



**HAL**  
open science

# The analysis of data from on-farm research network : Statistical approaches to test the efficacy of management practices and data visualization

Anabelle Laurent

## ► To cite this version:

Anabelle Laurent. The analysis of data from on-farm research network : Statistical approaches to test the efficacy of management practices and data visualization. Bacteriology. Université Paris-Saclay; Iowa State University, 2020. English. NNT : 2020UPASB022 . tel-03464120

**HAL Id: tel-03464120**

**<https://pastel.hal.science/tel-03464120>**

Submitted on 3 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The analysis of data from on-farm research network: Statistical approaches to test the efficacy of management practices and data visualization

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 581 Agriculture, alimentation, biologie, environnement et santé (ABIES)

Spécialité de doctorat : Sciences agronomiques

Unité de recherche : Université Paris-Saclay, AgroParisTech, INRAE, UMR Agronomie, 78850, Thiverval-Grignon, France.

Référent : AgroParisTech

Thèse présentée et soutenue à Iowa State University, le 9 Décembre 2020, par

**Anabelle Claudine LAURENT**

## Composition du Jury

**Kendall LAMKEY**

Professor, Department of Agronomy, Iowa State University

**Peter KYVERYGA**

Director of Analytics, Iowa Soybean Association

**Nicolas MARTIN**

Assistant professor, Crop Sciences, University of Illinois

**Emily HEATON**

Professor, Department of Agronomy, Iowa State University

**Heike HOFMANN**

Professor, Department of Statistics, Iowa State University

**Nicolas GUILPART**

Maître de conférences, AgroParisTech

Président

Rapporteur & Examineur

Rapporteur & Examineur

Examinatrice

Examinatrice

Examineur

**David MAKOWSKI**

Directeur de recherche, INRAE, UMR Agronomie

**Fernando MIGUEZ**

Professor, Department of Agronomy, Iowa State University

**Mark LICHT**

Assistant professor, Department of Agronomy, Iowa State University

Directeur de thèse

Co-Directeur de thèse

Invité

## DEDICATION

A ma famille, pour votre soutien, votre confiance et votre amour qui m'a permis d'arriver jusque-là aujourd'hui.

## TABLE OF CONTENTS

	Page
NOMENCLATURE	vi
ACKNOWLEDGMENTS	vii
ABSTRACT	viii
CHAPTER 1. GENERAL INTRODUCTION	1
1.1 On-Farm Research: A Cooperation Between the Research Community and Farmers	1
1.2 The Different Purposes of On-Farm Research Network	2
1.3 The Importance of an Analytical Approach	2
1.4 Challenges and Needs	3
1.5 Dissertation and Organization	3
1.6 References	4
CHAPTER 2. A FRAMEWORK FOR VISUALIZATION AND ANALYSIS OF AGRONOMIC FIELD TRIALS FROM ON-FARM RESEARCH NETWORKS	6
2.1 Abstract	6
2.2 An Introduction to On-Farm Research Network	7
2.3 Developing Research Networks	7
2.3.1 Analyzing Data Across Experiments	8
2.4 Creating a Data-Analytics Framework	12
2.4.1 Preliminary Analysis	13
2.4.2 Defining Yield Responses	15
2.4.3 Importance of Replication	16
2.4.4 Yield Limiting Factors and Yield Response	17
2.4.5 Statistical Modeling	17
2.4.6 Calculating Economic Responses	21
2.4.7 Visualization	21
2.5 Case Studies	26
2.5.1 Foliar Fungicide Impact on Soybean Yields	26
2.5.2 Row Spacing Impact on Soybean Yields	31
2.5.3 Soil-Applied Insecticide Impact on Bt Corn Yields	34
2.6 Limitations of our Data-Analytics Framework	37
2.7 Summary	37
2.8 Software Availability	38
2.9 Acknowledgments	39
2.10 References	39
2.11 Supplemental Materials	42
CHAPTER 3. INTERACTIVE WEB-BASED DATA VISUALIZATION AND ANALYSIS TOOL FOR SYNTHETIZING ON-FARM RESEARCH NETWORKS DATA	43

3.1 Abstract	43
3.2 Introduction	44
3.3 Materials and Methods	47
3.3.1 General Architecture	47
3.3.2 Data Description	49
3.3.3 Statistical Model Description	50
3.3.4 Economic Responses	52
3.3.5 Tool Description	52
3.3.6 Technical Details	54
3.4 Case-study	56
3.5 Discussion	61
3.6 Conclusion	64
3.7 Acknowledgments	65
3.8 References	65
3.9 Supplemental Materials	68
CHAPTER 4. GOING BEYOND MEAN EFFECT SIZE; PRESENTING PREDICTION INTERVALS FOR ON-FARM NETWORK TRIAL ANALYSES	72
4.1 Abstract	72
4.2 Introduction	73
4.3 Materials and Methods	79
4.3.1 Data Description and Experimental Design	79
4.3.2 Data Analysis	80
4.3.2.1 Frequentist Method	83
4.3.2.2 Bayesian Method	86
4.3.2.3 Sensitivity Analysis of Priors for Between-Trial Variance	88
4.4 Results	91
4.4.1 A Detailed Example	91
4.4.2 Results Obtained for the On-Farm Research Networks with the Frequentist Method	93
4.4.3 Results Obtained for the On-Farm Research Networks with the Bayesian Method	95
4.4.4 Comparison Between Frequentist and Bayesian Methods	101
4.5 Discussion	104
4.6 Conclusion	109
4.7 Acknowledgments	109
4.8 References	109
4.9 Supplemental Materials	114
CHAPTER 5. ON-FARM TRIALS REVEAL SIGNIFICANT BUT UNCERTAIN CONTROL OF BOTRYTIS CINEREA BY AUREOBASIDIUM PULLULANS AND POTASSIUM BICARBONATE IN ORGANIC GRAPEVINES	119
5.1 Abstract	119
5.2 Introduction	120
5.3 Materials and Methods	124
5.3.1 Data Description and Experimental Design	124

5.3.2 Data Analysis	126
5.3.3 Statistical Inference	129
5.4 Results	131
5.4.1 Description of the Observed Proportions of Diseased Bunches	131
5.4.2 Statistical Modeling Evaluation	132
5.4.3 Estimated efficacy of the biopesticides	135
5.5 Discussion	141
5.6 Acknowledgments	145
5.7 References	145
5.8 Supplemental Materials	150
 CHAPTER 6. COMPARISONS OF CORN AND SOYBEAN YIELD RESPONSE IN SMALL-PLOT AND ON-FARM FOLIAR FUNGICIDE TRIALS	 158
6.1 Abstract	158
6.2 Introduction	160
6.3 Materials and Methods	165
6.3.1 Data Description	165
6.3.2 Data Analysis	169
6.4 Results	173
6.4.1 Descriptive Results	173
6.4.2 Results Obtained with the Bayesian Approach	178
6.5 Discussion	182
6.6 Acknowledgments	185
6.7 References	186
6.8 Supplemental Materials	189
 CHAPTER 7. GENERAL CONCLUSION	 193
7.1 Supplemental Materials	195
 LONG RESUME	 201

**NOMEMCLATURE**

CCB	Clariva Complete Beans
CMV	CruiserMaxx advanced plus Vibrance
GDD	growing degree days
GLRM	generalized linear random-effect model
GM	genetically modified
ISA	Iowa Soybean Association
ISOFAST	interactive summaries of on-farm strip trials
OFRN	on-farm research network
OFT	on-farm trial
RHR	relative heterogeneity ratio
se	standard error
SPT	small-plot trial
WAIC	widely applicable information criterion

## ACKNOWLEDGMENTS

This dissertation was made possible through the support and input of many people.

I would like to begin by acknowledging my co-advisors for their support, mentorship, and giving me the opportunity to conduct this research. I am grateful to have learned from you; I couldn't have hoped for better. Fernando, thank you for always striving to make me a better scientist. David, thank you for your guidance, for challenging me, and joining the cotutelle adventure.

I would also like to thank the other committee members, Peter Kyveryga, Emily Heaton, Heike Hofmann, Kendal Lamkey, Mark Licht, Nicolas Martin, and Nicolas Guilpart for their insightful comments and encouragement. I would particularly like to recognize Peter Kyveryga for our collaboration and his input in various aspects of my research. I thank Nicolas Martin and Nicolas Guilpart, for their time and for contributing to my cotutelle.

I would also like to thank the support staff in the Agronomy Department and especially MaryAnn Grapp, for happily providing me with the resources I needed.

Thank you to my fellow grad students and the Miguez lab group for their companionship, general support, and helpful discussions.

Thank you to my friends in France and to the ones I have made in Iowa (and who are here and there). You made my experience full of encouragement, kindness, and laughter. More than ever, friendships have been crucial to me!

I extend my thanks to my brother and my parents, especially my mom, my biggest cheerleader.



**ABSTRACT**

An on-farm research network is an organization of farmers that conducts agronomic experiments under local conditions. There is growing interest in on-farm research networks because they provide the infrastructure needed to test new products and management practices in farmers' fields. Often, the results are usually presented as individual reports (i.e., a report summarizing the outcome for one trial), but this provides limited information difficult to generalize and does not allow presenting, in a synthetic way, all the results collected from the different trials. Moreover, there is unexplored potential in detecting yield response variability patterns for better decision making. The overall objective of this thesis is to demonstrate the importance of identifying appropriate statistical methods for analyzing and visualizing on-farm research network data. Specifically, I focused on analyzing the on-farm research networks managed by the Iowa Soybean Association, and an adaptation was made with a French case-study. A data-analytics framework was developed to analyze multiple trials that use a common protocol and identify the conditions where an imposed treatment may or may not be effective. This framework used a random-effect model through a Bayesian approach and returned yield response estimates at the network and trial levels. The framework was implemented through a web-application for 51 different management practices on corn and soybean. The web-application includes dynamic data visualization features to enhance communication and information sharing, and is accessible to a broad audience to improve accessibility to on-farm research insights. A random-effects statistical model was used to compute prediction intervals describing a range of plausible yield response for a new (out-

of-sample) trial, and compute the probability that the tested management practice will be ineffective in a new field. Depending on the level of between-trial variability, the prediction intervals were 2.2–12.1 times larger than confidence intervals for the estimated mean yield responses (i.e., at the network level) for all tested management practices. Using prediction intervals and the probability of ineffective treatment will prevent farmers from over-optimistic expectations that a significant effect at the network level will lead with high certainty to a yield gain on their farms. The data-analytic framework was adapted to a French on-farm research network focusing on the efficacy of biocontrol agent products against *Botrytis cinerea*, potassium bicarbonate and *Aureobasidium pullulans*, on organic vine. The results favored potassium bicarbonate as its efficacy on incidence at the network level is higher for diseased intensities between 0% and 10% than for *Aureobasidium pullulans*. For both biocontrol agents, the efficacy on incidence for a new trial is highly uncertain for intensity levels higher than 15%. Finally, this research investigated the impact of experimental plot scale (i.e., small-plot scale and field scale) on the effect of management practice on crop yield and identified the cause of potential discrepancies to inform on-farm decision-making better and adapt the extrapolation of the results. Taken together, this research represents the first major effort in consolidating results from on-farm research network and provides insight to make better farming management decisions.

## CHAPTER 1. GENERAL INTRODUCTION

### 1.1 On-Farm Research: A Cooperation Between the Research

#### Community and Farmers

On-farm research enables conducting experiments under farmers' environments, which plays an essential role in gaining knowledge about production systems (Marchant et al., 2019), and can lead to robust agronomic recommendations. While the concept of on-farm research is not new, it is becoming more widespread due to a growing interest among farmers and the research community, and thanks to the availability of combine-mounted yield monitors that enable farmers to measure crop yield responses (Alesso et al., 2019; Kyveryga, 2019). On-farm research is based on a collaboration between the research community (university researchers, extension services, local agronomists, and agribusiness firms) and farmers. As a result, farmers are considered as active research participants and beneficiaries of this endeavor (Thompson et al., 2019). On-farm research can be used to answer farmers' questions, validate outputs from research stations, determine economic profitability, and explore spatial variability of crop responses within fields (Alesso et al., 2019). On-farm research is also useful to encourage participatory learning to enhance the adoption of sustainable agriculture technologies (Snapp et al., 2019) and quantify genotype x environment interactions (Hernández et al., 2019). Finally, on-farm research is used to test new products and management practices in farmer's fields (Kyveryga et al., 2018) mainly through the concept of an on-farm research network (hereafter called OFRN).

## **1.2 The Different Purposes of On-Farm Research Network**

An OFRN is an organization of farmers who exchange experiences, share their knowledge, and test important questions using common protocols and commercial-scale field equipment (Matthewson et al., 2013). OFRNs can help farmers improve their productivity, efficiency, and profitability (Pruss et al., 2005; Moayedi and Azizi, 2012). Most of the existing OFRNs are led by public institutions such as universities, extension services, or private companies, but in the latter case, access to data and results summaries are limited. The most common experimental design is the “replicated strips” where two treatments, a new management practice (e.g., seeding rate, row spacing, new pest, and disease treatments) and a standard farmer practice, are applied in several strips. On-farm trials are often repeated across locations and growing seasons to capture different pedo-climatic conditions in a given agricultural area.

## **1.3 The Importance of an Analytical Approach**

The analytical approach is a critical component of OFRN to effectively analyze data collected and communicate the results to farmers and the research community. OFRNs can provide at least two levels of information about the performance of a management practice: at the network level and at the individual farm level. The strips in an on-farm trial help capture the within-trial variability and the collection of trials, characterized by different environmental conditions and farmer management practices, capture the between-trial variability. This analytical approach helps summarize many on-farm trials containing similar treatments across locations and years (Moore and Dixon, 2015).

## 1.4 Challenges and Needs

Often, the results are usually presented as individual reports (*i.e.*, a report summarizing the outcome for one trial) (Kyveryga, 2019), but this provides limited information difficult to generalize and does not allow presenting, in a synthetic way, all the results collected from the different trials. Moreover, the comparison of the trial performances can be misleading as the outcomes arise from distinct statistical analyses. There is a crucial need to develop appropriate statistical methods that can simplify multi-trial analyses and estimate the effectiveness of a management practice at the trial and network levels. A new way to disseminate the results from OFRNs is required, which allows for effective and simultaneous summarization, analyses, interpretation and communication of the results. In the end, these statistical analyses and the development of a new visualization tool should help farmers make decisions for the future.

## 1.5 Dissertation and Organization

The overall objective of this research is to demonstrate the importance of identifying appropriate statistical methods for analyzing and visualizing on-farm research network data. Specifically, this research will focus on the analyses of the OFRNs managed by the Iowa Soybean Association, and an adaptation was made with a French case-study. This research represents the first major effort in consolidating results from OFRN. The first chapter focuses on developing of a data-analytics framework that allows analyzing multiple studies and identifying the conditions where an imposed treatment may or may not be effective. It includes a random-effect model that allows estimating and quantifying the

uncertainty of the performance of a specific management practice at the individual location and across all on-farm trials to understand the performance at the network level. The second chapter concerns the development of an interactive web tool that allows to visualize simultaneously all trial results pertaining to the same management practice and simplify the interpretation of multi-site and multi-year summaries. The third chapter looks at the computation of prediction intervals that describe a range of plausible yield response for a new (out-of-sample) field at the trial level, and the probability that the tested management practice will be ineffective in a new field (i.e., absence of positive yield gain resulting from the new management practice). This work provides an operational workflow to fit a random-effects model and compare frequentist and Bayesian prediction intervals and the probability of an ineffective treatment. During this research, I spent an extensive stay at INRA in France to enhance the generic framework by meeting with French stakeholders. As a result, the fourth chapter constitutes an adaptation of the generic framework by using data from a French OFRN focusing on the efficacy of biocontrol agent products on the organic vine. Finally, the sixth chapter investigates the impact of experimental plot scale (i.e., small-plot scale and field scale) on the effect of management practice on crop yield and identify the cause of potential discrepancies. It is crucial to understand the causes of yield gaps to better inform on-farm decision-making and adapt the extrapolation of the results.

## 1.6 References

Alesso, C.A., P.A. Cipriotti, G.A. Bollero, and N.F. Martin. 2019. Experimental Designs and Estimation Methods for On-Farm Research: A Simulation Study of Corn Yields at Field Scale. *Agron.j.* 111(6): 2724–2735. doi: 10.2134/agronj2019.03.0142.

- Hernández, M.V., I. Ortiz-Monasterio, P. Pérez-Rodríguez, O.A. Montesinos-López, A. Montesinos-López, et al. 2019. Modeling Genotype × Environment Interaction Using a Factor Analytic Model of On-Farm Wheat Trials in the Yaqui Valley of Mexico. *Agron.j.* 111(6): 2647–2657. doi: 10.2134/agronj2018.06.0361.
- Kyveryga, P.M. 2019. On-Farm Research: Experimental Approaches, Analytical Frameworks, Case Studies, and Impact. *Agron.j.* 111(6): 2633–2635. doi: 10.2134/agronj2019.11.0001.
- Kyveryga, P.M., T.A. Mueller, and D.S. Mueller. 2018. On-Farm Replicated Strip Trials. In: Kent Shannon, D., Clay, D.E., and Kitchen, N.R., editors, *ASA, CSSA, and SSSA Books. American Society of Agronomy and Soil Science Society of America, Madison, WI, USA.* p. 189–207
- Marchant, B., S. Rudolph, S. Roques, D. Kindred, V. Gillingham, et al. 2019. Establishing the precision and robustness of farmers' crop experiments. *Field Crops Research* 230: 31–45. doi: 10.1016/j.fcr.2018.10.006.
- Matthewson, M., M. Fery, and M. Powell. 2013. *Creating Farmer Networks: A Toolkit for Promoting Vibrant Farm Communities.* <https://www.sare.org/content/download/73146/1058808/file/pnw638.pdf>.
- Moayedi, A.A., and M. Azizi. 2012. Improvement of Knowledge and Skills Level of Wheat-Cultivating Farmers using On-Farm Researches. *Procedia - Social and Behavioral Sciences* 46: 2258–2261. doi: 10.1016/j.sbspro.2012.05.465.
- Moore, K.J., and P.M. Dixon. 2015. Analysis of Combined Experiments Revisited. *Agronomy Journal* 107(2): 763. doi: 10.2134/agronj13.0485.
- Pruss, J., D. Beegle, A. Turgeon, R. Day, and R. Weaver. 2005. Using On-Farm Data to Validate Crop Management Recommendations and Implementation. *Agronomy Journal* 97: 113–117.
- Snapp, S.S., J. DeDecker, and A.S. Davis. 2019. Farmer Participatory Research Advances Sustainable Agriculture: Lessons from Michigan and Malawi. *Agron.j.* 111(6): 2681–2691. doi: 10.2134/agronj2018.12.0769.
- Thompson, L.J., K.L. Glewen, R.W. Elmore, J. Rees, S. Pokal, et al. 2019. Farmers as Researchers: In-depth Interviews to Discern Participant Motivation and Impact. *Agron.j.* 111(6): 2670–2680. doi: 10.2134/agronj2018.09.0626.

## **CHAPTER 2. A FRAMEWORK FOR VISUALIZATION AND ANALYSIS OF AGRONOMIC FIELD TRIALS FROM ON-FARM RESEARCH NETWORKS**

Modified from a paper published in *Agronomy Journal*, special issue “On-farm research”

Anabelle Laurent<sup>1</sup>, Peter Kyveryga<sup>2</sup>, David Makowski<sup>3</sup>, Fernando Miguez<sup>1</sup>

<sup>1</sup> Department of Agronomy, Iowa State University, Ames, IA

<sup>2</sup> Analytics, Iowa Soybean Association, Ankeny, IA

<sup>3</sup> INRAE, Centre Île-de-France, Paris, France

### **2.1 Abstract**

An on-farm research network is an organization of farmers that conducts agronomic experiments under local conditions. It is common that an elementary statistical analysis be conducted for the individual studies. However, there is an unexplored potential in detecting yield response variability patterns for better decision making. We developed a data-analytics framework and web-application program that allows users to analyze multiple studies that use a common protocol and can identify the conditions where an imposed treatment may or may not be effective. The development of this data-analytics framework is needed to improve the predictions at the farm level that can lead to more cost-effective, sustainable and environmentally sound agricultural production. Data visualization is an important part of data-analytics. In this paper, we have developed and tested a Bayesian hierarchical model that can be used to assess the general agronomic performance of different management practices. Decision making related to new management practices should be based on the complete evidence, local conditions, and economic considerations. The web application includes dynamic data visualization features to enhance communication and sharing of information with the goal of reaching a broader audience.



## **2.2 An Introduction to On-Farm Research Network**

A farmer network is an organization of farmers who exchange experiences, share their knowledge, and test important questions using common protocols and commercially available field equipment (Matthewson et al., 2013). There is increasing interest in On-Farm Research Networks (OFRNs) because they provide the infrastructure needed to test new products and management practices in farmers' fields (Kyveryga et al., 2018). In addition, data from these experiments can be used to validate simulation models and determine the economic profitability of new technologies. Within this infrastructure, the most common design is to compare a new management practice (e.g. seeding rate, row spacing, new pest and disease treatments etc.) to a standard farmer practice. This new generation of OFRNs can help farmers improve their productivity, efficiency and profitability (Pruss et al., 2005; Moayedi and Azizi, 2012) and, create a novel communication platform between farmers, agronomists, and scientists.

## **2.3 Developing Research Networks**

Farmer networks can arise from diverse motivations, and they can start by recruiting cooperating farmers and defining the network's missions. Once a group of farmers is identified, the next important step is to define a problem and research question (Kyveryga et al., 2018). The question should be simple enough to be approached through standard experimental designs and executed using farmers' available equipment. The collaboration among farmers, researchers, local agronomists, or crop consultants makes the implementation of common experiments and protocols possible by defining the number of treatments, the variables to be measured (e.g. crop yield, grain moisture, and

protein content), and the experimental design. Usually, scientists or research agronomists assist farmers with data collection, data analysis, interpretation, and the communication of results to the general public.

### **2.3.1 Analyzing Data Across Experiments**

Increasingly, scientists and farmers are using on-farm testing as an approach to build locally adapted recommendations. However, the scientific community is challenged by combining results from studies conducted on different soils and climatic conditions. Integrating yield and climatic data has the potential to improve recommendations. For example, Kyveryga et al. (2013) combined on-farm, weather and soil data to analyze the risk of yield losses resulting from a reduction of farmer normal N fertilizer rates applied to corn (*Zea mays* L.). Bissonnette et al. (2018) used on-farm data from 18 strip-trial experiments located in the northern half of Iowa over three years to study the effect of nematicide seed treatment, Clariva Complete Beans (CCB) (*Pasteuria nishizawae*, sedaxane, thiamethoxam, fludioxonil and mefenoxam as active ingredients) compared with CruiserMaxx Advanced plus Vibrance (CMV) (thiamethoxam, mefenoxam, fludioxonil and sedaxane as active ingredients), on soybean cyst nematode (*Heterodera glycines*) reproduction and soybean (*Glycine max* (L.) Merr.) yield. They found that CCB seed treatment had a variable effect on soybean cyst nematode reproduction and soybean yield. Kyveryga et al. (2013) analyzed data from 282 on-farm strip-trial experiments across Iowa over five years of experimentation to identify when a foliar application of pyraclostrobin fungicide produced profitable soybean yield responses. They found that greater yield

responses were observed for trials that received more than 30.5 cm of cumulative March through May rainfall.

To our knowledge, most of the existing OFRNs are based in the United States and led by public institutions such as universities or extension services (Table 1). Private companies also manage their own OFRNs, but access to data and results summaries are limited. Some OFRNs have been implemented for decades, such as the Practical Farmers of Iowa led by Cooperators' Program since 1987, and the Nebraska On Farm Research Network led by the University of Nebraska since 1990. Most of them have similarities regarding the crops of interest, the experimental design, and the topics of research. Current large-scale equipment makes some experimental designs such as replicated strip trial design with two treatments (the new management practice and the control) more practical than others. The implementation of this experimental design was made easier in recent years due to the wide adoption of precision agriculture technologies that enable farmers to measure yield with mass flow sensors and GPS technology, which generally produced similar results as weigh wagons (Nelson et al., 2015). The management practices tested typically involve crop management (e.g. planting date, seeding rate, tillage, row spacing), crop protection (e.g. pesticide, GM resistant cultivars), plant nutrition (e.g. fertilizer, manure, lime) and plant growth regulators (e.g. auxin, gibberellic acid, cytokinin). In some cases, OFRNs are crop-specific (e.g., Minnesota Wheat's On-Farm Research Network and On-Farm Soybean Management Network) or management practice-specific (e.g., the Indiana Infield Advantage focuses on nutrients in corn).

Results of on-farm trials are usually presented as individual field reports (i.e. a report summarizing the outcome for one trial) showing replicate yield values and

treatment averages in the form of tables or histograms. Some other basic information (e.g. planting date, variety, soil texture, weather data, location) are also typically provided. In an effort to develop more practical communication methods, some OFRNs such as Minnesota Wheat's On-Farm Research Network, Nebraska On-Farm Research Network and Pennsylvania On-Farm Soybean Network have compiled all trial reports into an annual report format. An example from Dupont Pioneer (Jeschke and Ahlers, 2018) studied the effect of foliar fungicides (alone or combined with an insecticide) on soybeans across 279 on-farm trials and shared the trials' average yield differences through a histogram and ranking of trials by decreasing yield response values. Despite the number of trials involved, only the average yield response per trial was reported and without explanations of variability in yield response. In another example, the South Dakota On-farm Research program allows for sorting experiments into different categories. The Nebraska On-Farm Research Network and the Iowa Soybean Association On-Farm Network® have online searchable databases that allow users to query individual summary trial data by year, crop and, management practice, but this is not sufficient to understand general patterns in treatment effects and gain novel insights from the data.

Currently, for most OFRNs, individual trial summaries provide descriptive information and elementary statistical analysis. Even though this information is highly valuable, it does not directly lead to a better understanding of the overall agronomic performance of the treatment or product. Also, they do not allow for the detection of patterns that can explain the yield response variability for different soil textures, rainfall amounts, planting dates, or seed varieties. Finally, individual trial summaries cannot provide an estimate of the probability that a new management practice will or will not

outperform standard practices in following growing seasons or in new environments. To overcome these limitations, a new framework for the analysis of OFRN data is needed which is not simply limited to a multi-location analysis (Moore and Dixon, 2015), although it should contain common elements found in mixed-effect models and meta-analyses (Pinheiro and Bates, 2000; Philibert et al., 2012).

The evolution and recent expansion of OFRNs (Table 1) present a unique opportunity to fill this gap by developing a data-analytics framework and an easy-to-use tool for decision making which would allow effective and simultaneous summarization, analyses, interpretation, and communication of the results. The development of such a data-analytics framework is necessary to improve predictions at the farm level that can ultimately lead to more cost-effective, sustainable, and environmentally sound agricultural production. Data visualization is an important element of the data-analytics framework, useful for identifying trends and clusters, spotting patterns, evaluating model outputs, and communicating results (Unwin et al., 2006). Visualization tools are needed to allow farmers and agronomists to detect patterns across sites and years. Data visualization has the potential to revolutionize sharing and communication of analysis (Wojciechowski et al., 2015) and is more convenient and informative than individual summaries. So far, this approach has not been used in the context of OFRN.

## 2.4 Creating a Data-Analytics Framework

The main goals of our data-analytics framework called Interactive Summaries of On Farm Strip Trials (ISOFAST) are: i) assess the general agronomic performance of different practices, ii) explain yield response variability using field-level covariates, and iii) use interactive and dynamic visualization to enhance communication and decision making by farmers. The utility of this framework is illustrated using three case studies testing specific agronomic questions about a foliar fungicide on soybean, row spacing on soybean and a soil-applied insecticide on corn.

Table 1: Examples of on-farm research networks.

<b>Name</b>	<b>Managing organization</b>	<b>Experimental design</b>	<b>Starting date</b>	<b>Crops</b>
<b>On-Farm Network®</b>	Iowa Soybean Association	RST	2005	soybean; corn
<b>Pennsylvania On-Farm Soybean Network</b>	Pennsylvania Soybean Board.	RST	2009	soybean
<b>Nebraska On-Farm Research Network</b>	University of Nebraska	RST RCBD	1990	soybean; corn wheat; pea sorghum; beans wheat
<b>Minnesota Wheat's On-Farm Research Network</b>	Minnesota Association of Wheat Growers.	RST	2014	wheat
<b>Practical Farmers of Iowa</b>	Cooperators' Program	RST RRST	1987	corn; soybean; oat winter rye; horticulture soybean; corn
<b>Purdue Collaborative On-Farm Research</b>	Indiana Certified Crop Advisers (CCAs) and Purdue Extension	RRST	2006	
<b>South Dakota Soybean On-farm Research Program</b>	South Dakota Soybean Research and Promotion Council	RST	2014	soybean
<b>California Collaborative Research and Extension network</b>	University of California Santa Cruz	Split-plot	2014	vegetables; strawberry
<b>DuPont Pioneer Soybean Fungicide Research</b>	Dupont Pioneer	NA	2007 to 2014	soybeans

Our framework is implemented through a web application accessible to a broad audience to improve accessibility to on-farm research insights.

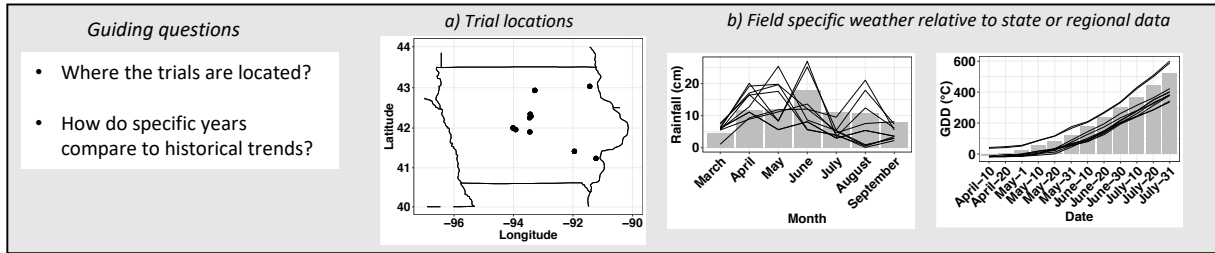
#### **2.4.1 Preliminary Analysis**

The data-analytics framework starts by providing a brief summarization of background information and rationale for testing a new management practice under on-farm conditions. Specific agronomic objectives, details about management practices, product chemistry, application rates, timing of applications and number of locations are also included. Our data-analytics framework provides a map which displays the trial locations and general attributes (Figure 1.a).

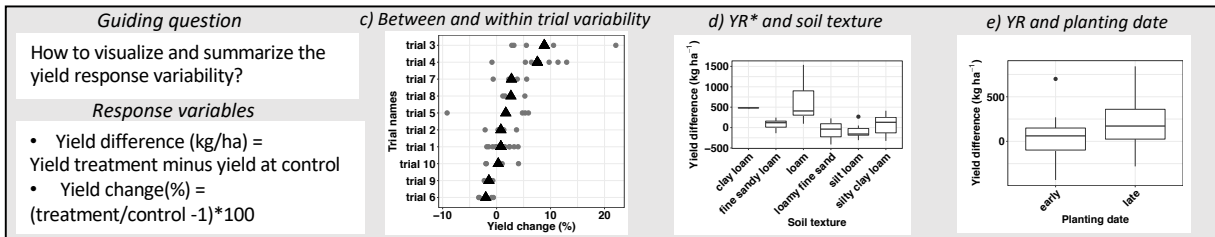
Because precipitation and temperature are important to understanding yield responses, the data-analytics framework allows for the simultaneous display of in-season monthly rainfall and growing degree day observation for each trial. Growing degree day (GDD) is a common temperature index used to estimate plant development, and accumulation of GDD values determines the maturity of the crop, yield, and yield components (Qadir et al., 2007). Reference rainfall (average over the duration of all the trials) is included which help to identify wet, dry and average seasons (Figure 1.b, on the left). The same visualization is given for the cumulative GDD over the growing season with reference values (Figure 1.b on the right).

## Framework for visualizing and analyzing on-farm research network data

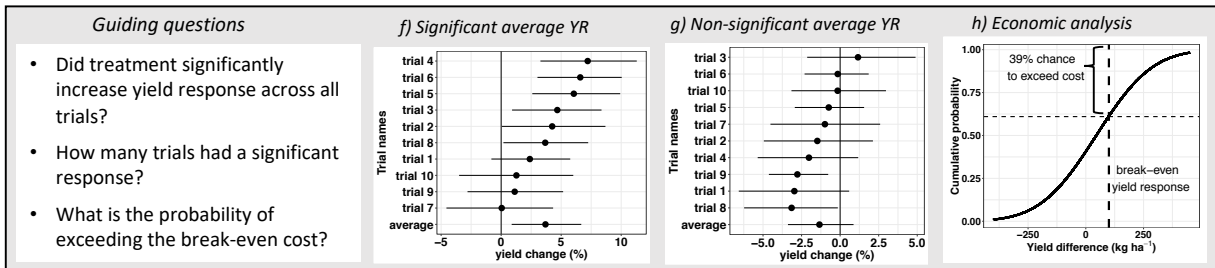
### Step 1: Location and description of environmental conditions



### Step 2: Descriptive analysis of the yield response



### Step 3: Statistical modeling: treatment effect within and across all trials



### Step 4: Statistical modeling: explain the yield response variability

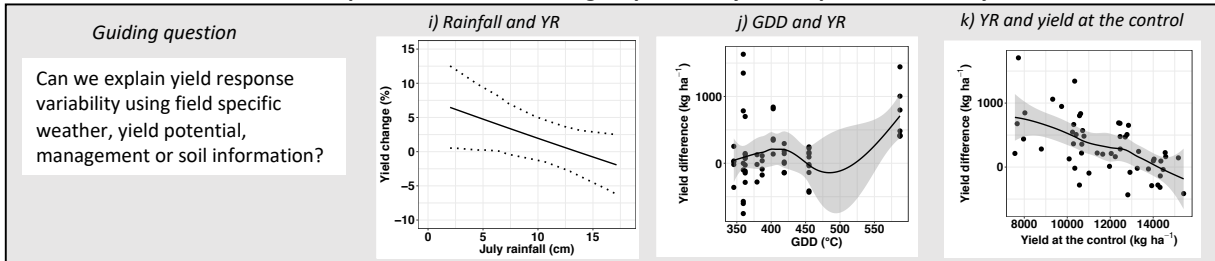


Figure 1: Example of visualization of trial results from on-farm research network. a) dots represent trial locations; b) a trial is represented by a black line, and the reference values (using historical data or climatology) are represented by grey bars; c) dots represent individual replicates or experimental units and the black triangle represents the trial mean; f and g) estimated mean yield response (YR) and the associated 95% credible intervals for the individual trials and the overall yield response at the bottom (posterior means from Bayesian analysis); h) the dashed line represents the break-even yield response, the vertical bar represents the probability to exceed breakeven cost, and the curve represents the cumulative distribution of yield response; i) the black line represents the estimated yield response and the dashed lines represent the 95% credible intervals (from Bayesian analysis); j and k) the dots represent trial means; the black line, the estimated yield



response; and the grey shade represents the 95% confidence intervals from the local regression.

### **2.4.2 Defining Yield Responses**

The main objective of our framework is to quantify the effect of a new management practice on yield compared to a control (i.e. a control corresponding to a common cropping practice or a product normally used by a farmer). Two different metrics are proposed to measure yield response; the yield ratio (a ratio of yield obtained with new management practice to yield at the control) and the yield difference (yield obtained with the new management practice minus yield at the control).

The yield difference measures the effect of the new management practice in absolute yield units. It can be easily expressed as the economic gain or loss (in \$ per ha), but it is unit-dependent. The yield ratio measures the effect of the management practice relative to the yield obtained at the control. It is unitless, and thus, it does not depend on yield units or a moisture content adjustment, or other similar factors. Its value can be used in different contexts characterized by different productivity levels. It is thus possible to multiply the estimated yield ratio by low to high reference yield values to obtain the range of yield gains or losses. A yield ratio higher than 1, or a yield difference higher than 0, means a yield gain using the new product or management practice. A yield ratio lower than 1, or a yield difference lower than 0, means a yield loss using the new management practice. The main consideration for favoring one metric over the other is whether the management practice scales with yield (yield change) or if it is invariant to yield levels (absolute yield difference). For this reason, our data-analytics framework provides both metrics.

### 2.4.3 Importance of Replication

It is very common in agronomic experiments to have replicates (i.e. multiple measurements for a single variable or multiple experimental units) in order to reduce variability and increase the statistical power of experiments. Also, replicates are important to quantify variability within each experiment (i.e. within a trial) and between experiments (i.e. between trials). Sometimes, an observation can be judged far from its group average and thus be considered as an outlier (Ramsey and Schafer, 2013). Outliers may be due to natural variation, equipment problems, human error, or can be caused by hail, flooding or extreme heat. Graphics display all replicate values and describe yield response variability between and within trials (Figure 1.c). Additionally, trials are ranked by increasing mean yield responses. Displaying the means helps to summarize data and identify replicates that deviate from the overall mean or general trend. Ranking trials by decreasing average yield response provides a first impression of the effectiveness of the treatment. Our framework does not remove outliers if they can be explained by natural, physiological or agronomic mechanisms. We consider all outliers because they often show important source of yield variability.

Yield variability can also be explained by environmental and management variables. Since trials are generally located across the state and farmers apply their own management preferences, some characteristics such as soil texture, seed variety, and crop planting dates can vary substantially. In our framework, yield responses are also presented for different soil texture and planting date categories (early and late planting date) using a boxplot (Figure 1.d and e). The planting date threshold corresponds to the midpoint between the earliest and latest planting dates related to a specific management practice.

#### **2.4.4 Yield Limiting Factors and Yield Response**

Yield limiting factors (e.g. weather stress, pest pressure, soil characteristics) can influence crop yields directly or by interacting with each other. When crop damage by pests is not observed, then yield at the control can be used as a proxy of yield limiting stress factors. If yield is low in the control strips, this might indicate that a limiting factor or pest pressure has prevented the crop from reaching its potential. A consistent negative relationship between yield response and yield in the control strips would suggest that the product or practice studied has directly addressed a yield-limiting factor (Figure 1.k). For example, Salvagiotti et al. (2008) used yield measurements from fertilized plots (N application) and unfertilized control plots and demonstrated that yield response to N fertilization was positive when the yield potential was low. The reason behind the lower yield potential was different for each specific site-year, such as low soil pH or fertility or water limitations. Another way to assess pest pressure or other major limiting factors in on-farm trials is to use crop scouting data. This requires rigorous knowledge of pest and crop biology, pest identification, and sampling methods. Consequently, at the moment, our framework does not uniformly provide a specific analysis and visualization of scouting data. After a descriptive step to visualize and describe yield response variability, statistical modeling can be used to help to explain the heterogeneity, improve the understanding of the data and quantify the uncertainty of the treatment effects.

#### **2.4.5 Statistical Modeling**

Appropriate statistical analyses should focus on different but related questions: What is the performance of a specific treatment in an individual trial or location? And what

is performance across all trials or the overall mean yield response? Answering these questions will be beneficial i) toward understanding the effectiveness of management practices at the network level, ii) clarify the specific questions that farmers have about their own farm and iii) help make future management decisions.

Since data are collected for several individual trials, we used a hierarchical model to estimate the mean effect size at the network level, the individual effect sizes for all trials, and their credible intervals through a Bayesian approach. The network level represents the whole group of on-farm trials testing the same management practice. The Bayesian analysis has an advantage over classical statistical analyses because it can use prior information derived from literature or expert knowledge. The Bayesian approach integrates the observed data with priors and returns a posterior distribution of the parameters of interest. Another advantage of the Bayesian approach is that it allows incorporation of full uncertainty in all parameters. The uncertainty in parameter estimates is quantified by using credible intervals.

The hierarchical model uses yield ratios or yield differences as the response variable. The yield ratio generally benefits from log transformation for normality and stabilization of variances. The results are then back-transformed for interpretation as percent change (i.e. yield change expressed in % = (yield ratio-1) \* 100). Trials are represented by site-years as they are rarely repeated at the same location over time. The Bayesian hierarchical model was implemented using the R package, MCMCglmm, through R Studio (Hadfield, 2010; RStudio Team, 2015).

For a continuous explanatory variable, the statistical model is:

$$\log(R_{ij}) = \mu + \beta X_{ij} + \alpha_i + \varepsilon_{ij}$$

Where  $\log (R_{ij})$  represents the natural log of the  $j^{\text{th}}$  yield ratio (or yield difference without log transformation) in the site-year  $i^{\text{th}}$ ;  $\mu$  represents the intercept of the log transformed ratio;  $\beta$  represents the regression parameter (equal to zero if there is no continuous variable such as rainfall);  $\alpha_i$  represents the random effect of the site-year  $a$  and  $\varepsilon_{ij}$  represents the residual error.  $\alpha_i$  and  $\varepsilon_{ij}$  are assumed to follow a normal distribution with mean zero and some variance.

$$\alpha_i \sim N(0, \sigma_\alpha^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

We defined the priors for the three parameters of the model (i.e.  $\mu$ ,  $\sigma_\alpha^2$  and  $\sigma_\varepsilon^2$ ). The prior for the intercept  $\mu$  represents the distribution of the mean of the log ratio (or the yield difference) and it follows a Gaussian distribution with a mean of zero and a variance of 2. With a variance of 2, the log-ratio can take either a high positive or a low negative value depending on the dataset.

The priors of the variances of the random effect,  $\sigma_\alpha^2$ , and of the residual error,  $\sigma_\varepsilon^2$ , are independent inverse Gamma distribution with parameter  $(\nu/2, \nu/2)$  where the degree of belief (i.e.  $\nu$ ) is equal to 0.002. The parametrization of the priors is specific to the R package *MCMCglmm* (Hadfield, 2010).

For a categorical variable (as the soil texture), the statistical model is:

$$\log (R_{ij}) = \mu + \sum_{k=1}^K \beta_k X_{ij}^{(k)} + \alpha_i + \varepsilon_{ij}$$

Where  $X_{ij}^{(k)}$  equals to 1 (an indicator variable) if  $R_{ij}$  belongs to the  $k^{\text{th}}$  category, zero otherwise and  $\beta_k$  represents parameter for the  $k^{\text{th}}$  category.

We used the same visual approach as in meta-analysis (*i.e.*, forest plot) to show estimated posterior yield responses from individual trials. The forest plots show variation between and within trials, as well as overall posterior means (Figure 1.f-g). Individual trial posterior means are statistically significant if their credible intervals do not cross the vertical line (*i.e.* yield change or yield difference equal to zero) corresponding to a threshold between yield increase and yield loss from a new management practice or treatment in question. Trials are ranked in increasing order to easily distinguish potential groups of trials with similar positive or negative yield response. Different credible interval levels (*i.e.* 0.80, 0.90 or 0.95) are available to satisfy farmers' and scientists' expectations and risk preferences.

Cumulative probabilities of yield response at the regional level can be calculated from the posterior distributions of yield response or yield change provided by the Bayesian model. The cumulative distribution function represents the probability that the yield response is less than or equal to a certain value (Figure 1.h). For example, if the probability of having a 4% yield increase is equal to 70% it means that there is a 70% chance of reaching a 4% yield increase or less. Cumulative distribution of yield response can be useful for decision making for farmers.

Our data-analytics framework provides two different ways to attribute yield response variability using explanatory variables. The first approach is for continuous or categorical variables in the Bayesian hierarchical model (see equations above) (Figure 1.i) and the second approach is to use a local polynomial regression (Figure 1.j) (Cleveland, 1979). For each method, 95% credible intervals or 95% confidence intervals, respectively, are displayed to describe the uncertainty in yield response.

### **2.4.6 Calculating Economic Responses**

Economic analysis is important to decide if a new practice or product should be adopted. Cumulative distribution functions of yield response are used to conduct a break-even economic analysis (Figure 1.h). The on-line tool, allows users to enter grain price and treatment cost (i.e. cost of product and application). Based on this information, a break-even yield response and expected profit are calculated. The break-even line is plotted on the graph and the probability of exceeding the break-even cost is estimated as the distance on the Y-axis between 1 and the intersection of the cumulative distribution curve with the break-even line. The range in expected average profit is calculated using 25th and 75th percentiles from the cumulative distribution function.

### **2.4.7 Visualization**

The graphical features are implemented through an interactive and dynamic graphical web-tool. We used Shiny, an R package from RStudio (RStudio Team, 2015; Chang et al., 2016) that combines numerous extension packages. The web-application has a user interface divided in two parts: the sidebar menu on the left and the main panel on the right (Figure 2). On the sidebar menu, the user can select the crop and then a specific management practice. Then, the user has access to different components organized into a list on the sidebar menu.

The main panel, located on the right side of the interface, returns visuals described in Figure 1. The main panel has interactive features such as zoom-in, zoom-out, filter, select and pointer-hovering in order to interact with data and graphical information. Zoom-in and zoom-out are interactive features available for all the visual graphics but are most useful

for the trial locations maps if users want more precision regarding trial location. The tool blocks identifying exact trial location due to data privacy issues. Because of the large amount of data for some management practices, it can be inconvenient to observe summaries of all data at once.



Figure 2: The interface of the web application, on the left side bar is a menu with different studies and description of the components available for that study. On the right is the main panel, displaying components which have been selected for review and interaction.

To overcome this visualization issue, data can be filtered by year to allow users to focus on a specific data subset. Selection is another important visualization process in our web-application. When a graphic represents the yield response, users can choose between the yield change and the yield difference. For the relationship between yield response and monthly rainfall, users can select a specific month or cumulative months by start, middle or end of crop season. By hovering the pointer over the dot on a visual graphic, a label reports



extra information such as the exact numerical value of the dot. For example, for the visual graphic representing the overall and trial yield responses (Figure 1.f), a label reports the exact numerical value of the point estimate and the boundaries of the credible interval for yield difference or yield change.

The web-application also provides interactive boxes located below each graphic to report extra information, results of statistical analyses and key messages. Some boxes are updated in real time after users' action. For example, the number of trials that had a significant positive yield response is updated after the selection of a significant level for the credible interval. The web-tool is comprehensive and intuitive enough to be easily used by a broader audience that will include farmers and non-specialists. More detail about the structure of our web-application and how to use it are available in the supplementary material.

Our data-analytics framework was implemented for a total of 19 different management practices tested by the Iowa Soybean Association. The data related to the different management practices are stored in different datasets (one dataset per management practice) and differed by number of trials, yield value and years of experiments (Table 2).

The following section provides examples of the implementation of the data-analytics framework for three case studies: foliar fungicide on soybean, row spacing on soybean and soil-applied insecticide on corn.

Table 2: Description of different on-farm trial categories included in our framework.

Treated	Control	Crop	Number of trials	Number of experimental units per treatment	Number of years
Nematicide Clariva Complete Beans® ( <i>Pasteuria nishizawae</i> )	Cruiser Maxx Advanced + Vibrance ( <i>thiamethoxam</i> , <i>mefenoxam</i> , <i>fludioxonil</i> and <i>sedaxane</i> )	soybean	32	223	3
Seed treatment Ileva® ( <i>fluorpyram</i> ) + Acceleron®	Acceleron® ( <i>pyraclostrobin</i> and <i>imidacloprid</i> )	soybean	26	353	2
Row spacing 15 inches	30 inches	soybean	18	120	4
Foliar fungicide Headline® ( <i>pyraclostrobin</i> )	untreated	soybean	206	1088	9
Foliar fungicide Stratego® ( <i>prothioconazole</i> and <i>trifloxystrobin</i> )	untreated	soybean	29	328	5
Foliar fungicide Stratego YLD® ( <i>prothioconazole</i> and <i>trifloxystrobin</i> )	untreated	soybean	37	200	3
Foliar fungicide Priaxor® ( <i>pyraclostrobin</i> and <i>fluxapyroxad</i> )	untreated	soybean	43	191	6
Foliar fungicide Priaxor and Fastac® ( <i>alpha-cypermethrin</i> )	untreated	soybean	22	97	5
Foliar fungicide Quadris® ( <i>azoxystrobin</i> )	untreated	soybean	18	93	1
Hero® pyrethroid insecticide ( <i>bifenthrin</i> and <i>zeta-cypermethrin</i> )	untreated	soybean	7	43	1
Inoculant Terramax®	untreated	soybean	15	99	1
Biostimulant Vitazyme® (1- <i>triacentanol</i> and <i>brassinosteroids</i> )	untreated	soybean	10	44	2
Biological co-product Tryptophan®	untreated	soybean	16	89	2
Seed treatment Nemastrike® ( <i>tioxazafen</i> ) + Acceleron®	Acceleron® ( <i>pyraclostrobin</i> and <i>imidacloprid</i> )	soybean	6	34	1
High density seeding (normal rate + 30k)	normal rate between 140K and 170K (rate commonly used in Iowa)	soybean	20	12	4

Table 2: continued.

Treated	Control	Crop	Number of trials	Number of experimental units per treatment	Number of years
Low yield density (normal rate – 30k)	normal rate between 140K and 170K (rate commonly used in Iowa)	soybean	21	140	4
Winter rye cover crop	untreated	soybean	32	166	6
Oats cover crop	untreated	soybean	12	50	2
Sulfur SuperCal SO <sub>4</sub>	untreated	soybean	15	75	1
Residual sulfur SuperCal SO <sub>4</sub>	untreated	soybean	6	27	1
Soil-applied insecticide Aztec® ( <i>tebupirimphos</i> and <i>cyfluthrin</i> )	untreated	corn	36	195	8
Fertilizer anhydrous ammonia	UAN	corn	26	127	4
Fall-applied anhydrous ammonia	Spring-applied anhydrous ammonia	corn	66	360	6
Nitrification Inhibitor (Instinct®) on manure	untreated	corn	29	115	4
Nitrification Inhibitor (Instinct®) on UAN	untreated	corn	19	96	3
Foliar fungicide Headline® ( <i>pyraclostrobin</i> )	untreated	corn	143	703	9
Foliar fungicide Stratego® ( <i>propiconazole</i> and <i>trifloxystrobin</i> )	untreated	corn	32	153	2
Foliar fungicide Stratego YLD® ( <i>propiconazole</i> and <i>trifloxystrobin</i> )	untreated	corn	82	444	6
Foliar fungicide Quilt® ( <i>azoxystrobin</i> and <i>propiconazole</i> )	untreated	corn	28	144	3
Biological co-product Tryptophan®	untreated	corn	14	68	2
Seed treatment NemaStrike® ( <i>tioxazafen</i> ) + Acceleron®	Acceleron® ( <i>pyraclostrobin</i> and <i>imidacloprid</i> )	corn	8	53	1
Mycorrhizal fungi Endoprime®	untreated	corn	17	148	2
Winter rye cover crop	untreated	corn	42	241	7
Oats cover crop	untreated	corn	19	107	3
Sulfur SuperCal SO <sub>4</sub>	untreated	corn	48	214	4
Residual sulfur SuperCal SO <sub>4</sub>	untreated	corn	16	77	2

## 2.5 Case Studies

### 2.5.1 Foliar Fungicide Impact on Soybean Yields

Hypothesis: foliar fungicides increase soybean yields

#### *Background*

Foliar fungicides help to manage several common foliar diseases in soybean such as anthracnose, Septoria brown spot, Cercospora leaf blight, frogeye leaf spot, pod and stem blight, and soybean rust. Foliar pathogens reduce green leaf area causing a reduction of photosynthetic activity which can affect crop growth and yield (Bassanezi et al., 2001). In Iowa, foliar diseases typically result in minor yield losses which explains why applying foliar fungicide has not been a common practice (Swoboda and Pedersen, 2009; Wrather and Koenning, 2006). However, the use of foliar fungicide has increased since 2004, especially during periods of high market grain prices. As a consequence, better information was needed about fungicides in managing Septoria brown spot or frogeye leaf spot (Kyveryga et al., 2013). The objective of these trials was to study the effect of a foliar fungicide (Headline®) on soybean compared to a control by quantifying the yield response across a wide range of environmental and management practices.

#### *Materials and Methods*

The foliar fungicide Headline® was tested in 206 trials (Figure 3 on the left) over nine years (2006-2013, 2015) and compared to a control (untreated). These experiments used yield data collected on a combine equipped with GPS. The active ingredient of Headline® is pyraclostrobin. Most of the applications were done by farmers using ground

sprayers, but in about 20% trials the foliar fungicide was applied by airplanes. The time of application varied between trials but most of them were done at the crop stage R3 (beginning pod development).

The experimental design was the replicated strip, where the two treatments are applied in strips without randomization (Figure 3 on the right). A pair of strips (foliar fungicide and the control) constitutes a replicate. Each trial which was part of the network had a minimum of three replicates. Some trials required more than three replicates to capture the entire field for spatial analysis of yield responses. The width of the individual strip depends on the size of application equipment and can range from 4.6 to 27.4 m. The length of the strips depends on the field size. For example, a typical field in Central Iowa is about 32 ha, which would have a dimension of about 457 x 701 m.

Trials were well distributed throughout Iowa, with a majority located in the Des Moines Lobe (Figure 3, on the left). Therefore, our data covers a broad set of environmental conditions and field management across Iowa.



Figure 3: Schematic illustration of an on-farm research network where foliar fungicide applications on soybean were compared to an untreated control. Each trial is represented by a dot (described on the left) on the map where the year of measurement is distinguished by color. All the trials follow a replicated strip trial design (described on the right) and have at least 3 replicates.

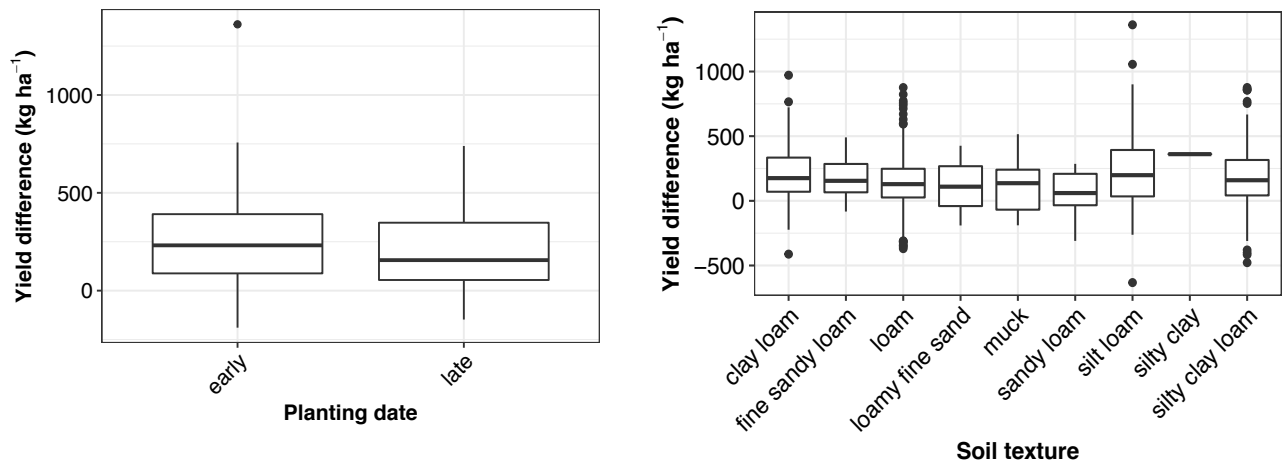


Figure 4: Yield difference by planting date (on the left) and soil texture (on the right) for the foliar fungicide on soybean dataset. A planting date before May 20th is considered as early; after as late.

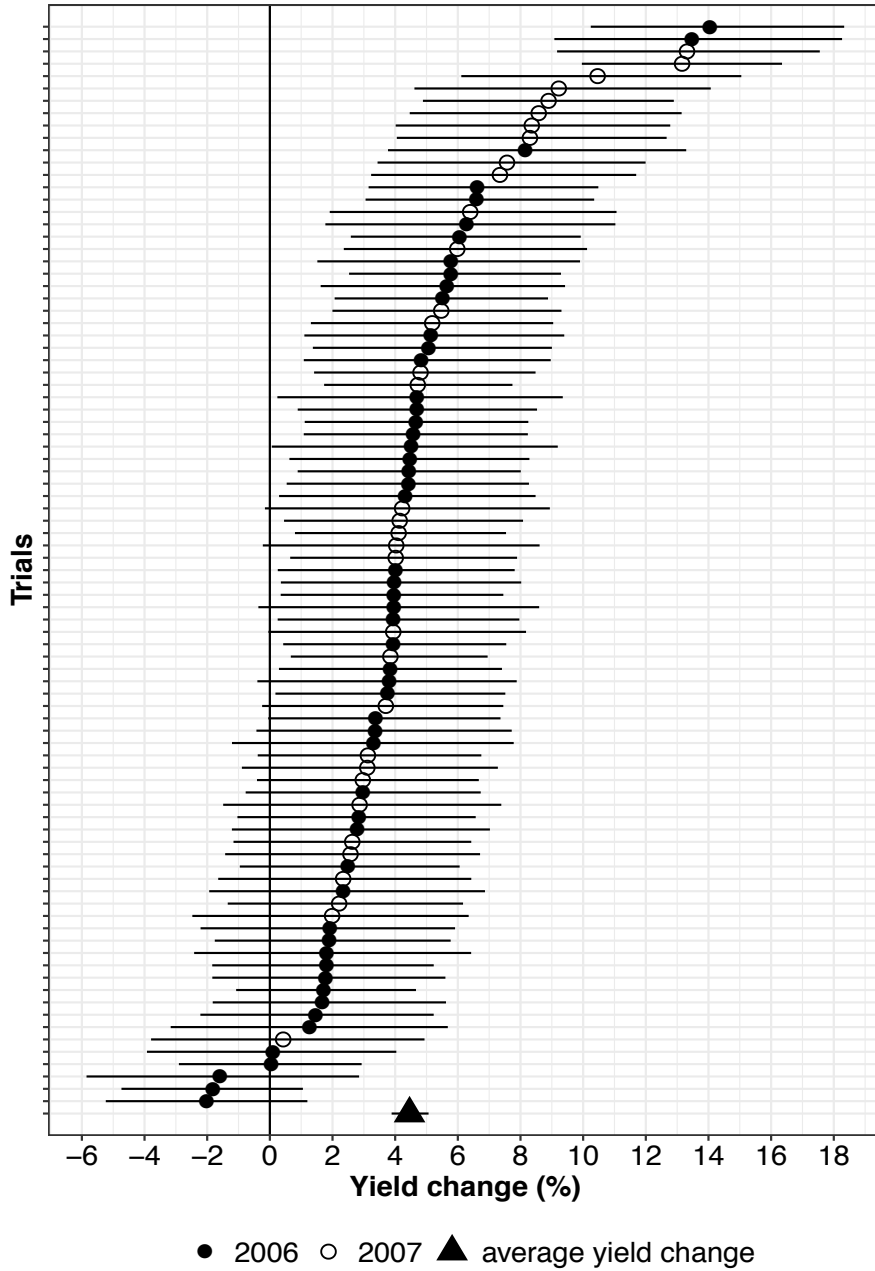


Figure 5: Estimated posterior mean yield change for individual trials and their 95% credible intervals for data collected in 2006-2007 for the foliar fungicide on soybean dataset. The average yield change is estimated using the whole data set. For simplicity, only two years of measurements are displayed in this figure.

### *Results and Discussion*

Disease development in soybean fields can be affected by different environmental conditions or management practices. The tool provides the ability to display the effect of planting date and soil texture on yield response to fungicide. In this case, yield difference was not significantly affected by planting date or soil texture (Figure 4). However, the mean yield difference was higher for the early planting date than for the late planting date (246 kg ha<sup>-1</sup> (se=28) and 195 kg ha<sup>-1</sup> (se=24), respectively).

The average yield change was statistically significant and equal to 4.5 % [3.9; 5.1] indicating a 95% probability that the posterior yield response would fall in a range from 3.9 to 5.1 % of yield increase (Figure 5). Considering all years, 54% of the trials (112 trials of 206) had a significant positive yield response to the foliar fungicide Headline®. These results confirm that this management practice provided consistent yield benefits under the evaluated conditions.

Our results are in general agreement with previous studies looking at the difference between Headline® and an untreated control. For example, results from small plot research trials over 5 years managed by Dupont Pioneer (Jeschke and Ahlers, 2018) showed an average yield response of 249 kg ha<sup>-1</sup> when Headline was applied at the R3 growth stage and a total of 78% of the trials presented a positive yield response. Bestor et al. (2014) had similar results and reported that Headline had a higher yield (276 kg ha<sup>-1</sup>) than the untreated control. The average yield difference, based on 7 locations, was equal to 276 kg ha<sup>-1</sup>. Wise and Buechley (2010) and Mahoney et al. (2015) reported a yield for Headline® and untreated control of 202 kg ha<sup>-1</sup> and 180 kg ha<sup>-1</sup>, respectively.



### **2.5.2 Row Spacing Impact on Soybean Yields**

Hypothesis 1: Narrow row spacing produces higher yields than wide row spacing on soybean.

Hypothesis 2: Wet conditions increase diseases in narrow row spacing on soybean.

#### *Background*

A common soybean row width spacing is equal to 76 cm, however, many farmers have been testing whether yield will increase by planting narrower rows (De Bruin and Pedersen, 2008). Soybean often yields higher when planted in narrow versus wide row spacing. For example, De Bruin and Pedersen (2008) advocate the adoption of 38-cm row spacing based on a 5.6% yield increase in a 38-cm vs. 76-cm in a 3-year study at five locations in Iowa. Iowa State University Extension and Outreach (n.d.) showed a 309 kg/ha advantage of 38-cm over 76-cm in a two-year study at 17 locations. However, many farmers are still hesitant to switch to narrow row spacing due to the required investment in new planters and the higher risk of soybean diseases in narrow rows. In fact, narrow spacing increases the canopy area development, light interception, growth rate, dry matter accumulation, and seed yield but also results in higher soil moisture or relative humidity which may create favorable conditions for the development of white mold (*Sclerotinia stem rot*). The objectives were i) to study the impact of narrow row spacing compared to wide row spacing by quantifying the yield response and ii) to study the effect of rainfall amounts on yield differences.

### Materials and Methods

Wide row spacing (76 cm) and narrow row spacing (38 cm) were tested in 18 trials in Iowa conducted during four years (2010, 2014-2016). Wide row spacing is considered as the control treatment since it is used more commonly. To achieve the narrow, 38-cm, row spacing treatment, a 76-cm row planter was used twice in the same treatment using autosteering or GPS guidance systems. This is feasible for research trials, but not practical for typical commercial use. The experimental design is the same as the one described on Figure 3.

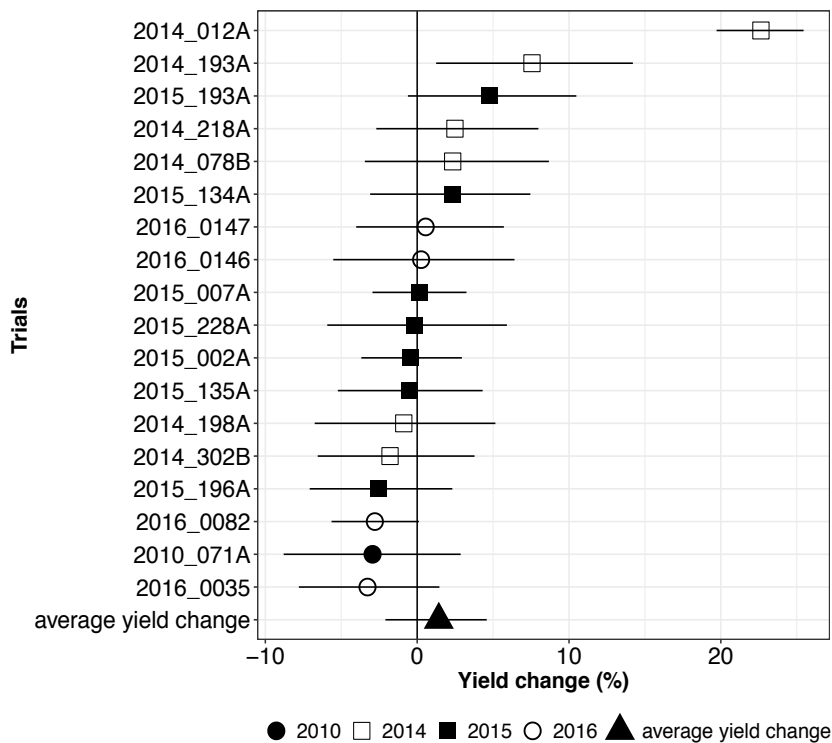


Figure 6: Estimated posterior mean yield changes for on farm trials comparing narrow to wide soybean row spacing and their 95% credible intervals for row spacing on soybean dataset. The codes on the y-axis are the identifiers for different strip trials (fields). All trials were conducted in Iowa.

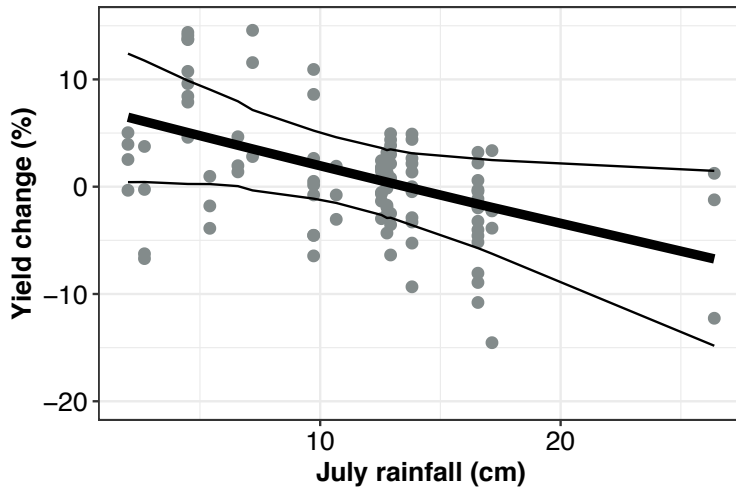


Figure 7: Yield change for all the trials as a function of July rainfall and 95% credible intervals (thinner lines) for the mean change for the row spacing on soybean dataset. Grey dots represent the data points (for all the trials) used to adjust the Bayesian hierarchical model with July rainfall as a continuous variable.

### *Results and Discussion*

The overall mean yield change as a result of switching from wide to narrow row spacing was estimated at 1.4 % [-2.1; 4.6]. The treatment difference is not significant as the low boundary of the credible intervals is negative (Figure 6). The trial 2014\_012A at the top of Figure 6 reached the highest estimated yield change and deviated substantially from other trial results. A plausible explanation is that this trial was affected by hail in early July and these conditions favored the ability of plants in the 38-cm row spacing to recover over that of plants in 76-cm row spacing. Only two of the 18 trials had a significant positive yield response which favored the 38-cm row spacing compared to the 76-cm row spacing.

Our results do not agree with the findings of De Bruin and Pedersen (2008) as they found that 38-cm row spacing yielded 248 kg/ha higher than 76-cm row spacing in Iowa. They advocated the adoption of 38-cm row spacing based on a 5.6% yield increase in a 38-cm vs. 76-cm in a 3-year study at five locations in Iowa. Another study led by the ISU

Extension and Outreach (n.d.) including more than 30 experiments found that the average yield response for the narrow row was higher than 303 kg ha<sup>-1</sup> compared with wide row spacing. Differences between studies were attributed to soil dryness. For example, the relationship between yield response and rainfall in July (Figure 7) suggests that there is an advantage of using 76-cm row spacing when rainfall amounts exceed about 15 cm. Under wet conditions, the 38-cm row spacing results in excessive moisture build-up in the canopy favoring the development of Sclerotinia stem rot. Consistent with this result, Andrade et al. (2019) found that, in the central US region, July rainfall was higher in the experiments showing a yield advantage using wide row spacing. Soybean producers should be aware that the conclusion given by published data using small plot studies do not necessarily agree with the conclusions from OFRN. There is a need to understand why sometimes the results from OFRN and small plot research are not consistent.

### **2.5.3 Soil-Applied Insecticide Impact on Bt Corn Yields**

Hypothesis 1: soil-applied insecticide to BT-corn protect yield from corn rootworm (*Diabrotica virgifera virgifera*) damage.

Hypothesis 2: Soil-applied insecticide reduces the impact on corn root mass.

#### *Background*

The western corn rootworm is one of the most destructive corn pests in the Midwestern United States (Park and Tollefson, 2006). Corn rootworm feeds on corn roots and can drastically reduce root mass and grain yield (Oleson et al., 2005). Planting genetically modified corn, as BT-corn, can be a management strategy to reduce pest

pressure, as they produce insecticidal proteins. In 2009, a study including 64 trials on continuous corn showed that soil insecticides could boost yields of corn rootworm hybrids with the BT trait (Swoboda, 2009). The average yield increase was greater than 672 kg/ha for 40% of the trials. In addition, some farmers used soil-applied insecticide on Bt-corn to reinforce their protection strategy despite a significant cost.

The objectives are i) to study the impact of a soil-applied insecticide to Bt-corn compared to an untreated control by quantifying the yield response and ii) to quantify root damage by measuring root injury (eaten nodes) and root weight.

### *Materials and Methods*

The commercial soil-applied insecticide Aztec® (active ingredients tebupirimphos and cyfluthrin) was compared with an untreated control in 36 trials over eight years (2008-2015). All the trials had corn as a previous crop. The two treatments were applied to corn rootworm resistant corn hybrids (containing the Bt trait). Aztec® was applied in-furrow with farmer equipment. The experimental design was the same as the one described in Figure 3.

### *Results and Discussion*

The Bayesian hierarchical model estimated a yield increase across all trials equal to 1.5 %, with corresponding 95% credible intervals [0.5; 2.6] (Figure 8 on the left). These results were different from a study conducted by Petzold-Maxwell et al. (2013) where yield differences were not detected.

Nine out of 36 trials had a significant yield response and four of them occurred in 2012. This is likely because 2012 was dryer and warmer than normal years, leading to conditions where corn without a soil-applied insecticide suffers greater yield losses (Figure 8 on the right) and the insecticide had a positive impact on corn yield. Our scouting data related to root injury did not show a clear difference between corn with or without a soil-applied insecticide. In the previous study conducted by Petzold-Maxwell et al. (2013) there was no significant difference in root injury between Bt corn with or without a soil-applied insecticide while Gassmann (2012) found that root injury was significantly lower for Bt corn with soil-applied insecticide compared to the control.

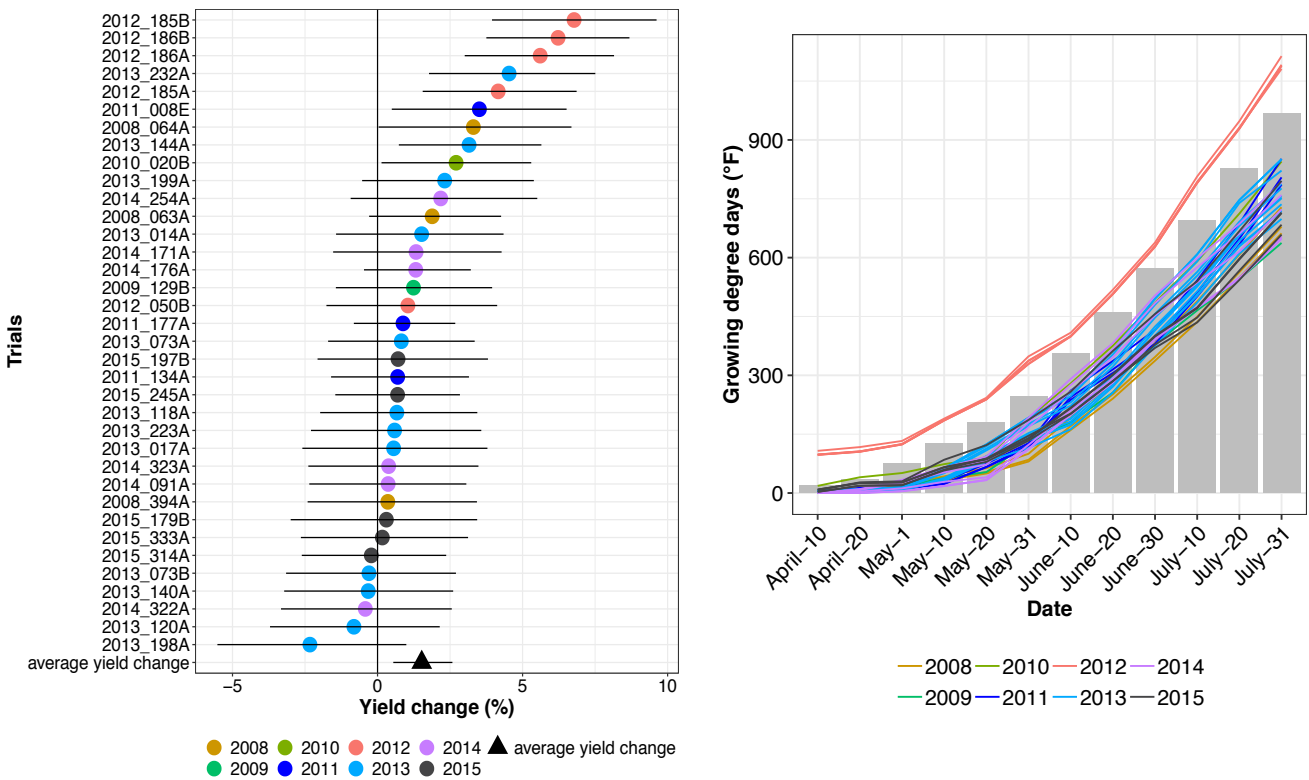


Figure 8: Estimated yield change for individual trials and the corresponding 95% credible intervals for the soil-applied insecticide treatment (Bayesian hierarchical model outputs) (on the left). Each trial is represented by a line and reference values of growing degree days in Central Iowa are represented by gray bars (on the right).

## **2.6 Limitations of our Data-Analytics Framework**

Some caveats of our approach should be highlighted. Limited access to environmental and management variables prevented us from gaining a deeper understanding of yield response variability. Data collection of environmental and management variables should be a crucial step in the ORFN to improve the analysis. Since the analysis of scouting data is specific to each new management practice and research question, our data-analytic framework does not provide uniform visual graphics and statistical methods to summarize this type of data. Nevertheless, we highly recommend studying the relationship between scouting data and other variables collected through the OFRN, such as growing degree days, cumulative rainfall, soil texture, and planting date.

To facilitate the adoption of the new platform, we wrote a manual guide for farmers that explains how to use the web-application and how to interpret the graphics. Next steps will be to include contextual tooltips into the web-application and to provide training to improve and facilitate the adoption of the web-application.

Our web application will continue to evolve as needed with upgrades to existing plots, summaries, and the addition of new management practices. In the future, our web-application could be improved by interviewing users to receive their feedback and to ensure proper interpretation and understanding of the graphics and information available.

## **2.7 Summary**

In this paper, we presented an interactive data-analytics framework for analyses and visualization of data from OFRNs. The aim of our data-analytics framework is to communicate and share descriptive information and statistical summaries of on-farm

research to a broader audience. Our framework is well adapted to a replicated strip trial design using two treatments with or without strip randomization. Most of the visual graphics can be applied to other experimental designs. Graphics and statistical methods were implemented for 36 different management practices tested on soybean and corn. We used statistical approaches that differed from those commonly applied to OFRN data. Trials were analyzed together and not individually, which provides a better understanding of the overall effectiveness of a new management practice. In addition, the uncertainty of the yield response was estimated to include the range of plausible values. Decision making about the new management practice should be based on combining different outputs and summaries from the data-analytics framework in a proper economic and agronomic context.

## **2.8 Software Availability**

Software name: ISOFAST

Developers: Anabelle Laurent, Xiaodan Lyu, Suzanne Fey, Samantha Tyner, Halley Jeppson, Eric Hare.

Year of release: 2018

Hardware required: PC, tablet, mobile

Software required: Web browser. Firefox, Chrome, Safari, Internet Explorer

Programming language: R

Availability: Currently hosted at <https://analytics.iasoybeans.com/cool-apps/ISOFAST/>

License: Free for non-commercial use



## 2.9 Acknowledgments

The authors thank Xiaodan Lyu, Suzanne Fey, Samantha Tyner, Eric Hare and Halley Jeppson for their technical assistance programming ISOFAST. We are thankful to many farmer participants and agronomists along with the staff of the Iowa Soybean Association On-Farm Network for executing on-farm trials and collecting data. This research was funded by the Iowa Soybean Association. The work of DM was partly funded by the STIMUL (Scenarios towards integrating multiscale land-use tools) flagship project as part of the “Investments d’Avenir” Programme (LabEx BASC; ANR-11-LABX-0034). This work was partly based upon research supported by the Chateaubriand Fellowship of the Office for Science & Technology of the Embassy of France in the United States. The work of FM and AL was also supported by the Plant Sciences Institute and Iowa Soybean Research Center at Iowa State University.

## 2.10 References

- Andrade, J.F., J.I. Rattalino Edreira, S. Mourtzinis, S.P. Conley, I.A. Ciampitti, et al. 2019. Assessing the influence of row spacing on soybean yield using experimental and producer survey data. *Field Crop Res.* 230:98-106.
- Bassanezi, R.B., L. Amorim, A.B. Filho, B. Hau, and R.D. Berger. 2001. Accounting for photosynthetic efficiency of bean leaves with rust, angular leaf spot and anthracnose to assess crop damage. *Plant Path.* 50:443-452.
- Bestor, N.R.C., A.E. Robertson, and D.S. Mueller. 2014. Effect of Foliar Fungicides on Late-season Anthracnose Stem Blight on Soybean. *Plant Health Prog.* 15:118-121.
- Bissonnette, K.M., C.C. Marett, M.P. Mullaney, G.D. Gebhart, P. Kyveryga, et al. 2018. Effects of Clariva Complete Beans Seed Treatment on *Heterodera glycines* Reproduction and Soybean Yield in Iowa. *Plant Health Prog.* 19:1-8.
- Chang, W., J. Cheng, J. Allaire, Y. Xie, and J. McPherson. 2016. shiny: Web Application Framework for R.
- Cleveland, W.S. 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. *J. Amer. Stat. Assoc.* 74:829-836.

- De Bruin, J.L., and P. Pedersen. 2008. Effect of Row Spacing and Seeding Rate on Soybean Yield. *Agron. J.* 100:704-710.
- Gassmann, A.J. 2012. Field-evolved resistance to Bt maize by western corn rootworm: Predictions from the laboratory and effects in the field. *J. Invertebr. Path.* 110:287-293.
- Hadfield, J.D. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *J. Stat. Software.* 33.
- ISU Extension and Outreach. n.d. Row spacing is important to maximize your yield. Iowa State University, Extension and Outreach. [http://crops.extension.iastate.edu/soybean/production\\_spacing.html](http://crops.extension.iastate.edu/soybean/production_spacing.html) (accessed 5 April 2017).
- Jeschke, M., and R. Ahlers. 2018. Foliar Fungicide and Insecticide Use in Soybeans. <https://www.pioneer.com/home/site/us/agronomy/library/foliar-fungicide-insecticide-soybeans/> (accessed 3 October 2018).
- Kyveryga, P.M., T.M. Blackmer, and D.S. Mueller. 2013. When Do Foliar Pyraclostrobin Fungicide Applications Produce Profitable Soybean Yield Responses? *Plant Health Prog.* 14.
- Kyveryga, P.M., T.A. Mueller, D.S. Mueller, D.K. Shannon, D.E. Clay, et al. 2018. On-Farm Replicated Strip Trials. ACSESS Publications. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Inc.
- Mahoney, K.J., R.J. Vyn, and C.L. Gillard. 2015. The effect of pyraclostrobin on soybean plant health, yield, and profitability in Ontario. *Can. J. Plant Sci.* 95:285-292.
- Matthewson, M., M. Fery, and M. Powell. 2013. Creating Farmer Networks: A Toolkit for Promoting Vibrant Farm Communities. <https://www.sare.org/content/download/73146/1058808/file/pnw638.pdf> (accessed 30 November 2018).
- Minnesota Association of Wheat Growers. 2018. NW Minnesota On-Farm Research Network. NW Minnesota On-Farm Research Network. <http://smallgrains.org/farm-research-network/> (accessed 5 March 2018).
- Moayedi, A.A., and M. Azizi. 2012. Improvement of Knowledge and Skills Level of Wheat-Cultivating Farmers using On-Farm Researches. *Procedia – Soc. and Behav. Sci.* 46:2258-2261.
- Moore, K.J., and P.M. Dixon. 2015. Analysis of Combined Experiments Revisited. *Agron. J.* 107:763-771.

- Nelson, B.P., R.W. Elmore, and A.W. Lenssen. 2015. Comparing Yield Monitors with Weigh Wagons for On-farm Corn Hybrid Evaluation. Agron. & Horticulture Faculty Publications. 841
- Oleson, J.D., Y.-L. Park, T.M. Nowatzki, and J.J. Tollefson. 2005. Node-Injury Scale to Evaluate Root Injury by Corn Rootworms (Coleoptera: Chrysomelidae). *J. Econ. Entomol.* 98:1-8.
- Park, Y.-L., and J.J. Tollefson. 2006. Development and economic evaluation of spatial sampling plans for corn rootworm *Diabrotica* spp. (Col., Chrysomelidae) adults. *J. Appl. Entomol.* 130:337-342.
- Pennsylvania Soybean Board. 2018. PA On-Farm Network. <http://pasoybean.org/checkoff-at-work/research/on-farm-network/> (accessed 5 March 2018).
- Petzold-Maxwell, J.L., L.J. Meinke, M.E. Gray, R.E. Estes, and A.J. Gassmann. 2013. Effect of Bt Maize and Soil Insecticides on Yield, Injury, and Rootworm Survival: Implications for Resistance Management. *J. Econ. Entomol.* 106:1941-1951.
- Philibert, A., C. Loyce, and D. Makowski. 2012. Assessment of the quality of meta-analysis in agronomy. *Agric., Ecosyst. & Environ.* 148:72-82.
- Pinheiro, J., and D. Bates. 2000. Linear Mixed-Effects Models: Basic Concepts and Examples. *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag, New York. p. 3–56
- Pruss, J., D. Beegle, A. Turgeon, R. Day, and R. Weaver. 2005. Using On-Farm Data to Validate Crop Management Recommendations and Implementation. *Agron. J.* 97:113-117.
- Qadir, G., F.-U. Hassan, and M.A. Malik. 2007. Growing Degree Days and Yield Relationship in Sunflower (*Helianthus annuus* L.). *Int. J. Agric. Biol.* 9:564-568.
- Ramsey, F.L., and D.W. Schafer. 2013. *The statistical sleuth: a course in methods of data analysis*. 3rd ed. Brooks/Cole, Cengage Learning, Australia; Boston.
- RStudio Team. 2015. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.
- Salvagiotti, F., K.G. Cassman, J.E. Specht, D.T. Walters, A. Weiss, et al. 2008. Nitrogen uptake, fixation and response to fertilizer N in soybeans: A review. *Field Crop Res.* 108:1-13.
- Swoboda, R. 2009. Can Soil Insecticide Boost Yield of CRW Trait Hybrids? *WallacesFarmer*. <https://www.wallacesfarmer.com/story-can-soil-insecticide-boost-yield-of-crw-trait-hybrids-9-22332> (accessed 10 April 2018).
- Swoboda, C., and P. Pedersen. 2009. Effect of Fungicide on Soybean Growth and Yield. *Agronomy Journal* 101(2): 352.

University of Nebraska. 2018. Cropwatch.

<https://cropwatch.unl.edu/farmresearch/resultshome> (accessed 3 April 2018).

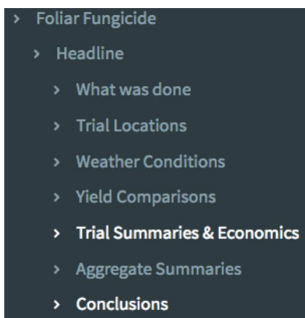
Unwin, A., M. Theus, and H. Hoftman. 2006. Graphics of large datasets: visualizing a million. New York: Springer.

Wise, K., and G. Buechley. 2010. Evaluation of foliar fungicides for soybean disease management in southeastern Indiana, 2010. Plant Dis. Manage. Reports 5:FC116.

Wojciechowski, J., A. Hopkins, and R. Upton. 2015. Interactive Pharmacometric Applications Using R and the Shiny Package. CPT Pharmacometrics Syst. Pharmacol. 4:146-159.

Wrather, J., and S. Koenning. 2006. Estimates of disease effects on soybean yields in the United States. J. Nematol. 38:173-180.

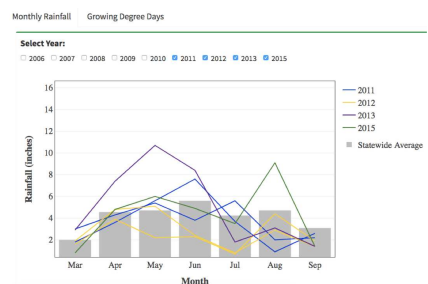
## 2.11 Supplemental Materials



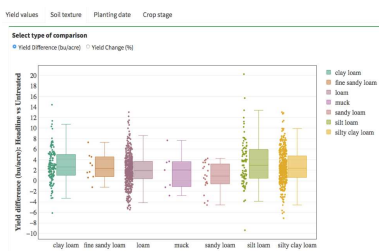
### What was done:

Foliar fungicide pyraclostrobin (Headline) applications on soybean and yield response. Most of the applications were done by farmers using ground sprayers but in about 20% trials the foliar fungicide was aerially applied.

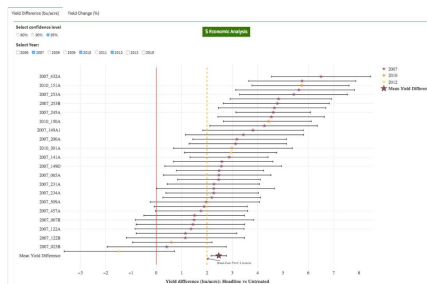
### Weather conditions:



### Yield comparisons (by soil texture for example):



### Individual summaries and economic analyses:



### Aggregate summaries:

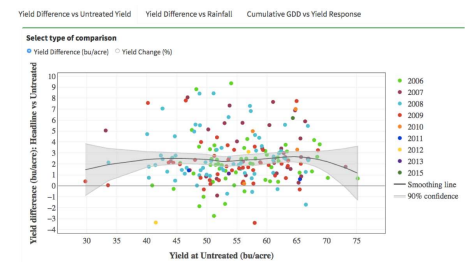


Figure S1. Screenshots of the web- application interface and examples of product outcomes

### CHAPTER 3. INTERACTIVE WEB-BASED DATA VISUALIZATION AND ANALYSIS TOOL FOR SYNTHETIZING ON-FARM RESEARCH NETWORKS DATA.

Modified from a paper published in *Research Synthesis Methods*, special issue “Data Visualization”,

Anabelle Laurent<sup>1</sup>, Xiaodan Lyu<sup>2</sup>, Peter Kyveryga<sup>3</sup>, David Makowski<sup>4</sup>,

Heike Hofmann<sup>2</sup>, Fernando Miguez<sup>1</sup>

<sup>1</sup> Department of Agronomy, Iowa State University, Ames, IA, USA

<sup>2</sup> Department of Statistics, Iowa State University, Ames, IA, USA

<sup>3</sup> Center for Farming Innovation, Iowa Soybean Association, Ankeny, IA, USA

<sup>4</sup> INRAE, Centre Île-de-France, Paris, France

#### 3.1 Abstract

The on-farm research network concept enables a group of farmers to test new agricultural management practices under local conditions with support from local researchers or agronomists. Different on-farm trials based on the same experimental design are conducted over several years and sites to test the effectiveness of different innovative management practices aimed at increasing crop productivity and profitability. As a larger amount of historical trial data are being accumulated, data of all the trials require analyses and summarization. Summaries of on-farm trials are usually presented to farmers as individual field reports, which are not optimal for the dissemination of results and decision making. A more practical communication method is needed to enhance result communication and decision making. R Shiny is a new rapidly developing technology for turning R data analyses into interactive web applications. For the first time for on-farm research networks, we developed and launched an interactive web tool called ISOFAST

using R Shiny. ISOFAST simultaneously reports all trial results about the same management practice to simplify interpretation of multi-site and multi-year summaries. We used a random-effects model to synthesise treatment differences at both the individual trial and network levels and generate new knowledge for farmers and agronomists. The friendly interface enables users to explore trial summaries, access model outputs, and perform economic analysis at their fingertips. This paper describes a case-study to illustrate how to use the tool and make agronomic management decisions based on the on-farm trial data. We also provided technical details and guidance for developing a similar interactive visualization tool customized for on-farm research network. ISOFAST is currently available at <https://analytics.iasoybeans.com/cool-apps/ISOFAST/>.

### **3.2 Introduction**

The on-farm research network concept enables a group of farmers to test a new agricultural management practice, or product, under local conditions with the support of researchers, local agronomists, or crop consultants<sup>1,2</sup>. Farmers are in charge of setting-up their own on-farm trial using commercially available field equipment to develop locally adapted recommendations. As a consequence, on-farm research networks help farmers improve their productivity, efficiency, and profitability. The experimental design commonly used is the replicated strips where a new management practice (e.g., plant nutrition, disease management, pest control, plant population) is compared to a standard farmer practice, and the crop yield is the outcome variable<sup>3</sup>. Different on-farm trials based on the same design are located at multiple sites across Iowa in different years to capture different soil and climatic conditions. In addition to the scientific aspect, the social dimension of on-

farm research networks is essential as farmers can exchange their experience and knowledge. Also, the results of the experiments are shared with other farmers, researchers, and extension staff in the network over multiple years and on-farm trials.

Individual trial reports are a standard way to share data collected in the set of trials included in a given on-farm network. Each report summarizes the results of each trial, in particular, treatment yield averages, environmental characteristics (e.g., soil texture, weather data, location), and cropping practices (e.g., planting date, variety). Some institutions such as the Iowa Soybean Association (hereafter called ISA), the Nebraska On-Farm Research Network and, the South Dakota On-Farm Research program have online searchable databases which allow users to query individual trial reports by year, crop and management practices<sup>1</sup>. These individual field reports are useful, but they provide limited information and they are not able to present in a synthetic way all the results collected from the different trials and, in particular, to describe the between-farm variability of the effect of the new management practice tested.

Therefore, there is a need to develop a visualization tool that allows farmers, agronomists, and researchers to share, visualize and analyze data collected over multiple years and sites. Graphical displays represent an essential part of exploratory data analysis that can help identify patterns such as trends and clusters<sup>4</sup>. Moreover, a graphical summary is a crucial complement of statistical approaches as it provides an intuitive understanding of the model outputs<sup>5,6</sup>. Effective data visualization enhances sharing and communicating insights from large amounts of data<sup>7</sup>.

Historical trial data is being accumulated from on-farm research trials and provides an outstanding resource from which we developed a web-based publicly available

interactive tool, called ISOFAST based on R Shiny<sup>8,9</sup>. We selected appropriate R packages which implement Data-Driven Documents (also known as D3) from the JavaScript library to produce interactive charts so that users can explore trial summaries through mouse clicking or hovering. The tool simplifies the interpretation of multi-location and multi-year summaries by simultaneously reporting all trial results about the same management practice tested.

We used a random-effect model that allows estimating and quantifying the uncertainty of the performance of a specific management practice. These estimates were calculated for individual locations and also across all on-farm trials, in order to understand the practice performance at the network level. This statistical model helps to summarize existing knowledge and provides more reliable conclusions to farmers and agronomists. Users can also use the online tool to customize a static report by choosing their questions of interest through different filters or other interactive widgets. To our knowledge, this is the first time that data from on-farm research networks are shared through a web-application using interactive graphics and advanced statistical analysis.

We introduced the data, the model, and the tool in more detail in Section 2. A case study is presented in Section 3 to illustrate the functionalities of the tool. We discussed the pros and cons of the tool and provided some general suggestions in Section 4 for developing similar agronomic summarization tools. Conclusion remarks are given in Section 5.



### 3.3 Materials and Methods

#### 3.3.1 General Architecture

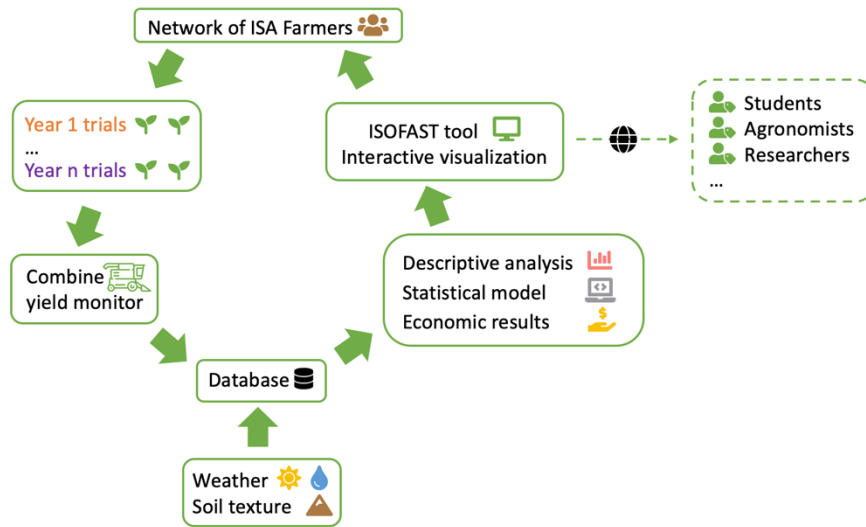


Figure 1. System architecture of the ISA on-farm research network.

ISOFAST stands for Interactive Summaries of On-Farm Strip Trials and plays an important role in the on-farm research network (Figure 1). The target audience of ISOFAST is farmers, who have carried out the on-farm strip trials, and also crop consultants and agronomists. The trials have been located at multiple sites across Iowa in different years. Depending on the management practice or product tested, between 6 to 206 trials were conducted and about the same number of farmers were involved (see Figure A1, supplementary material). Crop yield in bushel/acre is collected using a yield monitor installed on combine harvester and then data are processed and cleaned by ISA staff (Figure 1). Cleaning protocols are based on late-season aerial imagery which allows detecting different within-field problems in farm equipment, extremes in combine harvester speed and grain moisture, impact of flooding, plant stand reduction and,

waterways or water buffers on yield monitor observations<sup>10-12</sup>. Most of the analyses are based on yield differences or on yield ratios in paired strips instead of absolute yield values, which reduces the impact of yield monitor calibration for grain weight and moisture by farmers. Besides yield observations, ISA collects relevant auxiliary information such as field management, weather (growing degree days and monthly rainfall), and soil characteristics. With a growing database, the ISA analytics team conducts some statistical analyses (descriptive and algorithmic) and translates the statistical model into economic results as presented in the following case study. ISOFAST allows farmers to access different types of statistical analyses and to interact with the underlying data (raw and processed) through a web browser, thus providing insight to make better farming management decisions. In addition, trials are continuously being conducted and data analyses is expanded at the same time. Moreover, the accessibility of ISOFAST reaches a broader audience beyond farmers, such as students, agronomists, researchers, policy makers, and ag industry. The ISOFAST tool is accessible for any user without restrictions on permission such as IP address. For farmer privacy concerns, each trial is labeled by an anonymous identifier and the exact coordinates are not identifiable at the current zoom level on any map of trial locations. Participants' names and trial locations are not identified in the tool. All data, including but not limited to field and farm management information, natural resource and spatial data provided to ISA and/or collected by ISA are owned by enrolling participants. ISA will retain all data and information only for additional analyses aimed to improve crop management, natural resources, and environmental quality and to advance scientific knowledge.

### 3.3.2 Data Description

We consider here two types of networks of on-farm trials conducted in Iowa (USA) on two major crops, soybean, and maize, testing treatments (management practices) related to plant nutrition, crop management, and crop protection. In total, 27 and 24 management practices were tested on soybean and maize, respectively (see Figure A1, supplementary material). Each network includes on-farm trials related to one specific treatment conducted at multiple locations across multiple years. The data collected in a given network are stored in a specific Excel file in which each row includes the results obtained in one replicate of one trial (one site-year), i.e., a trial identifier, a replicate identifier, the experiment year, the geographic location, planting date, yield in bushel/acre of the treatment group, and yield in bushel/acre of the control group. The Analytics team of the ISA uses geodatabases in the data processing and cleaning stages with ESRI ArcView Desktop GIS. In addition, Microsoft Access and SQL databases are used for data wrangling. Farmer contact and field management information are collected using a Microsoft CRM (Customer Relationship Management) database.

Environmental factors (soil, water, and temperature) are essential to plant growth. To account for variability in the crop yields due to soil condition, location-level soil texture data (loam, sand, or clay) are matched from the soil survey geographical (SSURGO) database<sup>13</sup>. The Iowa Mesonet<sup>14</sup> maintains year specific weather data for the five Iowa regions (Central North, Northwest, Northeast, Southwest, and Southeast) in Iowa. We used 15-year monthly regional average precipitation in inches as a normal rainfall condition. Growing degree days (GDD) is a measure of heat accumulation used to predict plant and pest development rates, including crop maturity. We computed average GDD in Fahrenheit

degrees every ten days from April to July at three reference stations (Central, North, and South) across the experiment years. The referenced GDD is then contrasted with the trial-specific GDDs to show a relative temperature condition. Four-kilometer grid daily rainfall data from the Iowa Environmental Mesonet are used to assigned daily rainfall to each trial location then averaged to monthly estimates. After fitting the model described in the following subsection to the data, point estimates together with confidence intervals of trial-level and overall yield response (difference or ratio) are computed for each network dataset. Three confidence levels (0.8, 0.9 and 0.95) for intervals are available in the tool.

### 3.3.3 Statistical Model Description

We used a random-effect model with nested errors<sup>1</sup> to estimate the yield of a treatment compared to a control in individual on-farm trials and the overall mean yield response across all on-farm trials. Trials were represented by site-years as they were rarely repeated at the same location over time. The model was fitted using a Bayesian approach, which considers previous knowledge as prior distributions and better quantifies the uncertainty in variation of yield response within and across trials <sup>1</sup>.

Both yield ratio and yield difference, as the response variable, were modeled. In fact, farmers prefer yield difference as it is unit-dependent and can be easily expressed as economic gain or loss. Nevertheless, yield ratio, typically used in meta-analysis, are useful for agricultural scientists as it is unitless and might be less dependent on other factors.

The statistical model is defined as follows:

$$\ln(R_{ij}) = \mu + \alpha_i + \varepsilon_{ij} \quad (1)$$

where  $\ln(R_{ij})$  represents the logarithm of the  $j^{\text{th}}$  yield ratio in the  $i^{\text{th}}$  site-year ( $R_{ij}$  is estimated as yield of the new management practice treatment in the  $i^{\text{th}}$  trial and  $j^{\text{th}}$  replicate divided by yield of the control in the  $i^{\text{th}}$  trial and  $j^{\text{th}}$  replicate);  $\mu$  represents the true mean log yield ratio in the considered population;  $\alpha_i$  represents the random effect of the  $i^{\text{th}}$  site-year, and  $\varepsilon_{ij}$  represents the random noise. The random variables  $\alpha_i$  and  $\varepsilon_{ij}$  are assumed to follow independent Gaussian distributions with mean zero and constant variances,  $\alpha_i \sim N(0, \sigma_\alpha^2)$  and  $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ .

Using the yield difference as the effect size, the statistical model (1) becomes:

$$D_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (2)$$

where  $D_{ij}$  represents the  $j^{\text{th}}$  yield difference in bushel/acre in the  $i^{\text{th}}$  site-year ( $D_{ij}$  is estimated as the yield of the new management practice treatment in the  $i^{\text{th}}$  trial and  $j^{\text{th}}$  replicate minus yield of the control in the  $i^{\text{th}}$  trial and  $j^{\text{th}}$  replicate);  $\mu$  represents the true mean yield difference in the considered population; other notations remain unchanged (see model (1) described above).

Model (1) includes three parameters,  $\mu$ ,  $\sigma_\alpha^2$ , and  $\sigma_\varepsilon^2$ . The variance  $\sigma_\varepsilon^2$  quantifies the variability of  $\varepsilon_{ij}$  within trial (between replicates) and  $\sigma_\alpha^2$  quantifies the variability of  $\alpha_i$  across trials (site-years).

A continuous or categorical explanatory variable can be added to the random-effect model, such as rainfall or soil texture, respectively. The random-effect model was fitted using a Bayesian method implemented in the R package, MCMCglmm<sup>1,15</sup>.

### 3.3.4 Economic Responses

Estimated empirical cumulative distribution functions of trial average yield responses are used to conduct a break-even economic analysis where a break-even yield response and expected profit are calculated. A break-even yield difference in bushel/acre is calculated as the treatment cost in dollars per bushel divided by grain market price (i.e., cost of product and application) in dollars per acre. For example, if maize is sold \$3.50/bushel and the treatment (applying 50 lbs /acre less N fertilizer) cost \$17.50/acre, then the break-even is 5 bushel/acre. In other words, \$17.50 is saving from not applying an extra 50 lbs N/acre compared to the control, the break-even yield difference (yield of treatment unit minus the yield of the control unit) is minus 5 bushel/acre, which means there is no economic loss to apply less N fertilizer as long as yield loss is 5 bushels/acre fewer than the standard practice with the full N rate. The probability of exceeding the break-even cost is calculated as the distance on the y-axis between 100% and the intersection of the cumulative distribution curve with the break-even line.

### 3.3.5 Tool Description

The menu bar (see Figure 2) on the left of the ISOFAST tool mimics the hierarchical structure of the treatments (see Figure A1, supplementary material). Users can explore experiment data related to a specific treatment by navigating to the corresponding subcategory. For example, if we are interested in trials testing the effect of reducing normal farmer N application rates by 50 lbs/acre or 30%, we can click “Corn”, “Nitrogen Studies”, and “Normal N minus 50 lbs N” in the menu bar sequentially. Under each treatment, the

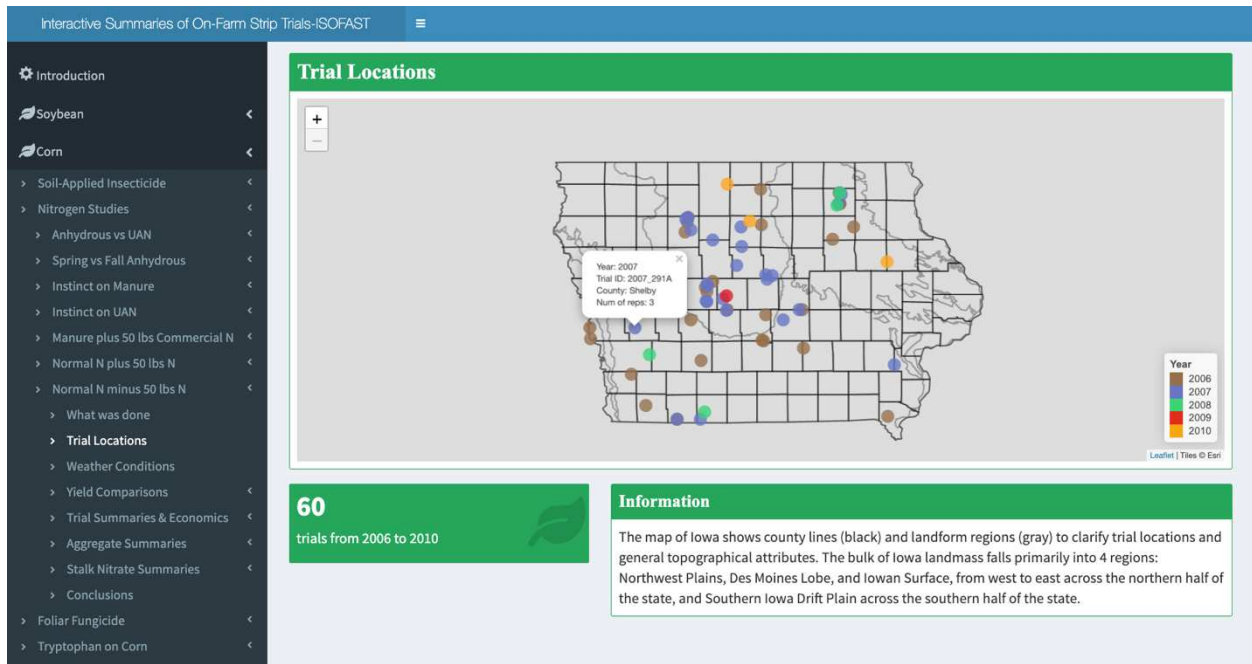


Figure 2. Layout of ISOFAST: menu bar on the left; visuals on the right

sub-menu items are linked to a set of almost identical analysis. The sub-menu items help users to explore in-depth data summaries. In particular, starting with introducing some background information of the considered treatment, the tool then informs the user about geographic locations of the trials with an interactive map of trial locations colored by year (see Figure 2), followed by a spaghetti plot of the trial-specific weather data stacked over a bar plot of the referenced long-term weather condition for specific region of Iowa.

Before visualizing the random-effect model estimates with a forest plot<sup>16</sup>, a forest plot of the raw (individual experimental units) yield ratio or yield response is presented (see Section 3). To show the distribution of replicate-level yield responses (difference or ratio) for different soil texture classes or planting date (“early” or “late” in relative to a specific date), two boxplots with jitters are available. After clicking the “Aggregate Summaries” menu for each subcategory treatment, the user can view scatter plots of the

trial-level average yield responses versus yield for the control treatment, monthly/short-term rainfall or cumulative GDDs. To help users understand potential effects of other factors on yield responses, locally weighted regression and smoothing line<sup>17,18</sup> (also denoted LOWESS) with 90% confidence regions are added to the scatter plot. The last sub-menu item takes the user to the study conclusion with main findings in text and a download widget, allowing the user to download a static report in pdf version for the selected sections (see Figure 3). Table A1 in the supplementary material shows an example of visuals of a typical analytics process under each subcategory treatment.

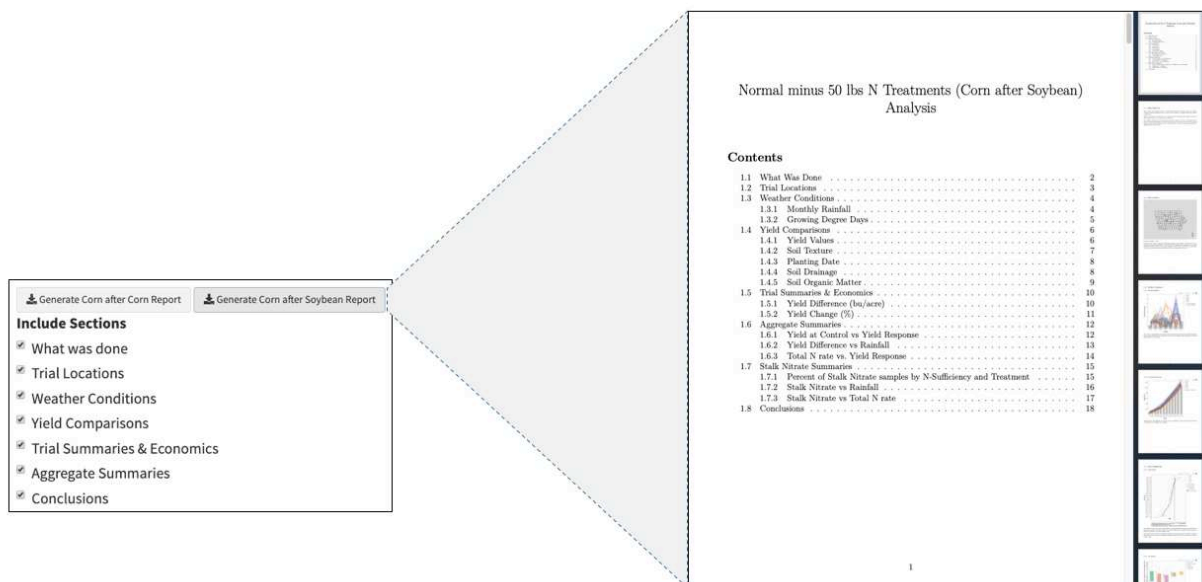


Figure 3: Static report download button in the conclusion tab (on the left) to generate a PDF report (on the right). Users can select/deselect sections of interest.

### 3.3.6 Technical Details

R Shiny is a web development framework that can turn R data analyses and summaries into interactive webpages<sup>9</sup>. Usually, a Shiny program consists of two major R functions, “ui” and “server”, where “ui” defines the user interface including layout, input



widgets, position and size of output visual components (usually plot, table, and map); “server” defines how the output visual components react to user’s inputs. For example, for the visual presented in Figure 4, the radio button asking the user to choose yield difference (bushel/acre) or yield change (%) is defined in “ui”, as well as arranging the output plot below the radio button and with a height of 600 pixels and a width equal to the window width. And “server” outputs the forest plot with variable on the horizontal axis being reactive to the user’s choice of the comparison type.

Besides “ui” and “server”, the user can define some global variables evaluated before running the application. The corresponding R code for global variables can be saved as a separate R file (e.g. “global.r”). For ISOFAST, data import, plot theme (such as layout and font size), and some utility functions are defined in this file. By our design, each subcategory treatment is illustrated with an almost identical set of visuals. As the experimental data are being accumulated, a set of customized R plotting functions are designed such as that plots of the same style are produced given an input dataset. To fulfill that functionality, each input dataset is pre-processed to follow a consistent data structure. In this way, we make the tool highly scalable to a growing database and highly flexible to changes in the future.

The user interface design of ISOFAST mainly relies on the R package `shinydashboard`<sup>19</sup>. Considering that accessibility is key (reaching a wide audience), different users might prefer a larger font size than the default, thus we used a customized Cascade Style Sheet (CSS) to achieve that. We use the R package `leaflet`<sup>20</sup> for designing interactive maps in the tool and `plotly`<sup>21</sup> for interactive figures. Those visual components are interactive in that users can hover a data point (to view pre-defined tooltips), zoom in

or out, and filter data (by hitting color legend). Figures 4 and 5 show how the tooltip creates an interactive figure. We use the `modelDialog` function in the R package `shiny`<sup>9</sup> for designing pop-up windows when users hit the “economic analysis” button (see Figure 6). A download button, together with `rmarkdown`<sup>22</sup> rendering, enables producing a PDF report containing complete or partial visuals and texts upon the user request for a selected menu item (see Figure 3).

### 3.4 Case-study

Nitrogen (N) is a primary nutrient for crop growth and development. Maize needs N fertilizer because the soil system typically cannot supply the full N demand of a maize plant<sup>23,24</sup>. Producers should apply optimal and environmentally safe N application rates that enhance yield and profit but reduce N overapplications. Indeed, a high N application rate can lead to an elevated level of  $\text{NO}_3$  in the soil and high risk of potential loss by leaching, especially in a high rainfall area<sup>25</sup>. However, an appropriate rate of N can minimize the negative environmental impacts.

In that context, an on-farm research network study was focused on the impact of reducing farmers’ normal N fertilizer application rates to maize by about one third (around 50 lbs/acre) without economic loss. A reduced N fertilizer application rate (treatment) was tested in 44 trials over four years (2006-2008, 2010) and compared with a normal N fertilizer application rate (control). The average normal N rate was 145 lb/acre, and the reduced application rate was 95 lb/acre. All the trial locations had soybean as a previous crop. These experiments produced yield data that were collected on combine harvesters operated by farmers and equipped with Global Position System (GPS). Each trial included

in the on-farm research network had a minimum of three replicates. Some trials required more than three replicates to capture spatial variability of yield responses within an entire field. The width of an individual strip depends on the size of application equipment and can range from 5 to 20 m. The length of a strip depends on the field size. For example, a typical field in Central Iowa is about 32 hectares, approximately 457 meters by 701 meters. Trials were randomly distributed throughout the different Iowa landform regions, with a majority located in the Des Moines Lobe. Therefore, our data covers a broad set of environmental conditions and field managements across Iowa.

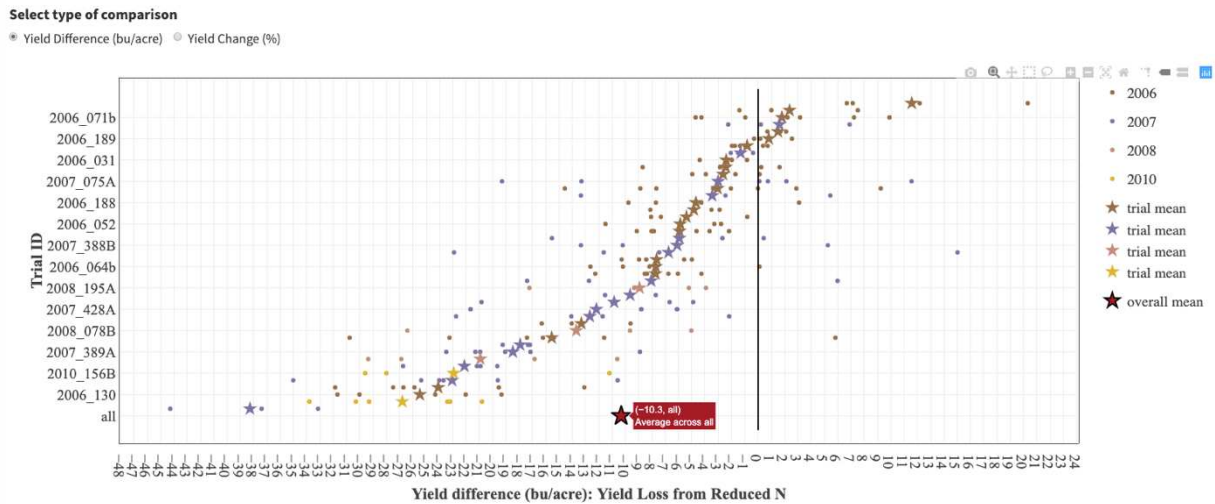


Figure 4. Observed yield difference in bushel/acre (yield at normal rate minus yield at reduced rate) at each trial. Dots indicate the replicates and the star corresponds to the average measured yield difference. The dots and stars are colored by year trials were conducted. The overall yield difference (at the network level, using the replicates from all trials) is indicated by a red star (at the bottom of the graph). Trials are ranked by the increasing average yield difference. The vertical black line represents a yield difference equal to 0 bushel/acre.

It is important to show observed yield differences before describing the statistical model outputs as a way to guide the user through the analysis. Presenting the observed

yield differences in bushel/acre (or yield change in percentage) for each on-farm trials with a forest plot (Figure 4) helps understand the existence and magnitude between-trial and within-trial variabilities in the observed yield responses. As one way to present the distribution of the yield responses, trial IDs in the vertical axis of the plot are ordered from bottom to top by increasing average yield responses.

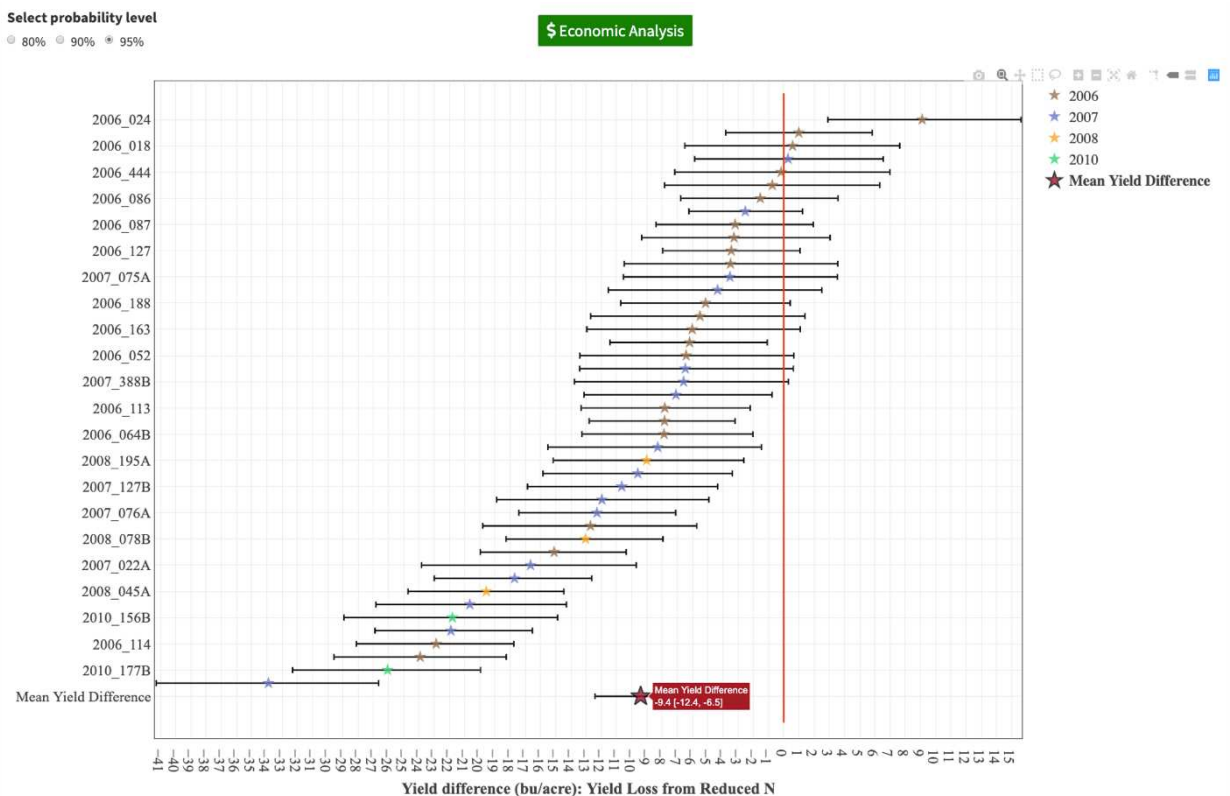


Figure 5. Estimated random effects of yield differences for individual trials with their 95% confidence intervals (brown, lilac, orange, and blue stars indicating point estimate; black horizontal lines indicating confidence interval) and mean yield difference with its 95% credible interval (red star and black horizontal line) for the overall effect across all on-farm research network of 44 trials assessing the effect of a reduced N fertilizer application rate compared to a standard N fertilizer application rate. The vertical red line represents a yield difference equal to 0 bushel/acre. Computations were done using the model (2) fitted using a Bayesian method.

The estimated individual and overall effects are visualized simultaneously in the ISOFAST (Figure 5). The trials are ranked by increasing order of estimated mean effect to

show the variability between trials. Cleveland and McGill<sup>26</sup> (1984) found positions along a common scale help comparison of quantitative data better than nonaligned scales. The overall effect (across all years and trials) is always at the bottom of the graph. In our case study, the Bayesian random-effect model estimated a yield difference or yield loss from reduced N across all trials equal to negative 9.4 bushels/acre with corresponding 95% credible interval ranging between negative 12.4 bushels/acre and negative 6.5 bushels/acre. In Figure 5, 25 out of 44 trials had a statistically significant yield loss relative to the treatments with the normal N fertilizer rate. To make reliable decisions for the future, information about both overall and individual effects are necessary.

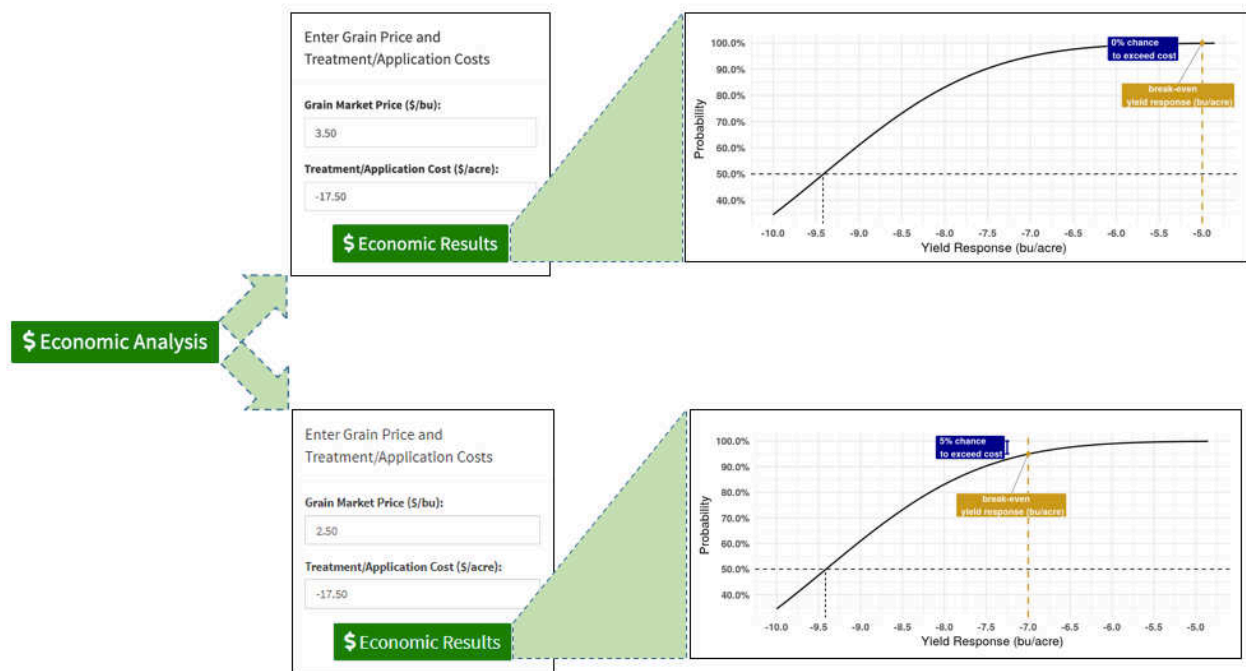


Figure 6: The user interacts with the “\$ Economic Analysis” widget (on the left) by entering a maize market price (\$/bushel) and a treatment application cost or cost of reduced N fertilizer that we did not apply (\$/acre) (in the middle) to generate the cumulative distribution of potential yield loss due to reduced N application, calculate the break-even yield difference or loss (yellow vertical dashed line) and the probability of exceeding the

direct application cost (blue window). The first scenario (on the top) is defined by a grain market price equal to 3.50 \$/bushel and a treatment application cost equal to negative 17.50 \$/acre, meaning that the cost of saved fertilizer is \$17.50 if the decision is to reduce farmers' normal rate by 30%. The second scenario (at the bottom) is defined by a grain market price equal to 2.50 \$/bushel and the cost from saved fertilizer is 17.50 \$/acre.

ISOFAST economic analysis is presented in Figure 6. With a maize price at \$3.50 per bushel and 50 lb N application cost at \$17.50 per acre, the breakeven yield loss is equal to negative 5 bushels/acre, meaning that we can lose yield but this yield loss will be less than the cost of saved N fertilizer. It is very unlikely to recover the cost of yield loss from a reduction of 50 lb N/acre as the probability of reaching the breakeven yield difference (-5 bushels/acre) with 50 lb N/acre reduced N is equal to 0%. If the user chooses a different grain market price and treatment application cost, the probability of economic return will be adjusted. For example, Figure 6 shows that the chance to exceed the cost N in this case changes from 0% to 5% when the market price of maize per bushel decreases from \$3.50 to \$2.50 given that the treatment cost is kept the same. Both the probability of exceeding the cost enables farmers to better understand the uncertainty better and evaluate economic decisions in the future. In a similar study, the University of Nebraska-Lincoln tested the effect of a fertilizer N reduction by 5% on marginal profit. They found out that the fertilizer N reduction resulted in a grain yield reduction of 6 bushels/acre for cropping history as maize following soybean<sup>27</sup>.

Considering early season rainfall, yield losses from reduced N tended to be greater by about 1 bushel/acre (a slope of the regression line and considering the effect of N rate applied) with each additional inch of cumulative May through June rainfall (Figure 7). These results confirm that reducing farmers' normal N rate by 30% provided consistent

yield losses under the evaluated conditions, but the decision regarding whether or not to reduce N application by 50 lb/acre would also depend on early season (May through June) rainfall.

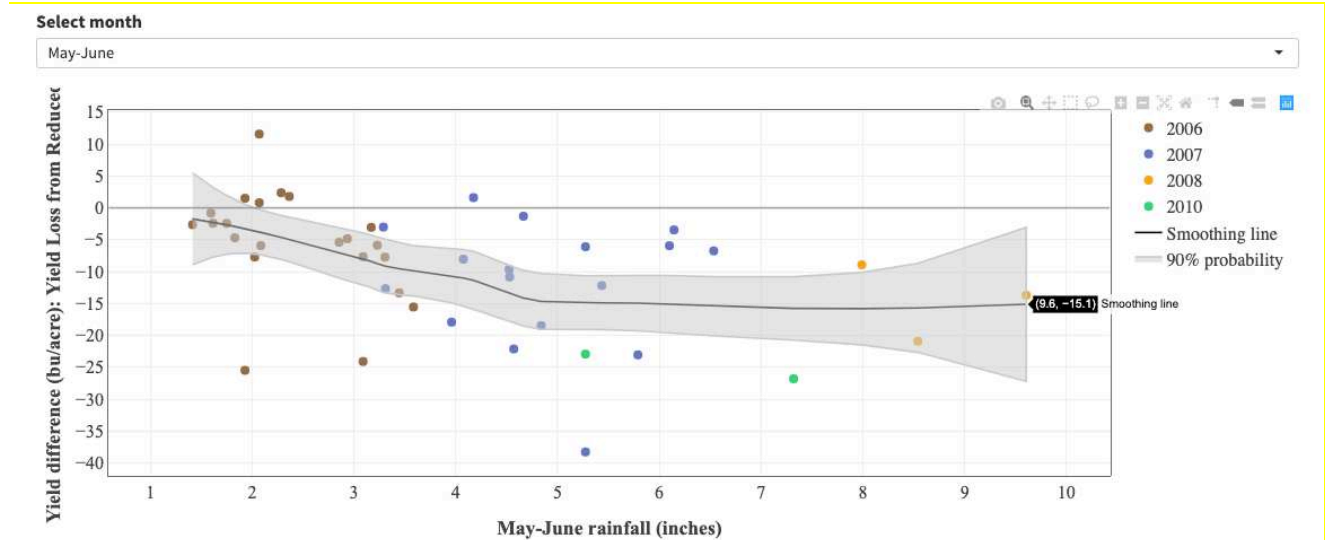


Figure 7. Averaged trial yield difference (bushel/acre) per trial as a function of cumulative May and June rainfall (inches). Year of experiment is displayed with different colors. The user can select a month of interest, or a cumulative period of rainfall, using the scroll down menu on the top of the graph.

### 3.5 Discussion

A significant advantage of ISOFAST is that it simplifies the understanding of advanced statistical methods for summarizing yield responses from agronomic treatments from several trials conducted over time and assessing the uncertainty in yield response for each treatment. The economic analysis bridges the complex statistical analysis and practical economic outcomes by translating abstract statistical language into intuitive human language. Farmers' feedback and a review made by agronomists from ISA during the development of the ISOFAST resulted in a friendly user interface. As a potential study, it could be interesting to understand how ISOFAST may play a role in changing (or not)

farmers' management practices and how they convert ISOFAST analyses summaries into decision making.

The carefully designed system architecture and tool structure make the process reproducible and scalable. ISOFAST integrates complex information from multiple sources and reduces the cognitive load on users. It is possible to navigate through experiment summaries of interest in one study to another using a web browser. Even though the widespread use and adoption of ISOFAST will take time as farmers and other stakeholders are used to individual static reports, a complete or partial PDF report is downloadable from the tool. The Iowa Soybean Association and Iowa State University have published several farmer friendly articles, newsletters and developed two Youtube tutorials to help users better understand the ISOFAST graphs and visual displays. Since launched in April 2019, ISOFAST has been visited 2,400 times by 1,700 unique users.

However, some work needs to be done to speed up the response of ISOFAST. Indeed, users can experience a slow loading of some webpages, as R is single-threaded. The background R program of ISOFAST usually answers user's requests one by one instead of doing multiple tasks at one time. Continuous progress have been made by the Shiny developer team to promote shiny in production<sup>28</sup>, including UI testing, load testing, plot caching and so on. The R package profvis<sup>29</sup> can be used to identify the slowest part in the R computing. After this has been identified, we can either move part of the work out of Shiny or optimize the R code to accelerate computation. Other minor issues should be addressed in future development. For example, we need to optimize the ISOFAST layout for mobile devices, especially for farmers who are frequent users of mobile phones rather than desktops or laptops. Indeed, we surveyed four groups of farmers, about 90 individuals. Two



groups insisted that they would like to use the tool on mobile phones, while the other two indicated their preference for laptops and desktop computers.

In general, by using the free and open-sourced R language, developing Shiny web application is of low cost and little technical demand for agronomists, researchers, or data analysts who have some experience with R. But it also suffers from the disadvantage of slow computation. The advantage is fast implementation and flexible visualization. The design, development, and deployment of ISOFAST were achievable for R users without past experience of web development nor knowledge of HTML or JavaScript. During developing the ISOFAST platform, both Proof of concept, and Production, were intermingled. We developed concepts and produced on-line summaries for different trial categories depending on whether we had sufficient number of trials locations and years of data collection. Visualization tools in R, such as `ggplot2`<sup>30</sup> and `plotly`<sup>21</sup> increase the possibility of designing creative and domain-specific graphs (*e.g.*, forest plots) for visualizing raw data and Bayesian analysis outputs. Commercial visualization tools may be able to produce more mobile-friendly and more responsive applications, but they are usually costly and do not necessarily support some special chart types such as forest plots. Given a limited budget and a low requirement of software engineering from ISA, ISOFAST fulfills its goal of a user-friendly visualization tool for on-farm strip trials made for farmers and agronomists. Although little software engineering is needed for developing a practical web tool for agronomic applications, we recommend following the best practices in software development, including optimizing the data structure and the code for stability due to large traffic volume and for robustness due to potential programming errors. Those principles are extremely important, especially when the final web tool needs to be maintained over

many years. A testing stage asking for feedback from end-users is also helpful before launching the tool to the public. Programming bugs or large revisions after launching and putting into use the tool might result in a bad user experience. ISOFAST is being maintained by the ISA staff and will be continuously modified to add new trial categories, address geographic differences across the state, improve speed and clarity of the online summaries.

### **3.6 Conclusion**

The ISOFAST web application is currently available to the public and hosted by the Iowa Soybean Association at <https://analytics.iasoybeans.com/cool-apps/ISOFAST/>. This web-application represents the first attempt to share on-farm research data analyses to a broad audience, including farmers and agronomists with different backgrounds and interests. In fact, to our knowledge, there is no visualization tool that provides multi-site and multi-year analyses with the aim of helping decision-making. Multiple functionalities and interactive features allow users with different interests to perform self-assisted assessment and exploration of study results to assist in education and decision making. We hope that the tool will be used by farmers, agronomists, and researchers who can provide feedback and suggestions for improving its functionality. We believe that practitioners, such as farmers without data analysis or statistical background, can benefit from exploring and understanding large datasets using ISOFAST. From a pedagogical point of view, ISOFAST can be a useful tool for teaching agronomy, data science and applied statistics to students. A collaboration between farmers, agronomists, and visualization researchers is promising to benefit on-farm network research as a discipline and improve decision-making for future growing seasons. ISOFAST will continue to evolve as needed with the

addition of data from new or existing management practices and products and with upgrades to existing visualization technologies.

### 3.7 Acknowledgment

The authors thank Suzanne Fey and Eric Hare for their technical assistance programming ISOFAST. The combine harvester icon displayed in Figure 1 was made by [Freepik](#) from [www.flaticon.com](http://www.flaticon.com). This research was funded by the Iowa Soybean Association checkoff. The work of FM, AL and XL was also supported by the Plant Sciences Institute and Iowa Soybean Research Center at Iowa State University. The work of DM was supported by the French institute of convergence CLAND (16-CONV-0003). AL and XL contributed equally to the writing of this paper.

### 3.8 References

1. Laurent A, Kyveryga P, Makowski D, Miguez F. A Framework for Visualization and Analysis of Agronomic Field Trials from On-Farm Research Networks. *Agronomy Journal*. 2019;111(5):1-12. doi:10.2134/agronj2019.02.0135
2. Matthewson M, Fery M, Powell M. Creating Farmer Networks: A Toolkit for Promoting Vibrant Farm Communities. Published online 2013. <https://www.sare.org/content/download/73146/1058808/file/pnw638.pdf>
3. Kyveryga PM, Mueller TA, Mueller DS, Shannon DK, Clay DE, Kitchen NR. On-Farm Replicated Strip Trials. In: *ACSESS Publications*. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Inc.; 2018. doi:10.2134/precisionagbasics.2016.0096
4. Unwin A, Theus M, Hofmann H. *Graphics of Large Datasets : Visualizing a Million*. New York : Springer; 2006.
5. Young FW, Valero-Mora PM, Friendly M. *Visual Statistics. Seeing Data with Dynamic Interactive Graphics*. Vol 914. 1st ed. John Wiley & Sons, Incorporated; 2011.

6. Heer J, Shneiderman B. Interactive dynamics for visual analysis. *Communications of the ACM*. 2012;55(4):45. doi:10.1145/2133806.2133821
7. Kelleher C, Wagener T. Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software*. 2011;26(6):822-827. doi:10.1016/j.envsoft.2010.12.006
8. RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc.; 2015. <http://www.rstudio.com/>
9. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. *Shiny: Web Application Framework for R*; 2016. <https://CRAN.R-project.org/package=shiny>
10. Kharel T, Swink S, Youngerman C, et al. Processing/Cleaning Corn Silage and Grain Yield Monitor Data for Standardized Yield Maps across Farms, Fields, and Year. Published online 2018. [http://nmsp.cals.cornell.edu/publications/extension/ProtocolYieldMonitorDataProcessing2\\_8\\_2018.pdf](http://nmsp.cals.cornell.edu/publications/extension/ProtocolYieldMonitorDataProcessing2_8_2018.pdf)
11. Kyveryga PM, Mueller TA, Mueller DS. On-Farm Replicated Strip Trials. In: Kent Shannon D, Clay DE, Kitchen NR, eds. *ASA, CSSA, and SSSA Books*. American Society of Agronomy and Soil Science Society of America; 2018:189-207. doi:10.2134/precisionagbasics.2016.0096
12. *Guide to On-Farm Replicated Strip Trials*. Iowa Soybean Association; 2015:51. <https://www.iasoybeans.com/upl/downloads/publications/a-farmer-s-guide-to-replicated-strip-trials-2016.pdf>
13. USDA. Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database. <https://sdmdataaccess.sc.egov.usda.gov>
14. Iowa Environmental Mesonet. Iowa State University, Iowa Environmental Mesonet. Published 2018. <https://mesonet.agron.iastate.edu/>
15. Hadfield JD. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software*. 2010;33(2). doi:10.18637/jss.v033.i02
16. Sedgwick P. How to read a forest plot in a meta-analysis: *BMJ*. Published online July 24, 2015:h4028. doi:10.1136/bmj.h4028
17. Cleveland WS. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*. 1979;74(368):829-836. doi:10.2307/2286407

18. Cleveland WS. LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression. *The American Statistician*. 1981;35(1):54. doi:10.2307/2683591
19. Winston C, Borges Ribeiro B. *Shinydashboard: Create Dashboards with "Shiny."*; 2018. <https://CRAN.R-project.org/package=shinydashboard>
20. Cheng J, Karambelkar B, Xie Y. *Leaflet: Create Interactive Web Maps with the JavaScript "Leaflet."*; 2018. <https://CRAN.R-project.org/package=leaflet>
21. Sievert C, Parmer C, Hocking T, et al. *Plotly: Create Interactive Web Graphics via "Plotly.js"*; 2017. <https://CRAN.R-project.org/package=plotly>
22. Allaire J, Xie Y, McPherson J, et al. *Rmarkdown: Dynamic Documents for R*; 2020. <https://rmarkdown.rstudio.com>
23. Blumenthal J, Baltensperger D, Cassman KG, Mason S, Pavlista A. Importance and effect of nitrogen on crop quality and health. In: *Nitrogen in the Environment: Sources, Problems, and Management*. 2nd ed. Elsevier; 2008.
24. Abebe Z, Feyisa H. Effects of Nitrogen Rates and Time of Application on Yield of Maize: Rainfall Variability Influenced Time of N Application. *International Journal of Agronomy*. 2017;2017:1-10. doi:10.1155/2017/1545280
25. Gehl RJ, Schmidt JP, Maddux LD, Gordon WB. Corn Yield Response to Nitrogen Rate and Timing in Sandy Irrigated Soils. *Agron J*. 2005;97(4):1230-1238. doi:10.2134/agronj2004.0303
26. Cleveland WS, McGill R. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*. 1984;79(387):531-554. doi:10.1080/01621459.1984.10478080
27. Wortmann C. Effects on Profitability and Nitrogen Loss from Reducing N Rate for Corn. CropWatch. Published 2019. <https://cropwatch.unl.edu/2019/effects-reduced-n-rate-corn-profitability-and-nitrogen-losses>
28. Cheng J. *Supplement to Shiny in Production.*; 2019.
29. Chang W, Luraschi J, Mastny T. *Profvis: Interactive Visualizations for Profiling R Code.*; 2019. <https://CRAN.R-project.org/package=profvis>
30. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Second edition. Springer; 2016.

### 3.9 Supplemental Materials



Figure S1: Hierarchical structure of the management practices tested for each of the two categories (soybean or maize). Active ingredients tested for each brand of fungicide or insecticide are given in brackets.

Table S1: Template visuals for each sub-menu item available in ISOFAST. Names of sub-menu items and the corresponding variables are described on the left column. An example of visual is given on the right column. Yield difference is expressed in bushel/acre; yield change is expressed in percentage, rainfall is expressed in inches, growing degree days (GDD) is expressed in °F.

<p>What was done</p>	<div data-bbox="874 405 1350 562"> <p><b>What was done</b></p> <p>The objective of this on-farm study was to identify when and where farmers can reduce their normal Nitrogen fertilizer application rates to corn by about one third, or roughly 50 lbs/acre, without economic loss.</p> <p>Sixty on-farm replicated strip trials were conducted between 2006 and 2010. Farmers compared their normal N rate to a rate reduced by 50 lb N/acre.</p> <p>The average normal N rate in the 16 trials in corn after corn was 175 lb/acre, (150-200 lb/acre), and the average normal minus 50 N rate was 125 lb/acre. The average normal N rate in the 44 trials in corn after soybean was 145 lb/acre, (120-195 lb/acre), and the average normal minus 50 N application rate was 95 lb/acre.</p> </div>
<p>Trials location</p>	<div data-bbox="874 600 1350 931"> <p><b>Trial Locations</b></p> <p><b>60</b> Trials from 2006 to 2010</p> <p><b>Information</b></p> <p>The map of Iowa shows county lines (black) and landform regions (gray) to clarify trial locations and general topographical attributes. The bulk of Iowa landmass falls primarily into 4 regions: Northwest Plains, Des Moines Loess, and Iowa Surface, from west to east across the northern half of the state, and Southern Iowa Drift Plain across the southern half of the state.</p> </div>
<p>Weather conditions (per trial and 15-year regional average)</p> <ul style="list-style-type: none"> <li>• Monthly Rainfall</li> <li>• Growing Degree Days</li> </ul>	<div data-bbox="874 969 1350 1294"> <p><b>Select Year:</b> 2006 2007 2010 2008 2009</p> <p><b>Select Reference Region:</b> CN NE NW SE SW</p> <p><b>Monthly Rainfall Information</b></p> <p>The reference rainfall (indicated by the gray bar) represents the 15-year mean monthly rainfall for the selected region of the state. Regions are based on Iowa landform areas and historical weather patterns. Source: Iowa Mesonet.</p> </div> <div data-bbox="874 1317 1350 1630"> <p><b>Select Year:</b> 2006 2007 2010 2008 2009</p> <p><b>Select Reference Station:</b> Central North South</p> <p><b>Growing Degree Days Information</b></p> <p>Growing degree days (GDD) is a measure of heat accumulation used to predict plant and pest development rates, including crop maturity.</p> </div>

Table S1: Continued

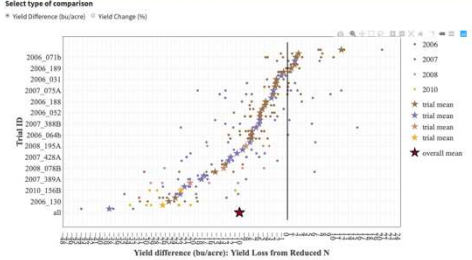
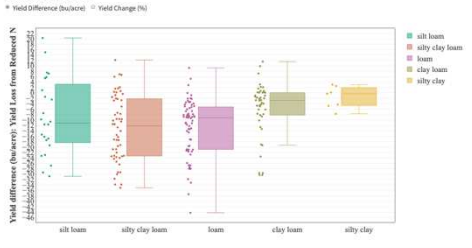
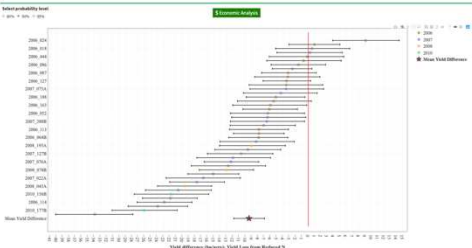
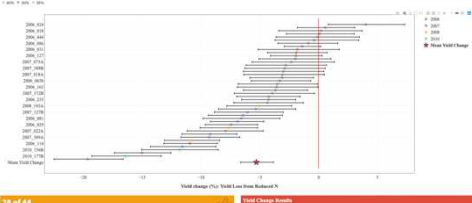
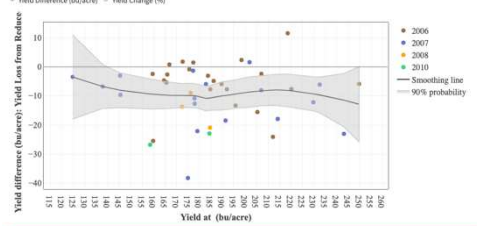
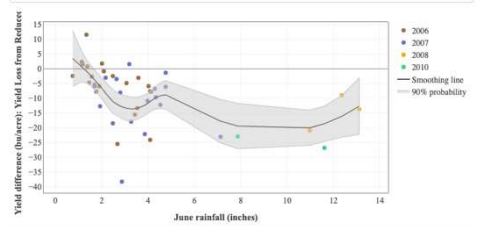
<p><b>Yield Comparison</b></p> <ul style="list-style-type: none"> <li>• Yield difference</li> <li>• Yield change</li> <li>• Soil texture</li> <li>• Planting date</li> </ul>	 <p><b>Information</b></p> <p>Yield difference is the yield of the reduced Nitrogen treatments minus the yield of the paired Normal Nitrogen treatments. The stars represent average yield differences for individual trials (bu/acre). Each dot represents one replicate within a trial.</p> <p>Yield change is the percentage of yield gain (if positive) or yield loss (if negative) using the treatment. The stars represent average yield changes for individual trials (%). Each dot represents one replicate within a trial.</p>
<p><b>Trial Summaries &amp; Economics</b></p> <ul style="list-style-type: none"> <li>• Yield difference</li> <li>• Yield change</li> </ul>	 <p><b>Information</b></p> <p>There is no clear trend in yield loss from the effect of soil texture in corn after soybean.</p>  <p><b>30 of 44</b></p> <p><b>Key result</b></p> <p>Mean of Yield Loss from Reduced N is 11.1 bu/acre. Mean of Yield Difference is 1.1 bu/acre. Mean of Yield Change is 1.1%.</p>  <p><b>28 of 44</b></p> <p><b>Yield Change Benefit</b></p> <p>Mean of Yield Change is 1.1%. Mean of Yield Loss from Reduced N is 11.1 bu/acre. Mean of Yield Difference is 1.1 bu/acre.</p>



Table S1: Continued

<p><b>Aggregate Summaries</b></p> <ul style="list-style-type: none"> <li>• Yield difference vs yield at the control</li> <li>• Yield change vs yield at the control</li> <li>• Yield at control vs yield difference</li> <li>• Yield difference vs rainfall</li> <li>• Yield difference vs cumulative GDD</li> <li>• Yield change vs cumulative GDD</li> </ul>	<p><b>Select type of comparison</b></p> <p>• Yield Difference (bu/acre) • Yield Change (%)</p>  <p><b>Information</b></p> <p>This graph shows the relationship between the Yield at the Normal N rate and the mean yield response to Normal minus 50 lbs N nitrogen. The flatter the smoother line, the weaker the relationship. Yield loss from reduced N does not correlate to Yield and the Normal N rate.</p>
<p><b>Conclusions</b></p>	<p><b>Select month</b></p> <p>June</p>  <p><b>Information</b></p> <p>There is a clear trend of greater yield loss with higher May through June of about 1.0 bu/acre per inch of rain.</p> <p><b>Conclusions</b></p> <ol style="list-style-type: none"> <li>The objective of this on-farm study was to identify when and where farmers can reduce their normal N fertilizer application rates on corn without economic loss. Farmers compared their normal rate to a rate lowered by about one third, or roughly 50 lb/acre.</li> <li>In corn after corn, the average 'Normal N rate' was 175 lb/acre, (120-200 lb/acre), and the average 'Normal minus 50 N rate' was 125 lb/acre. In corn after soybeans, the average 'Normal N rate' was 145 lb/acre, (120-195 lb/acre), and the average 'Normal minus 50 N rate' was 95 lb/acre.</li> <li>Between 2006 and 2010, 13 of the 30 trials in corn after corn and 30 of the 44 trials in corn following soybeans showed a significant effect on yield loss from the reduction of Normal N rates by 50 lb/acre.</li> <li>In corn after corn, the average yield loss from 'normal minus 50' lb N/acre was 7.5 bu/acre, with a 90% probability of being between 5.7 to 9.3 bu/acre.</li> <li>In corn after soybeans, the average yield loss from normal minus 50 lb N/acre N/acre, with a 90% probability of being between 3.8 to 8.5 bu/acre.</li> <li>With corn at \$3.50 per bushel and Commercial N application at \$17.50 per acre, the break-even yield response is a loss of only 5.9 bu/acre. In corn following corn, the probability to reach the break-even with 50 lb/acre reduced N is only 7%, and the average expected loss is \$8.50 per acre. In corn following soybeans, it is very unlikely to recover the cost of yield loss from a reduction of 50 lb N/acre, and the average expected loss is \$15.44 per acre. *Use the Economic Analysis button in the application to adjust for the current market price and your application costs.</li> <li>In corn after corn, yield loss from reduced N tended to be about 0.7 bu/acre greater with each additional inch of May through June rainfall.</li> <li>In corn after corn, overall, N rate did not affect yield loss, however SideDress applications tended to have higher yield loss than other combinations of form and timing.</li> <li>In corn after soybeans, yield loss from reduced N tended to be greater by about 1.0 bu/acre with each additional inch of May-June rainfall, and the yield loss tended to be lower by about 0.1 bu/acre for each additional pound of N.</li> <li>The decision to reduce N application by 50 lb N/acre would depend on both on normal N rates and rainfall. For corn after corn, application timing also needs to be considered.</li> </ol> <p><b>Include Sections</b></p> <ul style="list-style-type: none"> <li>What was done</li> <li>Trial Locations</li> <li>Weather Conditions</li> <li>Yield Comparisons</li> <li>Trial Summaries &amp; Economics</li> <li>Aggregate Summaries</li> <li>Conclusions</li> </ul>

## CHAPTER 4. GOING BEYOND MEAN EFFECT SIZE; PRESENTING PREDICTION INTERVALS FOR ON-FARM NETWORK TRIAL ANALYSES

Modified from a paper published in *European Journal of Agronomy*,  
special issue “Evidence synthesis in agronomy”

Anabelle Laurent<sup>1</sup>, Fernando Miguez<sup>1</sup>, Peter Kyveryga<sup>2</sup>, David Makowski<sup>3</sup>

<sup>1</sup> Department of Agronomy, Iowa State University, Ames, IA

<sup>2</sup> Analytics, Iowa Soybean Association, Ankeny, IA

<sup>3</sup> INRAE, Centre Île-de-France, Paris, France

### 4.1 Abstract

The aim of on-farm research is to identify and test new technology, product or management practice (e.g. more efficient seeding rate, enhanced row spacing, better disease management treatment, etc.) suited to local conditions by comparing it to a standard farmer practice across several farmers' fields. Typically, each trial includes two treatments (new practice vs. standard control practice) replicated at least three times in each field. The statistical analysis of yield data collected in such trials provides growers with useful information about the effectiveness of the tested farming practice on crop productivity and its uncertainty. We used a random-effects model to i) estimate the performance of a treatment compared to a control in individual trials, ii) estimate the overall mean yield response across all trials, iii) compute prediction intervals describing a range of plausible yield response for a new (out-of-sample) field at the trial level, and iv) compute the probability that the tested management practice will be ineffective in a new field. We used frequentist (classical) and Bayesian approaches for data collected in 26 on-

farm trial categories managed by the Iowa Soybean Association. Depending on the level of between-trial variability, we found that prediction intervals were 2.2 to 12.1 times larger than confidence intervals for the estimated mean yield responses for all tested management practices. We conclude that prediction intervals should be systematically reported to provide additional information about future trials or experiments with associated uncertainties. Nevertheless, prediction intervals should be interpreted with caution when the between-trial variance is small. Using prediction intervals and, when appropriate, the probability of ineffective treatment will prevent farmers from overoptimistic expectations that a significant effect at the overall population level will lead with high certainty to a yield gain on their own farms.

## **4.2 Introduction**

There is an increasing interest in conducting research directly on commercial farms under local conditions. On-farm research networks (OFRNs) allow evaluating the efficacy of new products, managements and technologies under field conditions with all the inherent variation present in a specific cropping system (Johnston et al., 2003). As on-farm experiments are repeated across locations and growing seasons, the scope of the statistical inference about the treatment effect is expanded over time and space (Moore and Dixon, 2015) with the goal of improving confidence in the selection of better agronomic management practices in future growing seasons and new environments (Crossa, 1990). To successfully conduct on-farm trials and improve the decision making process, all stakeholders need to work together to provide support for execution, data collection and analysis (Nelson et al., 2016).

In order to achieve valid results and reliable recommendations, OFRN must adhere, as best as possible, to the standard principles of experimental design (Casler, 2015). The “replicated strip” is a common and convenient experimental design where two treatments, a new management practice (e.g. crop management, crop protection, plant nutrition, and plant growth regulator), and a local or farmer ‘control’, are applied in several strips. Often, treatment randomization is omitted for greater convenience during technical operations such as treatment applications and crop harvest. Usually, the “farmer’s control” corresponds to a management practice commonly used by farmers in the considered agricultural area or can simply be left untreated. A set of strips located in a given field in a given year is called a trial. A pair of strips (the new management practice and the control) constitutes a replicate and, ideally, each trial should have a minimum of three (more is better) replicates for capturing within-field variability (Iowa Soybean Association, 2018). For example, in Iowa, an OFRN was conducted for 4 years to test the effect of lower seeding density on soybean yield compared to the seed density commonly used by farmers (Laurent et al., 2019). This OFRN included 21 trials with an average of 6 replicates in each trial. It showed that the low seeding density increased yield by 0.6% but that the yield gain was not statistically significant. This would imply that it can be economically beneficial to reduce the seeding rate due to lower expenses related to the purchase of seeds associated with similar yield levels. Another OFRN was conducted in Iowa for 9 years to estimate the effect of a foliar fungicide (*pyraclostrobin*) application on maize yield compared to an untreated control. It included 143 trials with an average of 5 replicates in each trial and showed that the foliar fungicide significantly increased yield by 2.3%. The implication of

these results is that foliar fungicide application could be profitable when maize prices are relatively high compared to fungicide costs.

Typical OFRNs have two main levels of replication: the strips within a trial, capturing within-trial variability and the trials representing different environments and characterizing between-trial variability across a specific region. Since trials are located in different areas characterized by different environmental conditions and farmer management practices, a substantial degree of between-trial variability is expected. Previous management, soil properties, weather conditions, farmers' individual choices, such as seed variety, will vary across sites and years and produce different production and environmental outcomes. The within-trial variability is a result of heterogeneous soil, field topography within a given trial and, also, the ability of farmers to repeat the established treatments in the same manner across different strips. Treatment replicates are essential to quantify the 'background noise' (or error) in a study and thus evaluate the effectiveness of the proposed new product or practice. It is, thus, important to repeat the treatment and control several times in each trial.

The analysis of data generated by OFRNs can follow similar methodologies to those developed in the field of meta-analysis for estimating the mean effect sizes of treatments (Borenstein et al., 2009; Gates, 2002; Philibert et al., 2012). In OFRNs, the "effect size" is defined as the yield ratio (a ratio of yield obtained with a new management practice to yield in the control) or the yield difference (yield obtained with a new management practice minus yield in the control) (Laurent et al., 2019; Nakagawa et al., 2017).

The random-effects model is a useful statistical tool to synthesize data provided by OFRNs (Laurent et al. 2019). This model, commonly used for meta-analysis (Gurevitch et

al., 2018), accounts for both within and between-trial variability through the use of random trial and random replicate (nested within-trial) effects (Partlett and Riley, 2017; Veroniki et al., 2019). As a consequence, each trial is assumed to have a different 'true' effect size (*i.e.* deviation from the overall mean) due to heterogeneity (between-trial variance) across trials and that the observed effect sizes differ from the true values due to within-trial variability (Borenstein et al., 2017; Riley et al., 2011). For example, the yield response from a foliar fungicide application will differ among trial locations as the disease pressure (Wise and Mueller, 2011) and crop susceptibility, environmental conditions (rainfall, humidity etc.), and physiological responses from fungicides in the absence of disease vary (Kyveryga et al., 2013; Mallowa et al., 2015). In that case, it is not realistic to assume a homogeneous fungicide effect for all trials, and a random effect model may offer a better description of the data distribution than a fixed-effect model assuming a constant fungicide effect. In a preliminary analysis, the significance of the between-trial variance could be tested using a variety of methods such as the Cochran's Q test (Borenstein et al., 2009).

With the random-effects model, the overall effect of the tested management practice is summarized by estimating a mean effect size measuring its average effect, or population-level effect, across all trials. The uncertainty about the mean effect size parameter is usually described by computing a confidence interval (see table 1). This approach has already been used to analyze OFRNs (Laurent et al., 2019). In this case, mean effect size estimates and their confidence intervals may provide farmers with an incomplete or too optimistic range of plausible effects of tested management practices in future environments (e.g. next growing season).

In addition to confidence intervals, random-effects models can be used to derive prediction intervals (see table 1 ), which are rarely used in agricultural research (Higgins et al., 2009; Partlett and Riley, 2017). In OFRNs, a prediction interval can be interpreted as a range of effects (e.g., yield response) of a given management practice compared to a control for a new trial similar to those included in the studied sample (Higgins et al., 2009; IntHout et al., 2016; Partlett and Riley, 2017). When there is a large between-trial variability, the effect of the treatment in an new trial might differ considerably from the overall mean effect (Higgins et al., 2009). Thus, prediction intervals can provide farmers, agronomists and other stakeholders with a more complete picture of the range of possible effects of a given management practice for a new field, not initially included in the sample of trials. To our knowledge, the usefulness of prediction interval has been recently emphasized in medical sciences (Graham and Moran, 2012; Guddat et al., 2012; Higgins et al., 2009; IntHout et al., 2016; Partlett and Riley, 2017; Riley et al., 2011; Veroniki et al., 2019), but prediction intervals are not currently reported in agricultural research, except in a few examples, such as in Kyveryga et al. (2013) where predictive probability distributions of yield loss from reduced nitrogen applications in maize are reported.

Parameters of random-effects model can be estimated using Bayesian or frequentist (classical) statistical approaches. In the frequentist approach, model parameters are viewed as fixed, and data are random. In the Bayesian approach, all unknown model parameters can be viewed as random variables and the results of the inference are expressed through a probability distribution (called posterior distribution) of these parameters given the observed data and prior knowledge about the parameters (Wakefield, 2013). The Bayesian approach has some advantages over frequentist statistical

analyses. One of these is the use of prior information derived from literature or expert knowledge and the incorporation of full uncertainty in all parameters in the posterior distribution (Higgins et al., 2009; Partlett and Riley, 2017). A rigorous uncertainty analysis is important when the goal is to make a statement about a potential outcome in a “new” trial (i.e. a field similar to those included in the studied sample but not observed yet) which was not part of the trials included in the OFRN.

Here, our objective is to show that prediction intervals and probabilities that a treatment will be ineffective provide relevant information when explaining results of on-farm trials and making management decisions. We show that prediction intervals are useful to assess the range of possible outcomes of a new management practice, especially when the between-trial variability is large. Our study provides an operational workflow to fit a random-effects model and compare frequentist and Bayesian prediction intervals and probability of an ineffective treatment. Both approaches are compared across a large range of OFRNs, testing various agricultural practices related to nutrient management, sowing dates and densities, and crop protection products.



Table 1: key definition of statistical terms

<b>Statistical term</b>	<b>Specific definition</b>
<b>Confidence interval</b>	Classical frequentist interpretation. Range of values within which an unobserved fixed parameter value would fall with a particular probability. For example, in repeated sampling (same sample size, large number of repetitions, and same conditions) a 95% confidence interval is expected to include the true unknown parameter 95% of the time (Krzywinski and Altman, 2013; Sedgwick, 2014).
<b>Credible interval</b>	Bayesian interpretation. A range of values within which an unobserved parameter value falls with a particular probability given the observed data and prior assumptions. Contrary to the confidence interval, the parameter is treated as a random variable, described by a specific distribution. These intervals are derived from summaries of the posterior distribution of the parameter such as the 0.025 and the 0.975 quantiles (Gelman et al., 2014).
<b>Prediction interval</b>	A range of values which is expected to include a future study (similar to the previously observed ones) with a specify probability level (Higgins et al., 2009). In our example, a future study refers to a new on-farm trial which is not part of the network but conducted under similar conditions. This interval will include additional variability to account for unobserved situations.

## 4.3 Materials and Methods

### 4.3.1 Data Description and Experimental Design

We used data collected in 27 different On-Farm Research Network trial categories managed by the Iowa Soybean Association. Each trial category compared the effect of an alternative management practice to a standard practice on maize or soybean yields and included anywhere from 6 to 206 on-farm individual trials. The tested new management practices were: fungicide and insecticide applications, lower plant densities, or comparisons of timing of fertilizer applications. Each individual on-farm trial was based on a replicated strip trial design including anywhere from 3 to 39 replicates, with a median of

5 replicates (Figure 1). Each replicate was composed of two strips corresponding to the tested management practice and the standard common (or farmer-chosen) practice. The widths of the individual strips were adapted to the size of the application equipment and ranged from 4.6 to 27.4 m. The length of the strips was adjusted to the field size. For example, a typical field in Central Iowa is about 32 ha, which would have a dimension of about 460 x 700 m. Yield data were collected using a combine equipped with a yield monitor and GPS in each strip.

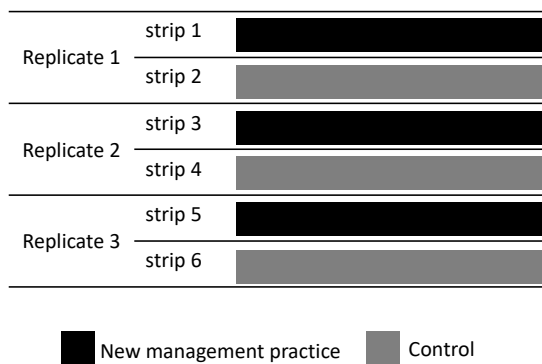


Figure 1: Schematic illustration of an on-farm trial with replicated strips for two treatments (here the new management practice is in black and the control is in grey) and three replicates within a farmer's field. The OFRNs considered here include between 6 and 206 individual trials.

#### 4.3.2 Data Analysis

We used a random-effects model (Laurent et al. 2019) to: *i*) estimate the performance of a treatment compared to a control in individual trials, *ii*) estimate the overall mean yield response across all trials, *iii*) compute prediction intervals for estimating plausible yield responses for a new field not included in the sample of studied trials but following the same model, and *iv*) compute the probability of ineffective treatment (i.e. absence of positive yield gain resulting from the new management practice).

Our model includes yield ratio as a response variable, but it can be easily adapted to other response variables, including a yield difference expressed in absolute units. The yield ratio was log-transformed to normalize the distribution as is typically done in meta-analysis. The yield difference is unit-dependent and can be easily expressed as the economic gain or loss while the yield ratio is unitless and might be less dependent on other factors. As long as an event, factor or issue affects the yield of the control and treatment to the same extent, it will not have an impact on the effect size whether the response ratio or the difference is used. Trials were represented by site-years as they were rarely repeated at the same location over time.

The statistical model is defined as follows:

$$\ln(R_{ij}) = \mu + \alpha_i + \varepsilon_{ij} \quad (1)$$

where  $\ln(R_{ij})$  represents the logarithm of the  $j^{\text{th}}$  yield ratio in the  $i^{\text{th}}$  trial ( $R_{ij}$ =yield in the new management practice treatment of the  $i^{\text{th}}$  trial and  $j^{\text{th}}$  replicate divided by yield in the control of the same trial and replicate);  $\mu$  represents the true mean log yield ratio in the considered population;  $\alpha_i$  represents the random effect of the  $i^{\text{th}}$  trial, and  $\varepsilon_{ij}$  represents the residual error. The random variables  $\alpha_i$  and  $\varepsilon_{ij}$  are assumed to follow independent Gaussian distributions with mean zero and constant variances,  $\alpha_i \sim N(0, \sigma_\alpha^2)$  and  $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$  (the character  $\sim$  means “is distributed as”).

Using the yield difference as effect size, the statistical model (1) becomes:

$$D_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (2)$$

where  $D_{ij}$  represents the  $j^{\text{th}}$  yield difference in the  $i^{\text{th}}$  trial ( $D_{ij}$ = yield in the new management practice treatment of the  $i^{\text{th}}$  trial and  $j^{\text{th}}$  replicate minus yield in the control of the same trial

and replicate);  $\mu$  represents the true mean yield difference in the considered population, and the interpretations of other parameters remain unchanged (see model (1) described above).

Models (1) and (2) include three parameters,  $\mu$ ,  $\sigma_\alpha^2$ , and  $\sigma_\varepsilon^2$ . The true value of the effect size in the  $i^{\text{th}}$  trial (individual log yield ratio) can be represented by  $\theta_i = \mu + \alpha_i$ . The residual  $\varepsilon_{ij}$  represents the deviation of the  $j^{\text{th}}$  replicate from  $\theta_i$ , and the value of  $\alpha_i$  measures the deviation of  $\theta_i$  from the overall mean  $\mu$  (Figure 2). The variance  $\sigma_\varepsilon^2$  quantifies the variability of  $\varepsilon_{ij}$  within trial (i.e., within field or between replicate variability) and  $\sigma_\alpha^2$  quantifies the variability of  $\alpha_i$  across trials (Figure 2). This model can be easily expanded to handle different within-trial variances, i.e. unequal  $\sigma_{\varepsilon_i}^2$  across trials and  $\varepsilon_{ij} \sim N(0, \sigma_{\varepsilon_i}^2)$ .

The fraction of the total variance of the effect size that is due to between-trial heterogeneity can be defined as the relative heterogeneity ratio (RHR) as follow:

$$\text{Relative Heterogeneity Ratio} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}$$

RHR is equal to the ratio of the between-trial variance to the total variance (sum of the between-trial and within-trial variances), and ranges from 0 to 1. This is also known as the intra-class correlation, but here we prefer this term as it relates more closely to our research questions. The hierarchical model described above was fitted to the data of each network using both Bayesian and frequentist methods.

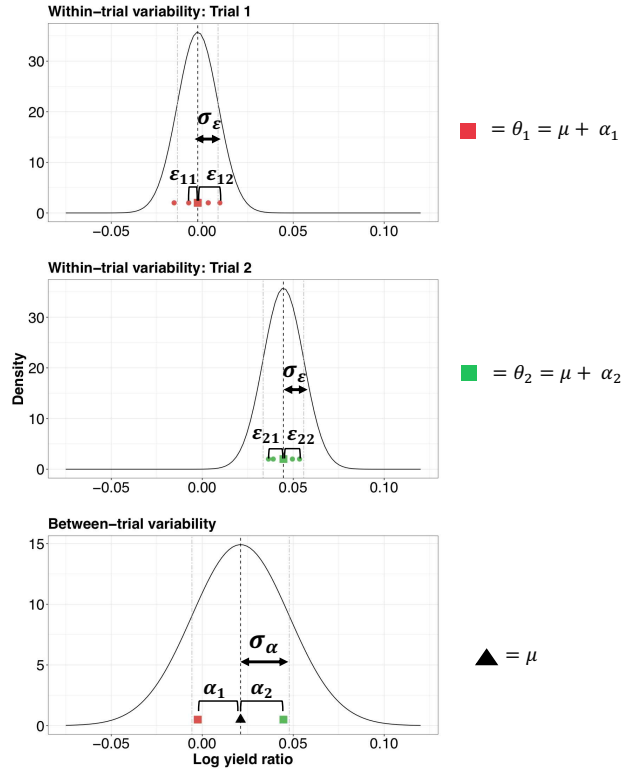


Figure 2: Graphical description of the model (1). Dots represent measured yield ratios in replicates. Trials 1 and 2 are shown in red and green, respectively. The colored squares represent  $\theta_1$  and  $\theta_2$ , the true expected values of log yield ratio in the corresponding trials based on the model (1). The black triangle represents the true mean log yield ratio  $\mu$  in the considered or target population of trials. Parameters  $\sigma_\alpha$  and  $\sigma_\epsilon$  represent the between-trial standard deviation and the within-trial standard deviations, respectively. The residual  $\epsilon_{ij}$  is the difference between the yield ratio measurement in replicate  $j$  trial  $i$  and the true log yield ratio  $\theta_i$  in the trial  $i$ .

#### 4.3.2.1 Frequentist Method

The model (1) was fitted using a frequentist approach (restricted maximum likelihood method) using the lmer function of the R package lme4 (Bates et al., 2015; RStudio Team, 2015). The 95% confidence interval of the estimated value of  $\mu$  was computed from the output of lmer directly using the profile likelihood method. The 95% confidence interval of the individual estimated value of  $\theta_i = \mu + \alpha_i$  (i.e., of the effect size

obtained for each individual trial) was calculated for each trial by bootstrap (1,200 samples).

A frequentist prediction interval for  $\theta_{new}$  in a new trial was computed following the approach proposed by Higgins et al. (2009). The key assumption is that

$$\frac{\theta_{new} - \hat{\mu}}{\sqrt{\{\widehat{\sigma}_\alpha^2 + \widehat{SE}(\hat{\mu})^2\}}} \sim t_{k-2}^\alpha$$

Based on this assumption, an approximate  $100(1 - \alpha)\%$  prediction interval was defined as follows

$$\hat{\mu} \pm t_{k-2}^\alpha \sqrt{\{\widehat{\sigma}_\alpha^2 + \widehat{SE}(\hat{\mu})^2\}} \quad (3)$$

where  $t_{k-2}^\alpha$  is the  $100(1 - \alpha/2)\%$  percentile of the  $t$ -distribution with  $k-2$  degrees of freedom,  $k$  represents the number of trials,  $\widehat{\sigma}_\alpha^2$  represents the estimated between-trial variance and  $\widehat{SE}(\hat{\mu})^2$  is the variance of the estimated mean effect  $\hat{\mu}$ . The probability of ineffective treatment in a new trial  $P(\theta_{new} < 0)$  was computed from this distribution.

We present here a simple example illustrating the conceptual difference between a confidence interval and a prediction interval. We consider a study including  $n$  observed values of a specific effect size (e.g., log of a ratio of yield obtained with a new management practice to yield in the control). To keep this example simple, the standard deviation (denoted by  $\sigma$ ) of the observations is assumed to be known and the observations are assumed normally distributed. Let  $\hat{\mu}$  denote the average of the  $n$  observations. Then, the 95% confidence interval of the mean is defined by:

$$\hat{\mu} - 1.96 \times \sigma / \sqrt{n} , \hat{\mu} + 1.96 \times \sigma / \sqrt{n}.$$

This interval gives a range of values which is likely to contain the true mean. Its width decreases as the sample size increases (Figure 3). This is logical because the higher the number of observations, the more accurate the estimated mean value. Actually, if the sample size becomes very large, the width of the confidence interval will tend towards zero because the mean value will be then almost exactly known. However, this interval does not give information on the plausible range of values for a new observation. To obtain this information, it is necessary to compute a prediction interval. Here, the 95% prediction interval is expressed as

$$\hat{\mu} - 1.96 \times \left(1 + \frac{1}{\sqrt{n}}\right) \times \sigma, \hat{\mu} + 1.96 \times \left(1 + \frac{1}{\sqrt{n}}\right) \times \sigma.$$

The prediction interval is larger than the confidence interval as it includes an extra term. The differences between the two types of interval are clearly illustrated in Figure 3, for two sample sizes.

The confidence interval and the prediction interval are both centered on the average log yield ratio. Like the confidence interval, the width of the prediction interval also decreases as the sample size  $n$  increases. However, the width of the prediction interval does not tend towards zero when the sample size becomes very large but rather to

$$\hat{\mu} - 1.96 \times \sigma, \hat{\mu} + 1.96 \times \sigma.$$

This indicates that, even if the mean value is perfectly known, an uncertainty will remain about the possible value for a new observation, and this uncertainty will depend on the standard deviation  $\sigma$ .

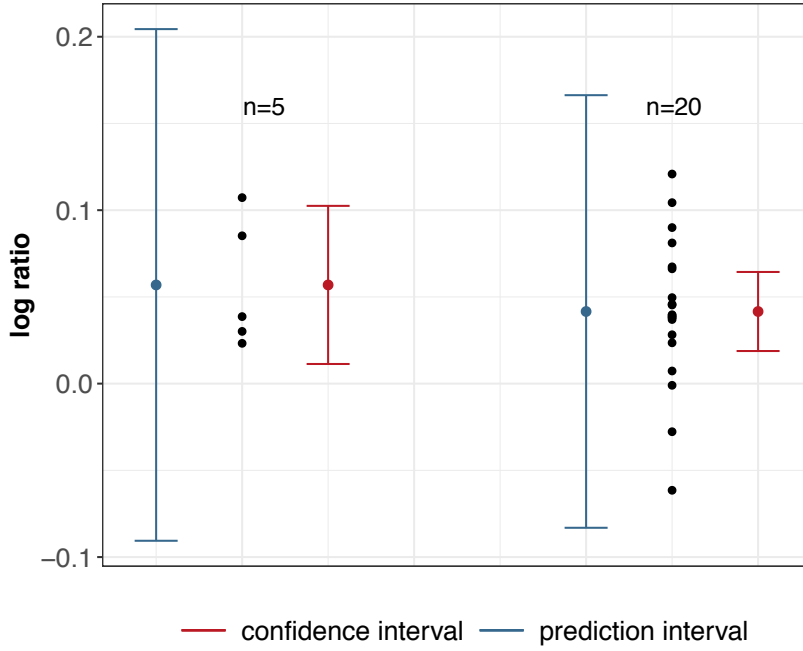


Figure 3: Confidence interval and prediction interval for two different sample sizes ( $n=5$  on the left and  $n=20$  on the right).  $\sigma = 0.052$ . For a sample size equal to 5,  $\hat{\mu} = 0.057$ . For a sample size equal to 20,  $\hat{\mu} = 0.042$ .

#### 4.3.2.2 Bayesian Method

Posterior distributions for  $\mu$ ,  $\theta_i$ ,  $\sigma_\alpha^2$  and  $\sigma_\varepsilon^2$  were computed using Markov chain Monte Carlo with the R package *MCMCglmm* (Hadfield, 2010; RStudio Team, 2015). We provided a diagram flow that describes all steps and associated R functions when estimating posterior distribution parameters (Figure 4).

Parametrization of the priors for the three parameters of the model (i.e.  $\mu$ ,  $\sigma_\alpha^2$  and  $\sigma_\varepsilon^2$ ) is specific to the R package *MCMCglmm*. A weakly informative prior for  $\mu$  was defined by a Gaussian distribution with zero mean and with a variance equal to 2. With a variance of 2, the log ratio can take either a high positive or a low negative value depending on the dataset. For the type of datasets that we were analyzing a value of 2 for the variance is



several orders of magnitude larger than the variances observed. This means that this prior is virtually uninformative, because the agronomic effect in these trials are usually very small, typically in the order of 0 to 20% (in the log response ratio scale a range of 0 to 0.2). The priors for the variance of the random effect,  $\sigma_{\alpha}^2$ , and the variance of the residual error,  $\sigma_{\varepsilon}^2$ , were defined as inverse Gamma distributions,  $\sigma_{\alpha}^2 \sim IG(0.001, 0.001)$  and  $\sigma_{\varepsilon}^2 \sim IG(0.001, 0.001)$ . These priors give more importance to small variance values. Hadfield (2016) mentioned that this distribution is frequently used for variance components. In addition, Chung et al. (2013) recommended the use of this prior distribution, especially when the number of studies (here, the number of site-years) is small. In *MCMCglmm*, these priors are specified using the variance of the inverse Gamma distribution (called V and fixed to 1) and the degree of belief parameter (called nu and fixed to 0.002).

The convergence of the MCMC chains (Toft et al., 2007) was checked with R package *coda* and *parallel* was used to support parallel computation (Plummer et al., 2006). The warmup iterations (or 'burnin', the number of iterations that will be discarded) was determined with the Gelman-Rubin diagnosis test (Raftery and Lewis, 1995) implemented with three chains. Finally, we ran MCMC chains for 200,000 iterations and discarded the first 20,000 iterations representing the initial warmup, and kept one sampled value every 10 in order to obtain independent samples. A total of 18,000 posterior samples were stored. Several outputs were computed from the samples representing posterior distributions, namely the posterior population mean of  $\mu$ , 95% credible interval (see table 1), 95% prediction interval, and the probability of ineffective treatment in the population and in a new trial. The 95% credible interval was directly computed from the samples from

the posterior distribution generated for  $\mu$ . This interval was used to quantify the uncertainty associated with the estimation of  $\mu$ . The probability  $P(\mu < 0)$  was computed to inform the chance of having a yield ratio of less than 1 for the population, called elsewhere in the manuscripts as the probability of ineffective treatment.

The 95% prediction interval was computed to estimate a likely effect in a ‘new’ trial not originally included in the trial network (IntHout et al., 2016). According to our model, the distribution of a mean yield ratio in a new trial is given by  $\theta_{new} \sim N(\mu, \sigma_{\alpha}^2)$ ; but we need to consider the uncertainty in the estimation of both the mean and variance. This distribution was used to generate new  $\theta$  values from the samples from the posterior distributions produced for  $\mu$  and  $\sigma_{\alpha}^2$ . The 95% prediction interval and the probability of ineffective treatment in a new trial, i.e.  $P(\theta_{new} < 0)$ , were then computed from the sample of values of  $\theta$ . Note that  $P(\theta_{new} < 0)$  measures the risk of yield loss using the new management compared to the standard technique for a new trial.

#### 4.3.2.3 Sensitivity Analysis of Priors for Between-Trial Variance

The between-trial variance  $\sigma_{\alpha}^2$  indicates how the yield response differs across trial averages. A high (low) between-trial variance reflects a high (low) heterogeneity across trials. This variance plays a key role when computing frequentist prediction intervals because it determines the lower and upper bounds of this interval (see eq.(2)). A large (small) between-trial variance will tend to produce a large (narrow) prediction interval. It is thus useful to analyze the sensitivity of the results to the chosen prior distribution for  $\sigma_{\alpha}^2$  and to check that the main conclusions of the analysis are robust to the prior characteristics.

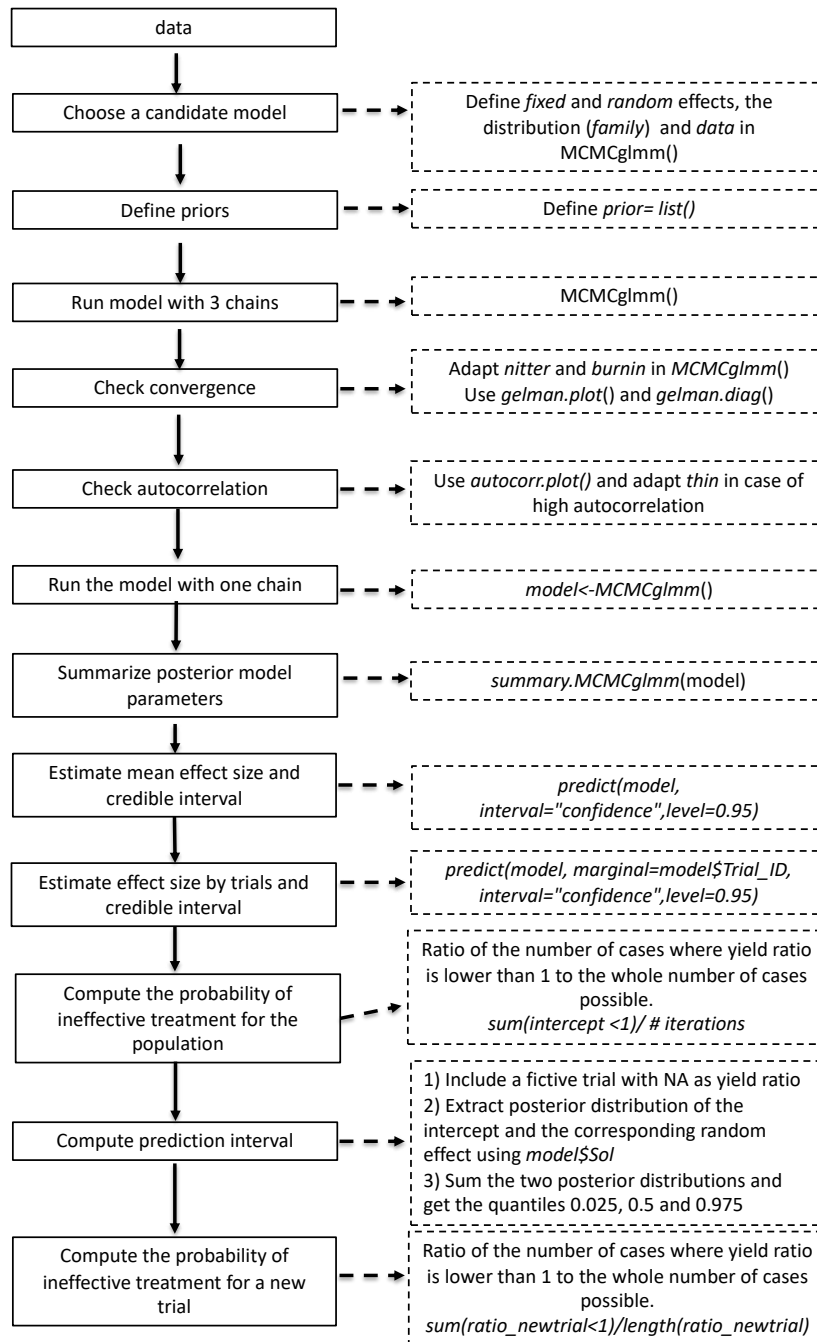


Figure 4: Workflow to fit the Bayesian hierarchical model (1) and estimation of credible and predictive intervals for population mean yield ratio parameters. Thick arrows represent major steps and dashed arrows represent the corresponding R code. Italic style represents specific R code. The full R code and a demo training dataset are archived at [https://github.com/femiguez/EJA\\_OFRN\\_prediction\\_intervals](https://github.com/femiguez/EJA_OFRN_prediction_intervals).

As commonly done in Bayesian statistics (Gelman, 2006; Lambert et al., 2005), the between-trial variance was assumed to follow an Inverse-Gamma (IG) distribution (Gelman, 2006; Lambert et al., 2005). Initially, we assumed that  $\sigma_{\alpha}^2 \sim IG(0.001, 0.001)$  and  $\sigma_{\varepsilon}^2 \sim IG(0.001, 0.001)$ . These priors are weakly informative and this parametrization is a popular choice (Chung et al., 2013; Hadfield, 2016; Lambert et al., 2005) because it does not strongly favor any particular value and specifies that the variance is positive and not equal to zero. We tested different variants of this inverse-gamma distribution to assess its impact on the prediction interval. More specifically, we examined two alternative weakly informative priors on the between-trial variance ( $\sigma_{\alpha}^2$ ) and three more informative priors. The five considered set of priors are defined below:

prior 1:  $\sigma_{\alpha}^2 \sim IG(0.001, 0.001)$  and  $\sigma_{\varepsilon}^2 \sim IG(0.001, 0.001)$

prior 2:  $\sigma_{\alpha}^2 \sim IG(0.01, 0.01)$  and  $\sigma_{\varepsilon}^2 \sim IG(0.001, 0.001)$

prior 3:  $\sigma_{\alpha}^2 \sim IG(0.1, 0.1)$  and  $\sigma_{\varepsilon}^2 \sim IG(0.001, 0.001)$

prior 4:  $\sigma_{\alpha}^2 \sim IG(0.5, 0.5)$  and  $\sigma_{\varepsilon}^2 \sim IG(0.001, 0.001)$

prior 5 :  $\sigma_{\alpha}^2 \sim IG(0.246, 0.001)$  and  $\sigma_{\varepsilon}^2 \sim IG(0.746, 0.0003)$

Priors 1 and 2 give more importance to small between-trial values (their modes are equal 0.00099 and 0.00990, respectively). Priors 3 to 4 give more importance to higher values of between-trial variance (their modes are equal to 0.09090 and 0.33333, respectively). For priors 1 to 4, the within-trial variance remains the same. The prior 5 was derived from the on-farm research network testing a foliar fungicide Headline (active ingredient: *pyraclostrobin*) on soybean by fitting the IG distribution to observed data. This network included 206 trials and its between-trial variance may be representative enough

for other networks in the same study area. Its mode is equal to 0.00079, which is relatively close to the mode of prior 1.

### 4.4 Results

#### 4.4.1 A Detailed Example

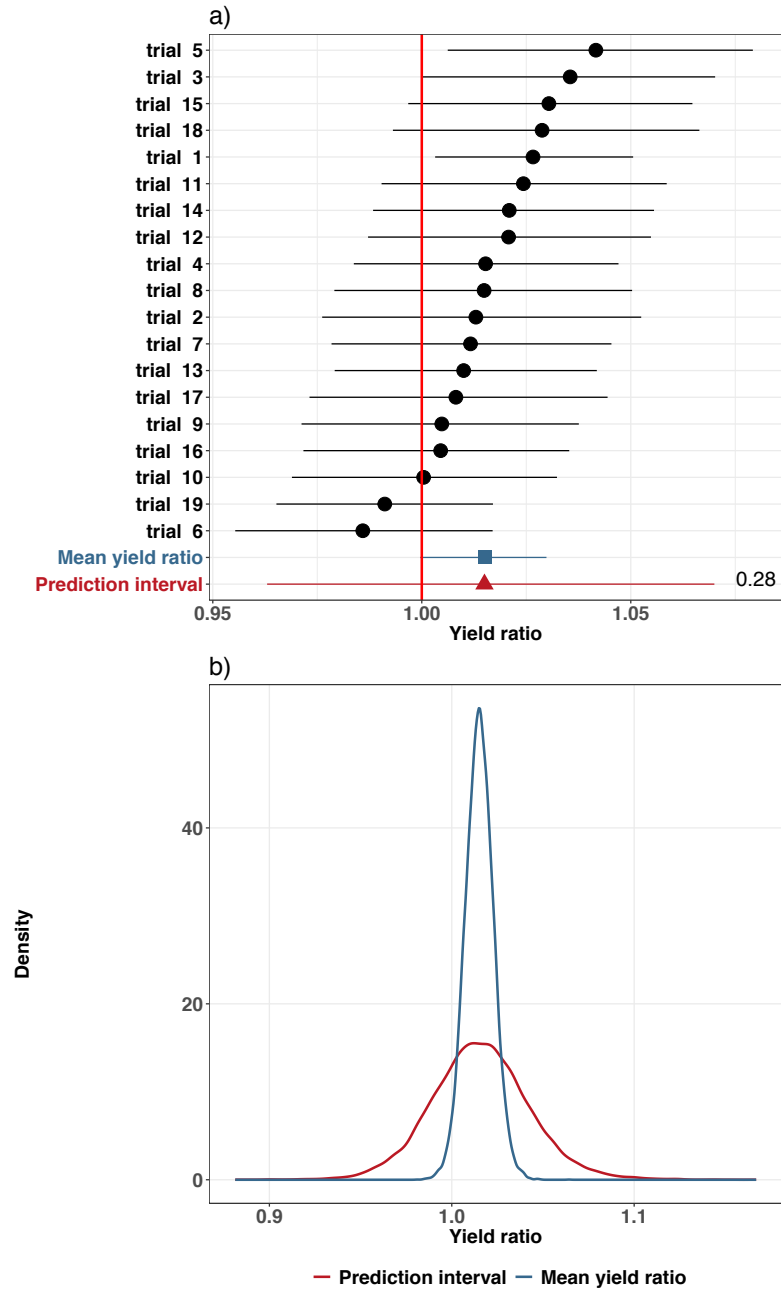


Figure 5: Analysis of an on-farm network of 19 trials assessing the effect of a higher seeding density (18 to 21%) compared to the seeding density commonly used by farmers. a)

Estimated individual effects with their 95% credible intervals (black), mean yield ratio with its 95% credible interval (blue), and 95% prediction interval (red). The number of replicates are displayed in brackets on the left side of the graphic. The numerical value at the bottom-right indicates the probability of ineffective treatment (probability of yield ratio less than 1) in a new trial; b) Posterior distribution mean yield ratio (in blue) and prediction distribution for a new trial (in red). Computations were done using the model (1) fitted using a Bayesian method.

The proposed approach is illustrated in Figure 5 for a specific on-farm network focusing on the effect of high seeding density (plus 74K seed/ha compared to the control) on soybean. The control was the seeding density commonly used by farmers in Iowa (i.e. between 346K-420K seed/ha. The mean yield ratio was equal to 1.014 with a 95% credible interval of [0.999; 1.030]. Three trials had a significant and positive estimated posterior individual yield ratio (Figure 5a). The predictive posterior probability distribution covers a much larger range of values than the posterior of the mean yield ratio (Figure 5b). Consequently, the 95% prediction interval is larger than the credible interval and ranges from 0.963 to 1.070 (Figure 5a). Based on the prediction interval, there is a 28% chance that a high seeding density on soybean will not increase yield compared to the seeding density commonly used in a new trial. Yield reductions from higher seeding density could be caused by soybean lodging and increase plant competition for water, nutrients, and light. In addition, high seeding density could create a more favorable environment for some diseases, such as white mold, due to the increased humidity within the canopy (Kolkman and Kelly, 2002). Under drought conditions, high seeding density is excessive relative to water availability and consequently, soybean yields are affected under such adverse conditions (Carciochi et al., 2019).

The 95% credible (confidence, with the frequentist approach) and prediction intervals illustrated here were computed for the 26 different OFRN, and are presented in the two following sections (3.2 and 3.3).

### 4.4.2 Results Obtained for the On-Farm Research Networks with the Frequentist Method

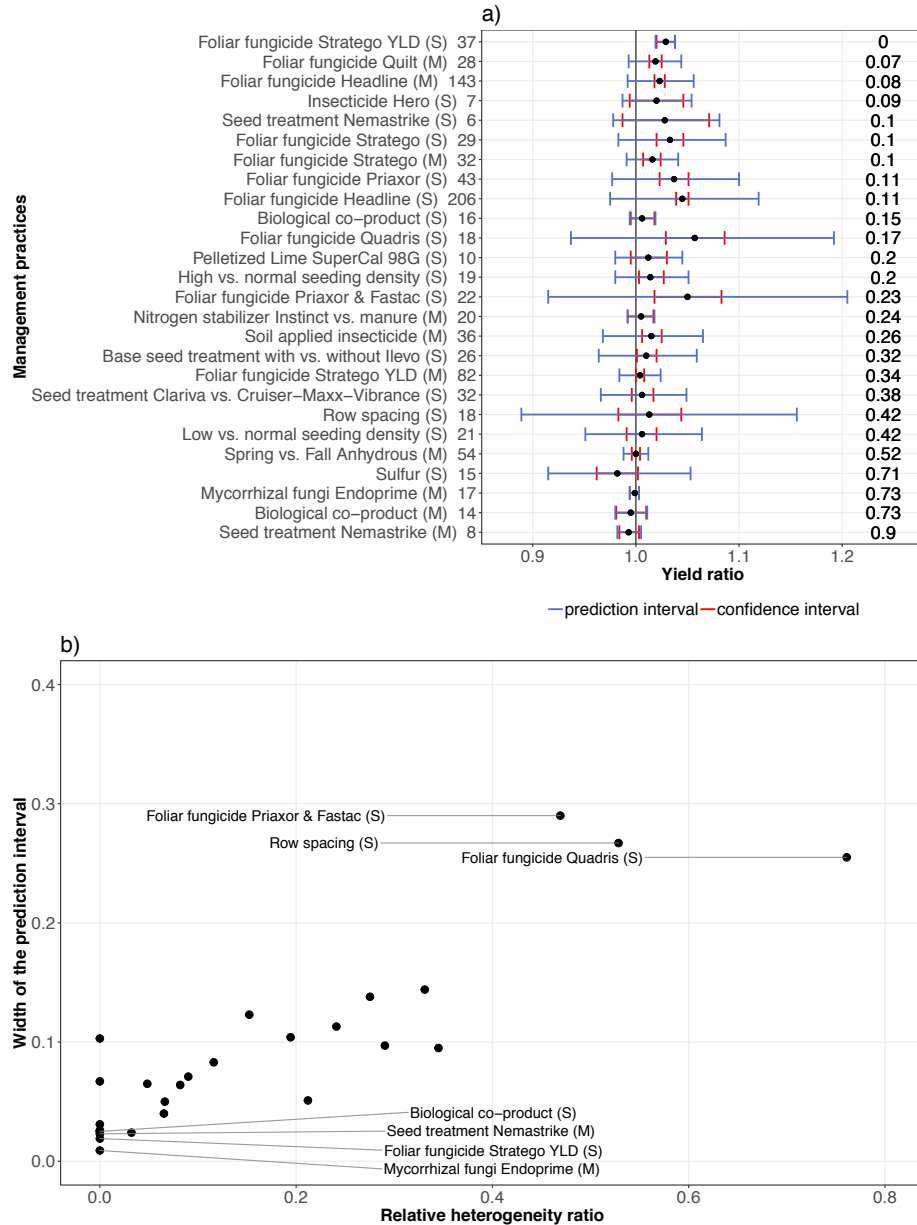


Figure 6: a) Estimated mean yield ratios with their 95% confidence (in red) and the prediction (in blue) intervals. Probabilities that the treatment will produce ineffective

results in a new trial (here, probability of a yield ratio less than 1) are displayed on the right side of the plot. Management practices are ranked by increasing the probability of ineffective treatment. S = soybean, M = maize. Both management practices, Spring vs. Fall anhydrous (M), and Nitrogen stabilizer Instinct vs. manure (M) had soybean as a previous crop. The number of trials are displayed on the left side of the graphic (a); b) Relationship between the width of the 95% prediction interval (the difference between the upper and lower bounds of the prediction interval) and the value of relative heterogeneity ratio measuring the fraction of between-trial variance in the total effect size variance. A relative heterogeneity ratio of zero indicates a null between-trial variance.

All OFRNs were analyzed with the frequentist method described in section 2.2.2 (Figure 6.a). As expected, the prediction intervals are much larger than the confidence intervals for several OFRNs, especially for Row spacing, Foliar fungicide Priaxor and Fastac, and Foliar fungicide Quadris on soybean. For example, for Foliar fungicide Quadris, the confidence interval ranges from 1.03 to 1.09, indicating a significant positive mean effect across all trials that are part of the network, but the prediction interval ranges from 0.94 to 1.19 revealing that the Quadris effect in a new situation would not be systematically positive for all trials. Its probability of ineffective result is equal to 17%, showing that, although Quadris is expected to increase yield in most situations, it could be ineffective in almost one case out of five.

For some networks, prediction intervals were almost identical to the 95% confidence intervals (Figure 6.a). These networks had a small number of trials (e.g. seed treatment NemaStrike on soybean and on maize, insecticide Hero on soybean) and/or an estimated between-trial variance close or equal to zero (e.g. foliar fungicide Stratego YLD on soybean) (Figure 6.b). In that situation, the prediction intervals given by eq.(2) were equal or similar to the confidence intervals (Figure 6.b). For the networks with a small heterogeneity (i.e. a small between-trial variability), the prediction intervals tend to be narrower than for the networks with a high heterogeneity (Figure 6.b).



As most prediction intervals for yield ratios include one (except for the foliar fungicide Stratego YLD on soybean), a systematic yield gain is not guaranteed in new situations for any of the tested management practices (Figure 6.a). For example, the network that focuses on the seed treatment Nemastrrike (active ingredient: tioxazafen) on maize had a probability of ineffective treatment equal to 0.90 which indicates a strong evidence of inefficiency in a new trial (Figure 6.a). In a contrasting example, the foliar fungicide Stratego YLD on maize (active ingredients: prothioconazole and trifloxystrobin) had a probability of ineffective treatment equal to zero as the prediction interval does not overlap one. However, the result obtained for Stratego should be interpreted with caution as the between-trial variance was estimated as zero, suggesting that we were unable to estimate the between-trial variance from this dataset correctly (Figure 6.b, and see also next section). Apart from Stratego, the chance of yield gain in a new trial was never close to 100%.

#### **4.4.3 Results obtained for the On-Farm Research Networks with the Bayesian method**

With the Bayesian method (Figure 7.a), the prediction intervals were systematically larger than the 95% credible intervals. Given the prior specification, estimated between-trial variances were always greater than zero and the prediction intervals were always larger than the confidence intervals. For example, the between-trial variance was estimated to be equal to zero with the frequentist method for the networks Seed treatment Nemastrrike on maize, Foliar fungicide Stratego YLD on soybean, and Mycorrhizal fungi Endoprime on maize (Figure 6.b), but the estimated between-trial variances were positive

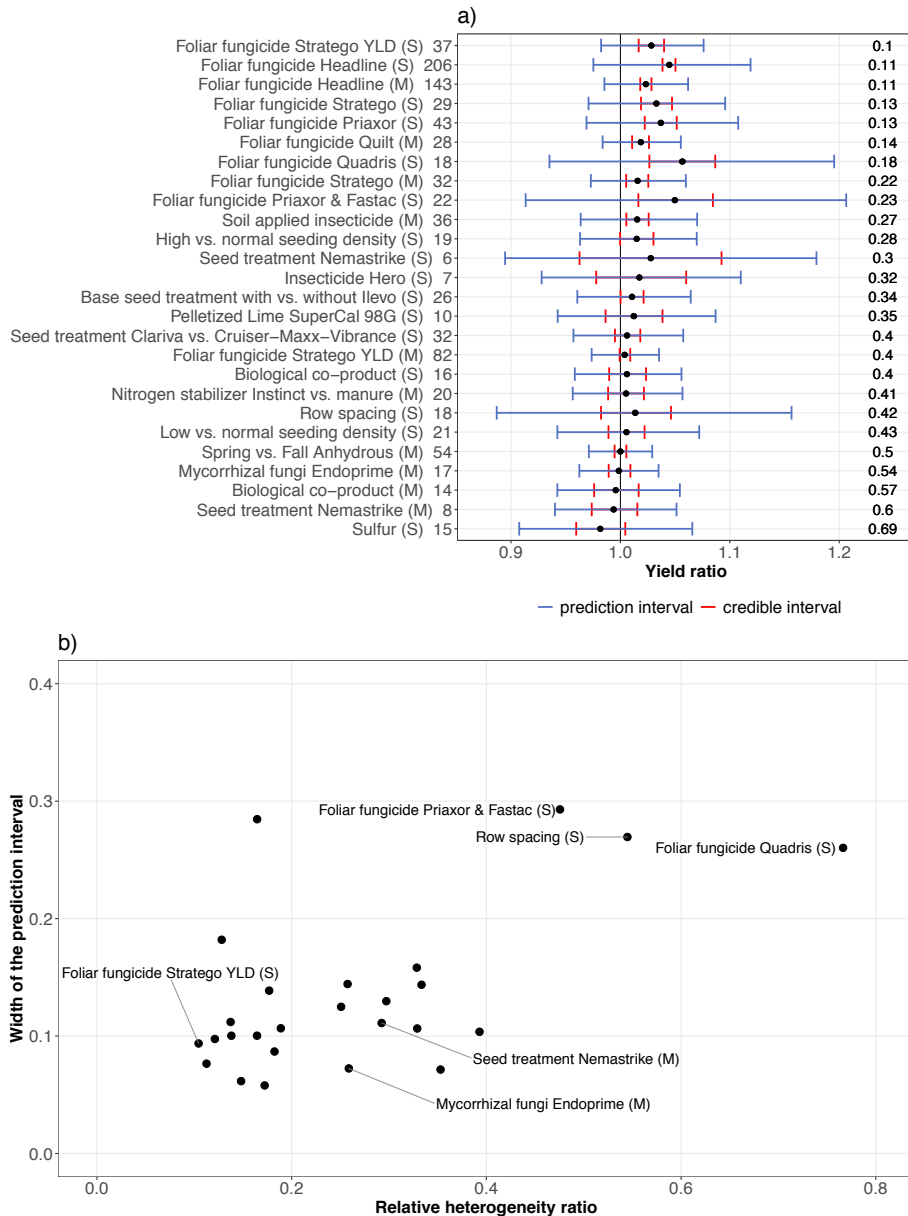


Figure 7: a) Bayesian: posterior mean yield ratios with their 95% credible intervals (in blue) and 95% prediction intervals (in red). The probability that each treatment will be ineffective in a new trial (probability of having a yield ratio less than 1) are displayed on the right side of each plot. Management practices are ranked by increasing probability of ineffective treatment. S = soybean, M = maize. The number of trials are displayed on the left (a); b) Relationship between the width of the 95% prediction interval (difference between the upper and lower bounds of the prediction interval) and the value of the relative heterogeneity ratio measuring the fraction of between-trial variance to the total effect size variance. A value of relative heterogeneity ratio equal to zero indicates a null between-trial variance.

for these networks when estimated with the Bayesian method (Figure 7.b). Our results also show that the relationship between the width of prediction intervals and the relative heterogeneity ratio is weaker with the Bayesian (Figure 7.b) compared to the frequentist approach (Figure 6.b).

Prediction intervals include the value of one in all cases (Figure 7.a), showing that a yield gain is not guaranteed for any of the tested management practices. In this context, it is clearly relevant to compute the probability of ineffective treatment in new trials. The network that focuses on the seed treatment NemaStrike (active ingredient: tioxazafen) on maize had a probability of ineffective treatment equal to 0.60 which indicates a strong evidence of inefficiency of this product in a new trial (Figure 6.b). The low seeding density compared to normal seeding density is also characterized by a high probability of ineffective treatment (0.43). In a contrasting example, all the foliar fungicides had relatively low probability of ineffective results, ranging between 0.1 and 0.23 (except for the foliar fungicide Stratego YLD on maize (active ingredients: prothioconazole and trifloxystrobin)). These results reveal that a farmer applying one of these foliar fungicides in a new trial will have between 3.3 to 9 times higher chance to have a yield gain than having a yield loss.

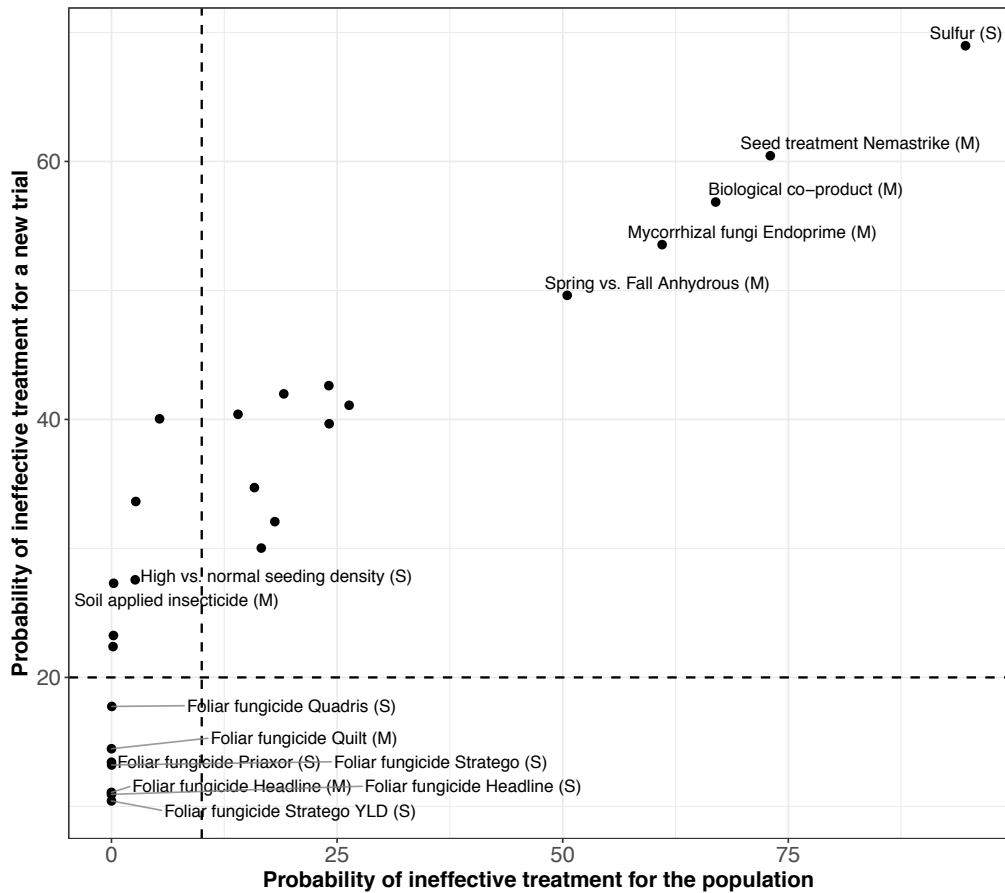


Figure 8: Probability of ineffective treatment for the population  $P(\mu < 0)$  and for new trials  $P(\theta_{new} < 0)$  for all considered networks using a Bayesian approach. One dot represents one management practice and one network. Horizontal and vertical black dashed lines indicate the threshold to define the four categories.

In addition to credible and predictive intervals, we computed the probabilities of ineffective treatment for the population  $P(\mu < 0)$  and for new trials  $P(\theta_{new} < 0)$  for all considered networks. The networks were then grouped into four categories according to their probabilities of ineffective treatment at the population level and in a new trial (Figure 8). We fixed the threshold at 10% and 20% for the probabilities of ineffective treatment in the population and in a new trial, respectively. In fact, we considered that above these thresholds, the probability of ineffective treatment indicated a strong inefficiency of the management practice. This classification can be used to identify the

management practices that are expected to produce yield gains on average in the population of fields and/or for a new field, not initially included in the network. This classification can help farmers choose management practices with strong evidence of yield gain.

Category 1 is defined by a small probability of ineffective treatment in the considered population and in a new trial (bottom left corner) (Figure 8). This category includes treatments that are expected to be efficient both on average for the population and for a new field. All the foliar fungicides, except one, belong to this category. Category 2 is defined by a high probability of ineffective treatment in the population and a low probability in a new trial (see bottom right on the Figure 10). No management practices belong to this category, which was expected because the predictive intervals are systematically larger than the confidence/credible intervals. Category 3 is represented by a low probability of ineffective treatment in the population but a high probability of ineffective treatment in a new trial (see top left on the Figure 10). The practices falling in this category are expected to perform well on average but fail to produce a yield gain in 80% or more of new situations. Category 4 is defined by a high probability of ineffective treatment both in the considered population and in a new trial; see top right on the Figure 8) and indicated inefficient management practices on yield at both levels. About half of the studied management practices belong to this latter category.

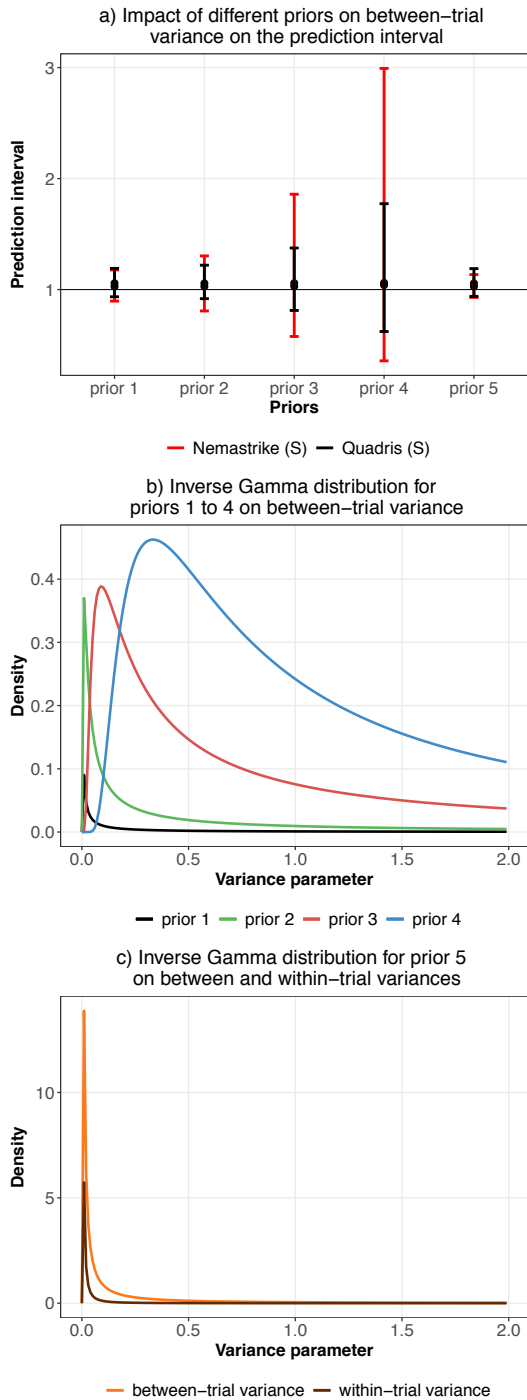


Figure 9: a) Impact of five different priors defined for the between-trial variance on the prediction intervals, for two networks (Nemastrike and Quadris); b) inverse gamma (IG) distributions used for the priors 1-4 defined for the between-trial variance; c) inverse gamma distributions used for the prior 5 defined for the between and within-trial variances and based on the on-farm research network testing the foliar fungicide Headline. Prior 1:  $\sigma_{\alpha}^2 \sim IG(0.001, 0.001)$ . Prior 2:  $\sigma_{\alpha}^2 \sim IG(0.01, 0.01)$ . Prior 3:  $\sigma_{\alpha}^2 \sim IG(0.1, 0.1)$ . Prior 4:  $\sigma_{\alpha}^2 \sim IG(0.5, 0.5)$ . Prior 5 :  $\sigma_{\alpha}^2 \sim IG(0.246, 0.001)$  and  $\sigma_{\varepsilon}^2 \sim IG(0.746, 0.0003)$ . For priors 1 to 4, the prior of the within-trial variance is unchanged and defined by  $\sigma_{\varepsilon}^2 \sim IG(0.001, 0.001)$ . Prior 5 was derived from a network including 206 trials.

Figure 9.a shows prediction intervals obtained for two networks using five different priors. The priors 1, 2 and 5 led to similar prediction intervals for both networks. The results were thus robust to the assumptions made by these three priors. The use of priors covering higher values of between-trial variances (priors 3 and 4, Figure 9b) increased the width of the prediction intervals. The prediction intervals obtained with priors 3 and 4 were larger compared to the others, especially for the seed treatment Nemastrrike on soybean, likely due to the small number of trials (6 trials only). Because of this small number of trials, the use of a prior covering high values for the between-trial variance had a stronger impact on the Nemastrrike network prediction interval compared to the Quadris network where the higher number of trials (18) was able to partially offset the influence of the prior.

#### **4.4.4 Comparison Between Frequentist and Bayesian Methods**

The results obtained with the Frequentist and the Bayesian methods are compared in Figure 10. The widths (upper bound minus lower bound) of the confidence/credible and prediction intervals obtained with the two methods are strongly correlated but the intervals were wider with the Bayesian method than with the Frequentist method (Figure 10.a and b); especially, for the seed treatment Nemastrrike on soybean and the insecticide Hero on soybean. These two management practices included only six and seven trials, respectively, and their between-study variances were thus poorly estimated and highly uncertain (Figure 7.b). The uncertainty in the estimated between-study variance is unaccounted for with the Frequentist method leading to unrealistic narrow probability

intervals. The size of the intervals derived with this approach are larger, partly because the Bayesian method takes this source of uncertainty into account using prior information.

The differences in the probability of ineffective treatment between the two approaches were smaller, but the probabilities obtained with the Bayesian approach were sometimes lower and sometimes higher, depending on the network (Figure 10.c).

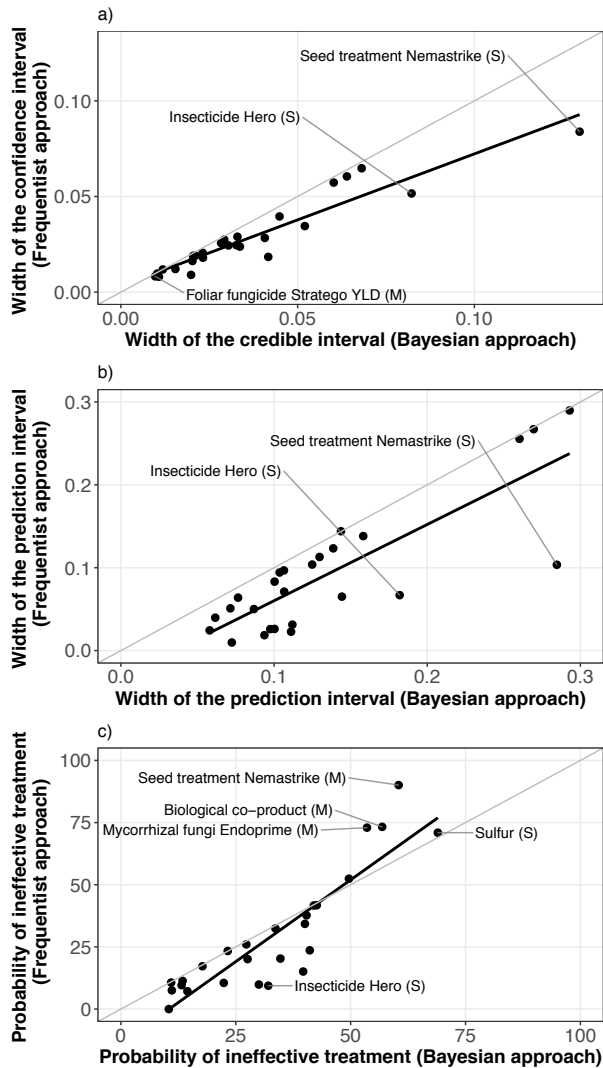


Figure 10: a) Widths of confidence intervals (frequentist approach) and of credible intervals (Bayesian approach); b) Widths of prediction intervals with frequentist and Bayesian methods; c) probability of ineffective treatment with frequentist and Bayesian methods. The regression line and the identity line are indicated in black and grey, respectively.



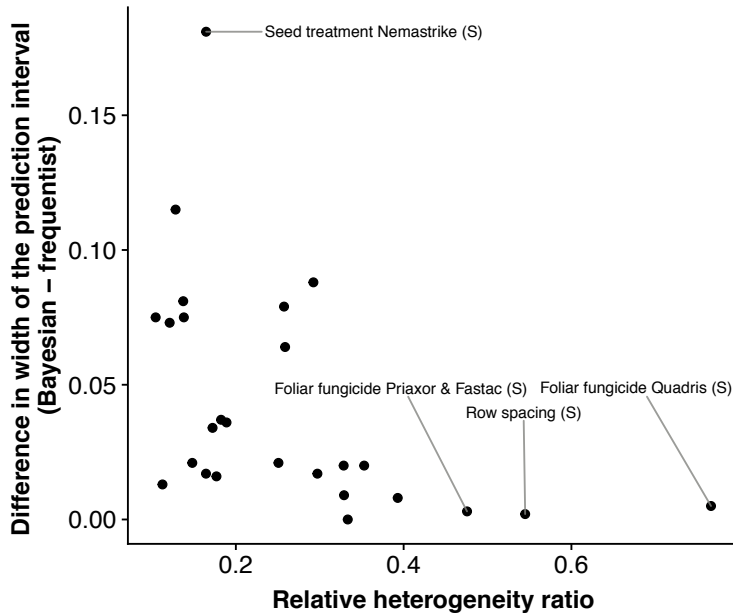


Figure 11: Difference in the width of the 95% prediction intervals between the Bayesian and frequentist methods as a function of the relative heterogeneity ratio. The width of each interval was computed by taking the difference between the upper and lower bounds of the interval. Then, the difference between the width obtained with the Bayesian method and the width obtained with the frequentist was computed.

The difference between the two approaches can be related to the level of heterogeneity (between-trial variance) in the network. In the case of low heterogeneity (i.e., RHR < 30%), the prediction interval was wider with the Bayesian method compared to the frequentist method (Figure 11). In that particular case, the frequentist method was not always able to provide a reliable estimate of the between-trial variance. The reason is that the frequentist approach is not appropriate when the estimate for the between-study variance is close to zero, which is more likely when there is a small number of trials and/or low heterogeneity in the network. In contrast, for larger heterogeneity (RHR > 30%), the two methods generated similar prediction intervals.

In practice, when heterogeneity is relatively high (RHR > 30%), both methods provided similar prediction intervals.

#### 4.5. Discussion

The confidence/credible interval of the mean effect size (e.g. mean yield ratio) provides farmers, agronomists, and advisors with information about the uncertainty in the estimated average performance of a given management practice. This is useful, but to extend inferences to a new environment (i.e., farmers' fields not originally included in the network of trials), the prediction interval matters, especially when the between-trial variability is high. Therefore, we advise data analysts to communicate the information given by the prediction intervals and the probability of ineffective treatment. We would like to emphasize that the main objective of on-farm research is to evaluate the efficacy of a product and management practice at the field level. Products used through the network are already on the market, which means they have been previously tested and approved. Before being tested through on-farm research, the performance of management practices and products have been assessed under lab and/or small-plot conditions. To scale a product or management practice at the farm level, we need to consider the interaction between biophysical, social, economic, and institutional factors (Wigboldus et al., 2016). One approach for evaluating these interactions could be farm surveys or other alternative methods conducted at the farm scale.

In the networks considered here, our results showed that the probability of ineffective treatment in a new trial often exceeded 0.2, even in the case of significant positive average effects. It is thus relevant to provide farmers and their advisors with this probability as it is expected that this will improve the transparency of the results. In particular, the computation of prediction intervals and of the probability of ineffective treatment will prevent farmers from believing that a significant effect on average will lead

with certainty to a yield gain on their own farms. Beyond the management practices and chemical products analyzed in this paper, the method could be easily extended for agroecological management practices such as bio-control agent products, companion crops (Verret et al., 2017), low impact tillage methods (e.g. no-till, strip-till) and cover crops. In addition, performances of agricultural decision support tools aiming to optimize and manage inputs such as fertilizer can also be analyzed with our proposed method.

Higgins et al. (2009) mentioned that prediction intervals are “potentially the most relevant and complete statistical inference to be drawn from random-effect meta-analyses”. Prediction intervals represent the most complete summary of a random-effects meta-analysis (Higgins et al., 2009; Partlett and Riley, 2017; Riley et al., 2011) and should be routinely reported in meta-analysis to provide robust conclusions (Graham and Moran, 2012; IntHout et al., 2016). One advantage of prediction intervals is that they are in the same metric as the original effect size and that their interpretation is more straightforward compared to the between-trial variance. However, Partlett and Riley (2017) mentioned that prediction interval should be interpreted with caution, especially when the heterogeneity is small because the interval coverage might be lower than expected. In this case, the between-trial variance is sometimes poorly estimated, especially with the frequentist method, and this can reduce the reliability of the prediction interval. Agronomically, this can be interpreted in two ways. One possibility is that there is truly very little variability in the effect of a management practice from one trial to another. The other possibility is that there is substantial variability associated with this practice but we were not able to capture it with this particular set of trials. In the latter, trials conducted under different circumstances could have revealed much greater heterogeneity. In our

study, we found that the size of the prediction interval depends on the between-trial variance, but the relationship between these two quantities was weaker when the inference is conducted with a Bayesian method due to the influence of the priors. The prediction interval is less dependent on the size of the dataset and on the level of the between-trial heterogeneity when computed with a Bayesian method, but its width may depend on the priors.

The analysis of 26 OFRNs located in Iowa showed that foliar fungicides had a significant effect on yield and, for these management practices, the probability of ineffective treatment was small for a new situation. Even though farmers can expect a yield gain in most situations, an economic analysis should be considered in recommendations, such as in the ISOFAST on-line tool (Laurent et al. 2019). The efficacy of foliar fungicides tested through on-farm trials is in accordance with the outcomes from small-plot scale research. Indeed, in his meta-analysis, Wise et al. (2019) showed that fungicide applications, compared to non-treated plots, resulted in a positive yield effect in 68% out of 436 studies. Some management practices had a significant effect on yield but the risk of having a yield loss was higher than 0.2 due to a large between-trial variance (i.e., one chance of yield loss out of four). These practices are thus expected to produce a yield gain on average, but could lead to yield loss in 20% or more of the farmers' fields. Actually, nearly half of the management practices tested through OFRNs were not effective both on average (the effect on yield was not significant) and for new situations. For example, our results showed that the mycorrhizal fungi did not increase maize yield compared to untreated plots. This negative result can be explained by different factors such as species compatibility, habitat niche availability for the mycorrhizal fungi and high competition with

native fungi (Cely et al., 2016) that determine the success of a mycorrhizal fungi in field conditions. The on-farm research network focusing on sulfur on soybean reached the highest probability of ineffective treatment. The effectiveness of sulfur could be related to some soil properties (i.e. organic matter and texture), baseline fertility or plot history such as previous manure application (Sawyer et al. 2015). In general, an ineffective management practice could be explained by, for example, an incorrect timing or dose application of a product, product contamination, absence or low pressure of pest or disease when testing pesticide applications, higher rates of nitrogen applications than optimal when testing different nitrogen practices or treatments, crop genetics, due to moisture stress (drought or excess), and other soil and environmental factors which might be limiting yield. Note that some management practices may have positive effects in some conditions and negative in others. This can be detected by including covariates in the statistical analysis (Laurent et al., 2019).

We showed that in case of a large degree of heterogeneity ( $RHR > 30\%$ ), Bayesian and frequentist methods lead to similar 95% prediction intervals. In contrast, for a low degree of heterogeneity ( $RHR < 30\%$ ), the coverage of the 95% prediction interval was narrower with the frequentist method than with the Bayesian method. As demonstrated by Partlett and Riley (2017) in a random effects meta-analysis based on a frequentist method, 95% prediction intervals, using the conventional method proposed by Higgins et al. (2009), are inaccurate when variation across trials is small. They advised considering prediction intervals derived from a frequentist method only if there is substantial between-study heterogeneity. Otherwise, a Bayesian method should be preferred.

The poor coverage of the Higgins prediction interval was also pointed out by Nagashima et al. (2018) when the heterogeneity parameter is replaced with its point estimate, but its validity depends on a large sample size which is problematic if few studies are included in the meta-analysis. Nagashima et al. (2018) proposed a parametric bootstrap method to account for the uncertainty in the heterogeneity and stated that the new prediction interval is valid even if the number of studies is small. We adapted the proposed method to our on-farm research networks, by considering trials as independent, and we compared Nagashima prediction intervals to the prediction intervals derived from our random-effect model using the frequentist and Bayesian methods. The coverage of the prediction interval was wider with the Bayesian method for 24 networks out of 26.

The Bayesian approach provides a more intuitive method to estimate prediction intervals and offers a more direct expression of uncertainty (van de Schoot et al., 2014). In addition, for some management practices, the frequentist approach estimated a between-trial variance of zero when the number of trials was small. Chung et al. (2013) mentioned that zero between-study variance estimates in random-effect models are not unusual with frequentist methods. Bayesian methods avoid obtaining a zero between-trial variance in case of small data sets and allow the uncertainty of the estimated variance to be considered (Bayarri and Berger, 2004)

When there is a small number of studies, a Bayesian approach with an informative prior distribution for the between-study variance may be a viable option (Borenstein et al., 2009; Chung et al., 2013; Higgins et al., 2009). Nevertheless, the Bayesian method we used has some limitations. In particular, the prior specification is not always straightforward (van de Schoot et al., 2014) and may have some influence on results and interval

estimations (Van Dongen, 2006). Our sensitivity analysis showed that the prediction interval is sensitive to the between-trial variance prior when the between-trial variability is small. Consequently, the prediction interval should be interpreted with caution in this case. In addition, past studies showed that the credible and prediction intervals obtained with Bayesian methods do not always have good frequentist properties, i.e. their coverage is not necessarily close to the target value (e.g. 95%). This depends on several factors and, in particular, on the prior distributions used (Williams et al., 2018; Bayarri and Berger, 2004 section 3.4.3).

#### **4.6 Conclusion**

In conclusion, we advise researchers dealing with field experiments or on-farm trials to compute both confidence/credible interval and prediction intervals because they convey different highly relevant information. Nevertheless, we should be cautious in interpreting prediction intervals when between-trial heterogeneity is low and recommend the use of a Bayesian method in this case with informative priors whenever possible.

#### **4.7 Acknowledgements**

The work of FM and AL was also supported by the Iowa Soybean Research Center and the Plant Science Institute at Iowa State University. The work of DM was partly funded by the Institut Carnot Plant2Pro (project LICITE) and by the CLAND project (ANR-CONV-0006).

#### **4.8 References**

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67. <https://doi.org/10.18637/jss.v067.i01>

Bayarri, M.J., Berger, J.O., 2004. The Interplay of Bayesian and Frequentist Analysis. *Statist. Sci.* 19, 58–80. <https://doi.org/10.1214/088342304000000116>

Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., 2009. *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd, Chichester, UK. <https://doi.org/10.1002/9780470743386>

Borenstein, M., Higgins, J.P.T., Hedges, L.V., Rothstein, H.R., 2017. Basics of meta-analysis:  $I^2$  is not an absolute measure of heterogeneity:  $I^2$  is not an absolute measure of heterogeneity. *Res. Syn. Meth.* 8, 5–18. <https://doi.org/10.1002/jrsm.1230>

Carciochi, W.D., Schwalbert, R., Andrade, F.H., Corassa, G.M., Carter, P., Gaspar, A.P., Schmidt, J., Ciampitti, I.A., 2019. Soybean Seed Yield Response to Plant Density by Yield Environment in North America. *Agron. J.* 111, 1923–1932. <https://doi.org/10.2134/agronj2018.10.0635>

Casler, M.D., 2015. Fundamentals of Experimental Design: Guidelines for Designing Successful Experiments. *Agronomy Journal* 107, 692. <https://doi.org/10.2134/agronj2013.0114>

Cely, M.V.T., de Oliveira, A.G., de Freitas, V.F., de Luca, M.B., Barazetti, A.R., dos Santos, I.M.O., Gionco, B., Garcia, G.V., Prete, C.E.C., Andrade, G., 2016. Inoculant of Arbuscular Mycorrhizal Fungi (*Rhizophagus clarus*) Increase Yield of Soybean and Cotton under Field Conditions. *Front. Microbiol.* 7. <https://doi.org/10.3389/fmicb.2016.00720>

Chung, Y., Rabe-Hesketh, S., Choi, I.-H., 2013. Avoiding zero between-study variance estimates in random-effects meta-analysis. *Statistics in Medicine* 32, 4071–4089. <https://doi.org/10.1002/sim.5821>

Crossa, J., 1990. Statistical Analyses of Multilocation Trials, in: *Advances in Agronomy*. Elsevier, pp. 55–85. [https://doi.org/10.1016/S0065-2113\(08\)60818-4](https://doi.org/10.1016/S0065-2113(08)60818-4)

Gates, S., 2002. Review of methodology of quantitative reviews using meta-analysis in ecology. *J Anim Ecology* 71, 547–557. <https://doi.org/10.1046/j.1365-2656.2002.00634.x>

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 515–533.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2014. *Bayesian data analysis*. CRC Press, Boca Raton.

Graham, P.L., Moran, J.L., 2012. Robust meta-analytic conclusions mandate the provision of prediction intervals in meta-analysis summaries. *Journal of Clinical Epidemiology* 65, 503–510. <https://doi.org/10.1016/j.jclinepi.2011.09.012>

Guddat, C., Grouven, U., Bender, R., Skipka, G., 2012. A note on the graphical presentation of prediction intervals in random-effects meta-analyses. *Syst Rev* 1, 34. <https://doi.org/10.1186/2046-4053-1-34>



Gurevitch, J., Koricheva, J., Nakagawa, S., Stewart, G., 2018. Meta-analysis and the science of research synthesis. *Nature* 555, 175–182. <https://doi.org/10.1038/nature25753>

Hadfield, J.D., 2016. MCMCglmm Course Notes.

Hadfield, J.D., 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software* 33. <https://doi.org/10.18637/jss.v033.i02>

Higgins, J.P.T., Thompson, S.G., Spiegelhalter, D.J., 2009. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172, 137–159. <https://doi.org/10.1111/j.1467-985X.2008.00552.x>

IntHout, J., Ioannidis, J.P.A., Rovers, M.M., Goeman, J.J., 2016. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open* 6, e010247. <https://doi.org/10.1136/bmjopen-2015-010247>

Iowa Soybean Association, 2018. Power Calculator for Multilocation Trials.

Johnston, Renteria, A., Hannon, M.R., 2003. Improving validity of on-farm research. *Journal of Swine Health Production* 11, 240:246.

Kolkman, J.M., Kelly, J.D., 2002. Agronomic Traits Affecting Resistance to White Mold in Common Bean. *Crop Science* 42, 693. <https://doi.org/10.2135/cropsci2002.0693>

Krzywinski, M., Altman, N., 2013. Error bars. *Nat Methods* 10, 921–922. <https://doi.org/10.1038/nmeth.2659>

Kyveryga, P. M., Blackmer, T.M., Mueller, D.S., 2013. When Do Foliar Pyraclostrobin Fungicide Applications Produce Profitable Soybean Yield Responses? *Plant Health Progress* 14, 6. <https://doi.org/10.1094/PHP-2013-0928-01-RS>

Kyveryga, P. M., Caragea, P.C., Kaiser, M.S., Blackmer, T.M., 2013. Predicting Risk from Reducing Nitrogen Fertilization Using Hierarchical Models and On-Farm Data. *Agronomy Journal* 105, 85. <https://doi.org/10.2134/agronj2012.0218>

Lambert, P.C., Sutton, A.J., Burton, P.R., Abrams, K.R., Jones, D.R., 2005. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statist. Med.* 24, 2401–2428. <https://doi.org/10.1002/sim.2112>

Laurent, A., Kyveryga, P., Makowski, D., Miguez, F., 2019. A Framework for Visualization and Analysis of Agronomic Field Trials from On-Farm Research Networks. *Agronomy Journal* 111, 1–12. <https://doi.org/10.2134/agronj2019.02.0135>

Mallowa, S.O., Esker, P.D., Paul, P.A., Bradley, C.A., Chapara, V.R., Conley, S.P., Robertson, A.E., 2015. Effect of Maize Hybrid and Foliar Fungicides on Yield Under Low Foliar Disease Severity Conditions. *Phytopathology* 105, 1080–1089. <https://doi.org/10.1094/PHYTO-08-14-0210-R>

Moore, K.J., Dixon, P.M., 2015. Analysis of Combined Experiments Revisited. *Agronomy Journal* 107, 763. <https://doi.org/10.2134/agronj13.0485>

Nagashima, K., Noma, H., Furukawa, T.A., 2018. Prediction intervals for random-effects meta-analysis: A confidence distribution approach. *Stat Methods Med Res* 096228021877352. <https://doi.org/10.1177/0962280218773520>

Nakagawa, S., Cuthill, I.C., 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews* 82, 591–605. <https://doi.org/10.1111/j.1469-185X.2007.00027.x>

Nakagawa, S., Noble, D.W.A., Senior, A.M., Lagisz, M., 2017. Meta-evaluation of meta-analysis: ten appraisal questions for biologists. *BMC Biol* 15, 18. <https://doi.org/10.1186/s12915-017-0357-7>

Nelson, R., Coe, R., Haussmann, B.I.G., 2016. Farmer research networks as a strategy for matching diverse options and contexts in smallholder agriculture. *Ex. Agric.* 55, 125–144. <https://doi.org/10.1017/S0014479716000454>

Partlett, C., Riley, R.D., 2017. Random effects meta-analysis: Coverage performance of 95 % confidence and prediction intervals following REML estimation. *Statistics in Medicine* 36, 301–317. <https://doi.org/10.1002/sim.7140>

Philibert, A., Loyce, C., Makowski, D., 2012. Assessment of the quality of meta-analysis in agronomy. *Agriculture, Ecosystems & Environment* 148, 72–82. <https://doi.org/10.1016/j.agee.2011.12.003>

Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 1 6, 7–11.

Raftery, A.E., Lewis, S.M., 1995. The Number of Iterations, Convergence Diagnostics and Generic Metropolis Algorithms, in: W. R. Gilks, D. J. Spiegelhalter, & S. Richardson (Eds.). Chapman and Hall, London, UK, pp. 115–130.

Riley, R.D., Higgins, J.P.T., Deeks, J.J., 2011. Interpretation of random effects meta-analyses. *BMJ* 342, d549–d549. <https://doi.org/10.1136/bmj.d549>

RStudio Team, 2015. RStudio: Integrated Development Environment for R. RStudio, Inc., Boston, MA.

Sawyer, J., Lang, B., Barker, D., 2015. Sulfur Management for Iowa Crop Production. Extension and Outreach Publications.

- Sedgwick, P., 2014. Understanding confidence intervals. *BMJ* 349, g6051–g6051. <https://doi.org/10.1136/bmj.g6051>
- Toft, N., Innocent, G.T., Gettinby, G., Reid, S.W.J., 2007. Assessing the convergence of Markov Chain Monte Carlo methods: An example from evaluation of diagnostic tests in absence of a gold standard. *Preventive Veterinary Medicine* 79, 244–256. <https://doi.org/10.1016/j.prevetmed.2007.01.003>
- van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J.B., Neyer, F.J., van Aken, M.A.G., 2014. A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. *Child Development* 85, 842–860. <https://doi.org/10.1111/cdev.12169>
- Van Dongen, S., 2006. Prior specification in Bayesian statistics: Three cautionary tales. *Journal of Theoretical Biology* 242, 90–100. <https://doi.org/10.1016/j.jtbi.2006.02.002>
- Veroniki, A.A., Jackson, D., Bender, R., Kuss, O., Langan, D., Higgins, J.P.T., Knapp, G., Salanti, G., 2019. Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. *Res Syn Meth* 10, 23–43. <https://doi.org/10.1002/jrsm.1319>
- Verret, V., Gardarin, A., Pelzer, E., Médiène, S., Makowski, D., Valantin-Morison, M., 2017. Can legume companion plants control weeds without decreasing crop yield? A meta-analysis. *Field Crops Research* 204, 158–168. <https://doi.org/10.1016/j.fcr.2017.01.010>
- Wakefield, J., 2013. Bayesian and Frequentist Regression Methods, Springer Series in Statistics. Springer New York, New York, NY. <https://doi.org/10.1007/978-1-4419-0925-1>
- Wigboldus, S., Klerkx, L., Leeuwis, C., Schut, M., Muilerman, S., Jochemsen, H., 2016. Systemic perspectives on scaling agricultural innovations. A review. *Agron. Sustain. Dev.* 36, 46. <https://doi.org/10.1007/s13593-016-0380-z>
- Williams, D.R., Rast, P., Bürkner, P.-C., 2018. Bayesian Meta-Analysis with Weakly Informative Prior Distributions (preprint). *PsyArXiv*. <https://doi.org/10.31234/osf.io/7tbrm>
- Wise, K., Mueller, D., 2011. Are Fungicides No Longer Just For Fungi? An Analysis of Foliar Fungicide Use in Corn. *APSnet Features*. <https://doi.org/10.1094/APSnetFeature-2011-0531>
- Wise, K.A., Smith, D., Freije, A., Mueller, D.S., Kandel, Y., Allen, T., Bradley, C.A., Byamukama, E., Chilvers, M., Faske, T., Friskop, A., Hollier, C., Jackson-Ziems, T.A., Kelly, H., Kemerait, B., Price, P., Robertson, A., Tenuta, A., 2019. Meta-analysis of yield response of foliar fungicide-treated hybrid corn in the United States and Ontario, Canada. *PLoS ONE* 14, e0217510. <https://doi.org/10.1371/journal.pone.0217510>

## 4.9 Supplemental Materials

R code and full demo:

```

Required R packages and corresponding version
knitr::opts_chunk$set(echo = TRUE)
rm(list=ls())
library(dplyr)      # version 0.8.3

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
library(ggplot2)   # version 3.2.1
library(MCMCglmm)  # version 2.29

## Loading required package: Matrix
## Loading required package: coda
## Loading required package: ape
library(lme4)       # version 1.1-21
library(Hmisc)     # version 4.2-0

## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
##
## Attaching package: 'Hmisc'
## The following object is masked from 'package:ape':
##
##   zoom
## The following objects are masked from 'package:dplyr':
##
##   src, summarize
## The following objects are masked from 'package:base':
##
##   format.pval, units
library(forcats)   # version 0.4.0

simdata<-read.csv("~/Dropbox/PhD/Thesis dissertation/sections/Appendix/appendix_C/simdata.csv")

For the purpose of the supplementary material, we used a data frame called "datasim" including simulated
yield ratio (lrr) of a hypothetical management practice to a control for each replicate (rep) within one trial
identifier (Trial_ID). Please find the csv file under the folder called data

head(simdata)

##   Trial_ID rep      lrr
## 1      T1  1 -1.5664706
## 2      T1  2 -0.2764872

```

```
## 3      T1    3 -0.7414822
## 4      T1    4  0.9210677
## 5      T1    5  0.8332006
## 6      T1    6  0.7745080
```

## Frequentist approach using lme4 R package (Bates et al., 2015)

First, run the random-effect model

```
mod_fq=lmer(lrr ~ 1 + (1|Trial_ID),data= simdata)
```

- point estimate of log ratio (median)

```
fit = fixef(mod_fq)
```

- Lower bound of confidence interval (2.5%)

```
low = as.data.frame(confint(mod_fq))[3,1]
```

- Upper bound of confidence interval (97.5%)

```
up = as.data.frame(confint(mod_fq))[3,2]
```

- prediction interval based on the method of Higgins et al. (2009)

```
bs_var= (as.data.frame(VarCorr(mod_fq))[1,5])^2 # estimated between-trial variance
se = as.data.frame(coef(summary(mod_fq)))[1,2] # standard error of the intercept
var = se^2 # variance of the estimated mean effect
k = length(unique(simdata$Trial_ID)) # number of trials
mu=fixef(mod_fq) # quantile 0.5 of the prediction interval
low = mu-qt(.975, df=k-2)*sqrt(bs_var+var) # quantile 0.025 of the prediction interval
up = mu+qt(.975, df=k-2)*sqrt(bs_var+var) # quantile 0.975 of the prediction interval
```

- probability of ineffective treatment

```
var_total<-sqrt(bs_var+var) # total variability
q<-mu/var_total # mean = q*total variability
pt(q, df=k-2) # student t distribution
proba=1-pt(q, df=k-2) # probability of ineffective treatment
```

## Bayesian approach using MCMCglmm R package (Hadfield, 2010)

- credibility interval and individual trial yield response

First, run the random-effect model

```
prior1<- list(B = list(mu=0,V=2), # prior on the fixed effect
             G=list(G1=list(V=1, nu=0.002)), # prior on the between-trial variance
             R = list(V = 1, nu = 0.002)) # prior on the within-trial variance

mod_bayes<-MCMCglmm(lrr~1, # fixed effect
                  random=~Trial_ID, # random effect
                  data=simdata,
                  family="gaussian",
                  prior=prior1, # priors (defined above)
                  thin=10, # thinning interval
                  nitt= 200000, # number of iterations
                  burnin = 20000, # burnin
```

```

verbose=FALSE,      # if true MH diagnostics are printed to screen
pr=TRUE,            # posterior distribution of random effects are saved
DIC=TRUE)           # deviance information criterion

```

- credibility interval of the mean log ratio

```
cred_int<-predict(mod_bayes,interval="confidence",level=0.95)[1,]
```

fit represents the point estimate of log ratio (median)

lwr represents the lower bound of credibility interval (quantile 0.025)

upr represents the upper bound of credibility interval (quantile 0.975)

- credibility interval of individual log yield ratio

```

pred<-predict(mod_bayes,interval="confidence", marginal=mod_bayes$Trial_ID,level=0.95)
pred<-as.data.frame(pred)
pred$trial<-simdata$Trial_ID

```

```

fit<-pred$fit
lwr<-pred$lwr
upr<-pred$upr
pred$trial <- reorder(pred$trial, pred$fit)

```

```
duplication<-which(duplicated(pred))
```

```
pred<-pred[-duplication,]
```

```
print(pred)      # dataframe returning the credibility interval of individual log ratio
```

```

##          fit          lwr          upr trial
## 1    0.03017861 -0.6187245722 0.703096327   T1
## 7    0.36527045 -0.3073941588 1.026622655   T2
## 13   -0.02984426 -0.7224213287 0.611527105   T3
## 19    0.91973970  0.2145759155 1.608352290   T4
## 25    0.37002046 -0.2844214390 1.051100788   T5
## 31   -0.64276247 -1.3514431294 0.044183806   T6
## 37    0.27574911 -0.3815017917 0.934770685   T7
## 43   -0.01524471 -0.6691457939 0.665777735   T8
## 49   -0.16458629 -0.8507691450 0.498099267   T9
## 55   -0.27190691 -0.9409554930 0.387783326  T10
## 61   -0.52198664 -1.2044730628 0.158288868  T11
## 67   -0.48081198 -1.1736078630 0.191935473  T12
## 73    1.04841216  0.3118909431 1.725425483  T13
## 79   -0.65457631 -1.3478837959 0.031545325  T14
## 85    0.54645374 -0.1191727725 1.233794661  T15
## 91    0.68815028  0.0194888869 1.381003571  T16
## 97    0.18770461 -0.4819972397 0.840268623  T17
## 103  -0.19745905 -0.8534295390 0.487277300  T18
## 109  0.88181239  0.2237714724 1.588294891  T19
## 115  -0.68624803 -1.3793391594 0.008250905  T20
## 121  0.95281852  0.2633895370 1.662107413  T21
## 127  -0.16313033 -0.8285265449 0.512344119  T22
## 133  0.22344044 -0.4451294485 0.888321462  T23
## 139  0.66425947  0.0005631111 1.370772180  T24
## 145  -0.39801074 -1.0732504504 0.262715751  T25

```

The column “fit” represents the point estimate of log ratio for each individual trial (see column “trial”) (median)

The column “lwr” represents the lower bound of credibility interval for each individual trial (quantile 0.025)  
 The column “upr” represents the upper bound of credibility interval for each individual trial (quantile 0.975)

- prediction interval for a new trial

In order to compute a prediction interval, you need to add a new trial with missing value was added (here called IDNew) and run the random effect model.

```
simdata<-simdata %>% add_row(Trial_ID="IDNew")
tail(simdata)
```

```
##      Trial_ID rep      lrr
## 146     T25  2 -0.75740552
## 147     T25  3 -0.01329639
## 148     T25  4 -0.61918290
## 149     T25  5  0.38288358
## 150     T25  6 -0.59204070
## 151   IDNew NA          NA
```

```
mod_bayes_new<-MCMCglmm(lrr~1, # fixed effect
  random=~Trial_ID, # random effect
  data=simdata,
  family="gaussian",
  prior=prior1, # priors (defined above)
  thin=10, # thinning interval
  nitt= 200000, # number of iterations
  burnin = 20000, # burnin
  verbose=FALSE, # if true MH diagnostics are printed to screen
  pr=TRUE, # posterior distribution of random effects are saved
  DIC=TRUE)
```

Instead of using the function predict, the MCMC chains were used to compute the prediction interval

```
REnew<-as.data.frame(mod_bayes_new$Sol[, "Trial_ID.IDNew"]) # MCMC chain for the random effect IDNew
intercept<-as.data.frame(mod_bayes_new$Sol[, "(Intercept)"]) # MCMC chain for the the intercept
newtrial<-REnew+intercept # sum the two MCMC chains
predinterval<-quantile(newtrial[,1], c(.025, .5, .975)) # get quantiles
```

- probability of ineffective treatment

```
X<-sum(newtrial$var1<0) # number of cases where log ratio <0
n<-length(newtrial$var1) # total number of possible cases
proba<-X/n # probability of ineffective treatment
```

You do need an exponential transformation to convert the Frequentist and Bayesian outputs from log yield ratio to yield ratio.

## Figures

We use the exponential transformation for the figures

```
# outputs = dataframe combining trial estimations, credible interval and prediction interval
outputs <- pred %>%
  add_row(fit=cred_int[1],
    lwr=cred_int[2],
    upr=cred_int[3],
    trial="mean yield ratio") %>%
  add_row(fit=predinterval[2],
    lwr=predinterval[1],
```

```

upr=predinterval[3],
trial="prediction interval")

outputs<- outputs %>%
  mutate(trial = forcats::fct_reorder(trial,fit)) %>%
  mutate(trial = forcats::fct_relevel(trial, "prediction interval", after = 0)) %>%
  mutate(trial = forcats::fct_relevel(trial, "mean yield ratio", after = 1))

ggplot(outputs, aes(x = exp(fit), xmin = exp(lwr), xmax = exp(upr), y = trial )) +
  geom_point(size=2) +
  geom_segment( aes(x = exp(lwr), xend = exp(upr), y = trial, yend=trial)) +
  theme_bw() +
  xlab("Yield ratio") +
  ylab("") +
  geom_text(size=5,aes(x=5.2,y=1.2,label= ifelse(trial=="prediction interval",
                                                round(proba,2),""))) +
  geom_vline(xintercept = 1, color = "black", size=0.8) +
  theme(axis.text=element_text(size=12,face="bold"),
        axis.title =element_text(size=12,face="bold"),
        legend.text=element_text(size=12,face="bold"),
        legend.title=element_blank())

```

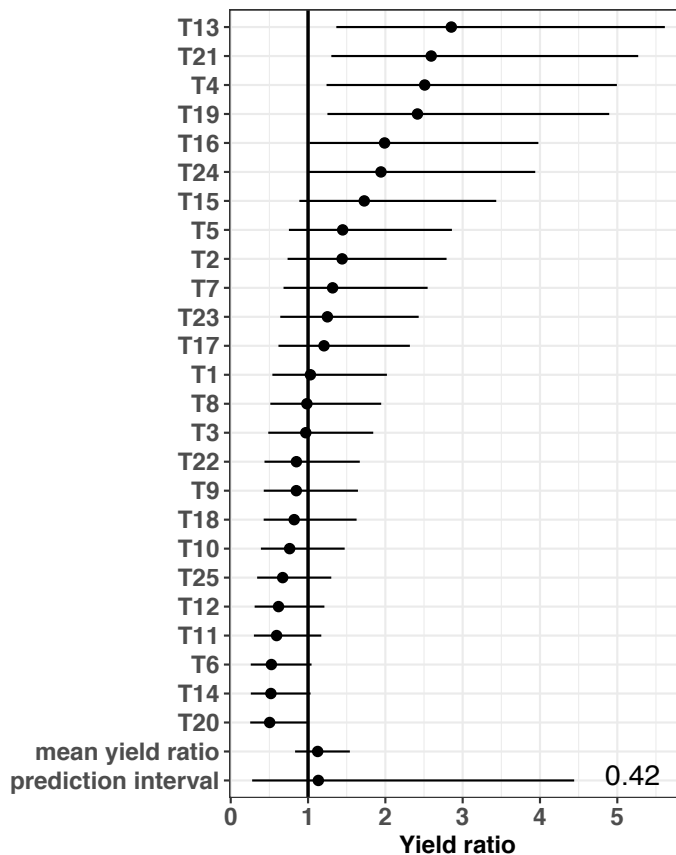


Figure S1. Estimated individual effects (denoted by T and a numerical value) with their 95% credible intervals, mean yield ratio and its 95% credible interval, and 95% prediction interval. The numerical value at the bottom-right indicates the probability of ineffective treatment (probability of yield ratio less than 1) in a new trial.



## CHAPTER 5. ON-FARM TRIALS REVEAL SIGNIFICANT BUT UNCERTAIN CONTROL OF BOTRYTIS CINEREA BY AUREOBASIDIUM PULLULANS AND POTASSIUM BICARBONATE IN ORGANIC GRAPEVINES.

Modified from a paper under review in *Frontiers in Plant Sciences*

Anabelle Laurent<sup>1</sup>, David Makowski<sup>2</sup>, Nicolas Aveline<sup>3</sup>, Séverine Dupin<sup>4</sup>, Fernando Miguez<sup>1</sup>

<sup>1</sup>Department of Agronomy, Iowa State University, Ames, IA, USA

<sup>2</sup>INRAE, Centre Île-de-France, Paris, France

<sup>3</sup>Institut Français de la Vigne et du Vin–Vinopôle Nouvelle Aquitaine, France

<sup>4</sup>Chambre d’Agriculture de la Gironde-Vinopôle Nouvelle Aquitaine, France

### 5.1 Abstract

*Botrytis cinerea*, a fungal pathogen that causes gray mold, can decrease grape yield, substantially reduce wine quality, and therefore cause significant economic losses. In a context of increasing awareness of environmental and human health, biopesticides are potential alternatives to synthetic chemical treatments to produce grapes and wine with high food standards. However, the effectiveness of biopesticides is not well known, and more research is needed to help winemakers assess biopesticide ability to control wine diseases. Our study aims to assess the efficacy of two commercial biopesticides, based on potassium bicarbonate and *Aureobasidium pullulans*, in reducing the incidence of gray mold (*i.e.*, the proportion of grape bunches that are diseased) in grapes. We use data from an on-farm trial network managed over three years in a major wine-producing region located in Southwestern France, and fit Bayesian generalized linear multilevel models able to take into account the variability of treatment effect across trials. The fitted models were then used to estimate the efficacy on incidence as a function of the intensity (*i.e.*, the proportion

of diseased grape berries in a bunch) in an untreated plot in order to determine if the effectiveness of the treatments depends on the disease pressure. At average disease intensity (*i.e.*, 3%), the efficacy on disease incidence at the network level was equal to 20% (95%CI=[-0.1; 37.3]) and 13% (95%CI=[0.2; 24.7]) for potassium bicarbonate and *Aureobasidium pullulans*, respectively. For both biopesticides, the efficacy on incidence for a new site-year is highly uncertain, but potassium bicarbonate had a lower uncertainty and a lower application cost compared to *Aureobasidium pullulans*. Our results confirm that potassium bicarbonate is a promising biopesticide for organic vineyards in southwestern France, but the large uncertainty suggests the need for further research.

## 5.2 Introduction

Reducing the use of synthetic pesticides has become a major objective in Europe, particularly in France, where their use has remained high for many years despite government actions (Hossard et al., 2017). On vines, synthetic pesticides are used several times during a growing season to control various diseases caused mostly by pathogenic fungi (Chen et al., 2019). Grapevine accounts for around 14% of product purchases for less than 4% of the agricultural area occupied (Butault et al., 2011). The intensive use of synthetic pesticides has negative consequences on the environment by contaminating soil, surface and ground water (Komarek et al., 2009; Mailly et al., 2017), and on human health (Coleman et al., 2012; Chen et al., 2020).

For environmental and human health reasons, there is a growing interest in using alternatives to synthetic pesticides to produce grapes and wine without hazardous chemical residues in wine (Jacometti et al., 2010; Rotolo et al., 2018). Alternative methods

rely on the use of biopesticides based on macro/micro-organisms (also called biocontrol agents) and substances of natural origin such as plant and microbial extracts, mineral and organic compounds (EPA, 2016; Nicot et al., 2016; Amichot et al., 2018). Biopesticides have been mostly evaluated under controlled conditions (Nicot, 2011; Haidar et al., 2016) and their performance under field conditions can be lower, or ineffective and highly variable, compared to controlled conditions, due to climatic variations, unstable quality of the product and difficulty in maintaining living organisms in good conditions (*i.e.*, survival and colonization ability) (Nicot, 2011; Bardin et al., 2015; Pertot et al., 2017a). Thus, there is a need to test biopesticides under field conditions to assess their actual effectiveness and adapt control strategies against grey mold in vineyards (Nicot, 2011; Pertot et al., 2017a).

*Plasmopara viticola* (the causal agent of downy mildew), *Erysiphe necator* (the causal agent of powdery mildew), and *Botrytis cinerea* (hereafter called *B. cinerea*) require the highest number of chemical treatments in several major wine-producing regions, in particular in France. *B. cinerea*, a fungal pathogen that causes gray mold (also called botrytis bunch rot), can decrease yield, highly reduce wine quality, and cause significant economic losses (Jacometti et al., 2010). Botrytis bunch rot can affect the berries during ripening, just a few days before harvest. As the use of synthetic pesticides is not allowed at that time of the growing season, biopesticides represent an opportunity for winegrowers to control this disease late in the season.

Previous studies, conducted under field conditions (*i.e.*, research vineyard or commercial vineyard), have reported mixed results on the efficacy of biopesticides against *B. cinera* (Supplementary Material, Table S1). Pertot et al., (2017b) evaluated anti-*B.cinerea* biopesticides with different mechanisms of action, alone and in mixtures, and showed that

the average efficacy on disease intensity (*i.e.*, the percentage of reduction of diseased grapes berries using the treatment compared to the control) was high, with little variability between years and locations (Supplementary Material, Table S1). Similar levels of efficacy were achieved with a single biopesticide and with multiple biopesticides. The experiments of Pertot et al., (2017b) were characterized by low levels of disease intensity, and the use of management practices decreasing grape susceptibility (*i.e.*, defoliation) which could explain the high efficacy levels obtained in these experiments. Other studies showed lower levels of efficacy. Rotolo et al. (2018) found that disease incidence (*i.e.*, the percentage of reduction of diseased grapes bunches using the treatment compared to the control) was reduced by less than 30% under high disease pressure with a biopesticide treatment while the use of mixtures of or the alternate use of biopesticides and synthetic fungicide (including fluopyram as an active ingredient) (Supplementary Material, Table S1) showed efficacy up to 96%. They suggested combining biopesticides with limited use of synthetic fungicides as an alternative strategy to reduce the risk of fungicide resistance and residue levels. Calvo-Garrido et al. (2019) found that several commercial biopesticides products achieved 21% to 58% of efficacy in botrytis bunch rot intensity, but not on every trial. Finally, Aziz et al. (2016) has reported high reduction levels of disease incidence and on intensity (up to 68% and 93%) with treatments combining several biopesticides (Supplementary Material, Table S1).

The above-mentioned studies provide useful information on the efficacy of different biopesticides but none of them attempted to relate the level of efficacy to the disease pressure. As biopesticide efficacy may depend on disease pressure, it would be valuable to establish a quantitative relationship between the effect of biopesticides on disease control

and the disease pressure in the vineyard considered. Moreover, the results of past studies suggest that the effect of pesticides is highly variable between vineyards, but so far, this variability has not been properly quantified. Finally, none of the previous studies relied on actual on-farm trials conducted by winegrowers using their own equipment in their own vineyards. Some of these studies were conducted in commercial vineyards but using a non-commercial sprayer (*i.e.*, small equipment spraying pesticides on grapes directly). To address these limitations, we assessed the efficacy of biopesticides on the disease incidence with the goal of understanding how the efficacy was affected by the disease pressure. We analyzed data from two on-farm research networks in organic grapevine production, based in Southwestern France, for which biopesticides on *B. cinerea* have been tested. In our analysis, disease intensity in untreated vineyards was used as a proxy for disease pressure. All the data were collected in commercial vineyards conducted according to standard farming practices during three years in 23 and 26 sites-years, regarding the biopesticide. This dataset offers a unique opportunity to assess the efficacy of biopesticides under realistic conditions as winegrowers used commercially available field equipment and managed the on-farm trial with their standard management practices apart from the biopesticide treatment. Two biopesticides are considered, namely *Aureobasidium pullulans* (hereafter called *A. pullulans*) and potassium bicarbonate. Both were registered on the French market in 2012 and 2011, respectively. Their modes of action are very different as the first one is based on a living organism while the second is a non-biological product. Before 2011, two biopesticides, composed of *Bacillus Subtilis*, were approved for organic production (Serenade® and Serenade Max®), but the efficacy was low when disease pressure was relatively high, and the market price was high (110 €/ha) (RESAQVitiBio,

2015). In this context, the Vinopôle de Bordeaux-Aquitaine conducted several on-farm trials in order to assess the effectiveness of new products under farming conditions. To our knowledge, our study is the first to report the results of trials assessing potassium bicarbonate in field trials on vine to control *Botrytis cinerea*.

Here, our objective is to evaluate whether the two above-mentioned biopesticides are effective in controlling the incidence of gray mold in commercial vineyards and whether their efficacy depends on the disease intensity in untreated controls. We assessed the efficacy on incidence at the population-level (i.e., average across all trials) and for each trial separately in order to analyze the between-trial variability of the treatment efficacy. We also rigorously analyze the levels of uncertainty associated with our estimates using a Bayesian approach. Finally, we describe the range of plausible levels of treatment efficacy for a new vineyard (out of sample) taking into account the observed variability of the treatment effect in our dataset.

## 5.3 Materials and Methods

### 5.3.1 Data Description and Experimental Design

The Vinopôle Bordeaux-Aquitaine managed an on-farm research network on the grapevine, called RESAQ VitiBio, to evaluate the efficacy of biopesticides against *Botrytis* bunch rot in organic production. Two biopesticides, potassium bicarbonate (Armicarb®) and *A. pullulans* (Botector®) were tested in 23 and 26 trials, respectively, over three years (2014 to 2016) and nine locations. Each trial, defined as a unique combination of location and vintage, was composed of two replicates of two strips, one strip including untreated vines and the other including vines treated with one biopesticide. A pair of strips

(biopesticide and the control) constitutes a replicate. Within each strip, four plots including 10 consecutive vine stocks (*i.e.*, 5m<sup>2</sup>) were used for measurements of incidence and intensity. Trials were distributed throughout Southwestern France, a major wine-producing area. Therefore, this study covers typical conditions of the oceanic climate in this area with commonly used cultivars, management techniques and production targets. Armicarb® is composed of 85% potassium bicarbonate and interacts directly with the pathogen by killing mycelium and spores while disrupting pH and osmotic pressure. The first treatment was applied at veraison (change of berry color and accumulation of sugar), and up to two additional treatments were applied two weeks before harvest. Botector® is composed of two fungus strains of *Aureobasidium pullulans*, DSM 14940 and DSM 14941, competing for nutrients and space with the pathogen. Potassium bicarbonate was applied 1 to 3 times during ripening according to weather and the risk of *Botrytis cinerea* development. The use of *A. pullulans* required a mandatory application just before the bunch closure (Westover, 2018) and up to two applications during ripening. During ripening, biopesticides were applied after heavy rainfall (>20mm) or consecutive days of rain (3mm per day) as it creates favorable conditions for *Botrytis cinerea* development. Treatment applications were stopped on 1 and 3 days before the harvest for potassium bicarbonate and *A. pullulans*, respectively. As a prophylactic method, some leaves were removed during the growing season for all the trials, as frequently done in organic systems (Weigle and Carroll, 2014).

At the end of each growing season, few days before the cultivar-specific harvest date, *Botrytis* bunch rot incidence and intensity were assessed by visually rating 50 grape bunches per plot. The incidence represents the percentage of bunches with symptoms (*i.e.*,

the number of bunches with symptoms divided by 50). The intensity (also called severity in the literature) represents the percentage of diseased berries (*i.e.*, the sum of rotten berries per bunch divided by 50), and it is indicative of disease pressure. Disease intensity is a variable with high interest as it reflects the impact on wine yield and gustative quality (Hill et al., 2010; Ky et al., 2012).

### 5.3.2 Data Analysis

The biopesticides bicarbonate potassium and *A.pullulans* were assessed separately as they were not tested side-by-side. For each type of biopesticides, two statistical models, called GLRM\_0 and GLRM\_Int, were used:

The model GLRM\_0, with trial and block as random effects, is defined as (Agresti, 2002; Makowski et al., 2014):

$$Y_{ijkl} \sim \text{BetaBinomial}(n, \pi_{ijk}, \phi) \quad (1)$$

$$\text{logit}(\pi_{ijk}) = \mu_0 + \omega_i + \gamma_{j\{i\}} + (\mu_T + \theta_i + \tau_{j\{i\}})X_k \quad (2)$$

$$\omega_i \sim \mathcal{N}(0, \sigma_\omega^2) \text{ iid}$$

$$\gamma_{j\{i\}} \sim \mathcal{N}(0, \sigma_\gamma^2) \text{ iid}$$

$$\theta_{ik} \sim \mathcal{N}(0, \sigma_\theta^2) \text{ iid}$$

$$\tau_{ijk} \sim \mathcal{N}(0, \sigma_\tau^2) \text{ iid}$$

where  $i$  is the on-farm trial index (1, ..., 23 or 26),  $j$  is the block index (1,2),  $k$  is the treatment index (1 for control, 2 for biopesticide),  $l$  is the plot index (1,2,3,4),  $n$  is equal to 50 and represents the number of bunches evaluated in the  $i^{\text{th}}$  on-farm trial,  $j^{\text{th}}$  block,  $k^{\text{th}}$



treatment for the  $l^{th}$  plot,  $\pi_{ijk}$  is the probability that a bunch is affected by botrytis in the  $i^{th}$  on-farm trial and the  $j^{th}$  block for the  $k^{th}$  treatment,  $Y_{ijkl}$  is the number of botrytis-affected bunches in a sample 50 bunches,  $\phi$  is an overdispersion parameter (see the section “statistical inference” below for more information),  $X_k$  is a binary variable equal to one if a biopesticide treatment was applied on the  $l^{th}$  sample in the  $j^{th}$  block of the  $i^{th}$  on-farm trial and equal to zero otherwise,  $\mu_0$  is the mean value of the logit of  $\pi_{ijk}$  in the untreated vineyard over all on-farm trials,  $\mu_T$  is the mean biopesticide effect over on-farm trials,  $\omega_i$  is the random trial effect,  $\gamma_{j\{i\}}$  is the random block effect nested within the  $i^{th}$  on-farm trial,  $\theta_i$  is the random interaction between the  $i^{th}$  on-farm trial and the biopesticide effect,  $\tau_{j\{i\}}$  is the random interaction between the  $j^{th}$  block nested with the  $i^{th}$  on-farm trial and the biopesticide effect.

According to this model, the global incidence (*i.e.*, proportion of affected fruits based on mean parameter values) is expressed as  $I_0 = \frac{e^{\mu_0}}{1 + e^{\mu_0}}$  without biopesticide and  $I_T = \frac{e^{\mu_0 + \mu_T}}{1 + e^{\mu_0 + \mu_T}}$  with biopesticide. The incidence in the  $i^{th}$  on-farm trial and block  $j^{th}$  is expressed as  $I_{0ij} = \frac{e^{\mu_0 + \omega_i + \gamma_{j\{i\}}}}{1 + e^{\mu_0 + \omega_i + \gamma_{j\{i\}}}}$  without biopesticide and  $I_{Tij} = \frac{e^{\mu_0 + \mu_T + \omega_i + \gamma_{j\{i\}} + \theta_i + \tau_{j\{i\}}}}{1 + e^{\mu_0 + \mu_T + \omega_i + \gamma_{j\{i\}} + \theta_i + \tau_{j\{i\}}}}$  with biopesticide.

Based on these quantities, several standard measures of treatment efficacy can be derived, both at the global average and block within-trial levels. The risk ratio  $\frac{I_T}{I_0}$  (or  $\frac{I_{Tij}}{I_{0ij}}$  at the block within-trial level) is expressed as the ratio of the proportion of affected fruits with biopesticide to the proportion of affected fruits without. The disease control efficacy (hereafter called CE) on incidence (%), expressed as  $100 \left(1 - \frac{I_T}{I_0}\right)$ , describes the percentage

of reduction in incidence among treated bunches to untreated bunches. It is expressed as  $100 \left( 1 - \frac{I_{Tij}}{I_{0ij}} \right)$  for the block within trial  $ij$ . The odds ratio is another standard metric commonly used to assess treatment efficacy. The global Odds ratio is expressed as  $e^{\mu_T}$  and is equal to the ratio of the Odds of disease with and without biocontrol, *i.e.*,  $\frac{I_T}{1-I_T} / \frac{I_0}{1-I_0}$ . The specific Odds ratio for the block within-trial  $ij$  is equal to  $\frac{I_{Tij}}{1-I_{Tij}} / \frac{I_{0ij}}{1-I_{0ij}}$ . Risk ratio and Odds ratio lower than one indicates disease reduction with biocontrol.

The model (2) can be updated to take disease pressure (measured here by the intensity at the control) into account. To do so, eq.(2) should be replaced by

$$\text{logit}(\pi_{ijk}) = \mu_0 + \beta_1 Z_{j\{i\}} + \omega_i + \gamma_{j\{i\}} + (\mu_T + \theta_i + \tau_{j\{i\}} + \beta_2 Z_{j\{i\}}) X_k \quad (3)$$

The model description remains the same as in (2) except for the variable  $Z_{j\{i\}}$  corresponding to the disease pressure measured by the intensity (%) in the untreated control of the  $j^{\text{th}}$  block of the  $i^{\text{th}}$  on-farm trial.  $\beta_1$  is the regression parameter representing the effect of a 1% intensity increase on the disease incidence in the control.  $\beta_2$  is a regression parameter for the interaction between the intensity at the control and the biopesticide effect. With this model, the effect of a 1% intensity increase on the disease incidence in the treated vine is equal to  $\beta_1 + \beta_2$ . With this model, it is also possible to measure the efficacy of the treatment through the computation of risk ratio, CE, and Odds ratio. However, with the model based on eq.(3), these quantities are not constant but depend on the value of  $Z_{j\{i\}}$ , *i.e.*, on the disease pressure. Indeed, according to eq.(3), the proportion of affected fruits in the  $i^{\text{th}}$  on-farm trial and block  $j^{\text{th}}$  is expressed as

$\frac{e^{\mu_0 + \beta_1 Z_{j\{i\}} + \omega_i + \gamma_{j\{i\}}}}{1 + e^{\mu_0 + \beta_1 Z_{j\{i\}} + \omega_i + \gamma_{j\{i\}}}}$  without biocontrol and  $\frac{e^{\mu_0 + \mu_T + \beta_1 Z_{j\{i\}} + \beta_2 Z_{j\{i\}} + \omega_i + \gamma_{j\{i\}} + \theta_i + \tau_{j\{i\}}}}{1 + e^{\mu_0 + \mu_T + \beta_1 Z_{j\{i\}} + \beta_2 Z_{j\{i\}} + \omega_i + \gamma_{j\{i\}} + \theta_i + \tau_{j\{i\}}}}$  with

biocontrol. The Odds ratio is then expressed as  $e^{\mu_T + \beta_2 Z_{j\{i\}} + \theta_i + \tau_{j\{i\}}}$  and, thus, decreases or increases as a function of  $Z_{j\{i\}}$  depending on the sign of  $\beta_2$ .

### 5.3.3 Statistical Inference

Models based on (2) and (3) are further denoted to as GLRM\_0 and GLRM\_Int. They were fitted using a Bayesian approach with the R package *brms* (Bürkner 2017). The priors for the variance of the random effects were defined as a truncated Student's t distribution with 3 degrees of freedom, zero lower bound and a 2.5 scale. The prior for the overdispersion was defined as a gamma distribution,  $\phi \sim \text{Gamma}(0.01, 0.01)$ . Posterior distributions for  $\mu_0$ ,  $\mu_T$ ,  $\beta_1$ , and  $\beta_2$  were computed using the No-U-Turn Samplers (Hoffmann and Gelman, 2014) as implemented in the Stan software (Stan Development Team, 2018). The convergence of the MCMC chains was checked with R packages *coda* and *brms* (Plummer et al., 2006; Bürkner, 2017). We ran four independent chains for 3,000 iterations and the metric *Rhat*, which compares the between- and within-chain estimates for model parameters, were between 1 and 1.02 which indicates convergence. Several other variants of GLRM were fitted and evaluated using the Widely Applicable Information Criterion (WAIC) values. The WAIC is a more general criterion than Akaike Information Criteria (Watanabe, 2013). As these alternative models showed a higher WAIC and less precise parameter estimates (*i.e.*, wider credibility intervals), we did not select them. In particular, the overdispersion was found to be strong (Table 1), as frequently observed in plant disease epidemiology. Indeed, in many cases, Binomial models exhibit overdispersion

with a variance much higher than the one predicted by this type of model (Bolker et al., 2009). Overdispersion can occur due to technical variation related to error measurements from the experimental design, missing covariates, an excess frequency of zero, and biological variation between the subjects of interest. More information is given in the supplementary material (see Supplementary Material Tables S2 and S3). Overdispersion can arise when models have been incorrectly specified by failing to include important predictors or by using other link functions (Richards, 2007; Harrison, 2014). Not accounting for overdispersion in a model can result in biased parameter estimates. One way to deal with overdispersion is to use the Beta-Binomial model (Harrison, 2015; McElreath, 2020), which includes an overdispersion term and assumes that each binomial count observation has its own probability of success.

At the on-farm network level, several outputs were computed from the samples representing posterior distributions, such as the posterior median and 95% credibility intervals of incidence under-treated and untreated conditions, risk ratio, Odds ratio, and CE. The 95% credibility interval was directly computed from the samples drawn from the posterior distribution and used to describe uncertainty. We also computed 95% predictive intervals of the same quantities in order to describe plausible ranges of values for a new situation, similar to those included in our sample assuming similar conditions (Higgins et al., 2009; IntHout et al., 2016).

## 5.4 Results

### 5.4.1 Description of the Observed Proportions of Diseased Bunches

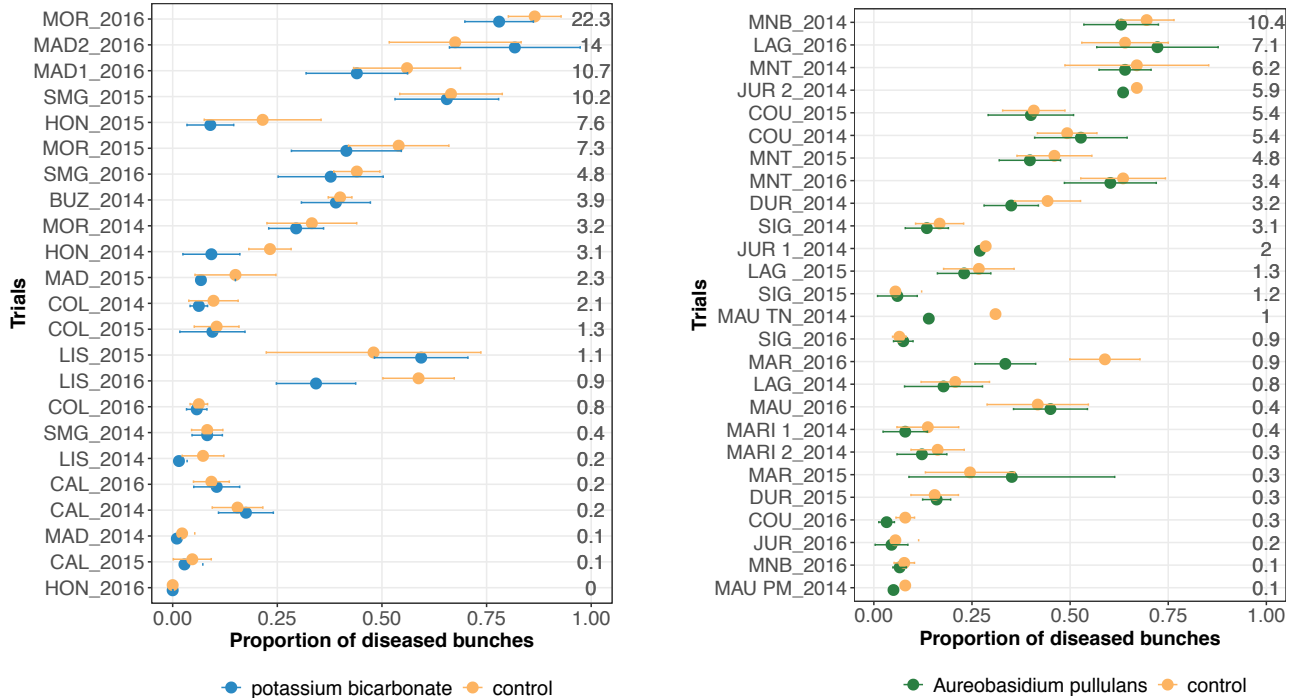


Figure 1: Measured disease incidences in treated bunches with potassium bicarbonate (blue dots, on the left) or with *Aureobasidium pullulans* (green dots, on the right) and in untreated bunches (yellow points) for each trial (site-year). Horizontal bars indicate 95% frequentist confidence intervals. Mean intensity in the control is displayed for each trial on the right. Trials are ranked by increasing mean intensity in the untreated control (%).

Mean observed disease incidences in treated and untreated bunches varied between trials for both biopesticides (Figure 1). The mean intensity at the control per trial varied from 0 to 22.3% and from 0 to 10.4% for the on-farm network testing bicarbonate potassium and *A. pullulans*, respectively. About a third of the trials had a mean intensity in the control lower than 5%.

Mean disease incidences in untreated bunches are mostly higher than the mean disease incidence in treated bunches except for MAD2\_2016, LIS\_2015 (Figure 1, on the left), MAR\_2015, and LAG\_2016 (Figure 1, on the right). Most of the trials showed overlapping confidence intervals, between the biopesticide treatment and the untreated control, for the mean incidences (Figure 1). However, in LIS\_2016 (Figure 1, on the left) and MAR\_2016 (Figure 1, on the left), the mean incidence in untreated bunches was substantially higher than the disease incidence observed in treated bunches.

#### **5.4.2 Statistical Modeling Evaluation**

Regarding the WAIC values (Table 1), results were in favor of the model including disease pressure as a covariable (GLRM\_Int) (i.e., GLRM\_Int got a lower WAIC than GLRM\_0). We also performed an efficient approximate Leave-one-out Cross-validation (LOO) which was in favor of GLRM\_Int for both biopesticides. The graphical analysis of the observed vs. expected incidences derived from the posterior predictive distributions for GLRM\_Int showed a better fit compared to GLRM\_0 (Figure 2). Nevertheless, the model GLRM\_0 remains relevant for estimating the global level of efficacy (averaged over all levels of disease pressure) of the biopesticides, especially because its WAIC is higher only by a narrow margin.

Table 1: Widely Applicable Information Criterion (WAIC), parameter estimates and 95% credible intervals for GLRM\_0 and GLRM\_Int for potassium bicarbonate and *Aureobasidium pullulans*.

Model	Parameter estimates and WAIC	Potassium bicarbonate	<i>Aureobasidium pullulans</i>
GLRM_0	WAIC	2023	2275
	$\mu_0$	-1.28 [-2.09; -0.48]	0.96 [-1.48; -0.44]
	$\mu_T$	-0.32 [-0.56; -0.07]	-0.17 [-0.35; -0.00]
	$\sigma_\omega$	1.85 [1.34; 2.62]	1.26 [0.93; 1.72]
	$\sigma_\gamma$	0.19 [0.01; 0.41]	0.22 [0.02; 0.43]
	$\sigma_\theta$	0.23 [0.02; 0.48]	0.10 [0.01; 0.27]
	$\sigma_\tau$	0.27 [0.04; 0.46]	0.24 [0.07; 0.37]
	$\phi$	18.96 [14.66; 24.11]	25.37 [19.71; 32.37]
GLRM_Int	WAIC	1935	2249
	$\mu_0$	-1.79 [-2.47; -1.12]	-1.24 [-1.71; -0.79]
	$\mu_T$	-0.16 [-0.47; 0.17]	0 [-0.23; 0.24]
	$\beta_1$	0.12 [0.09; 0.15]	0.11 [0.07; 0.16]
	$\beta_2$	-0.04 [-0.08; -0.01]	-0.06 [-0.12; -0.01]
	$\sigma_\omega$	1.51 [1.04; 2.17]	1.10 [0.80; 1.50]
	$\sigma_\gamma$	0.21 [0.01; 0.44]	0.17 [0.01; 0.39]
	$\sigma_\theta$	0.30 [0.04; 0.56]	0.16 [0.01; 0.37]
	$\sigma_\tau$	0.32 [0.13; 0.50]	0.26 [0.11; 0.40]
	$\phi$	30.33 [22.59; 40.26]	29.08 [22.06; 37.95]

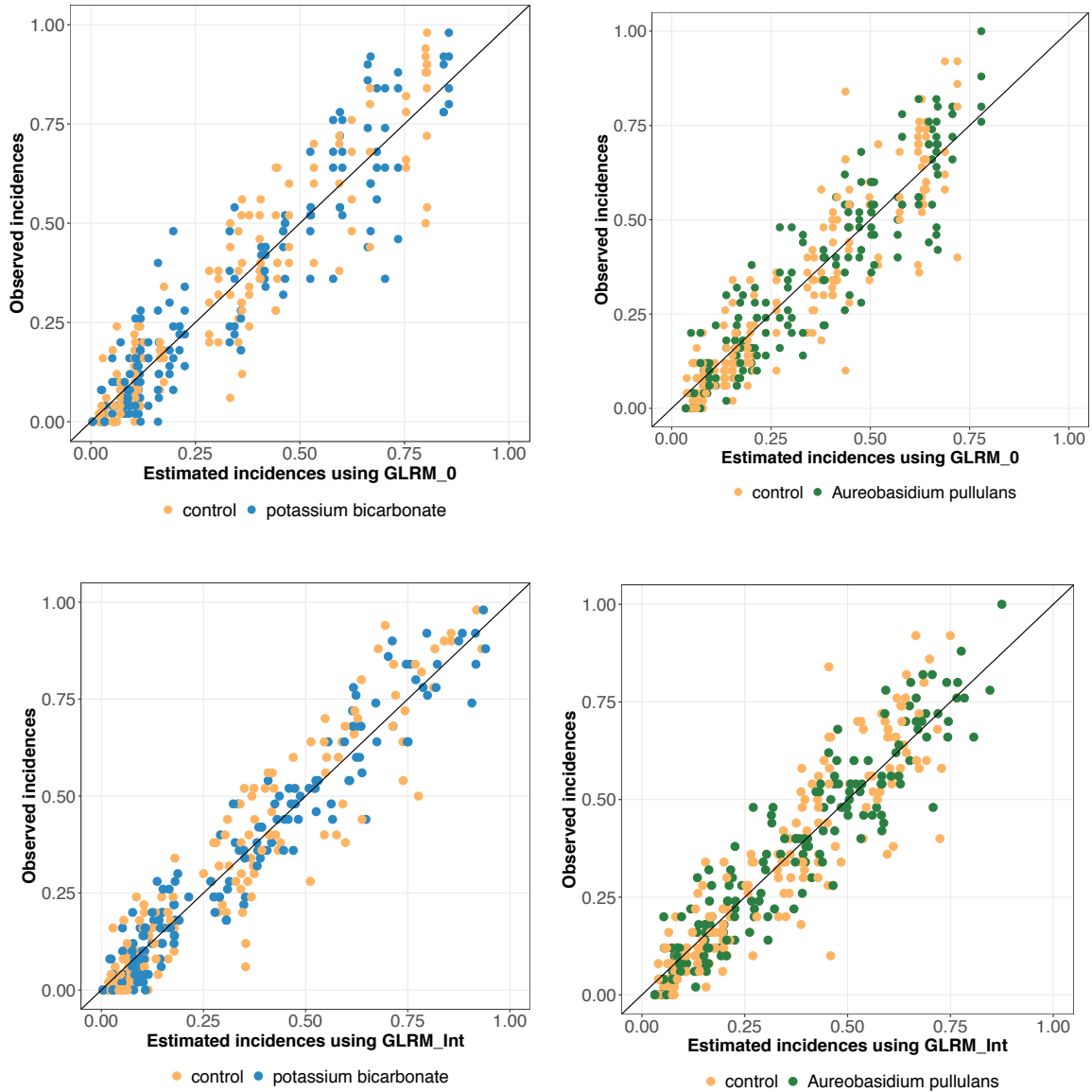


Figure 2: Observed vs. estimated incidences (posterior means) derived from the generalized linear multilevel model including disease pressure (GLRM\_Int). Incidence for bunches treated with potassium bicarbonate or *Aureobasidium pullulans* are represented by blue or green dots respectively. Incidence for untreated bunches are represented by yellow dots. One dot represents one replicate.



### 5.4.3 Estimated Efficacy of the Biopesticides

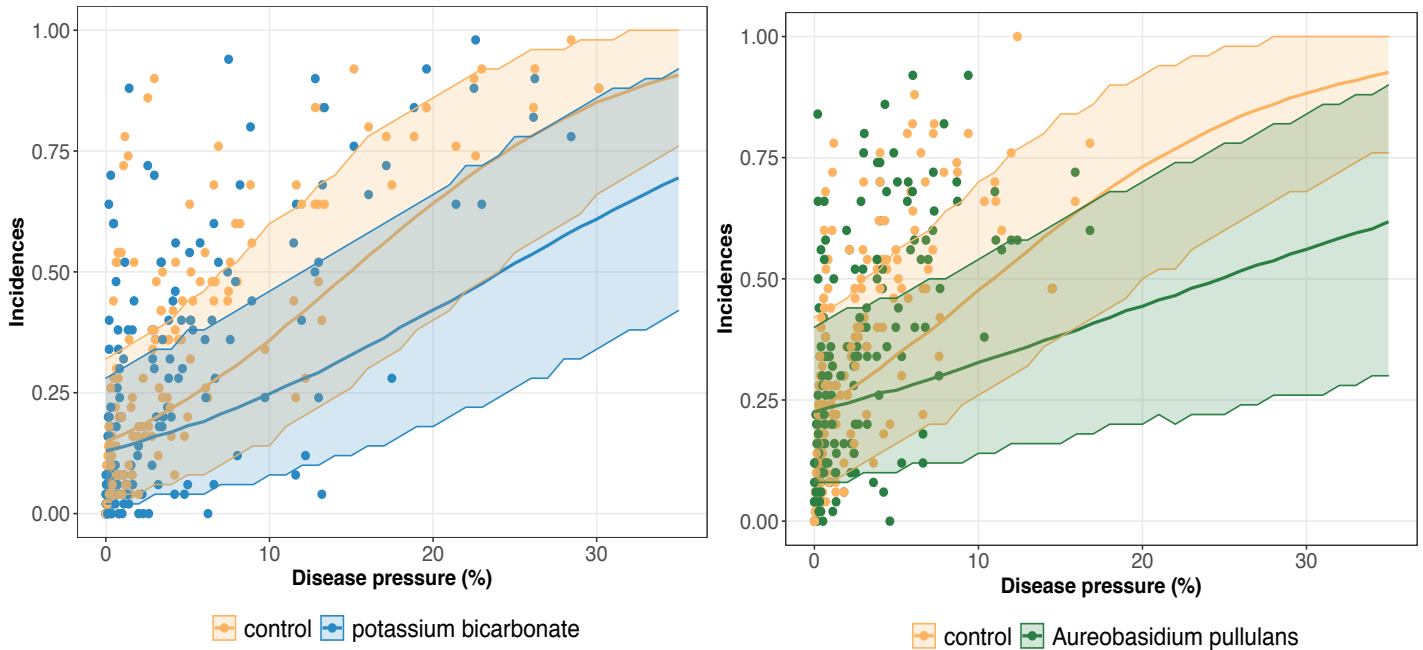


Figure 3: Estimated global disease incidence in bunches treated with potassium bicarbonate (blue line) or *Aureobasidium pullulans* (green line) and untreated bunches (yellow line) as a function of the disease pressure (disease intensity in the untreated control) and their corresponding 95% credibility band (blue, green and yellow curves and shadows). Blue and yellow dots represent the observed incidences in treated and untreated bunches, respectively.

The results of model GLRM\_Int revealed that global disease incidences (i.e., across all trials) increase as a function of the disease pressure (Figure 3). The estimated disease incidences in untreated vineyards were higher in both networks regardless of the disease pressure level, although the uncertainty was large (Figure 3). For low disease pressure, no substantial difference was found between the incidence in the treated and untreated vineyards. For higher disease pressure, the gap between treated and untreated vineyards increased, showing a reduction in incidence when a biopesticide is applied. For disease intensity levels between 0% and 10%, the incidence is lower when using potassium

bicarbonate than *A.pullulans*. For example, when the intensity at the control reaches 10%, the incidence with potassium bicarbonate is equal to 0.25 while the incidence with *A.pullulans* is equal to 0.32. There is a high heterogeneity in the measured incidences (dots on Figure 3) resulting from a strong between- and within-trial variability. The variability on the incidence between trials is illustrated on Figures 4 and 5 for potassium bicarbonate and *A.pullulans*, respectively. To simplify the visualization of the results, only a subset of trials is presented but all the trials are displayed in the Supplementary Material (see Figures S1-4 in the Supplementary Material). Each trial has different incidences as a function of the disease pressure. For example, at disease intensities of 10%, the estimated incidence in treated bunches with potassium bicarbonate were 18% in trial HON\_2014 and 61% in trial SMG\_2015 (Figure 4, on the left) while the global estimated incidence was equal to 25% (Figure 3 on the left). When the intensity in the control was 10%, the estimated incidence in treated bunches with *A.pullulans* was 53% in trial COU\_2014 and 44% in trial DUR\_2014 (Figure 5, on the left), while the global estimated incidence was 32% (Figure 3 on the right).

A risk ratio lower than one indicates disease reduction with biocontrol. Above a threshold of 15-20% of intensity, the individual risk ratio reached one as intensity increased (Figures 4 and 5). The 95% credible band of the risk ratio is wide when the observed incidences are widely spread (Figure 4 on the right, trial called CAL\_2014) or when there is no observed values for a specific range of intensity (Figure 4 on the right, trial called MOR\_2016 for intensity lower than 10%).

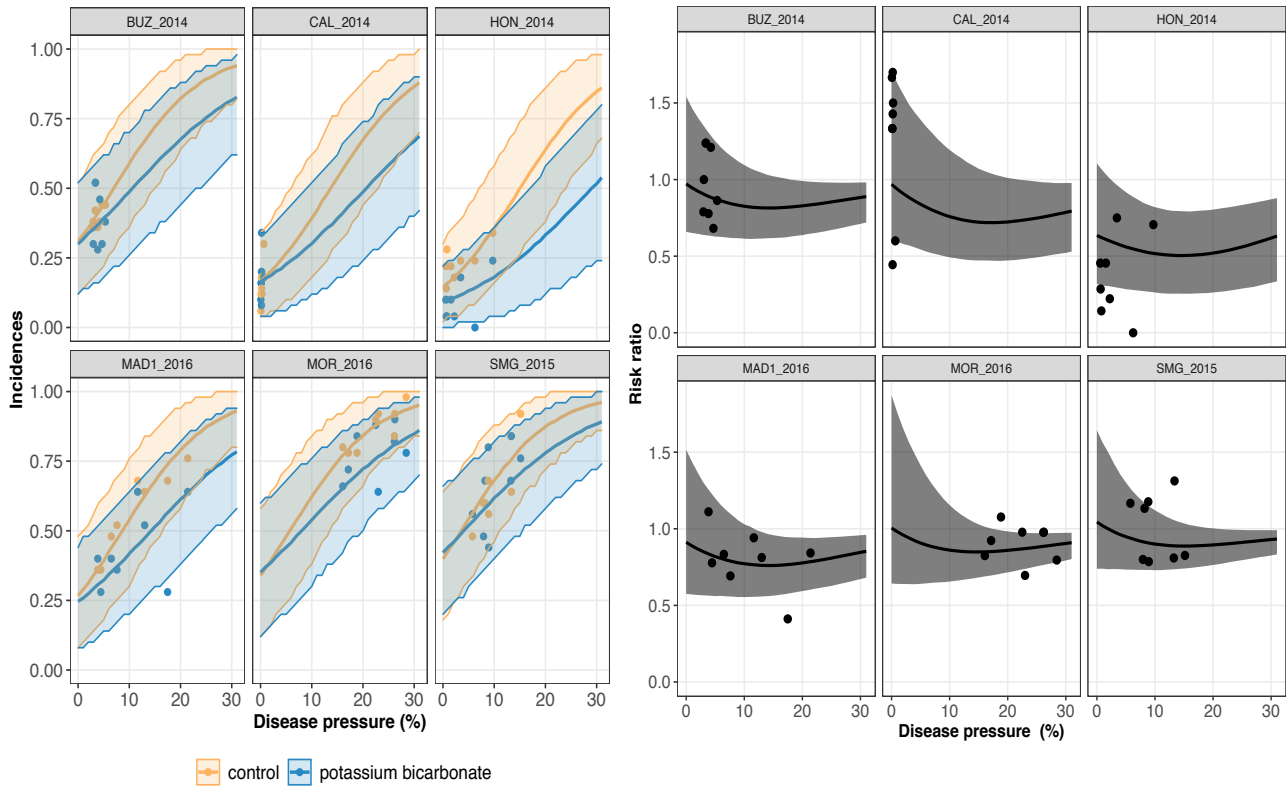


Figure 4: On the left, trial-specific estimated incidences in bunches treated with potassium bicarbonate (blue line) and untreated bunches (yellow line) as a function of the disease pressure (disease intensity in the untreated control) and their corresponding 95% credible bands (blue and yellow shadows) for six different trials (out of 23). Blue and yellow dots represent the observed incidences in treated and untreated bunches, respectively. On the right, the corresponding estimated risk ratios (black line) with their 95% credible bands (grey shadow). Observed risk ratios are represented by black crosses and observed disease pressure levels are represented by black dots.

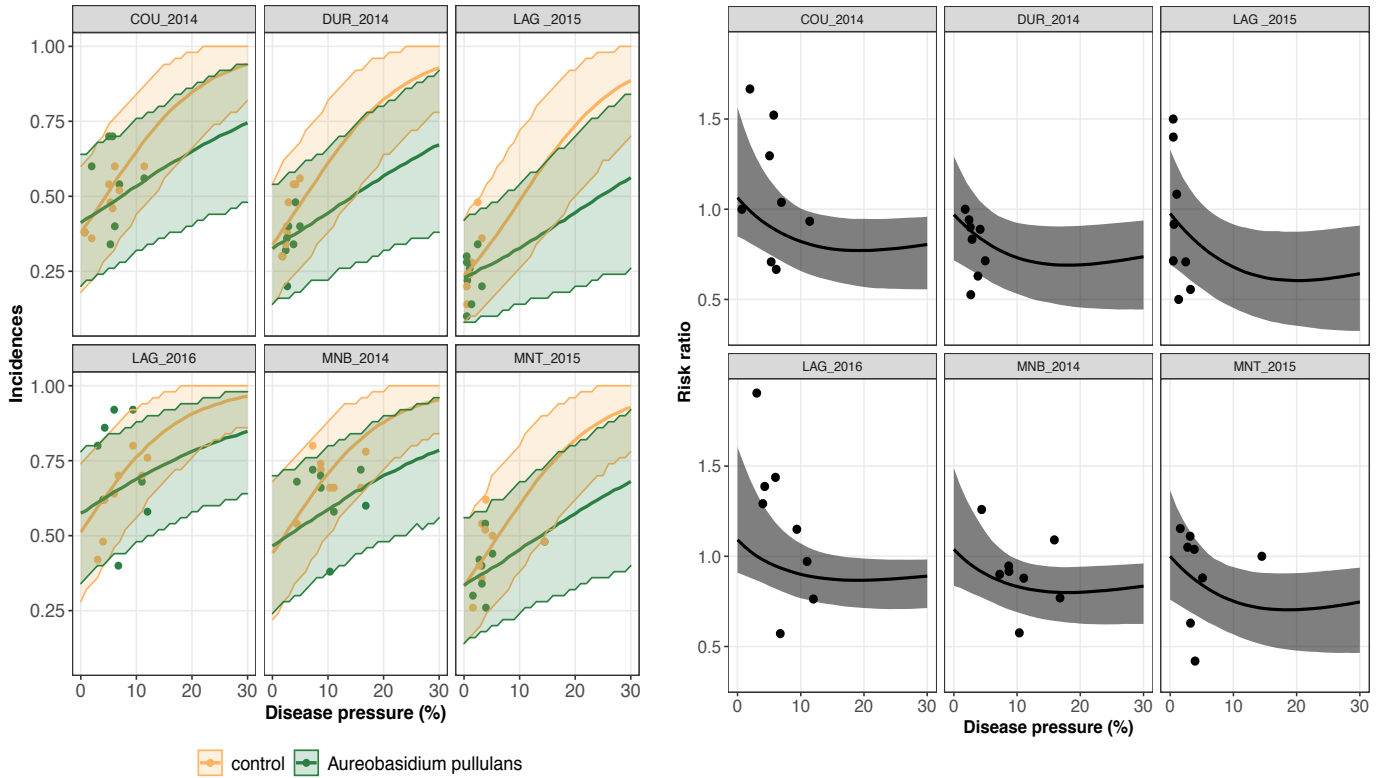


Figure 5: On the left, trial-specific estimated incidences in bunches treated with *Aureobasidium pullulans* (green line) and untreated bunches (yellow line) as a function of the disease pressure (disease intensity in the untreated control) and their corresponding 95% credible bands (green and yellow shadows) for six different trials (out of 26). Green and yellow dots represent the observed incidences in treated and untreated bunches, respectively. On the right, the corresponding estimated risk ratios (black line) with their 95% credible bands (grey shadow). Observed risk ratios are represented by blue crosses and observed disease pressure levels are represented by black dots.

Table 2: Estimated values of risk ratio (RR) and odds ratio (OR) obtained with GLRM\_0 and GLRM\_Int (for three levels of disease intensity in the control). Values in bold indicate the lowest between the two biopesticides considered.

Model	Treatment efficacy	Estimated values [95% CI] for potassium bicarbonate	Estimated values [95%CI] for <i>A. pullulans</i>
GLRM_0	RR	<b>0.78 [0.63; 0.94]</b>	0.88 [0.77; 0.99]
	OR	<b>0.73 [0.57; 0.93]</b>	0.84 [0.71; 0.99]

Table 2: Continued

Model	Treatment efficacy	Estimated values [95% CI] for potassium bicarbonate	Estimated values [95%CI] for <i>A. pullulans</i>
GLRM_Int	RR 2%	<b>0.82 [0.64; 1.05]</b>	0.90 [0.79; 1.05]
	RR 10%	0.69 [0.52; 0.86]	<b>0.68 [0.51; 0.87]</b>
	RR 15%	0.65 [0.47; 0.84]	<b>0.62 [0.42; 0.86]</b>
	OR 2%	<b>0.79 [0.59; 1.06]</b>	0.88 [0.72; 1.07]
	OR 10%	0.58 [0.41; 0.80]	<b>0.53 [0.34; 0.80]</b>
	OR 15%	0.48 [0.29; 0.73]	<b>0.39 [0.19; 0.73]</b>

Odds ratio and risk ratio at the network level are in favor of potassium bicarbonate according to the results obtained with model GLRM\_0 (Table 2). Based on model GLRM\_Int, results are also in favor of potassium bicarbonate for low disease pressure, i.e., 2%. But for higher disease pressure levels, values tend to be in favor of *A. pullulans* but with large credible intervals revealing high uncertainty.

The global disease efficacy on incidence for bicarbonate potassium was twice as important as *A. pullulans* according to the estimated values derived from GLRM\_0 (Figure 6). The overall disease control efficacy on incidence was equal to 22.3 [5.7; 36.7] and to 11.7 [0.06; 22.7] for potassium bicarbonate and *A. pullulans*, respectively. With the

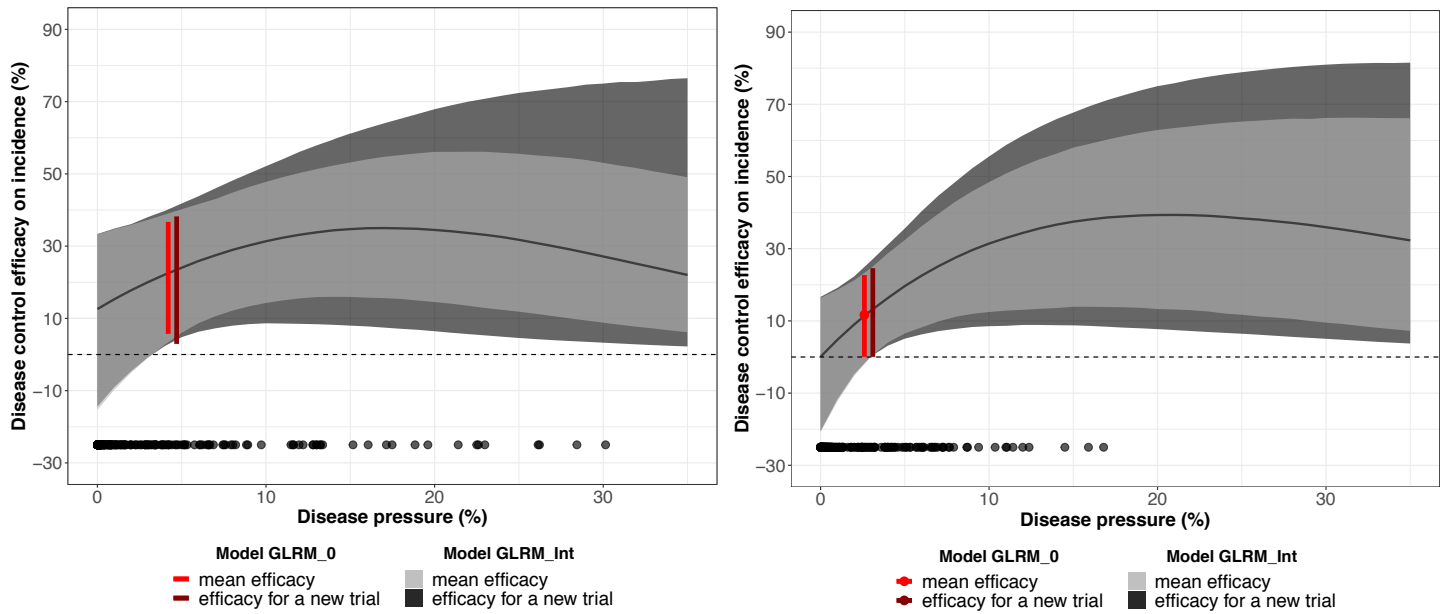


Figure 6: Mean disease control efficacy on incidence and its 95% credible interval (in light grey) estimated from the network testing the potassium bicarbonate (on the left) and *Aureobasidium pullulans* (on the right). The dark grey shadow represents 95% predictive intervals (plausible ranges for a new trial). The red vertical interval represents the mean disease control efficacy for a generalized linear multilevel model without covariate, and its x-axis represents the disease pressure (disease intensity in the untreated control). Black dots represent the observed disease pressure for all the replicates.

GLRM\_Int model, the global level of efficacy is higher with bicarbonate potassium than with *A. pullulans* for a disease pressure level between 0% and 10%. Above 10%, the global level of efficacy is higher with *A. pullulans*, but the uncertainty is large as most of the measured intensity at the control is smaller than 10%. At 17% and 21% of intensity, the global level of efficacy reached the highest value equal to 34.9% [15.6; 54.7] and 39.3% [13.2; 63.4] for potassium bicarbonate than *A. pullulans*, respectively, and then decreased. The uncertainty around the global level of efficacy is narrower for potassium bicarbonate as the range of observed intensity is larger (from 0% to 31%) while the range of observed intensity for *A.*

*pullulans* was equal to 0% to 17%. Thus, for intensities higher than 17%, outputs for *A. pullulans* can be seen as an extrapolation and results should be interpreted with caution. The efficacy for a new trial describes a plausible range of efficacy on incidence for a new vineyard (out-of-sample) conducted under similar conditions. For an intensity at the control higher than 10%, the efficacy for a new trial is larger than the 95% credible interval at the network level, meaning a substantial uncertainty regarding the efficiency of these biopesticides for a new trial.

## 5.5 Discussion

In the present study, we assessed the efficacy on the incidence of two biopesticides, *A. pullulans* and potassium bicarbonate, to an untreated control for the treatment of a grape disease (*B. cinerea*) through an on-farm research network managed using organic practices. Our assessment confirmed that the data were better described by including the intensity at the control as a covariable.

Our results are in favor of potassium bicarbonate as its efficacy on incidence at the network level is higher for intensities between 0% and 10% than for *A. pullulans* and for intensities higher than 10% the credible band is narrower for potassium bicarbonate than *A. pullulans*. Regarding the treatment application, potassium bicarbonate costs 45€/ha and *A. pullulans* costs 80€/ha for recommended doses (Caboulet et al., 2020). The predictive distribution for a new trial allows us to assess the heterogeneity and uncertainty in the treatment effect and represents a more appropriate treatment effect than the 95% credible interval at the population level.

Our results are consistent with those from Rotolo et al. (2018) as they found efficacies on incidence ranging from 7.6% to 23% for some biopesticides under high

disease intensity and, more specifically, the efficacy on incidence for *A. pullulans* was equal to 11%. The main difference is that 11 sprays were applied (off-label spray numbers for research purposes) while only 3 were applied in our trials. In another study conducted by Calvo-Garrido et al. (2019), the efficacy on incidence reached 18% and 17% for *A. pullulans* with 5 or 6 applications for an intensity at the control equal to 17% and 13, respectively. Our results do not agree with the study conducted by Pertot et al. (2017b) as an efficacy on incidence with *A. pullulans* was equal to 75%. This high value was explained by the relatively low-medium level of the disease and an optimal timing of application allowing *A. pullulans* to survive until harvest at concentrations sufficient to prevent *B. cinerea*. Comparisons against other studies should be interpreted with caution as they came from environments where indicators like temperature and the relative humidity differed, which impact the establishment of living organisms such as *A. pullulans* (Nicot, 2011). Unfortunately, similar studies regarding potassium bicarbonate on the grapevines in field conditions are not available.

Our results show a lower efficacy on incidence compared to the biopesticides combined mentioned in the literature (Supplementary Material, Table S1) (O'Neill et al., 1996; Elmer et al., 2005; Cañamás et al., 2011; Pertot et al., 2017b) and this might suggest the presence of publication bias (Rothstein et al., 2005). Those experiments might have an advantage in terms of application timing while our trials were managed in real conditions by winegrowers who were having a first experience regarding the application of biopesticides. As a result, it is possible that the biopesticides were not applied in a timely fashion as farmers experienced lack of knowledge, technical or management issues, or unfavorable climatic conditions. As mentioned by Pertot et al. (2017b), the application



timing is a crucial step and *A.pullulans* should be applied at veraison, when sugar starts to increase, as it consumes the sugar needed for *B. cinerea* to grow and colonize grapevine wounds. The first application of potassium bicarbonate appears to be crucial, and if the fungus becomes visible, it will be difficult to control it (Wenneker and Kanne, 2009); thus, potassium bicarbonate should be used as a preventive method. Regarding synthetic chemical fungicides used against *B. cinerea*, Gabriolotto et al. (2009) compared the effectiveness of five different combinations of synthetic active ingredients including boscalid, pyrimethanil, cyprodinil, fludioxonil, fenhexamid, and iprodione to an untreated control in two commercial vineyards. Results showed an efficacy on disease incidence ranging from 23% to 73% with a mean of 56% using the synthetic active ingredients. Despite a high control of *B. cinerea* using synthetic chemical fungicides, there is a risk of resistance development and residue accumulation, as mentioned by their analysis.

Our results covered a range of intensity at the control before harvest from 0% to 16.8% and 0% to 30.1% for bicarbonate potassium and *A. pullulans*, respectively. Those values are in accordance with previous studies where the values of intensity ranged, all trials combined, between 4.3% to 20.3% (Reglinski et al., 2005; Cañamás et al., 2011; Aziz et al., 2016; Pertot et al., 2017b; Calvo-Garrido et al., 2019) except for Rotolo et al. (2018) who reported 30.5% and 53.3% of intensity at the control before harvest in two different trials.

The winegrowers part of this network use commercially available field equipment and standard management practices apart from the biopesticide treatment. Analyzing the trials from different vineyards together using a generalized linear multilevel model allowed us to understand the overall efficacy of the tested biopesticides besides the efficacy at the

trial level. Also, the uncertainty was estimated to provide a range of plausible values of efficacy and help decision making about biopesticides. In addition to the overall efficacy, the efficacy for a new trial provides a more complete picture of the range of plausible effects of the tested biopesticides in a new environment (*i.e.*, not initially included in the network of trials) (Laurent et al., 2020).

Our data showed a substantial between-trial variability (Table 1), which can be explained by the diversity of agronomic management, pedoclimatic conditions, the sensitivity of grapes varieties to *B. cinerea*, and vine density.

Prior to the arrival of biopesticides on the market for controlling *B.cinerea*, organic vine producers only benefited from prophylactic methods (*e.g.*, leaf or bunch removal). *B.cinerea* can lead to a strong qualitative depreciation on wines from only 5% of Botrytis-affected grapes that cause degradation of color, aroma, and structure (Ky et al., 2012), which reinforces the need to provide additional and reliable management practice for organic producers. Biopesticides represent a promising tool in Integrated Pest Management as they can be combined with chemical fungicides to reduce the risk of fungicide resistance and high residue levels. Indeed, the chemical can be applied early in the season (*e.g.*, flowering, pre-closure, and veraison), while biopesticides can be planned for later before harvesting (Rotolo et al., 2018).

In conclusion, the on-farm trial network confirmed a partial efficacy of the bicarbonate potassium when the disease pressure does not exceed 20%. The efficacy on disease incidence for *A.pullulans* is lower and the uncertainty is relatively high. Compared to the inorganic compound potassium bicarbonate, *A.pullulans* is a living organism and needs to colonize and survive in an environment with fluctuating climatic conditions

(Bardin et al., 2015) that can affect its efficacy. Cost applications are also in favor of the use of potassium bicarbonate. The biopesticide based on potassium bicarbonate represents a promising management practice for controlling *B. cinerea* in organic grapevine production. In the wine-growing area of our study, the interest for testing management practices through an on-farm research network is increasing, thus our proposed approach should be useful in future research efforts.

### 5.6 Acknowledgments

The authors are thankful to many winegrowers and RESAQ network's members: Chambres d'Agriculture Dordogne, Landes, Pyrénées Atlantiques and Lot-et-Garonne, Agrobio Périgord, ALTEMA Madiran, and Vini Vitis Bio.

### 5.7 References

- Agresti, A. 2002. Categorical Data Analysis. John Wiley & Sons, Inc., Hoboken, NJ, USA.
- Amichot, M., P. Joly, F. Martin-Laurent, D. Siaussat, and A.-V. Lavoit. 2018. Biocontrol, new questions for Ecotoxicology? Environmental Science and Pollution Research 25: 33895–33900.
- Aziz, A., B. Verhagen, M. Magnin-Robert, M. Couderchet, C. Clément, et al. 2016. Effectiveness of beneficial bacteria to promote systemic resistance of grapevine to gray mold as related to phytoalexin production in vineyards. Plant Soil 405(1–2): 141–153. doi: 10.1007/s11104-015-2783-z.
- Bardin, M., S. Ajouz, M. Comby, M. Lopez-Ferber, B. Graillot, et al. 2015. Is the efficacy of biological control against plant diseases likely to be more durable than that of chemical pesticides? Front. Plant Sci. 6. doi: 10.3389/fpls.2015.00566.
- Bolker, B.M., M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, et al. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution 24(3): 127–135. doi: 10.1016/j.tree.2008.10.008.
- Bürkner, P.-C. 2017. brms : An R Package for Bayesian Multilevel Models Using Stan. J. Stat. Soft. 80(1). doi: 10.18637/jss.v080.i01.

Butault, J.-P., N. Delame, F. Jacquet, and G. Zardet. 2011. L'utilisation des pesticides en France : état des lieux et perspectives de réduction. Centre d'études et de prospective.

Caboulet, D., M. Guisset, R. Ribera, L. Berard-Delay, and J. Thiery. 2020. Le coût des fournitures en viticulture et oenologie. 49th ed. Chambre D'Agriculture des Pyrénées-Orientales.

Calvo-Garrido, C., J. Roudet, N. Aveline, L. Davidou, S. Dupin, et al. 2019. Microbial Antagonism Toward Botrytis Bunch Rot of Grapes in Multiple Field Tests Using One *Bacillus ginsengihumi* Strain and Formulated Biological Control Products. *Front. Plant Sci.* 10: 105. doi: 10.3389/fpls.2019.00105.

Cañamás, T.P., I. Viñas, R. Torres, J. Usall, C. Solsona, et al. 2011. Field applications of improved formulations of *Candida sake* CPA-1 for control of *Botrytis cinerea* in grapes. *Biological Control* 56(2): 150–158. doi: 10.1016/j.biocontrol.2010.11.007.

Chen, M., F. Brun, M. Raynal, and D. Makowski. 2019. Timing of Grape Downy Mildew Onset in Bordeaux Vineyards. *Phytopathology*® 109(5): 787–795. doi: 10.1094/PHYTO-12-17-0412-R.

Chen, M., F. Brun, M. Raynal, and D. Makowski. 2020. Delaying the first grapevine fungicide application reduces exposure on operators by half. *Sci Rep* 10(1): 6404. doi: 10.1038/s41598-020-62954-4.

Coleman, M.D., J.D. O'Neil, E.K. Woehrling, O.B.A. Ndunge, E.J. Hill, et al. 2012. A Preliminary Investigation into the Impact of a Pesticide Combination on Human Neuronal and Glial Cell Lines In Vitro (P. Mukhopadhyay, editor). *PLoS ONE* 7(8): e42768. doi: 10.1371/journal.pone.0042768.

Elmer, P.A.G., S.M. Hoyte, J.L. Vanneste, T. Reglinski, P.N. Wood, et al. 2005. Biological control of fruit pathogens. *New Zealand Plant Protection* 58: 47–54.

EPA. 2016. What are Biopesticides? <https://www.epa.gov/ingredients-used-pesticide-products/what-are-biopesticides>.

Gabrielotto, C., M. Monchiero, M. Nègre, D. Spadaro, and M.L. Gullino. 2009. Effectiveness of control strategies against *Botrytis cinerea* in vineyard and evaluation of the residual fungicide concentrations. *Journal of Environmental Science and Health, Part B* 44(4): 389–396. doi: 10.1080/03601230902801117.

Haidar, R., M. Fermaud, C. Calvo-Garrido, J. Roudet, and A. Deschamps. 2016. Modes of action for biological control of *Botrytis cinerea* by antagonistic bacteria. *Phytopathologia Mediterranea* (Preprint Articles). doi: 10.14601/Phytopathol\_Mediterr-18079.

Harrison, X.A. 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ* 2: e616. doi: 10.7717/peerj.616.

Harrison, X.A. 2015. A comparison of observation-level random effect and Beta-Binomial models for modelling overdispersion in Binomial data in ecology & evolution. *PeerJ* 3: e1114. doi: 10.7717/peerj.1114.

Higgins, J.P.T., S.G. Thompson, and D.J. Spiegelhalter. 2009. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172(1): 137–159. doi: 10.1111/j.1467-985X.2008.00552.x.

Hill, G.N., R.M. Beresford, and K.J. Evans. 2010. Tools for accurate assessment of botrytis bunch rot (*Botrytis cinerea*) on wine grapes. *NZPP* 63: 174–181. doi: 10.30843/nzpp.2010.63.6560.

Hoffmann, M., and A. Gelman. 2014. The No-U-Turn Sampler: Adaptively Setting Path Lengthsin Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 15(1): 1351–1381.

Hossard, L., L. Guichard, C. Pelosi, and D. Makowski. 2017. Lack of evidence for a decrease in synthetic pesticide use on the main arable crops in France. *Science of The Total Environment* 575: 152–161. doi: 10.1016/j.scitotenv.2016.10.008.

IntHout, J., J.P.A. Ioannidis, M.M. Rovers, and J.J. Goeman. 2016. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open* 6(7): e010247. doi: 10.1136/bmjopen-2015-010247.

Jacometti, M.A., S.D. Wratten, and M. Walter. 2010. Review: Alternatives to synthetic fungicides for *Botrytis cinerea* management in vineyards. *Australian Journal of Grape and Wine Research* 16(1): 154–172. doi: 10.1111/j.1755-0238.2009.0067.x.

Komarek, M., E. Cadkova, V. Chrastny, F. Bordas, and J.-C. Bollinger. 2009. Contamination of vineyard soils with fungicides: A review of environmental and toxicological aspects. *Environment International* 36(1): 138–151. doi: 10.1016/j.envint.2009.10.005.

Ky, I., B. Lorrain, M. Jourdes, G. Pasquier, M. Fermaud, et al. 2012. Assessment of grey mould (*Botrytis cinerea*) impact on phenolic and sensory quality of Bordeaux grapes, musts and wines for two consecutive vintages: *Botrytis cinerea* effect on grape and wine quality. *Australian Journal of Grape and Wine Research* 18(2): 215–226. doi: 10.1111/j.1755-0238.2012.00191.x.

Laurent, A., F. Miguez, P. Kyveryga, and D. Makowski. 2020. Going beyond mean effect size: Presenting prediction intervals for on-farm network trial analyses. *European Journal of Agronomy* 120: 126127. doi: 10.1016/j.eja.2020.126127.

Mailly, F., L. Hossard, J.-M. Barbier, M. Thiollet-Scholtus, and C. Gary. 2017. Quantifying the impact of crop protection practices on pesticide use in wine-growing systems. *European Journal of Agronomy* 84: 23–34.

Makowski, D., A. Vicent, M. Pautasso, G. Stancanelli, and T. Rafoss. 2014. Comparison of statistical models in a meta-analysis of fungicide treatments for the control of citrus black spot caused by *Phyllosticta citricarpa*. *Eur J Plant Pathol* 139(1): 79–94. doi: 10.1007/s10658-013-0365-6.

McElreath, R. 2020. *Statistical rethinking: a Bayesian course with examples in R and Stan*. 2nd ed. Taylor and Francis, CRC Press, Boca Raton.

Nicot, P.C. 2011. Classical and augmentative biological control against diseases and pests: critical status analysis and review of factors influencing their success. IOBC/WPRS, S.I.

Nicot, P.C., A. Stewart, M. Bardin, and Y. Elad. 2016. Chapter 9 Biological Control and Biopesticide Suppression of Botrytis-Induced Diseases. In: Fillinger, S. and Elad, Y., editors, *Botrytis – the Fungus, the Pathogen and its Management in Agricultural Systems*. Springer International Publishing, Cham. p. E1–E1

O’Neill, T.M., Y. Elad, D. Shtienberg, and A. Cohen. 1996. Control of Grapevine Grey Mould with *Trichoderma harzianum* T39. *Biocontrol Science and Technology* 6(2): 139–146. doi: 10.1080/09583159650039340.

Pertot, T. Caffi, V. Rossi, L. Mugnai, C. Hoffmann, et al. 2017a. A critical review of plant protection tools for reducing pesticide use on grapevine and new perspectives for the implementation of IPM in viticulture. *Crop Protection* 97: 70–84. doi: 10.1016/j.cropro.2016.11.025.

Pertot, I., O. Giovannini, M. Benanchi, T. Caffi, V. Rossi, et al. 2017b. Combining biocontrol agents with different mechanisms of action in a strategy to control *Botrytis cinerea* on grapevine. *Crop Protection* 97: 85–93. doi: 10.1016/j.cropro.2017.01.010.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* 6: 7–11.

Reglinski, T., P.A.G. Elmer, J.T. Taylor, F.J. Parry, R. Marsden, et al. 2005. Suppression of Botrytis bunch rot in Chardonnay grapevines by induction of host resistance and fungal antagonism. *Austral. Plant Pathol.* 34(4): 481. doi: 10.1071/AP05057.

RESAQVitiBio. 2015. Demie Journée Technique: Visite des Essais Biocontrôle - Pourriture Grise.

Richards, S.A. 2007. Dealing with overdispersed count data in applied ecology: Overdispersed count data. *Journal of Applied Ecology* 45(1): 218–227. doi: 10.1111/j.1365-2664.2007.01377.x.

Rothstein, H.R., A.J. Sutton, and M. Borenstein. 2005. *Publication Bias in Meta-analysis: Prevention, Assessment and Adjustments*. John Wiley & Sons, Ltd, Chichester, UK.

Rotolo, C., R.M. De Miccolis Angelini, C. Dongiovanni, S. Pollastro, G. Fumarola, et al. 2018. Use of biocontrol agents and botanicals in integrated management of *Botrytis cinerea* in table grape vineyards: Management of *Botrytis cinerea* in table-grape vineyards. *Pest. Manag. Sci* 74(3): 715–725. doi: 10.1002/ps.4767.

Stan Development Team. 2018. *Stan Modeling Language Users Guide and Reference Manual*.

Watanabe, S. 2013. A Widely Applicable Bayesian Information Criterion. *Journal of Machine Learning Research* 14: 867–897.

Weigle, T., and J. Carroll. 2014. *Production Guide for Organic Grapes*. New York State Integrated Pest Management Program (224).

Wenneker, M., and J. Kanne. 2009. Use of potassium bicarbonate (Armicarb) on the control of powdery mildew (*Sphaerotheca mors-uvae*) of gooseberry (*Ribes uva-crispa*). *Communications in agricultural and applied biological sciences* 75(4): 563–568.

Westover, F. 2018. Grapevine Phenology Revisited: Using growth stages to improve vineyard management. *Wines & Vines* 99(3).  
<https://winesvinesanalytics.com/features/article/196082/Grapevine-Phenology-Revisited> (accessed 4 May 2020).

## 5.8 Supplemental Materials

Table S1: Summary of previous published studies evaluating the efficacy of biopesticides to control *Botrytis cinerea* on grapevine based on outdoor experiments.

Reference	Experiment	Active ingredient(s)	Efficacy	Comments
<b>Pertot et al., 2017b</b>	<ul style="list-style-type: none"> <li>field trials (commercial vineyards) in 3 locations in Italy for 4 years.</li> <li>RCBD with 3 or 4 reps (minimum 8 vines per rep)</li> <li>20 bunches/rep for assessing symptoms</li> </ul>	<ul style="list-style-type: none"> <li><i>Trichoderma atroviride</i> applied at 'berries beginning to touch' stage</li> <li><i>Aureobasidium pullulans</i> (Botector®) applied at the 'beginning of ripening' stage</li> <li><i>Bacillus subtilis</i> applied 20 and 7 days before harvesting</li> <li><i>T.atroviride+B.subtilis+A.pullulans</i> (combined strategy)</li> </ul>	<ul style="list-style-type: none"> <li>mean Ei*: 75% to 85%</li> <li>mean Es*: 85% to 95%</li> <li>efficacy with single biopesticide comparable to a combined strategy</li> </ul>	<ul style="list-style-type: none"> <li>high level of efficacy can be explained with the relatively low-medium level of the disease</li> <li>biopesticide applied at the specific stage fully controlled the disease (they survived until harvest at concentrations sufficient to prevent <i>B. cinerea</i>)</li> </ul>
<b>Rotolo et al., 2018</b>	<ul style="list-style-type: none"> <li>two field trials conducted in two table grape vineyards</li> <li>RCBD with 4 reps. One plot = 12 vines</li> <li>11 sprays (off-label spray numbers for research purpose)</li> <li>150-200 bunches/plot for assessing symptoms</li> </ul>	<ul style="list-style-type: none"> <li><i>Bacillus subtilis</i></li> <li><i>Bacillus amyloliquefaciens</i></li> <li><i>Aureobasidium pullulans</i> (Botector®)</li> </ul>	<ul style="list-style-type: none"> <li>mean Ei ranged from 7.6 to 23%</li> <li>mean Es ranged from 19% to 50%</li> </ul>	<ul style="list-style-type: none"> <li>biopesticides not effective (&lt;30%) under high disease pressure compared to alternate use of biopesticide with chemical fungicide</li> </ul>
<b>Calvo-Garrido et al., 2019</b>	<ul style="list-style-type: none"> <li>4 vineyard sites between 2015 to 2017</li> <li>leaf removal in 2016 and 2017 only</li> <li>RCBD 4 or 5 reps</li> <li>two different application strategies: "full season" strategy = 5 or 6 applications and "late season"= after veraison and following determination of a Disease Risk Index</li> </ul>	<ul style="list-style-type: none"> <li><i>Bacillus subtilis</i> QST713</li> <li><i>Bacillus amyloliquefaciens</i></li> <li><i>Aureobasidium pullulans</i> (Botector®)</li> <li><i>Ulicladium oudemansii</i></li> <li><i>Bacillus subtilis</i> IAB/BS03</li> </ul>	<ul style="list-style-type: none"> <li>Es for <i>B. amyloliquefaciens</i> ranged from 37 to 58%</li> <li>Es for <i>A.pullulans</i> ranged from -24 to 48%</li> <li>Es for <i>B. subtilis</i> QST713 ranged from -43 to 54%</li> <li>Es for <i>B. subtilis</i> IAB/BS03 ranged from -17 to 25%</li> <li>Es for <i>U. oudemansii</i> ranged from -59 to 47%</li> </ul>	<ul style="list-style-type: none"> <li><i>pullulans</i> significantly reduced intensity in one out of 4 sites-years</li> <li><i>B. subtilis</i> QST713 significantly reduced the intensity in 4 out of 6 sites-years</li> <li><i>B. amyloliquefaciens</i> significantly reduced the intensity in 3 out of 5 site-years</li> </ul>



Table S1: Continued

Reference	Experiment	Active ingredient(s)	Efficacy	Comments
<b>Aziz et al., 2016</b>	<ul style="list-style-type: none"> <li>• 2 vineyards in 2006 in France</li> <li>• canopy management was applied</li> <li>• RCBD with 12 plants per plot and 3 reps</li> <li>• intensity and incidence were performed on 50-60 clusters</li> <li>• treatments were applied twice in July (individually and in mixture)</li> </ul>	<ul style="list-style-type: none"> <li>• <i>Bacillus subtilis</i></li> <li>• <i>Pseudomonas fluorescens</i></li> <li>• <i>Pantoea agglomerans</i></li> <li>• <i>P. fluorescens</i> + <i>P. agglomerans</i></li> <li>• <i>P. fluorescens</i> + <i>B. subtilis</i></li> <li>• <i>P. agglomerans</i> + <i>B. subtilis</i></li> </ul>	<ul style="list-style-type: none"> <li>• Ei and Es for <i>P. fluorescens</i> were equal to 40.3% and 78%, respectively</li> <li>• Ei and Es for <i>P. agglomerans</i> were equal to 18.2% and 72%, respectively</li> <li>• Ei and Es for <i>B. subtilis</i> were equal to 43.6% and 87%, respectively</li> <li>• Ei ranged from 53.1% to 67.6% for mixtures</li> <li>• Es ranged from 78% to 93% for mixtures</li> </ul>	<ul style="list-style-type: none"> <li>• combination of two bacteria which were least effective when used alone (<i>P. fluorescens</i> and <i>P. agglomerans</i>) provided the best protection</li> </ul>
<b>Elmer et al., 2005</b>	<ul style="list-style-type: none"> <li>• 2 vineyards over three growing season</li> <li>• RCBD with 4 or 6 reps (depending on the year)</li> <li>• treatment applied at 5% and 90% flowering</li> </ul>	<ul style="list-style-type: none"> <li>• <i>Ulocladium oudemansii</i> (BOTRY-Zen®)</li> </ul>	<ul style="list-style-type: none"> <li>• efficacy was equal to 92%, 83% and 81% in 2003, 2004 and 2005, respectively (Ei or Es was not mentioned)</li> </ul>	
<b>Reglinski et al., 2005</b>	<ul style="list-style-type: none"> <li>• one field trial (commercial vineyard) in New Zealand</li> <li>• RBD</li> <li>• 10 bunches/vines and 5 vines/treatment</li> </ul>	<ul style="list-style-type: none"> <li>• <i>Ulocladium oudemansii</i></li> </ul>	<ul style="list-style-type: none"> <li>• Ei = 78%</li> </ul>	<ul style="list-style-type: none"> <li>• <i>U. oudemansii</i> aggressively suppressed the growth and development of <i>B. cinerea</i> on necrotic tissue</li> </ul>

Table S1: Continued

Reference	Experiment	Active ingredient(s)	Efficacy	Comments
O'Neill et al., 1996	<ul style="list-style-type: none"> <li>• 133 field experiments in 9 commercial vineyards (19 countries) between 1988 and 1994.</li> <li>• RBD with 4-6 reps</li> <li>• 3 to 30 plants per plot</li> <li>• 4 applications (end of flowering, closure bunches, veraison and 2-3 weeks after veraison)</li> <li>• incidence assessed on 50-100 bunches/plot</li> </ul>	<ul style="list-style-type: none"> <li>• <i>Trichoderma harzianum</i> T39</li> </ul>	<ul style="list-style-type: none"> <li>• on average, Ei = 36 +/- 2.7%</li> <li>• average disease incidence in the untreated plots of all experiments = 42 +/- 2.3%</li> </ul>	<ul style="list-style-type: none"> <li>• if disease incidence in untreated plot &gt; 60%, biopesticide resulted in a 20% reduction</li> </ul>
Magnin-Robert et al., 2013	<ul style="list-style-type: none"> <li>• research vineyard treated over 3 years.</li> <li>• RCBD with 5-12 plants/plot and 3 reps</li> <li>• measurement performed with 100 clusters/treatment</li> <li>• biopesticide applied in mixture only</li> </ul>	<ul style="list-style-type: none"> <li>• <i>Acinetobacter lwoffii</i> (two strains)= AL</li> <li>• <i>B. subtilis</i> (1 strain)= Bs</li> <li>• <i>P. agglomerans</i> (two strains)= PA</li> <li>• <i>P. fluorescens</i> (two strains)= PF</li> </ul>	<ul style="list-style-type: none"> <li>• treatments with mixtures of both strains of <i>P. agglomerans</i> had a leaf protection** equal to 53%</li> <li>• treatments combining <i>Acinetobacter</i> spp. and <i>Pseudomonas</i> spp. = poor leaf protection (AL+PF, AL+Bs+PF, or AL+PA+PF+Bs)</li> <li>• treatments combining AL+PA+PF and all other combinations strongly protected the leaves: from 45 to 75%</li> </ul>	<ul style="list-style-type: none"> <li>• treatments with bacterial mixtures were shown to induce systemic resistance against <i>B. cinerea</i> in the second year</li> <li>• efficacy and duration of such a disease control seemed to be reinforced on year three without renewal of bacterial treatments</li> </ul>
Cañamás et al., 2011	<ul style="list-style-type: none"> <li>• field trials in 2005 and 2006 (Spain)</li> <li>• RCBD with 4 reps</li> <li>• 5-7 vines per replication</li> <li>• 5 sprays: flowering, pea sized, veraison, 21 days before harvest and 7 days before harvest</li> </ul>	<ul style="list-style-type: none"> <li>• <i>Candida sake</i> CPA-1 (trying different formulations)</li> <li>• one fungicide</li> </ul>	<ul style="list-style-type: none"> <li>• Ei for all <i>C. sake</i> treatments ranged between 36% and 40%</li> </ul>	<ul style="list-style-type: none"> <li>• no significant difference between the <i>C. sake</i> treatments and the conventional fungicide in field trials</li> </ul>

\*E<sub>i</sub> = efficacy based on incidence (percentage of bunches with *B. cinerea* rot symptoms) ; E<sub>s</sub> = efficacy based on intensity (percentage of *B. cinerea* rotten berries per bunch)  
 RBD= randomized block design ; RCBD = randomized complete block design  
 \*\* leaf protection = reduction of the lesion size relative to the control (%)

Below, please find the R code and Figures 1-4

Required R packages and corresponding version

```
library(brms) # version 2.13.0
# You do need to install the R package rstanarm to run brms
```

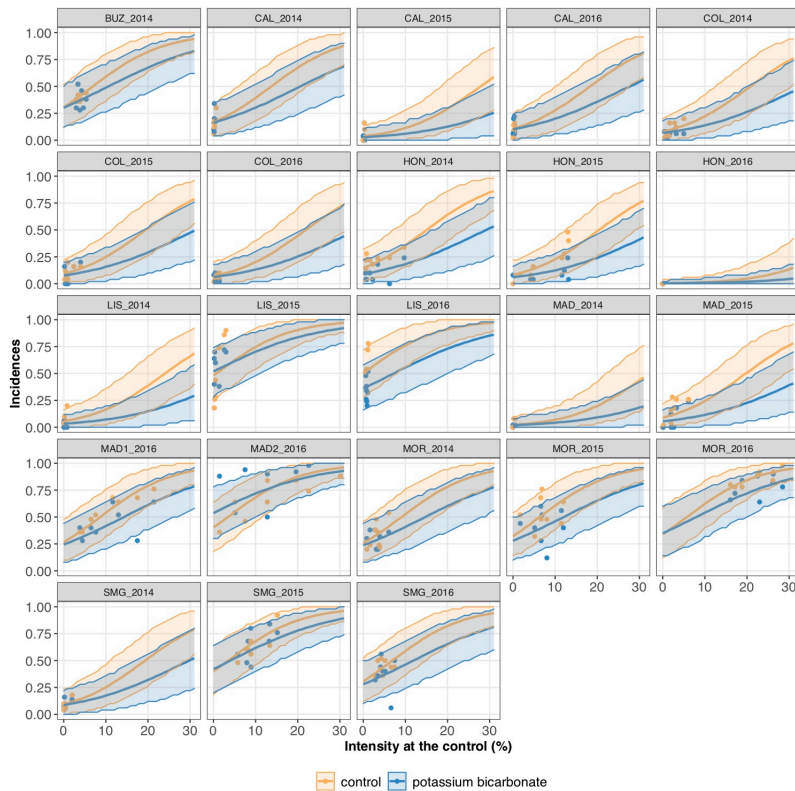


Figure S1: Individual incidence in treated bunches with potassium bicarbonate (blue line) and untreated bunches (yellow line) as a function of the intensity at the control and their corresponding 95% credible band (blue and yellow shadows). Blue and yellow dots represent the observed incidences in treated and untreated bunches, respectively.

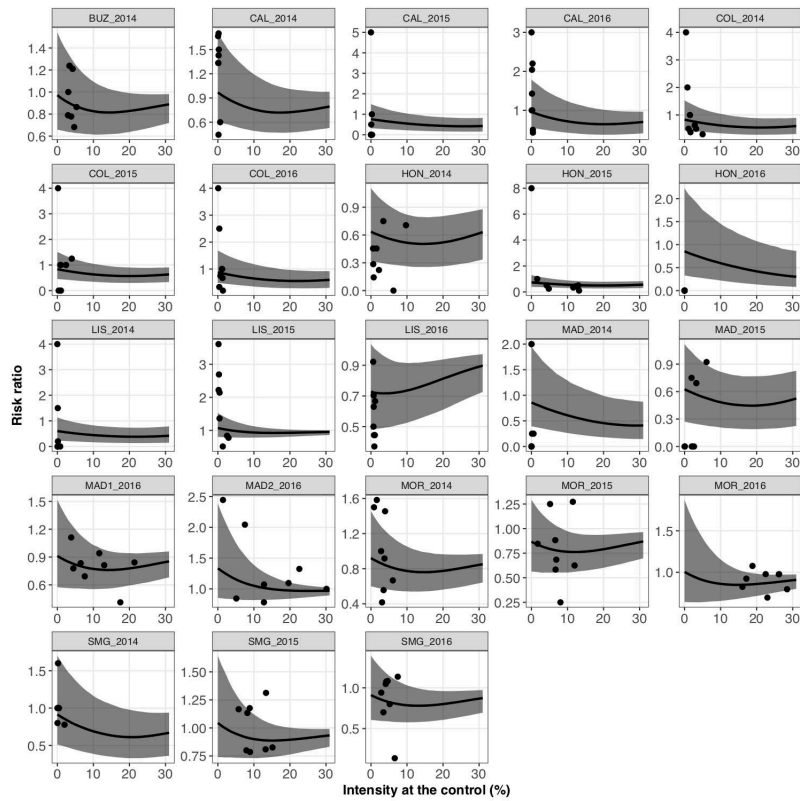


Figure S2: Individual risk ratios with potassium bicarbonate (black line) with their 95% credible bands (grey shadow) as a function of the intensity at the control. Observed risk ratios are represented by black dots.

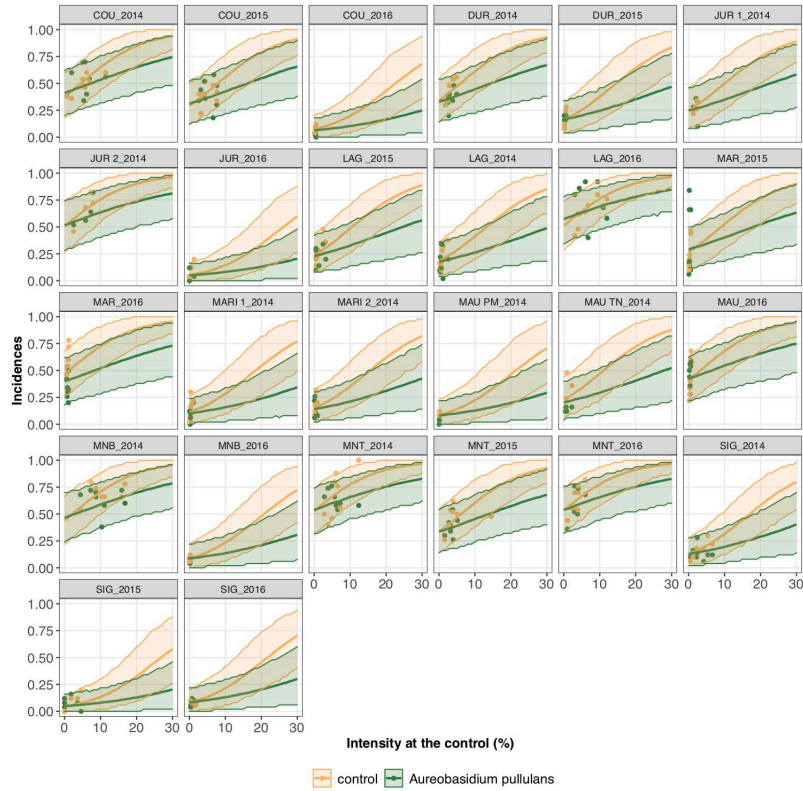


Figure S3: Individual incidence in treated bunches with *Aureobasidium pullulans* (green line) and untreated bunches (yellow line) as a function of the intensity at the control and their corresponding 95% credible band (green and yellow shadows). Green and yellow dots represent the observed incidences in treated and untreated bunches, respectively.

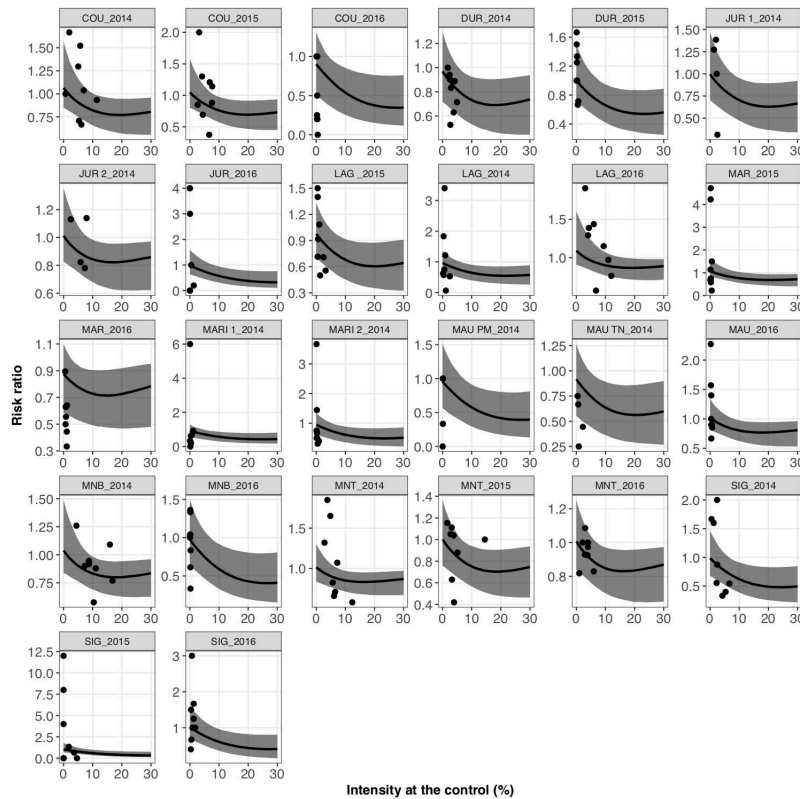


Figure S4: Individual risk ratios with *Aureobasidium pullulans* (black line) with their 95% credible bands (grey shadow) as a function of the intensity at the control. Observed risk ratios are represented by black dots."

#### Models comparison:

```
GLRM_0 <- brm(ndisbunches | vint(nbunches) ~ TRT +
  (1 | Trial_ID/Bloc) +
  (1 | Trial_ID:TRT) + (1 | Trial_ID:TRT:Bloc),
  data = dataA, family = beta_binomial2,
  iter = 3000,
  cores = 4, seed = 1234, stanvars = stanvars,
  control = list(adapt_delta = 0.95))
```

```
GLRM_Int <- brm(ndisbunches | vint(nbunches) ~ TRT + CTR_Int +
  TRT:CTR_Int + (1 | Trial_ID/Bloc) +
  (1 | Trial_ID:TRT) + (1 | Trial_ID:TRT:Bloc),
  data = dataA, family = beta_binomial2,
  cores = 4, seed = 1234, stanvars = stanvars,
  control = list(adapt_delta = 0.90))
```

```
GLRM_3 <- brm(ndisbunches | trials(nbunches) ~ TRT + CTR_Int +
  TRT:CTR_Int + (1 | Trial_ID/Bloc) +
  (1 | Trial_ID:TRT) + (1 | Trial_ID:TRT:Bloc),
  data = dataA, family = binomial,
  cores = 4, seed = 1234,
```

```
control = list(adapt_delta = 0.9))
```

```
GLRM_4 <- brm(ndisbunches | vint(nbunches) ~ TRT + CTR_Int +
  TRT:CTR_Int + (1 | Trial_ID/Bloc),
  data = dataA, family = beta_binomial2,
  cores = 4, seed = 1234, stanvars = stanvars,
  control = list(adapt_delta = 0.90))
```

```
GLRM_5 <- brm(ndisbunches | vint(nbunches) ~ TRT + CTR_Int +
  TRT:CTR_Int + (1 | Trial_ID/Bloc) +
  (1 | Trial_ID:TRT) ,
  data = dataA, family = beta_binomial2,
  cores = 4, seed = 1234, stanvars = stanvars,
  control = list(adapt_delta = 0.90))
```

GLRM\_0 and GLRM\_Int are the models used in our manuscript to analyze the datasets.

GLRM\_3 is similar to GLRM\_Int but with a binomial family, meaning the overdispersion is not taking into account.

GLRM\_4 has no random interactions (i.e. (1 | Trial\_ID:TRT) and (1 | Trial\_ID:TRT:Bloc) are missing).

GLRM\_5 is similar to GLRM\_Int but with no random interaction between the block nested with trial and the treatment (i.e. (1 | Trial\_ID:TRT:Bloc) is missing).

Table S2: Efficient approximate leave-one-out cross-validation for potassium bicarbonate.

model	elpd_diff	se_diff	WAIC
<b>GLRM_Int</b>	0.0	0.0	1935
<b>GLRM_5</b>	-4.3	4.6	1946
<b>GLRM_4</b>	-23.4	8.6	1986
<b>GLRM_0</b>	-42.9	11.1	2023
<b>GLRM_3</b>	-105.3	22.9	2134

Table S3: Efficient approximate leave-one-out cross-validation for *Aureobasidium pullulans*.

<b>model</b>	<b>elpd_diff</b>	<b>se_diff</b>	<b>WAIC</b>
<b>GLRM_Int</b>	0.0	0.0	2249
<b>GLRM_5</b>	-5.7	4.7	2262
<b>GLRM_4</b>	-11.3	8.0	2275
<b>GLRM_0</b>	-12.6	6.6	2275
<b>GLRM_3</b>	-122.7	29.6	2481

We compared the five fitted models (see Tables 2 and 3), for each biocontrol agent, by estimating the difference in their expected predictive accuracy by the difference in the expected log predictive density (elpd\_loo).

A negative elpd\_diff favors the first model.

We also computed the Widely Applicable Information Criterion (WAIC). A lower WAIC indicates a better fit.



## **CHAPTER 6. COMPARISONS OF CORN AND SOYBEAN YIELD RESPONSE IN SMALL-PLOT AND ON-FARM FOLIAR FUNGICIDE TRIALS**

Modified from a paper to be submitted to Agronomy Journal

Anabelle Laurent<sup>1</sup>, David Makowski<sup>2</sup>, Peter Kyveryga<sup>3</sup>,

Alison Robertson<sup>4</sup>, Fernando Miguez<sup>1</sup>

<sup>1</sup>Department of Agronomy, Iowa State University, Ames, IA, USA

<sup>2</sup>INRAE, Centre Île-de-France, Paris, France

<sup>3</sup>Analytics, Iowa Soybean Association, Ankeny, IA

<sup>4</sup>Department of Plant Pathology and Microbiology, Iowa State University, Ames, IA, USA

### **6.1 Abstract**

Small-plot and on-farm trials are used in agronomic research to provide management recommendations. These two types of agronomic experiments differ in machinery used and experimental scales. Small-plot trials (SPT) are traditionally conducted at agricultural experiment stations by research institutes or universities, while on-farm research trials (OFT) are conducted under local conditions and managed by farmers using their own equipment and specific management practices. As several authors have mentioned discrepancies between these experimental scales, extrapolating the outputs from small-plot trials to a larger scale can be challenging. There is a need to understand the impact of the experimental scale on the effect of agronomic management practices. In this study, we evaluated the impact of the experimental scale on foliar fungicide on crop yield and identified the cause of potential discrepancies. We utilized data

from 976 corn and soybean trials collected in five states across the United States. For soybean, our comparison of 479 on-farm trials and 83 small-plot trials showed no meaningful difference between mean yield responses. For corn, our comparison of 300 on-farm trials and 114 small-plot trials resulted in similar mean yield responses as well, but the within-trial standard deviation for the on-farm trials was four times smaller than the value estimated from the small-plot trials. The differences in the effects of different types of fungicide were similar whether they were estimated with on-farm trials or small-plots, for both crops. As on-farm and small-plot trials often have different objectives, a decision about the experimental scale choice should be based on research questions, land resources, and economic, technical, and human means.

## 6.2 Introduction

Agronomic research commonly produces knowledge of production-system performance and provides valuable management recommendations based on field experiments (Kravchenko et al., 2017). The experimental plots used differ in terms of management, equipment, and scales. For example, small-plot research trials, which can range from 0.005 to 0.01 ha, are the most widely used (Kravchenko et al., 2017) and they are traditionally conducted at agricultural experiment stations by research institutes or universities and are better suited for advanced or complex experimental designs (*i.e.*, randomized complete block or split-plot). Small-plot research trials might not always represent farming systems and farming practices (Marchant et al., 2019), but they are easier to manage for evaluating several treatments (Kandel et al., 2018). On-farm research trials (which can range from 1.9 to 11 ha (Laurent et al., 2019)) correspond to another type

of trials conducted on larger plots, directly on farmer fields and managed by farmers. They are becoming more widespread due to a growing interest among farmers and the research community (Kyveryga, 2019). On-farm research trials are conducted under local conditions and managed by farmers using their own equipment to answer field-level questions (Robertson et al., 2007). The standard experimental design for OFTs is the replicated strips where a new management practice, or product, is compared to a standard farmer practice (Kyveryga et al., 2018; Laurent et al., 2019). On-farm research trials represent a more relevant commercial-scale regarding farming practices, soil properties, and terrain attributes such as topography (Kravchenko et al., 2005, 2017; Terra et al., 2006; Marchant et al., 2019).

Even though on-farm trials can be used to validate small-plot trials at a larger field-scale (Licht and Witt, 2019), several authors have mentioned discrepancies between these scales (Table 1), especially with regards to crop yield. Therefore, extrapolating the outputs from small-plot trials to a larger scale can be challenging, and there is a need to understand the impact of experimental scale on yield. In his study, Tedford et al. (2017) analyzed the yield benefits from fungicides on corn at different scales. They found a yield difference between fungicide and untreated control equal to 1132 kg/ha at the field-scale (8.1 ha) while at the small-plot scale (14 to 45 m<sup>2</sup>), the yield difference was equal to 378 kg/ha. The main explanation offered was that near the alley and edge, yields are higher than towards the center of the plot as they experience less competition and have better light interception (Kleczewski, 2017). Tedford et al. (2017) further argued that there is a limited yield benefit when using fungicides in these conditions because within a large field, and away from the edge, plant yields are lower, which provides room for yield improvement when fungicides

are applied. In contrast, Vincelli and Lee (2015) demonstrated that open allies did not affect foliar fungicide on corn yields in field trials (Table 1). Schmer et al. (2008) compared bioenergy crop OFTs (6.7 ha on average) with SPTs. They found that agricultural energy inputs (fertilizer, herbicide, seed, diesel, other) were lower than previous studies based on SPTs and that switchgrass from OFTs, (managed for biomass yield with moderate levels of N fertilizer) produced 93% and 471% more ethanol per hectare than reconstructed prairies on SPTs and low-input switchgrass on SPTs, respectively. Conversely, Kandel et al. (2018) reported similar yield responses between SPTs and OFTs to foliar fungicides application on soybean over a control. On average, yield responses were identical, while the within variance was smaller in on-farm trials than in small-plot trials. This study also performed a power analysis showing that small plots required more replications than on-farm trials to detect the same overall treatment differences (probably greater noise due to larger soil variability or increased combine operator error). Kravchenko et al. (2017) showed that small-plot scale yields matched field-scale yields for conventional management but not for alternative systems (reduced-input and organic). Alternative farming systems are more dependent on timely management interventions such as weed pressure, the greatest management challenge for farmers at the field-scale. In contrast, mechanical weed control can generally be conducted at an optimal time at a small-plot scale.

Also, in field conditions, producers might experience logistic constraints that lead to yield losses or management practices (*e.g.*, late sowing) that result in an undetectable yield benefit (Andrade et al., 2019). Another reason concerns weather and spatial variabilities. Crop yields are highly variable across a field as a result of interaction between topography

(*i.e.*, terrain slope), soil properties (*i.e.*, soil moisture), weather conditions, and management practices (Kravchenko and Bullock, 2000; Kravchenko et al., 2005; Terra et al., 2006), while for SPTs the yield variability tends to be minimized by blocking factors, judicious plot layouts, and positing the experimental units on favorable soils (Kravchenko et al., 2017).

Beyond small-plot trials and on-farm trials, Andrade et al. (2019) suggested analyzing data collected from producer field surveys to better understand this discrepancy. They analyzed the yield difference between the narrow row and wide row spacing on soybean using field survey data and small-plot research data in different US soybean regions. They found similar yield differences across many producer fields and in small plots within the central region but not for the north region.

At the small-plot scale only, a meta-analysis focused on foliar fungicide-treated corn showed that of 436 trials, conducted across 13 states and Ontario during 2014 and 2015, 68% had a positive yield response (the difference between treated and untreated) (Wise et al., 2019). Fungicide formulation, disease level, and timing applications affected the yield response. Paul et al. (2011) conducted a meta-analysis using 187 trials in the US Corn Belt, where the treatments were the four most widely used foliar fungicides. They found a significant increase in yield relative to the untreated control, and the effect size (*i.e.*, yield difference) was generally greater for the higher disease severity category than the lower severity category. In field conditions, producers might experience logistic constraints that lead to yield losses or management practices (*e.g.*, late sowing) that result in an undetectable yield benefit (Andrade et al., 2019). Another reason concerns weather and spatial variabilities. Crop yields are highly variable across a field as a result of interaction

between topography (*i.e.*, terrain slope), soil properties (*i.e.*, soil moisture), weather conditions, and management practices (Kravchenko and Bullock, 2000; Kravchenko et al., 2005; Terra et al., 2006), while for small-plot research the variability tends to be minimized by blocking factors, judicious plot layouts, and location of experimental units within favorable soil types (Kravchenko et al., 2017).

Here, we compare small-plot vs. on-farm trials conducted to assess the effect of fungicide treatments on yields. For this study, we collected data from published summaries and raw data from small-plot trials and on-farm trials evaluating the foliar fungicides' effect on yield. Data were derived from five states across the United States on corn and soybean, including 53 active ingredients (alone and in combination) present in commercial foliar fungicides. Thus, our results should cover broader conditions than the ones already published. In our study, we want to evaluate the impact of experimental scale on crop response to commercial fungicides and identify the cause of potential discrepancies. It is crucial to understand the causes of discrepancy to inform on-farm decision-making better and interpret results at appropriate spatial scales.

Table 1: Main differences between small plot and on-farm trials.

<b>Topic of interest</b>	<b>Small plot trial (SPT)</b>	<b>On-farm trial (OFT)</b>
Plot size	- From 0.005 to 0.01 ha (Kravchenko et al., 2017)	- From 1.9 to 11 ha depending on the size of application equipment (Laurent et al., 2019)
Experimental design	- Usually randomized completely block design with multiple treatment to be tested simultaneously - Split-plot design is also frequent - Variability is minimized by blocking, contiguous designs, judicious plot layouts	- Replicated strip trial design with two treatments (the new management practice and the control) is the most popular and practical design - Implemented by farmers with their large-scale equipment

	and location on favorable soil (Kravchenko et al., 2017)	- Variability in soil properties is not minimized (Kyveryga et al., 2018)
--	--	---

Table 1: Continued

Topic of interest	Small plot trial (SPT)	On-farm trial (OFT)
Experimental design	- Monitor and pest control	- Replicated strip design usually presents a lack of randomization (Laurent et al., 2019). Thus, bias from unknown factors that could affect yield is not minimized  - Management practices are consistent across the trial and the treatment is the only management factor to change
Harvest	- Small-plot combine	- Farmers measure yield with mass flow sensors and GPS technology
Soil	- Greater noise (yield variability) in SPT could be explained by larger soil variation (Kandel et al., 2018)  - SPTs are usually located on favorable soils (Kravchenko et al., 2017)	- Soil and topographic diversity which leads to spatial variability in plant growth and crop yield (Kravchenko et al., 2005)
Weed control	- Mechanical weed control can generally be conducted at optimal times (Kravchenko et al., 2017)  - SPT represents an artificial system (hand-seeded, hand-weeded, and irrigated during establishment) (Schmer et al., 2008)	- OFTs resemble closer to farming systems and technologies (Marchant et al., 2019)

## 6.3 Materials and Methods

### 6.3.1 Data description

The data were extracted from foliar fungicides reports published in Fungicide and Nematicide Tests (F&N Tests) and Plant Disease Management Reports (PDMR), Iowa State Research Farm Progress Reports, Nebraska OFRN, Iowa Soybean Association OFRN, Ohio OFRN, South Dakota OFRN, Pennsylvania OFRN, and a few peer-reviewed papers. The keywords *corn*, *soybean*, and *foliar fungicide* were used to find the published reports or

peer-reviewed articles. More information about the search strategy can be found in supplemental materials (Section 6.8.1).

We consider a trial as a unique combination of year and location. Published summaries and raw data were collected from 978 trials conducted from 1999 to 2019, representing 6280 rows in the data set (one row represents a pair of treatment and control). Most of the time, a published summary included a single trial. In some cases, published summaries included data from near-by locations or at the same research station but testing, for example, different hybrids. In that case, we considered each trial as a separate observation. To be selected, the experimental design must include replicates, an untreated control, and the treatment(s) should be defined by the use of a single foliar fungicide on corn or soybean. The crop of interest must be corn or soybean, and studies must report the yield (raw data or summaries in weight or volume per unit area) for the untreated control and the treatment(s). Seed fungicide and combination of distinct foliar fungicides were not included in the data set. Other variables were not required for inclusion but were included in the data set if available (e.g., hybrid, plot length, plot width, rate of application; see Table 2 for more information).

Table 2: Description of the variables included in our data set.

<b>Variable</b>	<b>Description</b>	<b>Unit or extra information</b>
plot_scale	experiment scale	SPT or OFT
year	year of harvest	
crop	harvested crop	corn or soybean
variety	crop seed variety	
location	location of the trial	City or county
state	U.S.A. state	IA, NE, PA, SD, or OH
previous_crop	previous crop	
exp_design	type of experimental design	
irrigation	trial irrigation	yes/no/na
plot_long	length of the experimental unit	feet



plot_wide	width of the experimental unit	feet
experiment	number of the experiment reported in the study	
total_rep	total number of replicates	

Table 2: Continued

<b>Variable</b>	<b>Description</b>	<b>Unit or extra information</b>
rep	specific replicate number	for raw data only
seeding_rate	seeding density	seeds/acre
planting_date	planting date	
harvest_date	date of harvest	
treatment	foliar fungicide	commercial name
act_ing	active ingredient(s)	
group_name	fungicide group name for the active ingredients based on chemical structure and site of action	Grouping done by the Fungicide Resistance Action Committee (FRAC, 2020)
rate	fungicide rate of application	fl oz
stage	crop stage at the time of the application	
yield_treated	yield for the treatment	bushel/acre
yield_untreated	yield at the control	bushel/acre
diff	difference between yield_treated and yield_untreated	bushel/acre
ratio	ratio of yield_treated over yield_untreated	
lratio	log of the ratio	
plant_disease1	targeted foliar disease #1	
plant_disease2	targeted foliar disease #2	
LSD	least significant difference	
alpha	alpha level	.05 or .1
CV	coefficient of variation	%
author	first author of the report	
title	title of the report	
source	source of the report	scientific journal, website, etc.
ID	name of the pdf file	unique ID per report
exp_ID	merging variables experiment and ID	

Most of the on-farm-trials (90%) were conducted using a replicated strip design.

Small-plots were conducted using a randomized complete block design (84%) and

randomized block design (9%). The number of replicates or blocks ranged from 2 to 20 and 3 to 8 for the on-farm and small-plot trials, respectively (Table 3). The median number of replicates was 4 or 5 for both on-farm trials and small-plots (Figure 1). For the on-farm trials, the experimental units were 76 to 762 m long by 3 to 36 m wide, while for the small-plots, the experimental units were 1.8 to 35 m long by 1.2 to 12 m wide (Table 3). A large proportion of SPTs compared a several fungicides (i.e., at least two different fungicide Group Names) than OFTs (Table 3). In most cases, a nonionic surfactant was added with the foliar fungicide. Data were collected from 1999 to 2019 in 5 states (Iowa, Nebraska, Pennsylvania, South Dakota, and Ohio). For corn, 31%, 26%, and 12% of the trials included pyraclostrobin, prothioconazole + trifloxystrobin, and azoxystrobin + propiconazole as a treatment, respectively. For soybean, 32%, 16%, and 11% of the trials included pyraclostrobin, prothioconazole plus trifloxystrobin, and pyraclostrobin plus fluxapyroxad as a treatment, respectively. From the active ingredients, the foliar fungicides were classified into Group Names based on chemical relatedness of structures (chemical structure, site of action) (FRAC, 2020). The most represented fungicides group names were QoI (quinone inside inhibitors), DMI (demethylation inhibitors), and SDHI (succinate-dehydrogenase inhibitors). As some commercial fungicides included active ingredients from different group names, the group names are separated by an underscore (*e.g.*, QoI\_DMI).

Table 3: Description of the two plot scales: small-plot trial and on-farm trial

<b>Variables</b>	<b>Small-plot trials</b>	<b>On-farm trials</b>
Number of trials for corn	114	300
Number of trials for soybean	83	481
Range plot length (m)	1.8 to 35	76 to 762

Table 3: Continued

<b>Variables</b>	<b>Small-plot trials</b>	<b>On-farm trials</b>
Range plot width (m)	1.2 to 12	3 to 36
Mean number of replicate (min-max)	5 (3-8)	5 (2-20)
Number of trials in IA	64	640
Number of trials in NE	55	24
Number of trials in OH	62	55
Number of trials in PA	3	18
Number of trials in SD	13	44
Number of trials having one fungicide Group Name for corn	15	297
Number of trials having one fungicide Group Name for soybean	39	465
Number of trials having at least two different fungicides Group Names for corn	99	3
Number of trials having at least two different fungicides Group Names for soybean	44	16

### 6.3.2 Data analysis

We homogenized the data by calculating the mean log yield ratio (treated over untreated) for each trial, in case of the studies reported replicates values, and we kept the aggregated data otherwise. Thus, the dataset was reduced to 2353 rows (one row represents one mean yield ratio for a treatment over the control). When several foliar fungicides per trial are available, there is one mean log yield ratio per foliar fungicide.

For the statistical analysis, the mean log yield ratio was used as an effect size.

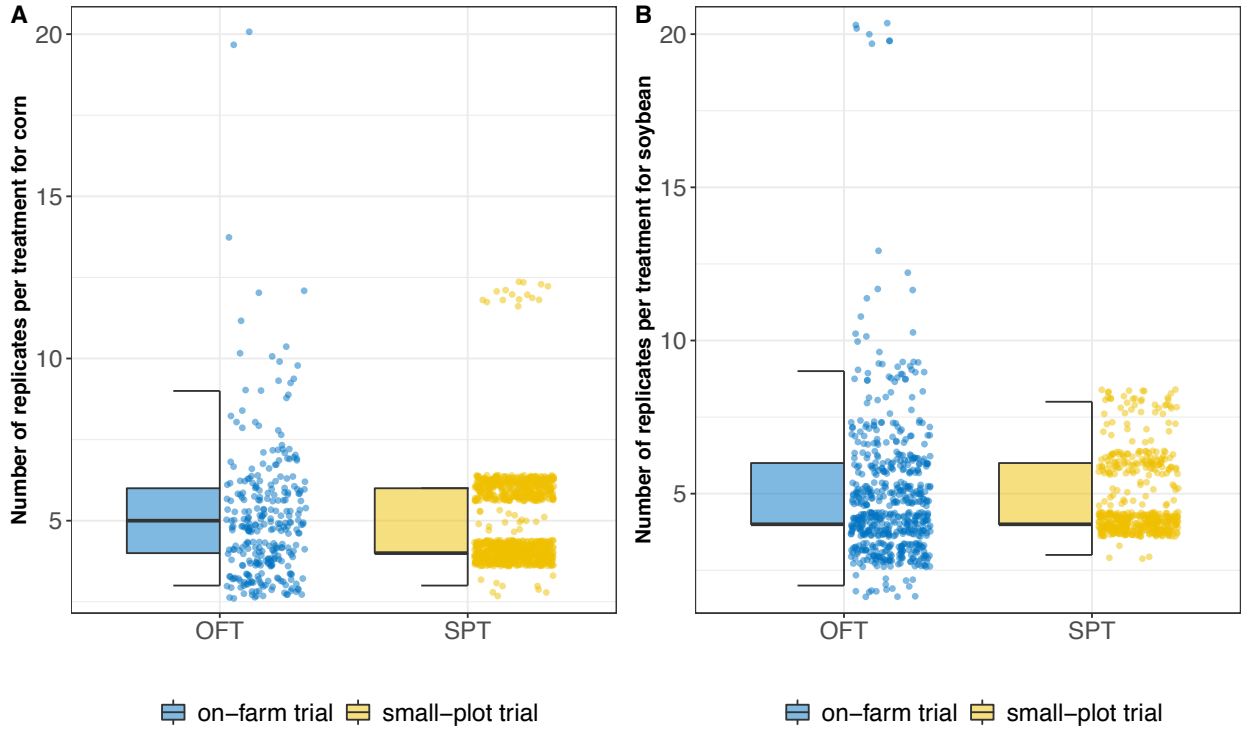


Figure 1: Boxplots of the number of replicates for corn (A) and soybean (B) for the different plot scales. On-farm trials are represented in blue and small plot trials in yellow. One dot represents the number of replicates for one combination of trial and treatment.

$$\log(R_{ijk}) = \mu_{OFT} + \mu_{SPT}X_i + \alpha_{ij} + \varepsilon_{ijk} \quad (1)$$

$$\alpha_{ij} \sim N(0, \sigma_{\alpha_i}^2)$$

$$\varepsilon_{ijk} \sim N(0, \sigma_{\varepsilon_i}^2)$$

Where  $\log(R_{ijk})$  represents the natural log of the  $j^{\text{th}}$  trial of category  $i$  ( $i=SPT$  or  $i=OFT$ ) for the  $k^{\text{th}}$  yield ratio.

$\mu_{OFT}$  represents the mean value of the log ratio for the OFT category.

$\mu_{SPT}$  represents the mean effect of the SPT category on the log ratio compared to OFT.

$X_i$  is a dummy variable equal to 1 if  $R_{ijk}$  belongs to the SPT category; zero otherwise.

$\alpha_{ij}$  represents the random effect of the  $j^{\text{th}}$  trial for the  $i^{\text{th}}$  category of trials.

$\varepsilon_{ijk}$  represents the residual error.

$\alpha_{ij}$  is assumed to follow independent Gaussian distributions with mean zero and unequal between-trial variance  $\sigma_{\alpha i}^2$  for the i=SPT and i=OFT categories.

$\varepsilon_{ijk}$  is assumed to follow independent Gaussian distributions with zero mean and unequal within-trial variance  $\sigma_{\varepsilon i}^2$  for the i=SPT and i=OFT categories.

Model (1) includes five parameters,  $\mu_{SPT}$ ,  $\mu_{OFT}$ ,  $\sigma_{\alpha i}^2$  and  $\sigma_{\varepsilon i}^2$ . The variance  $\sigma_{\alpha i}^2$  and  $\sigma_{\varepsilon i}^2$  quantifies the between- and within-trials variabilities, respectively. Separate variances are considered for OFT and SPT.

For categorical variables indicating the fungicide group name, the statistical model is:

$$\log(R_{ijk}) = \mu_{OFT} + \mu_{SPT}X_i + \sum_{g=2}^G(\beta_g + \theta_{SPTg}X_i)Z_i^{(g)} + \alpha_{ij} + \varepsilon_{ijk}$$

(2)

The model (2) remains the same as (1) except for  $\beta_g$  representing the parameter for the  $g^{\text{th}}$  fungicide group name,  $\theta_{SPTg}$  representing the interaction between the fungicide group name and the SPT category, and the  $Z_i^{(g)}$  representing a binary variable equal to 1 if  $R_{ijk}$  belongs to the  $g^{\text{th}}$  fungicide group name; zero otherwise.

Models (1) and (2) were fitted using a Bayesian approach with the R package brms (Bürkner, 2017). The priors for  $\sigma_{\alpha i}^2$  were defined as a truncated Student's t distribution with 10 degrees of freedom, zero mean and a scale parameter of 0.1. The prior for  $\log(\sigma_{\varepsilon OFT})$  was defined as a truncated Student's t distribution with 10 degrees of freedom, a mean of -4.4 and a scale parameter of 0.1. The prior for  $\log(\sigma_{\varepsilon SPT})$  was defined as a truncated Student's t distribution with 5 degrees of freedom, a mean of -3.2 and a scale

parameter of 0.5. These priors were based on previous datasets (Laurent et al., 2019). Priors for  $\mu_{SPT}$  and  $\mu_{OFT}$  were defined as a Normal distribution centered on zero and with a variance of 0.2. Posterior distributions for  $\beta_g$  and  $\theta_{SPTg}$  were computed using the No-U-Turn Samplers (Hoffmann and Gelman, 2014) as implemented in the Stan software (Stan Development Team, 2018). The convergence of the MCMC chains was checked by running three independent chains for 3,000 iterations, and the metric Rhat, which compares the between- and within-chain estimates for model parameters, were between 1 and 1.02 which indicates convergence. Two variants of the model (1) were fitted and evaluated using the Widely Applicable Information Criterion (WAIC) values. The WAIC is a more general criterion than Akaike Information Criteria (Watanabe, 2013) and appropriate for Bayesian models. The first model included equal between and within-trial variances, a second model included unequal within-trial variance only, and the third model considered unequal between-trial variances only for the plot-scale factor. The model (1) and the model with unequal within-trial variances performed better but, considering the structure of the data, we selected the model (1). We considered other models, but they resulted in higher WAIC values, so we did not select them. We performed a Pareto-Smoothed Importance Sampling cross-validation (PSIS) (Vehtari et al., 2017; McElreath, 2020) to detect highly influential observations on the WAIC. Two yield observations for both corn and soybean had high values of  $k$ , the Pareto shape parameter which indicates potentially influential observations and were discarded from the data set. These observations came from Ohio OFRN, South Dakota OFRN, Iowa State Research Farm Progress Reports and, Plant Disease Management Reports.

The four moments for the log ratio of corn and soybean and for each plot scale were computed using the R package moments (Komsta and Novomestky, 2015).

The mean yield difference was estimated as a function of the yield at the control using a linear regression.

## 6.4 Results

### 6.4.1 Descriptive results

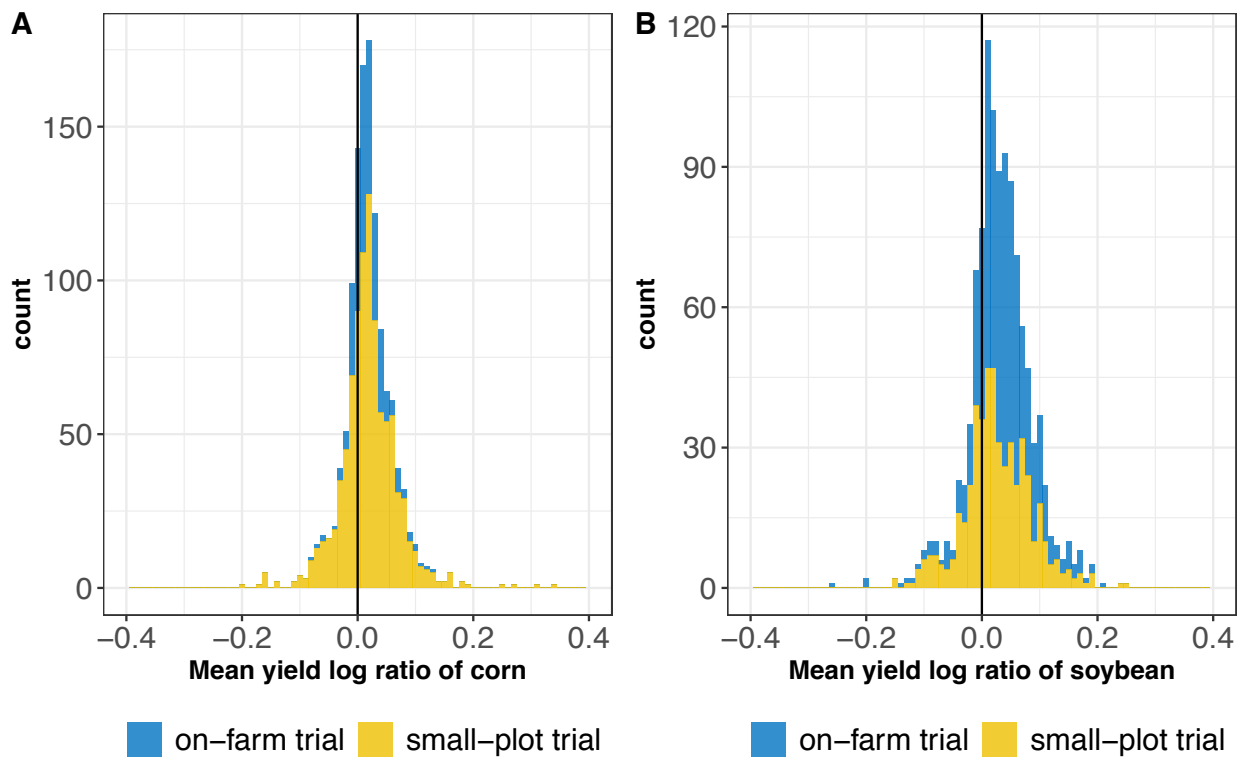


Figure 2: Distribution of the log yield ratio for corn(A) and soybean (B). On-farm trials are represented in blue and small plot trials in yellow. One count represents one replicate from the raw data or the mean yield difference from the report summaries.

Table 4: Description of the four moments for the log ratio of corn and soybean and for each plot scale (moments for the yield ratio are indicated in parenthesis).

<b>Crop</b>	<b>Plot scale</b>	<b>mean</b>	<b>variance</b>	<b>skewness</b>	<b>kurtosis</b>
Corn	OFT	0.017 (1.018)	0.001	0.66 (0.83)	5.6 (5.8)
	SPT	0.019 (1.020)	0.003	-0.008 (0.73)	10.6 (11.2)
Soybean	OFT	0.036 (1.038)	0.003	-0.47 (-0.05)	7.2 (6.2)
	SPT	0.027 (1.029)	0.003	0.18 (0.44)	3.9 (4.2)

For corn, the mean yield ratio for the two scales were similar (1.018 and 1.020 for OFT and SPT, respectively) (Table 4). The variance was three times higher for SPT compared to OFT. Regarding the skewness, the data are relatively symmetrical for OFT and SPT (Table 4 and Figure 2.A). The kurtosis was 5.6 and 10.6 for OFT and SPT, respectively, which means the data are Leptokurtic (peak is high and sharp and distribution is longer). The peak was higher for OFT than for SPT (Figure 2.A). For soybean, the mean yield ratio and the variances for the two scales were similar (1.038 and 1.029) for OFT and SPT, respectively (Table 4). Regarding the skewness, the distributions were fairly symmetrical for SPT and slightly skewed to the left for OFT (Figure 2.B). The kurtosis was 7.2 and 3.9 for OFT and SPT, respectively, which means the distributions are Leptokurtic (peak is high and sharp) for OFT and Mesokurtic (similar to a normal distribution) for SPT.



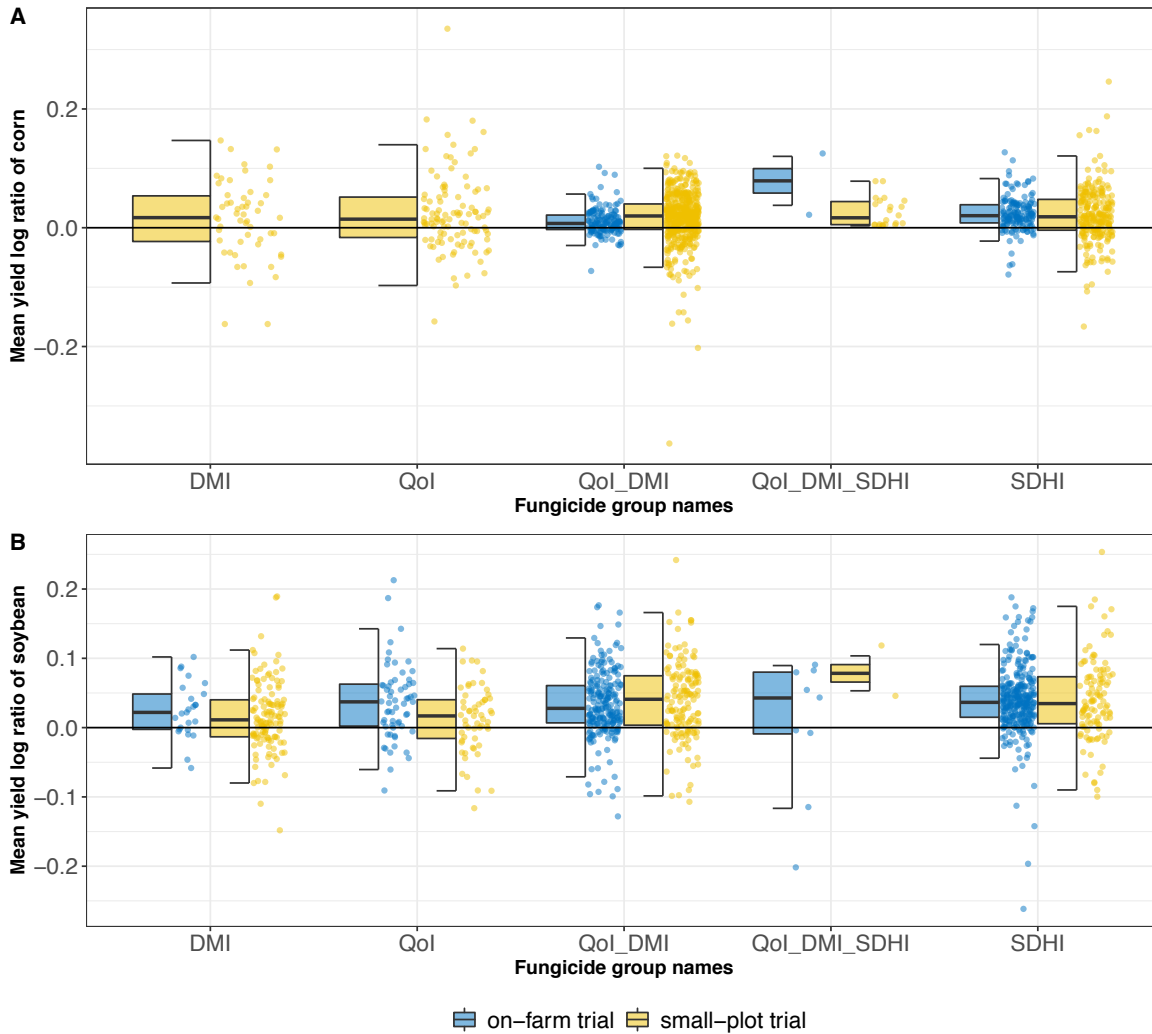


Figure 3: Boxplots of the mean log ratio of corn (A) and soybean (B) for the five most represented fungicides group names in the data set. On-farm trials are represented in blue and small plot trials in yellow. One dot represents the mean yield difference for one combination trial and foliar fungicide.

Across all the fungicide group names, the median log-ratio was positive for SPT and OFT (Figure 3). For the OFT, the group names DMI and QoI were not represented for OFT and, QoI\_DMI\_SDHI was poorly represented. For QoI\_DMI and SDHI, the median log-ratio for corn were similar but the data were more dispersed. Across all the fungicide group names, the mean yield ratios for corn were less dispersed for OFTs than SPTs while soybean mean yield ratios were as dispersed for OFTs as SPTs.

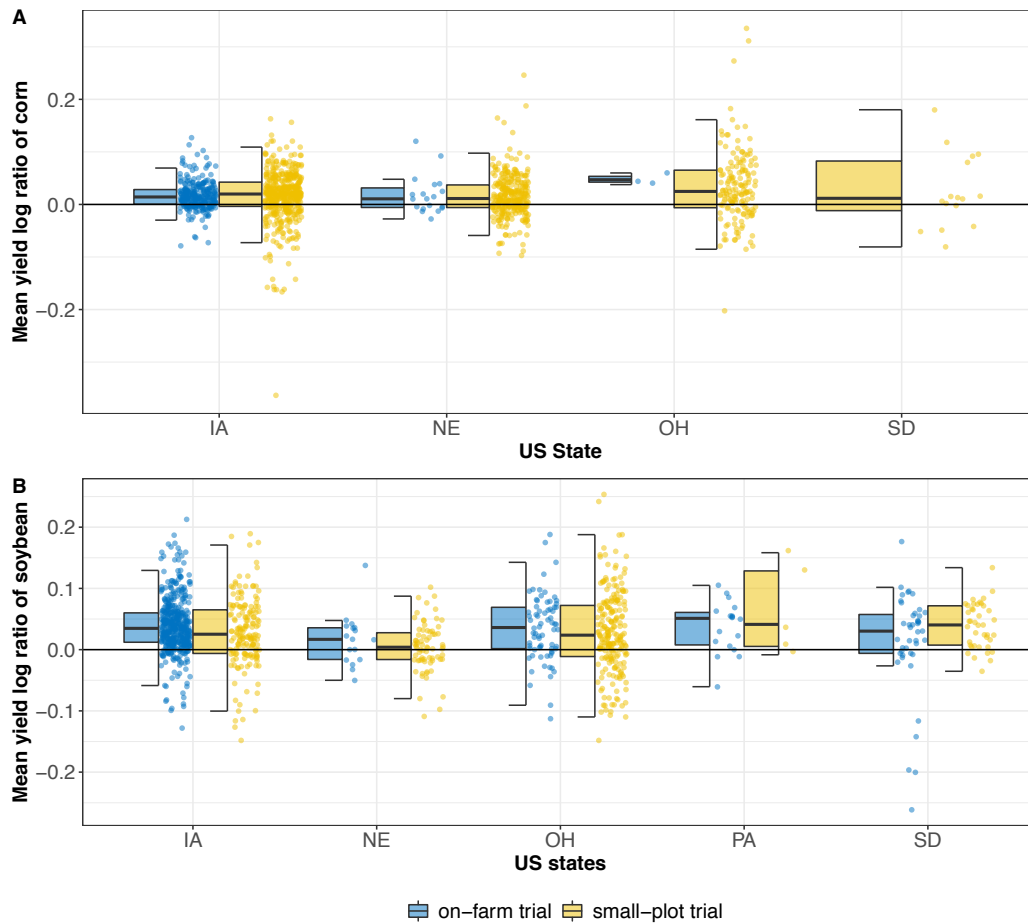


Figure 4: Boxplots of the log ratio of corn (A) and soybean (B) for each state. On-farm trials are represented in blue and small plot trials in yellow. One dot represents the mean yield difference for one combination trial and foliar fungicide.

Most of the data came from Iowa and Ohio for soybean (Figure 4). For corn, only Iowa and Nebraska had enough data in SPT and OFTs. South Dakota had some low values of the log yield ratio for soybean in OFTs. Other boxplots comparing the foliar fungicide application timings and year of harvest are available in supplemental materials (Section 6.8.2).

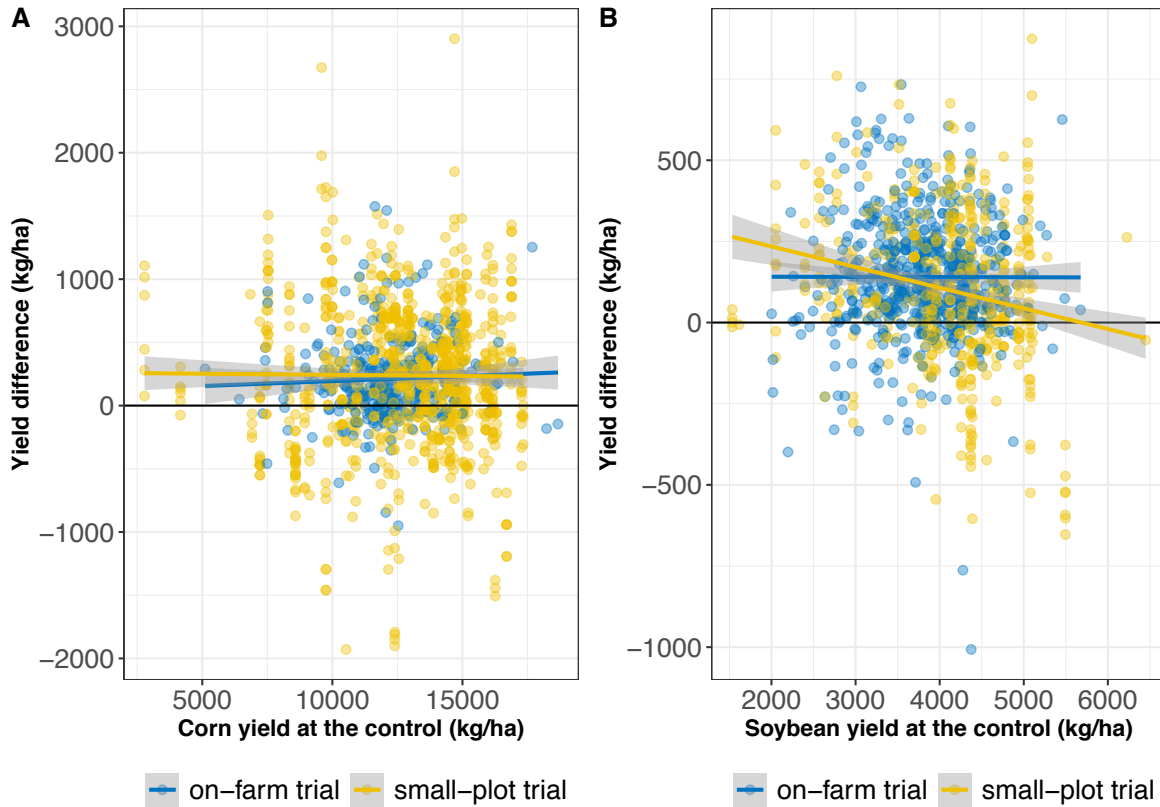


Figure 5: Yield difference of corn (A) and soybean (B) as a function of yield at the control (kg/ha). On-farm trials are represented in blue and small plot trials in yellow. One dot represents one replicate from the raw data or the mean yield difference from the report summaries. The lines represent the linear regression and their associated grey shade represents the 95% confidence intervals from the regression.

Often a management practice such as a fungicide application is effective when yields are relatively lower, and thus it can be expected that, in more general terms, a low yield at the control might indicate that a limiting factor or pest pressure has prevented the crop from reaching its potential (Figure 5). For soybean, there is a negative relationship between the yield response and the yield at the control for SPTs (Figure 5, on the right). When the yield at the control was low (*i.e.*, 2000-4000 kg/ha), the mean yield difference was in the 110-240 kg/ha range, indicating a beneficial effect of foliar fungicide on soybean

yield. When the yield at the control was high (*i.e.*, >4000 kg/ha), fungicide showed minimal benefits. For the OFTs, the relation between the mean yield difference and the yield at the control is fairly constant, and the mean yield difference was about 140 kg/ha. For corn, the relationship between the yield difference and yield at the control is fairly constant for SPTs (the mean yield difference was around 240 kg/ha) (Figure 5, on the left). For corn conducted on OFTs, the relationship between the yield difference and yield at is slightly positive and the mean yield difference was in the 150-280kg/ha range when the yield at the control ranged between 5000-16000 kg/ha.

#### 6.4.2 Results obtained with the Bayesian approach

The between-trial standard deviation is twice larger for SPT than OFT for corn and soybean (Table 5). The within-trial standard deviation was similar between OFT and SPT for soybean, but for corn, the within-trial standard deviation was four times larger for SPT than OFT . Parameters estimates for model (2) are presented in Table 6.

Table 5: Parameter estimates of response ratios and 95% credible intervals for model (1) for soybean and corn.  $\mu$  are for population estimates (means) and  $\sigma$  for standard deviations.

Parameter	Soybean	Corn
$\mu_{SPT}$	1.03 [1.02; 1.04]	1.02 [1.01; 1.02]
$\mu_{OFT}$	1.04 [1.03; 1.04]	1.02 [1.01; 1.02]
$\sigma_{\alpha SPT}$	0.04 [0.03; 0.05]	0.04 [0.03; 0.04]
$\sigma_{\alpha OFT}$	0.02 [0.02; 0.03]	0.02 [0.02; 0.03]
$\sigma_{\epsilon SPT}$	0.05 [0.04; 0.05]	0.04 [0.04; 0.04]
$\sigma_{\epsilon OFT}$	0.04[0.04; 0.05]	0.01 [0.01; 0.01]

Table 6: Parameter estimates of response ratios and 95% credible intervals for model (2) for soybean and corn.  $\mu$  are for population estimates (means) and  $\sigma$  for standard deviations.

Parameter	Soybean	Corn
$\mu_{SPT}$	1.01 [1.00; 1.04]	1.02 [1.01; 1.01]
$\mu_{OFT}$	1.02 [1.00; 1.03]	1.01 [1.01; 1.03]
$\sigma_{\alpha SPT}$	0.04 [0.03; 0.05]	0.04 [0.03; 0.04]
$\sigma_{\alpha OFT}$	0.03 [0.02; 0.04]	0.02 [0.03; 0.04]
$\beta_{Qoi}$	1.01 [1.01; 1.03]	not applicable
$\beta_{Qoi\_DMI}$	1.01 [0.99; 1.03]	not applicable
$\beta_{Qoi\_DMI\_SDHI}$	0.98 [0.94; 1.02]	not applicable
$\beta_{SDHI}$	1.02 [1.01; 1.04]	1.01 [1.01; 1.02]
$\theta_{SPT Qoi}$	0.99 [0.96; 1.02]	not applicable
$\theta_{SPT Qoi\_DMI}$	1.00 [0.98; 1.03]	not applicable
$\theta_{SPT Qoi\_DMI\_SDHI}$	1.05 [0.97; 1.14]	not applicable
$\theta_{SPT SDHI}$	1.01 [0.98; 1.03]	0.99 [0.98, 1.00]
$\sigma_{\epsilon SPT}$	0.05 [0.04; 0.05]	0.04 [0.03; 0.04]
$\sigma_{\epsilon OFT}$	0.04 [0.04; 0.05]	0.01 [0.01; 0.01]

The global yield ratio was similar between OFTs and SPTs for soybean (Figure 6). For corn, the global yield ratio was 3 times larger for SPT compared to OFT. It is important to mention that in our dataset, 99% of the OFTs on corn included only one fungicide group name (Table 3) which can explain the smaller credible interval in OFTs.

The mean yield ratio per trial and its associated uncertainty are represented in Figure 7 for both crops. The trials are ranked by increasing mean yield ratio to better understand the between-trial variability. The 95% credible interval is given by the horizontal bars and it represents the within-trial variability. For corn, the uncertainty is narrower for OFT than

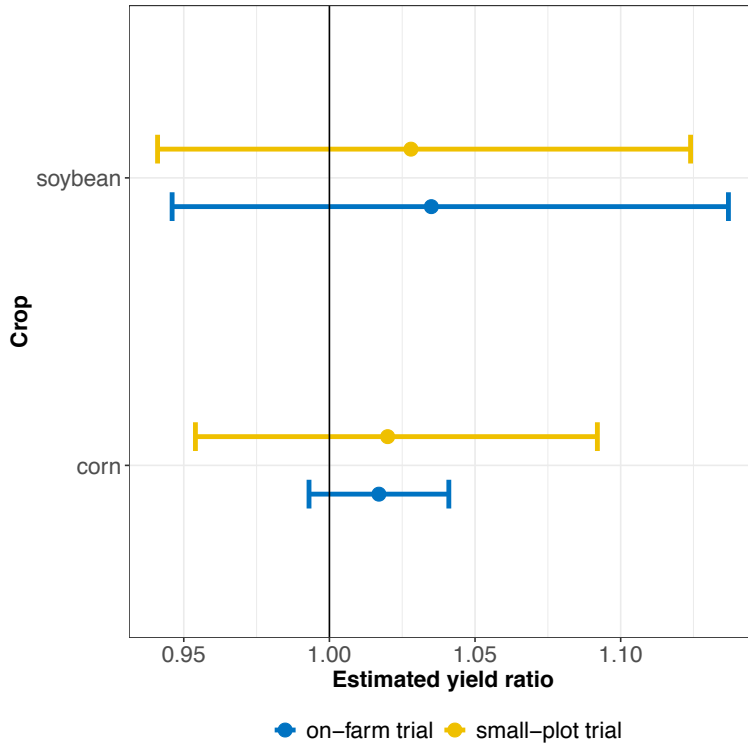


Figure 6: Global estimated yield ratio with its 95% credible intervals for small-plot trials (in yellow) and on-farm trials (in blue) for corn and soybean. Computations were done using model (1).

SPT which is explained by a smaller within-trial standard deviation (Table 5). The uncertainty around the mean yield ratio for soybean conducted on OFTs and SPTs were similar as the within-trial standard deviations were similar (Table 5). For soybean, most of the mean yield ratio for SPTs are at the bottom or the top of the forest plot (Figure 7, on the right).

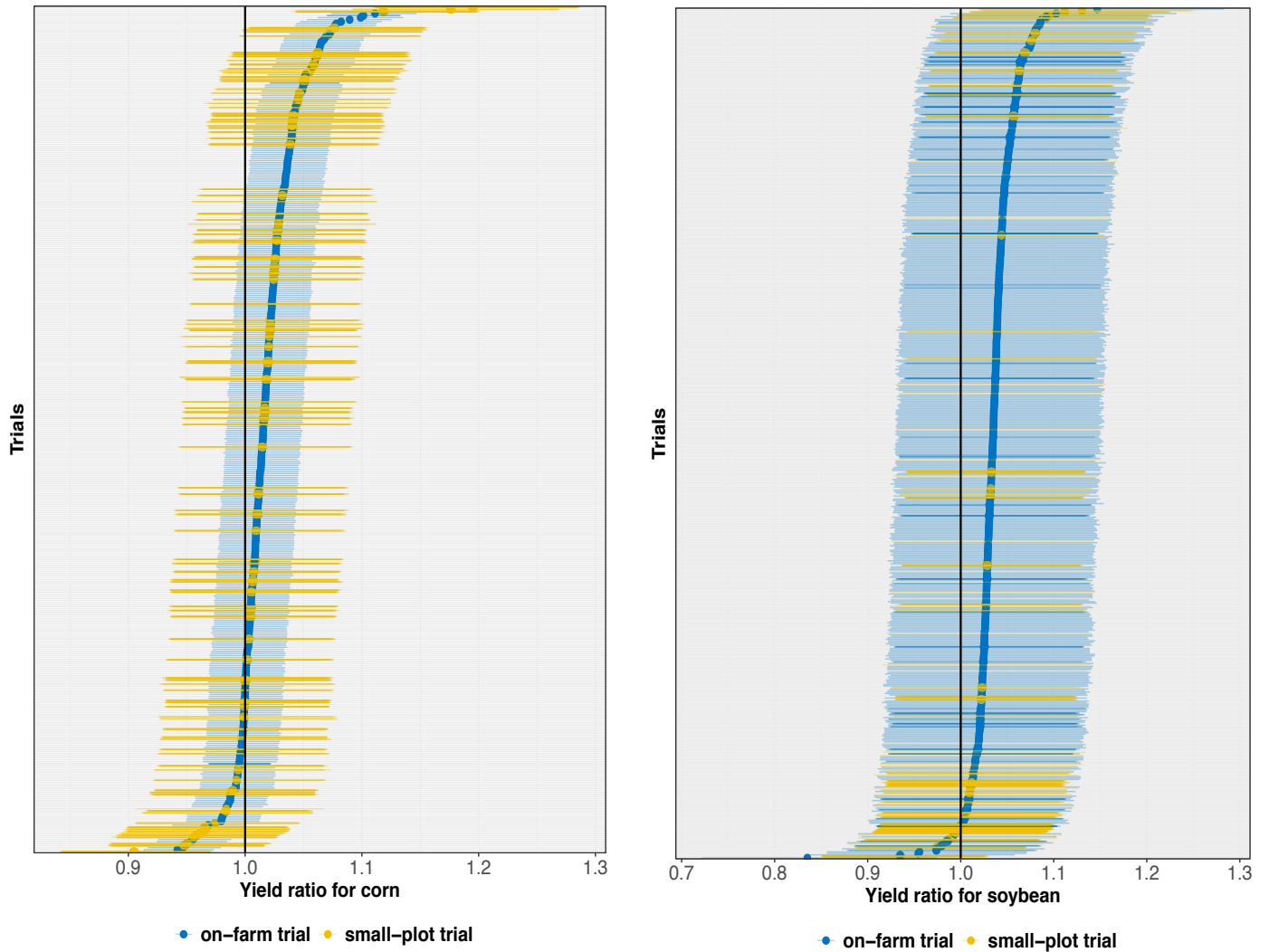


Figure 7: Estimated mean yield ratio with its 95% credible intervals for small-plot trials (in yellow) and on-farm trials (in blue) for corn (a) and soybean (b). Computations were done using model (1).

The mean yield ratio for each fungicide group name was similar between plot scales (Figure 8). Only the fungicide group names having enough trials in both plot scales were included in the analysis. For corn conducted on OFTs, SDHI fungicides had a yield ratio equal to 1.02 [0.99;1.05].

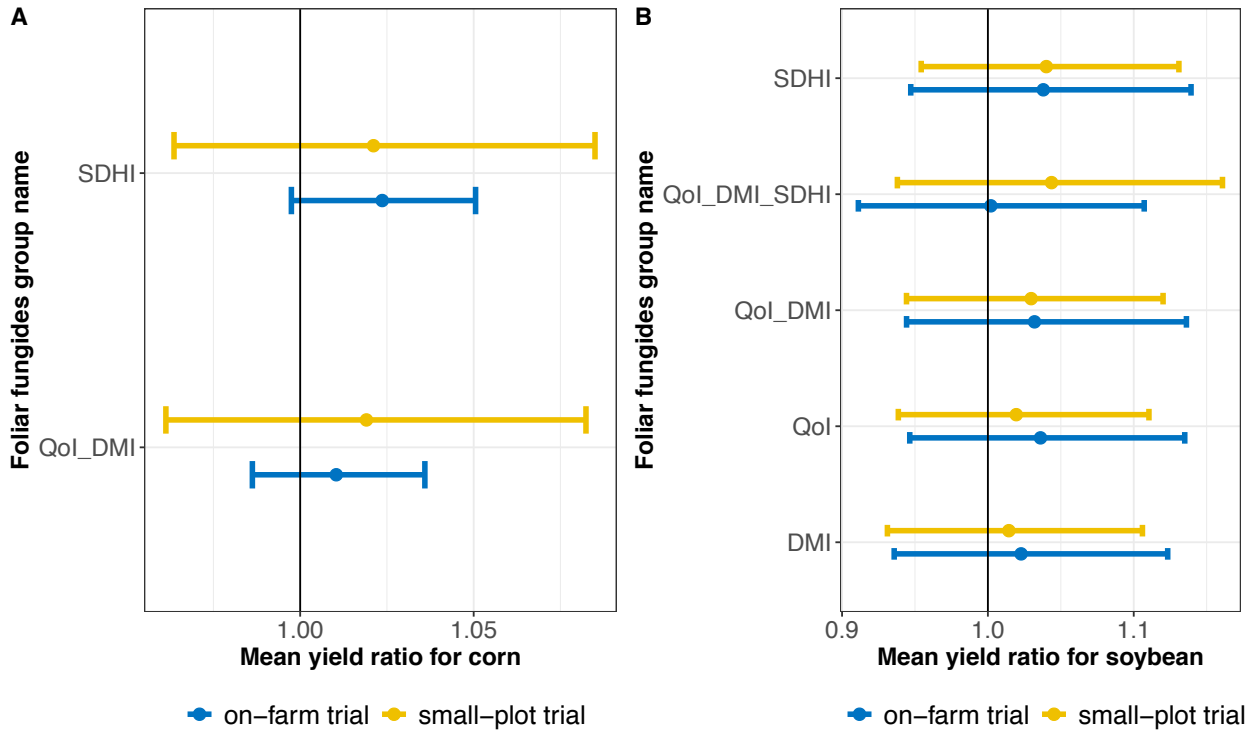


Figure 8: Estimated mean yield ratio with its 95% credible intervals per fungicide group name for small-plot trials (in yellow) and on-farm trials (in blue) for soybean (on the left) and corn (on the right). Computations were done using the model (2).

## 6.5 Discussion

In the present study, we assessed the impact of experimental scale (*i.e.*, OFTs vs. SPTs) on the effect of foliar fungicide on corn and soybean yields compared to an untreated control. Our comparison of 479 OFTs and 83 SPTs for soybean showed no statistical difference between mean yield responses. For corn, our comparison of 300 OFTs and 114 SPTs resulted in similar mean yield responses but the within-trial standard deviation for OFT was four times smaller than SPT. Comparisons between the fungicide groups, found no differences between OFTs and SPTs for both crops. Our results showed a negative relationship between the yield response and the yield at the control for the SPTs on



soybean, which suggests that the foliar fungicide had directly addressed a yield-limiting factor.

Our results are consistent with Kandel et al. (2018) as they compared the effect of foliar fungicides on soybean yield in 230 OFTs and 49 SPTs and found similar soybean yield responses. However, by analyzing a subset of 12 OFTs and 12 SPTs, each pair of trials within the same geographic area, they found a smaller residual variance in OFTs than SPT. This outcome is in accordance with our model outputs for corn, but residual standard deviations in OFTs and SPTs for soybean were similar (Table 5). Kandel et al. (2018) showed, by performing a power analysis, that SPTs need more replications than OFTs, which is not always possible regarding land, economic, technical, and labor resources, to detect the same overall treatment difference. A larger within-trial variability in SPTs could be explained by a larger impact of the spatial heterogeneity or increased combined operator errors. In our study, SPTs differed from OFTs by including more foliar fungicide treatments in the experimental design while the majority of OFTs included a single foliar fungicide treatment. Thus, the estimated uncertainty around the mean yield could also be explained by the variation between foliar fungicide treatments. Tedford et al. (2017) reported a greater yield response of foliar fungicides on corn in OFTs than SPTs, which contradicts our findings. Indeed, the mean yield responses in SPTs and OFTs were similar, but the uncertainty for OFTs was smaller. The main explanation given by Tedford et al. (2017) is that plants in SPTs are closer to the edge of the plot and could have less competition and/or better light interception. Thus, the edge located plants are likely near their maximum yield potential, and the benefit of a foliar fungicide is small. In another study, Paul et al. (2011) performed a meta-analysis of corn yield responses to foliar

fungicides in SPTs conducted in 14 states across the US between 2002 and 2009. They found a significant yield response compared to an untreated control on corn. For two foliar fungicides belonging to the group QOI-DMI, they found a higher yield difference when the mean yield in the untreated control was low. A SDHI foliar fungicide, representing 52% of the fungicide groups in their dataset, showed no effect across different yield levels for the untreated control. Our results are inconsistent with Paul et al., (2011) as we found no difference for the different foliar fungicide groups. Finally, Yan et al. (2002) compared non-replicated OFTs and SPTs for winter wheat in Ontario and found that both OFTs and SPTs provided valid data for cultivar evaluation.

OFTs and SPTs do not share the same purposes, and neither should be discounted. SPTs are more appropriate for advanced or complex experimental designs (e.g., randomized complete block, split-plot), allowing multiple treatments with or without different levels simultaneously. SPTs are suitable for testing innovative management practices and increasing agronomic knowledge before an on-farm evaluation, thus reducing the risk of yield loss by farmers. OFTs are useful for comparing a new management practice to a standard farmer practice using a replicated strip design (Laurent et al., 2019), are executed under local conditions and, use farmers' equipment. Thus, farmers are more likely to trust outputs from OFT than SPT or commercial fields (Thompson et al., 2019) which can facilitate the adoption of new management practices. OFTs allow in-depth investigation across a wider variety of soils and landscape position, which is usually limited to station-based research. OFTs can answer specific field-level questions such as nutrient management, environmental issues (Schnepf and Cox, 2007) and investigate runoff and soil erosion (Kuhn et al., 2014). Regarding financial and land resource constraints, OFTs

represent a more feasible option than SPTs conducted in experiment station research (Thompson et al., 2019).

Even if our study focuses on comparing the yield response in OFTs and SPTs, it is essential to consider the price of fungicides, and their application cost, for better decision making. A yield benefit is not always enough to offset the fungicide cost in the absence of disease (*i.e.*, assumed here by a high yield at the control). Kandel et al. (2016) suggested that although fungicides can increase soybean yield in the absence of disease, only 14% of the trials had profitable response to fungicides.

Some limitations of our study should be highlighted. While our goal was to compare OFTs and SPTs across different states, in order to have different pedoclimatic conditions, our data set was unbalanced. For example, OFTs for corn mostly came from Iowa, and few from Nebraska, which can explain the smaller uncertainty as most of the trials are from the same farming area. For soybean, data were unbalanced as 82% of the trials were OFTs. A decision about the experimental scale choice should be based on research questions, land resources, and economic, technical and human means. Future research could include site-specific covariates, such as fungicide application timing, corn hybrid, soybean variety, or crop stage to better predict yield responses to fungicides.

## 6.6 Acknowledgments

The work of FM and AL was also supported by the Iowa Soybean Research Center and the Plant Science Institute at Iowa State University.

## 6.7 References

- Andrade, J.F., J.I. Rattalino Edreira, S. Mourtzinis, S.P. Conley, I.A. Ciampitti, et al. 2019. Assessing the influence of row spacing on soybean yield using experimental and producer survey data. *Field Crops Research* 230: 98–106. doi: 10.1016/j.fcr.2018.10.014.
- Bürkner, P.-C. 2017. brms : An R Package for Bayesian Multilevel Models Using *Stan*. *J. Stat. Soft.* 80(1). doi: 10.18637/jss.v080.i01.
- FRAC. 2020. FRAC Code List© 2020:Fungal control agents sorted by cross resistance pattern and mode of action(including FRAC Code numbering). [https://www.frac.info/docs/default-source/publications/frac-code-list/frac-code-list-2020-finalb16c2b2c512362eb9a1eff00004acf5d.pdf?sfvrsn=54f499a\\_2](https://www.frac.info/docs/default-source/publications/frac-code-list/frac-code-list-2020-finalb16c2b2c512362eb9a1eff00004acf5d.pdf?sfvrsn=54f499a_2).
- Hoffmann, M., and A. Gelman. 2014. The No-U-Turn Sampler: Adaptively Setting Path Lengthsin Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 15(1): 1351–1381.
- Kandel, Y.R., C.L. Hunt, P.M. Kyveryga, T.A. Mueller, and D.S. Mueller. 2018. Differences in Small Plot and On-Farm Trials for Yield Response to Foliar Fungicide in Soybean. *Plant Disease* 102(1): 140–145. doi: 10.1094/PDIS-05-17-0697-RE.
- Kandel, Y.R., D.S. Mueller, C.E. Hart, N.R.C. Bestor, C.A. Bradley, et al. 2016. Analyses of Yield and Economic Response from Foliar Fungicide and Insecticide Applications to Soybean in the North Central United States. *Plant Health Progress* 17(4): 232–238. doi: 10.1094/PHP-RS-16-0038.
- Kleczewski, N. 2017. Small vs Large Plot Studies on Fungicides and Yield — Which is Better? <https://sites.udel.edu/weeklycropupdate/?p=11252>.
- Komsta, L., and F. Novomestky. 2015. Moments, cumulants, skewness, kurtosis and related tests.
- Kravchenko, A.N., and D.G. Bullock. 2000. Correlation of Corn and Soybean Grain Yield with Topography and Soil Properties. *Agron. J.* 92(1): 75. doi: 10.1007/s100870050010.
- Kravchenko, A.N., G.P. Robertson, K.D. Thelen, and R.R. Harwood. 2005. Management, Topographical, and Weather Effects on Spatial Variability of Crop Grain Yields. *Agronomy Journal* 97(2): 514. doi: 10.2134/agronj2005.0514.
- Kravchenko, A.N., S.S. Snapp, and G.P. Robertson. 2017. Field-scale experiments reveal persistent yield gaps in low-input and organic cropping systems. *Proc Natl Acad Sci USA* 114(5): 926–931. doi: 10.1073/pnas.1612311114.

Kuhn, N.J., P. Greenwood, and W. Fister. 2014. Use of Field Experiments in Soil Erosion Research. *Developments in Earth Surface Processes*. Elsevier. p. 175–200

Kyveryga, P.M. 2019. On-Farm Research: Experimental Approaches, Analytical Frameworks, Case Studies, and Impact. *Agron.j.* 111(6): 2633–2635. doi: 10.2134/agronj2019.11.0001.

Kyveryga, P.M., T.A. Mueller, and D.S. Mueller. 2018. On-Farm Replicated Strip Trials. In: Kent Shannon, D., Clay, D.E., and Kitchen, N.R., editors, *ASA, CSSA, and SSSA Books*. American Society of Agronomy and Soil Science Society of America, Madison, WI, USA. p. 189–207

Laurent, A., P. Kyveryga, D. Makowski, and F. Miguez. 2019. A Framework for Visualization and Analysis of Agronomic Field Trials from On-Farm Research Networks. *Agronomy Journal* 111(5): 1–12. doi: 10.2134/agronj2019.02.0135.

Licht, M., and M. Witt. 2019. Conducting On-Farm Trials. Iowa State University, Extension and Outreach. <https://crops.extension.iastate.edu/cropnews/2019/03/conducting-farm-trials>.

Marchant, B., S. Rudolph, S. Roques, D. Kindred, V. Gillingham, et al. 2019. Establishing the precision and robustness of farmers' crop experiments. *Field Crops Research* 230: 31–45. doi: 10.1016/j.fcr.2018.10.006.

McElreath, R. 2020. *Statistical rethinking: a Bayesian course with examples in R and Stan*. 2nd ed. Taylor and Francis, CRC Press, Boca Raton.

Paul, P.A., L.V. Madden, C.A. Bradley, A.E. Robertson, G.P. Munkvold, et al. 2011. Meta-Analysis of Yield Response of Hybrid Field Corn to Foliar Fungicides in the U.S. Corn Belt. *Phytopathology*® 101(9): 1122–1132. doi: 10.1094/PHTO-03-11-0091.

Robertson, G.P., L.W. Burger, C. Kling, R. Lowrance, and D. Mulla. 2007. *New approaches to environmental management research at landscape and watershed scales. Managing agricultural landscapes for environmental quality: strengthening the science base*. Soil and Water Conservation Society, Ankeny, Iowa

Schmer, M.R., K.P. Vogel, R.B. Mitchell, and R.K. Perrin. 2008. Net energy of cellulosic ethanol from switchgrass. *Proceedings of the National Academy of Sciences* 105(2): 464–469. doi: 10.1073/pnas.0704767105.

Schnepf, M., and C.A. Cox, editors. 2007. *Managing agricultural landscapes for environmental quality: strengthening the science base*. Soil and Water Conservation Society, Ankeny, Iowa.

Stan Development Team. 2018. Stan Modeling Language Users Guide and Reference Manual.

Tedford, E.C., A.B. Kriss, C. Geater, M. Saini, B. Battles, et al. 2017. Plot size can influence yield benefits from fungicides on corn. *Crop Protection* 91: 66–73. doi: 10.1016/j.cropro.2016.09.017.

Terra, J.A., J.N. Shaw, D.W. Reeves, R.L. Raper, E. van Santen, et al. 2006. Soil Management and Landscape Variability Affects Field-Scale Cotton Productivity. *Soil Science Society of America Journal* 70(1): 98. doi: 10.2136/sssaj2005.0179.

Thompson, L.J., K.L. Glewen, R.W. Elmore, J. Rees, S. Pokal, et al. 2019. Farmers as Researchers: In-depth Interviews to Discern Participant Motivation and Impact. *Agron.j.* 111(6): 2670–2680. doi: 10.2134/agronj2018.09.0626.

Vehtari, A., A. Gelman, and J. Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 27(5): 1413–1432. doi: 10.1007/s11222-016-9696-4.

Vincelli, P., and C. Lee. 2015. Influence of Open Alleys in Field Trials Assessing Yield Effects from Fungicides in Corn. *Plant Disease* 99(2): 263–266. doi: 10.1094/PDIS-04-14-0415-RE.

Wickham, H. 2016. *ggplot2: elegant graphics for data analysis*. Second edition. Springer, Cham.

Wise, K.A., D. Smith, A. Freije, D.S. Mueller, Y. Kandel, et al. 2019. Meta-analysis of yield response of foliar fungicide-treated hybrid corn in the United States and Ontario, Canada (D.A. Lightfoot, editor). *PLoS ONE* 14(6): e0217510. doi: 10.1371/journal.pone.0217510.

Yan, W., L.A. Hunt, P. Johnson, G. Stewart, and X. Lu. 2002. On-Farm Strip Trials vs. Replicated Performance Trials for Cultivar Evaluation. *Crop Science* 42(2): 385. doi: 10.2135/cropsci2002.0385.

## 6.8 Supplemental materials

### 6.8.1 Search strategy

For on-farm trials data, all the websites of the on-farm research network listed in Table 1 in Laurent et al., (2019) were consulted and five of them are sharing data through individual trial reports, annual reports or searchable database.

Below are the websites used for extracting on-farm trial reports:

- Nebraska On-Farm Research Network (<http://resultsfinder.unl.edu/>). In the searchable database, all the counties, year and irrigation methods were selected. The item crop was filtered by corn and soybean, topic was restricted to crop protection and the subtopic was restricted to fungicide.
- Agronomic Crops Network of Ohio State University (<https://agcrops.osu.edu/on-farm-research>). A search was done using the fungicide as a keyword.
- On-farm trial annual reports managed by Pennsylvania Soybean Board were investigated (<https://pasoybean.org/checkoff-at-work/research/on-farm-network/>).
- South Dakota Soybean On-Farm Research Program ([https://onfarmresearch.sdsoybean.org/trials?fwp\\_trial\\_type=fungicide&fwp\\_paged=3](https://onfarmresearch.sdsoybean.org/trials?fwp_trial_type=fungicide&fwp_paged=3)). In the searchable database, the trial type was restricted to Fungicide, other items were kept by default.
- On-farm trial reports were given by P.K, a co-author of this study, but data are available on <https://analytics.iasoybeans.com/cool-apps/ISOFAST/>

For small-plot trials, two research equations, given below, were used on Iowa State University Library website:

- soybean (title) AND "foliar fungicide" (title) AND "yield" (any field)
- corn (title) AND "foliar fungicide" (title) AND "yield" (any field)

Many technical reports for small-plot trials were extracted from the Plant Disease Management Reports website

(<http://www.plantmanagementnetwork.org/pub/trial/PDMR/default.asp>) by

investigating each volume separately. This website also provides reports from Fungicide and Nematicide Tests.



## 6.8.2 Additional descriptive results

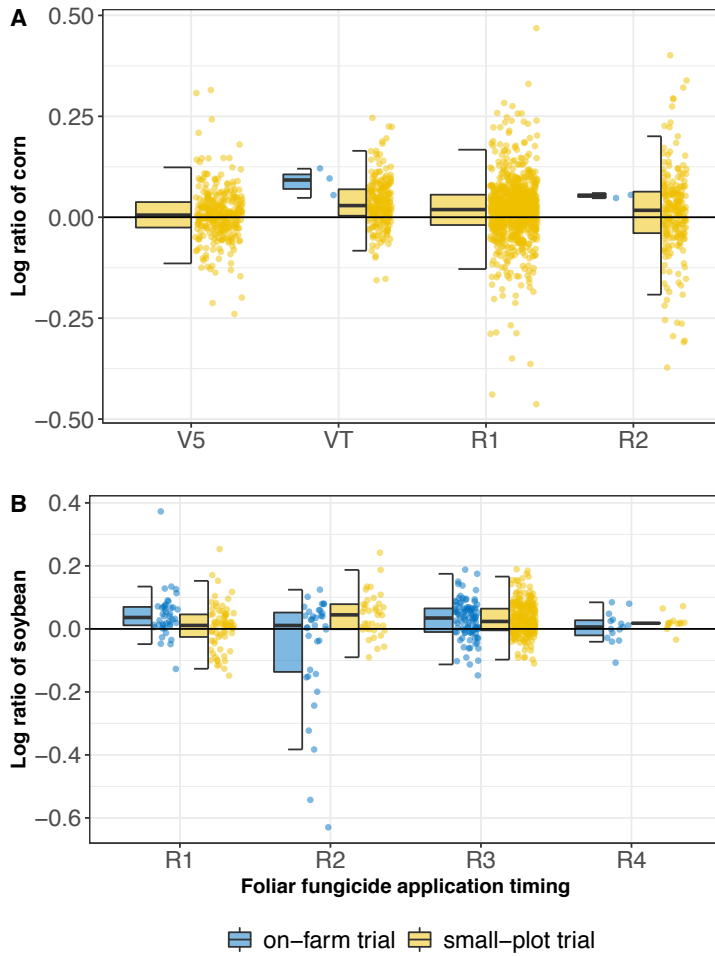


Figure S1: Boxplot of the log ratio of corn (A) and soybean (B) for the four most represented foliar fungicide application timings in the data set. On-farm trials are represented in blue and small plot trials in yellow. One dot represents the mean yield difference for one combination trial and foliar fungicide.

The foliar fungicide application timing was not available for more than 50% of the studies for both crops. Regarding the four most frequent timing application, few data were available for corn in OFTs (Figure S1.A). At R2, the soybean yield response is higher for SPTs than OFTs, but no difference was found for other application timings (Figure S1.B).

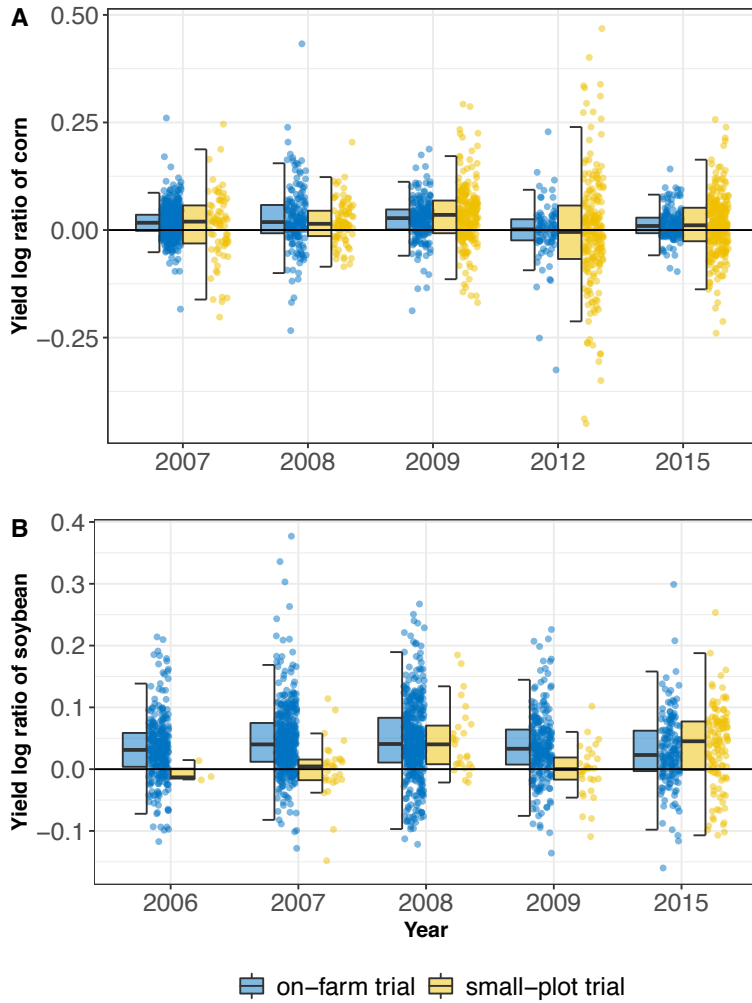


Figure S2: Boxplot of the log ratio of corn (A) and soybean (B) for different years of harvest included in the data set. On-farm trials are represented in blue and small plot trials in yellow. One dot represents the mean yield difference for one combination trial and foliar fungicide.

OFTs and SPTs showed some difference over time especially for soybean in 2007 and 2009 (Figure S2.B). For corn, the yield responses are quite similar but the data are more dispersed for SPTs (Figure S2.A).

## CHAPTER 7. GENERAL CONCLUSION

The overall objective of this research was to advance the field of agricultural statistics by identifying appropriate statistical methods for on-farm research data. This research field is exciting because it connects research community and farmers, leads to real-world applications, shares results with a broad audience, and helps farmers make better choices. As this work includes more than 50 OFRNs delivered by the Iowa Soybean Association, one French OFRN on grapevine and explores OFRNs from four different states in the US, focusing on generic methods was a key to achieve this general objective.

In Chapter 2, I provided a data-analytics framework adapted to a replicated strip trial design using two treatments, but the methods can be easily extended to other experimental designs. In Chapter 3, I described the interactive web application ISOFAST for OFRN that enhanced the communication of the results with an intuitive understanding of the model outputs. This web application simplifies the interpretation of multi-site and multi-year summaries to help decision making. In Chapter 4, I introduced prediction intervals and found that they can be about 2.2-12.1 times wider than confidence intervals for the estimated yield responses for all tested management practices considered in this research. This chapter made the case that prediction intervals and the probability of ineffective treatments should be systematically reported to provide information about outcomes for a future trial in an OFRN. In Chapter 5, I demonstrated that an adaptation of the data-analytics framework developed in Chapter 2 was extended to a French case-study by adapting the variable response and the statistical model. In Chapter 5, I found that two biopesticides, based on potassium bicarbonate and *Aureobasidium pullulans*, had a partial efficacy in reducing the incidence of gray mold in organic vineyards when the disease

pressure does not exceed 20%. In the wine-growing area of our study, the interest for testing management practices through an on-farm research network is increasing; thus, the proposed statistical model might be useful for future research projects. In Chapter 6, I found similar mean yield responses to foliar fungicides compared to an untreated control for soybean. For corn, mean yield responses were similar, but the within-trial variance for OFT was four times smaller than for SPT.

The body of knowledge gained from this research is two folds: to provide advanced and extendable statistical methods for researchers and agronomists responsible for analyzing OFRN data, and to improve accessibility to on-farm research insights using data visualization. This work is mostly based on Bayesian approaches as it provides a more intuitive framework than the frequentist approach to estimate yield responses at the network and trial level. Besides, the Bayesian approach allows to better estimate and quantify the uncertainty of the performance of management practices. Overall, this research provides a better understanding of the overall effectiveness of a new management practice and represents the first major effort in consolidating results from on-farm research network. Indeed, ISOFAST is the first interactive web tool developed and launched for communicating OFRN data, and the use of meta-analyses techniques (*i.e.*, mixed-effect model, prediction intervals) was, to our knowledge, not standard in the field of OFRN.

Some research projects in close relation to this dissertation or projects expanding the concept of OFRN might arise in the future. This work demonstrates the importance of the within-trial variability compared to between-trial variability. It seems crucial to investigate and understand the different sources of variability to help reduce the noise and improve statistical inferences and predictions. Another challenge will be to improve the generic

aspect of the proposed methods to match other agricultural areas having different research questions. For example, during my research I spent an extensive stay at INRA in France (supplemental materials 7.1) and French institutes expressed interest in testing decision support tools or a combination of different management practices and also in assessing agronomic sustainability by including indicators other than crop yield. This work was done at different scales from a major wine producing region located in France to five US states. Still, future projects, gathering data from the OFRNs from the University of Nebraska and the Iowa Soybean Association, will emerge to evaluate the effect of different management practices at a wider scale. Access to a larger quantity of OFRN data will represent an opportunity to understand the sources of uncertainty taking into account weather, topography, soil properties and agronomic management. It will also be interesting to compare the yield response at the field scale, the region scale (NE and IA), and the farm scale using farm-survey data. Other management practices other than foliar fungicide (*e.g.*, fertility and soil management) could also be compared at the small-plot and field scales to understand the causes of yield gaps and inform better on-farm decision-making.

### **7.1 Supplemental Materials**

As part of the cotutelle between Iowa State University and Université Paris Saclay, I spent an extensive stay in France at INRA under the direction of David Makowski. The overall objective was to enhance the generic framework (developed in Chapter 2) to a French case study.

### **Workshop with French stakeholders**

To conduct this project, it was important to answer a few key questions: What has been done? Are there any OFRNs in France? What are the specificities of the French context? For that, we identified stakeholders from French extension agencies and private companies, and we organized a workshop to explain the concept of OFRN in the US, present the web-application ISOFAST (Chapter 3), and discuss the pros and cons of our framework in the French context. I also discussed individually with some French extension agencies. In total, I met with 10 French institutes:

- Terres Inovia: Institute for legumes and oleaginous crops
- IFV: Institute of Vine and Wine production
- ITB: Institute of sugar beet
- Agricultural Chamber
- ACTA: Head of the Agricultural Technical Institute network
- ITAB: Institute of organic production
- Dijon Céréales: Agricultural cooperative
- INRAE: National Research Institute for Agriculture, Food and Environment
- Arvalis: Agricultural technical institute for crop.
- FRAB Bretagne: Regional Federation of organic farmers in Bretagne

The French institutes showed a great interest in the concept of OFRN. Some already managed OFRNs, but it appeared that some deficiencies in the experimental design prevent them from performing advanced statistical analysis, or sometimes very few sites were part of the network. Farmers own GPS-yield monitors, but it is less common in France than in

the US. Most of the U.S. OFRNs look at the yield as an indicator but some collaborators were interested in combining the concept of OFRN with environmental indicators or by combining different management practices as a treatment.

This workshop was helpful in identifying the French Institute of Vine and Wine (IFV) as a collaborator to share datasets to test the framework. The following section 7.2 gives a summary of this collaboration and more information can be found in the Chapter 5.

### **Project on biocontrol with the Institute of Vine and Wine production**

The French Institute on the Wine and Vine managed an OFRN on the grapevine from 2014 to 2016 to evaluate the efficacy of biocontrol agent products against *Botrytis* bunch rot in organic production. The goal was to enhance the acquisition of knowledge in farming conditions as most of the studies on biocontrol agents available before 2012 were done by phytosanitary companies.

Two biocontrol agents, potassium bicarbonate (Armicarb®) and *A. pullulans* (Botector®) were tested in 23 and 26 trials, respectively, over three years, and compared to an untreated control. Armicarb® is composed of 85% potassium bicarbonate and interacts directly with the pathogen by killing mycelium and spores while disrupting pH and osmotic pressure. Botector® is composed of two fungus strains of *Aureobasidium pullulans*, DSM 14940 and DSM 14941, and is competitive for nutrients and space. As a prophylactic method, leaf removal was applied during the growing season for all the trials, as frequently done in organic systems (Weigle and Carroll, 2014).

Trials were distributed throughout Southwestern France, a major wine-producing area. Therefore, this study covers typical conditions of oceanic climate and grapevine cultivars in this given area. The experimental design were strips with two replicates, where the two treatments were applied without randomization. A pair of strips (biocontrol agent product and the control) constitutes a replicate. Within each strip, four plots were used for measurements.

Unlike the yield data analyzed for Chapters 2 to 4, the variable responses for this project are the incidence and the intensity. Those variables were assessed by visually rating 50 grape bunches per plot at the end of each growing season. The incidence represents the percentage of bunches with symptoms (*i.e.*, the number of bunches with symptoms divided by 50), and the intensity (also called severity in the literature) represents the percentage of rotten berries (*i.e.*, the sum of rotten berries per bunch divided by 50).

From a statistical point of view, the binomial or Poisson distributions are commonly used in epidemiology as the disease occurs with a probability  $p$ . In our case, we used a Beta-Binomial distribution as the expected variance was higher than the one predicted by the Binomial model (*i.e.*, overdispersion). A logit link was used to constraint the probability to lie between zero and one. More information about the statistical model and the results of this project can be found in Chapter 5.

### **Project with Dijon Cereal**

I also had an opportunity to collaborate with Dijon Cereal (hereafter called DC), an agricultural cooperative whose capital is carried by nearly 3800 farmers members. DC is based in Central-Eastern France and gathers breeders and farmers growing cereals from a wide diversity of farming environments (weather, soil).



Rapeseed is a crop largely grown in this area, but yields have decreased by 40% due to lack of water in recent years. Rapeseed must be sown in late summer (by the end of August), but the lack of water affects the seed setting and filling process. The overall objective of DC is to test alternative crops that are water stress tolerant under rainfed conditions through an OFRN as the ones developed in the US.

To achieve this objective, David Makowski, Martin Lechenet (data analyst at DC), and I submitted a research proposal to the ANR Project Cland. Laurine Chevalier, a master's degree student, worked on implementing the OFRN and was co-advised by Martin Lechenet and I.

Farmer members participated in choosing of the alternative crops and were able to select which one to implement on their farm. The experimental design was the replicated strip. A total of 15 farmers were involved in the network: seven and eight farmers tested soybean and sunflower, respectively, with two seeding rates (the lower seeding rate as the control). In addition to the crop productivity, the yield components for the soybean and sunflower are part of the protocols. Farmers will handle only the yield measurement and the yield components were measurement by the graduate student and some staff. At this stage of the project only the number of plant/m<sup>2</sup> was measured. The results will be shared with the farmers through an interactive Html report made with Rmarkdown where farmers will have access to a summary of all the trials and specific information to their trial.

I contributed to this project by presenting the concept of OFRN to a group of farmers members and technicians from DC, providing recommendations (*i.e.*, recruiting members, experimental design, equipment needed, communication of the results) for running an

OFRN and co-advising Laurine Chevalier. A short stay at DC scheduled in spring 2020 to help with the implementation of the trials was canceled due to the pandemic situation.

The first implementation of an OFRN by DC was successful and well appreciated by farmers and technical staff. They plan to use this concept to test different catch crops for a methanization project.

### **Acknowledgments**

This research was supported by the Chateaubriand Fellowship – Embassy of France in the United States. I acknowledge Martin Lechenet and Laurine Chevalier to turn our collaboration into a great experience. I acknowledge the French stakeholders for their time, help, and interesting insight. I also thank the INRAE UMR Agronomy, in particular Michelle Fanucci, for organizing my stays in France.

## LONG RESUME EN FRANÇAIS

### INTRODUCTION GENERALE

L'expérimentation en parcelle agricole permet de conduire des essais en conditions réelles de production, ce qui joue un rôle essentiel dans l'acquisition de connaissances des systèmes agricoles, et peut amener à de solides recommandations agronomiques. Le concept d'expérimentation en parcelle agricole n'est pas nouveau, mais il se généralise en raison d'un intérêt croissant parmi les agriculteurs, chercheurs et agronomes, et grâce à la mise à disposition de moniteurs de rendements par les agriculteurs. L'expérimentation en parcelle agricole permet de tester des nouveaux produits et de nouvelles pratiques culturales dans les champs des agriculteurs, principalement par le biais du concept d'expérimentation en réseau d'agriculteurs (ci-après appelé ERA).

L'ERA regroupe plusieurs agriculteurs qui échangent leurs connaissances et expériences, et répond à des questions agronomiques par la mise en place d'un protocole expérimental commun en utilisant l'équipement agricole des agriculteurs. L'ERA aide les agriculteurs à améliorer la productivité, l'efficacité, et la rentabilité (Pruss et al., 2005; Moayedi and Azizi, 2012). La plupart des ERA existant sont dirigés par des universités, des instituts de recherche ou technique ou des entreprises privées. Le dispositif expérimental en bandes avec répétitions est le plus répandu et comprend une nouvelle pratique agricole (par exemple : densité de semis, écartement des rangs, traitement contre les maladies ou ravageurs) et la pratique standard de l'agriculteur (ou témoin non traité). Les essais sont répétés chez plusieurs agriculteurs et plusieurs années afin de capturer différentes conditions pédoclimatiques d'une zone agricole donnée.

L'approche analytique est un élément essentiel des ERA pour analyser efficacement les données collectées et communiquer les résultats aux agriculteurs et aux chercheurs-agronomes. Les ERA peuvent fournir au moins deux niveaux d'information sur les performances d'une pratique agricole : au niveau du réseau et au niveau de l'exploitation agricole. Généralement, les résultats sont présentés sous la forme de rapports individuels (*i.e.*, un rapport résumant les résultats d'un essai en parcelle agricole) (Kyveryga, 2019), mais les informations fournies sont limitées, difficiles à généraliser, et cela ne permet pas

de synthétiser les résultats recueillis dans les différents essais du réseau. De plus, la variabilité de la réponse de rendement est inexplorée alors qu'elle constitue un potentiel pour améliorer la prise de décision par les agriculteurs.

Il est essentiel de développer des méthodes statistiques appropriées qui peuvent simplifier les analyses multi-locations et estimer l'efficacité d'une pratique agricole au niveau de la parcelle et du réseau. De plus, une nouvelle façon de communiquer les résultats est nécessaire pour analyser simultanément les résultats des différents essais, et de faciliter la synthèse, l'interprétation et la communication des résultats. Les méthodes d'analyses statistiques et le développement d'un nouvel outil de visualisation devraient aider à la prise de décisions par les agriculteurs.

L'objectif général de cette thèse est de démontrer l'importance d'identifier des méthodes statistiques appropriées pour analyser et visualiser les données issues d'expérimentations en réseaux d'agriculteurs. Plus précisément, je me suis consacrée à l'analyse des ERAs gérés par Iowa Soybean Association, et d'un cas d'étude français. Ce travail de recherche représente le premier effort majeur de consolidation des résultats des ERAs.

Ce document contient un résumé et des figures clefs par chapitre de thèse. Une brève discussion générale est donnée à la fin de ce document ainsi que la liste des articles et communications produits pendant la thèse.

## CHAPITRE 1. CADRE DE VISUALISATION ET D'ANALYSE DE DONNEES ISSUES D'EXPERIMENTATIONS EN RESEAUX D'AGRICULTEURS

Les ERAs sont des organisations d'agriculteurs conduisant des essais agronomiques dans leurs champs avec des conditions réelles de production et suivant un protocole commun (Figure 1). Souvent, les résultats des essais sont partagés sous la forme d'un rapport individuel (*i.e.*, un rapport par essai) fournissant des informations descriptives et un test statistique simple. Bien que ces informations soient nécessaires, cela ne permet pas d'estimer la réponse de rendement à l'échelle du réseau, de comprendre la variabilité des réponses de rendement à l'aide de covariables (par exemple, la texture du sol, la précipitation, la date de semis) et, la comparaison de l'efficacité de la pratique agricole testée entre parcelles agricoles peut être biaisée. Nous avons développé un cadre d'analyse pour analyser simultanément plusieurs essais agricoles ayant un protocole expérimental commun et identifier les conditions dans lesquelles un traitement peut ou non être efficace (Figure 2). Ce cadre d'analyse utilise un modèle statistique à effets aléatoires par une approche bayésienne et fournit des estimations de réponse de rendement, et l'incertitude associée, à l'échelle du réseau et des essais. Ce cadre d'analyse a été mis en application pour 27 et 24 pratiques agricoles sur le maïs et le soja, respectivement. De plus, pour faciliter la communication des résultats, un application-web a été développée. Celle-ci comprend des fonctions de visualisation dynamique et interactive des données afin d'améliorer le partage d'informations, et est accessible à un large public pour faciliter l'accès aux résultats issus des réseaux d'expérimentations. Le développement de ce cadre d'analyse des données issues de ERA est nécessaire pour améliorer les prédictions à l'échelle des essais pour conduire à une production agricole plus rentable, durable et respectueuse de l'environnement. La prise de décision relative à pratique agricole donnée doit être basée sur l'ensemble des résultats issus du cadre d'analyse ainsi que la prise en compte d'un contexte économique et agronomique adéquat.

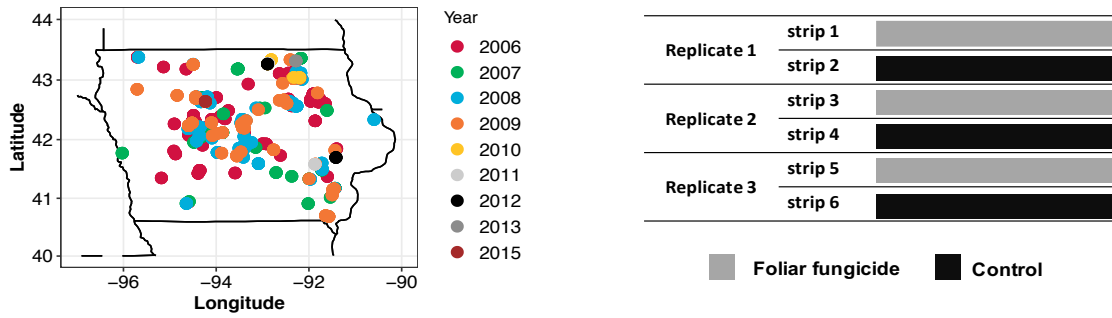
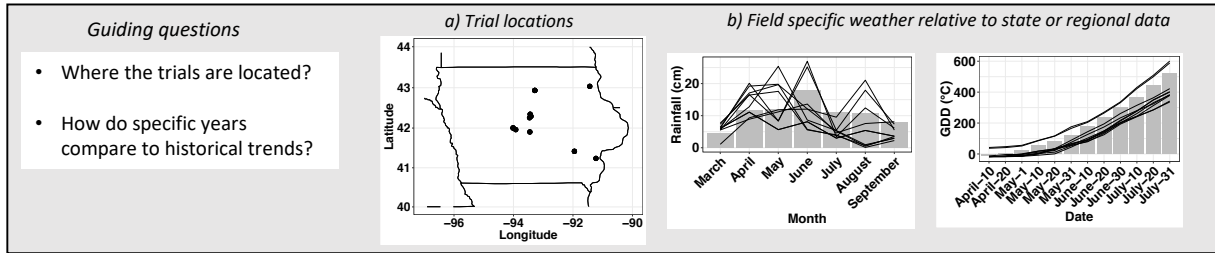


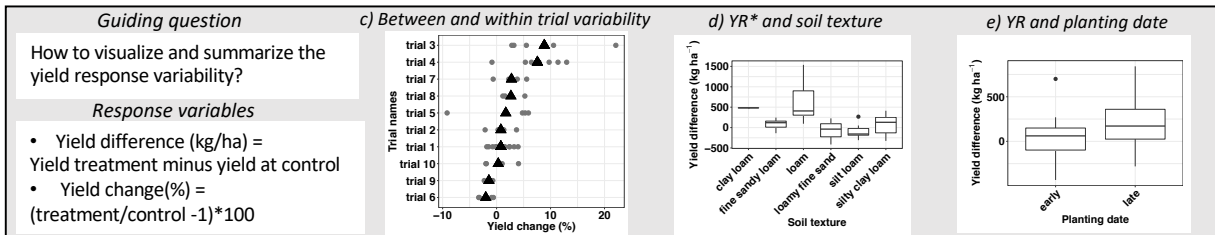
Figure 1 : Illustration schématique d'un réseau ERA où les applications de fongicides foliaires sur le soja ont été comparées à un témoin non traité. Chaque essai est représenté par un point (décrit à gauche) sur la carte de l'Iowa où l'année de mesure est distinguée par une couleur. Tous les essais suivent un modèle d'essai en bandes répétées (décrit à droite) et ont au moins trois répétitions.

## Framework for visualizing and analyzing on-farm research network data

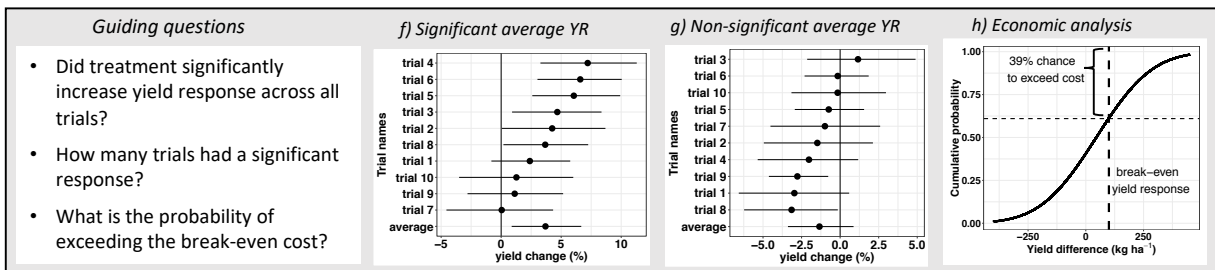
### Step 1: Location and description of environmental conditions



### Step 2: Descriptive analysis of the yield response



### Step 3: Statistical modeling: treatment effect within and across all trials



### Step 4: Statistical modeling: explain the yield response variability

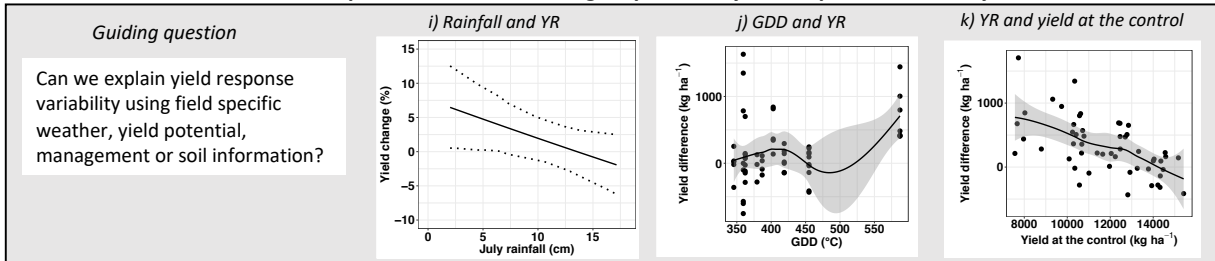


Figure 2 : Exemple de visualisation des résultats issus d'ERA. a) un point représente un essai en parcelle agricole appartenant à un ERA ; b) un essai est représenté par une ligne noire et des valeurs de référence (données historiques sur les 15 dernières années) sont représentées par les barres grises; c) un point représente une répétition et un triangle noir représente la valeur moyenne pour un essai; f et g) réponse de rendement estimée (YR) et l'incertitude associée représentée par un intervalle de crédibilité à 95% pour chaque essai et à l'échelle du réseau (en bas) (à partir d'un modèle Bayésien); h) la ligne en pointillée représente le seuil de rentabilité, la barre verticale représente la probabilité de dépasser le seuil de rentabilité et, la courbe représente la distribution cumulative de la réponse de rendement; i) une ligne noire représente la réponse de rendement estimée et la ligne en pointillée représente l' intervalle de crédibilité à 95% (à partir du modèle Bayésien); j) and k) les points représentent les réponses de rendement brutes et la ligne noire en pointillée

et la zone ombrée grise représentent l'estimation de la réponse de rendement et l'intervalle de crédibilité à 95% , respectivement, suite à une régression locale.

## **CHAPITRE 2. OUTIL INTERACTIF POUR VISUALISER, ANALYSER ET SYNTHÉTISER LES DONNÉES ISSUES D'EXPERIMENTATIONS EN RESEAUX D'AGRICULTEURS**

De nombreux ERA ont été conduits par Iowa Soybean Association et de ce fait, un volume conséquent de données a été accumulé depuis 2006. Les résultats de ces expérimentations sont présentés sous la forme de rapports individuels (i.e., un rapport résumant les résultats d'un essai en parcelle agricole), ce qui n'est pas optimal pour le partage des résultats à l'échelle du réseau. Une méthode de communication plus adaptée est nécessaire pour améliorer la dissémination des résultats et aider à la prise de décision. R Shiny, un package du logiciel R Studio) est une nouvelle technologie qui permet de convertir les analyses de données en applications web interactives. Pour la première fois pour les ERA, un outil web interactif appelé ISOFAST (« Interactive Summaries of On-Farm Strip Trials ») a été développé. ISOFAST présente simultanément l'ensemble des expérimentations traitant de la même pratique agricole afin de simplifier l'interprétation des résultats multi-locations et pluriannuels (Figure 3). Les analyses statistiques sont basées sur un modèle à effets aléatoires qui synthétise la réponse de rendement à la fois à l'échelle de la parcelle agricole et du réseau et génère des connaissances nouvelles pour les agriculteurs et agronomes. L'ensemble des analyses proposées dans le chapitre 1 sont inclus dans ISOFAST. L'interface d'ISOFAST (Figure 4) est intuitive et permet aux utilisateurs d'explorer les résumés des expérimentations, d'accéder à une analyse descriptive, aux résultats statistiques, et d'effectuer des analyses économiques. Au total, 27 et 24 pratiques agricoles sont renseignées dans l'application web pour le soja et le maïs, respectivement. La collaboration entre agriculteurs, agronomes et chercheurs est bénéfique aux ERA et améliore la prise de décision pour les futures campagnes agricoles. ISOFAST continuera d'évoluer selon les besoins, avec l'ajout de données provenant de pratiques culturelles et de produits nouveaux ou existants et avec des mises à jour des technologies de visualisation existantes. ISOFAST est disponible à l'adresse suivante : <https://analytics.iasoybeans.com/cool-apps/ISOFAST/>.



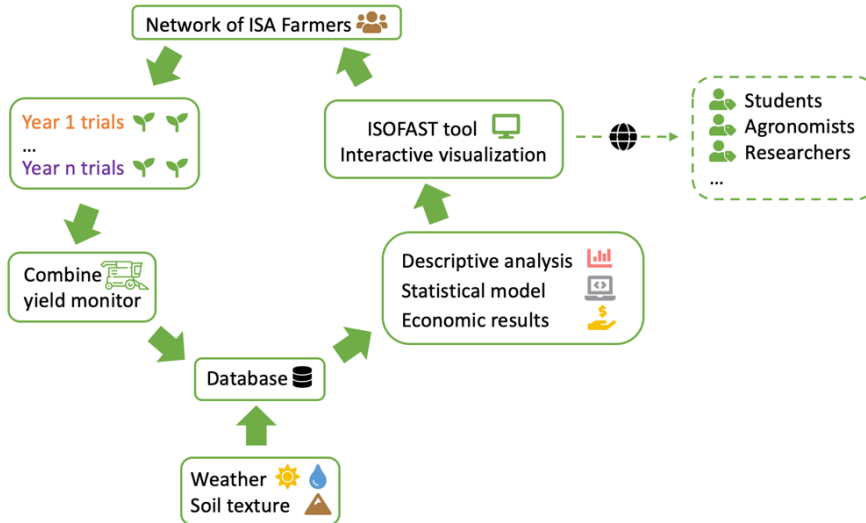


Figure 3. Architecture du système du réseau d'expérimentations conduit par Iowa Soybean Association

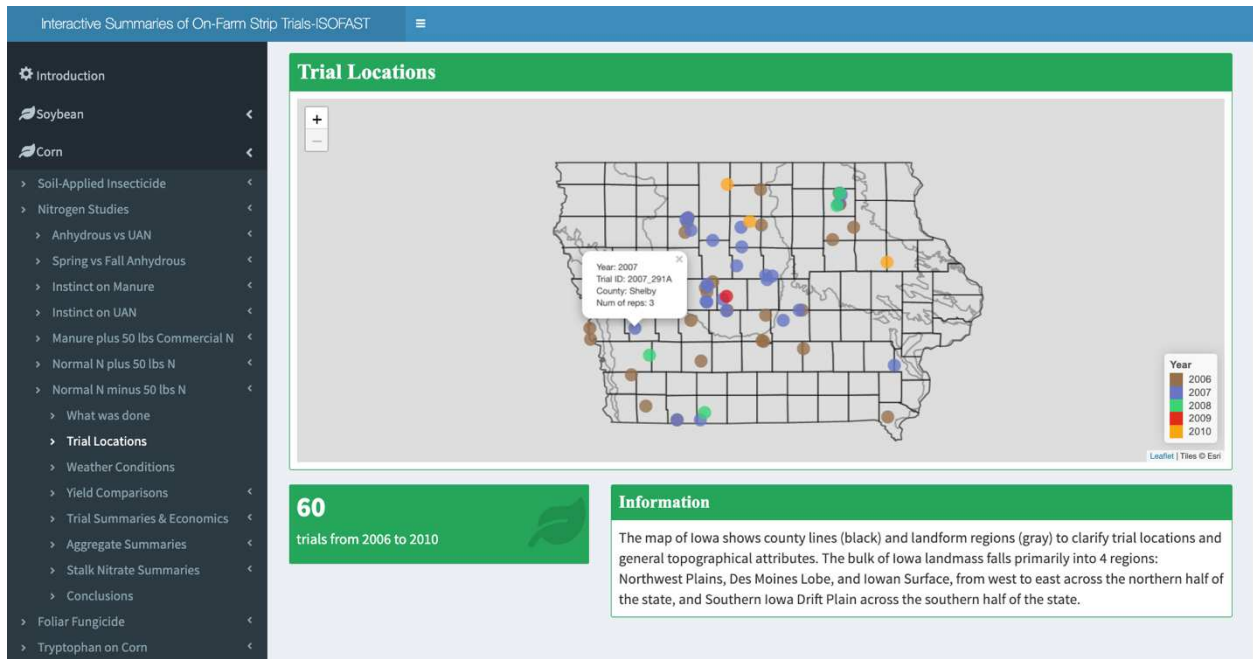


Figure 4. Capture d'écran de l'outil ISOFAST : barre de menu à gauche ; visuels à droite.

### **CHAPITRE 3. AU DELA DE LA TAILLE D'EFFET MOYENNE ; PRESENTER LES INTERVALS DE PREDICTIONS POUR LES ANALYSES D'EXPERIMENTATIONS EN RESEAU D'AGRICULTEURS**

Comme énoncé dans les deux chapitres précédents, un modèle à effets aléatoires a été utilisé pour estimer l'efficacité d'une pratique agricole comparée à un témoin non traité (ou pratique standard employée par l'agriculteur) dans chaque expérimentation du réseau, mais aussi d'estimer la performance moyenne à l'échelle du réseau. Ce chapitre se base sur ce même modèle pour calculer des intervalles de prédiction décrivant la gamme des réponses de rendement plausibles pour une nouvelle expérimentation (en dehors des expérimentations faisant partie du réseau) et de calculer la probabilité que la pratique agricole testée soit inefficace (perte de rendement) pour une nouvelle expérimentation (Figure 5). Ces calculs ont été réalisés par une approche fréquentiste et Bayésienne pour 26 pratiques agricoles (soit 26 différents réseaux dirigés par Iowa Soybean Association). Selon le niveau de variabilité inter-expérimentations, les intervalles de prédiction étaient 2,2 à 12,1 fois plus larges que les intervalles de confiance des réponses de rendement moyen (c'est-à-dire à l'échelle du réseau) pour toutes les pratiques agricoles testées (Figure 6). L'intervalle de prédiction devrait être systématiquement communiqué afin de fournir des informations supplémentaires pour des futures expérimentations. Néanmoins, les intervalles de prédiction doivent être interprétés avec prudence lorsque la variabilité inter-expérimentations est faible. L'utilisation des intervalles de prévision et de la probabilité qu'un traitement soit inefficace empêchera les agriculteurs d'avoir des attentes trop optimistes concernant le gain de rendement dans le cas où celui-ci est significatif à l'échelle du réseau.

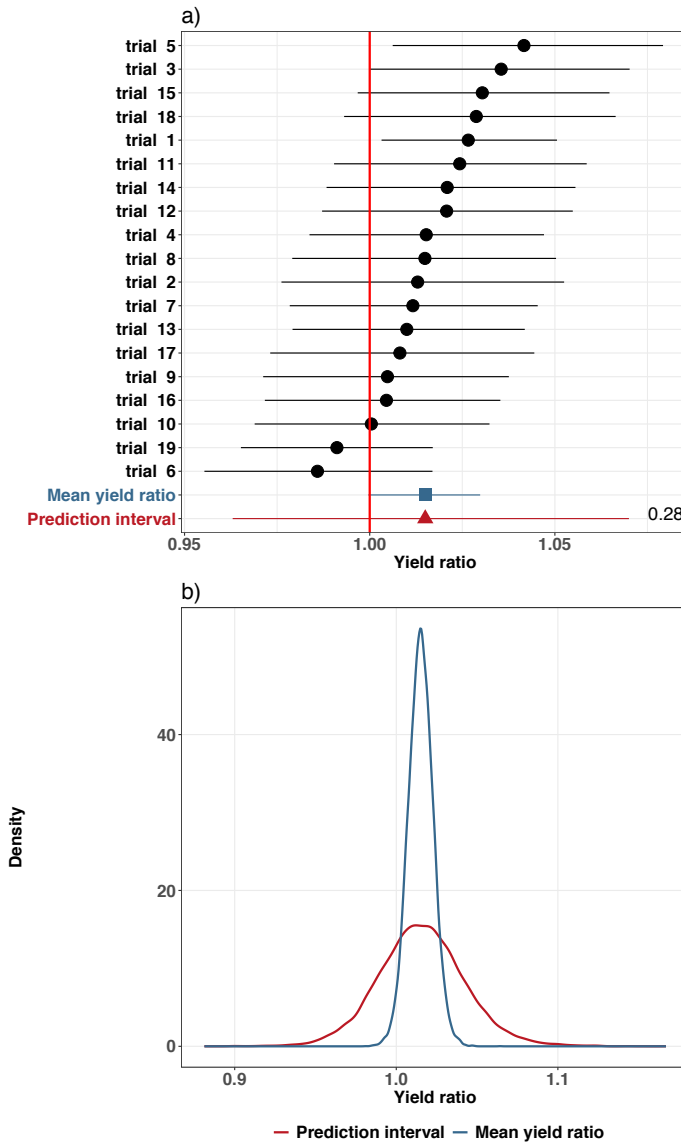


Figure 5 : Analyse d'un réseau de 19 expérimentations évaluant l'effet d'une densité de semis plus élevée (de 18 à 21 %) par rapport à la densité de semis couramment utilisée par les agriculteurs. a) réponse de rendement estimées pour chaque expérimentation et l'intervalle de crédibilité à 95% associé (en noir), réponse de rendement moyenne et l'intervalle de crédibilité à 95% associé (en bleu), et l'intervalle de prédiction à 95% (en rouge). Le nombre de répétitions est affiché entre parenthèses sur le côté gauche du graphique. La valeur numérique en bas à droite indique la probabilité d'un traitement inefficace (probabilité de ratio de rendement inférieur à 1) dans une nouvelle expérimentation ; b) distribution postérieure de la réponse de rendement moyenne (en bleu) et de l'intervalle de prédiction (en rouge). Les calculs ont été effectués à d'une approche Bayésienne.

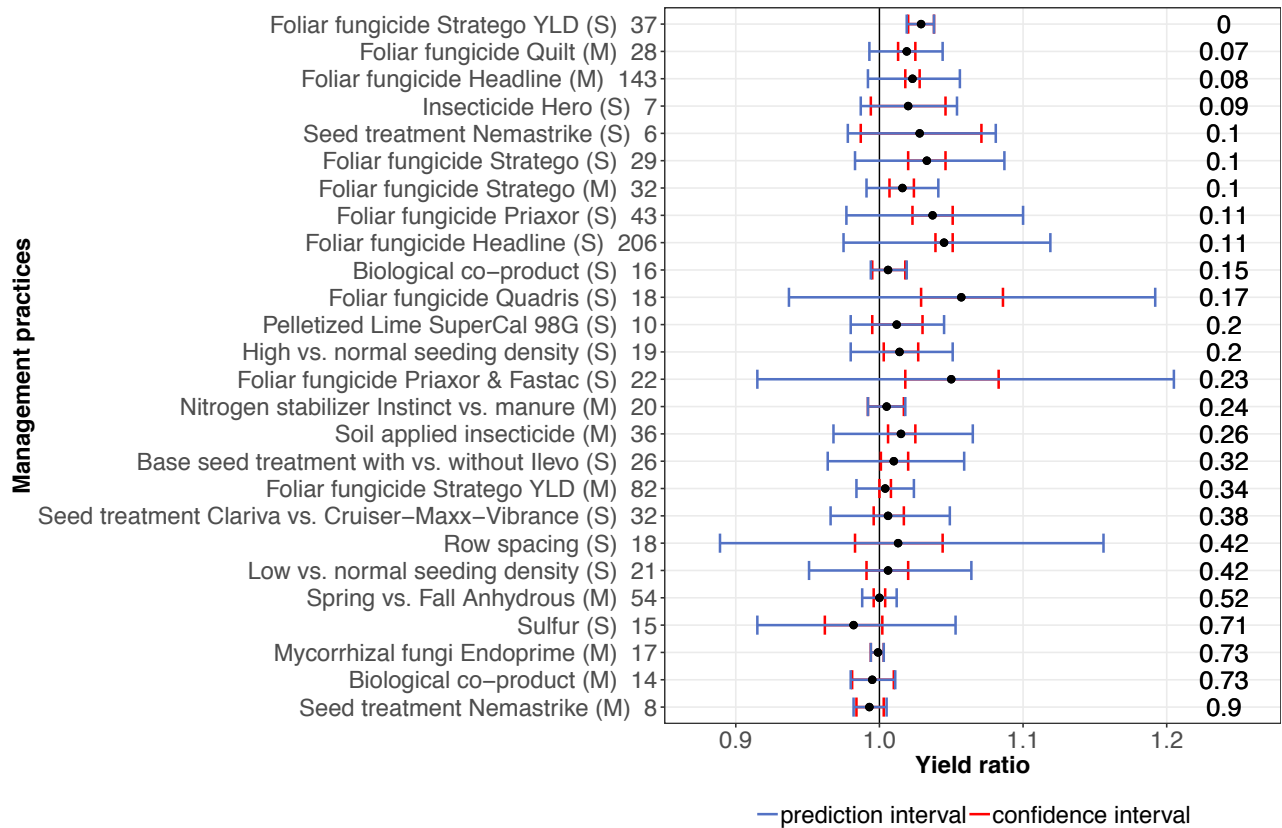


Figure 6 : Estimation des ratios de rendement moyens avec leurs intervalles de crédibilité à 95% (en rouge) et les intervalles de prédiction (en bleu). La probabilité que la pratique agricole testée soit inefficace dans un nouvel essai (ici, la probabilité d'un rapport de rendement inférieur à 1) est affichée à droite du graphique. Les pratiques de agriculteurs sont classées en fonction de l'augmentation de la probabilité d'inefficacité. S = soja, M = maïs. Les deux pratiques "anhydre de printemps vs. anhydre d'automne" (M) et "stabilisateur d'azote « Instinct » vs. fumier" (M) avaient le soja comme précédent de culture. Le nombre d'expérimentation est affiché sur le côté gauche du graphique

**CHAPITRE 4. DES EXPERIMENTATIONS EN RESEAU D'AGRICULTEURS REVELENT UN  
CONTROLE SIGNIFICATIF MAIS UNCERTAIN DU BOTRYTIS CINEREA PAR L'EMPLOI  
D'AUREOBASIDIUM PULLULANS ET DE BICARBONATE DE POTASSIUM  
EN VITICULTURE BIOLOGIQUE**

*Botrytis cinerea*, un champignon pathogène qui provoque la pourriture grise, peut diminuer le rendement de la vigne, réduire considérablement la qualité du vin et donc entraîner des pertes économiques importantes. Dans un contexte de sensibilisation accrue à la santé humaine et à l'environnement, les biopesticides sont une alternative potentielle aux traitements chimiques de synthèse pour produire du raisin et du vin dans le respect de normes alimentaires élevées. Toutefois, l'efficacité des biopesticides n'est pas bien connue et la recherche est nécessaire pour aider les viticulteurs à évaluer leur capacité à lutter contre les maladies de la vigne. Cette étude évalue l'efficacité de deux biopesticides présents dans le commerce, à base de bicarbonate de potassium et d'*Aureobasidium pullulans*, pour réduire l'incidence de la pourriture grise (c'est-à-dire la proportion de grappes de raisin malades). Cette étude utilise des données d'expérimentation en réseau, où des viticulteurs en agriculteur biologique, basés dans une grande région viticole du sud-ouest de la France, ont comparé dans leur vigne le potassium ou *Aureobasidium pullulans* à un témoin non traité pendant trois ans. Un modèle Bayésien linéaire généralisé multiniveaux, capable de prendre en compte la variabilité inter-expérimentation, a été utilisé pour évaluer l'efficacité de ces biopesticides sur l'incidence en fonction de l'intensité dans le témoin non-traité (c'est-à-dire la proportion de baies de raisin malades dans une grappe non traitée) et afin de déterminer si l'efficacité des biopesticides dépend de la pression de la maladie. Pour une intensité moyenne (c'est-à-dire 3 %), l'efficacité moyenne sur l'incidence de la maladie (c'est à dire à l'échelle du réseau) était égale à 20 % (95 % IC = [-0,1 ; 37,3]) et 13 % (95 % IC = [0,2 ; 24,7]) pour le bicarbonate de potassium et *Aureobasidium pullulans*, respectivement (Figure 7). Pour les deux biopesticides, l'efficacité sur l'incidence pour une nouvelle expérimentation (intervalle de prédiction) est très incertaine, mais le bicarbonate de potassium a eu une incertitude plus faible et un coût d'application plus bas par rapport à *Aureobasidium pullulans*. Nos résultats confirment que le bicarbonate de potassium est un biopesticide prometteur dans les conditions de culture

des vignobles biologiques du sud-ouest de la France, mais le degré d'incertitude indique la nécessité de poursuivre les recherches.

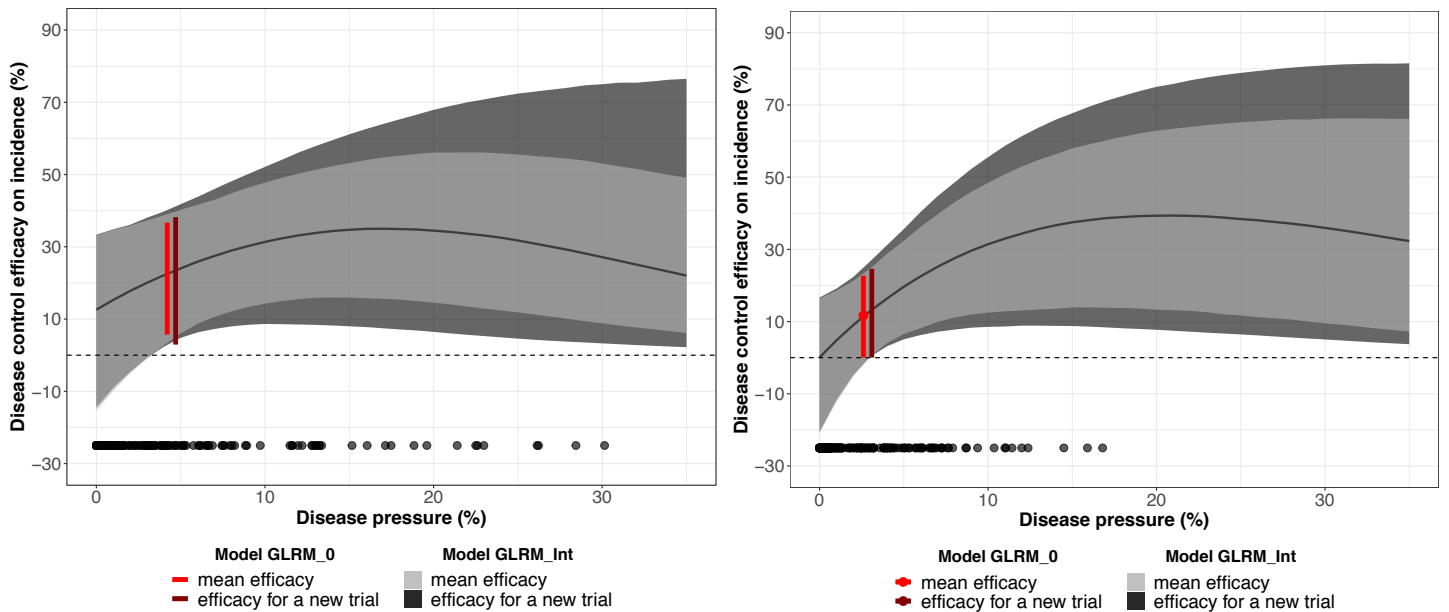


Figure 7 : Efficacité moyenne sur l'incidence et son intervalle de crédibilité à 95% (en gris clair) estimée à partir du réseau testant le bicarbonate de potassium (à gauche) et *Aureobasidium pullulans* (à droite). L'ombre gris foncé représente les intervalles de prédiction à 95 % (valeurs plausibles pour un nouvel essai). L'intervalle vertical rouge représente l'efficacité moyenne du contrôle de la maladie pour un modèle linéaire généralisé à plusieurs niveaux sans intensité comme covariable et l'axe x représente la pression de la maladie (intensité de la maladie dans le contrôle non traité). Les points noirs représentent la pression de la maladie observée pour toutes les répétitions.

## **CHAPITRE 5. COMPARAISON DES RENDEMENTS DU SOJA ET LE MAÏS ENTRE LES ESSAIS TESTANT LES FONGICIDES FOLIAIRES EN STATIONS EXPERIMENTALES ET EN PARCELLES AGRICOLE**

Les essais conduits en stations expérimentales et en parcelles agricoles sont utilisés en recherche agronomique afin de fournir, entre autres, des recommandations sur les pratiques agricoles. Ces deux d'échelles expérimentales se différencient par le matériel agricole employé et la taille de l'essai. Les essais en stations expérimentales sont principalement menés par des instituts de recherche ou des universités, tandis que les essais en parcelles agricoles sont conduits dans des conditions locales et gérés par les agriculteurs à l'aide de leur propre équipement agricole et de leurs pratiques agricoles. Comme des différences entre ces deux échelles expérimentales ont été mentionnées dans la littérature, l'extrapolation des résultats des essais en stations expérimentales à une plus grande échelle peut être difficile. Il est donc nécessaire de comprendre l'impact de l'échelle expérimentale sur l'effet des pratiques agricoles. Dans cette étude, nous avons évalué l'impact de l'échelle expérimentale des essais testant des fongicides foliaires sur le rendement du soja et du maïs, et identifié la cause des divergences potentielles. Nous avons utilisé des données de 976 essais sur le maïs et le soja recueillies dans cinq États des États-Unis (Figure 8). Pour le soja, notre comparaison de 479 essais en parcelles agricoles et de 83 essais en stations expérimentales n'a montré aucune différence significative entre les réponses moyennes de rendement (Figure 9). Pour le maïs, notre comparaison de 300 essais en parcelles agricoles et 114 essais en stations expérimentales a également donné des réponses similaires en termes de rendement moyen, mais l'écart type intra-essai pour les essais en parcelles agricoles était quatre fois plus petit que l'écart type intra-essai des essais en stations expérimentales (Figure 9). Comme les essais en parcelles agricoles et les essais stations expérimentales ont souvent des objectifs différents, la décision de choisir l'échelle expérimentale doit être basée sur les questions de recherche, les ressources foncières ainsi que les moyens économiques, techniques et humains.

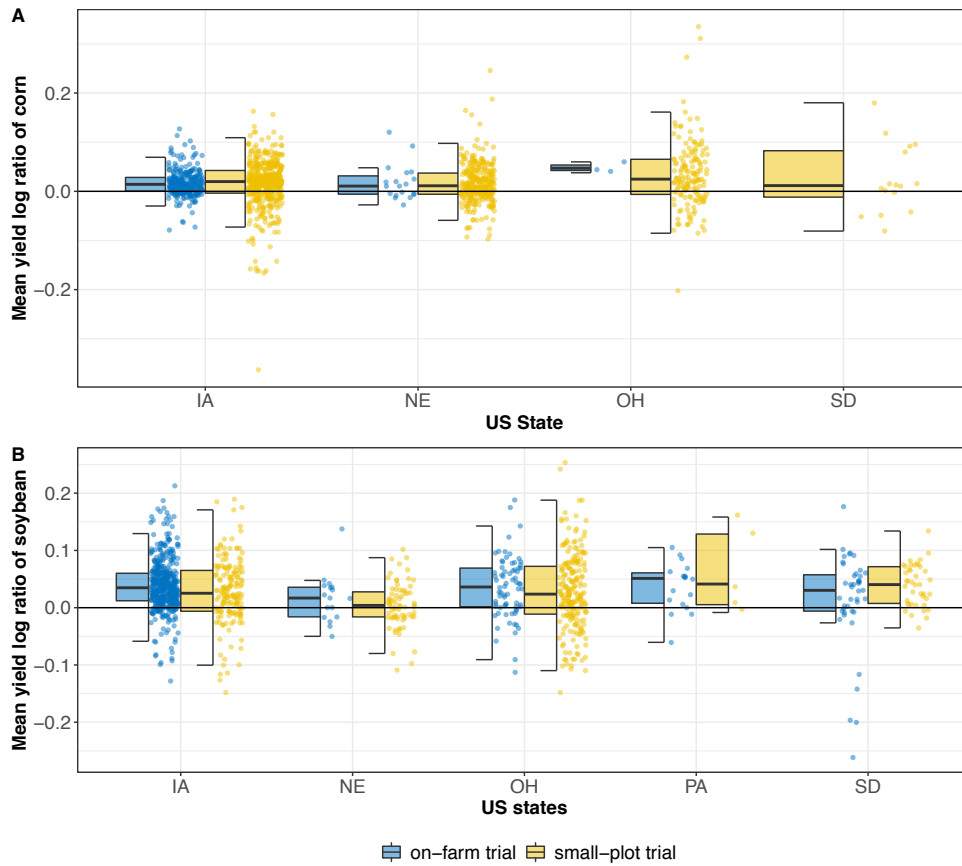


Figure 8 : Diagramme en boîte du log de rendement du maïs (A) et du soja (B) pour chaque État. Les essais en parcelles agricoles sont représentés en bleu et les essais en stations expérimentales en jaune. Un point représente la différence de rendement moyenne pour une combinaison essai (localisation) et traitement (fongicide foliaire).



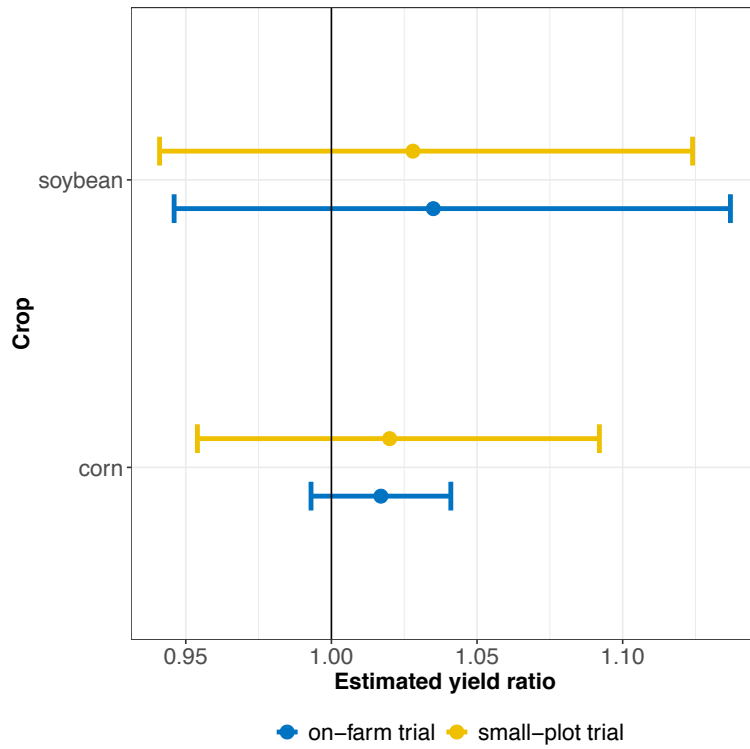


Figure 9 : Estimation du ratio de rendement global, et l'intervalle de crédibilité à 95% correspondant, pour les essais en stations expérimentales (en jaune) et en parcelles agricoles (en bleu) pour le soja (en haut) et le maïs (en bas).

## DISCUSSION GENERALE

L'objectif général de ces travaux de recherche était de faire avancer le domaine des statistiques agricoles en identifiant des méthodes statistiques appropriées pour les données issues d'ERA. Ce domaine de recherche est particulièrement intéressant car il met en relation la communauté des chercheurs et les agriculteurs, débouche sur des applications concrètes, permet de partager les résultats avec un large public, et aide les agriculteurs dans leur prise de décision. Comme ce travail comprend plus de 50 OFRN délivrés par l'Iowa Soybean Association, un OFRN français sur la vigne et explore les OFRN de quatre états différents des États-Unis, il était essentiel de se concentrer sur les méthodes génériques pour atteindre l'objectif général de cette thèse.

Ce travail est principalement basé sur des approches bayésiennes car il fournit un cadre plus intuitif que l'approche fréquentiste pour estimer les réponses de rendement au niveau du réseau et des essais. En outre, l'approche bayésienne permet de mieux estimer et quantifier l'incertitude des performances des pratiques agricoles. Dans l'ensemble, ces travaux de recherche permettent de mieux comprendre l'efficacité globale d'une nouvelle pratique agricole et représente le premier effort majeur de consolidation des résultats issues d'ERA. En effet, ISOFAST est le premier outil web interactif développé et lancé pour communiquer sur ERA, et l'utilisation de techniques de méta-analyses (c'est-à-dire modèle à effets mixtes, intervalles de prédiction) n'était, à notre connaissance, pas standard dans le domaine des ERA.

Certains projets de recherche en relation étroite avec cette thèse ou des projets élargissant le concept d'OFRN pourraient voir le jour à l'avenir. Par exemple, il semble nécessaire d'étudier et de comprendre les différentes sources de variabilité pour aider à réduire l'erreur résiduelle et ainsi améliorer les inférences et les prévisions statistiques. Un autre défi consistera à améliorer l'aspect générique des méthodes proposées afin de les adapter à d'autres zones agricoles ayant des questions de recherche différentes. Par exemple, au cours de ma thèse, j'ai passé un long séjour à l'INRA en France et des instituts français ont exprimé leur intérêt pour tester des outils d'aide à la décision, ou une combinaison de différentes pratiques agricoles, et aussi d'évaluer la durabilité

agronomique en incluant des indicateurs autres que le rendement des cultures. Ce travail de thèse été réalisé à différentes échelles, d'une grande région viticole située en France à cinq états américains. Néanmoins, de futurs projets, rassemblant des données des ERA de l'Université du Nebraska et de Iowa Soybean Association, verront le jour pour évaluer l'effet de différentes pratiques agricole à une plus grande échelle. L'accès à une plus grande quantité de données représentera une opportunité de comprendre les sources d'incertitude en tenant compte de la météo, de la topographie, des propriétés du sol et des pratiques agronomiques. Il sera également intéressant de comparer la réponse des rendements à l'échelle du champ, de la région (Nebraska et Iowa) et de l'exploitation agricole en utilisant les données d'enquêtes agricoles. D'autres pratiques agricoles autres que les fongicides foliaires (par exemple, la fertilisation et le travail du sol) pourraient également être comparées à l'échelle des essais de stations expérimentales et de parcelles agricoles afin de comprendre les causes des écarts de rendement et de mieux éclairer la prise de décision au niveau de l'exploitation.

### Liste des articles

On-farm trials reveal significant but uncertain control of *Botrytis cinerea* by *Aureobasidium pullulans* and potassium bicarbonate in organic grapevines

A. Laurent, D. Makowski, N. Aveline, S. Dupin, F.E. Miguez  
Frontiers in Plant Sciences, section Plant Pathogen Interactions. Accepted with revisions in December 2020

Interactive web-based data visualization and analysis tool for on-farm research networks. 2020, Research Synthesis Method, special issue Data Visualization

A. Laurent, X. Lyu, P. Kyveryga, D. Makowski, H. Hofmann, F.E Miguez  
DOI : 10.1002/jrsm.1440

Going beyond mean effect size; presenting prediction intervals for on-farm network trial analyses. 2020, European Journal of Agronomy, special issue Evidence Synthesis

A. Laurent, F. Miguez, P. Kyveryga, D. Makowski.  
DOI : 10.1016/j.eja.2020.126127

A framework for visualization and analysis of agronomic field trials from on-farm research networks. 2019, Agronomy Journal, special issue On-Farm Research

A. Laurent, P. Kyveryga, D. Makowski, F.E Miguez  
DOI : 10.2134/agronj2019.02.0135

### **Liste des communications**

Going beyond mean effect size; assessing the risk of ineffective treatment in on-farm trial network. ASA, CSSA & SSSA International meeting, November 04-07 2019, San Antonio, USA

A. Laurent, P. Kyveryga, F. Miguez, D. Makowski.

Interactive graphics and analysis of on-farm research network data. ASA, CSSA & SSSA International meeting, November 10-13 2018, Baltimore, USA

A. Laurent, F. Miguez, D. Makowski, P. Kyveryga.

Analysis of soybean data from on-farm research network to assess the efficacy of new management practices Research Seminar "Recent advances in soybean agronomic research", June 28 2018, Paris, France

A. Laurent, P. Kyveryga, D. Makowski, F. Miguez.

Share and summarize data from on-farm research network through a web application: a way to make agronomic decision

Production Agriculture Symposium, University of Minnesota, February 15 2018, Minneapolis, USA

A. Laurent, P. Kyveryga, D. Makowski, F. Miguez.

Analysis of on-farm data using a Bayesian hierarchical model, ASA, CSSA & SSSA International meeting, October 22-25 2017, Tampa, USA

A. Laurent, P. Kyveryga, D. Makowski, F. Miguez.

### **Liste des posters**

On-line summaries of the ISA On-farm network: How to make agronomic decisions from on-farm trials?, Production Agriculture Symposium, February 15th 2018, University of Minnesota, Minneapolis, USA

A. Laurent, F. Miguez, P. Kyveryga.

**Titre :** Visualisation et analyse de données issues d'expérimentations en réseaux d'agriculteurs pour tester l'efficacité de pratique agricole

**Mots clés :** modèle statistique, réseau d'agriculteurs, expérimentation en parcelles agricoles, analyse Bayésienne, visualisation de données

**Résumé :** L'expérimentation en réseau d'agriculteurs permet de conduire des essais agronomiques dans des conditions réelles de production. L'intérêt est grandissant car cela représente une opportunité pour mettre en place des essais testant de nouveaux produits et de nouvelles pratiques agricoles dans les champs des agriculteurs. Généralement, les résultats sont présentés sous la forme de rapports individuels (i.e., un rapport résumant les résultats d'un essai en parcelle agricole), mais les informations fournies sont limitées, difficiles à généraliser, et cela ne permet pas de synthétiser les résultats recueillis dans les différents essais. De plus, la variabilité de la réponse de rendement est inexplorée alors qu'elle constitue un potentiel pour améliorer la prise de décision par les agriculteurs. L'objectif général de cette thèse est de démontrer l'importance d'identifier des méthodes statistiques appropriées pour analyser et visualiser les données issues d'expérimentations en réseaux d'agriculteurs. Plus précisément, je me suis consacrée à l'analyse des réseaux gérés par Iowa Soybean Association, et d'un cas d'étude français. Un cadre d'analyse des données a été développé pour analyser simultanément plusieurs essais agricoles ayant un protocole expérimental commun et identifier les conditions dans lesquelles un traitement peut ou non être efficace. Ce cadre d'analyse utilise un modèle statistique à effets aléatoires par une approche bayésienne et fournit des estimations de réponse de rendement à l'échelle du réseau et des essais. Ce cadre d'analyse a été mis en application pour 27 et 24 pratiques agricoles sur le maïs et le soja, respectivement, au sein d'une application web. Celle-ci comprend des fonctions de visualisation dynamique et interactive des données afin d'améliorer le partage d'informations, et est accessible à un large public pour faciliter l'accès aux résultats issus des réseaux d'expérimentations.

Un modèle statistique à effets aléatoires a été utilisé pour calculer des intervalles de prédiction décrivant la gamme des réponses de rendement plausibles pour un nouvel essai (en dehors de l'échantillon testé), et calculer la probabilité que la pratique agricole testée soit inefficace sur le rendement dans une nouvelle parcelle agricole. Selon le niveau de variabilité inter-essais, les intervalles de prédiction étaient 2,2 à 12,1 fois plus larges que les intervalles de confiance des réponses de rendement moyen (c'est-à-dire à l'échelle du réseau) pour toutes les pratiques agricoles testées. L'utilisation des intervalles de prévision et de la probabilité qu'un traitement soit inefficace empêchera les agriculteurs d'avoir des attentes trop optimistes concernant le gain de rendement dans le cas où celui-ci est significatif à l'échelle du réseau. Le cadre d'analyse des données a été adapté à un réseau d'expérimentation français axé sur l'efficacité des produits de biocontrôle contre le *Botrytis cinerea*, le bicarbonate de potassium et *Aureobasidium pullulans*, sur la vigne en production biologique. Les résultats sont en faveur du bicarbonate de potassium car son efficacité sur l'incidence à l'échelle du réseau est plus élevée qu'*Aureobasidium pullulans*, pour une intensité de la maladie variant de 0 à 10%. Pour ces deux produits de biocontrôle, l'efficacité sur l'incidence pour un nouvel essai est très incertaine pour les niveaux d'intensité de maladie supérieurs à 15%. Enfin, ce travail de thèse a permis d'étudier la cause des écarts potentiels entre les résultats de rendement obtenus en stations expérimentales et en expérimentations aux champs. L'objectif est de mieux informer la prise de décision au niveau de l'exploitation et d'adapter l'extrapolation des résultats. Ce travail de thèse représente le premier effort majeur de consolidation des résultats d'expérimentation en réseaux d'agriculteurs et fournit des connaissances pour aider à la décision en termes de pratique agricole.

**Title :** The analysis of data from on-farm research network: Statistical approaches to test the efficacy of management practices and data visualization

**Keywords :** statistical model, farmers network, on-farm research, Bayesian analysis, data visualization

**Abstract :** An on-farm research network is an organization of farmers that conducts agronomic experiments under local conditions. There is growing interest in on-farm research networks because they provide the infrastructure needed to test new products and management practices in farmers' fields. Often, the results are usually presented as individual reports (i.e., a report summarizing the outcome for one trial), but this provides limited information difficult to generalize and does not allow presenting, in a synthetic way, all the results collected from the different trials. Moreover, there is an unexplored potential in detecting yield response variability patterns for better decision making. The overall objective of this thesis is to demonstrate the importance of identifying appropriate statistical methods for analyzing and visualizing on-farm research network data. Specifically, I focused on analyzing the on-farm research networks managed by the Iowa Soybean Association, and an adaptation was made with a French case-study. A data-analytics framework was developed to analyze multiple trials that use a common protocol and identify the conditions where an imposed treatment may or may not be effective. This framework used a random-effect model through a Bayesian approach and returned yield response estimates at the network and trial levels. The framework was implemented through a web-application for 51 different management practices on corn and soybean. The web-application includes dynamic data visualization features to enhance communication and information sharing, and is accessible to a broad audience to improve accessibility to on-farm research insights.

A random-effects statistical model was used to compute prediction intervals describing a range of plausible yield response for a new (out-of-sample) trial, and compute the probability that the tested management practice will be ineffective in a new field. Depending on the level of between-trial variability, the prediction intervals were 2.2–12.1 times larger than confidence intervals for the estimated mean yield responses (i.e., at the network level) for all tested management practices. Using prediction intervals and the probability of ineffective treatment will prevent farmers from over-optimistic expectations that a significant effect at the network level will lead with high certainty to a yield gain on their farms. The data-analytic framework was adapted to a French on-farm research network focusing on the efficacy of biocontrol agent products against *Botrytis cinerea*, potassium bicarbonate and *Aureobasidium pullulans*, on organic vine. The results favored potassium bicarbonate as its efficacy on incidence at the network level is higher for diseased intensities between 0% and 10% than for *Aureobasidium pullulans*. For both biocontrol agents, the efficacy on incidence for a new trial is highly uncertain for intensity levels higher than 15%. Finally, this research investigated the impact of experimental plot scale (i.e., small-plot scale and field scale) on the effect of management practice on crop yield and identified the cause of potential discrepancies to inform on-farm decision-making better and adapt the extrapolation of the results. Taken together, this research represents the first major effort in consolidating results from on-farm research network and provides insight to make better farming management decisions.