CoSAI
Commission on Sustainable Agriculture Intensification

MINING THE GAPS:

USING MACHINE LEARNING TO MAP A MILLION DATA POINTS FROM AGRICULTURAL RESEARCH FROM THE GLOBAL SOUTH

HAVOS.AI

Cornell University

CABI

Authors
Jaron Porciello, Leslie Lipper, Thomas Bourne, Maryia Ivanina, Sammi Lin, Sarah Langleben

Author Affiliations
Porciello (Havos.Ai and Cornell University); Lipper, Lin, and Langleben (Cornell University), Bourne (Havos.Ai), and Ivanina (Epam Systems, Inc.)

Editorial Services
Stacey Shackford

Graphic Design, Artwork and Layout
Walmazan Studios Inc.

Data partner
CABI

ACKNOWLEDGEMENTS

We wish to thank Mary O'Connor and Cristina Ashby from CABI for their partnership and compilation of the dataset. Thank you to the Commission on Sustainable Agriculture Intensification (CoSAI) including David Shearer for managing this study and especially Julia Compton for envisioning the concept and shepherding this work. A special thank you to the CoSAI Global Commissioners: Ruben Echeverria, Akissa Bahri, Aysegul Ozkavukcu, David Simon, Grethel Aguilar, Haris Gazdar, Irene Annor-Frempong, Jennifer Baarn, Jianguo Liu, Julio Berdegue, Madiodio Niasse, Mauricio Lopez, P.V. Vara Prasad, Pablo Tittonell, Rasheed Sulaiman V, Rodomiro Ortiz, Sarah Mbago-Bhunu, Uduak Edem Igbeka, Uma Lele, Varad Pande, and Ximena Rueda.

TABLE OF CONTENTS

FIGURES

DEFINITIONS

The Global South indicates countries that fall into the World Bank's Lending Classification Categories for Low-Income, Middle-Income & Upper-Middle Income Countries.

Artificial intelligence is the simulation of human intelligence processes by machines, especially computers.

Machine learning is the study of computer algorithms that can improve automatically through experience and by the use of data.

Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can. It combines linguistics with statistical, machine learning, and deep learning models.

While there is no precise definition of an **intervention** for most sectors outside of medicine and health, it is generally recognized that an intervention is an activity that is introduced in a population to produce a certain outcome. Often, the aim of identifying interventions in one context is to evaluate whether it could be reintroduced in another context and with similar results. Interventions is a proxy term to identify research programs, strategies, experiments and projects and other work that has been explored outside of a controlled experiment environment (e.g., a laboratory) and preferably with a target user group.

The OCED-DAC defines **outcomes** as, "likely or achieved short-term and medium-term change and effects of intervention outputs." We use a machine-learning model trained to identify and extract outcomes from scientific literature that is primarily based on how researchers have expressed it in their text.

Research outputs and results: The artefact of conducting research and codifying it in a format (usually written) that can be disseminated. The analysis conducted in this report relies on research outputs, often referred to herein as publications or collectively as our 'evidence base.'

INTRODUCTION

> *We're entering a new era in agriculture, one that moves beyond a purely production-oriented vision and recognizes its role in contributing to a food system that prioritizes people's livelihoods and nutrition, as well as environmental and climate outcomes.*

This shift in thinking will require major shifts in policy, research, and investment. But where should these investments go? What foundations should be strengthened? Which gaps need filling? What's working? What's not?

In order to answer these questions in an informed way, we need to examine the evidence that exists and identify areas where more research is needed.

But this is easier said than done.

The evidence base for agriculture is growing exponentially, and while the wider food systems literature may contain many of the solutions we are seeking, they need to be holistically integrated in order to find those needles in the proverbial haystacks.

Evidence synthesis reports, such as systematic and scoping reviews, provide much needed transparent, rigorous evidence for specific questions. But often, broader questions about the whole of the evidence base need to be answered first, the basic who-what-when-where-and-how that comes before trying to apply a more thorough lens.

State of the evidence reports like this one provide more coverage than what we can achieve in a more focused systematic review—a birds-eye view of the evidence base so that we know where to invest in the future. But until recently we lacked the technology to conduct a landscape scan of the millions of articles that are out there.

Earlier work in this area has suggested that the evidence base we have is not fit for the questions we need and want to ask (Lipper et al., 2020). We need additional efforts to help us understand what the current evidence base has found.

And we also need resources to be designed in such a way that we can seamlessly add new data as it emerges, from many partners and independent of the sources from which the data originated.

With the aid of artificial intelligence and machine learning technology, we took a deep dive into more than 1.2 million publications to assess the current landscape of research for the Global South.

The result is a clearer picture of what research has been conducted on small-scale farming and post-production systems from 2000 to the present, and where evidence gaps exist.

This, in turn, highlights potential areas for investment in research and innovation for small-scale farms in the Global South, and provides scope for future research questions.

CONTEXT AND KEY OBJECTIVES

MIND THE GAPS

Until recently, agricultural research and innovation has been largely focused on improving productivity, focused mainly on a small number of crops (Serraj & Pingali, 2018). While we've seen very high returns from this approach, we have also seen the unintended and negative consequences it can have on nutrition and diets, social inclusion, and the environment (Davidson, 2016; Webb & Kennedy, 2014).

We are now witnessing a major shift in thinking about agriculture, one which puts agriculture in the larger context of a system with complex interactions between food production, processing, consumption, and climate change (Barrett et al., 2020).

This same shift implies a need for rethinking the role of agricultural research and development efforts, and a push for innovations that go beyond productivity. There is a corresponding urgency to identify priority investments (Laborde et al., 2020; Reardon, Lu, et al., 2019). In order to do so, however, we must have an adequate and accessible evidence base on agricultural innovations and their potential in the context of a transformation (Herrero et al., 2020; Reardon, Echeverria, et al., 2019). And it has become increasingly clear that there are several gaps in evidence.

This study looks at the summaries of more than 1.2 million past publications and uses these to assess the current landscape of research for the Global South. The questions that we ask in this report were prioritized by CoSAI:

- What is the research output focused on the Global South? Which countries have had more research focused on them, and which have less? What about crop research?
- Can we identify major domains of research? Can they be organized in a way that will help us better interpret gaps across different research domains?
- What outcomes are being studied across research domain areas?
- Who are the user groups included within studies? How much of the research is targeted on solutions for small-scale farmers and other agricultural actors?
- Do we have information about the impact of agricultural innovation on communities such as indigenous and tribal communities, youth or elderly, extremely poor? How much research focuses on women?
- What does the evidence say about gaps in institutions, policy and finance?
- What does our current evidence base reveal about how we are studying our changing climate? What are the priority aims of research surrounding sustainable development and biodiversity?
- Who are the primary funders and research organizations that are consistently showing up in Global South focused research?
- Can we determine what scale of research is being undertaken—farm and household level, macro and/or landscape, enterprise and/or food system?

What is Innovation?

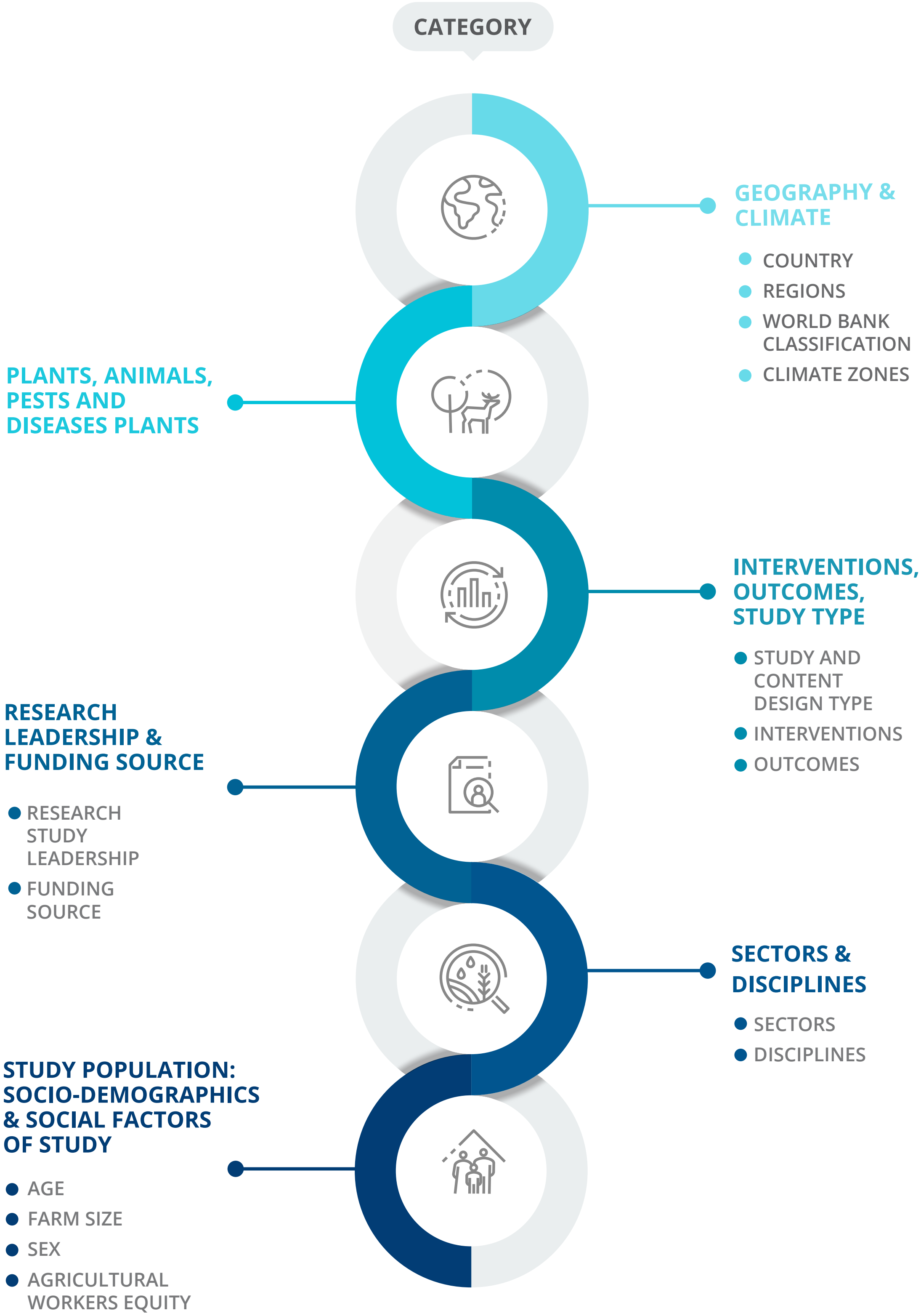
Innovation in the agricultural and food sectors is essential to achieve the types of transformative changes needed for the improved livelihoods, nutrition, and environmental performance of food systems. In agriculture, "innovation" has been used in a narrow sense to mean new technologies that increase productivity. CoSAI uses a much broader definition, such as the one from FAO 2018: *" a process whereby individuals or organizations bring new or existing products, processes or ways of organization into use for the first time in a specific context in order to increase effectiveness, competitiveness, resilience to shocks or environmental sustainability and thereby contribute to food security and nutrition, economic development or sustainable natural resource management."*

Clearly, in the context of innovation to support multiple objectives of food system transformation, the broader definition of innovation is more applicable. Yet we know that many of the research outputs in the evidence base may reflect a narrower definition. This gives rise to gaps in our knowledge which require attention.

We specifically seek to inform these questions based on what research has been conducted on small-scale farming and post-production food systems in the Global South, from 2000 to the present. The conceptual model (Figure 1) lays out these questions as a mega-research map.

Similar questions have been raised in the lead-up to the United Nations Food Systems Summit. The value of this report is a baseline mapping that will, we hope, aid the prioritization and coordination of international funding and research efforts.

FIGURE 1. The primary data points collected per article is outlined in the conceptual model.



OUR APPROACH

Every seven seconds, a new research paper is added to the treasure trove of scientific literature (Science, 2012). The volume of research has doubled in the past 10 years (Bornmann & Mutz, 2015). As the amount of published information continues to grow exponentially, it is increasingly difficult to get an accurate picture of what is out there, especially on a global scale.

We take a bottom-up approach to inform the questions in this report. Rather than being intimidated by the volumes of research out there, we dive head-first into it, using new technologies that are designed to handle classification tasks with speed and accuracy.

Advancements in artificial intelligence (AI) and machine learning (ML) can help us use the data we already have, and this can be a highly effective way of surfacing relevant insights from a large and representative dataset (Gil et al., 2014).

The academic and development sectors are far behind the business sector in using big data and ML approaches to improve decision making. But these approaches are essential. We believe that the next big thing will not only come from major scientific breakthroughs, but also from sequencing millions of data points over time to observe how they interact with each other and where major gaps exist.

Mining the Gap used Havos.Ai machine learning models and data summarizing more than 1.2 million reports and papers from development and research organizations, UN agencies, and peer-review journals to create a mega-map of agricultural research, as shown in Figure 1.

Data partner CABI provided access to 1.2 million citations from applied life sciences around the world. CABI was a natural partner because of their long standing commitment to open up the world's literature and catalogue resources from small publishers and repositories, Research from top global publishers *Science* and *Nature*, as well as hundreds of disciplinary giants like *Food Policy, Field Crops Research, Global Food Security, African Journal of Crop Science, and Phytopathology* are included in the dataset. The output of global institutions, including work spanning OneCGIAR, are also included. So are materials from smaller scholarly publications, many of them non-English, such as *Revista Mexicana de Ciencias Pecuarias* and *Atti dell'Accademia dei Georgofili.*

OUR METHODS

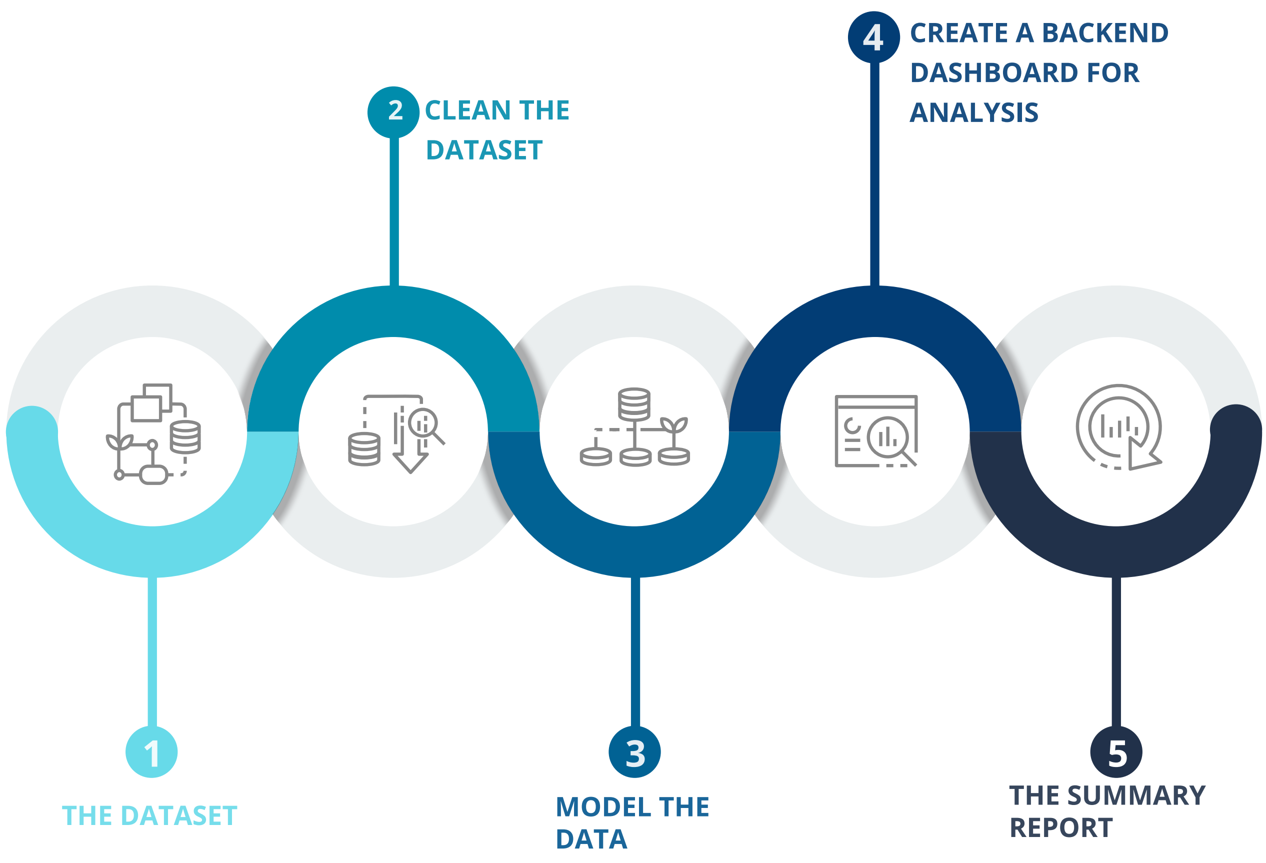
Working with CABI for this project enabled us to draw from a comprehensive, representative dataset to which we could seamlessly apply the Havos.Ai model and create a macro view of research from the past 20 years. The utility of this study is to provide a broad look and explore new ways of approaching the evidence base, and to add in new content from other sources over time.

Machine Learning (ML) seeks to detect patterns in data in a context of stochasticity ("noise"). At its most basic level, machine learning algorithms use historical data to learn patterns and uncover relationships. However, ML will almost always find a "pattern" — whether or not the identified pattern is insightful is revealed by the interpretation of humans.

Training is an important feature in machine learning. The more high-quality data that the system is presented with, the more refined the model. And, since ML generates data-driven predictions, knowledge about the sources of the data is essential for the results of ML to be useful. Scientific data gives us a higher probability of producing reliable, accurate results.

We create an ML pipeline for the more than 1.2 million articles in the dataset. Specific information is extracted from each article, all based on a series of questions and answers (Porciello et al., 2020). This helps us approach the literature to ask a series of modular questions, where we harmonize and clean the data before presenting the analysis to human experts. For instance, the kinds of questions we're interested in answering in this work include, does the research describe any interventions and outcomes include: and if so, what are they? What kind of study methods did the authors use? Who is the study population? Which crops or livestock are mentioned? Where is the research taking place? Additional technical details about the classification and the model ensemble are available in the annex.

FIGURE 2: An overview of the AI-assisted process



STUDY LIMITATIONS

There are important limitations of this study. First, the aim of this study is to surface relevant insights across studies using only summary title, abstract and other available metadata (such as keywords). (The exception is the analysis on funders and institutions[1].) We are necessarily limited in our observations and analysis based on what we can reasonably expect to learn from summary data.

Additional analysis of the corresponding full-text is possible using similar methods, and would provide additional insight into the quality of the research, including determining whether the underlying data supports the claims made in the summary data.

Second, additional analysis is needed to evaluate how the identified interventions and outcomes are supported within the body of the study. Ceres2030 global researchers tackled this problem across eight priority research areas, finding that only about 2% of the available evidence base had enough high-quality data to support their small-scale-producer-focused research questions.

Third, this is a representative dataset of 1.2 million summary titles, abstracts, and metadata from CABI databases and from other sources. CABI was targeted as a resource because of CABI's mission to identify and aggregate research from the Global South; it is among the best in the world for our purposes. But there are known gaps, such as large landscape reports from multinational agencies and NGOs. Thus, the research mapping will change over time if new information is incorporated in the dataset.

1 The exception is the analysis on funders and institutions. Use of summary metadata is specific for this report and a departure from other use cases, where the model has conducted full-text extraction and analysis, such as IFAD's Big Data Challenge and a scoping study, Agriculture in the Digital Age supported by the Bill & Melinda Gates Foundation and USAID.

OUR FINDINGS

PILLARS OF INNOVATION

Food systems can help us demonstrate the interconnectedness of the world of food and sustainable agriculture. For instance, technical innovation in crop breeding, seeds and storage facilities has increased productivity and yields so that fewer people will go hungry. Better management of limited natural resources through ecosystem services, like water and soil, protects biodiversity and fosters planetary health. We need thriving markets and roads to connect them to distribute and sell healthy, safe food that encourages diet diversity and food security. And, stable governments and other enabling systems are needed to continue to advance opportunities to increase education and eradicate poverty.

We identify how research publications cluster together, represented in Figure 3, across three pillars of agricultural innovation: technical, socioeconomic, and ecosystem services. Within each pillar are the top nine intervention areas, based on quantity of research.

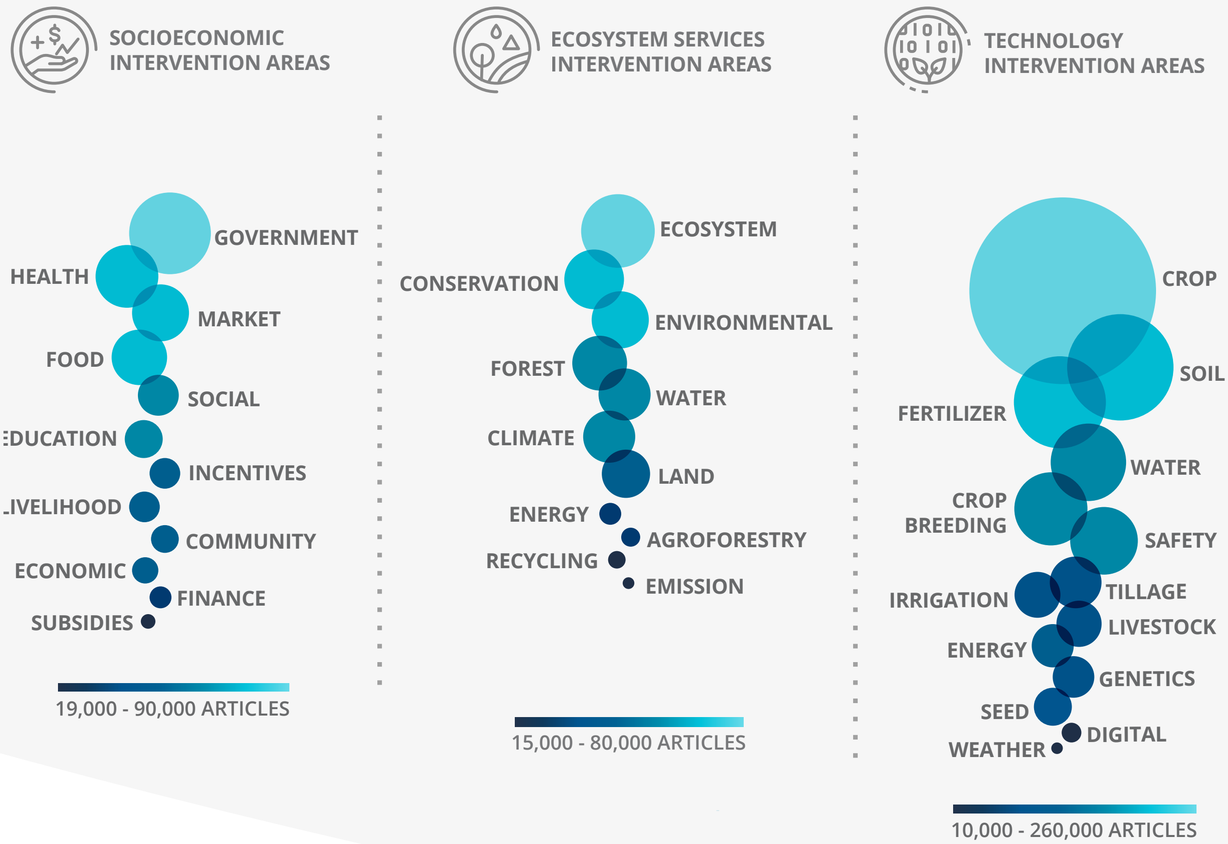
There are nearly twice as many research publications focused on technology innovations as compared to both ecosystem services and socioeconomic innovations. Research about crop and soil sciences, and the use of fertilizer, is well-represented, whereas outputs about emerging domain areas like digital agriculture are relatively small. There is less research being published about government, market, food and social interventions, as well as ecosystem services literature focused on conservation, water and forest.

Some of the names of the intervention area are the same, such as 'water' appearing as a domain areas in both the ecosystem services and technology pillar. In general, the underlying evidence base is comprised of a different suite of publications for each intervention area. Technological innovations for water (for instance) are more concerned with measuring soil respiration dynamics based on different precipitation patterns, whereas ecosystem services focuses more on management and use of water as a natural resources, such as nutrient recycling.

MORE ON METHODS

The pillars and domain areas were created using an AI-assisted clustering technique, where all of the supporting content (in this case, summary data) is examined in a vector space. We apply different algorithms in order to test different patterns that show us emergent relationships. Once these relationships emerge, we conduct an information extraction process to explore the underlying interventions that comprise each domain area. We can generate more specificity per domain area than what is pictured in Figure 3, down to the level of specific interventions. An expanded list of the domains is available in the annex, as well as more specifics about the methods; also provided in the annex are the specific outcomes that were captured in each category.

FIGURE 3. **The figure is organized first by pillar, and then intervention areas in each pillar. Each domain area is labeled with a single word that best represents the underlying research publications. The size of the intervention bubble corresponds to the number of articles supporting each intervention, and a scale indicating the general size of the document corpus is presented in each pillar.**



A selection of research outcomes—economic growth, healthy people, healthy planet, and gender & inclusion—were also captured across all articles that reported an outcome. Figure 4 shows the relationships between pillars, domains and outcomes, where each domain was assessed against each of the four outcome areas.

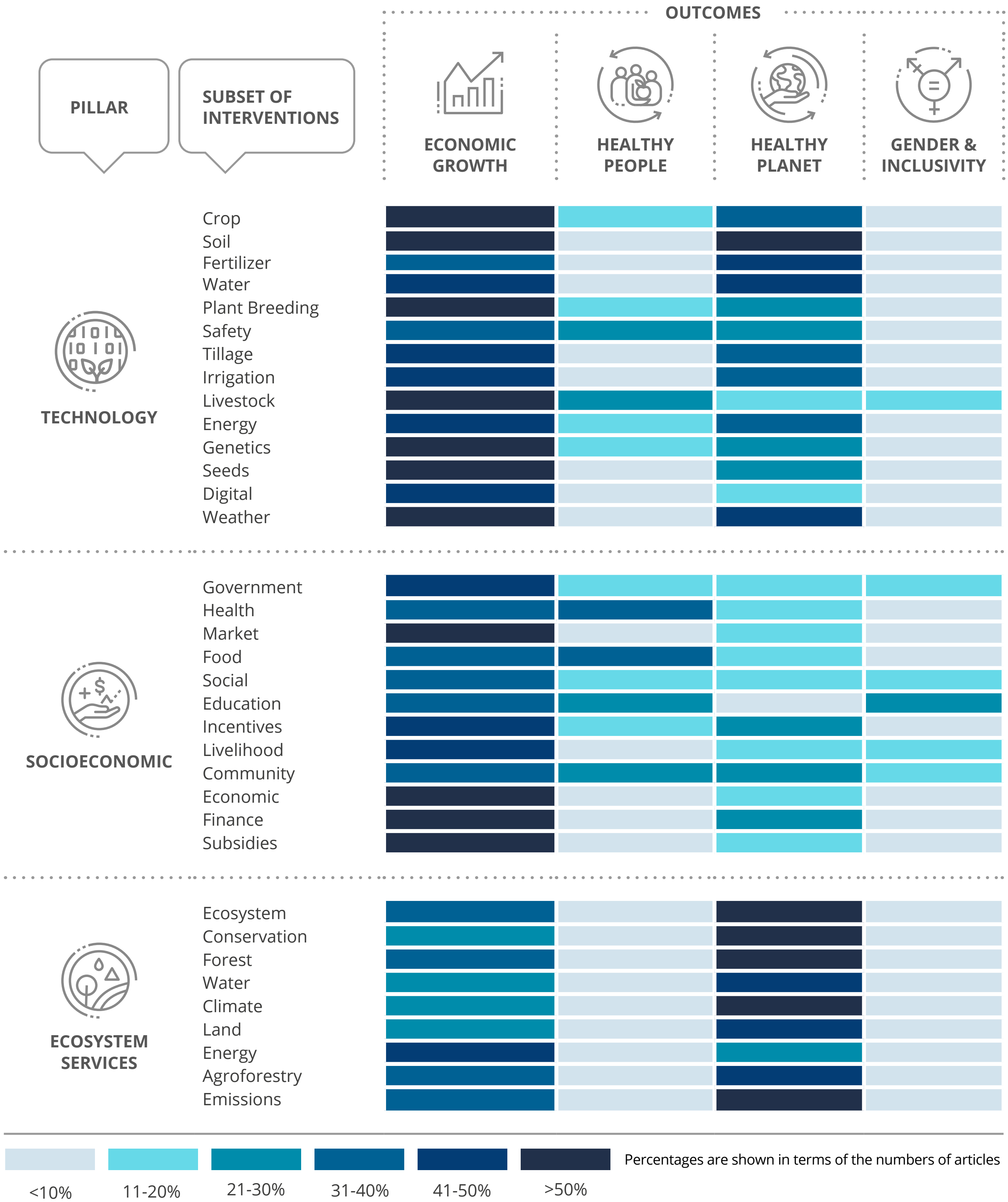
More research has focused on economic outcomes, such as productivity and yield, than any other area, and researchers are also actively trying to incorporate outcomes that measure water use and soil health (captured under Healthy Planet). But there are consistent gaps in the evidence for outcomes focused on nutrition, social inclusion, and gender empowerment across nearly every domain.

Such findings are not surprising. It reinforces the message that is continuously stated, that over-emphasis in any one domain area—or one pillar—cannot achieve gains across the entire system. We can see that focus on any single domain will be ineffective. Instead, an integrated approach across domains helps us achieve the greatest gains (Barrett et al., 2020; Laborde et al., 2020).

Over-emphasis in any one intervention area—or one pillar—cannot achieve gains across the entire system. We can see that focus on any single intervention will be ineffective. Instead, an integrated approach across interventions helps us achieve the greatest gains

FIGURE 4. A mapping of evidence pillars and intervention areas according to four outcome areas. Read the map from left to right, starting with domain area, and then from the top. This shows how each interventions area is faring in terms of studying specific outcomes.

Evidence Pillars & Gaps in Research Outputs



VIEW FROM WITHIN

The vision for the future of sustainable agriculture requires us to move fast, to identify innovation as it is emerging. A challenge we collectively face is how to identify when emerging research is in the pipeline, especially when the innovations come from specialized scientific and private sector research programs.

Most of the sciences have a 'death valley': a desert between scientific data and applied processes that can help us make use of those findings for innovation.

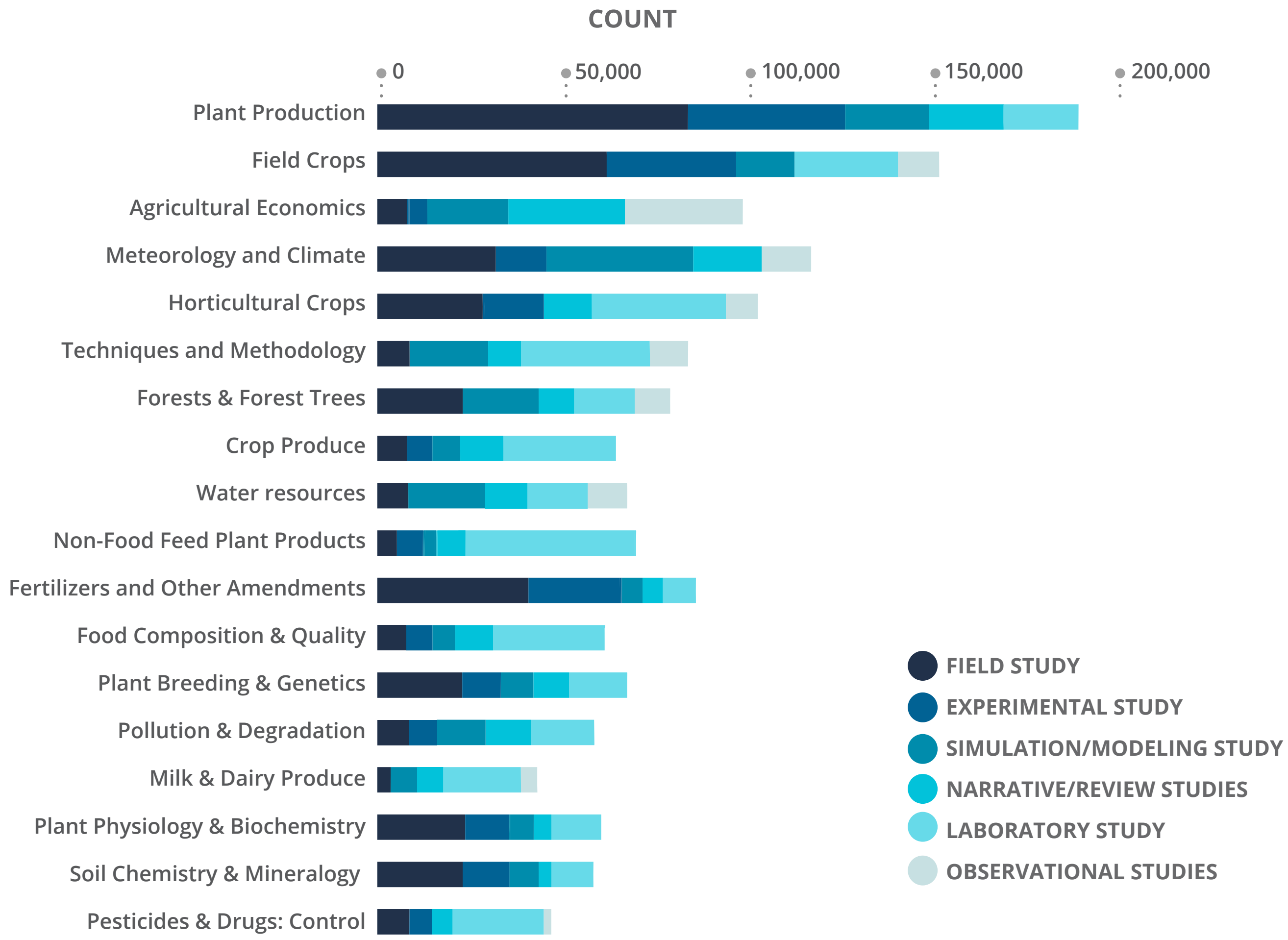
Most of the sciences have a 'death valley': a desert between scientific data and applied processes that can help us make use of those findings for innovation. Our universities and research centers are primarily equipped to support and conduct novel, basic research investigation.

What are the ways we see this emerging in the research? First, publication trends favor field, laboratory, simulation, and narrative studies (Figure 5). There are less experimental (clinical studies in nutrition, and impact evaluations) and observational studies. We show these trends using a selection of standardized CABI codes.

Other studies have highlighted that within the papers themselves, there tends to be a stronger focus on household and farm-level analyses and relatively little attention to the landscape or macro level analyses that are so important in the context of scaled-up interventions (Barrett et al., 2020; Liverpool-Tasie et al., 2020; Ricciardi et al., 2020).

In addition, despite an emphasis on household and farm-level outcomes, there are some serious gaps regarding what we know about the study populations themselves, where even basic demographic information about the study population, like age, sex, and education is missing (Acevedo et al., 2020).

FIGURE 5. **Research study types (definitions in the annex) are shown according to a selection of research disciplines. The names of the research disciplines come from standard CABI codes, which are applied to each article in the CABI's databases.**



Addressing the dearth of research that focuses on impact and causal pathways would require more opportunities to collect and harmonize data about innovations over time, and across different scales of research: farm-level, landscape or macro, and food systems. This is a long-standing discussion from organizations like Standing Panel on Impact Assessment (SPIA) that that aim to better link evaluation, interventions and outcomes across a common set of agreed-upon guidelines that, although they originate at the project or program level, could also be useful for exploring impact that is policy relevant.

Funders and governments are increasingly interested in seeing science's causal and applied impact, especially in countries where issues like poverty and food security loom largest. Among the most frequently acknowledged funders are the World Bank, European Commission, USAID, the Bill & Melinda Gates Foundation, and the Asian Development Bank. This is based on a sampling of the “acknowledgements section” across ~37,000 research papers because funding data is not systemically requested by all journals—though this is changing. But right now, it is difficult to generate an accurate mapping of research funding trends based solely on the scientific papers themselves.

There are some encouraging trends. The publications emerging about the Global South are primarily Global South led. The affiliations of the first author publishing the paper are overwhelmingly from institutions based in the Global South. This counters some of the narratives about Global North researchers dominating research production. Instead, there is a healthy representation of regional organizations (primarily academic organizations) that emerge as the top research producers in their own regions. Hopefully this will aid research prioritization so existing capacity can flourish.

WHERE IS THE EVIDENCE?

China, Brazil and India lead the way in publishing research outputs, with more modest results appearing across all of the other included countries. These trends follow even when we look at the cross-section views of crop, livestock and value chain research across the Global South. (Figure 6). Different countries and regions come into focus depending on the target crops, as highlighted in the maps below (Figures 7-12).

Of note is a particular lack of research about fruits and vegetables in Sub-Saharan Africa (Pingali, 2015). This is of concern because expanding fruits and vegetables in the food supply and reducing the concentration on cereals like rice, maize, and wheat is essential for improving health and reducing the incidence of non-communicable diseases (Fanzo et al., 2017). And given that Sub-Saharan Africa is where rural populations are predicted to increase significantly in the next 20 years, we would expect a considerable expansion in demand for fruits and vegetables.

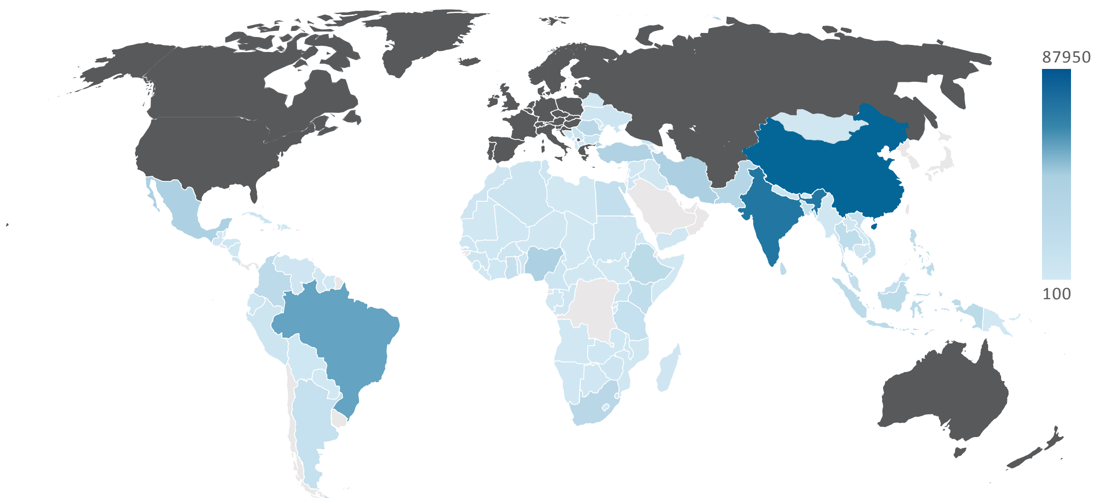
The expansion of fruit and vegetable production could provide increased opportunities for small-scale farmers, conditional on market access and technical capacity. Expansion of decent employment opportunities in food value chains will be key to improving livelihoods in areas of high rural poverty such as sub-Saharan African and South Asia, since the small farm sector is incapable of absorbing the expected increases in the labor force to provide opportunities for decent employment (*Rural Development Report 2021 – Transforming Food Systems for Rural Prosperity*, n.d.).

The extent to which research outputs focus on value chains and post-production processes including storage, distribution, or marketing channels, is a well-documented gap in the evidence base, and is reflected in the domain mapping (Liverpool-Tasie et al., 2020). A recent and exhaustive evidence synthesis examining post-harvest loss reduction concluded that there is a lack of studies on training, finance, infrastructure, policy and market interventions (Stathers et al., 2020).

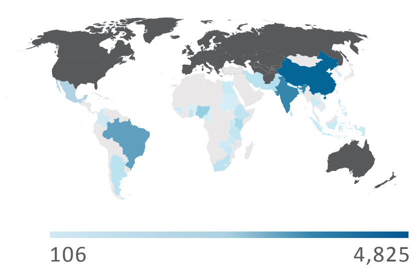
Value chains will play a key role in directing incentives and signals to the producers of food and agricultural products, as well as the consumers (Reardon, Echeverria, et al., 2019). Their fundamental role in the transformation of food systems to improve the livelihood, nutrition and environmental performance of the world population was a major theme in the recent UN Food System Summit.

FIGURE 6. Crop Research by Country shows research outputs by geography. Countries in black are excluded from this analysis. The calculation to identify countries within the research output is based on whether the geography is the area of focus where the research took place. Note that multiple countries can, and are, identified within one study. Specific Crop Research output is shown by country using the same calculations and coloring, and shown here by 1) maize; 2) rice; 3) wheat; 4) fruits and vegetables; 5) roots, tubers and bananas; and 6) livestock. The classifications are based on machine learning that uses a custom harmonized thesaurus based on plants and animals from AGROVOC and National Agricultural Library.

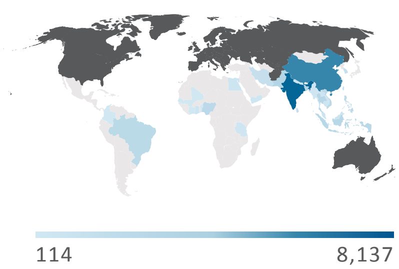
CROP RESEARCH BY COUNTRY



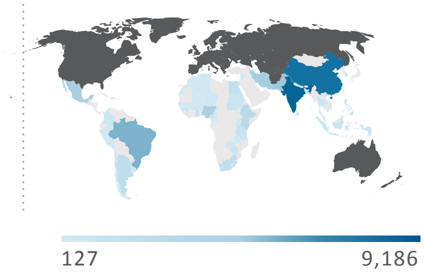
MAIZE



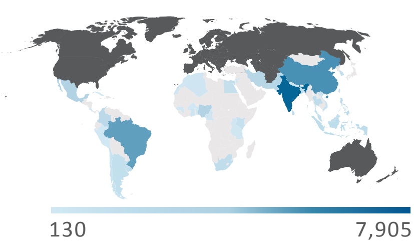
RICE



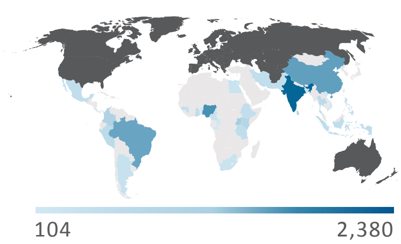
WHEAT



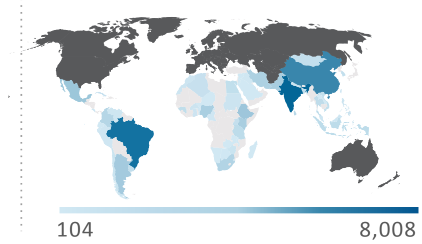
FRUITS AND VEGS



ROOTS, TUBERS AND BANANAS



LIVESTOCK



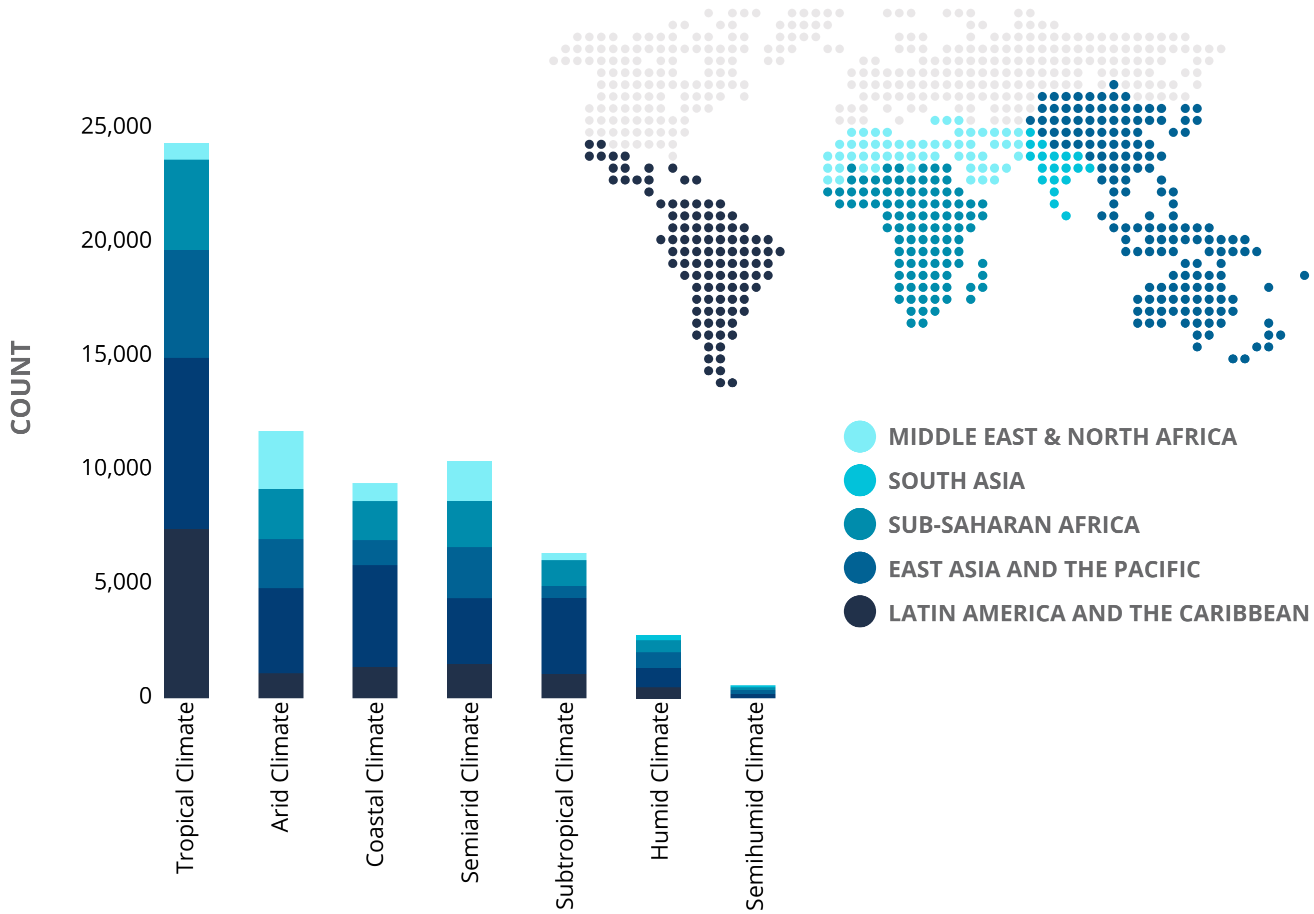
OUR CHANGING CLIMATE

We highlight the degree to which agriculture research outputs are focused on different climate zones such as tropical climate and arid climate (shown in Figure 14), per region. This is an early exploration to improve our understanding of the work researchers are doing to figure out agriculture's adaption to climate change. Here, we see that nearly twice as much work has been focused on issues facing tropical climates than arid climates, with some on coastal areas.

The acceleration of climate change means that the biodiversity inhabiting each climate zone will have less time to adapt to the climatic change. Ecosystem services will play an increasingly critical role to protect biodiversity and shared natural resources. It is also essential for food security for many indigenous communities that rely on food gathered from natural ecosystems, such as oceans. As shown in the pillars and evidence gap overview, ecosystem services emerged with the least amount of research inputs, and it is unclear how integrated these innovations really are within the other two socioeconomic and technology pillars.

The Earth's climate zones will continue to shift at an accelerated pace, and many climate scientists suggest that monitoring shifts in climate zones is a reasonable measure of 'reality' for living systems, including agriculture (Mahlstein et al., 2013).

FIGURE 7. **The frequency of articles per region according to different climate zones. The naming of the climate zones are recognized by FAO's AGROVOC.**



WHO ARE WE INCLUDING IN THE RESEARCH?

Understanding complex social factors about user groups is a cornerstone of both research and development.

We sought to identify who is included in research outputs by looking for information about the study populations, relying on basic demographic details as a proxy. For instance, we explored various employment: farmers or agricultural workers, including (but not limited to) small-scale producers, agribusiness dealers, value chain actors, extension service agents, and others. We investigated whether research outputs contained generalized descriptions, such as ages of a study population (adults, elderly, youth), or the sex/age range of a study population (women, men, girls, boys); and many other sociodemographic descriptions, such as mothers, indigenous, tribal or nomadic populations, and more[2].

The descriptions of farmers and agricultural workers is extremely ambiguous, and rarely includes contextual clues about farm size and type that are useful to discern who is really included. What we observe is that 10% of the literature mentions the general term 'farmer' without other contextual details, and 3% specifically identify small-scale producer. But, even this term is a somewhat complicated by the fact that a publication describing 'small-scale producers' from Brazil is featured alongside 'small-scale producers' from Malawi, even though farm sizes are quite different.

2 Livestock animals were excluded as a specific population of research

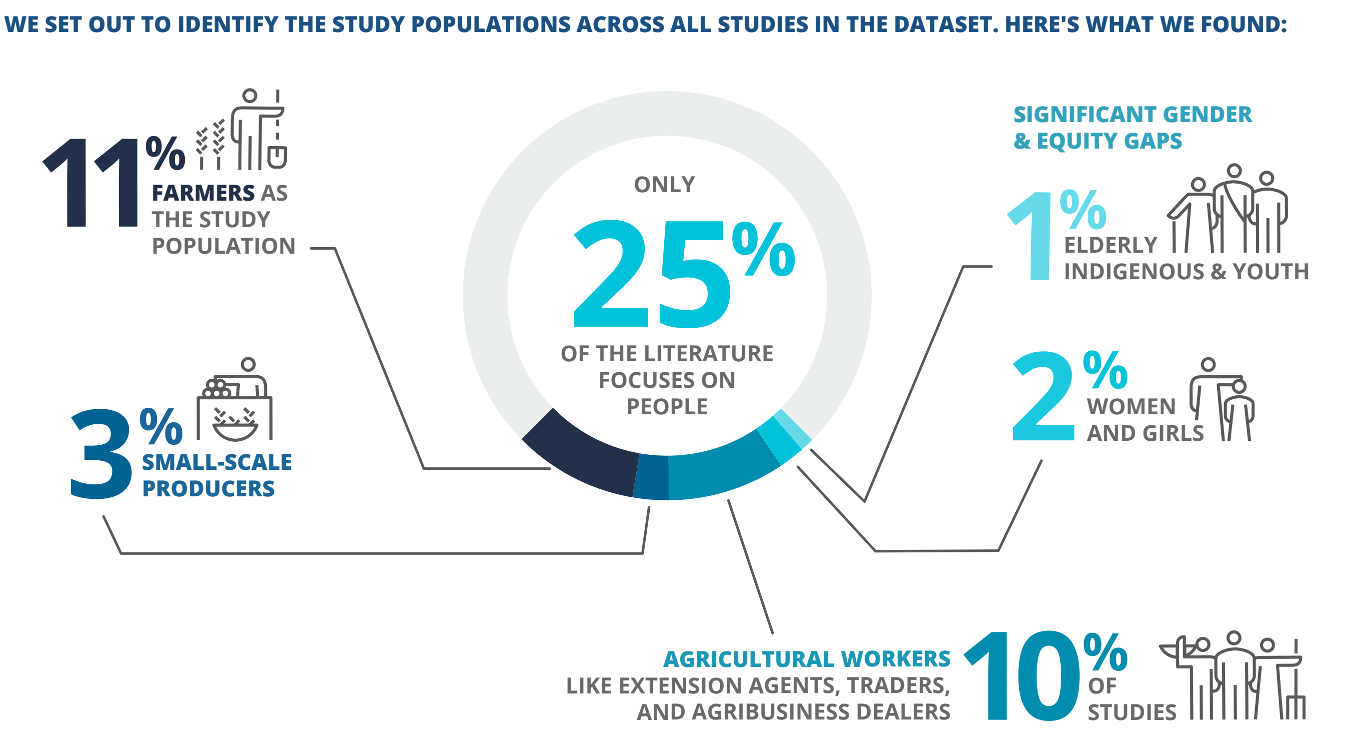
As highlighted in Figure 8, women are underrepresented, and both elderly and youth populations were rarely mentioned. Studies focusing on these groups are usually in association with health or nutrition outcomes. Information about communities such as tribal and indigenous populations, or research focused on other areas of equity, such as wealth, access to finance, education and literacy, is sparse, often in the low thousands.

More work needs to be done to capture information about all of the beneficiary communities. Overlapping social factors such as education, socioeconomic status, race, class and gender, can create interdependent systems of discrimination and disadvantage which reinforce the exclusion of some groups—particularly, but not only, women—from the benefits of agricultural innovations. Additional and sustained work in this area will reduce the likelihood of making generalized, homogenous assumptions for heterogeneous groups.

Given that the vast majority of published science is focused on basic, upstream experiments in which study populations are not part of the work, many of the gaps we identified make some sense. But, given that small-scale producers are the focal point of Sustainable Development Goal 2 (SDG2), these gaps are still startling from the perspective of research prioritization.

It is tempting to conclude that we must be missing information because we are only looking at titles and abstract data—and certainly this is a possibility we must keep in mind. But comparable and recent research by global research teams took a painstaking look at the underlying research papers and found similarly startling numbers, including the comprehensive Ceres2030 report that highlighted a massive under-investment in research for small-scale farms in the Global South ("Ending Hunger," 2020).

FIGURE 8. User Groups



SPOTLIGHT:

INSTITUTIONS, POLICY AND FINANCE

Research and innovation in policy, institutions, and finance will play a large part in developing the transformative changes needed to address complex challenges in agricultural systems.

Institutions and policy instruments have the power to promote or block broader transformation in this sector. In general, weak institutions are a key obstacle for small and poor farmers, and some of the research on policy instruments emphasizes single-solution instruments that may provide guidance and site-specific recommendations, but less about comparative analysis of multiple instruments, and its applicability in another contexts (Piñeiro et al., 2020; Zilberman et al., 2018).

Institutions, policies, and financing mechanisms are key to achieving change, especially to ensuring that farmers have the resources that they need to succeed. Within the research domains, however, the areas of focus are not immediately evident. One exception would be farmer organizations (FOs), such as associations, cooperatives, self-help and women's groups, and the extent to which they are empowered to work with all farmers (Bizikova et al., 2020).

Understanding the interlocking role that institutions, policies and finance have a critical need. They face different sets of constraints and opportunities, and thus they merit specific attention.

CONCLUSION

Agriculture (and, more broadly, food systems) is an incredible node that touches many issues and disciplines. However, such diffusion can make it incredibly challenging to work from the same evidence playbook. Agriculture cannot be either/or, it must be And.

By taking a birds eye view of research across disciplines spanning the three pillars of agricultural innovation (technical, socioeconomic, and ecosystem services), our findings reinforced the message that integrated approaches across interventions are more effective in achieving gains across the entire food system.

Our efforts to map and analyze the evidence pointed to some key gaps.

Not all underfunded areas can be treated equally. There are many areas of research that are underfunded, but some of those areas may result in more significant trade-offs than others. Research into fruits and vegetables (both in production and post-harvest), is one example where we risk greater challenges for healthy diets and diversity if this does not emerge as a key research priority. So is biodiversity. And whether we are studying traditional local systems that link to broader markets through intermediaries or larger industrialized and global systems, many food systems correlate with their location, so it is key to understand the geographic distribution of the evidence base. Likewise, it matters where in the world we set-up our research programs and the partnerships that are created.

The essentialness of equity. It is clear that too little is being captured and reported about study populations, including basic sociodemographic details, such as employment, age and sex. Equally important, however, is the capture of social factors that could underscore how barriers are systematic for some communities and not for others. As we look towards the future of research prioritization, equity outcomes need to become more pronounced.

Connecting research and innovation pathways. The research pipeline for agriculture is extremely long, and a decade or more can pass before some technologies (like nutritionally fortified crops) see results in farmers' fields. Despite this, it is challenging to connect and trace upstream and downstream research in any observable way, making it difficult to find pathways that can scale research to innovation in the market.

Beyond farm and household level outcomes. We also need to go beyond capturing research that reports impact at the household and farm-level to produce more evidence about impacts at the macro and food systems levels.

New technologies to share and unleash scientific potential. We know that the next big thing will not just come not from one idea or one platform, but by sequencing millions of small details on similar problems from researchers across the world. In the race to develop the food system of the future so that innovation can flourish, we need to analytic tools and databases that help us make short work of the hay and present a stack of needles.

About CoSAI:

The Commission on Sustainable Agriculture Intensification (CoSAI) was set up to promote more and better investment in innovation for Sustainable Agriculture Intensification (SAI) for the Global South, in support of the Sustainable Development Goals (SDGs). For CoSAI, innovation includes not only science and technology but also innovation in policies, finance, and social institutions. CoSAI has a timeline running up to December 2021.

CoSAI has six Commissioner Working Groups addressing Big Questions around innovation for SAI. Working Group 2 focuses on priorities for innovation. Some of the work already commissioned under this working group includes two studies on global funding for innovation in SAI (Investment Baseline and Investment Gap studies) and a study on instruments and approaches for innovation in SAI.

CoSAI is building up an evidence base to support the case for increased and better-targeted investment in agricultural innovation for the Global South. This includes studies on the investment baseline and projected investment gap, approaches and instruments, learning from case studies on pathways to innovation, and a Taskforce on Principles and Metrics.

About Havos Inc.

Havos.Ai builds software solutions and platforms for global organizations that want to use advanced computation for complex, open-ended problems that are beyond the scope of individual decision-making. Our approach taps into collective intelligence and wisdom of global experts, supported by artificial intelligence and the best scientific data.

Founded in 2021 by leaders in science, policy and industry, Havos.Ai improves decision-making for governments, multilateral agencies, funders, and research organizations. The company emerged as a start-up out of Cornell University.

List of Works Cited

Acevedo, M., Pixley, K., Zinyengere, N., Meng, S., Tufan, H., Cichy, K., Bizikova, L., Isaacs, K., Ghezzi-Kopel, K., & Porciello, J. (2020). A scoping review of adoption of climate-resilient crops by small-scale producers in low- and middle-income countries. *Nature Plants*, *6*(10), 1231–1241. https://doi.org/10.1038/s41477-020-00783-z

Barrett, C. B., Benton, T. G., Cooper, K. A., Fanzo, J., Gandhi, R., Herrero, M., James, S., Kahn, M., Mason-D'Croz, D., Mathys, A., Nelson, R. J., Shen, J., Thornton, P., Bageant, E., Fan, S., Mude, A. G., Sibanda, L. M., & Wood, S. (2020). Bundling innovations to transform agri-food systems. *Nature Sustainability*, *3*(12), 974–976. https://doi.org/10.1038/s41893-020-00661-8

Bizikova, L., Nkonya, E., Minah, M., Hanisch, M., Turaga, R. M. R., Speranza, C. I., Karthikeyan, M., Tang, L., Ghezzi-Kopel, K., Kelly, J., Celestin, A. C., & Timmers, B. (2020). A scoping review of the contributions of farmers' organizations to smallholder agriculture. *Nature Food*, *1*(10), 620–630. https://doi.org/10.1038/s43016-020-00164-x

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215–2222. https://doi.org/10.1002/asi.23329

Davidson, D. (2016). Gaps in agricultural climate adaptation research. *Nature Climate Change*, *6*(5), 433–435. https://doi.org/10.1038/nclimate3007

Ending hunger: Science must stop neglecting smallholder farmers. (2020). *Nature*, *586*(7829), 336–336. https://doi.org/10.1038/d41586-020-02849-6

Fanzo, J., Arabi, M., Burlingame, B., Haddad, L., Kimenju, S., Miller, G., Nie, F., Recine, E., Serra-Majem, L., & Sinha, D. (2017). Nutrition and food systems. *A Report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security*.

Gil, Y., Greaves, M., Hendler, J., & Hirsh, H. (2014). Amplify scientific discovery with artificial intelligence. *Science*, *346*(6206), 171–172. https://doi.org/10.1126/science.1259439

Herrero, M., Thornton, P. K., Mason-D'Croz, D., Palmer, J., Benton, T. G., Bodirsky, B. L., Bogard, J. R., Hall, A., Lee, B., Nyborg, K., Pradhan, P., Bonnett, G. D., Bryan, B. A., Campbell, B. M., Christensen, S., Clark, M., Cook, M. T., de Boer, I. J. M., Downs, C., … West, P. C. (2020). Innovation can accelerate the transition towards a sustainable food system. *Nature Food*, *1*(5), 266–272. https://doi.org/10.1038/s43016-020-0074-1

Laborde, D., Porciello, J., Smaller, C., Murphy, S., & Parent, M. (2020). *Ceres2030: Sustainable Solutions to End Hunger Summary Report* [Report]. Ceres2030. https://ecommons.cornell.edu/handle/1813/72799

Liverpool-Tasie, L. S. O., Wineman, A., Young, S., Tambo, J., Vargas, C., Reardon, T., Adjognon, G. S., Porciello, J., Gathoni, N., Bizikova, L., Galiè, A., & Celestin, A. (2020). A scoping review of market links between value chain actors and small-scale producers in developing regions. *Nature Sustainability*, *3*(10), 799–808. https://doi.org/10.1038/s41893-020-00621-2

Mahlstein, I., Daniel, J. S., & Solomon, S. (2013). Pace of shifts in climate regions increases with global temperature. *Nature Climate Change*, *3*(8), 739–743. https://doi.org/10.1038/nclimate1876

Piñeiro, V., Arias, J., Dürr, J., Elverdin, P., Ibáñez, A. M., Kinengyere, A., Opazo, C. M., Owoo, N., Page, J. R., Prager, S. D., & Torero, M. (2020). A scoping review on incentives for adoption of sustainable agricultural practices and their outcomes. *Nature Sustainability*, *3*(10), 809–820. https://doi.org/10.1038/s41893-020-00617-y

Pingali, P. (2015). Agricultural policy and nutrition outcomes – getting beyond the preoccupation with staple grains. *Food Security*, *7*(3), 583–591. https://doi.org/10.1007/s12571-015-0461-x

Porciello, J., Ivanina, M., Islam, M., Einarson, S., & Hirsh, H. (2020). Accelerating evidence-informed decision-making for the Sustainable Development Goals using machine learning. *Nature Machine Intelligence*, *2*(10), 559–565. https://doi.org/10.1038/s42256-020-00235-5

Reardon, T., Echeverria, R., Berdegué, J., Minten, B., Liverpool-Tasie, S., Tschirley, D., & Zilberman, D. (2019). Rapid transformation of food systems in developing regions: Highlighting the role of agricultural research & innovations. *Agricultural Systems*, *172*, 47–59. https://doi.org/10.1016/j.agsy.2018.01.022

Reardon, T., Lu, L., & Zilberman, D. (2019). Links among innovation, food system transformation, and technology adoption, with implications for food policy: Overview of a special issue. *Food Policy*, *83*, 285–288. https://doi.org/10.1016/j.foodpol.2017.10.003

Ricciardi, V., Wane, A., Sidhu, B. S., Godde, C., Solomon, D., McCullough, E., Diekmann, F., Porciello, J., Jain, M., Randall, N., & Mehrabi, Z. (2020). A scoping review of research funding for small-scale farmers in water scarce regions. *Nature Sustainability*, *3*(10), 836–844. https://doi.org/10.1038/s41893-020-00623-0

Rural Development Report 2021 – Transforming food systems for rural prosperity. (n.d.). IFAD. Retrieved October 18, 2021, from https://www.ifad.org/en/web/latest/-/rdr-2021-launch

Serraj, R., & Pingali, P. (2018). *Agriculture & Food Systems to 2050: Global Trends, Challenges and Opportunities* (Vol. 2). World Scientific.

Stathers, T., Holcroft, D., Kitinoja, L., Mvumi, B. M., English, A., Omotilewa, O., Kocher, M., Ault, J., & Torero, M. (2020). A scoping review of interventions for crop postharvest loss reduction in sub-Saharan Africa and South Asia. *Nature Sustainability*, *3*(10), 821–835. https://doi.org/10.1038/s41893-020-00622-1

Webb, P., & Kennedy, E. (2014). Impacts of Agriculture on Nutrition: Nature of the Evidence and Research Gaps. *Food and Nutrition Bulletin*, *35*(1), 126–132. https://doi.org/10.1177/156482651403500113

Zilberman, D., Lipper, L., McCarthy, N., & Gordon, B. (2018). Innovation in response to climate change. In *Climate smart agriculture* (pp. 49–74). Springer, Cham.

MINING THE GAPS:

USING MACHINE LEARNING TO MAP A MILLION DATA POINTS FROM AGRICULTURAL RESEARCH FROM THE GLOBAL SOUTH

Porciello, Jaron; Bourne, Thomas; Lipper, Leslie; Lin, Sammi; and Langleben, Sarah. 2021. *Mining the Gaps: Using Machine-Learning to Map a Million Data Points on Agricultural Research from the Global South*. Colombo, Sri Lanka: Commission on Sustainable Agriculture Intensification.

Copyright © 2021, Commission on Sustainable Agriculture Intensification

ATTRIBUTION: The work must be referenced according to international citation standards, while attribution should in no way suggest endorsement by WLE, IWMI or the author(s).

NON-COMMERCIAL: This work may not be used for commercial purposes.

Note: *This report uses proprietary datasets created through secondary research. Hence, unless mentioned otherwise, data and graphs in this report are derived from this database.*

havos.org

@AiHavosjp

@havos.org