluded because more than 95 per cent of respondents gave the correct answer.

were aggregated into a practice score[6]. We extend previous studies by including (post-)harvest recommendations and outcomes related to aflatoxin contamination, a qualitative yield dimension that is often overlooked because it is not directly visible (Ricker-Gilbert et al. 2022). We report impact on reported yields, but acknowledge that effects on the intermediate outcomes are expected to be more directly affected by our intervention.

Table 2.2: Outcome variable summary

Characteristic	**2021 Sample**			**2022 Sample**		
	Control N = 487[1]	**Treated** N = 304[1]	**p-value**[2]	**Control** N = 388[1]	**Treated** N = 402[1]	**p-value**[2]
Knowledge Index	0.54 (0.23)	0.60 (0.21)	<0.001	0.57 (0.27)	0.59 (0.26)	0.40
Practice Index	0.63 (0.22)	0.70 (0.18)	<0.001	0.64 (0.14)	0.64 (0.15)	0.71
Yield: Buckets per area (log)	2.26 (1.04)	2.49 (0.74)	0.39	2.63 (0.61)	2.65 (0.63)	0.32
Unknown	189	108		8	12	

[1] *Mean (SD)*
[2] *Wilcoxon rank sum test*

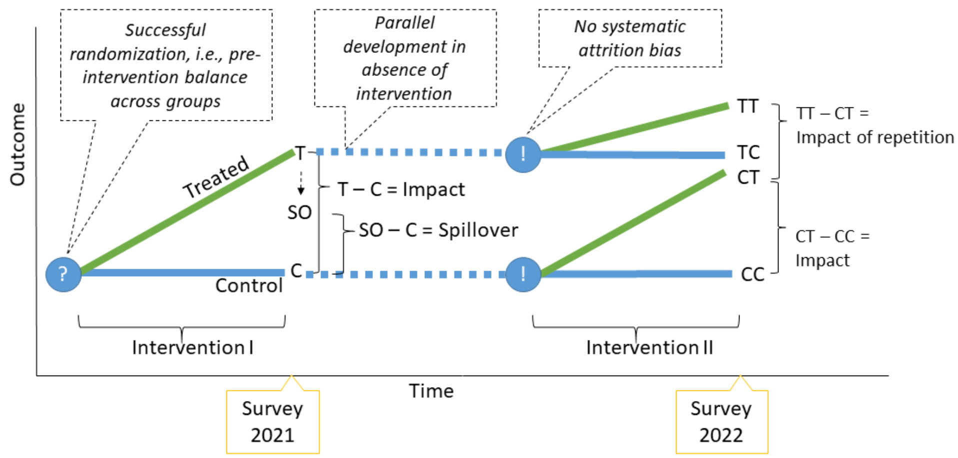
Note*: The difference in the unconditional means of outcome indices in 2021 ceases to be significant when controlling for the misallocated treatment and resulting covariate imbalance. Therefore, these differences rather indicate a selection bias in 2021 treatment allocation than genuine effects of treatment (see Section 2.4 for how we controlled for imbalances). In the 2022 sample, differences are no longer significant, which is consistent with the lack of impact we report.*

6 We included two questions regarding field preparation, pest management, harvest and post-harvest practices, respectively. Two questions regarding storage were excluded because more than 95 per cent of respondents had adopted them.

2.2.4 *Empirical Framework*

Conceptually, our study builds on the Neyman-Rubin framework of potential outcomes to identify causal effects (Neyman 1923; Rubin 1974). The internal validity of our estimation rests on the stable unit treatment value assumption, i.e. the potential outcome of one unit is not affected by the treatment status of another unit and treatment allocation is random[7]. Conceptual pathways to identify the main impact of the GAP-intervention, village level spillover effects, impact of repeated treatment and effect permanence are shown in Figure 2.3.

Figure 2.3: Two-period impact identification framework



***Note**: Green lines indicate hypothesized changes in outcomes due to the intervention, while blue lines indicate no change. Dashed boxes show implicit assumptions at different stages. Main impact is the difference between T and C (in 2021 without controlling baseline outcomes) and the difference between CT and CC (in 2022, with 2021 outcome levels as control). SO denotes the spillover sample of farmers that live in treatment villages, but did not receive GAP-messages in 2021. The intensity of indirect exposure to village neighbors*

[7] To avoid potential bias from spillovers in 2021, we excluded farmers who did not receive GAP messages but were located in treatment villages. The individual-level randomization in 2022 did no longer allow for this procedure, but the absence of a clear impact in 2021 suggests a negligible role of spillovers.

serves as an indicator of spillovers when compared to the pure control sample (C). In 2022, The difference between TT and CT quantifies the additionality of repeated treatment. In theory, the difference between TC and TT or between TC and CC would serve as an indicator of impact permanence over time, but it was not possible to assess this due to the null-result in the first period. The question mark indicates the unverified assumption of equal outcome levels across groups prior to treatment; the exclamation mark denotes that this assumption was verified.

Empirically, we estimated the average treatment effect on the annual cross-sectional samples via the linear model

$$Y_{ij} = \beta_0 + \beta_1 T_i + \beta_2 X_{ij} + \epsilon_{ij} \qquad \text{(Eq. 2.1)}$$

where y is the outcome of farmer i in village j, T is the binary treatment indicator with the parameter of interest β, X is a set of control variables, and ϵ is an identically distributed error term with mean zero. For the first year (2021), treatment T_i indicates the groups C and T in Figure 2.2, while for the second year (2022), T_i indicates the groups CC and CT, since the second year's specification is based on the subsample that did not receive treatment in 2021. For cross-sectional regressions, we clustered the standard errors at the village level. We included all observable covariates to account for remaining imbalances (see Table 2.1) to account for potential pre-treatment group differences. The covariates collected during 2021 are based on a five-year recall and thus reflect the pre-intervention year 2016. We include a sparse model specification with no covariates in the Figure S6 as a robustness check. For the regression using the cross-sectional sample of 2022, we add a control variable indicating treatment status in the year 2021 to account for different levels of treatment intensity.

To exploit the panel data structure, we employed fixed-effects to eliminate time-invariant unobserved heterogeneity:

$$Y_{ijt} = \alpha_i + \beta_1 T_{it} + \beta_2 X_{ijt} + u_{jt} + \epsilon_{ijt} \qquad \text{(Eq. 2.2)}$$

where Y_{ijt} is outcome i of farmer j at time t, α are individual fixed effects, while the other model elements are the same as in (2.1), with additional subscripts representing the time dimension. The treatment indicator T_{it} takes the value of zero for never treated farmers (CC), one for farmers treated in either year (CT and TC), and two for farmers treated in both years (TT). With this continuous treatment indicator, we assume constant effects, i.e., that going from never treated to once treated has the same average effect as going from once to twice treated. We report an alternative specification where we explicitly estimate the additive effect of repeated treatment indicated by a dummy variable in the supplementary material (Figure S4). To ensure our assumption of parallel trends between years, we controlled whether a farmer had a contract with the SE in season 2 of the year 2021, but this was only the case for 11% of farmers. Of the 791 farmers that completed the 2021 survey, only 592 also completed the 2022 survey. The attrition rate in the treatment group was at 36% twice as high as in the control group (18%). This was expected, since we oversampled some treatment villages in 2021 and decided to prioritize a more uniform number of farmers per village for the 2022 survey. To account for selective attrition, we followed Weuve (2012) and employed inverse attrition probability weights to adjust our FE estimates.

Finally, we created subgroups along relevant dimensions to study impact heterogeneity based on the internally valid CC vs. CT sample. To do so, we split continuous variables at their median, and binary variables into their respective levels to understand whether impacts differ along age, sex, education, farm size, income, road distance and contract farming scheme. The latter particularly served to identify the complementarity of digital and traditional extension, since the contract farming included four physical extension visits and access to inputs.

2.2.5 Deviations from study plan and robustness checks

During the year 2021, relevant deviations from the pre-registered study design occurred with potential implications for the internal validity. Given the unexpected result, our most relevant robustness check was the repetition of the experiment in the subsequent year. In addition, we conducted multiple robustness tests in addition to our main specification for the 2021 cross-section, as described below.

First, the SMS were intended to accompany the phenological stages during the season (i.e. content on planting, fertilization, weed- and pest management, maturity checks and post-harvest practices), but due to technical delays not all GAP-content arrived in time to be relevant for farmers. We therefore used a reduced practice adoption index in the Supplementary Material (Figure S1) that excludes early-stage practices, since no effect can be expected[8].

Second, the village-level covariate balancing was compromised, because the village-level information obtained from local exerts was partially inaccurate – a fact that we could only verify after collecting follow-up data with the exact GPS-based location of all farms. In addition, due to unreproducible technical glitches, the treatment was administered to 46 farmers in the planned control group, but there were also 93 farmers in the planned treatment group that did not receive SMS. The result was a significant imbalance between treatment and control group as shown in Table 2.1. To test whether our results are sensitive to selection bias we provide alternative estimates based on different purposefully chosen subsets of observations (Figure S2). First, we ran separate regressions for the subset of farmers

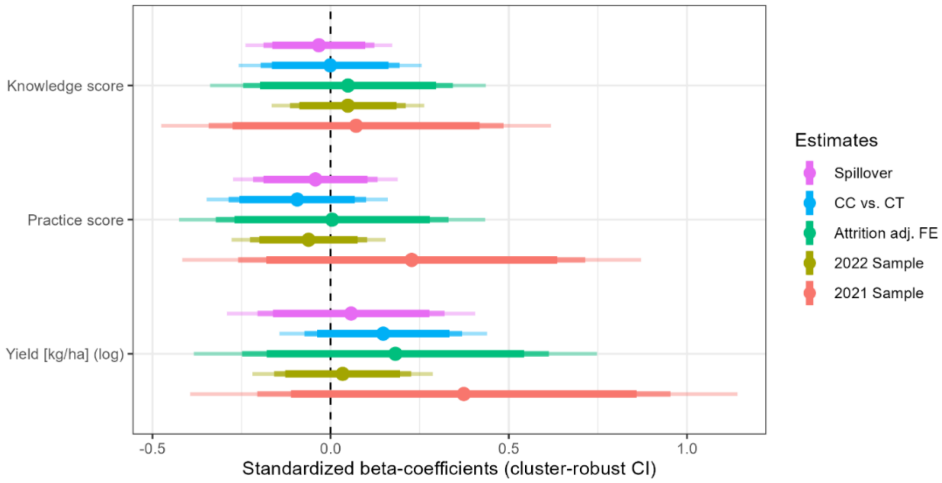
[8] We did not create a reduced knowledge index since farmers can still learn even if messages arrive later.

located in and outside of priority villages, respectively, to check for heterogeneous treatment effects. Second, we use nearest-neighbor propensity-score matching. Matching aims to reduce selection bias and relies on the assumption that all variables relevant for the selection process have been observed. We used the logit link with replacement and a caliper of 0.1 in the *MatchIt* package (Ho et al. 2011) in R (R Core Team 2020) and visually verified post-matching covariate balance. For time-variant covariates, we used the recall data collected during the 2021 survey, which refer to the year 2016 (pre-treatment). In this way we sought to avoid control variables being affected by treatment status.

2.3 Results

A simple comparison of means for the 2021 outcomes indicates significant differences between the treatment and control group (Table 2.2). However, attribution of these differences to the SMS-intervention is not straightforward because of potential selection bias resulting from potential non-random treatment allocation in 2021. Points in Figure 2.4 shows the estimated intention-to-treat effect GAP-messages had on the overall knowledge- and practice adoption scores as well as yields. Corresponding error bars show three levels of cluster robust confidence intervals, namely 90%, 95% and 99%. All estimates were divided by the standard deviation of the outcome and can be interpreted as standardized mean differences. Red and green estimates are based on Formula (2.1) using the 2021 and 2022 cross-sections, respectively. The light blue estimates are based on Equation (2.2) and the full panel data, i.e., the farmers with available observations in both years. Finally, the purple estimates are also based on Equation (2.2), but rely only on farmers from the 2021 control group. This last one is our preferred model specification, since it provides a clean before-after-control-intervention contrast. None of the estimates are significantly different from zero, and we cannot reject the hypothesis that GAP-messages did not increase knowledge, practice adoption and yields. Statistical significance aside (Heckelei et al. 2021), the estimated 90-99% confidence intervals suggest rather small to moderate effect ranges.

Figure 2.4: GAP SMS impact estimates



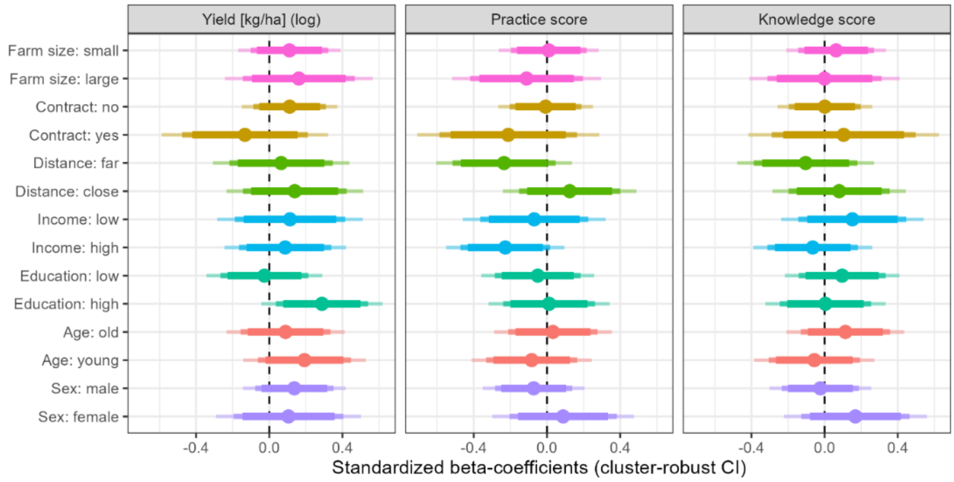
***Note**: For the three outcomes knowledge, practice adoption and yield, points indicate Intention to treat (ITT) estimates, lines are cluster-robust 90, 95, and 99% confidence intervals. Colors show different subsamples; red and olive-green employ cross-sectional regressions for 2021 and 2022, respectively (Eq.2.1). Dark green is an attrition-adjusted two-ways panel regression with all available observations, Light-blue uses only panel observations that did not receive treatment in 2021 (Eq. 2.2). Pink estimates compare farmers in control villages to farmers in treatment villages that did not receive GAP-messages. All regressions include household control variables presented in Table 2.1.*

The null-result prevented us from proceeding to assess within-village spillover effects as we had planned. We report spillover estimates in pink for completeness, but in the absence of an overall effect it would be surprising if they were anything but insignificant, so we do not consider them any further. Still, we tested whether estimates differed for a range of subgroups. To create the subgroups, we split the sample at the median of the respective variables (farm size, income, travel time, years of schooling, age), or into the two categories of binary variables (having a contract with the SE or not, being male or female). We did not find evidence for any heterogeneous effects (Figure 2.5). Furthermore, we used available secondary outcomes such as self-reported sales, harvest losses, and pest incidence, but did not find any impacts on any of them (Figure S1, Supplementary Information). Across a wide range of tested model specifications, including a covariate matching estimator, excluding subsets of control variables, and imputing missing covariate values, we found no coherent impact of the intervention (Figures S2-S6, Supplementary Information)

2.4 Discussion

The lack of evidence for any impact of the intervention was unexpected. Repeating the experiment in the consecutive year confirmed the null-result. This finding partially contrasts with prior research on this topic (see Section 2.1), which warrants a discussion of potential empirical and conceptual explanations. From a policy learning point of view, even experiments with methodological drawbacks are valuable. They can help to reduce the uncertainty around a given phenomenon and enable us to gauge the sensitivity of impacts to the violation of specific assumptions, design choices, and contextual factors. Moreover, even findings with very local external validity can be incorporated in meta-analyses, as long as they are target-equivalent and unbiased (Slough & Tyson 2022).

Figure 2.5: Heterogeneity impact estimates



Note: *For the three outcomes knowledge, practice adoption and yield, points indicate Intention to treat (ITT) estimates, lines are cluster-robust 90, 95, and 99% confidence intervals. For this heterogeneity analysis we split the data in high/low values of the respective variable of interest at the median (or 1/0 in case of binary variables). Colors show different subsamples based on the internally valid sample of farmers that did not receive treatment in 2021 and treatment was randomly allocated in 2022 (i.e. CC vs. CT groups).*

2.4.1 Empirical considerations

We sent out SMS and ensured that farmers were technically able to receive them. However, this does not imply that they actually read the messages and paid attention to its content. Therefore, rather than the actual treatment effect, we could only identify intention-to-treat (ITT) estimates, which is what other studies also did.

Our power calculations and resulting sampling strategy in 2021 were informed by local partners and literature on similar experiments. Our original sampling strategy was based on a power calculation aiming at identifying effect sizes in the order of 0.1 standard deviations at 95%

confidence with 80% power. The pre-analysis plan specifies the relevant assumptions we made in terms of number of villages (clusters), cluster sizes, intra-cluster correlation (ICC), detectable effect sizes and desired power. As a result of the deviations from our study design, the de-facto sample considerably deviated from our planned sample in terms of cluster group sizes due to organizational constraints, so we conducted an *ex-post* power calculation. We anticipated 46 clusters, each with 10-30 farmers, assuming an intra-cluster correlation of 0.2. De facto, we had higher variation in cluster size and therefore also higher intra-cluster correlation. Hence, the power dropped substantially, as illustrated in Figure S9. While we expected to detect changes in binary outcomes in the order of 10-20 percentage points with 80% power, we only achieved detectable changes in binary outcomes in the order of 15-30%. Similarly, for continuous outcomes we expected to detect effect sizes of 0.15 standard deviations, but in fact were only able to detect effect sizes of 0.38 standard deviations based on our cluster-randomization design. Notably, an effect size of 0.38 standard deviations translates to a 0.1 unit change in our main outcome indices, the knowledge and adoption scores. Importantly, an impact below that threshold is unlikely to have a relevant economic implication. In addition, we repeated the intervention in 2022, but randomized at the individual level. This allows for potential spillovers within villages, but if we assume that such spillovers do not exist, the power of that design increases dramatically, and we could have detected effect sizes greater than 0.12 standard deviations with 80% power in the 2022 cross-section. We thus argue that our study was still sufficiently powered.

Based on our pre-analysis plan, we tested our hypothesis using conventional levels of confidence, i.e., 95%. This implies by definition that one in 20 studies will not show a significant statistical effect even if it actually exists. As indicated by the additional confidence bars for 90% and 99% in Figure

2.4, our results are not sensitive to the chosen level of confidence. Moreover, since we repeated our experiment and did not find any impact on any of the measured outcomes, we are confident that we are not looking at a false negative. This conclusion is supported by the individual-level randomization we opted for in the second year, which gave us much more statistical power at the expense of being able to detect potential spillovers. Finally, we tested an alternative ANCOVA model specification in which we controlled for pre-intervention outcomes as suggested by McKenzie (2012), but did not find any significant impacts (Figure S8)[9].

2.4.2 Conceptual considerations

2.4.2.1 Study participants

The participating farmers in our study may have different characteristics from Haitian farmers in the general. In particular, they all had attended at least one workshop with the social enterprise, where they received peanut-related information and could sign up for a contract involving extension visits and input-provision in exchange for peanuts at the end of the season. That is, the farmers that attended these events are at least partially market-oriented land-owners, most with previous experience in peanut cultivation. Promotion of the events originally occurred via local groups (including churches), but since the organization has been active for almost a decade there is no particular bias to be expected with respect to who has heard of it. Table S3 contains a comparison between household- and outcome characteristics of our sample with representative secondary data sources (ICF International, 2013; Kostandini et al., 2021). The large majority of variables is within a ten percent margin of wider population estimates, indicating general representativeness. However, farmers in our sample had

[9] We thank an anonymous reviewer for this suggestion to improve statistical power.

higher levels of knowledge regarding peanut production than average farmers as indicated by higher levels of certified seed adoption and knowledge regarding post-harvest practices. Previous studies indicate that overall higher knowledge levels make incremental effects harder to achieve (Casaburi et al. 2013), especially via information that can only occupy 160 signs in a short text message (van Campenhout 2021). Nevertheless, the range of average knowledge- and adoption scores between 60-70% (Table 2.2) in our view still suggest sufficient opportunity for improvement and detectable impacts.

One study from Ecuador found heterogeneous impacts on practice use of an SMS-intervention along education levels although they did not find a significant overall effect on knowledge (Carrión-Yaguana et al. 2020). Although prevailing illiteracy is high in our study region, farmers had previously reported that at least one household member could read and understand the messages. Our results provide no evidence for heterogeneous effects across levels of education (Figure 2.5). Anecdotal evidence suggests no signs of message fatigue or mistrust in the context of this study.

2.4.2.2 Setting

Previous studies in Haiti found that subsidies did not improve input use or yields because farmers substituted rather than complemented their former input use for reasons of incorrect expectations (Gignoux et al. 2022). Specific for peanuts, another study found that a technology transfer program did not improve food security, production value or even the use of post-harvest technologies (Macours et al. 2018). Such findings may be specific to Haiti, which has been repeatedly exposed to external shocks over the past years and also during our study period, thus limiting generalizability. On July 7th 2021, Haiti’s president was assassinated, leading to political instability and social unrest that persisted throughout the following year. In addition, the southwestern part of Haiti was struck by an earthquake with a magnitude of 7.2 on August 14 followed by a Tropical Depression on August 16 of 2021. These shocks were not concentrated in our study region, but may have affected aspirations and thereby production and consumption decision of both treatment and control group (Tabe-Ojong et al. 2021). We speculate that as a result of social instability, farmers may put less trust in external information and their risk- and time preferences could become more attenuated. As a consequence, farmers could have paid less attention to messages and relied on sub-optimal, but known production practices.

We initially expected that effects may be stronger in priority villages due to other unobserved characteristics, but did not find clear evidence for such heterogeneous effects (Figure S2). As shown in Table 2.1, farmers in the treatment group had significantly more assets and shorter travel times. If these characteristics are also indicative of better pre-intervention managerial skills, impact estimates would be biased upwards. The fact that we did not find any impact thus supports the null-result. The likelihood that farmers were systematically unable to receive messages due to a lack of electricity or network coverage is low, since we verified message reception in both groups. Nonetheless, such infrastructural barriers could reduce the potential of ICT-based extension services in our study region and contribute to a digital divide through channels beyond the studied intervention (Mehrabi et al. 2021).

2.4.2.3 Design

Our evaluation design differed in various aspects from previous studies on the same topic. Although we aimed to control for pre-treatment differences by asking farmers to recall information on time-variant control variables five years ago, recall information may be subject to non-random measurement error. Therefore, we cannot be certain of the (im-) balances presented in

Table 2.1, left side, and rather rely on the robustness of our results with and without these control variables to derive our conclusions. Additionally, we did not collect outcome variables before the intervention in 2021. This in turn affected our intervention design, because identification of knowledge- and practice adoption gaps and the content design were necessarily obtained from local experts. Participatory development of message content in collaboration with the target group could have increased the effectiveness, but there is a limit to the complexity of short text messages. Previous studies found communication vehicles that allow for more content complexity, such as apps, to be more effective than SMS (Giulivi et al. 2022). In 2021, the SMS-intervention was delayed several weeks, which made the content of some messages useless because they did not arrive in a timely manner. Our robustness check using a reduced adoption score underlined that this source of error was not influential (Figure S1).

2.4.3 Lessons learned for experimental studies in presence of implementation constraints

Many constraints dictated by field reality may limit both the implementation of an intervention as well as the experimental design for its evaluation. Constraints are less of a problem if they can be effectively mitigated or do not drastically affect the validity of a research design. On the other hand, unanticipated constraints that compromise the validity of results can pose substantial challenges. In the following, we provide examples of both types of constraints and give practical strategies to avoid and address some limitations that may occur in experimental studies.

We anticipated that the context in which the intervention of this study took place would make it challenging to adhere to all aspects of our registered pre-analysis plan. However, a pre-registration does not mean that one cannot adapt a research design to unforeseen circumstances. Rather, it facilitates thoroughly study planning and enabled us to transparently put empirical weaknesses into perspective *vis-à-vis* a baseline scenario. Consequently, we recommend pre-registration under all circumstances, even under high risk of partial deviation from the study plan dictated by field conditions.

A more critical constraint includes the unavailability of baseline data and difficulty to identify baseline group imbalances at an early stage. One way to account for potential imbalances post-hoc is to collect recall data that can be used as control or to support covariate matching as we did during the follow-up survey. An additional possibility for outcomes such as yields and certain agricultural practices is to use remote sensing data to measure pre-intervention outcomes (Cole 2020); but this requires exact knowledge of the field boundaries while we only knew the coordinates of the farm. The evaluation procedure can be kept more flexible by collecting additional outcome variables or use different aggregation methods. In our case we created a reduced practice adoption index to account for unforeseen delays in message timing as a robustness check.

Finally, there are constraints that are both unforeseen and critical to a study's validity. In our case, this included not being able to conduct field visits and do qualitative work such as focus group discussions in person due to travel restrictions and security concerns. While digital communication can ameliorate some issues, it is not agile in contexts of limited connectivity and provides no first-hand opportunity to evaluate how a survey went or to collect spontaneous reactions of farmers. Building up trust and effective communication with implementers – ideally accompanied by real-time monitoring of all relevant field operations - thus remains crucial.

From a policy learning perspective, it is important to conduct experiments in challenging settings, since research bias could omit important policy implications. This is particularly the case for real-world interventions which are typically not designed by researchers. Furthermore, some constraints may be overcome by planning and empirical methods *ex-post*. But even when everything goes as planned, experiments have many limitations and it remains crucial to determine their adequacy for a given setting and research question (Bédécarrats et al. 2020).

2.5 Conclusion

This study found no evidence that ICT-based information affected knowledge, practice adoption or yields among Haitian peanut farmers. We discussed the validity of these findings and argue that two context-specific factors have contributed to these particular outcomes. First, our sample was relatively well-informed already before the intervention, suggesting diminishing marginal returns to SMS-based information provision. Second, exogenous shocks during the study period may have increased risk aversion and reduced trust in non-personal information sources.

In addition, similar studies in other contexts suggest that more personally tailored information based on bidirectional information flows provide larger benefits. Yet, the required technical means (e.g. Apps, Interactive Voice Recordings) have higher entry barriers in terms of required literacy and access. Therefore, it is important for policy makers to consider and limit the effect of how differentiated (digital) information access may intensify existing inequalities and marginalization of certain groups. More research is needed to understand how contextual factors shape the efficacy of digital farm advisory services.

Future studies would benefit from using measured as opposed to self-reported outcome data, such as via quantifying actual field-level yields as an indicator of productivity using remote sensing. This could help to provide more personalized and timely advice that would be of higher relevance to the receivers.

Supplementary material

Supplementary data are available at *Q Open* online.

Data availability

The data and code underlying this article are available along with the pre-analysis plan at https://osf.io/cazqw.

Acknowledgements: The authors would like to thank the employees of the social enterprise and all study participants for their contributions. In addition, the first author is grateful for insightful discussions with Matthias Sand, Yannic Damm, Bisrat Gebrekidan, and participants at the 2022 EAAE PhD Workshop in Parma.

2.6 References

Arouna, A., Michler, J. D., Yergo, W. G., and Saito, K. (2021). One Size Fits All? Experimental Evidence on the Digital Delivery of Personalized Extension Advice in Nigeria. *American Journal of Agricultural Economics*, *103*(2), 596–619.

Baumüller, H. (2018). The Little We Know: An Exploratory Literature Review on the Utility of Mobile Phone-Enabled Services for Smallholder Farmers. *Journal of International Development*, *30*(1), 134–154.

Bédécarrats, F., Guérin, I., and Roubaud, F. (Eds.). (2020). *Randomized control trials in the field of development: a critical perspective*. Oxford: Oxford University Press.

Camacho, A., and Conover, E. (2019). The impact of receiving SMS price and weather information on small scale farmers in Colombia. *World Development*, *123*(3), 104596.

Carrión-Yaguana, V. D., Alwang, J., and Barrera, V. H. (2020). Promoting Behavioral Change Using Text Messages: A Case Study of Blackberry Farmers in Ecuador. *Journal of Agricultural and Applied Economics*, *52*(3), 398–419.

Casaburi, L., Kremer, M., Mullainathan, S., and Ramrattan, R. (2013). *Harnessing ICT to Increase Agricultural production: Evidence from Kenya** (Working Paper).

Cole, S. A., Harigaya, T., Killeen, G., and Krishna, A. (2020). *Using Satellites and Phones to Evaluate and Promote Agricultural Technology Adoption: Evidence from Smallholder Farms in India* (Working paper).

Ding, J., Jia, X., Zhang, W., and Klerkx, L. (2022). The effects of combined digital and human advisory services on reducing nitrogen fertilizer use: lessons from China's national research programs on low carbon agriculture. *International Journal of Agricultural Sustainability*, *16*(1), 1–14.

Duflo, E., Glennester, R., and Kremer, M. (2006). Using Randomization in Development Economics Research: A Toolkit. Cambridge, MA: Center for International Development at Harvard University.

Fabregas, R., Kremer, M., and Schilbach, F. (2019). Realizing the potential of digital development: The case of agricultural advice. *Science (New York, N.Y.)*, *366*(6471).

Fulmer, A. M. (Ed.). (2018). Haiti Peanut Research Report: Summary Report of 2015 to 2017 Data: Efforts to improve peanut production in Haiti by investigating management options for foliar diseases, low soil fertility, and other yield-limiting agronomic issues. Athens, GA: Haiti Peanut Value Chain Intervention.

Fulmer, A. M., Brenneman, T. B., Kemerait, R. C., Macajoux, R., Carroll, D. A., Faroutine, G., … MacDonald, G. E. (2020). Evaluation of Improved Valencia Peanut Varieties for Production in Haiti. *Peanut Science*, *47*(1), 1–8.

Gignoux, J., Macours, K., Stein, D., and Wright, K. (2022). Input subsidies, credit constraints, and expectations of future transfers: Evidence from Haiti. *American Journal of Agricultural Economics*, ajae.12337.

Giulivi, N., Harou, A. P., Gautam, S., and Guereña, D. (2022). Getting the message out: Information and communication technologies and agricultural extension. *American Journal of Agricultural Economics*, ajae.12348.

Heckelei, T., Huettel, S., Odening, M., Rommel, J., Heckelei, T., Huettel, S., … Rommel, J. (2021). *The replicability crisis and the p-value debate – what are the consequences for the agricultural and food* *economics community?* (Discussion Paper No. 2021:2). Bonn: Institute for Food and Resource Economics. Retrieved from https://ageconsearch.umn.edu/record/316369

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, *42*(8), 1–28.

ICF International. (2013). *Haiti Baseline Survey*. Washington: USAID.

Imai, K., King, G., and Nall, C. (2009). The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation. *Statistical Science*, *24*(1).

Jack, B. K. (2011). Market Inefficiencies and the Adoption of Agricultural Technologies in Developing Countries. Center for International Development at Harvard University, and Agricultural Technology Adoption Initiative, J-PAL (MIT) and CEGA.

Kostandini, G., Rhoads, J., MacDonald, G. E., Tanellari, E., Johnson, R., Carroll, E., and Pressoir, G. (2021). Production, post-harvest management and gender dynamics among smallholder peanut farmers in Haiti. *Agriculture & Food Security*, *10*(1), 141.

Larochelle, C., Alwang, J., Travis, E., Barrera, V. H., and Dominguez Andrade, J. M. (2019). Did You Really Get the Message? Using Text Reminders to Stimulate Adoption of Agricultural Technologies. *The Journal of Development Studies*, *55*(4), 548–564.

Macours, K., Stein, D., Salazar, L., Gachot, S., Jacquet, B., Fahsbender, J., … de Salvo, C. P. (2018). *Technology Transfer to Small Farmers Program (PTTA) in Haiti: Implementation, Evaluation and Lessons Learned*. Inter-American Development Bank.

Maredia, M. K., Reyes, B., Ba, M. N., Dabire, C. L., Pittendrigh, B., and Bello-Bravo, J. (2018). Can mobile phone-based animated videos induce learning and technology adoption among low-literate farmers? A field experiment in Burkina Faso. *Information Technology for Development*, *24*(3), 429–460.

Mehrabi, Z., McDowell, M. J., Ricciardi, V., Levers, C., Martinez, J. D., Mehrabi, N., … Jarvis, A. (2021). The global divide in data-driven farming. *Nature Sustainability*, *4*(2), 154–160.

Neyman, J. (1923). On the application of probability theory to agricultural experiments: Essay on principles, section 9. (translated in 1990). *Statistical Science*, *5*, 465–480.

Oyinbo, O., Chamberlin, J., Abdoulaye, T., and Maertens, M. (2022). Digital extension, price risk, and farm performance: experimental evidence from Nigeria. *American Journal of Agricultural Economics*, *104*(2), 831–852.

R Core Team. (2020). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Rajkhowa, P., and Qaim, M. (2021). Personalized digital extension services and agricultural performance: Evidence from smallholder farmers in India. *PloS One*, *16*(10), e0259319.

Ricker-Gilbert, J., Omotilewa, O., and Kadjo, D. (2022). The Economics of Postharvest Loss and Loss-Preventing Technologies in Developing Countries. *Annual Review of Resource Economics*, *14*(1), 243–265.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.

Ruml, A., and Qaim, M. (2020). Effects of marketing contracts and resource-providing contracts in the African small farm sector: Insights from oil palm production in Ghana. *World Development*, *136*(1), 105110.

Slough, T., and Tyson, S. A. (2022). External Validity and Meta-Analysis. *American Journal of Political Science*, ajps.12742.

Spielman, D., Lecoutere, E., Makhija, S., and van Campenhout, B. (2021). Information and Communications Technology (ICT) and Agricultural Extension in Developing Countries. *Annual Review of Resource Economics*, *13*(1).

Tabe-Ojong, M. P., Heckelei, T., and Baylis, K. (2021). Aspiration Formation and Ecological Shocks in Rural Kenya. *The European Journal of Development Research*, *33*(4), 833–860.

Tambo, J. A., Aliamo, C., Davis, T., Mugambi, I., Romney, D., Onyango, D. O., … Byantwale, S. T. (2019). The impact of ICT-enabled extension campaign on farmers' knowledge and management of fall armyworm in Uganda. *PloS One*, *14*(8), e0220844.

TechnoServe. (2014). Smallholder Best Practice Guide: Growing Peanuts in Haiti: Final Draft. TechnoServe.

Tyroler, C. (2018). *Gender Considerations for Researchers Working in Groundnuts*. Washington: USAID & Feed the Future.

van Campenhout, B. (2021). The Role of Information in Agricultural Technology Adoption: Experimental Evidence from Rice Farmers in Uganda. *Economic Development and Cultural Change*, *69*(3), 1239–1272.

van Campenhout, B., Spielman, D. J., and Lecoutere, E. (2021). Information and Communication Technologies to Provide Agricultural Advice to Smallholder Farmers: Experimental Evidence from Uganda. *American Journal of Agricultural Economics*, *103*(1), 317–337.

Vandevelde, S., van Campenhout, B., and Walukano, W. (2021). Accounting for spillovers in assessing the effectiveness of video messages to improve potato seed quality: evidence from Uganda.

The Journal of Agricultural Education and Extension, *27*(4), 503–534.

von Braun, J., Afsana, K., Fresco, L. O., and Hassan, M. (2021). Food systems: seven priorities to end hunger and protect the planet. *Nature*, *597*(7874), 28–30.

Weuve, J., Tchetgen Tchetgen, E. J., Glymour, M. M., Beck, T. L., Aggarwal, N. T., Wilson, R. S., … Mendes de Leon, C. F. (2012). Accounting for bias due to selective attrition: the example of smoking and cognitive decline. *Epidemiology (Cambridge, Mass.)*, *23*(1), 119–128.

Chapter 3
On-demand multimedia farming advice increases knowledge on pest management among vegetable farmers in El Salvador[10]

3.1 Introduction

Digital technologies are emerging as important means of agricultural transformation towards food security and sustainability (von Braun et al. 2021). Information and Communication Technologies (ICT), in particular, have received increasing attention for their potential to disseminate information to farmers promptly, also in remote locations (Baumüller 2018; Fabregas et al. 2019). However, the effectiveness of various ICT formats on knowledge acquisition and adoption of practices varies.

Previous research investigated the impacts of text messages (Carrión-Yaguana et al. 2020; Casaburi et al. 2013; Larochelle et al. 2019), interactive voice recordings, and videos presented on tablets during personal meetings (Tambo et al. 2019; van Campenhout 2021). Overall, they report positive impacts of ICTs on agronomic know-how. In contrast, the effect on adoption

[10] This chapter is an unpublished report, drafted to complement the study presented in Chapter 2.

of recommended practices and production-related outcomes are less common and generally smaller in magnitude. Interventions based on short text messages struggle to effectively deliver more complex advice and a recent trial with short text messages conducted in Haiti found no impact on farmer's knowledge (Schulz and Börner 2023). Another limitation is that text messages and video screenings contain rather generic content and are pushed to recipients. At the same time, recent studies have shown a sizeable positive impact of personalized advice provided to farmers (Arouna et al. 2021; Rajkhowa and Qaim 2021). Therefore, there is a need to evaluate alternative ICT channels that are able to convey more complex advisory contents that are actionable and relevant.

With the diffusion of smartphones and steadily improving internet network coverage in many rural areas, farm advisory services can leverage more refined and visually interpretable information such as multimedia content and interactive communication platforms. However, previous studies have not rigorously evaluated interactive multimedia content such as audios, videos, infographics and pictures as an ICT-based communication vehicle for farming advice.

In this study we address this gap by examining the effectiveness of a real-world pilot intervention using demand-driven multimedia content delivered through an interactive chat tool. The pilot intervention was executed simultaneously to an SMS-based intervention in Haiti, which yielded no discernible impacts (Schulz and Börner 2023) and was implemented by the identical social enterprise. We present experimental evidence of the randomly phased-in digital extension service among vegetable farmers in El Salvador to assess the impact on farmers' knowledge and adoption of recommended practices. Several contextual factors, such as regular extension visits by the social enterprise, the trust level in information provided by them, and the provision of input and market access, were similar to the intervention in Haiti, enhancing the comparability of results alongside the harmonization of the follow-up survey; however, the distinct socioeconomic and biophysical setting as well as the unique intervention also pose limitations on direct comparison with Schulz & Börner (2023). Our preliminary results from the pilot study indicate significantly higher knowledge during the follow-up survey among farmers who requested multimedia content. We attribute this difference to the uptake, i.e., active participation by compliers, of the randomly phased-in intervention. These results suggest that demand-driven multimedia content has the potential to enhance agricultural knowledge acquisition and potentially improving farming practices in our study setting.

The following Section 3.2 describes the context, intervention, sampling design and empirical framework of our study. Section 3.3 shows the results which are discussed and set into perspective in Section 3.4 before we conclude with policy recommendations in Section 3.5.

3.2 Material and Methods

This study was approved by the Research Ethics Committee of the Center for Development Research (ZEF), University of Bonn. All study participants gave their prior informed consent in written form. An analysis plan for the study was pre-registered at the Open Science Foundation[11]. The study region is located in the Chalatanango district in the highlands of El Salvador. The climate is tropical, with a wet season between May and October. We partnered with a locally operating social enterprise that has been actively providing agricultural training, inputs, and market access to vegetable

[11] The pre-registration plan can be accessed via: https://doi.org/10.17605/OSF.IO/854K7.

farmers in the region for more than three years. To complement their physical extension activities, the organization tested the digital information intervention during a random phase-in to assess its efficacy. The author did not visit the study regions due to travel restrictions and instead relied on virtual meetings with the partner organization to plan and oversee intervention and data collection.

3.2.1 Intervention

The social enterprise actuating in the study region supports vegetable farmers with regular physical extension visits, material inputs, and market access. All farmers working with the social enterprise and owning a smartphone received regular weather forecasts via WhatsApp. This continued throughout the pilot intervention and was not affected by treatment allocation. The pilot intervention provided a mixture of texts, images, audio, and videos with Good Agricultural Practices (GAP) and was offered to farmers via WhatsApp in Spanish. All GAP content was developed by local agronomists. Based on a predefined schedule, treated farmers would first receive a text message offering further information on weekly topics. For example, “Hello [name of farmer], we want to inform you about the thrips plague. To receive daily information on how to combat it you only have to answer YES.” If farmers responded accordingly within a 24-hour time window, they would automatically receive further multimedia content, such as pictures containing the different growth stages of the plague, options of recommended pesticide products, and the recommended dosage. Similarly, a short video showing how and where to set up insect traps was offered to farmers. A screenshot of an exemplary chat is depicted in Figure 3.1 below. This study did not aim to assess whether these multimedia formats had different impacts.

Figure 3.1: Example of simple chatbot interaction

***Note**: Translation from top to bottom: 1) „Hello [name], Extensio brings you the week of paratrioza: Today and on Friday you receive information about this plague and you only have to respond ‚YES' to learn how to combat it. Greetings.". 2) „YES". 3) „Pest control Paratrioza" [close-up picture of paratrioza] „What to apply? For a 18 liter canister 1. Products with thiocyclam (preventive): Tryclan 50 SP (one cup); 2. Products with Spirotetramat: Movento 15 OD (1-2 cups); 3. Products with Flupyradifurone: Sivanto Prime 20 SL (25-50 CC)". One cup refers to the lid of the product container and is equivalent to 25 cubic centimeters (CC).*

We obtained message logs containing detailed delivery status (sent, delivered, read) and created a binary variable “GAP-request” that took the value of 1 if farmers responded affirmatively to the content offering and zero otherwise. Based on the delivery status of the weather messages that were sent to all farmers independent of treatment status, we verified the eligibility of the control group, namely whether farmers could receive WhatsApp messages at all (see

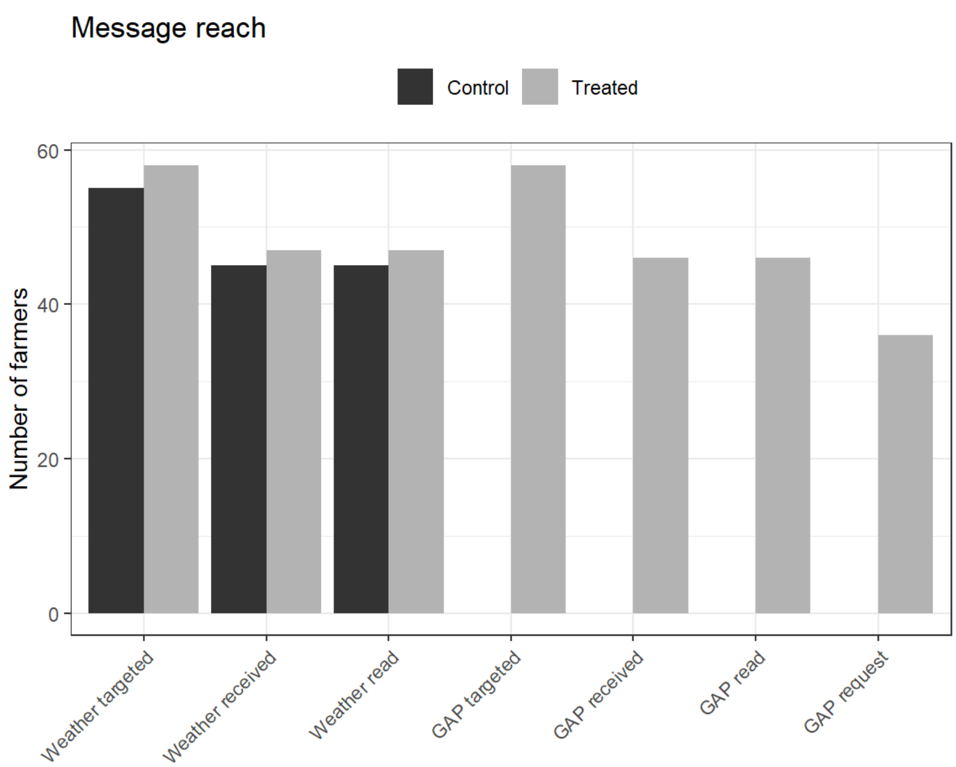
Figure 3.2 for more details).

3.2.2 Sampling approach

To be eligible to receive the intervention, farmers had to fulfill three criteria: 1) actively work with the local input provider in the study region; 2) give their prior informed consent to participate in the study; and 3) own a smartphone and be able to receive WhatsApp messages.

Within the scope of the regular extension visits, 233 farmers working with the local organization at the time were visited to obtain prior informed consent. After explaining the study design and objective, a subgroup of 140 farmers agreed to participate. The implication of this convenience sample in terms of external validity is discussed in Section 3.4. To avoid information spillovers and the associated violation of the stable unit treatment assumption, interventions are often randomized at the village level (Duflo et al. 2006). However, in our case, the available sample size only allowed for individual randomization for an acceptable statistical power. We address the concern of potential information spillovers below. Due to the small sample size, we applied non-bipartite covariate matching based on already available baseline data to increase statistical power (Beck, Lu, & Greevy, 2016; Imai, King, & Nall, 2009). The baseline data was collected via in-person household surveys in December 2020 by extension agents working for the social enterprise. Summary statistics of all baseline characteristics by intervention group are presented in Table 3.1. The WhatsApp-intervention occurred between May and June 2021 and the follow-up survey occurred between July and August 2021. The follow-up survey was also conducted by the regular extension agents during their visit to the farm. Four farmers moved away during the study period and did not finish the follow-up survey, so attrition was 3%, resulting in a final sample size of 136. Figure 3.2 shows the number of farmers in both groups that received weather messages and were thus considered eligible for the intervention. Around 80% of farmers with a smartphone received and read the weather messages; the missing 20% are likely due to a temporary lack of internet access and were transferred to the group of never-takers. We here define never-takers as farmers without smartphones or who did not receive the messages and use them as an alternative control group. The rationale to include this group in the analysis was to assess potential information spillovers because we assumed that messages could be easily forwarded to farmers in the control group, but not to never-takers without a smartphone. Overall, there were 43 farmers who did not have a smartphone or could not receive complementary weather messages during the intervention. Following the pre-analysis plan, one outlier was removed from the analysis due to heavily diverging production characteristics, leaving three final groups: treatment (n=47), control (n=45), and never-takers (n=43).

Figure 3.2: Eligibility and treatment compliance



3.2.3 Empirical Framework

Our primary outcome of interest is an agricultural knowledge index, calculated as the average score across six multiple-choice questions (one correct answer, one wrong answer and one opt-out) gauging farmers' understanding of recommended agricultural practices. To evaluate changes in application behavior, we created a correct dosage index. This index reflects the proportion of times farmers reported using the correct product and dosage when addressing a specific pest, based on three separate pests. Additionally, we collected data on self-reported production changes perceived by farmers after the intervention. This included their subjective assessments of yield losses (increased/decreased) and changes in spending on pesticides and fungicides (more/less).

Distributing the intervention via WhatsApp has the positive side effect that treatment delivery status can be measured by checking whether messages were delivered, visualized by the recipient, and responded to. This allowed us to estimate different effects for assigned treatment status and actual treatment status. First, we estimate the effect of being offered GAP as the intention-to-treat (ITT) effect and use as its indicator a binary variable that takes the value of one if the farmer received GAP-content in the smartphone and zero otherwise. We refer to this as ITT because it does not mean that the recipient paid attention to the message content. On the other hand, if the person actively requested additional information by answering to a GAP-content offering, we expect the probability that the person will also pay attention to subsequent multimedia content to be higher. We, therefore, use a binary treatment indicator variable that takes the value of one if a person requested GAP content to estimate a Local Average Treatment Effect (LATE).

Both treatment indicators are used within the same empirical framework, which can be written as

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \epsilon_i \qquad \text{(Eq. 3.1)}$$

Where Y is the outcome of interest for farmer i, T is the treatment indicator, X are control variables and ϵ is a randomly distributed error term. We include baseline controls to improve the precision of our estimate (Duflo et al. 2006). The estimated effect β is obtained via OLS regression.

Information spillovers can easily occur because our study region covers a relatively small area and the cost of sharing digital information with peers via the same communication channel is very low (Fabregas et al., 2019). Therefore, we assess potential spillover effects by comparing our control group with the group of never-takers that do not use a smartphone. If they are similar in both baseline and outcome characteristics, it is unlikely that the control group benefitted from information spillovers.

3.2.4 Sample characteristics

Table 3.1 summarizes the independent variables across different groups and test for differences. In general, we find good balance across the observed covariates. The treatment group seems to be somewhat better educated compared to the other groups. We consider the randomization to be successful, but nevertheless control for all baseline characteristics during impact estimation.

Following our pre-analysis plan, we excluded two practice adoption outcomes because more than 95% of farmers had adopted them. The excluded outcomes were 1) whether they sprayed pesticides during fresh hours and 2) whether they washed their spraying equipment between usage. Based on the remaining outcomes, we calculated scores for knowledge and correct dosage application by taking the unweighted mean across the respective index components. Regarding the outcome variables during the follow-up survey, we find significant differences between groups, as shown in Table 3.2. Notably, never-takers had significantly fewer correct answers among the knowledge questions than the control group. This could indicate information spillovers or existing information barriers and points towards a potential digital divide (Mehrabi et al., 2021). The comparison of means indicates no significant differences in pest management practices or pesticide use. However, treated farmers spent significantly more resources on pesticide products and their application than the control group, although both groups interacted with the input supplier in the same frequency. Notably, the control group spent the lowest amount of time and money on pesticide products in self-reported absolute terms but, at the same time, did not report to have spent less on it when asked for relative changes. Due to this inconsistency in measurements, we must interpret related results cautiously.

Table 3.1: Group-wise summary statistics of independent baseline variables

Variable	**N**	**All** N = 135[1]	**Control** N = 45[1]	**Never-Takers** N = 43[1]	**Treated** N = 47[1]	**Control vs. Never-Takers**	**Control vs. Treated**	**Never-Takers vs. Treated**
Married (1=yes)	122	67 (55%)	20 (48%)	24 (63%)	23 (55%)	0.17	0.51	0.45
Gender (1=female)	123	7 (5.7%)	2 (4.7%)	2 (5.3%)	3 (7.1%)	0.91	0.62	0.72
Owns greenhouse (1=yes)	123	53 (43%)	17 (40%)	15 (39%)	21 (50%)	>0.99	0.34	0.35
Has irrigation system (1=yes)	123	103 (84%)	35 (81%)	33 (87%)	35 (83%)	0.51	0.81	0.68
Owns tractor (1=yes)	123	5 (4.1%)	1 (2.3%)	2 (5.3%)	2 (4.8%)	0.51	0.57	0.91
Owns water pipe (1=yes)	123	62 (50%)	22 (51%)	21 (55%)	19 (45%)	0.72	0.59	0.38
Has natural gas (1=yes)	123	66 (54%)	24 (56%)	18 (47%)	24 (57%)	0.45	0.90	0.39
Has cable TV (1=yes)	123	83 (67%)	30 (70%)	22 (58%)	31 (74%)	0.26	0.69	0.13
SE primary market (1=yes)	122	96 (79%)	32 (76%)	28 (74%)	36 (86%)	0.79	0.29	0.19
With SE > 3 years (1=yes)	122	83 (68%)	25 (60%)	30 (79%)	28 (67%)	**0.065**	0.48	0.24
Tertiary education (1=yes)	122	6 (4.9%)	0 (0%)	2 (5.3%)	4 (9.5%)	0.28	**0.045**	0.38
Secondary education (1=yes)	122	34 (28%)	15 (36%)	7 (18%)	12 (29%)	**0.088**	0.47	0.31
Primary education (1=yes)	122	35 (29%)	10 (24%)	11 (29%)	14 (33%)	0.62	0.34	0.67
No formal education (1=yes)	122	47 (39%)	17 (40%)	18 (47%)	12 (29%)	0.53	0.26	**0.087**

Variable	**N**	**All** N = 135[1]	**Control** N = 45[1]	**Never-Takers** N = 43[1]	**Treated** N = 47[1]	**Control vs. Never-Takers**	**Control vs. Treated**	**Never-Takers vs. Treated**
Income < 1 min. wage (1=yes)	122	49 (40%)	15 (36%)	12 (32%)	22 (52%)	0.71	0.12	**0.059**
Income 1-2 min. wages (1=yes)	122	61 (50%)	22 (52%)	23 (61%)	16 (38%)	0.47	0.19	**0.046**
Income > 2 min. wages (1=yes)	122	12 (9.8%)	5 (12%)	3 (7.9%)	4 (9.5%)	0.55	0.72	0.81
Age (years)	119	41.20 (13.02)	39.80 (12.43)	44.42 (12.13)	39.59 (14.10)	0.12	0.94	0.10
log(Area cultivated) (ha)	122	0.37 (0.34)	0.40 (0.29)	0.33 (0.36)	0.37 (0.37)	0.36	0.68	0.61
Dependent people (#)	122	2.53 (1.42)	2.57 (1.48)	2.66 (1.38)	2.38 (1.41)	0.79	0.54	0.39
Household size (#)	122	3.87 (1.75)	3.81 (1.61)	3.89 (2.14)	3.90 (1.53)	0.83	0.81	0.98
Full-time employees (#)	122	1.41 (1.45)	1.43 (1.21)	1.08 (1.58)	1.69 (1.51)	0.28	0.41	0.060
Part-time employees (#)	122	2.02 (1.39)	2.05 (1.65)	2.16 (1.28)	1.88 (1.19)	0.72	0.58	0.38
Weather received (#)	113	3.47 (2.00)	4.36 (1.17)	0.00 (0.00)	4.17 (1.31)	**<0.001**	0.43	**<0.001**
log(Distance to road) (km)	125	0.29 (0.35)	0.34 (0.40)	0.26 (0.35)	0.29 (0.30)	0.30	0.52	0.70

***Note**: Group sizes given in the column label refer to sizes after imputing covariates by taking the mean of observed values. Statistics in this table refer only to observed sample characteristics with number of observations indicated on the left. For binary variables numbers indicate frequency (percentage) where value is equal to one; p-statistics are for Pearson's Chi-squared test. For continuous variables we report mean and standard deviation (in brackets); p-statistic refers to Wilcoxon rank sum test.*

Table 3.2: Groupwise summary statistics of outcome variables during the follow-up

Characteristic	Overall, N = 135	Control N = 45	Never-Takers N = 43	Treated N = 47	Control vs. Never-Takers	Control vs. Treated	Never-Takers vs. Treated
Knows material (y/n)	21 (16%)	4 (8.9%)	7 (16%)	10 (21%)	0.34	0.10	0.52
Knows shape (y/n)	70 (52%)	24 (53%)	19 (44%)	27 (57%)	0.39	0.70	0.21
Knows trap mechanism (y/n)	103 (76%)	35 (78%)	25 (58%)	43 (91%)	0.025	0.11	<0.001
Knows blue trap (y/n)	69 (51%)	22 (49%)	14 (33%)	33 (70%)	0.11	0.035	<0.001
Knows yellow trap (y/n)	82 (61%)	26 (58%)	17 (40%)	39 (83%)	0.065	0.010	<0.001
Knows mixture steps (y/n)	107 (79%)	36 (80%)	28 (65%)	43 (91%)	0.080	0.17	0.002
Knows paratrioza product (y/n)	109 (81%)	36 (80%)	28 (65%)	45 (96%)	0.067	0.048	<0.001
Knows blue trap product (y/n)	37 (27%)	9 (20%)	7 (16%)	21 (45%)	0.69	0.007	0.002
Knowledge (index)	0.50 (0.20)	0.48 (0.17)	0.43 (0.22)	0.60 (0.17)	0.15	0.005	<0.001
Used color traps (y/n)	23 (17%)	7 (16%)	5 (12%)	11 (23%)	0.63	0.32	0.14
Correct dosage: Paratrioza (y/n)	101 (75%)	35 (78%)	30 (70%)	36 (77%)	0.39	0.90	0.46
Correct dosage: Pulgon & MB (y/n)	78 (58%)	28 (62%)	24 (56%)	26 (55%)	0.55	0.51	0.96
Correct dosage: Trips (y/n)	98 (73%)	32 (71%)	29 (67%)	37 (79%)	0.70	0.42	0.24
Correct dosage (index)	0.70 (0.31)	0.72 (0.29)	0.66 (0.32)	0.72 (0.31)	0.35	0.94	0.31
Higher losses (SR; y/n)	24 (18%)	8 (18%)	3 (7.0%)	13 (28%)	0.18	0.21	0.010
Lower losses (SR; y/n)	81 (60%)	27 (60%)	26 (60%)	28 (60%)	0.96	0.97	0.93
Higher spending (SR; y/n)	27 (20%)	11 (24%)	9 (21%)	7 (15%)	0.68	0.26	0.48
Lower spending (SR; y/n)	71 (53%)	23 (51%)	17 (40%)	31 (66%)	0.27	0.15	0.012

Characteristic	Overall, N = 135	Control N = 45	Never-Takers N = 43	Treated N = 47	Control vs. Never-Takers	Control vs. Treated	Never-Takers vs. Treated
Phytosanitary spending [$]	137.65 (119.46)	108.11 (75.05)	134.30 (125.34)	169.00 (141.32)	0.30	0.014	0.16
Phytosanitary work (hours/week)	26.04 (20.71)	21.60 (16.43)	26.74 (23.81)	29.66 (20.98)	0.24	0.063	0.50

***Note**: For binary (y/n) variables numbers indicate frequency (percentage) where value is equal to one; p-statistics are for Pearson's Chi-squared test. For continuous variables we report mean and standard deviation (in brackets); p-statistic refers to Wilcoxon rank sum test. Indices were calculated as the inverse covariance weighted mean of the respective questions (i.e. knowledge, dosage). Changes in yield losses and input spending are self-reported (SR).*

3.3 Results

Among our sample, about 30% of farmers could not receive messages, which points out the relevance of alternative communication channels for agricultural extension. Among the farmers that received the intervention, we found that almost 80% decided to request GAP content, which constitutes a high rate of interaction (Figure 3.2). Table 3.3 shows the estimated intention to treat effect (ITT) and local average treatment effect (LATE) on the aggregated knowledge outcome and selected production-related outcomes. The ITT-estimator refers to the eligible sample and is based on the assigned treatment status (including non-compliers, i.e., farmers that could receive GAP-content but did not request it). The LATE estimator uses GAP-requests as the treatment indicator and therefore mirrors the treatment effect on those intended to be treated and decided to receive treatment (compliers). Our results indicate that the intervention increased complying farmer's knowledge by 9%. Regarding the adoption of recommended practices, we only report the impact on the use of color traps because all other practices were being used by more than 95% of the sample, so they were excluded as specified in the pre-analysis plan. We did not find any significant impact on the probability of using the recommended type and concentration of pesticide for a given pest, but we did find that farmers in the treatment group reported significantly higher pesticide spending.

Table 3.3: Estimated impacts of ICT-intervention on selected outcomes

Outcome	**ITT**			**LATE**			**N**
	Estimate[1]	**SE**	**p-value**	**Estimate**[1]	**SE**	**p-value**	
Knowledge (index)	0.05	0.04	0.15	0.09	0.04	**0.03**	92
Used color traps (y/n)	−0.02	0.08	0.84	−0.04	0.08	0.57	92
Reported any pest (y/n)	−0.11	0.09	0.23	−0.05	0.10	0.59	92
Correct product usage (index)	0.09	0.07	0.15	0.02	0.07	0.76	92
Higher losses (SR; y/n)	0.11	0.09	0.26	0.12	0.10	0.22	92
Higher spending (SR; y/n)	−0.03	0.09	0.77	−0.07	0.09	0.47	92
log Phytosanitary spending [SVC][2]	0.36	0.16	**0.03**	0.52	0.16	**0.00**	92
log Phytosanitary work (hours)	0.01	0.14	0.96	0.10	0.14	0.50	92

[1] *All control variables included.*

[2] *Currency: 1 SVC = 0.1143 USD (31.08.2021). The minimum salary for agricultural workers in El Salvador was approximately 240 US$ per month.*

***Note**: Indices were calculated as the inverse covariance weighted mean of the respective questions (i.e., knowledge, dosage). Changes in yield losses and input spending are self-reported (SR).*

Robustness checks

We compared the outcomes between the control group and never-takers to identify potential spillover effects. Results in Table 3.4 indicates no evidence for spillovers in outcomes. If, due to the limited sample size, we were unable to detect spillover effects, their presence would lead to an underestimation of the measured treatment effect. That is, if famers in the control group indirectly benefitted from the treatment the difference in estimated outcomes would be smaller. In that sense, spillovers may from a practical point of view even be desired.

Table 3.4: Spillover analysis

	ITT spillovers			
Outcome	Estimate	SE	p-value	N
Knowledge index	0.11	0.15	0.44	88
Used color traps	0.47	0.30	0.13	88
Reported any pest	0.07	0.29	0.80	88
Correct product usage (index)	−0.12	0.22	0.58	88
Higher losses (SR)	0.11	0.31	0.72	88
Higher spending (SR)	−0.20	0.34	0.57	88
Phytosanitary spending [SVC]	0.15	0.44	0.74	88
Phytosanitary work (hours)	0.35	0.58	0.54	88

***Note**: Spillovers were estimated as the difference between the control group and never-takers based on Eq. 3.1 where T took the value of 1 for never-takers. All variables reported in Table 3.1 were included as covariates.*

3.4 Discussion

The high rate of interaction with the simplistic chat-bot (i.e. demand for GAP-content by those to whom it was offered) indicates curiosity and openness by most farmers to test such communication channels. We received positive feedback from farmers regarding the pictures with pests since they allow for better identification of adequate pest management strategies. Multimedia content such as pictures and (animated) videos can thus be considered an inclusive tool for informing farmers.

The main results presented in the previous section align with previous studies that found positive impacts of digital information provision on farmer's knowledge on pest management (Larochelle et al. 2019). Contrary to expectations, we did not find evidence that the information intervention changed pest management practices. This is possibly due to the limited sample size and short time between intervention and follow-up survey.

Farmers may require additional time to acquire inputs and adapt their practices. Finally, we estimated a significantly higher pesticide spending in the treatment group; however, based on the observed differences between the control group and never-takers reported in Table 3.2 and Table 3.4, we are skeptical to what extent this estimate reflects an actual treatment effect. From a theoretical point of view, higher spending on pesticides is ex-ante, not an expected outcome of our information intervention, because the intervention did not affect pest incidence. It is, however, possible that treated farmers were more alerted to pests and thus acted either preventively or curatively. Since the messages originate from an input supplier, farmers may perceive the message itself (not its content) as an indicator or reminder of imminent pest risk. If farmers become more sensitive and capable of identifying pests, a reported increase in pesticide-related spending could translate into lower yield losses and overall benefits as long as the gains in yield exceed the additional cost of application. However, such changes may take more than a few weeks to materialize. Since the time between the intervention and follow-up survey was relatively short in this study (between three and six weeks depending on when follow-up was carried out), it is possible that farmers used up their existing stock of products before changing their pest management practices.

Digital divide

Only one farmer in our sample reported having broadband internet access at home; all others relied on mobile network coverage. Depending on available data plans, some multimedia content may be more appropriate than others if the goal is to overcome the digital divide. While farmers without smartphones are exempt from receiving multimedia content, high data volumes of videos could still be a de-facto barrier for information access, especially if perceived benefits are low.

Table 3.5: Probit selection results

	Dependent variable:	
	Smartphone user (y/n)	**Requested GAP (y/n)**
	(1)	(2)
Married (1=yes)	0.69** (0.35)	-2.33* (1.25)
Gender (1=female)	0.72 (0.72)	-4.82 (793.77)
Owns greenhouse (1=yes)	-0.03 (0.39)	-2.97* (1.77)
Has irrigation system (1=yes)	0.40 (0.48)	1.55 (2.54)
Owns tractor (1=yes)	-0.64 (0.87)	-14.51 (43,111.73)
Owns water pipe (1=yes)	-1.28** (0.62)	3.91 (866.31)
Has natural gas (1=yes)	0.63 (0.55)	-5.27 (866.31)
Has cable TV (1=yes)	-0.03 (0.35)	3.39** (1.56)
SE primary market (1=yes)	-0.66 (0.43)	-0.66 (1.61)
With SE > 3 years (1=yes)	-0.06 (0.35)	0.44 (0.86)
Tertiary education (1=yes)	0.21 (0.73)	36.04 (36,037.46)
Secondary education (1=yes)	-0.12 (0.45)	-1.20 (1.55)
Primary education (1=yes)	0.15 (0.37)	-0.08 (1.16)
Income < 1 min. wage (1=yes)	0.86** (0.38)	0.47 (1.10)
Income > 2 min. wages (1=yes)	-0.95 (0.65)	-0.64 (1.94)
Age (years)	-0.03** (0.02)	-0.06 (0.05)
log(Area cultivated) (ha)	2.02*** (0.73)	-0.87 (2.50)
Dependent people (#)	0.03 (0.11)	-0.30 (0.58)
Household size (#)	0.05 (0.10)	0.97 (0.74)
Full-time employees (#)	0.11 (0.12)	0.15 (0.34)
Part-time employees (#)	-0.08 (0.12)	0.54 (0.38)
log(Distance to road) (km)	-0.11 (0.45)	-0.86 (1.33)
Constant	0.96 (0.94)	-1.19 (3.27)
Observations	135	47
Log Likelihood	-66.29	-16.25
Akaike Inf. Crit.	178.57	78.50
Significance levels:	* p<0.1; ** p<0.05; *** p<0.01	

Note: Estimates in column 1 are based on the full sample, while column 2 is based on the subset of farmers that were offered the information (treatment group). We used a probit model to estimate the binary outcomes.

To address the extent to which smartphone-based ICTs overcome or contribute to a digital divide, Table 3.5 shows which characteristics were associated with using ICTs. This was not part of the pre-analysis plan and is therefore considered explorative. Results on the two binary outcome variables 1) whether or not a farmer used a smartphone (column 1) and 2) whether or not farmers actively requested GAP-contents (column 2) are reported. These estimated probit regression coefficients indicate that smartphone use is positively associated with area under cultivation, being married, and having less than one minimum wage of income. Especially the low income and higher area under cultivation are contradictory since they typically correlate, so a characterization of non-users as economically marginalized was not found in this case. At the same time, smartphone use was negatively associated with farmer's age, the latter of which can be expected. Among the farmers in the treatment group offered the GAP-intervention, farmers with cable TV at home were more likely to request GAP-contents. This could point towards the role of digital literacy and differences in utilized information sources. Since this sample is very small and not representative, future studies should assess the uptake of digital information sources by different user groups more systematically and align the targeting of interventions accordingly.

Limitations

The convenience sample employed here may differ from the general population of farmers since they are relatively well-trained and may have better access to inputs and markets. Since this pilot study was conducted among farmers working with the local agribusiness, the results must be extrapolated with caution. While we did control for training duration and many observed farm characteristics, treatment effects may differ for farmers with lower propensities of making a contract with the agribusiness. In particular, our results may not hold for less market-oriented farmers. In addition, the small sample size restricted the statistical power of this analysis. However, from a policy learning perspective, an imprecise measurement is better than no measurement because it reduces uncertainty. Since previous studies focus on different geographic regions (predominantly African countries and India) and different ICT interventions (predominantly SMS, Videos, and interactive voice recordings), this study complements and fills an important research gap. Self-reported outcomes are inherently prone to measurement error, which could explain the lack or counter intuitiveness of production-related impacts. Due to budgetary constraints, it was impossible to collect more detailed information on these outcomes within the scope of this pilot study.

3.5 Conclusion

The diffusion of smartphones among farmers in developing countries offers new possibilities for providing farming-related information that could stimulate the uptake of improved production practices. This study adds to the growing body of literature investigating the impacts of digital advisory services in agriculture. We conducted a randomized control trial among vegetable farmers in El Salvador to study the effect that digital information provision has on farmer’s knowledge and production behavior. Specifically, we implemented a simplistic chatbot that allowed farmers to request multimedia content with recommendations on agricultural practices. Results indicate a 9% increase in knowledge of recommended production practices. There was no evidence for impacts on production outcomes, but this study may also have been statistically underpowered to detect such effects. Future studies should compare different types of demand-driven multimedia content for increased information transfer. Developing more sophisticated chat-bots could be a cost-effective way to provide farmers personalized and locally relevant information. In addition, interactive information flows offer new opportunities for information collection at scale (Bartling et al. 2022). For example, geographically explicit information on crop status and production practices could help farm advisors and input suppliers better accommodate farmers' needs in a timely manner.

In sum, the available evidence from this pilot study points towards promising effects of demand-based multimedia advice. Future research should disentangle the synergistic and complementary effects of information format (image, video, audio), using a larger sample for statistical power and eliminating spillovers.

Data availability

The data and code underlying this article are available upon request.

Acknowledgements

The authors would like to thank the employees of the social enterprise and all study participants for their contributions. The author is grateful for helpful comments on a previous version of the manuscript from Anna Massfeller and support on conceptualization from Jan Börner.

3.6 References

Arouna, A., Michler, J. D., Yergo, W. G., and Saito, K. (2021). One Size Fits All? Experimental Evidence on the Digital Delivery of Personalized Extension Advice in Nigeria. *American Journal of Agricultural Economics*, *103*(2), 596–619.

Bartling, M., Robinson, A. C., Achicanoy Estrella, H., and Eitzinger, A. (2022). The impact of user characteristics of smallholder farmers on user experiences with collaborative map applications. *PloS One*, *17*(3), e0264426.

Baumüller, H. (2018). The Little We Know: An Exploratory Literature Review on the Utility of Mobile Phone-Enabled Services for Smallholder Farmers. *Journal of International Development*, *30*(1), 134–154.

Carrión-Yaguana, V. D., Alwang, J., and Barrera, V. H. (2020). Promoting Behavioral Change Using Text Messages: A Case Study of Blackberry Farmers in Ecuador. *Journal of Agricultural and Applied Economics*, *52*(3), 398–419.

Casaburi, L., Kremer, M., Mullainathan, S., and Ramrattan, R. (2013). *Harnessing ICT to Increase Agricultural production: Evidence from Kenya** (Working Paper).

Duflo, E., Glennester, R., and Kremer, M. (2006). Using Randomization in Development Economics Research: A Toolkit. Cambridge, MA: Center for International Development at Harvard University.

Fabregas, R., Kremer, M., and Schilbach, F. (2019). Realizing the potential of digital development: The case of agricultural advice. *Science (New York, N.Y.)*, *366*(6471).

Larochelle, C., Alwang, J., Travis, E., Barrera, V. H., and Dominguez Andrade, J. M. (2019). Did You Really Get the Message? Using Text

Reminders to Stimulate Adoption of Agricultural Technologies. *The Journal of Development Studies*, *55*(4), 548–564.

Rajkhowa, P., and Qaim, M. (2021). Personalized digital extension services and agricultural performance: Evidence from smallholder farmers in India. *PloS One*, *16*(10), e0259319.

Schulz, D., and Börner, J. (2023). No impact of repeated digital advisory service to Haitian peanut producers. *Q Open*, qoad023.

Tambo, J. A., Aliamo, C., Davis, T., Mugambi, I., Romney, D., Onyango, D. O., … Byantwale, S. T. (2019). The impact of ICT-enabled extension campaign on farmers' knowledge and management of fall armyworm in Uganda. *PloS One*, *14*(8), e0220844.

van Campenhout, B. (2021). The Role of Information in Agricultural Technology Adoption: Experimental Evidence from Rice Farmers in Uganda. *Economic Development and Cultural Change*, *69*(3), 1239–1272.

von Braun, J., Afsana, K., Fresco, L. O., and Hassan, M. (2021). Food systems: seven priorities to end hunger and protect the planet. *Nature*, *597*(7874), 28–30.

Chapter 4
Spatial diffusion of digital farm advisory app across India[12]

4.1 Introduction

Digital innovations are a central element to achieve food security, sustainable development and more resilient agri-food systems (von Braun et al. 2021; Finger 2023). Information and communication technologies such as digital agricultural advisory services have the potential to support farmers' decision making and contribute to higher productivity in low- and middle income countries (Fabregas, Kremer and Schilbach 2019; Spielman et al. 2021; Rajkhowa and Baumüller 2024). To leverage their full potential, it is important to understand adoption patterns. A vast body of literature evolves around the determinants of innovation adoption, highlighting the importance of environmental and socioeconomic context (Munshi 2004; Assunção, Bragança and Hemsley 2019; Schulz and Börner 2022). In

[12] This chapter is currently under peer-review. A previous version has been accepted for oral presentation at the International Conference of Agricultural Economists (ICAE 2024) in New Delhi, India. It is co-authored with Johannes Kopton, and Gustavo Magalhães de Oliveira.

Author contributions: DS: conceptualization, data collection and curation, formal analysis, writing original draft; JK: conceptualization, review and editing the paper; GMO: conceptualization, writing - review and editing the paper.

addition to structural drivers and barriers, inadequate information on the benefits of the innovations and knowledge about how to use them can hinder adoption and then null their potential positive impact (Abate et al. 2023). Therefore, particular attention has been given to the role of social networks on adoption patterns, showing how information shared within families, local communities or via formal extension networks can increase adoption rates in new agricultural technologies (Foster and Rosenzweig 1995; Maertens and Barrett 2013; Burlig and Stevens 2023). While some network studies highlight the role of local peer farmers and extension agents for information and innovation diffusion (Krishnan and Patnam 2014; Genius et al. 2014; Shikuku et al. 2019; Abdulai 2023), there also is empirical evidence for null effects (Duflo, Kremer and Robinson 2011). With respect to the adoption of digital information technologies such as smartphones and agricultural mobile applications, previous studies focus on farmer characteristics, and not on spatial network effects (Michels, Bonke and Musshoff 2020; Thar et al. 2021; Bounkham, Ahmad and Yaseen 2022; Soodan et al. 2023). However, as demonstrated by other forms of innovation (Lapple and Kelley 2015; Graziano and Gillingham 2015), social learning and the structure and composition of neighborhoods may play a key role in the diffusion of digital innovations such as agricultural mobile applications. Assessing neighborhood effects for digital innovations matters because the upfront investment costs associated with this type of digital technology may differ from those of traditional innovations such as new crop varieties. Hence, their adoption and diffusion among risk-averse farmers may differ from costlier innovations. Overall, empirical evidence on the role of contextual as well as socioeconomic drivers on the adoption of agricultural mobile applications remains restricted to a subset of surveyed adopters, limited in regional scope or relying on heavily aggregated spatiotemporal scales such as from census data.

Here, we address these gaps by investigating whether and how contextual socioeconomic factors and spatiotemporal spillover affect the diffusion of *Plantix* in India. We focus on India as a study region due to the high dependence by most of its rural households on agriculture while crop yield losses from insect pests and weeds are substantial (Sharma, Kooner and Arora 2017; Gharde et al. 2018). Existing data indicates that arthropods are responsible for a loss of approximately 18–20% in global annual crop production, with an estimated economic impact surpassing US$470 billion (see Sharma, Kooner and Arora 2017 for an overview). In the Indian agriculture, magnitudes of crop loss attributed to insect pests reach up to 30% in cotton and 25% in rice production (Dhaliwal, Jindal and Mohindru 2015). In this study we utilize the to our knowledge largest proprietary dataset of digital plant health advisory service usage by farmers across continental India. Based on more than 70 million GPS-referenced, timestamped, anonymized, and spatially aggregated uploaded images for pest and disease identification between 2017 and 2023 we characterize the spatial and temporal diffusion of *Plantix*, a free digital farm advisory mobile app widely used in India. *Plantix* aids farmers inter alia with the identification and management of plant diseases, pests, and nutrient deficiencies through the application of image recognition technology.

Our empirical analyses leverage a panel dataset with weekly temporal resolution and country-wide coverage of users in combination with contextual factors obtained from secondary data sources. First, we use a Cox proportional hazard model to estimate the probability of adoption of an agricultural mobile application at time t given that no previous usage was recorded in the region. This analysis quantifies contextual drivers and barriers of diffusion of agricultural mobile applications. Second, using fixed effects regressions, we estimate spatiotemporal spillovers effects at the extensive margin (adoption or non-adoption) and at the intensive margin (number of users) in the context of the agricultural mobile applications. In particular, we investigate the magnitude and duration of spillover effects depending on neighborhood structure, leveraging available information on user mobility to differentiate neighboring effects of local peers (stationary users) and extension agents (mobile users). The remainder of this article is organized as follows. Section 4.2 provides details about the data and explains the methodology and specification of the models. The empirical results and discussion are presented in Section 4.3. Finally, we bring some concluding remarks and limitations in Section 4.4.

4.2 Material and Methods

Our primary data consists of (not publicly available) records of *Plantix* app usage that was anonymised to protect privacy. *Plantix* is a smartphone app for Android, developed by PEAT GmbH, and to our knowledge currently the largest agricultural app worldwide in terms of users. Its core feature which is the focus of this analysis allows users to take a picture of a crop and identify pests and diseases. Other features not included in this analysis are local weather forecasts, fertilization calculator, a user forum, and market information. The app supports English and Hindi, as well as 9 local languages spoken in India. Disease identification is performed using an ensemble of state of the art image classification models including ConvNets and vision transformers and currently supports more than 350 diseases across 55 crops (i.e., 688 crop-disease combinations). Once a picture is taken and uploaded, *Plantix* users receive detailed information about the detected disease along with management recommendations and access to input markets, while the company retains user information and stores all geo-referenced data associated with the images. Since its initiation in 2017, *Plantix* has been used for image-based pest and disease identification on 13 million mobile devices in India, with a clear take-up in adoption since 2019, when user image uploads requests surpassed half a million per month during the Kharif season. As a free application, *Plantix* users incur low initial usage costs (i.e., smartphone and internet). We here rely on a subset of users that regularly used the pest-recognition tool via uploaded images, but exclude users that only used other functionalities of the app such as weather forecast. That is, we define frequent proactive users as those with more than five image uploads within more than one day.

4.2.1 Data aggregation

Our study region is the entire continental India. As identification of nearest neighbors is ambiguous when using a rectangular grid, we use a hexagonal grid to investigate spatial neighborhood effects. Furthermore, the crop disease occurrence which may drive *Plantix* usage is an ecological process and its spatial connectivity better modeled based on a hexagonal grid (Birch, Oom and Beecham 2007). We chose a cell width of four kilometers, resulting in a total of 231,473 grid cells. We discuss implications of the selected cell size below.

We first assess the land use and land cover attributes of the locations where *Plantix* was used (see next subsection). We confirm our intuition that 98.9% of images uploaded by users stem from grid cells containing at least some cropland, and 94.6% from grid cells with at least 10% cropland cover (Figure S2). Second, we calculate the crop-specific disease identification requests from frequent proactive users within the crop-specific growing region. Table 4.1 displays the total number of uploaded images, the number of frequent proactive users and to what extent these pest-identification requests overlap with the production area for 15 crop types matching those available in the secondary data source (Grogan et al. 2022). As shown in Column 5 in Table 4.1, the vast majority of user-supplied images originate from grid cells that overlap with the growing region of the respective crop according to the secondary data source. Our sampling grid is of higher resolution (4 km cell width) than the secondary data (5 arc minutes or approximately 10 km), so our spatial aggregation does not inflate the growing area reported in secondary data. Furthermore, the last column indicates the share of cells covering the respective growing region with at least one frequent proactive user. We display an aggregated vegetable category here[13], but do not use it in our subsequent analysis since it is not directly comparable.

Table 4.1: Plantix usage by crop type

(1) Crop Name	**(2) Image Uploads [1.000]**	**(3) Frequent Proactive Users [1.000]**	**(4) Production area [Mio. km^2]**	**(5) Share of image uploads in production area**	**(6) Share of production area with frequent proactive users**
VEGETABLE	6,997	1,398	2.50	84.4%	72.5%
RICE	2,066	500	2.49	97.0%	46.2%
COTTON	1,799	421	2.80	98.2%	28.5%
MAIZE	377	157	2.35	91.7%	34.4%
PEANUT	320	106	2.10	86.1%	25.2%
SOYBEAN	295	101	1.16	81.4%	28.7%
SUGARCANE	253	86	1.74	86.4%	18.4%
BANANA	233	81	0.66	31.4%	24.7%
POTATO	231	73	2.48	88.9%	16.3%
WHEAT	180	78	1.59	75.1%	28.9%
SORGHUM	45	27	2.82	99.8%	9.6%
MILLET	38	23	2.00	50.2%	6.0%

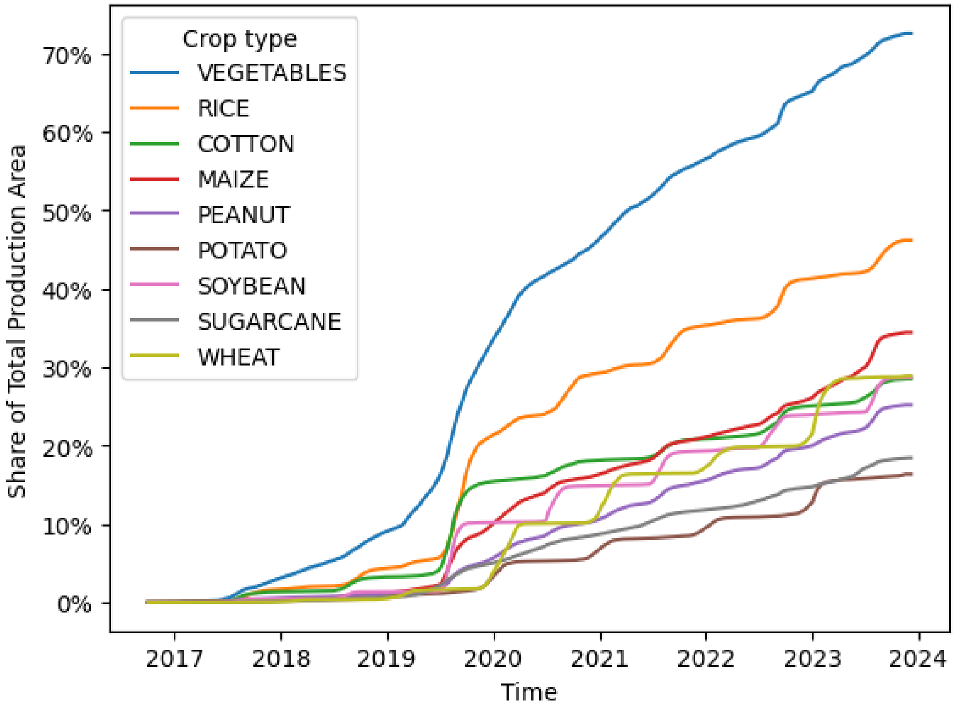
13 The crop category vegetable includes of the following crops: cucumber, pepper, eggplant, tomato, cabbage, pumpkin, onion, cauliflower, zucchini, garlic, okra, bitter gourd, pea, gram, chickpea, lentil, ginger, and turmeric.

TOBACCO	8	4	0.80	70.2%	2.7%
MANIOC	6	3	1.49	78.1%	1.6%
BARLEY	1	1	0.98	64.7%	0.8%

***Note**: Own calculations based on Plantix geo-referenced image uploads and crop-specific growing area (Grogan et al. 2022). Calculations are based on 4 km hexagonal grid cells across continental India (n=231,474). We consider growing area of the respective crop all cells that have any overlap with the growing area data, not only those that are fully covered. We count only Plantix requests fulfilling our minimum quality criteria regarding crop identification confidence (>85%) and location accuracy (<1.000m).*

In addition to the maximum share of production area with active users (Table 4.1, column 6), we also investigated how it developed over time, depicted in Figure 4.1. It shows a clear increase in spatial diffusion after 2019, and the temporal pattern related to how marked the different crop seasons are. For example, rice wheat and soybean show clearly marked seasonal patterns (plateaus followed by steep increases), while maize and peanuts show less pronounced seasonal variation (more monotonous slope).

Figure 4.1: Spatiotemporal innovation diffusion



***Note**: Share of crop-specific growing area (Grogan et al. 2022) with any Plantix image uploads for that crop over time, calculated on a 4 km hexagonal grid. The prominent staircase pattern results from concentrated usage and uptake during the respective main growing seasons. The steep increase after 2019 follows a country-wide marketing campaign.*

4.2.2 Time until adoption analysis

To characterize the spatiotemporal diffusion of *Plantix*, we first investigate the duration between the day it became first available and the active user appearing in a given grid cell. Figure 4.2 shows a non-random pattern indicating that some areas adopted rather early, while others lagged behind or never used it. Therefore, we are interested in exploring the association between contextual factors and the time until technology adoption. To quantify the relationship between contextual factors and *Plantix* diffusion we use spatially explicit secondary data. Based on previous studies pointing towards a possible digital divide in agriculture (Mehrabi et al. 2021), we are particularly interested in socioeconomic contextual factors as proxies of (digital) marginalization. The *Plantix* feature we assess here is conditional on access to the internet at some point to upload images, so we include data on network coverage and number of mobile devices (Ookla 2023). We also use population density as a proxy of potential users and degree of social network density (CIESIN 2017) and travel times to larger cities as a proxy of urban infrastructure and market access (Weiss et al. 2020).

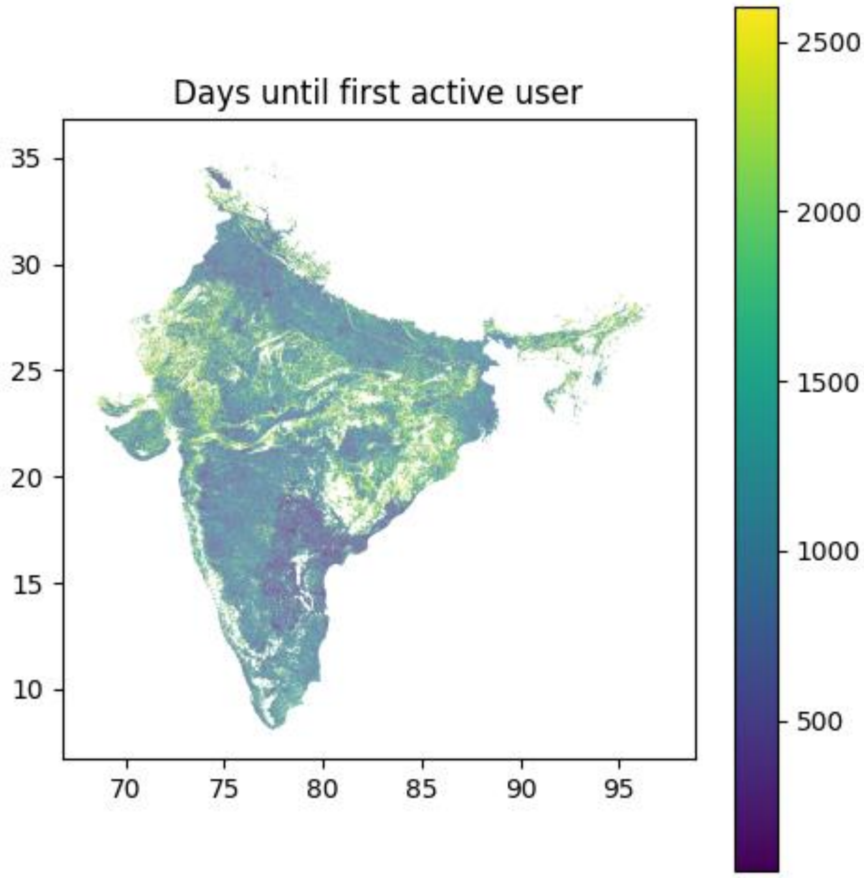
In addition to these key variables of interest, we control for geographic factors that may affect both agricultural production and therefore the potential benefits of using *Plantix*. To identify the potential areas of *Plantix* usage, we use data on land cover such as total cropland area and land use; i.e. production areas of specific crops for the years 2020 and 2015, respectively (Karra et al. 2021; Grogan et al. 2022). As a measure of agricultural suitability we additionally use soil and topography maps (Poggio et al. 2021; Hawker et al. 2022). We expect that crop choice can partially be explained by soil conditions, but also note that soil quality may act as a proxy of land value, which may be linked to farmer characteristics such as income and innovativeness. Since *Plantix* usage is likely to be driven by disease exposure, which – apart from crop and soil characteristics - is linked to climatic conditions, we also include temperature and precipitation data (Abatzoglou et al. 2018). Variable values from these datasets were extracted to our sampling grid. Multiple values within one grid cell were aggregated by taking the area-weighted mean. Given that many of the contextual variables are highly correlated, we perform feature selection by identifying the most important spatial predictors using random survival forest (Pölsterl 2020). Supplementary Table S1 shows summary statistics of all included contextual and control variables.

Empirically, we use time-to-event analysis, also known as duration analysis, on a cross-sectional version of our dataset. Duration analysis has been previously used to study agricultural innovation diffusion (Genius et al. 2014). We use the Cox proportional hazard model, which is given as:

$$h(X) = h_0(t)exp(\beta X) \qquad \text{(Eq. 4.1)}$$

where h(X) is the hazard of the event (first adoption by an active user) occurring in week t given X, h_0 (t) is the baseline hazard in week t, i.e. without any covariates, X are standardized covariates and exp(β) their estimated hazard ratios. The model assumed that the latter are constant over time, which in our case holds since we are interested in time-invariant contextual drivers. We use the non-parametric Breslow method as implemented in the *lifelines* Python package (Davidson-Pilon 2019).

Figure 4.2: Time until first frequent proactive user



Note: *Colors indicate the number of days between September 28, 2016, and the date, on which the first user made more than four image uploads for pest identification on at least one day in any given grid cell. White indicates no such image uploads as of December 2023, but there could be users for other features of the app in these areas.*

4.2.3 Spatial spillover analysis

Spatial spillover effects occur when events are linked by geographic proximity. That is, when the occurrence of an event in one region affects the probability of the event occurring in another region. In our case, at least two distinct mechanisms may contribute to spatiotemporal spillover of *Plantix*.

First, the natural occurrence and dispersion of pests and diseases is driven by interactions between weather, soil and management practices, which tend to be regionally confined. Therefore, the probability of finding a disease is itself geographically clustered. Second, information sharing and social learning among peers or other networks often occurs on a local basis, for example among neighbors or during farmer field days. Motivated by the literature on spatial information spillovers through networks, we assess how the previous use of *Plantix* in the neighborhood is linked to current usage.

Empirically, we use multiple fixed effects regression as implemented in the *fixest* package (Bergé 2018). For the extensive margin, i.e. whether any users used *Plantix* in a given grid cell i at time t, we use a logit model, whereas we use a poisson model for the intensive margin, i.e. the number of active users. For the extensive margin, we use all available observations, for the intensive margin we restrict the data to observations with at least one active user. Both models can then be written as:

$$Y_{it} = \alpha_i + \beta X_{it} + \gamma_{dt} + \epsilon_{it} \qquad \text{(Eq. 4.2)}$$

Where Y_{it} is the binary- or count-outcome value respectively for the extensive and intensive margin in grid cell *i* at time *t*, α_i is a time invariant grid cell fixed effect, γ_{dt} is a time-variant fixed effects interacting district with time, X_{it} is a set of spatiotemporal user lags with its associated coefficient β to be estimated, and ϵ_{it} is an identically distributed error term. We include fixed effects for each grid cell to account for time-invariant characteristics such as soil properties, climate and socioeconomic context factors including internet access, language, education and market access. The time-varying fixed effects capture unobserved characteristics at the district level including weather patterns that affect pest occurrence, regional marketing campaigns, different seasonal growth patterns of crops and district-level extension activities, fluctuations in agricultural commodities, all of which could affect the probability of using *Plantix*. By iteratively adding these fixed effects, we aim to isolate the effect of social networks. In all model specifications based on Eq. (4.2), we cluster our standard errors at the district and week level.

We count the number of unique users in the spatial neighborhood across increasing order by increasing the radius. That is, first order spatial neighbors (O1) refer to the six direct adjacent cells, second order neighbors (O2) to their twelve outer adjacent cells and so on, as illustrated in Figure S3. If a user was active across neighborhood orders, we only take into account the most inner observation to avoid double counting. We then estimate social network effects as the association between digital technology users in surrounding neighborhoods (i.e. order O1, O2, etc.) during previous time periods (i.e., week *t-1*, week *t-2* etc.) and the usage in the target cell in week t. Given the time-sensitive nature of plant health management, we consider four weeks a suitable time frame for effective information sharing. For our main specification we use the first spatiotemporal lag, i.e., the number of active users within the first order neighborhood during the previous week (O1, *t-1*) as our lag variable of interest (*X* in Eq. 4.2). In a second step, we include higher level spatiotemporal lags up to the fifth order to account for broader and longer lasting spillover effects.

We use a 10 kilometer sized grid for our main estimation and provide a higher resolution (i.e. 4 km grid) for a selected state[14] as a robustness check. The reason is that with increasing spatial resolution, the share of cells without any observed users increases and thereby reduces our effective sample size, i.e. cells with variation in the outcome variable. Considering the observed increase in diffusion after 2019, we focus the analysis on the

[14] We focus on Maharashtra in the robustness check, because it is the state with most observations and covered for the longest time period.

2020-2023 period and temporally aggregate the data to weekly time steps, resulting in 209 weeks. The spatiotemporal lag variables are then created by counting the active users during each of the previous four weeks. Summary statistics of all user spatiotemporal lags are shown in supplementary Table S2.

4.2.4 User classification

We estimate how the presence of different types of peers is associated with varying levels of adoption. The subset of frequent proactive users is characterized with respect to their geographic usage pattern by calculating the maximum geographic distances between their requests. We classify these frequent proactive users as a) “stationary” if their app usage radius is less than four kilometers; b) “mobile” if their app usage radius is more than 100 kilometers. In this way, we classify approximately 1.3 million frequent proactive users as stationary and 70,000 as mobile. While farms in India are on average much smaller than four kilometers in diameter (i.e. well below 2 ha), stationary farmers may use the app in a field in walking distance, so the minimum distance should not be too small. The upper cut-off is purely illustrative for identifying potential “innovation ambassadors” such as extension agents or farmer group leaders that visit several farms across a larger geographic area. We discuss implications and possible improvements of these thresholds in Section 4.3. Similar to the main analysis described in Section 4.2, we estimate whether a grid cell had any active users at a given time (binary) and how many active users it had (count), to differentiate adoption at the extensive and intensive margin. To estimate the association between stationary and mobile users during previous weeks, we use their first-order spatiotemporal lag, but counting only the number of stationary and mobile users in neighborhood cells, respectively. Empirically, we again use the multiple fixed effects model (Eq. 4.2).

4.3 Results and Discussion

4.3.1 Time to adoption

Our results from the time-to-event analysis using Eq. 4.1 show that socio-economic factors, in particular travel time to bigger cities and internet speed, are associated with digital innovation diffusion. Table 4.3 shows estimated hazard ratios related to the time until a cell had any request (left) or an active app user (right). We focus on the socioeconomic variables which are relevant for the discussion of the digital divide, while controlling for environmental and other geographic factors. We find that a one standard deviation higher mobile upload speed is associated with a 14% (95% CI: 12-15%) higher hazard ratio for having an active user. Furthermore, a standard deviation increase in travel time is associated with a 17% (95% CI: 15-18%) lower hazard ratio of a grid cell having the first active user. Our estimates for mobile devices and population density show a much smaller impact, as shown in Table 4.3.

These results corroborate some previous findings on agricultural technology adoption. Aparo, Odongo and De Steur (2022) show in their systematic literature review that poor quality mobile phones (battery), poor internet and mobile network connectivity and high cost of internet services are relevant constraints in adoption and usage of mobile phone technologies by farmers. Moreover, despite being a key driver in the diffusion of smartphone applications for agriculture (Michels, Bonke and Musshoff 2020; Thar et al. 2021), mobile internet coverage remains a challenge in rural areas even in developed countries (Michels et al. 2020).

Table 4.2: Results of time-to-event analysis

	Active user				
	Hazard ratio	Lower bound	Upper bound	z	p-value
Contextual variables					
Upload speed [kb/s]	1.14	1.12	1.15	24.40	<0.005
Mobile devices [#]	1.04	1.02	1.06	5	<0.005
Population density	1.00	1.00	1.01	0.93	0.04
Travel time [min]	0.83	0.82	0.85	-16.05	<0.005
Control variables					
Climate	*long term temperature (average, minimum, maximum), mean precipitation*				
Soil	*land cover shares (crop land, trees, built up, rangeland), soil clay content, volumetric fraction of coarse fragments*				
Land cover	*land cover shares (crop land, trees, built up, rangeland)*				
Topography	*slope, elevation*				
State dummies	*binary indicators for each state*				
Observations	231,474				
Events	173,948				
Partial log-likelihood	-1,891,339.31				
Concordance	0.72				
Partial AIC	3,782,776.61				

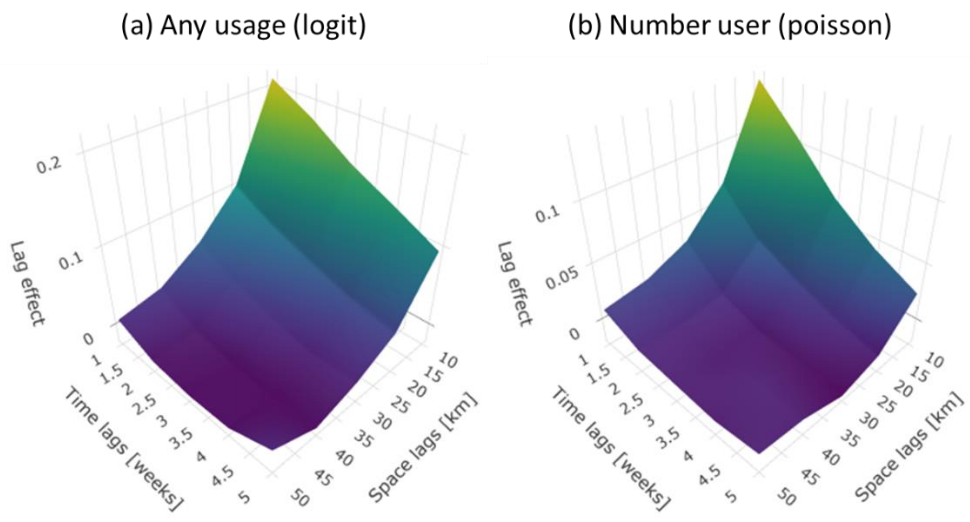
***Note**: Table shows results of own calculations based on Eq. (4.1). Standard errors used for confidence intervals are clustered at the municipal level.*

We interpret our findings as evidence of a spatial digital divide, where farmers in more rural locations with less internet connectivity clearly lag behind in terms of innovation adoption. This underlines the risk of further marginalizing disadvantaged communities. Our results are in line with studies on the global divide in digital farming technologies that show that while the majority of India is covered by 3G and 4G services, only 31% of Indian farming households have internet access (Mehrabi et al. 2021). This holds particular significance since a 2018 survey in India showed that only 6% of farmers received technical advice from agricultural extension workers, while a substantial 70% expressed distrust in the recommendations provided by these workers (Cole and Sharma 2018).

4.3.2 Neighborhood analysis

Our results using Eq. 4.2 show that the number of active users in the first order neighborhood during the previous week is positively associated with the usage and number of frequent proactive users in the target cell at present (t=0) (Table 4.3). Even after the inclusion of approximately 150,000 fixed effects capturing time invariant and time variant unobserved heterogeneity we find a standardized direct neighborhood effect of 0.3 (SE: 0.016) and 0.2 (SE: 0.015) on the usage and number of users in the target cell, respectively (Table 4.3, Model 3 and 6). This effect is large and statistically significant. In a next step, we include higher order spatiotemporal lags, namely up to the fifth neighborhood order and five previous weeks. Our estimated beta coefficients are shown in Figure 4.3 (see SI Tables S3 and S4 for tabular presentation and additional model specifications). The plot shows how the effect diminishes across time and space steps, such that the number of users in the third neighborhood order three weeks before observation is no longer statistically different from zero to explain innovation adoption (0.0124; SE: 0.007). This suggests that spatiotemporal spillover effects diminish over time and are with less than one month relatively short-lived in our case. Our results show a similar pattern although smaller in magnitude for the user count, modelled with a poisson distribution. Furthermore, using a higher spatial resolution of 4 kilometers in a robustness check, we find a similar overall pattern, although the effects on the intensive margin are considerably smaller in magnitude (Tables S5). This can be expected, since the 10 kilometers grid size may approximate actual social network size and mobility patterns more closely. Notably, including higher order lag effects using the smaller grid led to a similar decrease of the effect across time, but not across space (Tables S6 + S7). This corroborates our findings and choice of grid cell size, in that these results show that spatial effects are relatively constant in the close surrounding (<20km) and then drop quickly (20-50km).

Figure 4.3: Spatiotemporal lag effects



***Note**: Both plots show estimated beta coefficients of a fully saturated model, i.e. including all spatiotemporal lags from O1 T1 to O5 T5. Outcome variable is whether any active user was observed at the specific time (plot a), and how many active users were counted (plot b).*

Our findings support previous empirical studies on network effects on technology (Beaman et al. 2021; Abdulai 2023). These studies observe that signals from peer adoption decisions and experiences influence the adoption process through enhanced learning opportunities and reduced uncertainty. However, like most of the existing literature, their analysis focuses on annual windows to assess impact. Murage et al. (2011), for example, argue that farmers, on average, take more than two years to adopt a specific technology after the first learning exposure. Our study, however, illustrates that network effects can be considerably more time-sensitive. Our results suggest that peer effects may diminish dramatically within a month after exposure, highlighting the need for further research using high temporal resolution data.

Table 4.3: Fixed effects structures for lagged neighbour estimate

	Any usage			**Number of user**		
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
(Intercept)	-1.1401 (0.0292)***			0.2643 (0.0213)***		
Active users O1 T1	1.3474 (0.0267)***	0.7082 (0.0316)***	0.3092 (0.0166)***	0.5018 (0.0194)***	0.3854 (0.0189)***	0.2222 (0.0153)***
Num. obs.	7,569,771	6,719,977	6,481,774	2,226,931	2,226,931	2,226,931
Pseudo R^2	0.2393	0.3063	0.3196	0.1373	0.1916	0.2001
Num. groups: grd_id		32,153	32,147		32,153	32,153
Num. groups: year_week		209			209	
Num. groups: year_week × district			117,642			118,940

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: Table shows results of own calculations based on Eq. 4.2. Active user O1 T1 refers to the lagged count of frequent proactive users in the first order neighborhood (O1) in the first order time, i.e. one week before the observed outcome (T1). Model structures are without fixed effects (Model 1 and 4), with two-way grid and time fixed effects (Models 2 and 5), and grid plus time-district interaction fixed effects (Models 3 and 6). Fixed effects clusters without variation in the outcome were dropped, e.g. when all observations within a cell or district × week are equal to zero. Standard errors are clustered at district and week level.

4.3.3 Spillover effects by user type

Finally, we estimate neighborhood spill overs for different user groups and margins. Table 4.4 shows estimated coefficients for the number of stationary and mobile users in neighboring cells during the previous week, respectively. Estimates for the extensive margin relate to whether or not any active users were in a cell. Our standardized estimates indicate a sizeable and statistically significant ($p < 0.001$) association between the two user types and the adoption and number of adopters of the agricultural mobile application. Taken together, these findings illustrate that both types of peers - stationary and mobile - are relevant for the initial adoption of a new technology. Looking at the absolute effect magnitudes of mobile users at the extensive margin (0.043; SE: 0.0025) and intensive margin (0.02; SE: 0.0012), it seems that mobile users play a more important role for initial innovation adoption than for the number of adopters in a given region. This emphasizes the role of mobile peers in spatially diffusing innovation by encouraging the initial adoption of a new technology in new areas. We speculate that mobile users may be extension agents or farm advisors that travel to different farms, while stationary users are mostly farmers using the app on their own land. Overall, however, the estimated effect of stationary users is six to eight times larger in magnitude compared to mobile users. Therefore, stationary peers emerge as seemingly more relevant for the effects at both margins. We interpret these findings as indications for the important role of social ties and trust within local, stationary peers for the uptake and continued use of innovation.

Table 4.4: Neighborhood effect by user group

	Any frequent proactive usage	**Number of frequent proactive user**
Stationary users O1 T1	0.2346 (0.0126)***	0.1695 (0.0115)***
Mobile users O1 T1	0.0426 (0.0025)***	0.0204 (0.0012)***
Num. obs.	6,481,774	2,226,931
Num. groups: grd_id	32,147	32,153
Num. groups: year_week × district	117,642	118,940
Pseudo R^2	0.3192	0.2000

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

***Note**: Stationary users are the number of frequent proactive users whose requests originate within a radius of up to four kilometers. Mobile users are the number of frequent proactive users whose requests have a radius of more than 100 kilometers. O1 T1 refers to the first order neighborhood (O1) in the first order time, i.e. one week before the observed outcome (T1). Standard errors are clustered at district and week level.*

The role of extension agents for innovation adoption is well documented (Schulz and Börner 2022). Yet, our small effect estimate for mobile users supports the idea that trust in agricultural extension services may still be a barrier in technology adoption in India (Cole and Sharma 2018). Our study indicates that both types of ties—local and professional peers — should complement each other to drive agricultural technology diffusion. This view is in line with Fernando (2021), who finds that although mobile extension substantially lower the importance of peers as a source of information, it does not crowd-out peer interactions. Our findings complement previous research showing that the impact of extension agents on information diffusion – although high – diminished over time compared to peer effects (Krishnan and Patnam 2014). We also find that general peer effects diminish over time and space, and that both mobile and stationary users have a sizeable effect which is in line with reported complementarity of peer and extension network effects on innovation diffusion (Genius et al. 2014). Other studies pointed out the importance of information credibility (Abdulai 2023) and potential reputational gain from information sharing (Shikuku et al. 2019) as potential mechanisms. Digital innovations such as free advisory apps may have advantages in this respect. It is easy to recommend them to peers, since no upfront costs of adoption are required, and information quality and credibility (i.e. plant health status) can be assessed directly. Trust-building as a pre-requisite for adapting management decisions based on the gathered information is therefore a helpful trait.

4.4 Conclusion

This study provides empirical evidence on the spatial diffusion of a digital farm advisory app across India. We find that (i) remote regions with lower internet speeds clearly lag behind in terms of time until adoption, indicating a digital divide that should be addressed structurally by policymakers; (ii) neighborhood effects are considerably time-sensitive, diminishing dramatically over short time windows (weeks) in our case; (iii) while both extensive agents and local peers drive the adoption of agricultural innovation, the former is particularly relevant for the initial usage of new technological innovation, and the latter affects both the first use (extensive margin) and frequency of usage (intensive margin).

To the best of our knowledge, this is the first study to leverage a comprehensive dataset of digital technology users across entire India to characterize the spatial and temporal diffusion of innovation. *Plantix* adoption has occurred across most parts of India, with a clear increase in uptake speed since 2020. As of 2023, *Plantix* has been used for cotton, maize, peanut, wheat, soybean, and banana in more than a quarter of the grid cells that grow these crops and almost half of the rice growing grid cells.

One policy implication of our study is the need to improve infrastructure, in particular internet coverage in rural areas, to enable farmers to benefit from digital innovations. Promisingly, there were more new internet users in rural than in urban areas between 2019 and 2021, and the topic continues on the government agenda with ongoing initiatives in the suggested direction such as the Bharat Net project (Government of India 2023). However, access to digital infrastructure can only support those with the capacity to use it, so additional initiatives to foster digital literacy and trust may be needed. Along those lines, existing extension schemes could be spatially redirected to places where digital tools are not available or underutilized. Given the limited funding, physical extension services focusing on plant health may concentrate on regions currently not leveraging the full potential of digital farm advisory tools like *Plantix*.

Our results should be interpreted with the following caveats in mind. First, our analysis concentrates on specific contextual factors, but there may be further sociodemographic aspects worth considering. In particular, access to digital tools may differ by education, income, gender or social status (i.e. cast), as micro-level studies have shown (Thar et al. 2021), and a nuanced consideration thereof may explain more of the adoption patterns. Spatially disaggregated data, e.g. on district level, could shed light on the contribution of these factors, but this is beyond the scope of our work. Second, this study illustrates how information on user mobility can be used to classify them and draw conclusions about their potential roles within information networks. Unfortunately, we were unable to gather evidence on how suitable our classification was to differentiate the user groups. Therefore, the utilized spatiotemporal scales as well as our selected thresholds for differentiating users should be carefully examined. Apart from testing how sensitive our results are to alternative specifications, the data enables to evaluate how temporal and spatial resolution of spatial panel data affects spillover effects across scales by analyzing arbitrary combinations thereof.

Supplementary Information and codes

Supplementary Information containing additional figures and tables as well as the codes to replicate all results can be accessed at https://osf.io/4atku/?view_only=e0ceeb21cb83433ca33ac96e5d55c12a.

Data availability

The raw data is proprietary and sensitive. It can therefore not be made available. However, upon reasonable request, a subset of the aggregated (i.e., pre-processed) data can be made available. Note, that this subset will not allow for exact numerical, but rather qualitatively similar replication of the results.

Acknowledgements

The authors would like to thank Peat GmbH for providing access to their data. In addition, we are grateful to Robert Strey, Thomas Friedel, and Jan Börner for providing insights and comments on a previous version of this manuscript.

4.5 References

Abate, G. T., Abay, K. A., Chamberlin, J., Kassim, Y., Spielman, D. J., and Paul Jr Tabe-Ojong, M. (2023). Digital tools and agricultural market transformation in Africa: Why are they not at scale yet, and what will it take to get there? *Food Policy*, *116*, 102439.

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, *5*(1), 170191.

Abdulai, A. (2023). Information acquisition and the adoption of improved crop varieties. *American Journal of Agricultural Economics*, *105*(4), 1049–1062.

Aparo, N. O., Odongo, W., and De Steur, H. (2022). Unraveling heterogeneity in farmer's adoption of mobile phone technologies: A systematic review. *Technological Forecasting and Social Change*, *185*, 122048.

Assunção, J., Bragança, A., and Hemsley, P. (2019). Geographic Heterogeneity and Technology Adoption: Evidence from Brazil. *Land Economics*, *95*(4), 599–616.

Bandiera, O., and Rasul, I. (2006). Social Networks and Technology Adoption in Northern Mozambique. *The Economic Journal*, *116*(514), 869–902.

Beaman, L., BenYishay, A., Magruder, J., and Mobarak, A. M. (2021). Can Network Theory-Based Targeting Increase Technology Adoption? *American Economic Review*, *111*(6), 1918–1943.

Bergé, L. (2018). Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm. *CREA Discussion Papers*, (13).

Birch, C. P. D., Oom, S. P., and Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling*, *206*(3–4), 347–359.

Bounkham, P., Ahmad, M. M., and Yaseen, M. (2022). Determinants and benefits of using smartphone-based information for vegetable production and marketing: a case of Lao farmers. *Development in Practice*, *32*(7), 981–994.

Burlig, F., and Stevens, A. W. (2023). Social networks and technology adoption: Evidence from church mergers in the U.S. Midwest. *American Journal of Agricultural Economics*, ajae.12429.

CIESIN. (2017). Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11 [Data set]. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC).

Cole, S., and Sharma, G. (2018). The promise and challenges of implementing ICT in indian agriculture. *India Policy Forum. Conference*, *14*(1), 173–240.

Davidson-Pilon, C. (2019). lifelines: survival analysis in Python. *Journal of Open Source Software*, *4*(40), 1317.

Dhaliwal, G. S., Jindal, V., and Mohindru, B. (2015). Crop Losses due to insect pests: Global and Indian Scenario. *Indian Journal of Entomology*, *77*(2), 165.

Duflo, E., Kremer, M., and Robinson, J. (2011). Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya. *American Economic Review*, *101*(6), 2350–2390.

Fabregas, R., Kremer, M., and Schilbach, F. (2019). Realizing the potential of digital development: The case of agricultural advice. *Science (New York, N.Y.)*, *366*(6471).

Fernando, A. N. (2021). Seeking the treated: The impact of mobile extension on farmer information exchange in India. *Journal of Development Economics*, *153*, 102713.

Finger, R. (2023). Digital innovations for sustainable and resilient agricultural systems. *European Review of Agricultural Economics*, *50*(4), 1277–1309.

Foster, A. D., and Rosenzweig, M. R. (1995). Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *Journal of Political Economy*, *103*(6), 1176–1209.

Genius, M., Koundouri, P., Nauges, C., and Tzouvelekas, V. (2014). Information Transmission in Irrigation Technology Adoption and Diffusion: Social Learning, Extension Services, and Spatial Effects. *American Journal of Agricultural Economics*, *96*(1), 328–344.

Gharde, Y., Singh, P. K., Dubey, R. P., and Gupta, P. K. (2018). Assessment of yield and economic losses in agriculture due to weeds in India. *Crop Protection*, *107*, 12–18.

Government of India. (2023). *Economic Survey 2022-23*. New Delhi, India: Ministry of Finance. Department of Economic Affairs.

Graziano, M., and Gillingham, K. (2015). Spatial patterns of solar photovoltaic system adoption: The influence of neighbors and the built environment. *Journal of Economic Geography*, *15*(4), 815–839.

Grogan, D., Frolking, S., Wisser, D., Prusevich, A., and Glidden, S. (2022). Global gridded crop harvested area, production, yield, and monthly physical area data circa 2015. *Scientific Data*, *9*(1), 15.

Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., and Neal, J. (2022). A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, *17*(2), 024016.

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., and Brumby, S. P. (2021). Global land use / land cover with Sentinel 2 and deep learning. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 4704–4707. Brussels, Belgium: IEEE.

Krishnan, P., and Patnam, M. (2014). Neighbors and Extension Agents in Ethiopia: Who Matters More for Technology Adoption? *American Journal of Agricultural Economics*, *96*(1), 308–327.

Lapple, D., and Kelley, H. (2015). Spatial dependence in the adoption of organic drystock farming in Ireland. *European Review of Agricultural Economics*, *42*(2), 315–337.

Maertens, A., and Barrett, C. B. (2013). Measuring Social Networks' Effects on Agricultural Technology Adoption. *American Journal of Agricultural Economics*, *95*(2), 353–359.

Mehrabi, Z., McDowell, M. J., Ricciardi, V., Levers, C., Martinez, J. D., Mehrabi, N., … Jarvis, A. (2021). The global divide in data-driven farming. *Nature Sustainability*, *4*(2), 154–160.

Michels, M., Bonke, V., and Musshoff, O. (2020). Understanding the adoption of smartphone apps in crop protection. *Precision Agriculture*, *21*(6), 1209–1226.

Michels, M., Fecke, W., Feil, J.-H., Musshoff, O., Pigisch, J., and Krone, S. (2020). Smartphone adoption and use in agriculture: empirical evidence from Germany. *Precision Agriculture*, *21*(2), 403–425.

Munshi, K. (2004). Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. *Journal of Development Economics*, *73*(1), 185–213.

Murage, A. W., Obare, G., Chianu, J., Amudavi, D. M., Pickett, J., and Khan, Z. R. (2011). Duration analysis of technology adoption effects of dissemination pathways: A case of 'push–pull' technology for control of Striga weeds and stemborers in Western Kenya. *Crop Protection*, *30*(5), 531–538.

Ookla. (2023). Speedtest by Ookla Global Fixed and Mobile Network Performance Maps [Data set].

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D. (2021). SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL*, *7*(1), 217–240.

Pölsterl, S. (2020). scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn. *Journal of Machine Learning Research*, *21*(212), 1–6.

Rajkhowa, P., and Baumüller, H. (2024). Assessing the potential of ICT to increase land and labour productivity in agriculture: Global and regional perspectives. *Journal of Agricultural Economics*, 1477-9552.12566.

Schulz, D., and Börner, J. (2022). Innovation context and technology traits explain heterogeneity across studies of agricultural technology adoption: A meta-analysis. *Journal of Agricultural Economics*, 1477-9552.12521.

Sharma, S., Kooner, R., and Arora, R. (2017). Insect Pests and Crop Losses. In R. Arora and S. Sandhu, editors, *Breeding Insect Resistant Crops for Sustainable Agriculture*, pages 45–66. Singapore: Springer Singapore.

Shikuku, K. M., Pieters, J., Bulte, E., and Läderach, P. (2019). Incentives and the Diffusion of Agricultural Knowledge: Experimental Evidence from Northern Uganda. *American Journal of Agricultural Economics*, *101*(4), 1164–1180.

Soodan, V., Jamwal, M., Rana, N. P., Sharma, D., and Chakraborty, S. (2023). Modelling the adoption of agro-advisory mobile applications: a theoretical extension and analysis using result demonstrability, trust, self-efficacy and mobile usage proficiency. *Journal of Agribusiness in Developing and Emerging Economies*.

Spielman, D., Lecoutere, E., Makhija, S., and van Campenhout, B. (2021). Information and Communications Technology (ICT) and Agricultural Extension in Developing Countries. *Annual Review of Resource Economics*, *13*(1).

Thar, S. P., Ramilan, T., Farquharson, R. J., Pang, A., and Chen, D. (2021). An empirical analysis of the use of agricultural mobile applications among smallholder farmers in Myanmar. *The Electronic Journal of Information Systems in Developing Countries*, *87*(2), e12159.

von Braun, J., Afsana, K., Fresco, L. O., and Hassan, M. (2021). Food systems: seven priorities to end hunger and protect the planet. *Nature*, *597*(7874), 28–30.

Weiss, D. J., Nelson, A., Vargas-Ruiz, C. A., Gligorić, K., Bavadekar, S., Gabrilovich, E., … Gething, P. W. (2020). Global maps of travel time to healthcare facilities. *Nature Medicine*, *26*(12), 1835–1838.

Chapter 5
Innovation context and technology traits explain heterogeneity across studies of agricultural technology adoption: a meta-analysis[15]

5.1 Introduction

Innovations in agricultural production are essential to achieve global food security, affordable and healthy diets, and more sustainable use of natural resources (Rockström et al., 2017; Herrero et al., 2020; Braun et al., 2021). We conceptualize innovations as technologies and practices that change production factor composition or increase factor productivity. In developing countries, agricultural innovation has often resulted in positive impacts on productivity and food security (Ogundari and Bolarinwa, 2018; Stewart et al., 2015; Gollin et al., 2018), although heterogeneous social and ecological

[15] This chapter is published as: Schulz D, Börner J. 2023. Innovation context and technology traits explain heterogeneity across studies of agricultural technology adoption: A meta-analysis. *Journal of Agricultural Economics*. 74(2):570–90. DOI: 10.1111/1477-9552.12521. A previous version was presented at the 31st International Conference of Agricultural Economists, August 17-31, 2021 (online event).

Author contributions: DS: conceptualization, data collection and curation, formal analysis, writing original draft; JB: conceptualization, review and editing the paper, funding acquisition.

impacts have been reported (Pingali, 2012). Along with other innovations under the umbrella of digitalization and smart farming, remote sensors and robots performing autonomous operations in crop and livestock production are expected to power the next agricultural revolution (Lowenberg-DeBoer, 2015; Barrett and Rose, 2020; Torero, 2021). Therefore, a better understanding of the underlying diffusion patterns is needed to inform future rural development policy and agricultural extension strategies. Designing such strategies for adoption is challenging, especially in agriculture, because adoption depends on a wide range of interacting factors, such as biophysical context, farm structure, decision-maker characteristics, technology attributes, and institutions. The literature on the determinants of agricultural innovation is correspondingly rich in both theoretical and empirical studies from all over the world (Feder et al., 1985; Knowler and Bradshaw, 2007; Foster and Rosenzweig, 2010; Prokopy et al., 2019; Mwangi and Kariuki, 2015). Review studies have so far struggled to produce consistent evidence on the direction and magnitude of adoption determinants (Knowler and Bradshaw, 2007), or have done so with limited generalizability in terms of geography and types of innovation (Baumgart-Getz et al., 2012; Prokopy et al., 2019; Shang et al., 2021; Ruzzante et al., 2021).

We propose a theoretical framing that explicitly considers interactions between innovation traits and geographic contexts. This enables us to derive somewhat more generalizable insights than prior studies based on a meta-regression approach. Using a new comprehensive global data set of adoption studies and correlated hierarchical effects meta-regression analyses, we exploit variation across space and time to investigate how production contexts influence innovation adoption drivers depending on innovation traits.

We contribute to policy design and technology development. First, we estimate the magnitude of various farm-level innovation adoption determinants over a global range of contexts. Second, we quantify how innovation traits and key geographic context factors affect the relative importance of adoption determinants. This knowledge can inform R&D initiatives and policy makers in the design of locally adapted technologies and corresponding dissemination strategies that account for heterogeneous innovation contexts.

Section 5.2 motivates our theoretical framing. Section 5.3 describes the identification and information extraction from primary studies, and documents our empirical framework and secondary data. Section 5.4 presents our meta-regression results. In Section 5.5, we discuss policy implications and limitations of our study before we conclude with avenues for future research.

5.2 Conceptual Framework

Rather than looking at groups of similar innovations separately as other studies have done, we use some of the inherent economic innovation traits across innovation groups to derive more general, theory-informed insights into patterns of adoption. This is justified by prior reviews suggesting that innovations can be categorized meaningfully to relate their adoption determinants to specific traits (Fliegel and Kivlin, 1966; Rubas, 2004; Blair et al., 2021; Arslan et al., 2022) We expand global coverage by including OECD countries and a wider range of innovation traits, thereby adding to the meta-analysis by Ruzzante et al. (2021) who adopted a similar approach.

The induced innovation hypothesis (IIH) suggests that innovation is driven by the quest to use relatively more expensive production factors more efficiently (Hicks, 1932). The set of potential factor-augmenting technologies has been conceptualized as the innovation possibility frontier by early microeconomic theorists (Ahmad, 1966; Binswanger, 1974a; Funk, 2002). The first prominent empirical application by Hayami and Ruttan (1971) and related empirical work in agriculture found partial support for the hypothesis (Binswanger, 1974b; Cowan et al., 2015; Goldman, 1993). Based on improved methods and national datasets in the 1990s, several studies cast doubt on the general validity of the IIH (Olmstead and Rhode, 1993; Liu and Shumway, 2009). Clearly, a comprehensive understanding of innovation processes requires a broader theoretical approach linking micro-level, including behavioural, perspectives with system theories (Edler and Fagerberg, 2017). Still, the induced innovation rationale remains popular as a conceptual framework to motivate thinking about innovation processes in bio-based sectors (Asche and Smith, 2018; Stark et al., 2022).

As with most microeconomic optimization problems, the IIH can be formulated either in terms of highest gain (profit maximization) or in terms of least resistance (cost minimization). As such, technological change is usually factor-augmenting. A rational decision maker facing the choice between innovations augmenting different factors along the innovation possibility frontier (IPF) will choose the innovation that augments the most expensive factor, as it maximizes output (Funk, 2002). Similarly, one could argue that a rational decision maker would choose the technology along the IPF that minimise use of the more expensive factors.

In addition to the production factors (land, labour, and capital) that are commonly used in the literature (Blair et al., 2021; Pardey et al., 2010 we also consider knowhow as a proxy of management skills (Dawson and Lingard, 1982; Huffman, 2020). We make four related propositions:

- P1) The extent to which the farm size determines the adoption of land-intensive innovations is moderated by relative land-abundancy;

- P2) The extent to which labour availability determines the adoption of labour-intensive innovations is moderated by relative labour-abundancy;
- P3) The extent to which capital availability determines the adoption of capital-intensive innovations is moderated by relative capital-abundancy;
- and P4) The extent to which knowhow determines the adoption of knowhow-intensive innovations is moderated by relative knowhow-abundancy.

Framing the IIH in terms of factor intensities rather than relative factor prices allows us to use globally available data and a theoretically motivated classification of innovation options in our empirical approach below. Accordingly, we do not claim to test IIH directly – rather we seek to provide complementary economic evidence to explain adoption patterns of agricultural technologies at global scale.

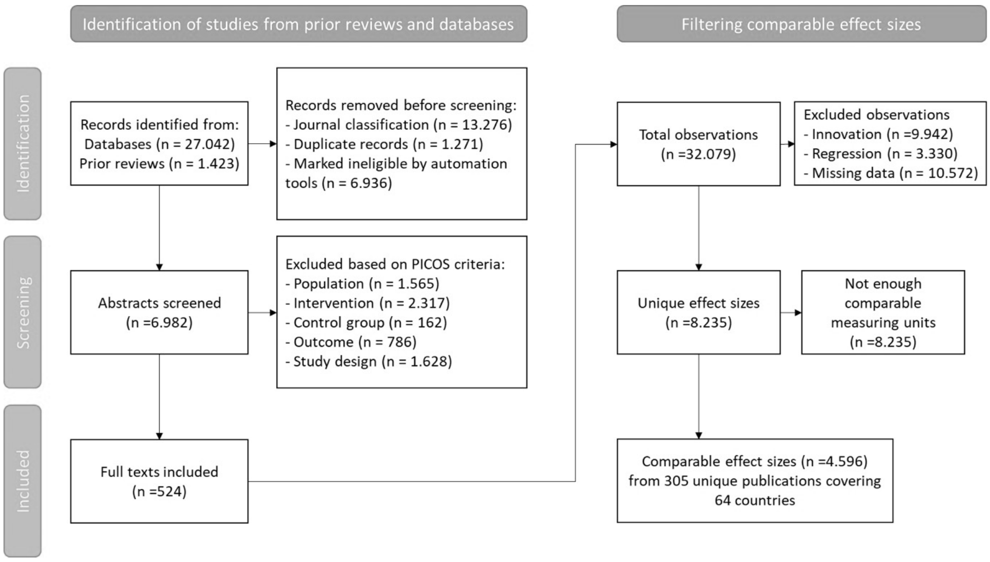
5.3 Materials and Methods

5.3.1 Primary Data Collection

We closely followed the guidelines for meta-analyses in economics by Havránek et al. (2020). A database containing agricultural innovation adoption determinants from prior studies was created in five steps (see Supplementary Information Text S1 and Text S2, on-line). First, we gathered and assessed the eligibility of 1,423 adoption studies from the reference lists of prior reviews (Table S1). Second, we followed Grames et al. (2019) and used text mining on the eligible studies to derive a data-driven systematic search string before we retrieved a total of 27,043 peer-reviewed articles from three literature databases, namely Web of Knowledge, EBSCOhost and AgEcon. Third, with the support of automation tools to prioritize relevant abstracts and titles, we screened all unique records according to the eligibility criteria presented in Table S2, on-line. Fourth, we extracted and coded the results of 534 randomly selected[16] primary studies along with meta data into a detailed spreadsheet, following Stanley and Doucouliagos (2012) and Floress et al. (2019). We thus base our analysis on a convenience sample of the innovation adoption literature similar to prior reviews (Oca Munguia and Llewellyn, 2020; Ruzzante et al., 2021). Apart from the estimated adoption coefficients and their precision estimates, we collected sample characteristics such as sample size, mean and standard deviation of independent variables, distribution of adopters/non-adopters, information about empirical specifications (e.g., logit, probit), and dependent variable characteristics (e.g. scale and innovation description). Fifth, we categorized all innovations and adoption determinants and expanded the approach of Floress et al. (2019) by including detailed information on measurement units, for example whether farm size was defined as total farm size or area cultivated, measured in hectares, acres, or a (non-) linear transformation of the same. An extended PRISMA diagram (Page et al., 2020) with the number of studies that were excluded at each stage of the screening process along with the filtering process of comparable effect sizes is shown in Figure 5.1.

[16] We expect that not all relevant studies were identified by our approach, but do not expect that non-identified studies differ systematically from the identified ones.

Figure 5.1: Extended PRISMA diagram of included studies and effect sizes



5.3.2 *Effect sizes*

The primary data for this study are estimated log odds ratios of adoption determinants, which can be used in meta-analysis without further standardization (Cooper et al., 2009; Stanley and Doucouliagos, 2012). As a measure of precision, this study used the variance of the log odds ratio, calculated from the standard errors, t-statistics, p-values or p-significance thresholds (typically coded as stars) depending on availability. Although we recognize that the majority of our observed effects is neither causal nor unbiased, we assume these estimates to be unbiased on average based on the central limit theorem applying to large samples (see Text S3 for further discussion).

Meta-regression relies on the condition that observations (effects) are measured in a homogeneous manner. We thus carefully ensure the comparability of adoption determinants by using a fine-grained categorization procedure and rigorous filtering. A total of 32079 beta coefficients of agricultural innovation adoption determinants were extracted from 524 unique studies (see SI Full list of included studies, on-line). Out of these, 22137 were eligible based on the reported outcome (Text S6, Table S3) and categorized into 42 categories of adoption determinants (Text S5, Table S4). For comparability, only studies using logit or probit estimation methods are included in this analysis, further reducing the number of observations to 18807. These restrictions could introduce a bias against lumpy innovations or those that can be partially or dynamically adopted, because their adoption is typically not studied as a binary outcome (Doss, 2006; Pannell and Claassen, 2020). We found no drastic differences between the frequency of innovations before and after filtering (Figure S4), but recognize that this check fails to account for studies that did not meet our PICOS criteria. Since some primary studies did not report test statistics or only effect estimates that could not be converted to log odds ratios, comparable effect size estimates and their variance could be calculated for 8235 observations. To ensure comparability, effect sizes within each category of adoption determinants were grouped into the respective measurement units whenever we could obtain sufficiently detailed information. For this analysis, we only used measures of adoption determinants that were used by at least five different studies.

5.3.3 Empirical Framework

5.3.3.1 Aggregation of dependent effect sizes

Meta-analysis without moderators is used to estimate a weighted mean for each adoption determinant, where the weights are inversely related to the variance. We used the estimated log odds ratios as the outcome measure and employed multilevel random effects models with robust variance estimation (RVE). Doing so requires us to deal with non-independent effects and correlated sampling errors. The correlated and hierarchical effects (CHE) model described by Pustejovsky and Tipton (2021) addresses these types of dependencies and can be written as follows for the average effect:

$$y_{ij} = \beta_0 + u_j + v_{ij} + e_{ij} \qquad \text{(Eq. 5.1)}$$

Where y_{ij} is the ith effect size (innovation) in study j (i=1…m, j=1…k), β_0 is the average population effect, u_j are study-level random effect with variance σ_1^2 (between study variance), v_{ij} are observation-level random effects with variation σ_2^2, and e_{ij} are the known sampling variances of the respective effect sizes with variance s_j^2 and Cov(e_{ij}, e_{ij})= ρs_j^2 where we assume a constant correlation[17] among estimates from the same study of ρ=0.5. The unknown variance components σ_1^2 and σ_2^2 are estimated using the Restricted Maximum Likelihood estimator (Viechtbauer, 2005).

5.3.3.2 Induced innovation: meta-regression framework

We use interaction terms between the country- and time-specific factor endowments and innovation-specific factor intensities to test the propositions outlined in Section 5.2. Table 5.1 provides an overview of the dependent variables (adoption determinants), the factor intensities assigned as binary variables to each innovation and the proxies for factor abundance used in the analysis. For the binary trait-indicators, we developed a coding scheme with predefined criteria to assign factor intensities. Four trained analysts independently assigned all innovation traits to all innovations based on the coding scheme, reaching a final inter-coder agreement of 96% (see SI Text S6, on-line, for further details). The selection of context indicators was informed by pragmatic criteria of comparability and availability across countries; we discuss the implications below. We consider quantity ratios

[17] The simplifying assumption of having the same constant correlation of outcomes within studies was taken because with the available data we were not able to model heteroscedastic variances.

adequate because they provide an intuitive proxy of relative scarcity, reflecting the material conditions of production, while being less sensitive to agricultural policies than price ratios, in the short term.

Empirically, interaction terms along with a set of control variables were added to the CHE model so that the extended model can be written as

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 C_{ij} + \beta_3 C_{ij} * T_{ij} + \beta_4 M_{ij} + u_j + v_{ij} + e_{ij} \qquad \text{(Eq. 5.2)}$$

Where β_0 is an intercept, β_1 the estimated coefficient for the factor intensity dummy T, β_2 the coefficients for country-year specific factor abundancies C, $\beta_3$the coefficient of the interaction between factor abundance and factor intensity, β_4 the coefficients for additional control moderators M; u_j v_{ij} and e_{ij} are defined as in Eq. 5.1. The aggregated effect was considered economically meaningful when its estimated 95% confidence interval did not include zero. To better interpret the magnitude, the aggregated log odds ratios were transformed to odds ratios. Results from estimating Eq. 5.2 must be interpreted with care given that context factors (C) may be endogenous to technology adoption. For example, if mechanisation reduces labour requirements, fewer workers per hectare of cropland are needed. We address this issue by adopting measures of context factors that were taken before levels of adoption were measured in the studies that enter our meta-analyses. Moreover, these studies largely focus on innovations at subnational scale in relatively early stages of dissemination, which are unlikely to affect context factors measured at national scales. That said, we do not claim to have found a strategy that rigorously identifies the causal effect of context factors on adoption factors, but expect to find plausible correlations. We show correlations and geographic distribution of each context factor in the on-line Figures S5 and S6, respectively. Summary statistics of all independent variables and adoption determinants are reported in on-line Tables S5 and S6, respectively.

Table 5.1: Definition of dependent and independent variables

Dependent variables		**Independent variables**	
Adoption determinant	Scale & measuring units	Innovation factor intensity (traits)	Geographic factor abundance (context)
Land	Continuous variables: hectares or acres of total farm size or area under cultivation	T1: Land intensity (i.e. 1 for contour farming, buffer strips, agroforestry, conservation practices, organic farming, 0 for all other)	C1: Land-abundance: (1) Log of hectares of cropland equivalent per worker (Fuglie 2012)
Labour	Continuous variables: number of women, men, adults or household members	T2: Labour intensity (i.e. 1 for permanent cover, contour farming, buffer strips, agroforestry, conservation practices, fertilizer, non-chemical pest control, nutrient intensity optimization, organic farming, soil analysis, 0 for all other)	C2: Labour-abundance: (1) Share of workforce employed in agriculture (ILO 2021)
Capital	Binary variables: access or use of formal credit	T3: Capital intensity (i.e. 1 for buffer strips, agroforestry, fertilizer, non-chemical pest control, chemical pest control, soil analysis, mechanization, precision farming analysis support, precision farming interventions, improved seeds,	C3: Capital-abundance: (1) Log of agricultural machinery stock per hectare of cropland

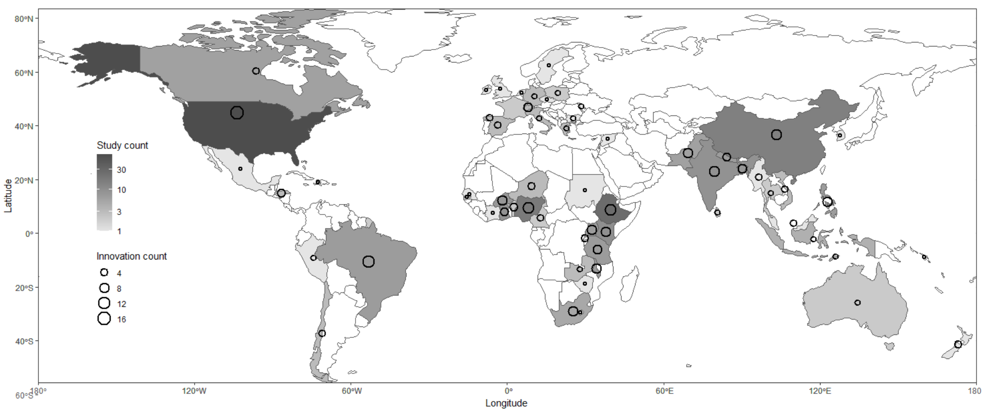
Dependent variables		Independent variables	
Adoption determinant	Scale & measuring units	Innovation factor intensity (traits)	Geographic factor abundance (context)
		GMOs, crop insurance, 0 for all other)	equivalents (Fuglie 2012)
Knowhow	Binary variables: access and use of traditional extension services	T4: Knowhow intensity (i.e. 1 for permanent cover, agroforestry, reduced tillage, conservation practices, non-chemical pest control, nutrient intensity optimization, chemical pest control, organic farming, soil analysis, analysis support for precision farming, contract farming, crop insurance, 0 for all other)	C4: Knowhow-abundance: (1) Education index (Smits & Permanyer 2019)

***Note**: Dependent variables (adoption determinants) and their measuring units were extracted from primary studies. Innovation traits were assigned to each innovation category based on predefined criteria. Country-level indicators of factor abundance were obtained for the year of data collection of the primary study. The indicator listed under (1) are our primary set of context variables.*

5.3.3.3 Robustness checks and publication bias assessment

We checked the robustness of our estimates by consecutively adding sets of control variables These were: 1) other innovation traits, 2) context indicators, 3) study- and regression characteristics, and 4) dummies indicating whether the primary study controlled for other selected adoption drivers such as assets, education, or income. Our main results are based on the full set of control variables. In the random effects model, true population effects may differ even in the absence of sampling error. We therefore tested within each outcome, whether the effect sizes belong to different populations by testing the significance of the Q statistic using a χ^2 distribution (Hedges and Olkin, 1985). We tested for the presence of publication bias using Egger's regression test with a significance threshold of $p=0.10$ (Egger et al., 1997; Sterne and Egger, 2005). Results of the main regressions with moderators after excluding influential observations are reported as robustness checks. Potentially influential observations were identified using Cook's distances larger than four standard deviations. As a constant correlation between estimates from the same study, we assumed a value of 0.5 and conducted a sensitivity analysis by varying this value between 0 and 1. We also tested whether results could be driven by studies that provide multiple estimates of the adoption of the same innovation (i.e. different model specifications) by considering only one average estimate per study (Gleser and Olkin, 1994). Finally, we tested whether results were sensitive to the choice of context-indicator by using an alternative set of context variables given in the SI.

Figure 5.2: Geographic distribution of comparable studies and innovation



__Note__: The colour of the country indicates the number of adoption studies in the respective country, while the size of the circles indicates for how many different innovations in a given *country adoption determinants were estimated. In the US, for example, we found 68 studies reporting adoption determinants for 16 different innovations.*

The analysis was conducted using the *metafor* package (Viechtbauer, 2010) and *clubSandwich* package (Pustejovsky, 2020) for R (R Core Team, 2020). Further information including summary statistics, variable descriptions, robustness checks, publication bias assessment, and a full list of included studies are provided in the online supplementary material.

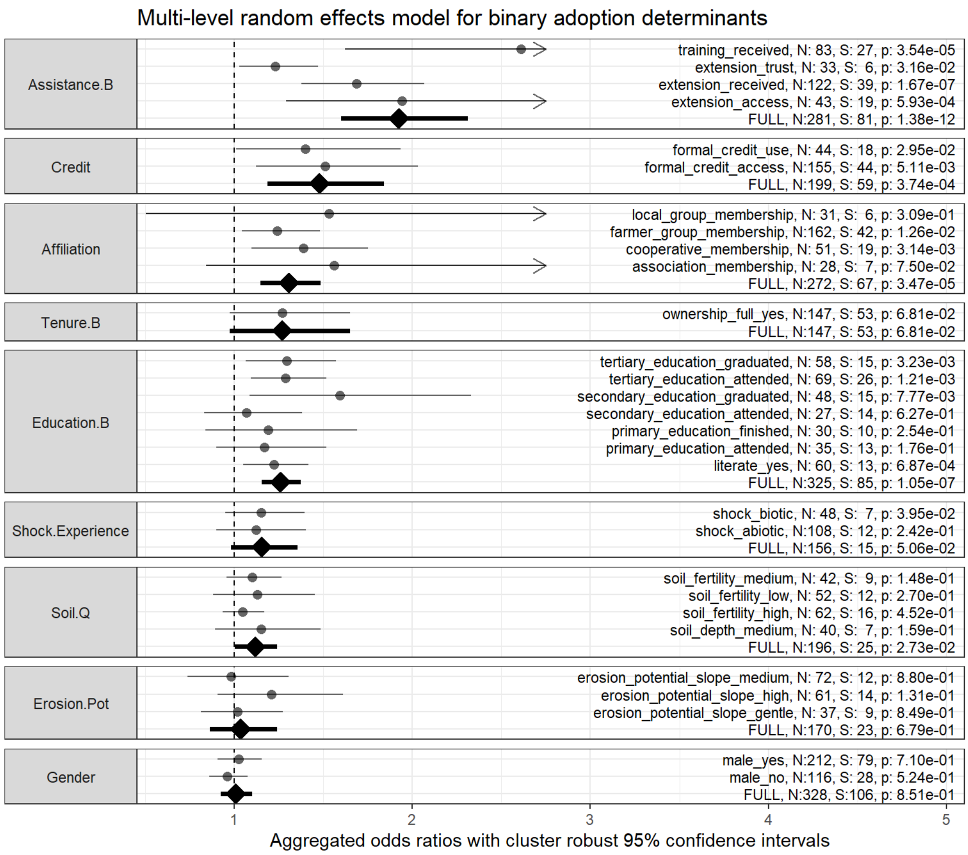
5.4 Results

This study synthesizes a total of 4596 estimated beta coefficients of innovation adoption determinants. They originate from 305 unique publications, of which 257 report results for a specific region, 46 at the country level, and 2 across countries. Figure 5.2 illustrates the geographic distribution of studies and innovations. Most studies focus on the United States, suggesting a potential language selection bias because we incorporated part of the data from Floress et al. (2019). In terms of number of observations, Sub-Saharan countries, predominantly Ethiopia, were strongly represented in the adoption literature, whereas Latin America, Europe, and Oceania are underrepresented in our dataset (see Table S6, online).

5.4.1 Effect Size Aggregation

Figure 5.3 and Figure 5.4 show the average odds ratios for comparable categories of binary and continuous adoption determinants respectively along with their robust 95% confidence intervals for all measuring units with at least 20 observations. The columns on the right indicate the number of effect sizes used for the estimate (N), the number of studies from which these effects were extracted (S) and the p-value indicating whether the estimated intercept significantly differed from zero. Odds ratios can be interpreted as changes in the odds of adopting the innovation against the reference of one, all else being equal. For example, binary variables indicating that extension services were received have an average odds ratio of 1.69, which translates to an increase of 69% (95% CI: 37.6-106.7%) in the odds of adoption. Similarly, binary variables indicating access to and use of formal credit were grouped together in the FULL model specification, resulting in an average increase in the odds of adoption by 48% (95% CI: 18.7-84.1%).

Figure 5.3: Weighted mean odds ratios (OR) for binary adoption determinants grouped by measuring unit

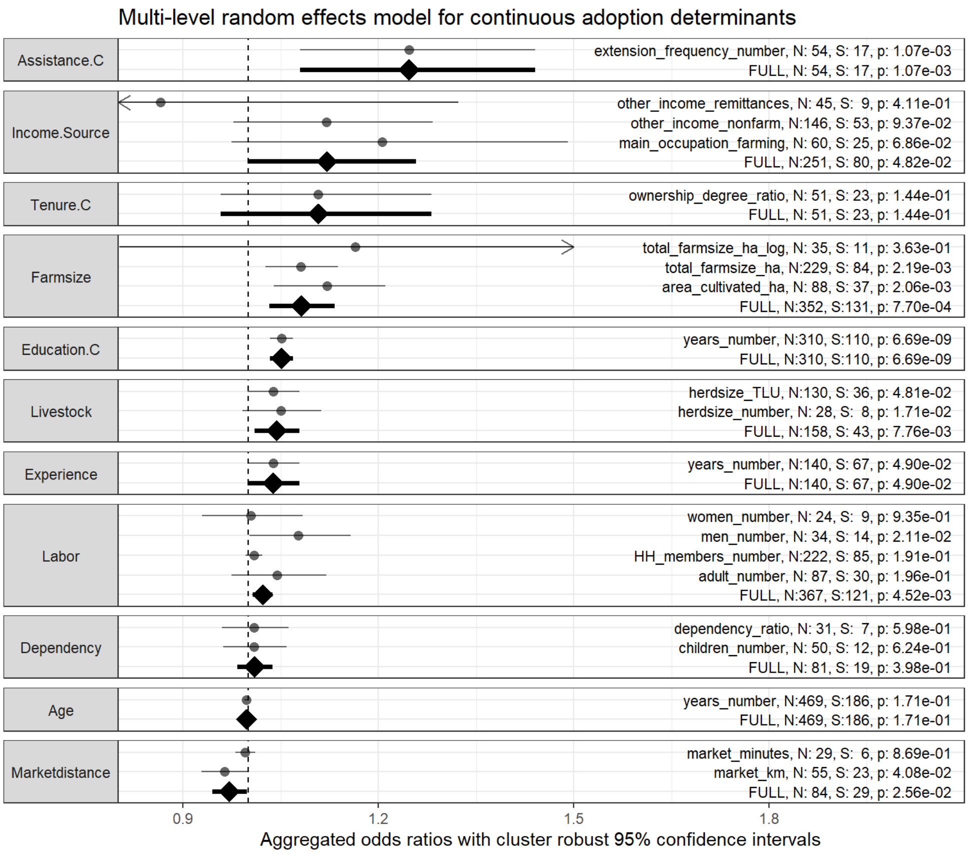


***Note**: Within each category (grey boxes on left side) we report separate regressions for each measuring unit (written on right side). The bold estimates (FULL) combine all* *measuring units with the respective category. N= number of observations, S=number of studies, p=Satterthwaite p-value of OR being equal to one. Dotted line indicates zero effect, namely odds of adoption equal odds of non-adoption. For exact numerical representation of estimates and additional model statistics see Table S7, on-line.*

The only adoption determinants that are consistently (i.e. for all measuring units) and significantly (i.e. $p<0.1$) different from zero were Assistance (binary and cont.), Credit (binary), Tenure (binary), Education (cont.), Experience, Livestock, while farm size – although generally positive and significant - exhibited substantial variation when measured in log hectare unit. These findings are more conclusive than those of previous vote-count analyses (Knowler and Bradshaw, 2007; Shang et al., 2021) Other commonly used determinants such as gender and age were not found to significantly differ from zero on average; the latter is contrary to findings by Baumgart-Getz et al. (2012), who focussed on North America. Market distance was the only measure negatively associated with adoption. In general, our findings are in line with those by Ruzzante et al. (2021), as expected. Binary measures tended to have larger magnitudes than related continuous measures. For example, having graduated from a university increases the odds of adoption by 29% (95% CI: 6.6-57.2%), while one additional year of education has an effect of 5% (95% CI: 3.3-6.9%). At the same time, variables measured on a continuous scale tended to have a lower variance. Binary measures must be interpreted with caution avoiding conclusions about relative magnitudes, because we could not control for the reference categories since this would drastically reduce the sample size. If tertiary school attainment is compared to secondary school attainment, one can expect a lower magnitude than when it is compared to another baseline category, for example, having received no primary education, which may be the case in developing countries. Hence, our estimates of categorical adoption determinants should be interpreted as upper-bound estimates.

Notably, even within the relatively fine-grained outcome measures, all estimates still have significant residual heterogeneity (p<0.01) (reported in Table S7, supplementary material on-line). We thus proceed to meta-regression analysis and assess whether moderators can explain this heterogeneity.

Figure 5.4 Weighted mean odds ratios for continuous adoption determinants grouped by measuring unit:



***Note**: Within each category (grey boxes on left side) we report separate regressions for each measuring unit (written on right side). The bold estimates (FULL) combine all measuring units with the respective category. N= number of observations, S=number of studies, p=Satterthwaite p-value of OR being equal to one. Dotted line indicates zero effect, namely odds of adoption equal odds of non-adoption. For exact numerical representation of estimates and additional model statistics see Table S7, on-line.*

5.4.2 *Revisiting the Induced Innovation Hypothesis*

The meta-regression results presented in Table 5.2 show the interaction effects between innovation specific factor intensity and country specific factor abundance for the four adoption categories land, labour, and capital and knowhow. The innovation traits T1-T4 are assigned the value of one if the innovations use the respective factor intensively, and zero otherwise, while the context factors C1-C4 are continuous measures (see Table 5.1 for details).

Table 5.2: Interaction effects of factor intensity and factor abundance for Land, Labour, Capital and Knowhow

	Farmsize	**Labour**	**Credit**	**Assistance.B**
Innovation traits				
T1: land-intensive	0.24***	0.00	0.00	0.21
	(0.07)	(0.02)	(0.12)	(0.17)
T2: labour-intensive	-0.10**	0.03	0.06	-0.12
	(0.07)	(0.06)	(0.12)	(0.11)
T3: capital-intensive	0.10*	-0.02	-0.48**	0.87***
	(0.06)	(0.03)	(0.25)	(0.34)
T4: knowhow-intensive	-0.03	0.02	-0.15	1.47***
	(0.04)	(0.02)	(0.11)	(0.54)
Context indicators				
C1: land-abundance	0.03	-0.06**	-0.25	-0.01
	(0.04)	(0.04)	(0.35)	(0.15)
C2: labour-abundance	0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.02)	(0.01)
C3: capital-abundance	0.00	-0.00	0.08	0.08
	(0.02)	(0.01)	(0.09)	(0.09)
C4: knowhow-abundance	-0.05	-0.14	1.18	0.24
	(0.25)	(0.11)	(2.36)	(1.16)
Interaction terms				
T1 * C1	-0.07***			
	(0.03)			
T2 * C2		-0.00		
		(0.00)		
T3 * C3			-0.12***	
			(0.05)	

	Farmsize	Labour	Credit	Assistance.B
T4 * C4				-1.79*
				(1.06)
Constant	0.11	0.26**	0.99	0.35
	(0.24)	(0.12)	(2.13)	(1.36)
Regression Type	Yes	Yes	Yes	Yes
Measurement Units	Yes	Yes	Yes	Yes
Model Specification	Yes	Yes	Yes	Yes
sigma2.1	0.04	0.00	1.06	0.18
sigma2.2	0.06	0.01	0.14	0.28
cochran.qe	5470.97	4085.72	1660.44	1154.95
p.value.cochran.qe	0	0	0.00	0.00
cochran.qm	41.89	25.75	21.03	31.78
p.value.cochran.qm	0.01	0.26	0.40	0.08
df.residual	343	338	172	165
logLik	-187.32	96.70	-190.79	-191.84
deviance	374.63	-193.41	381.59	383.68
AIC	428.63	-143.41	427.59	433.68
BIC	532.25	-47.83	499.98	511.33
AICc	433.43	-139.24	435.04	443.04
observations	368	361	193	188

*** p < 0.01;** p < 0.05;* p < 0.1

***Note**: Column labels indicate dependent variable, i.e. adoption determinant as specified in Table 5.1. Innovation traits (T1-T4) refer to binary variables of factor intensity, while context indicators (C1-C4) indicate factor abundance for land, labour, capital, and knowhow, respectively (see Table 5.1 for details). A set of control dummies accounts for model specifications in primary studies: regression type (logit, probit), scale of dependent variable (binary and multivariate), observation level (plot or farm), spatial level (regional or national); model specification dummies for whether the original model controlled for farm size, labour, credit, assets, income and education or not, and in case of farm size whether a squared term was included in the primary regression. Distribution of measurement units for each regression is shown in Table S5, on-line. Brackets contain robust standard errors clustered at the study-level. The sigmas refer to estimated variation components between studies (σ_1^2) and within study (σ_2^2). Cochrane test for residual heterogeneity (QE) and its significance (QEp) as well as omnibus moderator test statistic (QM) and its significance (QMp) are reported.*

We found significant (p<0.1) negative interaction terms for the proxies of land, capital and knowhow, i.e. farm size, credit access/use, and extension access. A negative interaction term indicates that a higher value of the context indicator is associated with a lower odds ratio of adopting innovations that use the interacted factor intensively. We did not find a significant interaction effect on labour, but note that the estimated between-study heterogeneity in true effects (σ_1^2) was close to zero, so that potential moderators that could explain this variation would have an economically rather insignificant magnitude. The omnibus moderator tests were marginally significant for farm size (p=0.08) and extension (p=0.08).

5.4.3 Robustness

Regarding the aggregated effects (Figure 5.3, Figure 5.4), Egger's regression test indicated no evidence for publication bias. However, the Q-statistic indicated significant heterogeneity in the true effects for all estimated average effects, which we attribute to differences in innovation, sample and study characteristics (Table S7, on-line). Hence, even though the average odds ratio is significantly greater than one, the distribution of true effects estimated by the random effects model may include effects smaller than one.

The QE-test for residual heterogeneity in the regression reported in Table 5.2 remained highly significant after the inclusion of all moderators. The moderators included in this analysis thus only explain a part of the variation in true effects. The interaction effects for farm size and credit shown in Table 2 remain stable after the exclusion of potentially influential studies identified via Cook's distance (Table S10, on-line). As shown in Table S8 and Table S9 both estimates and their p-values are sensitive to the assumed within-study correlation of estimates; at an extreme hypothetical intra-study correlation of 1 the effects are only marginally significant. Yet, the sensitivity analysis supports our belief that the within-study correlation of effects plays an important role and should be modelled accordingly. Our interaction effect estimates are robust to a variety of model specifications and when considering only one estimate per innovation per study (Table

S12-16). An alternative set of context variables produced similar magnitudes and directions of estimated interaction effects (Table S11), but the coefficients were no longer significant. The original set of context indicators was chosen to minimize the number of missing observations; for the alternative context variables much fewer observations are available, which explains the reduced statistical power.

5.5 Discussion

We found large and significant positive average effects for binary adoption determinants related to assistance, credit access and group affiliation. We also found statistically significant positive average effects, albeit an order of magnitude smaller, for continuously measured determinants related to years of formal education, experience, and livestock herd size. These findings are broadly in line with prior meta-analyses. Ruzzante et al. (2021), for example, reported similar effect directions, with only minor differences in absolute magnitudes. They used a different effect size measure (the partial correlation coefficient) and included both binary and continuous adoption measures, which may partially explain the difference.

We further found no evidence for a uniform effect of age or gender on adoption. We found mostly positive associations for labour endowment, farm size, risk preferences, and tenure status as adoption determinants, although some measuring units did not indicate a significant average impact. Importantly, we show that some of these factors matter more or less under a selected set of contextual conditions that reflect factor abundance and corresponding technology traits. Since direction and magnitudes of adoption determinants have been extensively discussed in earlier reviews (see Section 5.1 and Table S1, on-line), we focus here on the results of our moderation analysis. Regarding the interaction effect of innovation factor intensity and context factor abundance, we found that our propositions (Section 5.2), based on induced innovation can explain some of the variation in true effects across countries and innovations.

5.5.1 Land

Consistent with proposition P1, our results suggest that the extent to which land-availability at farm-level determines the adoption of land-intensive innovations decreases with increasing land-abundance. Our interpretation of land (farm size) as an adoption determinant deviates from Ruzzante et al. (2021) in that we do not interpret the positive sign as sufficient evidence for increasing returns to scale of the innovation. Prior studies have emphasized the role of fixed transaction costs involved in changing the production system. Often a critical scale of operation is needed to overcome an innovation threshold (Foster and Rosenzweig, 2017). Relying on fixed transaction costs as an impact channel for farm size is therefore consistent with the idea that the positive effect of farm size on adoption reflects economies of scale. Differences in the farm size estimate would then be attributable to different marginal cost of implementing the innovation. However, the fact that larger farms are more likely to adopt innovations may also relate to alternative mechanisms such as affluence, risk-affinity, management style, or bargaining power – which are likely endogenously linked to farm size.

Ruzzante et al. conjecture based on their findings that NRM-technologies may be implemented in a capital-intensive fashion in the USA, whereas labour-intensive implementation would dominate in developing countries. Indeed, within-innovation differences in factor intensities could explain heterogeneity in adoption determinants at various scales. Since our data allows us to empirically test for these relationships, we focus on the between-innovation variation in factor intensities by assuming each innovation to be homogenous with respect to factor intensities. Our theoretical framework based on induced innovation (IIH) provides a mechanistic interpretation of the relationship between factor abundance and factor intensity and for the macro-scale interpretation, it does not matter whether the variation occurs within or between innovations.

5.5.2 Labour

We did not find a statistically significant interaction effect for the factor labour. This may be due to a lack of statistical power to explain very small variations (see Figure 5.4). Furthermore, household size related variables are typically included in adoption studies as a proxy for farm labour usage in the presence of imperfect labour markets. Under functioning labour markets, the size of the household is not expected to have any influence on the farm labour usage, since labour supply and labour demand of the farm household are separable (Benjamin, 1992). Instead of low wages as a reason for farm labour being less of a driver for innovation, labour supply may actually be low due to imperfect markets, even though the country is labour abundant. Thus, the findings may point towards a discrepancy between labour abundance and actual labour supply. However, our data did not allow for tests with other farm labour indicators or an assessment of the role of seasonal fluctuations in labour availability.

5.5.3 Capital

Consistent with proposition P3, we find that a change in access to formal credit has a relatively smaller effect on the adoption decision in capital abundant contexts. This must be interpreted with caution because our observations are limited to non-OECD countries, which is not surprising given that capital markets work comparatively well in OECD countries and credit access is thus virtually never considered. This is in line with the sizeable impact of access to capital on US-agricultural productivity during the first half of the 20th century, when rural capital markets were less developed (Hutchins, 2022). Data limitations did not allow us to test the effect on adoption determinants such as debt-asset-ratio, which is more commonly measured in capital abundant OECD-countries. To corroborate our findings with respect to capital, we report additional moderation analyses for tenure status, livestock and gender in

Table 5.3. Following Arslan et al. (2020), we also used livestock (herd size measured in tropical livestock units or total heads) as a proxy of capital because of its function as collateral[18]. We tested whether the effect of this capital proxy is moderated by capital-intensity and abundance (Column 4 in Table 5.3) and found a positive interaction effect indicating that livestock is relatively more important for adopting capital-intensive innovations in capital-abundant settings. This result may seem at odds with the IIH. But, livestock often serves as a collateral and its effect is then indirect and mediated mainly by the availability of (in-) formal capital markets, which tend to be more developed in capital-abundant settings. In addition, we report results for tenure status (1=being full owner, 0 otherwise) and gender (1=male, 0=female). Similar to livestock, the effect of tenure status is positively influenced by capital abundance in the study context independent of capital intensity of the innovation. That is, being a full owner and having more livestock is more important where capital is available (and land can be employed as collateral on capital markets).

Interestingly, none of the context or trait-related variables were significantly associated with the estimated coefficients for gender as an adoption determinant. Although insignificant, we cautiously interpret the negative point estimate in the gender regression as a sign that identifying as male may be more important for the adoption of innovations in capital scarce contexts. This is intuitive considering the gender-based differences in access to capital and underlines to continued need for inclusive (esp. gender-sensitive) financial institutions.

[18] For formal credit, livestock has been argued to be a poor collateral for being prone to theft and disease (Binswanger & McIntire 1987), but some microcredit institutions do accept it nowadays (Chapoto & Aboagye 2017).

5.5.4 Knowhow

The large magnitude of the assistance effect (Figure 5.3) highlights that agricultural extension plays an important role in the innovation adoption process, even though a lack of accountability, performance gaps and distributional shortcomings were highlighted in the literature on agricultural knowledge systems (Anderson et al., 2006; Norton and Alwang, 2020). While we cannot rule out that some primary studies labelled "extension" relates to any type of professional in-person knowledge transfer, we assume that extension predominantly relates to the more traditional "Train & Visit" approaches. Such approaches were subject to criticism in favour of more bottom-up approaches (Chambers, 1998; Scoones and Thompson, 1994; Scoones et al., 2009), but our estimates show that they effectively enhance the adoption of agricultural innovations. Clearly, bottom-up approaches could have been even more effective in doing so and may have beneficial effects beyond promoting technology adoption.

Table 5.3: Additional regression results for adoption determinants related to capital

	Tenure.B	Tenure.B	Livestock	Livestock	Gender	Gender
Innovation traits						
T1: land-intensive	0.08	0.08	-0.00	-0.01	-0.07	-0.06
	(0.12)	(0.14)	(0.04)	(0.03)	(0.18)	(0.18)
T2: labour-intensive	0.10	0.10	0.04	0.04*	-0.24*	-0.25**
	(0.15)	(0.16)	(0.03)	(0.03)	(0.15)	(0.15)
T3: capital-intensive	-0.09	-0.10	0.04	0.18***	-0.28*	-0.39*
	(0.23)	(0.31)	(0.03)	(0.05)	(0.12)	(0.17)
T4: knowhow-intensive	0.21	0.21	0.00	0.01	0.05	0.05
	(0.19)	(0.22)	(0.03)	(0.03)	(0.14)	(0.14)
Context indicators						
C1: land-abundance	0.01	0.01	0.04	0.03	0.01	0.00
	(0.31)	(0.31)	(0.04)	(0.04)	(0.09)	(0.09)
C2: labour-abundance	-0.02	-0.02	-0.00	-0.00	0.00	0.01
	(0.02)	(0.02)	(0.00)	(0.00)	(0.01)	(0.01)
C3: capital-abundance	0.17	0.17	0.05***	0.04**	-0.04	-0.02
	(0.12)	(0.13)	(0.02)	(0.02)	(0.05)	(0.06)
C4: knowhow-abundance	-4.01*	-4.00*	-0.38	-0.36	0.53	0.49
	(1.55)	(1.55)	(0.21)	(0.24)	(0.95)	(0.94)
Interaction terms						
T3 * C3		-0.00		0.04***		-0.04
		(0.11)		(0.01)		(0.05)
Constant	4.94***	4.94***	0.22	0.23	0.27	0.35
	(2.72)	(2.71)	(0.17)	(0.18)	(0.99)	(0.98)
Regression Type	Yes	Yes	Yes	Yes	Yes	Yes
Measurement Units	Yes	Yes	Yes	Yes	Yes	Yes
Model Specification	Yes	Yes	Yes	Yes	Yes	Yes
sigma2.1	0.89	0.89	0.00	0.00	0.00	0.00
sigma2.2	0.15	0.15	0.01	0.01	0.48	0.48
cochran.qe	1415.33	1411.90	1674.14	1637.12	4274.09	4268.68
p.value.cochran.qe	0.00	0.00	0.00	0.00	0	0
cochran.qm	32.06	31.89	32.98	43.73	30.39	30.92
p.value.cochran.qm	0.04	0.06	0.03	0.00	0.08	0.10
df.residual	126	125	139	138	299	298
logLik	-135.55	-135.11	102.01	106.56	-373.77	-372.77
deviance	271.11	270.23	-204.02	-213.12	747.54	745.55

	Tenure.B	Tenure.B	Livestock	Livestock	Gender	Gender
AIC	317.11	318.23	-158.02	-165.12	795.54	795.55
BIC	382.34	386.11	-90.53	-94.86	884.35	887.97
AICc	327.93	330.23	-148.42	-154.50	799.92	800.33
Observations	147	147	160	160	321	321

*** $p < 0.01$;** $p < 0.05$;* $p < 0.1$

***Note**: Innovation traits (T1-T4) refer to binary variables indicating land intensive, labour intensive, capital intensive, and knowhow intensive, respectively (see Table 5.1 for details). A set of control dummies accounts for model specifications in primary studies: regression type (logit, probit), scale of dependent variable (binary and multivariate), whether the original model controlled for other independent variable categories or not, observation level (plot or farm), spatial level (regional or national). Brackets contain cluster robust standard errors. The sigmas refer to estimated variation components between studies (σ_1^2) and within study (σ_2^2). Cochrane test for residual heterogeneity (QE) and its significance (QEp) as well as omnibus moderator test statistic (QM) and its significance (QMp) are reported.*

We interpret the results in Table 5.2 (column 4) in favour our proposition P4, namely that extension as a proxy of knowhow positively influenced the adoption of knowhow-intensive innovations especially in knowhow scarce contexts. Our finding contradicts Ruzzante et al. (2021), who found that the education level in the same context was negatively associated with the effect extension has on the adoption of improved seeds (not knowhow-intensive), but positively associated with the effect it has on the adoption of natural resource management (knowhow-intensive). This is a surprise because we based our calculation on the same education index. In addition, we found that extension was more effective for the adoption of capital- and knowhow-intensive innovations opposed to land and labour-intensive ones. We speculate that our finding can be explained by factor mobility: capital and knowhow tend to move more freely than land and family labour (Binswanger and McIntire, 1987), making it easier for extension agents to successfully advocate capital and knowhow-intensive innovations. The finding may also indicate effectiveness of the facilitating role knowledge networks such as extension agents or associations play in promoting access to rural credit (Linh et al., 2019; Carrer et al., 2020; Balana and Oyeyemi, 2022).

Two factors limit our confidence in the coefficient of the corresponding interaction term. First, we did not find consistent results using alternative proxies of knowhow, namely experience and formal education (Table S17, on-line). The estimates of the interaction terms are also inconsistent across model specifications, outcomes, and particularly sensitive to the choice of alternative context indicators. As with labour, this may have to do with the difficulty of decomposing very limited heterogeneity in true effects. Second, the relative focus on non-OECD-countries in our sample, once again, limit generalizability to the global level. In OECD-countries, assistance is often measured by the presence of (self-paid) advisors rather than publicly financed extension agents. Including observations with this assistance proxy in the regressions led to inconsistent results. We excluded these observations due to the difference in their total heterogeneity (Figure 5.3: rows 1, 3, and 4); the corresponding dummy would have been correlated with the error term and thus biased our estimates.

5.5.5 Limitations

Our meta-analysis was constrained by the diversity of empirical strategies, (partially) unreported results, and notably by a lack of consistency in the measurement of commonly used adoption determinants. Overcoming comparability related issues by rigorously filtering out non-comparable observations and controlling for the exact measurement units increased the geographical imbalance in our final dataset. Our categorization of innovations and consecutive assignment of factor intensities did not account for potential heterogeneity of factor intensities, especially when endogenously influenced by the geographic context (see also Section

5.3.3.2). In addition, there may be a pre-existing geographic bias in terms of the innovations (and thus factor intensities) under study and thus covered in the literature. The same holds true for measurement scales and units. For example, capital was commonly measured as access to credit in developing countries and as debt-asset ratio in industrial countries.

Finally, the available context indicators may not be optimal for testing theory-based propositions. In addition, our indicators did not capture distributional asymmetries of context factors within countries, although we included the Gini-coefficient as a control (Table S13-16, column 7, on-line). Other proxies such as factor price ratios could be more intuitive in the context of induced innovation and would facilitate interpretability for policy makers, but to the best of our knowledge such data do not exist with global coverage.

5.5.6 Future Directions

Abstracting from specific innovations in terms of innovation traits merits closer attention both in meta-analyses and primary studies, because they may facilitate transferability of research findings. Our results point towards potential transferability of past research findings by abstracting traits and applying known (or assumed) combinations thereof to future innovations. The factor intensities employed in this study represent only a subset of distinct innovation traits and other dimensions should be considered more systematically. For the case of agricultural robots, for example, capital- and knowhow intensity may be useful traits, but they should be complemented by inherent impacts (e.g. environmental footprint), attributes relevant for social learning diffusion processes (e.g. observability) as well as differentiation between labour-augmenting and labour-replacing innovation (Marinoudi et al., 2019). For the *ex-ante* diffusion assessment of new technologies, a trait-based uncertainty reduction of adoption determinants could provide important insights. Interacting innovation characteristics and the affinity of innovators towards these characteristics has been proposed as a mediation mechanism in the ADOPT model by Kuehne et al. (2017; 2011). Our estimated ranges of odds ratios for the most commonly used categories of adoption determinants may serve as credible input ranges in agent-based models for modelling diffusion patterns of digital agricultural information (Shang et al., 2021). Finally, many environmental indicators (e.g. potential productivity or vulnerability to climate change impacts) as well as socioeconomic context indicators (e.g. population density, land prices, access to digital infrastructure) are available on subnational scales. Therefore, a promising avenue for policy-oriented future research would be a more regional analysis of certain subsets in terms of innovation and/or context to better understand the role of within-country production context variation.

Previously identified challenges for generalization include the various definitions of adoption, measures of adoption determinants, and representations of temporal adoption dynamics (Doss, 2006). Our findings therefore suggest that established minimum standards for agricultural adoption studies are needed to extract further generalized lessons from this important subfield in agricultural economics. Below we provide a non-exhaustive list of practical recommendations towards this goal, complementing previous attempts to create reporting guidelines for adoption studies. At a minimum, authors of adoption studies should:

1. Report all estimated effects in tabular format along with a measure of their sampling error independent of their significance.
2. Indicate model specifications and variables that were used in the regression, but omitted in the results table to save space.

3. Indicate any (non-)linear transformation of variables. Independent variables should be measured in or converted to the International System of Units (i.e. hectares, tons, years), where applicable.
4. Report the number of observations for each regression. Especially for data structures with multiple observations per individual (e.g. panel data, multiple plots), the unit of observation should be clearly indicated.
5. Provide summary statistics in tabular format and include at least the mean and standard deviation (proportion for binary variables) of all (in-)dependent variables for the entire sample as well as for different subgroups (e.g. adopters and non-adopters).
6. Provide a description of (a) the study area(s), (b) the innovation(s) considered, such as claimed advantages, historical adoption levels, and (c) the sample characteristics in terms of market orientation (e.g. subsistence vs. commercial), product specialization (e.g. rice farmers, mixed livestock farmers etc.).
7. Make preferential use of continuous independent variables as such and not recode them into categorical or ordinal scales.

Of course, following such guidelines is subordinate to a rigorous research design that contributes to better understand behaviour, as well as the role of gender, innovation characteristics and digitalization in agriculture (Pannell and Claassen, 2020).

5.6 Policy Implications

Innovation in agricultural production remains one of the most important strategic pillars in the transformation towards sustainable food systems, as pointed out during the UN Food Systems Summit (Braun et al., 2021).

Despite the large number of existing adoption studies worldwide, we still poorly understand why apparently beneficial agricultural technologies suffer from low or stagnating uptake. Here we have systematically taken stock of the existing mostly context-specific empirical evidence and found that agricultural knowledge and extension systems as well as alleviation of credit constraints may deserve more attention than they currently receive, especially in the developing world. In particular, our findings warrant more emphasis on the design of policies and interventions that improve technical knowledge, skills, and capital access.

For example, agricultural extension programmes could boost the uptake of new technologies by aligning dissemination strategies with regionally heterogeneous target group characteristics and agricultural factor scarcities. As digital technologies become increasingly available to farmers worldwide, the importance of technical knowledge and skills as adoption determinants will grow. Digital literacy thus also has to feature more prominently in the curricula of rural training and education programmes.

Agricultural extension is also often the vehicle for rural credit programmes. Our findings suggest considerable synergies from packaging agricultural extension and rural credit lines, such that they coherently promote technologies with attributes that address region and farm-specific output and input market constraints. Considering spatially heterogeneous endowments and access to production factors across prospective user groups may also allow technology developers to better tailor future innovations to local needs.

Beyond market-related factor scarcities, environmental policies are likely to play an increasingly important role as drivers of innovation in the transformation toward sustainable and climate-resilient agriculture (Ambec et al., 2011). For example, conservation policies that limit access to land in a context of land abundance, were shown to induce agricultural intensification (Koch et al., 2019). Similarly, smart environmental regulation could increase the attractiveness of eco-efficient technologies, such as weeding robots, if contextual variability in the abundance of other production factors were properly taken into account. A successful digital transformation thus implies increased and interdisciplinary collaboration between new and traditional stakeholders of agricultural knowledge systems in order to avoid innovation system failure (Hermans et al., 2015) .

Finally, the temporal dynamics of context factors imply that forward-looking policy design must be informed by a structural understanding of the embeddedness of production systems. This, at the same time, warrants great caution in the transfer of research findings across space and time, because differences in geographic context and accelerating climate change impacts clearly influence sample-specific findings of adoption studies. If we want to leverage innovation to overcome food system challenges, policy must take into account their context-specific (dis-) advantages and recognize macro-structural barriers of innovation diffusion.

Acknowledgements

The authors would like to acknowledge the contributions of Philipp Feisthauer, Uyen Tran, Abosede Jenrola and Ekaterina Teleshkan in the data collection process for this study. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2070 – 390732324. We are also grateful for the constructive comments of anonymous reviewers on an earlier draft. Open Access funding enabled and organized by Projekt DEAL.

Supplementary Information

Additional information including details on study selection, data extraction, summary statistics, description of variables, robustness checks, and the full list of included studies is provided in the online supplementary information (https://doi.org/10.1111/1477-9552.12521).

Data availability

The data and code to replicate all findings are available upon request.

5.7 References

Ahmad, S. (1966). On the Theory of Induced Invention. *The Economic Journal*, *76*(302), 344.

Ambec, S., Cohen, M. A., Elgie, S., and Lanoie, P. (2011). The Porter Hypothesis at 20: Can Environmental Regulation Enhance Innovation and Competitiveness? *SSRN Electronic Journal*, *41*(4), 797.

Anderson, J. R., Feder, G., and Ganguly, S. (2006, May). The Rise and Fall of Training and Visit Extension: An Asian Mini-drama with an African Epilogue. Washington, DC: World Bank.

Arslan, A., Floress, K., Lamanna, C., Lipper, L., and Rosenstock, T. S. (2022). A meta-analysis of the adoption of agricultural technology in Sub-Saharan Africa. *PLOS Sustainability and Transformation*, *1*(7), e0000018.

Arslan, A., Floress, K. M., Lamanna, C., Lipper, L., Asfaw, S., and Rosenstock, T. (2020). The adoption of improved agricultural technologies: A meta-analysis for Africa. IFAD.

Asche, F., and Smith, M. D. (2018). Viewpoint: Induced Innovation in Fisheries and Aquaculture. *Food Policy*, *76*(1), 1–7.

Balana, B. B., and Oyeyemi, M. A. (2022). Agricultural credit constraints in smallholder farming in developing countries: Evidence from Nigeria. *World Development Sustainability*, *1*(4), 100012.

Barrett, H., and Rose, D. C. (2020). Perceptions of the Fourth Agricultural Revolution: What's In, What's Out, and What Consequences are Anticipated? *Sociologia Ruralis*, *44*(1), 90.

Baumgart-Getz, A., Prokopy, L. S., and Floress, K. M. (2012). Why farmers adopt best management practice in the United States: a meta-analysis of the adoption literature. *Journal of Environmental Management*, *96*(1), 17–25.

Benjamin, D. (1992). Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models. *Econometrica*, *60*(2), 287.

Binswanger, H. P. (1974a). A microeconomic approach to induced innovation. *The Economic Journal*, *84*(336), 940–958.

Binswanger, H. P. (1974b). The Measurement of Technical Change Biases with Many Factors of Production. *The American Economic Review*, *64*(6), 964–976.

Binswanger, H. P., and McIntire, J. (1987). Behavioral and material determinants of production relations in land-abundant tropical agriculture. *Economic Development and Cultural Change*, *36*(1), 73–99.

Blair, G., Christensen, D., and Rudkin, A. (2021). Do Commodity Price Shocks Cause Armed Conflict? A Meta-Analysis of Natural Experiments. *American Political Science Review*, *115*(2), 709–716.

Carrer, M. J., Maia, A. G., de Mello Brandão Vinholis, M., and de Souza Filho, H. M. (2020). Assessing the effectiveness of rural credit policy on the adoption of integrated crop-livestock systems in Brazil. *Land Use Policy*, *92*(4), 104468.

Chambers, R. (Ed.). (1998). *Farmer first: Farmer innovation and agricultural research*. London: Intermediate Technology Publications.

Chapoto, T., and Aboagye, A. Q. Q. (2017). African innovations in harnessing farmer assets as collateral. *African Journal of Economic and Management Studies*, *8*(1), 66–75.

Cooper, H., Hedges, L. V., and Valentine, J. C. (Eds.). (2009). *Handbook of Research Synthesis and Meta-Analysis, The*. New York: Russell Sage Foundation.

Cowan, B. W., Lee, D., and Shumway, C. R. (2015). The Induced Innovation Hypothesis and U.S. Public Agricultural Research. *American Journal of Agricultural Economics*, *97*(3), 727–742.

Dawson, P. J., and Lingard, J. (1982). Management bias and returns to scale in a Cobb-Douglas production function for agriculture. *European Review of Agricultural Economics*, *9*(1), 7–24.

Doss, C. R. (2006). Analyzing technology adoption using microstudies: limitations, challenges, and opportunities for improvement. *Agricultural Economics*, *34*(3), 207–219.

Edler, J., and Fagerberg, J. (2017). Innovation policy: what, why, and how. *Oxford Review of Economic Policy*, *33*(1), 2–23.

Egger, M., Davey Smith, G., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical Research Ed.)*, *315*(7109), 629–634.

Feder, G., Just, R. E., and Zilberman, D. (1985). Adoption of Agricultural Innovations in Developing Countries: A Survey. *Economic Development and Cultural Change*, *33*(2), 255–298.

Fliegel, F. C., and Kivlin, J. E. (1966). Attributes of Innovations as Factors in Diffusion. *American Journal of Sociology*, *72*(3), 235–248.

Floress, K. M., Gao, Y., Gramig, B. M., Arbuckle, J. G., Church, S. P., Eanes, F. R., … Prokopy, L. S. (2019). *Meta-analytic data from agricultural conservation practice adoption research in the United States 1982-2018*. Fort Collins, CO: Forest Service Research Data Archive.

Foster, A. D., and Rosenzweig, M. R. (2010). Microeconomics of Technology Adoption. *Annual Review of Economics*, *2*.

Foster, A. D., and Rosenzweig, M. R. (2017). Are There Too Many Farms in the World? Labor-Market Transaction Costs, Machine Capacities and Optimal Farm Size. Yale: Economic Growth Center.

Fuglie, K. O. (2012). Productivity Growth and Technology Capital in the Global Agricultural Economy. In K. O. Fuglie, S. L. Wang, and V. E. Ball, editors, *Productivity growth in agriculture*, pages 335–368. Wallingford: CABI.

Funk, P. (2002). Induced Innovation Revisited. *Econometrica*, *69*(273), 155–171.

Gleser, L. J., and Olkin, I. (1994). Stochastically dependent effect sizes. In Cooper, H. & Hedges, L. V., editor, *The handbook of research synthesis*, pages 339–355. New York, NY, US: Russell Sage Foundation.

Goldman, A. (1993). Agricultural Innovation in Three Areas of Kenya: Neo-Boserupian Theories and Regional Characterization. *Economic Geography*, *69*(1), 44–71.

Gollin, D., Hansen, C. W., and Wingender, A. (2018, June). Two Blades of Grass: The Impact of the Green Revolution. Cambridge, MA.

Grames, E. M., Stillman, A. N., Tingley, M. W., and Elphick, C. S. (2019). An automated approach to identifying search terms for systematic reviews using keyword co–occurrence networks. *Methods in Ecology and Evolution*, *10*(10), 1645–1654.

Havránek, T., Stanley, T. D., Doucouliagos, H., Bom, P., Geyer–Klingeberg, J., Iwasaki, I., … Aert, R. C. M. (2020). Reporting guidelines for meta-analysis in economics. *Journal of Economic Surveys*, *34*(3), 469–475.

Hayami, Y., and Ruttan, V. W. (1971). *Agricultural development: An international perspective*. Baltimore: Johns Hopkins University Press.

Hedges, L. V., and Olkin, I. (1985). *Statistical Method for Meta-Analysis*. Burlington: Elsevier Science.

Hermans, F., Klerkx, L., and Roep, D. (2015). Structural Conditions for Collaboration and Learning in Innovation Networks: Using an Innovation System Performance Lens to Analyse Agricultural Knowledge Systems. *The Journal of Agricultural Education and Extension*, *21*(1), 35–54.

Herrero, M., Thornton, P. K., Mason-D'Croz, D., Palmer, J., Benton, T. G., Bodirsky, B. L., … West, P. C. (2020). Innovation can accelerate the transition towards a sustainable food system. *Nature Food*, *1*(5), 266–272.

Hicks, J. R. (1932). *The Theory of Wages*. London: Palgrave Macmillan UK.

Huffman, W. E. (2020). Human Capital and Adoption of Innovations: Policy Implications. *Applied Economic Perspectives and Policy*, *42*(1), 92–99.

Hutchins, J. (2022). The US farm credit system and agricultural development: Evidence from an early expansion, 1920–1940. *American Journal of Agricultural Economics*, *23*(7), 2757.

ILO. (2021). Employment in agriculture (% of total employment) (modeled ILO estimate).

Knowler, D., and Bradshaw, B. (2007). Farmers' adoption of conservation agriculture: A review and synthesis of recent research. *Food Policy*, *32*(1), 25–48.

Koch, N., Ermgassen, E. K. H. J., Wehkamp, J., Oliveira Filho, F. J. B., and Schwerhoff, G. (2019). Agricultural Productivity and Forest Conservation: Evidence from the Brazilian Amazon. *American Journal of Agricultural Economics*, *101*(3), 919–940.

Kuehne, G., Llewellyn, R. S., Pannell, D. J., Wilkinson, R., Dolling, P., and Ewing, M. A. (2011). ADOPT: a tool for predicting adoption of agricultural innovations. Unknown.

Kuehne, G., Llewellyn, R. S., Pannell, D. J., Wilkinson, R., Dolling, P., Ouzman, J., and Ewing, M. (2017). Predicting farmer uptake of new agricultural practices: A tool for research, extension and policy. *Agricultural Systems*, *156*(1), 115–125.

Linh, T., Long, H., Le Chi, Le Tam, and Lebailly, P. (2019). Access to Rural Credit Markets in Developing Countries, the Case of Vietnam: A Literature Review. *Sustainability*, *11*(5), 1468.

Liu, Y., and Shumway, C. R. (2009). Induced Innovation in U.S. Agriculture: Time–series, Direct Econometric, and Nonparametric Tests. *American Journal of Agricultural Economics*, *91*(1), 224–236.

Lowenberg–DeBoer, J. (2015). The Precision Agriculture Revolution: Making the Modern Farmer. *Foreign Affairs*, *94*(3), 105–112.

Marinoudi, V., Sørensen, C. G., Pearson, S., and Bochtis, D. (2019). Robotics and labour in agriculture. A context consideration. *Biosystems Engineering*, *184*(3), 111–121.

Mwangi, M., and Kariuki, S. (2015). Factors Determining Adoption of New Agricultural Technology by Smallholder Farmers in Developing Countries. *Journal of Economics and Sustainable Development*, *6*(5), 208–216.

Norton, G. W., and Alwang, J. (2020). Changes in Agricultural Extension and Implications for Farmer Adoption of New Practices. *Applied Economic Perspectives and Policy*, *42*(1), 8–20.

Oca Munguia, O. M., and Llewellyn, R. S. (2020). The Adopters versus the Technology: Which Matters More when Predicting or Explaining Adoption? *Applied Economic Perspectives and Policy*, *42*(1), 80–91.

Ogundari, K., and Bolarinwa, O. D. (2018). Impact of agricultural innovation adoption: a meta-analysis. *Australian Journal of Agricultural and Resource Economics*, *62*(2), 217–236.

Olmstead, A. L., and Rhode, P. (1993). Induced Innovation in American Agiculture: A Reconsideration. *Journal of Political Economy*, *101*(1), 100–118.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, n71.

Pannell, D. J., and Claassen, R. (2020). The Roles of Adoption and Behavior Change in Agricultural Policy. *Applied Economic Perspectives and Policy*, *42*(1), 31–41.

Pardey, P. G., Alston, J. M., and Ruttan, V. W. (2010). The Economics of Innovation and Technical Change in Agriculture. In *Handbook of the Economics of Innovation, Volume 2,*Vol. 2, pages 939–984. Elsevier.

Pingali, P. L. (2012). Green revolution: impacts, limits, and the path ahead. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(31), 12302–12308.

Prokopy, L. S., Floress, K. M., Arbuckle, J. G., Church, S. P., Eanes, F. R., Gao, Y., … Singh, A. S. (2019). Adoption of agricultural conservation practices in the United States: Evidence from 35 years of quantitative literature. *Journal of Soil and Water Conservation*, *74*(5), 520–534.

Pustejovsky, J. E. (2020). clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections.

Pustejovsky, J. E., and Tipton, E. (2021). Meta-analysis with Robust Variance Estimation: Expanding the Range of Working Models. *Prevention Science : The Official Journal of the Society for Prevention Research*.

R Core Team. (2020). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Rockström, J., Williams, J., Daily, G., Noble, A., Matthews, N., Gordon, L., … Smith, J. (2017). Sustainable intensification of agriculture for human prosperity and global sustainability. *Ambio*, *46*(1), 4–17.

Rubas, D. (2004). *Technology Adoption: Who is likely to adopt and how does the timing affect the benefits? Dissertation*. College Station, TX, USA: Texas A&M University.

Ruzzante, S. W., Labarta, R., and Bilton, A. (2021). Adoption of agricultural technology in the developing world: A meta-analysis of the empirical literature. *World Development*, *146*(1), 105599.

Scoones, I., and Thompson, J. (1994). *Beyond Farmer First*. Rugby, Warwickshire, United Kingdom: Practical Action Publishing.

Scoones, I., Thompson, J., and Chambers, R. (2009). *Farmer First Revisited*. Rugby, Warwickshire United Kingdom: Practical Action Publishing.

Shang, L., Heckelei, T., Gerullis, M. K., Börner, J., and Rasch, S. (2021). Adoption and diffusion of digital farming technologies - integrating farm-level evidence and system interaction. *Agricultural Systems*, *190*(1), 103074.

Smits, J., and Permanyer, I. (2019). The Subnational Human Development Database. *Scientific Data*, *6*, 190038.

Stanley, T. D., and Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. New York: Routledge.

Stark, S., Biber-Freudenberger, L., Dietz, T., Escobar, N., Förster, J. J., Henderson, J., … Börner, J. (2022). Sustainability implications of transformation pathways for the bioeconomy. *Sustainable Production and Consumption*, *29*, 215–227.

Sterne, J. A. C., and Egger, M. (2005). Regression Methods to Detect Publication and Other Bias in Meta-Analysis. In H. R. Rothstein, A. J. Sutton, and M. Borenstein, editors, *Publication Bias in Meta-Analysis*, pages 99–110. Chichester, UK: John Wiley & Sons, Ltd.

Stewart, R., Langer, L., Da Silva, N. R., Muchiri, E., Zaranyika, H., Erasmus, Y., … Wet, T. (2015). The Effects of Training, Innovation and New Technology on African Smallholder Farmers' Economic Outcomes and Food Security: A Systematic Review. *Campbell Systematic Reviews*, *11*(1), 1–224.

Torero, M. (2021). Robotics and AI in Food Security and Innovation: Why They Matter and How to Harness Their Power. In J. von Braun, M. S. Archer, G. M. Reichberg, and M. Sánchez Sorondo, editors, *Robotics, AI, and Humanity*,Vol. 87, pages 99–107. Cham: Springer International Publishing.

Viechtbauer, W. (2005). Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261–293.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

von Braun, J., Afsana, K., Fresco, L. O., and Hassan, M. (2021). Food systems: seven priorities to end hunger and protect the planet. *Nature*, *597*(7874), 28–30.

Chapter 6
Mow it or grow it: heterogeneous biodiversity-yield tradeoffs in grasslands[19]

6.1 Introduction

Grasslands are a major part of the global ecosystem, covering more than a third of the earth's terrestrial area and two thirds of its agricultural area (FAO 2022). Grasslands are a key source of feed for livestock production, but also provide a wide range of ecosystem services, such as erosion protection and water purification, carbon storage and sequestration as well as habitat for numerous species (Bai & Cotrufo 2022; Liu et al. 2022; Petermann & Buzhdygan 2021; Zhao et al. 2020). Thus, grasslands are an integral element

[19] A revised version of this chapter is published as Schulz, D., Stetter, C., Muro, J., Spekker, J., Börner, J., Cord, A., Finger, R. (2024) Trade-offs between grassland plant biodiversity and yields are heterogenous across Germany. *Communications Earth & Environment* 514(5). DOI: 10.1038/s43247-024-01685-0. A previous version was presented as a poster at the XVII Congress of the European Association of Agricultural Economists (EAAE) from August 29 to September 1, 2023, in Rennes, France. It is the result of a research visit at the Agricultural Economics and Policy group at the Federal Institute of Technology Zurich, supported by the PhenoRob international exchange program.

Author contribution: DS, CS, JB, and RF conceptualized the study, wrote and edited the original draft. DS and CS designed the methodology, DS in addition curated the data, implemented the formal analysis and created visualizations. JS contributed data. JM contributed data, supplementary visualizations and edited the draft.

of food security and to reach ambitious agri-environmental goals and to halt degradation of natural habitats and biodiversity loss as targeted by Sustainable Development Goal 15.5(United Nations 2015). Higher plant species richness is often associated with increased ecosystem services in both intensively and extensively managed grasslands. For example, plant species richness may contribute to increased biomass yield and yield stability, regulating services such as carbon sequestration and cultural ecosystem services (Binder et al. 2018; Daleo et al. 2023; Schaub et al. 2020). Management factors such as the fertilization, mowing, and grazing regimes are important determinants of plant species richness (Plantureux et al. 2005; Socher et al. 2012; Tälle et al. 2016; van Vooren et al. 2018). However, evidence on the direction and magnitude of these relations remains mixed and context-dependent (Le Clec'h et al. 2019; Le Provost et al. 2022; Tälle et al. 2018). Moreover, adjusting management towards higher plant species diversity may imply foregoing yields (Isselstein et al. 2005; Pan et al. 2022; Wang et al. 2019). While existing studies often relied on experimental evidence, the causal identification of large scale real world evidence of the management intensity-biodiversity-yield gradient as well as quantification of the underlying heterogeneity is missing.

Here we provide a new methodological approach to estimate causal, spatially explicit effects of mowing frequency as proxy of management intensity on plant species richness using large, satellite-sourced data across entire Germany for the period 2017-2020 (1,313,073 parcels observed for up to four years resulting in n=5,008,614 total observations). We also quantify the implied trade-offs of lower mowing frequency in terms of yield losses and quantify the opportunity costs associated with higher plant species richness.

We extend previous research by considering the impact of changing mowing frequency rather than focusing on the static relation between current (i.e. observed) mowing frequency and different biodiversity indicators. Modelling spatially explicit causal effects of changes in mowing frequency is important to guide policies to be effective and cost- efficient (Schlüter et al. 2023; Weber et al. 2023). Information on the spatial and temporal variability of a conservation policy's impact can help to target contexts where expected benefits of changing or maintaining certain management practices is largest (Gocht et al. 2016; Huber et al. 2022).

We exploit here the increasing availability of high resolution remote sensing products on mowing frequency and plant species richness that makes our approach scalable and allows us to build a unique dataset and evaluate grassland management policies at national scale (Ehlers et al. 2021; Klein et al. 2020). We consider the estimated number of mowing events (Schwieder et al. 2022) per year as a proxy of grassland use intensity (see Method Section for details). As a proxy of biodiversity we use annual plant species richness estimates (Muro et al. 2022). Both variables are remotely sensed. More specifically, plant species richness estimates refer to the number of plant species per 16 m^2 squared plots and are produced using time series of Sentinel-2 images and plant inventories across Germany (Fischer et al. 2010). Using generalized random forests (Athey et al. 2019) to analyze the resulting data set enables us to evaluate context-specific, spatially explicit causal impacts at field level, i.e. we estimate the effect of higher frequency of mowing events on species richness in the following year. The core principle behind this approach is to estimate non-observed counterfactual plant species richness under a hypothetical mowing frequency based on observed variables and their nonlinear interactions using inverse probability weights (see Section 5.5). We thus contribute to the causal identification of biodiversity impacts of agricultural management using a cutting edge combination of innovative remote sensing data and generalized random forests (Schlüter et al. 2023). The established empirical relation allows us to assess the effectiveness of different conservation strategies. Finally, we expand our analysis and additionally assess yield implications of lower mowing frequency. Thus, we provide a spatial explicit quantification of species richness-yield trade-offs, as moderated by mowing frequency. This allows to inform policy makers draw conclusions for the effective targeting of conservation policies accounting also for provisioning services and thus food security. More specifically, we estimate opportunity cost of marginally extensifying 30 percent of German grasslands to address the 30 by 30 goal.

6.2 Methods

This study covers a random subset of the available data (N=50,000) sampled from all permanent grassland parcels across the entire Federal Republic of Germany (n=1,313,073) observed over the four years 2017-2020. We use temporally and spatially explicit contextual variables to identify determinants of different mowing frequencies and to estimate heterogeneous treatment effects along the most relevant of these variables. Table S1 shows details for all data sources used. The following subsections provide a description of the outcome variable, treatment indicator, contextual variables and sampling strategy. Summary statistics of all input variables are presented in Table S2.

6.2.1 Outcome variable: Plant species richness

To approximate biodiversity, we use a novel dataset of plant species richness generated for entire Germany at 20-meter resolution for the years 2017-2021 based on a previously published method (Muro et al. 2022). The calibration of the model had already been performed with plant species inventories from Biodiversity Exploratories (Fischer et al. 2010; Muro et al. 2022). Time series of Sentinel-2 imagery for Germany were downloaded and interpolated using FORCE processing software (Frantz 2019), before applying the predictive model. The resulting product is a 20 m resolution raster estimating the number of species by 16m^2, which is the transect size. The spatial mismatch between transect size and raster size does not affect our interpretation of species number as an intensive variable. It implies, that we cannot sum up the plant species in a parcel, but only work with parcel averages. To ensure spatial transferability of the trained model for predicting across Germany, the feature space of the input data is used to mask out areas where the model is not applicable (Meyer & Pebesma 2021). That is, grasslands whose spectral-temporal signature differs from the training parcels are excluded from the prediction map. In addition, a grassland mask is applied (Riembauer et al. 2021). Based on the area of applicability, the prediction is valid for 70% of grassland areas. To avoid bias of non-applicable areas in our estimation, we use the percentage of overlap with grassland parcels as inverse weights in our analysis. We use independent secondary data (Hünig & Benzler 2017) to validate the spatial representativeness of our estimates (see Figure S8 for details).

6.2.2 Treatment variable: mowing frequency

The intensity of grassland use is commonly characterized along three dimensions: mowing frequency, fertilization input and grazing pressure (Blüthgen et al. 2012; Gómez Giménez et al. 2017). In this study, we use the number of mowing events per year as a proxy of grassland management intensity for the years 2017-2020(Schwieder et al. 2022). Mowing frequency represents management intensity because it directly reflects the human impact on the ecosystem and as a proxy has the advantage of a measurable temporal reflectance signature in satellite time-series. The temporal reflectance signature of a mowing event may be indistinguishable from an intensive rotational grazing scheme, whereas extensive grazing typically generates a less pronounced temporal reflectance signal and is thus not always identified as a mowing event. Nevertheless, the frequency of both events can still serve as a proxy of use intensity since the algorithm picks up drops in the near infrared signal caused by the removal of vegetation. The number of mowing events was mapped across Germany on a pixel-basis using a rule-based classification approach with dynamic thresholds varying across environmental conditions using imagery from Sentinel-2 and Landsat-8 satellites. The authors report a state-of-the-art overall accuracy of 60% with a slight tendency to underestimate mowing events in regions that were often covered by clouds and could thus be not observed with sufficient temporal resolution. We calculated the mean mowing frequency per parcel, resulting in a continuous indicator.

6.2.3 Contextual variables

We include a wide range of topographical, climatical, meteorological, and pedological properties to account for environmental conditions that may affect both management intensity and species richness (Le Clec'h et al. 2019) (see Table S1 for sources). Apart from these plot-specific environmental conditions, regional production structures as well as social networks may affect management intensity through market proximity or economies of scale (Spörri et al. 2023). Therefore, we include socioeconomic variables relating to the structure of the agricultural economy such as the average farm sizes or share of organic farms in the NUTS-3 regions in the year 2016. Farm size is often associated with larger fields and lower landscape heterogeneity. Landscape structure and heterogeneity in turn have been shown to increase bird biodiversity in Germany (Noack et al. 2021). As a proxy of landscape heterogeneity we calculate a land cover diversity index within a buffer of 1000 meters around each parcel. In addition, we control for parcel shapes since the distance to field boundaries has been shown to affect species composition and dispersion over time. Local and regional differences in regulations are accounted for by including binary indicators for strictly protected area status, Natura 2000 site, and each federal state. We also include latitude and longitude as well as binary year-indicators to control for unobserved heterogeneity. The selection of contextual variables follows the basic principles of causal identification to omit potential endogeneity problems (Pearl 2010).

6.2.4 Sampling strategy

We conduct a field-level analysis focusing on permanent grasslands across Germany. To identify fields with permanent grassland we use a unique, not yet publicly available dataset of field boundaries (Tetteh et al. 2023). We use only those field boundaries labeled permanent grassland. Within each parcel, we calculate the average number of mowing events. the spatial aggregation resulted in a continuous mowing frequency variable that serves as a proxy of grassland use intensity. In a robustness check we also consider the majoritarian mowing frequency per parcel as the treatment indicator. For every model run, including robustness checks, we draw a simple random sample of 50,000 fields for reasons of computational efficiency and to avoid spatial autocorrelation.

6.2.5 Empirical Framework

We are interested in the grassland species richness for the counterfactual scenario of a different mowing frequency. We build upon the potential outcomes framework where treatment effects are estimated by comparing observed outcomes to counterfactual outcomes under an alternative treatment (Neyman 1923; Rubin 1974).

To estimate the causal impact that the mowing frequency has on plant species richness we use causal forests, a specific case of generalized random forests (Athey et al. 2019; Athey & Imbens 2016). In particular, we used the augmented inverse probability-weighted estimator (Athey et al. 2019). This approach has been previously used to estimate, e.g., the impact of tillage on yields (Deines et al. 2019), the impact of agri-environmental schemes on environmental outcomes (Stetter et al. 2022), or the effect of cover crop adoption on maize and soybean yield losses in the United States (Deines et al. 2023). A key advantage of using generalized random forests is the possibility to learn about treatment effect heterogeneity. Furthermore, generalized random forests are doubly robust, i.e. they will give unbiased treatment effects as long as the assumption of unconfoundedness holds for either the treatment propensity model or the outcome model (Wager & Athey 2018). Another advantage of generalized random forests is their ability to partially capture the latent unobserved heterogeneity as long as these latent variables are some (non-)linear representation of the observed covariate space (see Figure S9).

In general terms, the approach consists of two steps. First, a prediction of the mowing frequency is created to serve as propensity score. In a second step, the propensity score is used as inverse weight to create the causal impact mowing has on species richness, while controlling for all other observed context variables. Since the treatment (mowing frequency) is not randomly distributed, we control for potential selection bias by including a range of contextual variables described in Table S1. In particular, we calculate a propensity score for the treatment variable to account for selection bias. We assume unconfoundedness, i.e. we expect that our chosen set of conceptual variables largely capture self-selection and serve as proxies for the effects of potential unobserved confounders that are not included in the analysis. We assess the validity of this assumption by comparing the distribution of propensity scores across quintiles of grassland mowing frequency (Figure S1). We identify and discuss the most important variables for prediction of the mowing propensity as well as the overall treatment effect.

6.2.6 Yield data

We model dry matter grass yields using the biophysical growth model LINGRA-N which was developed to model grass yields across the European Union (Bouman et al. 1996; Wolf 2012) and is available as open source software in R (Butikofer 2023; R Core Team 2023). The model takes daily weather data, hydraulic soil properties, nitrogen applications and mowing dates as inputs. Details on input data and validation can be found in the supplementary material, Text S3.

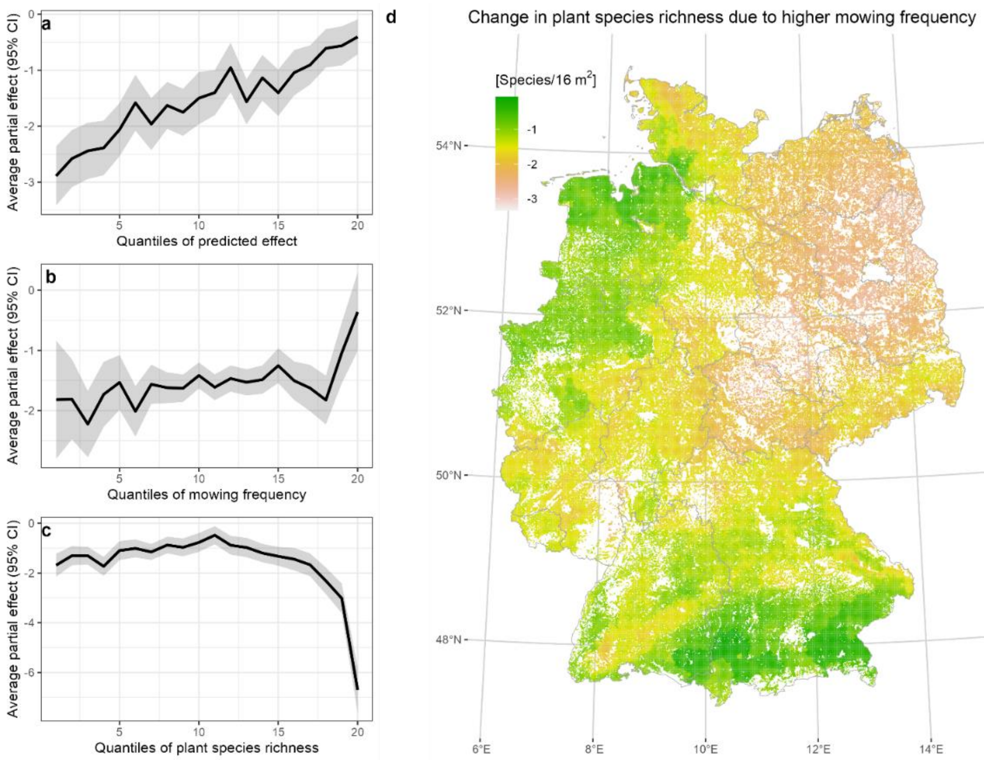
6.3 Results

Plant species richness estimations in our parcel level panel-dataset (N=5.008.614) across the four considered years 2017-2020 varies between 4 and 69 species across all mowing frequencies, with a mean of 23.3 (SD=5.7). The observed mowing frequency ranges between 0 and 5.5 with a mean of 1.9 (SD=0.82). This treatment indicator proxies the gradient of number of uses per year and is continuous due to within-field heterogeneity. Descriptive analysis already reveals a relationship between mowing frequency and species richness. Under the lowest mowing frequency (i.e. zero), the average species richness is 25.6 (SD=6.3), whereas under the highest mowing frequency (i.e. 5.5) this number decreases to 21.1 (SD=3.5).

6.3.1 Causal effect of mowing frequency on plant species richness

We find that plant species richness decreases with increasing mowing frequency. Our generalized random forest analysis shows that, on average, a unit increase in mowing frequency leads to 1.6 (SE: 0.05) fewer plant species, with a full range from -3.5 to 0.2 (SD: 0.39) (Fig. 1A). Using discrete mowing frequency instead of a continuous indicator as treatment variable leads to an effect of similar magnitude (Figure S5). The results of model calibration tests show a strong predictive capacity of the generalized random forest model ($p<0.001$). The omnibus tests of heterogeneity indicate that the effect varies across space and time ($p<0.001$, Table S3, row 1). The spatial heterogeneity identified in our analysis is highly policy relevant. For example, policy makers may be interested to spatially prioritize or avoid changes in mowing regimes in specific areas, e.g. depending on the expected impact of increased mowing frequency on species richness and the current level of species richness. Panel B and C in Figure 6.1 show how the effect varies along mowing frequency and original species richness. It shows that intensification has a slightly lower negative impact in already intensively used parcels, whereas it has a much higher impact in parcels that started off with a high species richness. Panel D shows the geographic distribution of extensification impacts; the largest gains from management extensification in terms of plant species richness occur in the northeast, while the lowest gains occur in the more intensively managed northwest and south of Germany.

Figure 6.1: Heterogeneous impacts of mowing frequency



***Note**: Panels on the left show doubly robust average partial effects across (a) quantiles of the predicted effect, (b) quantiles of mowing frequency and (c) outcome variable with 95 percent confidence intervals. The impact magnitude slightly diminishes with increasing mowing frequency. In contrast to that, the impact of mowing frequency becomes drastically more marked for very species rich parcels. Panel d shows the geographic distribution of predicted field-level impacts of a one-unit increase in mowing frequency aggregated to 1km grid cells. Source: own calculations.*

6.3.2 Drivers of impact heterogeneity

These spatial patterns are driven by spatial and temporal patterns in contextual factors, which we uncover using their variable importance. Average temperatures and precipitation in spring and summer months are among the most important variables to predict the impact of mowing frequency on plant species richness (Figure S4). Furthermore, municipal level structural variables such as average farm size and cattle densities but also variables describing field structure (size, diameter) are important impact predictors. On the contrary, soil properties, location dummies and surrounding landscape diversity are not found to be important predictors. In the context of generalized random forests, this does not necessarily imply that these predictors are irrelevant, but rather that their explanatory power may be captured by other variables.

The temporal variability of a mowing regime might affect species richness by creating specific ecological niches. To explore this underlying mechanisms, we test the sensitivity of our estimates using subsets of parcels that monotonously increased or decreased in mowing frequency over the four years previous to observation. Furthermore, we estimate the average mowing frequency over time and use it as a treatment indicator. All three subsets result in much larger effect magnitudes, implying that the impact proliferates over time (see Table S3, rows 4, 7, and 8).

6.3.3 Robustness checks

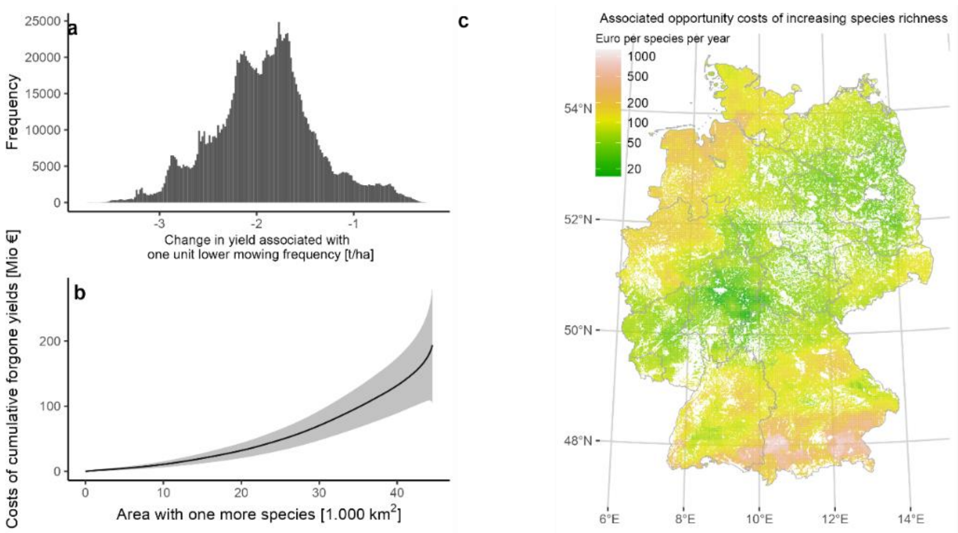
Our treatment variable (mowing frequency) may not capture all relevant dimensions of management that affect species richness. As an alternative proxy we therefore use a composite index containing mowing frequency, grazing intensity and fertilization intensity mapped across Germany available for 2017 and 2018(Blüthgen et al. 2012; Lange et al. 2022). We find a similar, although 30% smaller average partial effect size. This could be due to the temporal limitation of the alternative data; our main indicator also results in smaller effects when restricted to the years 2017 and 2018 (see Table S3, rows 18+19).

We visually confirm that our common support assumption holds, namely that the probabilities of different mowing frequencies overlap (Figure S1). Our results are robust to spatial and temporal restrictions of the training data, the omission of important predictor variables, as well as potential observation bias, i.e. different probabilities of plot to be covered by remote sensing (Table S3, rows 9-20). Furthermore, placebo tests show that our model is not sensitive to spurious patterns in the training data that could bias our estimates (Table S3, rows 23-24).

6.3.4 Quantifying yield trade-off implied by lower mowing frequency

Reducing mowing frequency increases plant species richness. However, this could imply yield reductions (relevant for food security) and thus also opportunity costs for farmers (i.e. they may lose money). To quantify this biodiversity-yield trade-off arising from lower mowing frequencies, we assess the implications of reduced mowing frequency on primary production. We aim to derive spatially explicit upper bound opportunity costs for parcel owners in three steps (see Section 6.2.6 for details). First, we use a biophysical growth model to estimate baseline plot level dry matter yields based on actual weather, soil and management (Wolf 2012). Second, we use generalized random forests to predict the change in yield associated with a change in mowing frequency (Athey et al., 2019). That is, we re-estimate yields under a marginally changed mowing frequency to get the difference between actual and counterfactual yields. Since mowing frequency and yields are highly endogenous, we use latitude, longitude and altitude as the only predictors, so our yield model has no causal interpretation. Third, we multiply the predicted changes in dry matter yield with an average hay price of 70€ per ton (KTBL 2023) to monetarize yield losses. Panel A in Figure 6.2 shows the distribution of yield losses associated with marginal changes in mowing frequency. They range between 1 and 3 tha^{-1} dry matter yield (mean: 2.4t ha^{-1}). Panel B in Figure 6.2 shows the cumulative monetary value of forgone yields (i.e., opportunity cost) from a unit decrease in mowing frequency averaged over the years 2017-2020. For example, having one additional plant species as a result of lower mowing frequency on the 30,000 km^2 grassland with lowest opportunity costs is associated with forgone yields worth approximately 70 million euro. Panel C shows the spatial distribution of opportunity cost. Notably, an additional plant species would incur ten times higher opportunity costs in terms of forgone yields in the southern parts of Germany compared to the center and eastern parts. This is likely partly due to the lower prevailing plant species richness in these intensively used regions, resulting in lower impacts of extensification on this outcome (see Figure 6.1), and potentially partly due to environmental factors that favor higher mowing frequencies.

Figure 6.2: Opportunity costs of higher plant species richness



***Note**: Panel A shows the distribution of forgone dry matter yields per hectare under a one-unit lower mowing frequency. Panel B shows the cumulative upper-bound opportunity cost curve in terms of forgone hay production, assuming an average hay price of 70€/t. The grey confidence band indicates annual variability and prediction error. Panel C shows the spatial distribution of opportunity costs in terms of monetary value of forgone hay yields per unit increase in plant species richness. Plot-level estimates are aggregated over four years (2017-2020) and to a 1 by 1 km grid for visualization. Source: own calculations.*

6.3.5 High potential for effective spatial targeting

Cost-effective targeting of biodiversity policies is crucial to efficiently balance biodiversity and food security as well as costs for farmers (if we impose measures) and taxpayers (if we compensate farmers for voluntary measures) and thus acknowledging that funding for protecting species is limited (McCarthy et al. 2012). However, the past Common Agricultural Policy of the European Union has been shown to be ineffective in targeting high value grasslands and promote biodiversity (Kaligarič et al. 2019). To shed light on the relevance of spatial targeting, we compare the environmental and economic performance of different contextual targeting policies. We consider the "30 by 30" goal (Parties to the United Nations Convention on Biological Diversity 2022), which aims to protect 30% of land by 2030, and was agreed at the 15th meeting of the Conference of Parties to the United Nations Convention on Biological Diversity. In Table 6.1 we compare the effectiveness and cost-efficiency of different policy scenarios that compensate for a marginal decrease in mowing frequency. The first column of Table 6.1 shows the baseline area protected, average plant species richness and yields (at status quo levels), while subsequent columns indicate differences relative to the baseline. In the second column, we consider a scenario in which grassland parcels located within currently existing protected areas marginally decrease their current mowing frequency by one unit. Our results indicate that this would affect 19.7% of permanent grasslands and increase the average species richness by 1.72 (7%), but it would also be associated with a decrease in dry matter biomass yields of, on average, 1.8 t ha^{-1} (20%). Column 3 shows that without any spatial targeting, reducing mowing frequency on randomly chosen30% of the grasslands in our sample causes an average increase of 1.46 plant species but also implies an average reduction in dry matter yield of 1.9 t ha^{-1}. When targeting the 30% of land with the highest predicted impact on plant species richness from extensification, the average species gain is 2.1 (i.e. 47% more effective, see column 4). Furthermore, column 5 shows that when targeting parcels with lowest yield losses when reducing the mowing frequency, the forgone yields can be reduced by 0.5 t ha^{-1} while achieving a higher impact on plant species richness as a non-targeting policy (1.68 vs. 1.46). Finally, we estimate in column 6 that the most cost-efficient policy targeting minimum forgone yields per gained species would incur only 40 percent of the compensation costs of a non-targeting policy.

Table 6.1: Comparison of targeting scenarios

	Base line	Protected areas	No spatial targeting	Target high plant species richness	Target low yield changes	Target low opportunity cost
Area extensified [%]	19.67	19.67	30	30	30	30
Average plant species richness	23.56	+1.72	+1.46	+2.16	+1.68	+2.05
Average dry matter yield [t/ha]	8.99	-1.82	-1.91	-1.4	-1.39	-1.49
Average opportunity cost [€/Species]	-	87.77	126.43	60.71	71.86	51.19

***Note**: baseline column indicates status quo levels; all other columns show scenario based changes relative to the baseline. Source: own calculations.*

6.4 Discussion

Our main contribution is the quantification of the causal relationship between changes in mowing frequency and species richness in grasslands. The here presented causal estimate of -1.6 species per additional mowing event is smaller than correlational estimates reported in previous studies (Socher et al. 2013; Weber et al. 2023), which underlines the importance to control for confounding context variables (Dee et al. 2023). This estimate does not consider the direction and variability of change, i.e. intensification or extensification. However, it may be reasonable to assume that plant species disappear quicker after intensification than they repopulate after extensification because of the altered habitat structure that could impede non-adapted species to grow (Socher et al. 2013). This is also of large policy relevance, e.g. where to support higher species richness or avoid its losses. While we find that impacts proliferate over time, our approach does not lend itself to directly test the above mentioned hypothesis.

Improvements in mapping indicators of land use intensity and biodiversity using remote sensing are required to enable more comprehensive assessments. In particular, future research should aim to overcome technical challenges, such as cloud coverage and refine the accuracy of indicators, and verify the spatial transferability of mapping products across diverse landscapes (Weber et al. 2023). Regarding the latter, we address the potential bias arising from geographically clustered training data (Fischer et al. 2010; Muro et al. 2022) by using an area of applicability mask (Meyer & Pebesma 2021). The validation of our plant species predictions against spatially representative independent data (Hünig & Benzler 2017) (SI Figure S8) shows no indication of systematic bias, but we recognize the need to harmonize data collection protocols for better comparability. Restricting our predictions to the area of applicability limits spatial error propagation, thereby avoiding spatially correlated errors (Meyer & Pebesma 2022). Given the large sample size, this approach addresses potential issues related to the unknown degree of individual prediction uncertainty. We also recognize the possibility of correlated measurement errors in both treatment and outcome indicators due to partial reliance on the same Sentinel-2 satellite imagery. Such errors could affect both indicators in the same way and thus introduce unobservable bias. In our particular case, we are not concerned that correlated measurement errors could invalidate our findings, because of fundamentally different mapping approaches (rule-based for mowing frequency versus neural network for species richness) used to generate outcome and treatment variables.

Future research could benefit from overcoming technical and other limitations of remote sensing products in order to increase measurement precision and allow for a larger range of conceptually relevant variable constructs. This could include ecologically more nuanced indicators that differentiate mowing from grazing as well as biodiversity indicators like relative abundances or the presence of endangered species (Roswell et al. 2021), since effects could vary by subgroups (Socher et al. 2013). Similarly, the focus on measuring outcomes should extend beyond singular dimensions of provisioning ecosystem services, like yield, to encompass the broader spectrum of potential impacts on various ecosystem services (Huber & Finger 2020; Le Clec'h et al. 2019; van Vooren et al. 2018). Finally, future investigations should refine the opportunity cost estimates by considering not only revenues but also production costs to enable more comprehensive economic analysis.

To the best of our knowledge, this study is the first to leverage a comprehensive set of remotely sensed, parcel-level estimates of management changes on biodiversity both measured in real world agricultural systems to inform conservation targeting at the national level (here Germany). In addition, our results contribute to better understand the spatial explicit trade-offs between different objectives such as high yields and high biodiversity. Our methodological framework illustrates the potential of integrating high-resolution remote sensing data with causal machine learning (Schlüter et al. 2023). It is flexible and can accommodate improved and more comprehensive data sources that are likely to emerge over time.

Overcoming current limitations could substantially improve the targeted design of conservation policies, but important implications arise already at this stage. Our results reveal that one-size-fits-all policy solutions, such as decreasing mowing frequency everywhere by the same magnitude, tend to be ineffective. This insight follows from the spatial heterogeneity of mowing impacts and corroborates findings from earlier studies (Gocht et al. 2016; McDonald et al. 2018). In practice, however, this impact heterogeneity has often been ignored, which resulted in heuristic, yet suboptimal, policy design and legislation (Armsworth et al. 2012). Since impacts of higher mowing frequency on plant species richness are highest in areas currently characterized by high plant species richness, policy should focus on maintaining mowing frequencies low in these areas. This is supported by research showing that agro-environmental schemes to be more effective in marginal areas than in intensively used farmland (Batáry et al. 2015). Conversely, extensification efforts in areas of high mowing frequency, such as the north-west and southern regions of Germany are likely less effective and cost-efficient. This implies a need for the development and sustainable use of technological and management innovations that raise the profitability of species rich grasslands. In other words, innovations that focus on grassland productivity only can affect the intensive and extensive margin of production leading to tradeoffs with species richness.

In sum, we show that considerable gains in policy efficacy and efficiency can be expected from leveraging large spatial data sets and digital tools for grassland management (Ehlers et al. 2021). For example, our parcel specific estimates of plant species richness and associated opportunity cost can be used to design and monitor more cost-efficient agri-environmental schemes. In particular, our spatially explicit impact predictions may help to design and implement change-based or result-based rather than action-based payment schemes (Bartkowski et al. 2021). Such schemes have been shown to be the preferred option among farmers (Šumrada et al. 2022) and could be implemented using a revealed cost approach, e.g. when combined with auctions. The policy analysis shows that targeting areas with low opportunity costs is not only the most cost-efficient, but also a very effective approach to increase plant species richness. Our results underscore the need for targeted, context-specific conservation policies, rather than one-size-fits-all solutions, enabled by the integration of large spatial datasets and digital tools. Our findings may support the design of tailored agri-environmental schemes, ultimately contributing to the broader efforts towards reconciling biodiversity conservation and provisioning ecosystem services in grasslands.

Data availability

For a detailed overview of data sources see Table S1. The parcel shapes and plant species estimations are available at https://zenodo.org/records/10619783; the indicator of mowing frequency is available at https://doi.org/10.5281/zenodo.5571613; the database of protected areas at www.protectedplanet.net; topography at https://doi.org/10.5523/bris.s5hqmjcdj8yo2ibzi9b4ew3sn; land cover at http://doi.org/10.1594/pangaea.910837; weather and climate data at https://opendata.dwd.de/climate_environment/CDC/; soil grids at https://soilgrids.org; soil hydraulic properties at https://esdac.jrc.ec.europa.eu/resource-type/datasets; soil depth at http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1304; and agricultural census data at https://www.regionalstatistik.de/genesis/online.

Supplementary Information and Code Availability

Supplementary material along with all codes to replicate figures and tables shown here can be accessed at https://osf.io/peh4a/?view_only=f4b37667a25745cdb6bf56f1116c9b56.

6.5 References

Armsworth, P. R., Acs, S., Dallimer, M., Gaston, K. J., Hanley, N., and Wilson, P. (2012). The cost of policy simplification in conservation incentive programs. *Ecology Letters*, *15*(5), 406–414.

Athey, S., and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(27), 7353–7360.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, *47*(2), 1148–1178.

Bai, Y., and Cotrufo, M. F. (2022). Grassland soil carbon sequestration: Current understanding, challenges, and solutions. *Science*, *377*(6606), 603–608.

Bartkowski, B., Droste, N., Ließ, M., Sidemo-Holm, W., Weller, U., and Brady, M. V. (2021). Payments by modelled results: A novel design for agri-environmental schemes. *Land Use Policy*, *102*, 105230.

Batáry, P., Dicks, L. V., Kleijn, D., and Sutherland, W. J. (2015). The role of agri-environment schemes in conservation and environmental management. *Conservation Biology*, *29*(4), 1006–1016.

Binder, S., Isbell, F., Polasky, S., Catford, J. A., and Tilman, D. (2018). Grassland biodiversity can pay. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(15), 3876–3881.

Blüthgen, N., Dormann, C. F., Prati, D., Klaus, V. H., Kleinebecker, T., Hölzel, N., … Weisser, W. W. (2012). A quantitative index of land-use intensity in grasslands: Integrating mowing, grazing and fertilization. *Basic and Applied Ecology*, *13*(3), 207–220.

Bouman, B. A. M., Schapendonk, A. H. C. M., Stol, W., and van Kraalingen, D. W. G. (Eds.). (1996). *Description of the growth model LINGRA as implemented in CGMS*. Wageningen: AB-DLO.

Butikofer, L. (2023). *LingraNR: R interface to LINGRA-N tool* (manual).

Daleo, P., Alberti, J., Chaneton, E. J., Iribarne, O., Tognetti, P. M., Bakker, J. D., … Hautier, Y. (2023). Environmental heterogeneity modulates the effect of plant diversity on the spatial variability of grassland biomass. *Nature Communications*, *14*(1), 1809.

Dee, L. E., Ferraro, P. J., Severen, C. N., Kimmel, K. A., Borer, E. T., Byrnes, J. E. K., … Loreau, M. (2023). Clarifying the effect of biodiversity on productivity in natural ecosystems with longitudinal data and methods for causal inference. *Nature Communications*, *14*(1), 2607.

Deines, J. M., Guan, K., Lopez, B., Zhou, Q., White, C. S., Wang, S., and Lobell, D. B. (2023). Recent cover crop adoption is associated with small maize and soybean yield losses in the United States. *Global Change Biology*, *29*(3), 794–807.

Deines, J. M., Wang, S., and Lobell, D. B. (2019). Satellites reveal a small positive yield effect from conservation tillage across the US Corn Belt. *Environmental Research Letters*, *14*(12), 124038.

Ehlers, M.-H., Huber, R., and Finger, R. (2021). Agricultural policy in the era of digitalisation. *Food Policy*, *100*, 102019.

FAO. (2022). FAOSTAT statistical database. Rome: Food and Agriculture Organization of the United Nations.

Fischer, M., Bossdorf, O., Gockel, S., Hänsel, F., Hemp, A., Hessenmöller, D., … Weisser, W. W. (2010). Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. *Basic and Applied Ecology*, *11*(6), 473–485.

Frantz, D. (2019). FORCE—Landsat + Sentinel-2 Analysis Ready Data and Beyond. *Remote Sensing*, *11*(9), 1124.

Gocht, A., Espinosa, M., Leip, A., Lugato, E., Schroeder, L. A., van Doorslaer, B., and Paloma, S. G. y. (2016). A grassland strategy for farming systems in Europe to mitigate GHG emissions—An integrated spatially differentiated modelling approach. *Land Use Policy*, *58*(1–2), 318–334.

Gómez Giménez, M., de Jong, R., Della Peruta, R., Keller, A., and Schaepman, M. E. (2017). Determination of grassland use intensity based on multi-temporal remote sensing data and ecological indicators. *Remote Sensing of Environment*, *198*, 126–139.

Huber, R., and Finger, R. (2020). A Meta-analysis of the Willingness to Pay for Cultural Services from Grasslands in Europe. *Journal of Agricultural Economics*, *71*(2), 357–383.

Huber, R., Le'Clec'h, S., Buchmann, N., and Finger, R. (2022). Economic value of three grassland ecosystem services when managed at the regional and farm scale. *Scientific Reports*, *12*(1), 4194.

Hünig, C., and Benzler, A. (2017). *Das Monitoring der Landwirtschaftsflächen mit hohem Naturwert in Deutschland* (BfN-Skripten No. 476). Bonn: Bundesamt für Naturschutz. Retrieved from https://doi.org/10.19217/skr476

Isselstein, J., Jeangros, B., and Pavlů, V. V. (2005). Agronomic aspects of biodiversity targeted management of temperate grasslands in Europe - A review. *Agronomy Research*, *3*, 139–151.

Kaligarič, M., Čuš, J., Škornik, S., and Ivajnšič, D. (2019). The failure of agri-environment measures to promote and conserve grassland biodiversity in Slovenia. *Land Use Policy*, *80*, 127–134.

Klein, N., Theux, C., Arlettaz, R., Jacot, A., and Pradervand, J.-N. (2020). Modeling the effects of grassland management intensity on biodiversity. *Ecology and Evolution*, *10*(23), 13518–13529.

KTBL. (2023). Standard grosss margins. Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V.

Lange, M., Feilhauer, H., Kühn, I., and Doktor, D. (2022). Mapping land-use intensity of grasslands in Germany with machine learning and Sentinel-2 time series. *Remote Sensing of Environment*, *277*, 112888.

Le Clec'h, S., Finger, R., Buchmann, N., Gosal, A. S., Hörtnagl, L., Huguenin-Elie, O., … Huber, R. (2019). Assessment of spatial variability of multiple ecosystem services in grasslands of different intensities. *Journal of Environmental Management*, *251*, 109372.

Le Provost, G., Schenk, N. V., Penone, C., Thiele, J., Westphal, C., Allan, E., … Manning, P. (2022). The supply of multiple ecosystem services requires biodiversity across spatial scales. *Nature Ecology & Evolution*.

Liu, H., Hou, L., Kang, N., Nan, Z., and Huang, J. (2022). The economic value of grassland ecosystem services: A global meta-analysis. *Grassland Research*, *1*(1), 63–74.

McCarthy, D. P., Donald, P. F., Scharlemann, J. P. W., Buchanan, G. M., Balmford, A., Green, J. M. H., … Butchart, S. H. M. (2012). Financial Costs of Meeting Global Biodiversity Conservation Targets: Current Spending and Unmet Needs. *Science*, *338*(6109), 946–949.

McDonald, J. A., Helmstedt, K. J., Bode, M., Coutts, S., McDonald-Madden, E., and Possingham, H. P. (2018). Improving private land conservation with outcome-based biodiversity payments. *Journal of Applied Ecology*, *55*(3), 1476–1485.

Meyer, H., and Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, *12*(9), 1620–1633.

Meyer, H., and Pebesma, E. (2022). Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nature Communications*, *13*(1), 2208.

Muro, J., Linstädter, A., Magdon, P., Wöllauer, S., Männer, F. A., Schwarz, L.-M., … Dubovyk, O. (2022). Predicting plant biomass and species richness in temperate grasslands across regions, time, and land management with remote sensing and deep learning. *Remote Sensing of Environment*, *282*(1), 113262.

Neyman, J. (1923). On the application of probability theory to agricultural experiments: Essay on principles, section 9. (translated in 1990). *Statistical Science*, *5*, 465–480.

Noack, F., Larsen, A., Kamp, J., and Levers, C. (2021). A bird’s eye view of farm size and biodiversity: The ecological legacy of the iron curtain. *American Journal of Agricultural Economics*, *106*(3), 199.

Pan, Q., Symstad, A. J., Bai, Y., Huang, J., Wu, J., Naeem, S., … Han, X. (2022). Biodiversity–productivity relationships in a natural grassland community vary under diversity loss scenarios. *Journal of Ecology*, *110*(1), 210–220.

Parties to the United Nations Convention on Biological Diversity. Kunming-Montreal Global Biodiversity Framework, CBD/COP/15/L25 (2022).

Pearl, J. (2010). 3. The Foundations of Causal Inference. *Sociological Methodology*, *40*(1), 75–149.

Petermann, J. S., and Buzhdygan, O. Y. (2021). Grassland biodiversity. *Current Biology*, *31*(19), R1195–R1201.

Plantureux, S., Peters, A., and McCracken, D. (2005). Biodiversity in intensive grasslands: Effect of management, improvement and challenges. *Agronomy Research*, *3*(2), 153–164.

R Core Team. (2023). *R: A language and environment for statistical computing* (manual). Vienna, Austria. Retrieved from https://www.R-project.org/

Riembauer, G., Weinmann, A., Xu, S., Eichfuss, S., Eberz, C., and Neteler, M. (2021). Germany-wide Sentinel-2 based Land Cover Classification and Change Detection for Settlement and Infrastructure Monitoring. In *Proceedings of the 2021 conference on Big Data from Space: 18-20 May 2021.* Virtual Event.

Roswell, M., Dushoff, J., and Winfree, R. (2021). A conceptual guide to measuring species diversity. *Oikos*, *130*(3), 321–338.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.

Schaub, S., Buchmann, N., Lüscher, A., and Finger, R. (2020). Economic benefits from plant species diversity in intensively managed grasslands. *Ecological Economics*, *168*(1), 106488.

Schlüter, M., Brelsford, C., Ferraro, P. J., Orach, K., Qiu, M., and Smith, M. D. (2023). Unraveling complex causal processes that affect sustainability requires more integration between empirical and modeling approaches. *Proceedings of the National Academy of Sciences*, *120*(41), e2215676120.

Schwieder, M., Wesemeyer, M., Frantz, D., Pfoch, K., Erasmi, S., Pickert, J., … Hostert, P. (2022). Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series. *Remote Sensing of Environment*, *269*, 112795.

Socher, S. A., Prati, D., Boch, S., Müller, J., Baumbach, H., Gockel, S., … Fischer, M. (2013). Interacting effects of fertilization, mowing and grazing on plant species diversity of 1500 grasslands in Germany differ between regions. *Basic and Applied Ecology*, *14*(2), 126–136.

Socher, S. A., Prati, D., Boch, S., Müller, J., Klaus, V. H., Hölzel, N., and Fischer, M. (2012). Direct and productivity-mediated indirect effects of fertilization, mowing and grazing on grassland species richness. *Journal of Ecology*, *100*(6), 1391–1399.

Spörri, M., El Benni, N., Mack, G., and Finger, R. (2023). Spatio-temporal dynamics of grassland use intensity in Switzerland. *Regional Environmental Change*, *23*(1), 23.

Stetter, C., Mennig, P., and Sauer, J. (2022). Using Machine Learning to Identify Heterogeneous Impacts of Agri-Environment Schemes in the EU: A Case Study. *European Review of Agricultural Economics*, *49*(4), 723–759.

Šumrada, T., Japelj, A., Verbič, M., and Erjavec, E. (2022). Farmers' preferences for result-based schemes for grassland conservation in Slovenia. *Journal for Nature Conservation*, *66*(1), 126143.

Tälle, M., Deák, B., Poschlod, P., Valkó, O., Westerberg, L., and Milberg, P. (2016). Grazing vs. mowing: A meta-analysis of biodiversity benefits for grassland management. *Agriculture, Ecosystems & Environment*, *222*, 200–212.

Tälle, M., Deák, B., Poschlod, P., Valkó, O., Westerberg, L., and Milberg, P. (2018). Similar effects of different mowing frequencies on the conservation value of semi-natural grasslands in Europe. *Biodiversity and Conservation*, *27*(10), 2451–2475.

Tetteh, G. O., Schwieder, M., Erasmi, S., Conrad, C., and Gocht, A. (2023). Comparison of an Optimised Multiresolution Segmentation Approach with Deep Neural Networks for Delineating Agricultural

Fields from Sentinel-2 Images. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, *91*(4), 295–312.

United Nations. (2015). Transforming our World: The 2030 Agenda for Sustainable Development.

van Vooren, L., Reubens, B., Broekx, S., Reheul, D., and Verheyen, K. (2018). Assessing the impact of grassland management extensification in temperate areas on multiple ecosystem services and biodiversity. *Agriculture, Ecosystems & Environment*, *267*(9), 201–212.

Wager, S., and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242.

Wang, Y., Cadotte, M. W., Chen, Y., Fraser, L. H., Zhang, Y., Huang, F., … Loreau, M. (2019). Global evidence of positive biodiversity effects on spatial ecosystem stability in natural grasslands. *Nature Communications*, *10*(1), 3207.

Weber, D., Schwieder, M., Ritter, L., Koch, T., Psomas, A., Huber, N., … Boch, S. (2023). Grassland-use intensity maps for Switzerland based on satellite time series: Challenges and opportunities for ecological applications. *Remote Sensing in Ecology and Conservation*, rse2.372.

Wolf, J. (2012). LINGRA-N: Simple generic model for simulation of grass growth under potential, water limited and N limited conditions.

Zhao, Y., Liu, Z., and Wu, J. (2020). Grassland ecosystem services: a systematic review of research advances and future directions. *Landscape Ecology*, *35*(4), 793–814.

Chapter 7
Modest forest and welfare gains from current REDD+ initiatives[20]

7.1 Introduction

In 2003, at the United Nations Framework Convention on Climate Change (UNFCCC) 9th Conference of the Parties (COP9), researchers from Brazil and the USA launched the notion of "compensated reduction": tropical countries should be *ex-post* rewarded for reducing their national forest loss below a pre-agreed baseline (Santilli et al. 2005). The European Union's (EU) Joint Research Centre also linked national forest-cover baselines to possible compensations between countries (Mollicone et al. 2007). Simultaneously, a high-level review of the economics of climate change concluded that for US$5-10 billon, two thirds of global deforestation could

[20] A revised version of this chapter is published as Wunder, S., Schulz, D., Montoya-Zumaeta, J.G., Börner, J., Ponzoni Frey, G., Betancur-Corredor, B. (2024) Modest forest and welfare gains from initiatives for reduced emissions from deforestation and forest degradation. Communications Earth & Environment 394(5). DOI: 10.1038/s43247-024-01541-1

Author contribution: SW conceptualized the study, collected data, wrote and edited the manuscript. DS collected and curated the data, performed the formal analysis, created figures, and wrote and edited the manuscript. JMZ collected and curated the data, edited and reviewed the manuscript. JB helped to conceptualize the study, wrote and edited the manuscript. GF collected data and created figures. BBC collected data, contributed to the analysis, and wrote and revised the manuscript.

be ‘bought out’, thus curbing one major source of global greenhouse gas emissions at low costs (Stern 2007). Since much land clearing in the forested hinterlands of the Global South only provides marginal economic returns, conservation opportunity costs there often remain modest, so allegedly they could be compensated or ‘bought out’ rather cheaply. The conceptual scope was later broadened towards an all-inclusive term of political consensus: “reduced emissions from deforestation, forest degradation and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries” (short: REDD+) (Turnhout et al. 2017).

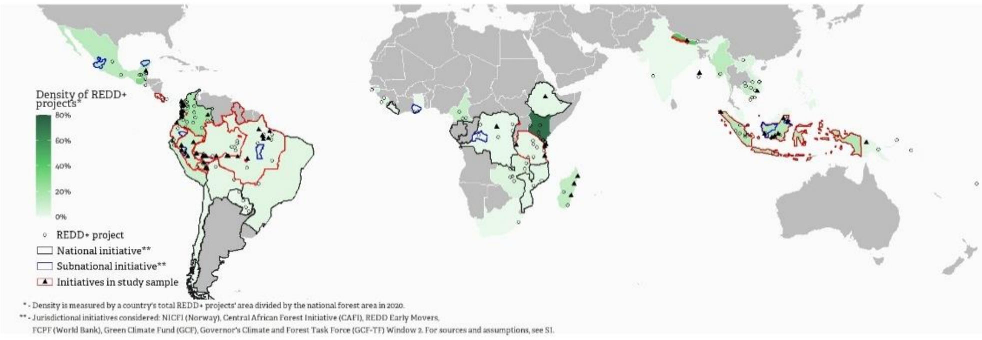
REDD+ would basically work as an international multilevel system of conditional, performance-based payments for environmental services (PES) (Angelsen 2017; Pagiola 2011; Wunder 2009). In this global architecture, carbon markets would mobilize funding, while recipient national governments would incentivize on-the-ground landholders and forest-dwelling indigenous populations, invest in economic alternatives, capacity building, and improve protected areas – thus delivering the enabling conditions for achieving emission reductions on the ground (Wertz-Kanounnikoff & Angelsen 2009). Hence, REDD+ as a model of intervention is usually associated with global performance-based forest-carbon funding, but implementation is de facto an umbrella term for a broad mix of ground-level initiatives, designed in contextually customized ways.

A swath of local-level REDD+ projects has been implemented since the COP13 Bali Action Plan in 2007 (Sills 2014; Simonet et al. 2018). Across governance scales, “nested approaches” were to resolve issues of attributing carbon credits between projects, subnational programmes, and the national level, including to avoid double-counting (Pedroni et al. 2009). REDD+ came out strengthened from the UNFCCC Paris Agreement (UNFCCC

2015). 377 REDD+ projects cover 53 million hectares in 56 countries, based on CIFOR's International Database on REDD+ Projects (ID-RECCO) (Atmadja et al. 2022). These projects were to avoid some 1% of annual forest-based emissions, but quite recently only 5% of their carbon credits had been sold (Simonet et al. 2018).

Across the Global South, Brazil (48), Colombia (33), and Peru (25) are project-leading; the density of REDD+ implementation (project area/national forest area) is highest in Kenya, Nepal, Central America, and the Andes region (Figure 7.1); conversely, REDD+ implementation is negatively biased against Central Africa and South Asia. Our map also clearly depicts the nesting challenges of REDD+ credits between overlapping national, subnational, and project scales of action (e.g., in Brazil). Finally, mapping the interventions evaluated by at least one of our included studies (black triangles; for sample selection, cf. Methods—case sources are listed in the Supplementary Information, SI) also points to some light research imbalances: compared to REDD+ implementation, only few rigorous impact evaluations have been done in Asia, Africa, Colombia, and Mexico.

Figure 7.1: Mapping REDD+: projects, programmes, and location of study sample



***Note**: Overlapping subnational, regional and national initiatives occur with variable implementation and research densities across the tropics.*

Rather than carbon markets assuming the lead role in financing large-scale jurisdictional-level implementation, REDD+ has remained ‘project-ified’, with bilateral or conservation donors financing only incipient actions (Nepstad et al. 2013). Bilateral donors and the UN-REDD programme implement so-called REDD+ “readiness” preparatory processes (increasing forest monitoring capacities, analysing deforestation drivers, etc.); 50+ countries have created national REDD+ programmes (Duchelle et al. 2019).

The originally envisaged model of REDD+ large-scale national implementation has in practice only in a few countries advanced towards large-scale conditional payments. Those comprise notably Norway’s International Climate and Forest Initiative (NICFI), launched as early as 2007 (Angelsen 2017), followed by Germany’s REDD Early Movers programme (Pistorius & Kiff 2014). More recently, two multilateral organizations started piloting large-scale, results-based payments: the Green Climate Fund (GCF) and the Forest Carbon Partnership Facility (FCPF), as well as efforts to mobilize public-private blended funding under the LEAF coalition.

A sometimes-harsh debate has accompanied REDD+ projects. Strong criticisms of REDD+ processes and impacts have focused inter alia on problems related to social inclusion, indigenous rights, and other welfare outcomes (Chomba et al. 2016; Corbera 2012; Fletcher et al. 2016; Griffiths 2007). Conversely, more optimistic outlooks stressed the experimental nature of project-scale REDD+ initiatives (with some encouraging outcomes), the time-consuming complexity of governance transitions, and the embryonic stage of the genuine national-level REDD+ that largely remains untested (Angelsen et al. 2017; Duchelle et al. 2019; Pham et al. 2018).

At this stage, how much do we know about the on-the-ground successes or failures of REDD+? Impact evaluations should answer this question; they are becoming standard tools in many sectors (Gertler et al. 2016). Following emphatic calls for solidly formalized empirical impact assessments also in environmental and biodiversity conservation (Ferraro & Pattanayak 2006), the field of environmental impact evaluation has recently expanded, as evidenced by various reviews and meta-studies (Börner et al. 2020; Samii et al. 2014; Snilsveit et al. 2019).

For REDD+ funded initiatives, so far less stylized evidence is available. The literature-synthesizing work has been mostly qualitative, ranging from reviews of the REDD+ literature (Turnhout et al. 2017), its governance challenges (Ravikumar et al. 2015) and perspectives for future REDD+ implementation (Duchelle et al. 2019). More formal evaluations include lessons from early carbon project (Caplow et al. 2011), and first REDD+ environmental and welfare effects (Duchelle et al. 2018).

Against this backdrop, this REDD+ meta-study systematically takes stock of the currently expanding evidence, which required careful delimitations (cf. Methods). In our Web-of-Knowledge based literature search using text-mining algorithms, we target rigorous quantitative evaluations of the environmental and welfare-related impacts of REDD+ interventions. This includes (corporate or NGO) projects, public programmes (e.g., national payments for environmental services (PES) schemes with forest-carbon components), and a few bilateral jurisdictional agreements (e.g. Norway's forest agreements with Guyana and Indonesia). Our focus is on avoided deforestation and degradation, rather than re-, afforestation, or restoration. Included evaluation studies sport impact estimates that can be scaled and ranked, i.e., effect sizes (and their precision) are comparable including across categories defined by relevant contextual and design variables. Finally, as often generically called for (Editorial 2021), we compare REDD+ impact sizes to those of other forest conservation instruments. To our knowledge, no such analysis pre-exists in the REDD+ literature.

7.2 Methods

7.2.1 Delineating the concept of REDD+

As we saw above, REDD+ is typically seen as a prototype type of action (i.e., a means) that generically remains described exclusively by its outcomes of reduced emissions (i.e., an end). This is fundamentally different from other interventions; for instance, “protected areas” or “forest certification”, describing means not ends. Observers can thus conceptually come to confuse a model for action (the alleged market-based offsetting strategy of REDD+) with an expected final goal (of having forests store more carbon) (Wunder et al. 2020b).

Here we thus explicitly walk through the typical stages and assumptions underlying a REDD+ intervention, using a Theory of Change (ToC) approach, designed for causally linking the stages of inputs, treatments, outputs, outcomes and impacts (Weiss 1997). Figure S2 (SI) outlines these stages going from left to right, with key critical assumptions flagged in bubble shapes. As for inputs, REDD+ is directly triggered by, and thus essentially dependent on the presence of external finance flows, be it from global markets for carbon credits (as originally envisaged), or from bilateral development/ environmental donors (such as Norway’s NICFI), multilaterals with a climate mandate (e.g. the Global Environment Facility or the Green Climate Fund), and private-sector non-market flows for direct emission offsets, based on notions of corporate social responsibility.

Generally, serious claims for REDD+ achievements can eventually only be made if knowledge about pre-existing carbon stocks, land-use trends, key drivers and stakeholders triggering forest loss (and protection) jointly can be merged into a credible baseline: what would have happened under the laissez-faire baseline assumption of ‘no-REDD+ intervention’? Notably, a proper assessment of levels/ changes in threat is quintessential: if threats from deforestation drivers are rising, treatment may have to be intensified. Conversely, if a projected threat was not to materialize at all, then neither the dynamic counterfactual nor the project will exhibit any deforestation.

REDD+ treatments are highly heterogeneous in their composition. We thus distinguish between the subcategories of incentives, disincentives, and enabling measures (Börner & Vosti 2013). First, invariably some incentives are present in REDD+ as a general local benefit-sharing mechanism, or compensation for the opportunity costs of newly introduced/ enforced restriction in forest use or conversion to alternative land uses. Incentives can either be conditioned upon compliance with certain land-use rules (e.g. PES-type of contracts), or unconditional investments into alternative, environmentally more benign livelihoods, social sectors (health, education), etc. Often, REDD+ interventions also entail disincentives, through newly introduced restrictions or a more thorough monitoring and sanctioning of incompliance with already existing ones. Typically, REDD+ has thus included both carrots and sticks. Third, enabling measures as a residual category include tools such as the free prior informed consent (FPIC) of local people’s participation in REDD+, a clarification of land tenure and access rules, etc.

Many real-world REDD+ projects and programmes, such as the Bolsa Floresta Programme in Brazil’s Amazonas State (Cisneros et al. 2019), or the Sustainable Settlements in the Amazon (PAS) project in the

Transamazon region of Pará State (Simonet et al. 2019; Carrilho et al. 2022) have been using the full spectrum of conditional and unconditional incentives, disincentives, and enabling measures. Pilot interventions experimented with different components, but an underlying belief prevailed that holistic, locally customized approaches carried higher probabilities of success, especially in market-remote, cash-strapped frontier regions. Unsurprisingly, many REDD+ projects are in their holistic range of actions 'ICDP-like', with a predominant focus on non-conditional livelihood enhancements (Sills 2014; Sunderlin et al. 2014). For the same reason, REDD+ projects have also had much to learn from ICDPs (Wunder et al. 2020b).

Public PES programmes with a partial focus on forest carbon goals constitute a second type of intervention. Often, carbon financing has helped to boost the funding of these national-level, or at least regional-scale programmes. Costa Rica's PSA, Peru's National Forest Conservation Programme, and Ecuador's Socio Bosque all constitute such examples, although the latter two combined PES with ICDP components (Giudice et al. 2019; Jones and Lewis 2015). Hence, with forest carbon enhancement for climate change mitigation being flagged as an explicit goal, these PES-based interventions need to be included as another pathway of implementing REDD+.

Outputs are to be understood as the immediate, often short-term results of the 'treatment': the treated recipients need to understand the goals and modalities of the intervention, the rules of the game (incl. land and resource tenure) are clarified, and (dis)incentives well-applied. Delivered outputs imply that stakeholder motivations have been successfully aligned with the goals of the intervention. For this to occur, treatments need to have been well-designed and carefully implemented. From the PES literature, we know that spatial targeting in the selection of participants and their to-be-treated land areas is an Achilles Heal, *vis-à-vis* two complementary dimensions: a) the site-specific environmental service density (here: forest carbon stocks per hectare), and b) the on-site projected threat (here: of deforestation/ degradation) of that stock to become endangered over time. Also, customization of the benefits (e.g. multiple payment levels) can help making the intervention more cost-effective and equitable (Ezzine-de-Blas et al. 2016; Wunder et al. 2018). The outcome level is where the REDD+ rubber hits the road: do critical stakeholders undertake the required behavioural on-the-ground changes? That is, do they reduce forest clearing, charcoal making, or timber harvesting in the by REDD+ required manner (environmental outcomes)? Similarly, do income, consumption, and assets increase among those targeted stakeholders (socioeconomic outcomes)? These are all measurable indicators that can potentialy be impact-evaluated.

The final transition towards impacts – the overarching primary carbon-related goal of reduced forest-based emissions, as well as ethically and politically important side-objectives related to biodiversity, self-perceived human wellbeing, equity, and tenure security – entail further subtleties. First, a reality check is to what extent intervention-targeted stakeholders and deforestation drivers have been adequately aligned. For instance, many REDD+ projects are focused on addressing smallholders to reduce their deforestation, but a local surge in land grabbing from more powerful external agents might render these efforts less fructiferous in terms of mitigating deforestation. Second, income and consumption outcomes trigger development feedback loops on the final impacts. Rebound effects refer to treatment-induced changes in household incomes potentially affecting consumption patterns (e.g., higher incomes stimulating meat and dairy consumption) that per se change ecological footprints. Magnet effects refer to the potential of these income changes to attract outside migrants, e.g.

through successful employment creation in REDD+ projects. Pull migration could have a bearing on land use, as migrants open up new land plots for subsistence agriculture. Both effects are well-established in the PES literature (Wunder et al. 2020a). Third, the goal of mitigating climate change is both universal and perpetual. Classical concerns *vis-à-vis* REDD+ projects are thus to what extent these time- and place-bound interventions contribute to the universal and perpetual impacts. As for permanence, the impact of a time-limited treatment on carbon stocks may also only be transitory – though as such still important for mitigating climate change in the short run. Conversely, to the extent the treatment triggers desirable structural changes at the output and outcomes level, permanence might be increased.

Likewise, a REDD+ treatment may not only reduce on-site deforestation, but also push some pressures outside the intervention area – a phenomenon known as leakage. This spillover effect will typically diminish, though not fully erase REDD+ mitigation impacts (Meyfroidt et al. 2020; Pfaff & Robalino 2017). The larger the scale of the REDD+ intervention, the less leakage we should expect – a key argument for favouring national programmes over REDD+ projects. The size of leakage in conservation incentive programmes is seldom quantified (Wunder et al. 2020a). For high-value products sold on international markets, such as harvesting precious timbers, leakage may be exceptionally high (Sohngen & Brown 2004). In general, the higher the price elasticity and the geographical mobility on output and input markets (incl. access to land), the larger leakage we should expect (Wunder et al. 2020a).

7.2.2 Sample delimitation

As mentioned, we aim to take stock of the currently available evidence from rigorous quantitative impact evaluations for REDD+ interventions. This means that we need to apply various apriori filters of inclusion (cf. Table

S1), related both to the underlying REDD+ intervention (Factors 1-4), and subsequently to the case study evaluating its impacts (Factors A-F).

As “REDD+ interventions” (1), we understand here, firstly, actions that implementers self-denominated using the RED(D)+ label, and secondly, other actions that fully or partially featured forest-based climate mitigation/ carbon outcomes in an explicit way. As mentioned above, this would include also national-level PES programmes that pretend to further forest-carbon objectives; in turn, some large watershed-focussed PES programmes (e.g. in China, Mexico, and Vietnam) remain excluded. As for actions (2), many forest carbon programmes include both conservation/ regeneration of standing forests and afforestation/reforestation (A/R) activities; those focused entirely on A/R do not functionally fit the REDD+ definition, and we thus excluded them. In terms of scale (3), we chose to be inclusive of both subnational REDD+ (incl. projects) and emerging national programmes, keeping in mind they likely have different characteristics – cf. also (1). Finally, as a temporal cut-off point for the start of REDD implementation (4), we used year 2007, coinciding with the Bali UNFCCC COP13: pre-2007 forest-carbon initiatives (Joint Implementation, Clean Development Mechanism, etc.) were of comparative interest (Caplow et al. 2011), but were inevitably bound to differ from REDD+.

A second layer of filters refers to the analytical level. First, we chose to include in the screened literature not only peer-reviewed but also grey-literature studies (A) – considering in a quickly-moving field also recent working paper-staged contributions (assessed by us as ‘high-quality'). As for analysed impacts (B), we looked at both forest carbon (main goal) and welfare effects (primary side-objective). As “bottom-line”, we understand effects to be observed at the right-hand side of the REDD+ ToC, i.e. both outcomes and impacts (see above). Impact evaluations are often stated in terms of outcomes (C), such as forest-cover (deforestation areas, rates) and land-use proxies (e.g. fire incidences), which are more precisely observable than forest carbon in the short-to-medium term. More process-oriented, intermediary outputs (middle part of ToC) are not of our interest (D): they are often more qualitative than quantitative, and less clearly (sometimes, ambiguously) linked to REDD+ bottom-line outcomes. Notably, we included subjectively stated wellbeing ("do you now feel better/worse-off/ unchanged than prior to the REDD+ project start?"), as a popular socioeconomic bottom line of evaluation (E). Admittedly, these indicators feature potential response biases, and are thus best triangulated with more objectively measurable socioeconomic outcomes.

The final, yet ponderous filtering criterion refers to the quality of impact evaluation (F). To rigorously attribute impacts to interventions, counterfactuals are needed: what would have happened without the REDD+ intervention? We only included impact studies using counterfactuals, i.e. experimental and quasi-experimental methods. This includes the alleged 'gold standard' of randomized controlled trials (RCT), and Before-After-Control-Intervention (BACI) designs. Various econometric techniques attempt to *ex-post* model counterfactuals, including using matching to identify adequate control observations, or selecting non-treated units to synthesize control units. Yet, different recall techniques can also be used to gather baseline data ex post in the field. To make impact estimates quantitatively comparable, we also needed standard deviation estimates. Many case-study authors did not publish these; we had to contact several for obtaining this supplement.

7.2.3 Study identification strategy

Our literature search strategy, data extraction procedures, and meta-analysis protocols were registered on the Open Science Framework (OSF)[21]. We started by screening our pool of studies from prior REDD+ reviews (Börner et al. 2020; Burivalova et al. 2019; Duchelle et al. 2018; Montoya-Zumaeta 2021). Initially, 15 eligible studies with quantitative estimates of REDD+ and carbon-focused PES projects using counterfactual impact evaluation methods was identified. A Boolean search string based on title and abstract of this initial sample was semi-automatically generated, following the method described by Grames et al. (2019) (cf. SI, Figure S1).

We extracted study characteristics such as location, intervention details, sample characteristics along with Hedge's G effect sizes. Our final sample comprises a total of 30 REDD+ interventions, analysed in 32 studies, with 52 effect sizes being included (35 forest-related, 17 socioeconomic outcomes). This includes disaggregated effects being used in the moderation analysis. For the main analyses, we aggregated effects resulting in 23 and 12 estimates for environmental and socioeconomic indicators, respectively. For a meta-study, this remains a fairly small sample, restricting also our analytical options: although the number of rigorous impact studies has expanded rapidly in recent years (more than half of the articles included have been published since 2017), more is needed to reach a critical mass for detailed statistical analysis. Our studies are just about equally divided between specialized REDD+ projects/programmes and PES schemes; yet the latter concentrate on fewer cases. In the former category, some studies are multi-case comparisons, e.g. a pool of Amazon Fund-financed and VCS-certified private REDD+ projects (West et al. 2020) and cases from CIFOR's

[21] Available at: https://osf.io/mydbk/?view_only=1cbc13ed180e4263962846605dacc510

Global Comparative Study on REDD+ (GCS-REDD) (Bos et al. 2017; Duchelle et al. 2017; Larson et al. 2018; Sunderlin et al. 2017).

How well does our final sample represent the REDD+ universe? For recall, it is shaped by the filters we have applied (cf. Tables 1), overlaying geographically an initial implementation bias (where have REDD+ investors gone?) with a research bias (where have scientists preferred to work, and found access to data?), and publication bias (is it more likely that positive results are published than negative or null results?). Our small sample mirrors an 'absolute' implementation bias towards Latin America (Brazil, Andes, Mesoamerica); it covers less well some 'high-density' REDD+ countries (Kenya, Colombia, Guatemala). In addition, we also find a significant research bias towards Latin America, but we found no evidence that evaluated REDD+ projects systematically differ in terms of their project area or annual deforestation pressure from those not evaluated (cf. SI Table S3). We did find evidence for a moderate publication bias based on Egger's regression test on funnel plot asymmetry (cf. SI): environmentally positive, significant results have a slightly higher likelihood of getting published. On aggregate, the external validity of our sample is deficient, yet still vastly exceeds that of earlier meta-studies of forest conservation incentives, having been based on smaller and geographically much more biased samples (Samii et al. 2014; Snilsveit et al. 2019).

7.2.4 Meta-analysis

The meta-analysis was carried out using the standardized mean difference (Hedges'g) as the outcome measure. We use the *metafor* along with *clubSandwich* packages in R (version 4.3.1) (R Core Team 2023). A multi-level random-effects model was fitted to the data, including random effects at the study and country level. For the main estimates, we assumed a correlation of 0.8 within studies and countries, and report robust variance estimates based on the correlated hierarchical effects procedure (Pustejovsky 2020; Pustejovsky & Tipton 2021). We conducted subgroup analyses, testing for differences between self-declared REDD+ and PES-cum-carbon programmes. Similarly, we tested for differences between the outcome and impact levels of the socioeconomic variables. For the moderation analysis we also included binary moderators indicating a) deforestation pressure (1 for high pressure; 0 otherwise); b) spatial targeting (1 if study explicitly mentions ecosystem service density and/or deforestation threat as determining factors for the location and/or intensity of the intervention; 0 otherwise); and c) benefit differentiation (1 if study explicitly mentions differently sized benefit levels within the same scheme; 0 otherwise). Our binary division between high and low deforestation threat was based on the position *vis-à-vis* the mean annual deforestation rate over the period 2001-21 across all countries (0.28% y^{-1}) from Global Forest Watch (GFW). We compare this threshold with the average case-level deforestation rate during the last five years prior to REDD+ start.

7.2.5 Environmental effects

Among the 23 observations in our main analysis. the observed standardized mean differences ranged from −0.1999 to 0.4623; most estimates were positive (91%). The Q-test indicated heterogeneity among the true outcomes ($Q(22)=98.2097$, $p<0.0001$, $\tau^2=0.0018$, $I^2=78.5150\%$): while the average outcome was estimated to be positive, in some studies the true outcome might be negative. Inspection of the studentized residuals did not reveal any values larger than ±3.0654, indicating the absence of outliers in this model. Additionally, based on Cook's Distance, one study (Roopsind et al. 2019) appeared to exert a notable influence. Figures S4a and b present funnel plot of the estimates. The regression test showed funnel plot asymmetry ($p=0.0321$), although the rank correlation test did not ($p=0.8346$).

7.2.6 Socioeconomic effects

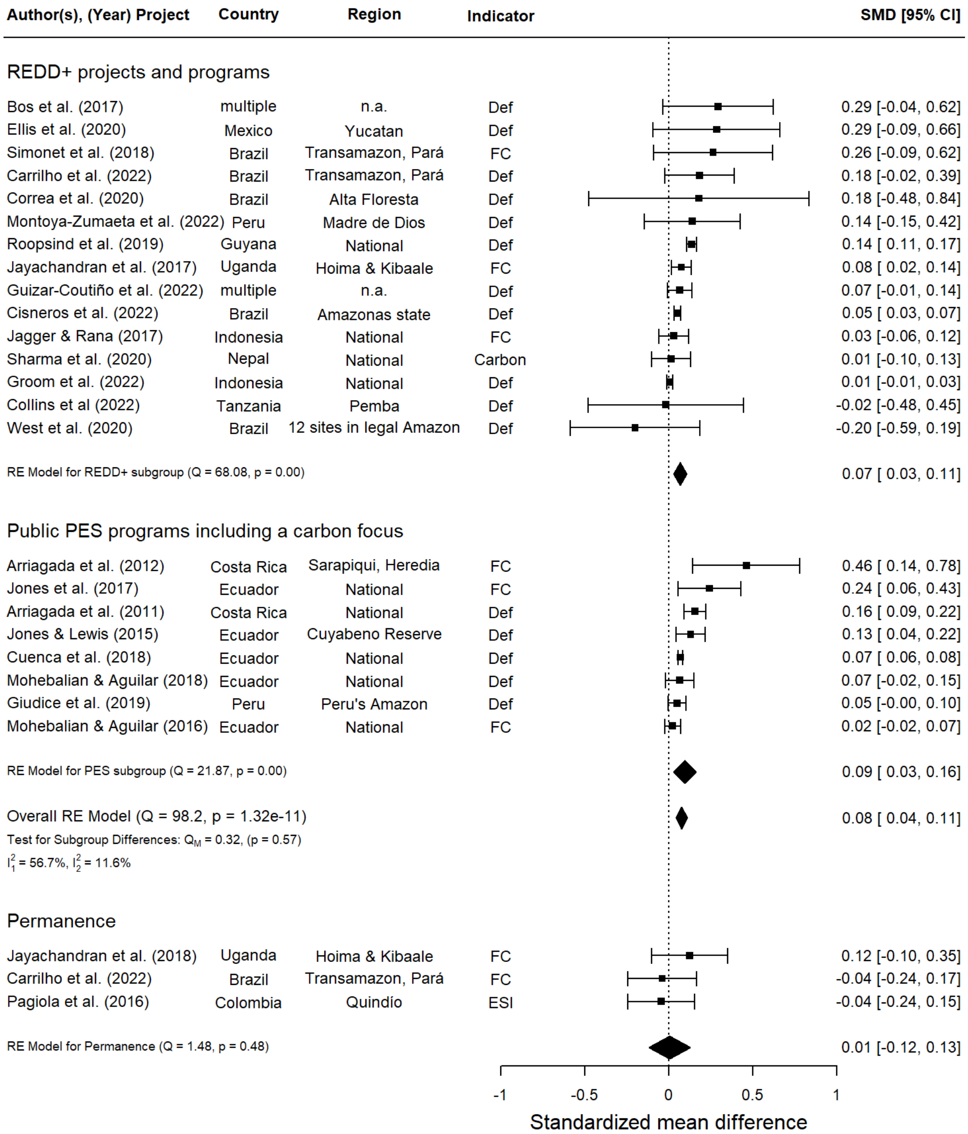
For the 12 included observations, the observed standardized mean differences ranged from −0.1249 to 0.2422; half of the estimates were negative (51%). The Q-test indicated heterogeneity among the true outcomes ($Q(11)=27.7393$, $p=0.0035$, $\tau^2=0.0045$, $I^2=62.8001\%$): although the average outcome was estimated to be positive, in some studies the true outcome may be negative. The studentized residuals showed no values exceeding ±2.8653, i.e., no indication of outliers. Based on Cook's Distance, none of the studies could be considered overly influential. Figures S4a and b present funnel plots of the estimates. Neither the rank correlation nor the regression test indicated any funnel plot asymmetry ($p=0.8406$ and $p=0.7048$, respectively).

7.3 Results

7.3.1 Environmental impacts

We used the correlated hierarchical effects model with random effects for our impact calculations. In total, 32 quantitative studies (listed in SI) with 26 forest-related and 12 socioeconomic primary effect sizes fulfilled our data selection criteria (cf. Methods). Figure 7.2 shows a forest plot of our meta-regression results for comparable forest impact sizes from REDD+ treatments (results fully reproduced in SI). We only have one (insignificant) estimate for forest carbon—the primary goal and final impact of REDD+ according to its Theory of Change (cf. SI). Most estimates are for forest-cover proxy outcomes (including absolute and relative forest loss) leading to these impacts, which can be more easily compared. Our mean overall estimated REDD+ effect of 0.08 (95% Confidence Interval: 0.04-0.11) can be considered "small". The Q-test indicates heterogeneity, meaning one can find true effects outside of this confidence interval. Yet, the positive significant estimate confirms modest forest conservation gains from REDD+.

Figure 7.2: REDD+ environmental impacts: projects, programmes, and permanence



***Note**: Indicator labels refer to Deforestation (Def) Forest cover (FC), and Ecosystem Services Index (ESI). Random Effects (RE) models without moderators, standard errors clustered at the country- and study level. We report Cochran's Q test statistic of residual* *heterogeneity (Q) along with its corresponding p-value. Random effect models of both subgroups show small but significant impacts of REDD+ initiatives, incl. carbon-focused PES schemes, with insignificant permanence.*

Two intervention subgroups can be distinguished in Figure 7.2: self-declared REDD+ projects (commercial, NGO-led, or national—cf. upper panel) versus carbon-inclusive multipurpose conservation PES (public, mostly national programmes schemes—middle panel). We found no significant effect difference between the two (p=0.57). The precision of estimates is lower among especially some of the smaller-sized projects. Even in public PES-for-carbon schemes, the same programmes evaluated in different studies reached quite divergent estimates – including studies carried out by the same authors (Arriagada et al. 2011, 2012; Mohebalian & Aguilar 2016, 2018), seemingly reflecting both variations in output variables and in matching methods (cf. Methods). Of special interest would also be the performance of larger-scale, jurisdictional-level REDD+, given the ongoing implementation shift towards those. These results are moderately encouraging, with conservation impacts in Guyana (Roopsind et al. 2019), Indonesia (Groom et al. 2022) (both NICFI) and Amazonas, Brazil (Cisneros et al. 2022) all being significantly positive, although the latter two very small-sized (all case references in SI).

Looking at the secondary impacts, i.e., indicators not directly comparable to forest-cover proxies (SI, Figure S3), we observe here also some larger, significant impacts, such as boosting tree-species richness, avoiding wildfire incidence or slowing forest encroachment. Yet, this extended picture remains variable, too. Notably, impacts on forest degradation, the second “D” in REDD+, are just like for deforestation small or statistically insignificant.

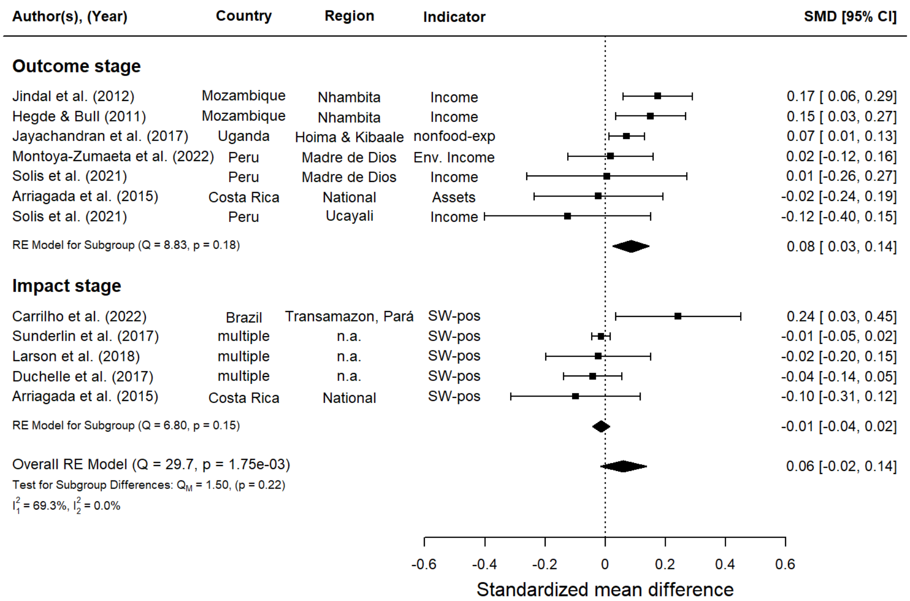
Finally, little is globally known so far about the permanence of REDD+, i.e. to what extent prospective conservation impacts last after the intervention has ended. Estimates across the three permanence studies (excluded from our overall REDD+ effect sizes in the upper panels) originating from Uganda (Jayachandran et al. 2018), Brazil (Carrilho et al. 2022), and Colombia (Pagiola et al. 2016) differ somewhat, but all coefficients are insignificant (Figure 7.3, lower panel). The dominant pattern here is thus that, following confirmed REDD+ deforestation reductions during implementation, post-REDD+ forest loss returns approximately to its pre-intervention speed, but without eliminating the temporary conservation and climate mitigation gains achieved.

7.3.2 Socioeconomic impacts

In our REDD+ Theory of Change (SI, Figure S2), the most important side-objective of REDD+ is it to improve local people’s wellbeing. Figure 7.3 thus shows comparable socioeconomic effects from rigorous REDD+ impact evaluations. Like for environmental impacts, these are divided between outcomes (changes in income, consumption, or asset holdings) and proper impacts: the self-stated subjective wellbeing, and changes herein, on behalf of REDD+ participants and other residents. Also here, our outcome variables are very close impact proxies – and can often be verified more objectively.

Empirically, the two types of indicators perform differently (p=0.03): at the outcome stage REDD+ has a significant positive, welfare-improving effect of 0.09 (95% CI: 0.03-0.15), while at the impact stage the mean effect of -0.01 is statistically indifferent from zero. Hence, the good news is that REDD+ on average tends to make benefit recipients slightly better off materially. The bad news is that this may not always boost self-perceived welfare. The few studies where outcomes and impacts are measured simultaneously (Arriagada et al. 2015; Montoya-Zumaeta et al. 2022) confirm this trend: the material benefits provided may come to fall short of community expectations, especially when these are ex-post assessed, after benefit flows have ceased. Thus, self-stated subjective wellbeing may also become a strategic vote of protest by local participants over REDD+ benefit sharing.

Figure 7.3: REDD+ socioeconomic outcomes and impacts



***Note**: Indicator label "SW-pos" refers to self-reported changes in subjective wellbeing. Random Effects (RE) models without moderators, standard errors clustered at the country- and study level. We report Cochran's Q test statistic of residual heterogeneity (Q) along with its corresponding p-value. Random effect models suggest small positive effect at outcome level (i.e. material welfare proxies) but no significant effect at impact level (i.e. subjective wellbeing).*

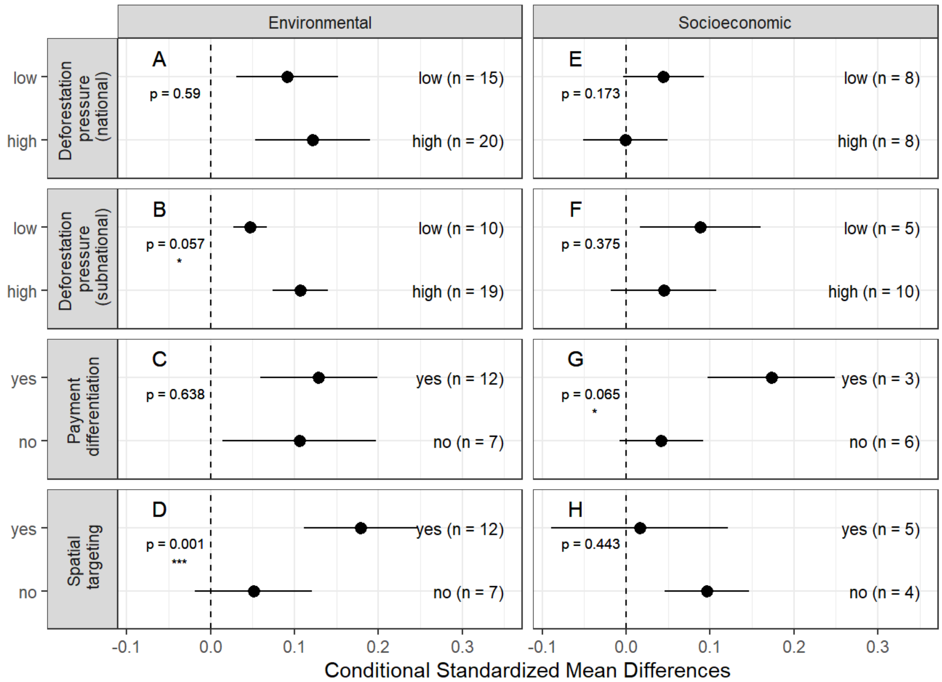
Looking at other socioeconomic outcomes (SI, Figure S4), we observe that several impacts for subgroups can come out negative, such as the subjective wellbeing of female REDD+ participants (Larson et al. 2018). This is a reminder that modest average gains in material welfare from REDD+ do not necessarily warrant equity or a do-no-harm principle: distributional and non-material effects may still create (objective or self-perceived) losers.

7.3.3 Contextual and design factors

We conducted a moderator analysis for potential hints about the implicit role different REDD+ context and design factors in co-determining the above-assessed effect sizes. Hence, we plot conditional impact sizes against selected variables (Figure 7.4, A-H). We used all available observations to account for within-study subgroup differences in the design variables. Figure 7.4 shows the estimated coefficients along with its 95% confidence intervals. Our number of effect-size observations is low for solidly exploring correlations, but we can set hypotheses for future research.

First, this includes baseline deforestation pressure (Figure 7.4 A,B,E,F), which for other conservation tools correlated positively with impact size (Börner et al. 2020; Wunder et al. 2020a). Intuitively, the lower *ex-ante* forest-loss threats are, the harder it would become to counterfactually demonstrate progress. For REDD+, indeed we confirm a positive relationship between threat and impact, statistically insignificant at national (Figure 7.4 A) but significant (p=0.057) at the zoomed-in scale of subnational deforestation pressure (Figure 7.4 B) (cf. Methods for classification). Socioeconomically, low-threat REDD+ might go along with higher welfare gains (e.g. through lower opportunity costs), as also indicated by the coefficient sign here, yet this correlation is insignificant (Figure 7.4 E,F).

Figure 7.4: Moderator analysis



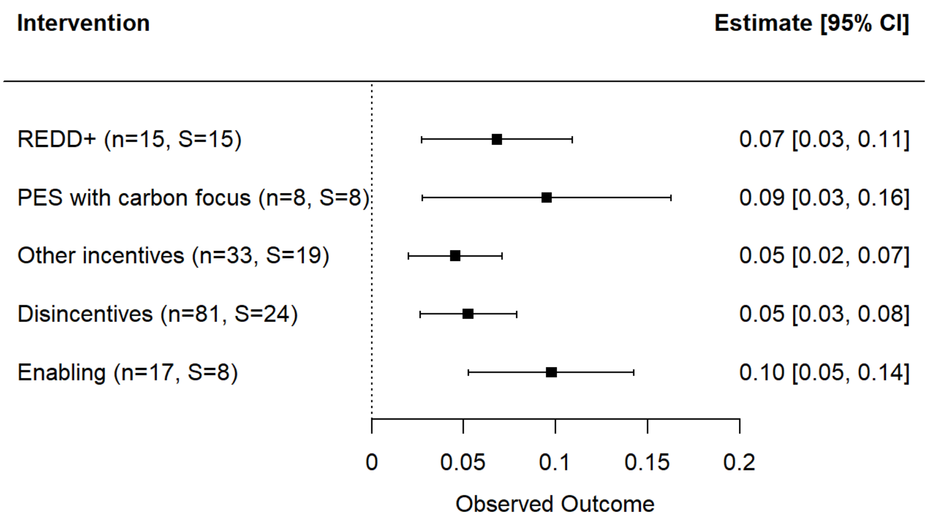
***Note**: The model did not include an intercept, so the coefficients can be directly interpreted as conditional mean effects. We report p-values from a Wald-test of moderator equality clustered at the study-level. Numbers of observations differ because a) we used disaggregated effect sizes, i.e. all available observations per study to exploit within-study variation, and b) we do not show observations with missing moderator values (in particular all cross-country studies). Pre-intervention deforestation pressures are positively associated with environmental REDD+ effects when measured at subnational (but insignificant at national) scale. Payment/ benefit differentiation in programme design is positively associated with socioeconomic benefits, but does not affect environmental outcomes. Spatial targeting strongly boosts environmental outcomes, but does not affect socioeconomic effects.*

Turning to design factors, the PES literature indicates that providing differentiated, beneficiary-customized, rather than uniform benefits can boost environmental effectiveness (Ezzine-de-Blas et al. 2016; Wunder et al. 2018). Again, we find the expected coefficient sign, but the correlation is insignificant (Figure 7.4 C). Yet, socioeconomic welfare improvements were significantly higher in programmes with differentiated, rather than uniform benefits (p=0.045). This indicates that social customization may be important in REDD+ benefit-sharing strategies. As for the second REDD+ design variable, spatially targeting lands with high density of/ high threat towards environmental benefits is again in the PES literature featured as key for additionality outcomes (Ezzine-de-Blas et al. 2016; Wunder et al. 2018). However, we only observed initiatives with spatial targeting in Latin America (cf. SI, Table S2), which restricts our ability to clearly distinguish between regional and design-induced differences. Given this caveat, spatial targeting for forest-carbon density and/or expected deforestation strongly (p=0.001) correlates with environmental impacts (Figure 7.4 D) and may thereby have contributed to the larger effects observed in Latin America. Meanwhile, it did not moderate the socioeconomic outcomes (Figure 7.4 H).

7.3.4 Comparing with other conservation instruments

Finally, we know generally too little about the comparative performance across conservation instruments (Editorial 2021). In Figure 7.5, we compare the normalized Hedges' G effect sizes recorded for the two types of REDD+ with those of pre-existing conservation instruments, such as other incentives, disincentives, and enabling actions—drawing on previous studies (Börner et al. 2020; Wunder et al. 2020a). As in Figure 7.2, we accounted for dependent effect sizes from the same study by assuming a correlation of 0.8 and used robust variance estimation.

Figure 7.5: Environmental impacts from REDD+ vs. other conservation tools



Note*: REDD+ data from own calculations; other tool effect sizes from Börner et al. (2020). "Other incentives" include certification and PES without carbon focus. Comparing Cohen's D effect sizes across environmental policy instruments, REDD+ mean impacts rank 2nd and 3rd among the five tools, but differences are insignificant.*

As for performance, the two REDD+ subgroups (defined as in Figure 7.2) compare fairly well to the other three instrument categories, in terms of mean effect sizes to protect forests (2nd and 3rd numerical ranks, among five). However, there is also a large variability of REDD+ outcomes underlying the rather small intervention samples. Consequently, no statistically significant differences between REDD+ and any of the alternative conservation instruments could be found. Comparatively speaking, REDD+ exhibited a middling, yet also changeable conservation performance.

7.3.5 Robustness checks

Lacking significant differences between the two forest-size outcomes (forest cover, deforestation rate), we included them both in the same primary-effect analysis. In addition, we found no evidence that impact estimates would vary systematically with programme duration. We tested to what extent our results were driven by a few influential studies by a) consecutively excluding studies with high weights, namely Groom et al. (2022) and Guizar-Coutiño et al. (2022); b) excluding studies using the synthetic control method, and c) excluding studies with a Cook's Distance larger than two standard deviations. The coefficient sizes slightly changed, but our conclusions remained robust.

Several studies employed matching techniques, and to calculate effect sizes one requires the correlation between pairs of observations (Borenstein 2009). Due to missing data, we assumed a correlation coefficient of 0.5 for our main specification, but tested also more extreme values (0.3 and 0.7) as a robustness check. Indeed, we found that REDD+ effect estimates are sensitive to the assumed parameter in the calculation method, but not enough to alter our findings in Figure 7.2 and Figure 7.3.

Based on the regional implementation and research bias towards Latin America, we were concerned that regional characteristics could affect the precision of our estimates, but found no evidence for error heteroscedasticity (cf. SI).

7.3.6 Risk of bias assessment

The risk of bias assessment (Supplement Text S5; Figures S5a-d) revealed some variations in methodological quality, both for studies reporting environmental and socioeconomic outcomes: some revealed low risk of bias, others some concerns, or high risk. Caution is needed in interpreting findings, particularly for studies with high risk of bias, as they may impact the overall strength of evidence. For environmental outcomes, bias sources included missing data and deviations from intended interventions; for socioeconomic outcomes, randomization and deviations from interventions were significant sources of bias.

7.4 Discussion

Since 2007, the world has incipiently gathered experiences with REDD+, a tool designed to conserve and enhance forest carbon in non-Annex I countries (i.e. largely developing/ emerging economies, plus China), in exchange for economic compensations from the industrialized Global North. REDD+ is an objectively desirable end (the goal of reducing forest-based emissions) but has equally become a controversial means of using ‘market-based’ international offsets to help accelerate climate change mitigation.

A broad range of REDD+ pilot projects has thus emerged. Jointly, they annually planned to avoid 84 million tCO2 of emissions (over 33 years of mean lifespan). or “around 1% of annual emissions from deforestation, forest degradation, harvesting and peat fires in the tropics” (Simonet et al. 2018). While potentially significant, leakage and credit performance apart only 5% of the correspondingly needed carbon credits had so far been sold on the voluntary market. In terms of de facto avoiding existing deforestation at scale, REDD+ projects have thus been but a drop in the sea (Simonet et al. 2018).

Effectively, REDD+ projects have been starved out by a grossly insufficient global willingness to pay for mitigating climate change. Uncertain funding prospects have also made many projects quasi-placeholders waiting whether funding flows would materialize, meanwhile keeping on-the-ground treatment intensities low (Duchelle et al. 2019; Sills 2014). In particularly, implementers have been hesitant to introduce PES-type of continuous compensations to landholders, since implementers currently cannot promise continuity (Sunderlin et al. 2014; Wunder et al. 2020b). Obviously, this has deteriorated the framework conditions under which REDD+ projects were expected to deliver efficient results.

Yet, this does not per se question the potential usefulness of REDD+ projects in providing valuable pilot lessons for the potential upscaling to jurisdictional REDD+. Above, we have taken stock of the experiences so far. We carefully delimited exactly which initiatives were to be labelled as REDD+ -- either by proponents, or by analysts. We have also screened which impact evaluations were sufficiently rigorous to ensure internal validity and deliver trustworthy results, based on realistic counterfactuals enabling credible causal attribution. With many new empirical studies emerging recently, our larger and geographically more balanced sample than in previous meta-studies should also increase confidence in our results.

We can thus shed some new light on the effectiveness and welfare implications of REDD+ initiatives. As for forest-cover and carbon effects – the ultimate raison d'être – REDD+ initiatives have had small-sized effects, similar to what other conservation instruments have (not) achieved (Börner et al. 2020; Wunder et al. 2020a). This holds for both specialized REDD+ projects/ programmes and cases where REDD+ has been integrated into national PES programmes. When interventions stop, prior pressures tend to resume, yet typically without fully undoing REDD+ gains (partial impermanence). Overall, given disappointing carbon-market flows, and the harsh critique against REDD+, environmental effectiveness is unimpressive, but probably still exceeds the expectations of many critical observers.

For lack of uniform cost data, we could above not systematically compare cost-efficiency parameters. The few available case studies with REDD+ costs data point to highly variable, in some cases elevated transaction costs, but also declining with scale (Nantongo & Vatn 2019; Rakatama et al. 2017; Rendón Thompson et al. 2013). A move towards larger-scale jurisdictional

REDD+ programmes may thus also push towards more ‘bang for the buck’ in climate-change mitigation.

On the socioeconomic side of local benefits provided – REDD+’s primary side-objective – our results on average portray small positive contributions to local livelihood outcomes (e.g. incomes, assets), yet insignificant impacts (e.g. subjective wellbeing). New incentive-based projects also tend to locally build easy-to-disappoint expectations regarding future benefit flows (Montoya-Zumaeta et al. 2019). Customized, rather than equal benefit transfers seem to improve socioeconomic outcomes. Single cases apart, no evidence points to REDD+ making local people systematically worse off: while not everybody locally may gain, a narrow outcome range from welfare neutrality to modest livelihood improvements is most common.

But why have the environmental impact results of this highly innovative tool overall not been much better? Above we have incipiently pointed to several design flaws, such as adverse selection bias and inadequate spatial targeting. Insufficient on-the-ground enforcement of contractual conditionality is, however, another commonly noted deficiency in REDD+ implementation: often implementers will prefer to tolerate land-use violations, safeguarding instead the social capital built with local communities (Cisneros et al. 2022; Giudice et al. 2019; Montoya-Zumaeta et al. 2022; Rosa Da Conceição et al. 2018). Moreover, by tolerating exaggerated baselines of future deforestation, the bar for REDD+ credits was set far too low; unsurprisingly, the majority of credits are non-performing (West et al. 2020, 2023), accelerating a public fatigue with environmental offsets.

Another critical qualitative issue surrounds complexity. REDD+ is to forest carbon what integrated conservation and development projects (ICDP) projects have been to biodiversity: an umbrella term under which a ratatouille of composite, heterogeneous interventions has gathered. Many

REDD+ initiatives are ‘ICDP-like’, in terms of using the same integrated, multifaceted approach: trying a bit of everything to satisfy multiple stakeholders and minimize risks of total failure. Unfortunately, the ICDP approach has had a dismal impact-producing record (Börner et al. 2020).

After 2007, many pre-existing ICDP projects looking for fresh funding were remodelled as REDD+ initiatives, producing an opportunistic self-selection bias. In Indonesia, for instance, many REDD+ projects were implemented by biodiversity-focused organizations; the targeted forest areas were more biodiversity-rich than carbon-dense, and only about one quarter of the project areas was truly threatened by deforestation (Murray et al. 2015). In turn, many genuinely new private sector initiatives adversely targeted de facto low-threat areas: avoiding deforestation would here become a low-hanging fruit (West et al. 2020). Hence, many REDD+ projects may have served more as a proof of concept than as a real test of whether the avoided-deforestation approach is holding water.

Arguably, it is no shame for pilot projects to underperform or fail, if useful lessons are learned for future initiatives. Did early REDD+ interventions thus maximize this learning and upscaling potential? Hardly so, mainly since too many projects were carried out in ‘high-and-far’, low-pressure settings, thus not taking the bull by the horns. Particularly the ICDP-type model was also too complex in design and transaction cost-heavy in implementation to replicate at scale. Furthermore, REDD+ implementers almost never facilitated impact evaluation through (semi-)experimental rollout of multiple design options of action. Hence, we stand back with many highly customized ICDP-like ‘boutique projects’, including multiple components of action; yet we know very little about how well these components worked, and why.

What about full national REDD+ programmes, as an alleged future of upscaled REDD+? For now, only the NICFI programmes in Guyana (Roopsind et al 2019) and Indonesia (Groom et al 2022) have been evaluated, finding for both small yet significant forest-protecting impacts. The Guyana case is not without controversy though: deforestation especially from gold mining actually increased during NICFI support, but less so than was predicted by a synthetic matching model for a no-REDD+ counterfactual, based on other comparable high-forest low-deforesting (HFLD) countries with significant mining sectors.

For future research, doing further analyses of larger-scale REDD+ programmes, be it NICFI or more recently GCF and FCPF interventions, looks promising, but impact evaluation needs to be integrated early into programme design. These impact evaluation analyses should in turn not just provide average effect estimates, but equally be challenged to investigate heterogeneous impacts, enabling us to tell causally plausible stories about where, how, and why REDD+ might work or fail.

Acknowledgments: We are grateful for financial support from Norad through CIFOR's Global-Comparative Study on REDD+, Germany's Federal Ministry for Food and Agriculture, (BMEL) through EFI's NewGo project, and the German Research Foundation under its Excellence Strategy [EXC-2070-390732324-PhenoRob]. We thank multiple authors of primary studies for sharing with us unpublished quantitative material from their respective impact evaluations, and commentators on previous versions of this document: Amy Duchelle, Steven Dörr, Leif-John Fosse, Yitagesu Tekle, Thales West, and Georg Winkel.

Supplementary Information

Supplementary information file with supporting results is available online at:
https://osf.io/mydbk/?view_only=1cbc13ed180e4263962846605dacc510.

Data and Code availability

All data and codes to replicate is available online at: https://osf.io/mydbk/?view_only=1cbc13ed180e4263962846605dacc510.

7.5 References

Angelsen, A. (2017). REDD+ as Result-based Aid: General Lessons and Bilateral Agreements of Norway. *Review of Development Economics*, *21*(2), 237–264.

Angelsen, A., Brockhaus, M., Duchelle, A. E., Larson, A., Martius, C., Sunderlin, W. D., … Wunder, S. (2017). Learning from REDD+: a response to Fletcher et al. *Conservation Biology*, *31*(3), 718–720.

Arriagada, R. A., Ferraro, P. J., Sills, E. O., Pattanayak, S. K., and Cordero-Sancho, S. (2012). Do Payments for Environmental Services Affect Forest Cover? A Farm-Level Evaluation from Costa Rica. *Land Economics*, *88*(2), 382–399.

Arriagada, R. A., Sills, E. O., Ferraro, P. J., and Pattanayak, S. K. (2015). Do Payments Pay Off? Evidence from Participation in Costa Rica's PES Program. *PLOS ONE*, *10*(7), e0131544.

Arriagada, R. A., Sills, E. O., and Pattanayak, S. K. (2011). *Payments for environmental services and their impact on forest transition in Costa Rica* (Working Paper). Retrieved from http://www.bioecon-network.org/pages/13th_2011/Arriagada.pdf

Atmadja, S. S., Duchelle, A. E., De Sy, V., Selviana, V., Komalasari, M., Sills, E. O., and Angelsen, A. (2022). How do REDD+ projects contribute to the goals of the Paris Agreement? *Environmental Research Letters*, *17*(4), 044038.

Borenstein, M. (2009). Effect Sizes for Continuous Data. In H. Cooper, L. V. Hedges, and J. C. Valentine, editors, *Handbook of Research Synthesis and Meta-Analysis, The*, pages 221–235. New York: Russell Sage Foundation.

Börner, J., Schulz, D., Wunder, S., and Pfaff, A. (2020). The Effectiveness of Forest Conservation Policies and Programs. *Annual Review of Resource Economics*, *12*(1), 45–64.

Börner, J., and Vosti, S. A. (2013). Managing Tropical Forest Ecosystem Services: An Overview of Options. In R. Muradian and L. Rival, editors, *Governing the Provision of Ecosystem Services*,Vol. 4, pages 21–46. Dordrecht: Springer Netherlands.

Bos, A. B., Duchelle, A. E., Angelsen, A., Avitabile, V., Sy, V. D., Herold, M., … Wunder, S. (2017). Comparing methods for assessing the effectiveness of subnational REDD+ initiatives. *Environmental Research Letters*, *12*(7), 074007.

Burivalova, Z., Allnutt, T. F., Rademacher, D., Schlemm, A., Wilcove, D. S., and Butler, R. A. (2019). What works in tropical forest conservation, and what does not: Effectiveness of four strategies in terms of environmental, social, and economic outcomes. *Conservation Science and Practice*, *1*(6), e28.

Caplow, S., Jagger, P., Lawlor, K., and Sills, E. (2011). Evaluating land use and livelihood impacts of early forest carbon projects: Lessons for learning about REDD+. *Environmental Science & Policy*, *14*(2), 152–167.

Carrilho, C. D., Demarchi, G., Duchelle, A. E., Wunder, S., and Morsello, C. (2022). Permanence of avoided deforestation in a Transamazon REDD+ project (Pará, Brazil). *Ecological Economics*, *201*, 107568.

Chomba, S., Kariuki, J., Lund, J. F., and Sinclair, F. (2016). Roots of inequity: How the implementation of REDD+ reinforces past injustices. *Land Use Policy*, *50*, 202–213.

Cisneros, E., Börner, J., Pagiola, S., and Wunder, S. (2022). Impacts of conservation incentives in protected areas: The case of Bolsa

Floresta, Brazil. *Journal of Environmental Economics and Management*, *111*, 102572.

Corbera, E. (2012). Problematizing REDD+ as an experiment in payments for ecosystem services. *Current Opinion in Environmental Sustainability*, *4*(6), 612–619.

Duchelle, A. E., de Sassi, C., Jagger, P., Cromberg, M., Larson, A. M., Sunderlin, W. D., … Pratama, C. D. (2017). Balancing carrots and sticks in REDD+: implications for social safeguards. *Ecology and Society*, *22*(3), art2.

Duchelle, A. E., Seymour, F., Brockhaus, M., Angelsen, A., Larson, A. M., MOELIONO, M., … MARTIUS, C. (2019). *Forest-based climate mitigation: lessons from REDD+ implementation* (World Resources Institute Issues Paper). Center for International Forestry Research. Retrieved from JSTOR.

Duchelle, A. E., Simonet, G., Sunderlin, W. D., and Wunder, S. (2018). What is REDD+ achieving on the ground? *Current Opinion in Environmental Sustainability*, *32*, 134–140.

Editorial. (2021). Weighing up policy tools. *Nature Sustainability*, *4*(7), 561–561.

Ezzine-de-Blas, D., Wunder, S., Ruiz-Pérez, M., and Moreno-Sanchez, R. D. P. (2016). Global Patterns in the Implementation of Payments for Environmental Services. *PLOS ONE*, *11*(3), e0149847.

Ferraro, P. J., and Pattanayak, S. K. (2006). Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments. *PLoS Biology*, *4*(4), e105.

Fletcher, R., Dressler, W., Büscher, B., and Anderson, Z. R. (2016). Questioning REDD+ and the future of market-based conservation. *Conservation Biology*, *30*(3), 673–675.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., and Vermeersch, C. M. J. (2016). *Impact Evaluation in Practice, Second Edition*. Washington, DC: Inter-American Development Bank and World Bank.

Giudice, R., Börner, J., Wunder, S., and Cisneros, E. (2019). Selection biases and spillovers from collective conservation incentives in the Peruvian Amazon. *Environmental Research Letters*, *14*(4), 045004.

Grames, E. M., Stillman, A. N., Tingley, M. W., and Elphick, C. S. (2019). An automated approach to identifying search terms for systematic reviews using keyword co–occurrence networks. *Methods in Ecology and Evolution*, *10*(10), 1645–1654.

Griffiths, T. (2007). *Seeing 'RED'? 'Avoided deforestation' and the rights of Indigenous Peoples and local communities*. Moreton-in-Marsh, UK: Forest Peoples Programme.

Groom, B., Palmer, C., and Sileci, L. (2022). Carbon emissions reductions from Indonesia's moratorium on forest concessions are cost-effective yet contribute little to Paris pledges. *Proceedings of the National Academy of Sciences*, *119*(5), e2102613119.

Guizar-Coutiño, A., Jones, J. P. G., Balmford, A., Carmenta, R., and Coomes, D. A. (2022). A global evaluation of the effectiveness of voluntary REDD+ projects at reducing deforestation and degradation in the moist tropics. *Conservation Biology*, *36*(6).

Jayachandran, S., de Laat, J., Audy, R., Pagiola, S. P., and Sedano Santamaria, F. (2018). *Evaluating the permanence of forest conservation following the end of payments for environmental services in Uganda* (No. AUS0000379). Washington: The World Bank. Retrieved from https://documents.worldbank.org/en/publication/documents-reports/documentdetail/867301589223292444/evaluating-the-permanence-of-forest-conservation-following-the-end-of-payments-for-environmental-services-in-uganda

Larson, A. M., Solis, D., Duchelle, A. E., Atmadja, S., Resosudarmo, I. A. P., Dokken, T., and Komalasari, M. (2018). Gender lessons for climate initiatives: A comparative study of REDD+ impacts on subjective wellbeing. *World Development*, *108*, 86–102.

Meyfroidt, P., Börner, J., Garrett, R., Gardner, T., Godar, J., Kis-Katos, K., … Wunder, S. (2020). Focus on leakage and spillovers: informing land-use governance in a tele-coupled world. *Environmental Research Letters*, *15*(9), 090202.

Mohebalian, P. M., and Aguilar, F. X. (2016). Additionality and design of forest conservation programs: Insights from Ecuador's Socio Bosque Program. *Forest Policy and Economics*, *71*, 103–114.

Mohebalian, P. M., and Aguilar, F. X. (2018). Beneath the Canopy: Tropical Forests Enrolled in Conservation Payments Reveal Evidence of Less Degradation. *Ecological Economics*, *143*, 64–73.

Mollicone, D., Achard, F., Federici, S., Eva, H. D., Grassi, G., Belward, A., … Schulze, E.-D. (2007). An incentive mechanism for reducing emissions from conversion of intact and non-intact forests. *Climatic Change*, *83*(4), 477–493.

Montoya-Zumaeta, J. G. (2021). *Reconciling conservation and development? Impacts of incentive-based forest conservation in Peru* (PhD Thesis). Australian National University, Canberra, Australia.

Montoya-Zumaeta, J., Rojas, E., and Wunder, S. (2019). Adding rewards to regulation: The impacts of watershed conservation on land cover and household wellbeing in Moyobamba, Peru. *PLOS ONE*, *14*(11), e0225367.

Montoya-Zumaeta, Javier G., Wunder, S., Rojas, E., and Duchelle, A. E. (2022). Does REDD+ Complement Law Enforcement? Evaluating Impacts of an Incipient Initiative in Madre de Dios, Peru. *Frontiers in Forests and Global Change*, *5*, 870450.

Murray, J. P., Grenyer, R., Wunder, S., Raes, N., and Jones, J. P. G. (2015). Spatial patterns of carbon, biodiversity, deforestation threat, and REDD+ projects in Indonesia: The delivery of biodiversity benefits in REDD+. *Conservation Biology*, *29*(5), 1434–1445.

Nantongo, M., and Vatn, A. (2019). Estimating Transaction Costs of REDD+. *Ecological Economics*, *156*, 1–11.

Nepstad, D., Irawan, S., Bezerra, T., Boyd, W., Stickler, C., Shimada, J., … Lowery, S. (2013). More food, more forests, fewer emissions, better livelihoods: linking REDD+, sustainable supply chains and domestic policy in Brazil, Indonesia and Colombia. *Carbon Management*, *4*(6), 639–658.

Pagiola, S. (2011). *Using PES to Implement REDD* (Payments for Environmental Services (PES) learning paper No. 2011–1). Washington, DC: World Bank. Retrieved from http://hdl.handle.net/10986/17892

Pagiola, S., Honey-Rosés, J., and Freire-González, J. (2016). Evaluation of the Permanence of Land Use Change Induced by Payments for Environmental Services in Quindío, Colombia. *PLOS ONE*, *11*(3), e0147829.

Pedroni, L., Dutschke, M., Streck, C., and Porrúa, M. E. (2009). Creating incentives for avoiding further deforestation: the nested approach. *Climate Policy*, *9*(2), 207–220.

Pfaff, A., and Robalino, J. (2017). Spillovers from Conservation Programs. *Annual Review of Resource Economics*, *9*(1), 299–315.

Pham, T. T., Moeliono, M., Angelsen, A., Brockhaus, M., Gallo, P., Hoang, T. L., … Bocanegra, K. (2018). Strategic alignment: Integrating REDD+ in NDCs and national climate policies. In *Transforming REDD+: lessons and new directions*, pages 69–80. Bogor, Indonesia: Center for International Forestry Research.

Pistorius, T., and Kiff, L. (2014). The Politics of German Finance for REDD+ *SSRN Electronic Journal*.

Pustejovsky, J. E. (2020). clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections.

Pustejovsky, J. E., and Tipton, E. (2021). Meta-analysis with Robust Variance Estimation: Expanding the Range of Working Models. *Prevention Science : The Official Journal of the Society for Prevention Research*.

R Core Team. (2023). *R: A language and environment for statistical computing* (manual). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rakatama, A., Pandit, R., Ma, C., and Iftekhar, S. (2017). The costs and benefits of REDD+: A review of the literature. *Forest Policy and Economics*, *75*, 103–111.

Ravikumar, A., Larson, A. M., Duchelle, A. E., Myers, R., and Gonzales Tovar, J. (2015). Multilevel governance challenges in transitioning towards a national approach for REDD+: evidence from 23 subnational REDD+ initiatives. *International Journal of the Commons*, *9*(2), 909.

Rendón Thompson, O. R., Paavola, J., Healey, J. R., Jones, J. P. G., Baker, T. R., and Torres, J. (2013). Reducing Emissions from Deforestation and Forest Degradation (REDD+): Transaction Costs of Six Peruvian Projects. *Ecology and Society*, *18*(1), art17.

Roopsind, A., Sohngen, B., and Brandt, J. (2019). Evidence that a national REDD+ program reduces tree cover loss and carbon emissions in a high forest cover, low deforestation country. *Proceedings of the National Academy of Sciences*, *116*(49), 24492–24499.

Rosa Da Conceição, H., Börner, J., and Wunder, S. (2018). REDD+ as a Public Policy Dilemma: Understanding Conflict and Cooperation in the Design of Conservation Incentives. *Forests*, *9*(11), 725.

Samii, C., Lisiecki, M., Kulkarni, P., Paler, L., Chavis, L., Snilstveit, B., … Gallagher, E. (2014). Effects of Payment for Environmental Services (PES) on Deforestation and Poverty in Low and Middle Income Countries: A Systematic Review. *Campbell Systematic Reviews*, *10*(1), 1–95.

Santilli, M., Moutinho, P., Schwartzman, S., Nepstad, D., Curran, L., and Nobre, C. (2005). Tropical Deforestation and the Kyoto Protocol. *Climatic Change*, *71*(3), 267–276.

Sills, E. O. (2014). *REDD + on the ground: a case book of subnational initiatives across the globe*. Bogor Barat: Centre for International Forest Research (CIFOR).

Simonet, G., Agrawal, A., Bénédet, F., Cromberg, M., de Perthuis, C., Haggard, D., … Vaillant, G. (2018). ID-RECCO, International Database on REDD+ projects and programs, linking Economic, Carbon and Communities data (Version 3.0) [Data set].

Snilsveit, B., Stevenson, J., Langer, L., Tannous, N., Ravat, Z., Nduku, P., … Ferraro, P. J. (2019). Incentives for climate mitigation in the land use sector—the effects of payment for environmental services on environmental and socioeconomic outcomes in low- and middle-income countries: A mixed-methods systematic review. *Campbell Systematic Reviews*, *15*(3), e1045.

Sohngen, B., and Brown, S. (2004). Measuring leakage from carbon projects in open economies: a stop timber harvesting project in Bolivia as a case study. *Canadian Journal of Forest Research*, *34*(4), 829–839.

Stern, N. (2007). *The Economics of Climate Change: The Stern Review*. Cambridge University Press.

Sunderlin, W. D., Ekaputri, A. D., Sills, E. O., A.E., A. E., Kweka, D., Diprose, R., … Toniolo, A. (2014). *The challenge of establishing REDD+ on the ground: Insights from 23 subnational initiatives in six countries*. Bogor, Indonesia: Center for International Forestry Research (CIFOR).

Sunderlin, W., de Sassi, C., Ekaputri, A., Light, M., and Pratama, C. (2017). REDD+ Contribution to Well-Being and Income Is Marginal: The Perspective of Local Stakeholders. *Forests*, *8*(4), 125.

Turnhout, E., Gupta, A., Weatherley-Singh, J., Vijge, M. J., De Koning, J., Visseren-Hamakers, I. J., … Lederer, M. (2017). Envisioning REDD+ in a post-Paris era: between evolving expectations and current practice. *WIREs Climate Change*, *8*(1), e425.

Weiss, C. H. (1997). How Can Theory-Based Evaluation Make Greater Headway? *Evaluation Review*, *21*(4), 501–524.

Wertz-Kanounnikoff, S., and Angelsen, A. (2009). Global and national REDD+ architecture: linking institutions and actions.

West, T. A. P., Börner, J., Sills, E. O., and Kontoleon, A. (2020). Overstated carbon emission reductions from voluntary REDD+ projects in the Brazilian Amazon. *Proceedings of the National Academy of Sciences*, *117*(39), 24188–24194.

West, T. A. P., Wunder, S., Sills, E. O., Börner, J., Rifai, S. W., Neidermeier, A. N., … Kontoleon, A. (2023). Action needed to make carbon offsets from forest conservation work for climate change mitigation. *Science*, *381*(6660), 873–877.

Wunder, S., Brouwer, R., Engel, S., Ezzine-de-Blas, D., Muradian, R., Pascual, U., and Pinto, R. (2018). From principles to practice in paying for nature's services. *Nature Sustainability*, *1*(3), 145–150.

Wunder, Sven. (2009). *Can payments for environmental services reduce deforestation and forest degradation?* (Realising REDD+), pages 213–224. Center for International Forestry Research. Retrieved from JSTOR.

Wunder, Sven, Börner, J., Ezzine-de-Blas, D., Feder, S., and Pagiola, S. (2020). Payments for Environmental Services: Past Performance and Pending Potentials. *Annual Review of Resource Economics*, *12*(1), 209–234.

Wunder, Sven, Duchelle, A. E., Sassi, C. D., Sills, E. O., Simonet, G., and Sunderlin, W. D. (2020). REDD+ in Theory and Practice: How Lessons From Local Projects Can Inform Jurisdictional Approaches. *Frontiers in Forests and Global Change*, *3*, 11.

Chapter 8
What we can(not) conclude from comparing yields and biodiversity across farm sizes[22]

8.1 Introduction and Re-analysis

In their research article, Ricciardi et al. (2021) (RICC) explore the relationship between farm size and six production-related outcomes. They report a negative association for yields and non-crop biodiversity after employing different methods to synthesize findings of prior studies. Seemingly interpreting empirical correlation as causal mechanism, they derive policy implications suggesting that more small farms improve socioeconomic and environmental performance of agriculture. Revisiting their empirical strategy, we find limited support for their empirical claims.

---

[22] This chapter is a comment to and re-analysis of: Ricciardi, V., Mehrabi, Z., Wittman, H., James, D. & Ramankutty, N. (2021). Higher yields and more biodiversity on smaller farms. *Nature Sustainability* 4, 651–657. The chapter was not peer-reviewed, but is available as a pre-print (http://dx.doi.org/10.13140/RG.2.2.19657.47205). It is first-authored by Hugo Storm, an additional co-author is Thomas Heckelei.

Author contribution: HS conceptualized the analysis, wrote and edited the draft. DS helped to conceptualize the analysis, implemented the formal re-analysis including replication material and visualizations, wrote, edited and reviewed the draft. TH conceptualized the analysis, wrote, edited and revised the draft.

In addition, we argue that policy conclusions need to carefully consider the mechanisms that drive differences in outcomes.

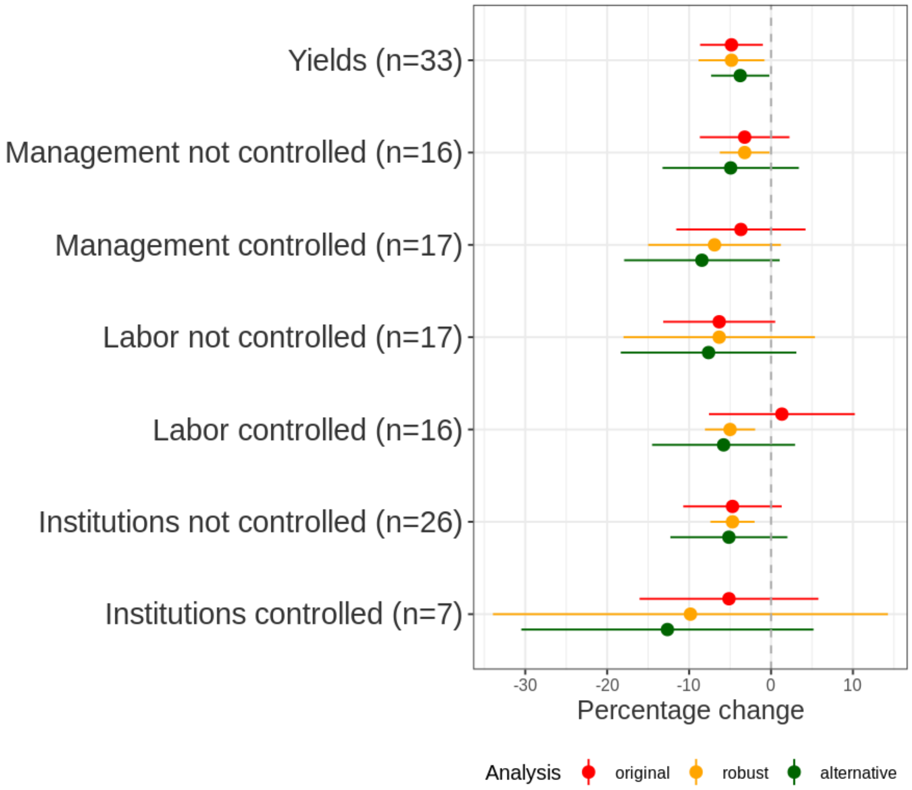
One of the claims by RICC is to provide strong empirical support that smaller farms have higher yields. Average yield decreases by 5% for each hectare increase in farm size. In their methodical approach deriving this result we see fundamental shortcomings. As we lay out in greater detail in the SI we consider the search terms used to identify studies included in the meta-analysis inappropriate. We consider it likely that the search terms miss relevant publications and find it problematic that no systematic derivation of search terms is provided. This is particularly worrying given that other meta-studies focusing on a more restrictive topic identified more than three times the number of studies (Garzón Delvaux et al., 2020). There is also no disclosure of who identified, read and coded the primary studies, as well as the level of agreement in case that more than one reviewer was involved which is considered good practice for rigorous meta-analysis (Havránek et al., 2020). Further, we are sceptical of the way the effect sizes are derived and documented. While the authors referred for the formulas to Rodríguez-Barranco et al. (2017) the provided supplementary material does not provide details about which formulas are specifically used. Inspecting the first two references listed in data provided in RICC we find discrepancies to the original articles (see SI). Further, we are sceptical about the magnitudes of the derived effect sizes. RICC studied relative reduction in yield for a one hectare increase in size, reporting effect sizes for the percentage change per 1 ha increase in farm size up to a -60% (RICC Figure 3). These magnitudes appear highly implausible beyond very small farms, while RICC reports an (unweighted) average farm size in the primary studies of 7.5 (SD=22.7). However, these results seem to be heavily driven by a single outlier (see SI), and more than 50% of observations are based on farmer samples having less than two hectares on average. Extrapolating on the basis of this data is therefore a sensitive issue to be communicated clearly.

For the mean estimate, the authors use a multi-level random effects model allowing the true effects to vary between studies and crop types but not accounting for potentially correlated effects and correlated sampling errors arising from overlapping samples in primary studies. In addition, they mix different outcome measures (yield as weight and value) and different measures for farm size (farm size as “plot size” and “area under cultivation”) reported in primary studies. In the last step, RICC use three separate mixed-effects regressions to determine whether the control of management, labor and institutions in the original studies moderate the mean effect. In Figure 2b of the original paper we believe that the results of those dummies are not correctly reported (see SI for details), we reproduce those estimates (red estimates in our Figure 8.1). We re-estimate their models and calculate predicted mean effect sizes with cluster-robust confidence intervals (CIs) and corrected the reporting error (orange estimates in our Figure 8.1). Lastly, we propose an alternative model that accounts for correlated sampling errors estimated robustly5 and adds dummies for Labor, Management, Institutions and a dummy indicating a value-based outcome measure2 (green estimates in our Figure 8.1). The results of the corrected reporting and alternative model differ somewhat from the RICC results, but also do not warrant strong empirical conclusions given the issues discussed above.

Apart from the inverse yield/size effect, RICC claim to provide empirical evidence “that smallholders are [...] stewards of biodiversity” and that biodiversity is higher on smaller farms. We are critical regarding the empirical evidence for this. The conclusion is based on a vote-counting analysis finding that 77% of studies report a significant negative association between farm size and non-crop biodiversity. In general, conclusions from vote-counting should be drawn cautiously because they assign equal weights to studies with different statistical power6. A sample size of n=87 is reported in their Figure 8.1, while the results are in fact based on 31 unique studies, counting studies several times if they report various indicators. Additionally, the almost exclusive regional focus of the included studies on Europe and North America needs a critical reflection. Most importantly, however, RICC readily equate farm size with field size and small scale landscape structures, with only three studies actually considering farm size while all others look at field size and/or landscape composition. Hence, a more precise conclusion would be that smaller field size and more fragmented landscapes seem to be associated with higher non-crop biodiversity in developed countries.

These methodological limitations call into question the validity of the empirical results RICC present. But even if the empirical results were valid, we would still question the policy conclusions drawn by RICC. It seems crucial to understand the underlying causes of yield or biodiversity differences between large and small farms in order to draw the right policy conclusion (see SI for a more detailed discussion). Finally, we ask whether comparing small and large farms is really the right approach to identify policies helping to achieve sustainable development. Different farm sizes are the (temporary) outcomes of complex, long-term behavioural processes. Maintaining or changing farm sizes may therefore not only be off-target, it may also require massive policy intervention coming at high risks of unintended side-effects. The more relevant questions are directly related to the outcomes of interest and the processes leading to their differentiation. For example, if fragmented landscapes support non-crop biodiversity, then scientists should learn to understand what leads to such landscape structures and policy should use such knowledge to support their maintenance or improvement. Understanding and empirical studying those processes is challenging, but we should not take an empirical correlation between farm size and outcomes as a shortcut and basis for policy conclusion.

Figure 8.1: Updated results of estimated effect from Ricciardi et al. (2021)



Note*: Updated results of estimated effect with 95% CI of farm size [ha] on yields [kg ha^{-1} and \$ ha^{-1}] along with moderation analysis corresponding to Figure 2b in the original publication. The first row gives the overall average effect; the six rows below show results of the moderation analysis. Red estimates are the original values published in RICC Figure 2b; orange estimates are based on the original models but including the intercept for the prediction and cluster-robust 95% CI; green estimates are predicted values based on our alternative model specification.*

Supplementary Information & Data availability

The data, re-analysis code, and supplementary information is available at https://osf.io/2zvwj/?view_only=c9da461b62004be0b7f3de57925cb282.

8.2 References

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.

Garzón Delvaux, P. A., Riesgo, L., and Gomez y Paloma, S. (2020). Are small farms more performant than larger ones in developing countries? *Science Advances*, *6*(41), eabb8235.

Hedges, L. V., Tipton, E., and Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 39–65.

Ricciardi, V., Mehrabi, Z., Wittman, H., James, D., and Ramankutty, N. (2021). Higher yields and more biodiversity on smaller farms. *Nature Sustainability*, *4*(7), 651–657.

Rodríguez-Barranco, M., Tobías, A., Redondo, D., Molina-Portillo, E., and Sánchez, M. J. (2017). Standardizing effect size from linear regression models with log-transformed variables for meta-analysis. *BMC Medical Research Methodology*, *17*(1), 44.

Stanley, T. D., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., … Rost, K. (2013). Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys*, *27*(2), 390–394.