OPTIMIZING SITE-SPECIFIC NITROGEN FERTILIZER MANAGEMENT BASED ON MAXIMIZED PROFIT AND MINIMIZED POLLUTION

by

Paul Briggs Hegedus

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Ecology and Environmental Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2022

DEDICATION

This work is dedicated to my wife Meg, who's unwavering support from start to finish has made this dissertation possible. Additionally, to our dog Ike, without whom I wouldn't have taken daily walks which served as a source of stress relief and inspiration.

ACKNOWLEDGEMENTS

I must first thank Dr. Cathy Zabinski, my undergraduate advisor, without whom I would not have knocked on the door of Dr. Bruce Maxwell, who immediately and enthusiastically welcomed me into his lab. There are not enough pages to describe my thanks to Bruce, who encouraged me to come to graduate school and became my major advisor, mentor, and friend.

I also must thank my co-advisor, Dr. Stephanie Ewing, for guiding my research into one of farmer's most important resources, soil. Thank you for always pushing me to learn more and for sharing your knowledge. This dissertation also wouldn't have come together without the feedback and edits from the rest of my committee, Dr. Patrick Carr, Dr. Robert Payn, and Dr. Eric Sproles. I also must thank Dr. Taro Mieno for his insights on agricultural economics.

This dissertation wouldn't have been possible without Gary and Paul Broyles, Chuck Merja, and Jess Wood. All of them have sacrificed their time (and sometimes money!) to implement my experiments and collect high quality data. I will cherish my conversations with them and the valuable knowledge about farming that I would never have received otherwise. I must also thank the undergraduate students who tirelessly aided in the preparation of thousands of soil samples crucial to this work. And a big thank you to Dr. Tobias Koffman who spent countless hours analyzing the mountain of soil samples.

Additionally, my utmost thanks to my Agroecology lab mates, Sasha Loewen and Hannah Duff, who helped me in countless ways and made this process fun and enjoyable. Similarly, thank you to the students in the Ewing and Sheppard Labs for their friendship and support. Finally, I must thank my parents, who instilled the value of education in me and provided me with the privilege and opportunity to pursue a graduate degree.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

TABLE OF CONTENTS CONTINUED

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

Table Page

LIST OF TABLES CONTINUED

LIST OF TABLES CONTINUED

LIST OF FIGURES

LIST OF FIGURES CONTINUED

LIST OF FIGURES CONTINUED

ABSTRACT

Application of nitrogen fertilizers beyond crop needs contributes to nitrate pollution and soil acidification. Excess nitrogen applications are most prevalent when synthetic fertilizers are applied at uniform rates across fields. Precision agroecology harnesses the tools and technology of variable rate precision agriculture, a common but underutilized management strategy, to make ecologically conscious decisions about field management that promote economic and environmental sustainability. On-farm precision experimentation provides the basis for making data driven ecological management decisions through the field-specific assessment of crop responses. This dissertation work used on-farm experimentation with variable nitrogen fertilizer rates, combined with intensive data collection and data science, to address the main objective of this dissertation: development and evaluation of optimized nitrogen fertilizer management on a subfield scale, based on maximization of farmer net-returns and nitrogen use efficiency. The response of winter wheat yield and grain protein concentration to rates of nitrogen fertilizer application varied among fields, and across time, which influenced the model form used to characterize the relationships of grain yield and quality to fertilizer within a field. Machine learning approaches, such as random forest regression, tended to provide the lowest degree of error when forecasting future crop responses. Machine learning also demonstrated its utility for use in agronomic applications, as a support vector regression model provided the most accurate predictions of nitrogen use efficiency on a subfield scale. Crop response and nitrogen use efficiency models were integrated into a decision-making framework for optimized site-specific based nitrogen fertilizer management based on between maximized profits and minimized potential of nitrogen loss. Simulations of optimized site-specific nitrogen fertilizer management compared to farmer's status quo management showed a 100% probability across all fields tested that that mean net-return from the site-specific approaches were more profitable than applications of farmer selected nitrogen fertilizer rates. However, even while considering minimization of the potential for nitrogen loss when identifying optimum nitrogen fertilizer rates, there was field specific variation in the probability that site-specific, compared to farmer selected, nitrogen fertilizer management reduced the total amount of nitrogen applied across a field.

CHAPTER ONE

INTRODUCTION

Prologue

Since the advent of agriculture, human culture and lifestyle have been influenced by innovations and technologies (Gross, 2013). Agriculture was partnership between humans and the land that endured for thousands of years; however, agriculture changed dramatically when the introduction of the potato to Europe in the 1400s catalyzed the first use of concentrated fertilizer to boost yields, in the form of guano, and the first use of pesticides, in the form of arsenic to control potato beetles (Wills, 2016). By the mid 1900's, the Green Revolution had initiated a new path for agriculture by introducing industrialized commodity food production to the world. The Green Revolution began in the 1940's and 1950's in Latin America, instigated by the Rockefeller Foundation's "Mexican Agricultural Program", but is more widely associated with developments in India during the 1960's and 1970's that revolutionized rice production. The Green Revolution led to a package of intensifying agronomic practices, including high yielding crop varieties, fertilizers, herbicides, and irrigation (Cassman, 1999; Talukder et al., 2020). New crop varieties had a reduced time from planting to maturity but required high amounts of nitrogen fertilizer to achieve high yields. This led to the increased application of fertilizers for greater production, made possible by genetic enhancements such as the strong stalks that prevent lodging (Cassman, 1999). Heavy investment in irrigation, to further increase yields and reduce water stress, also enhanced fertilizer response (Cassman, 1999). Practices of the Green Revolution spread across the globe, resulting in an increase in global cropping intensities and global food security (Struik, 2017).

The Green Revolution's intensification of wheat, rice, and corn cropping systems underpinned a global 56% increase in production between 1965 and 1985, and 20% between 1985 and 2005 (Foley et al., 2011). The increase in production was made possible by high yielding crop varieties and the package of practices that accompanied the new crops being planted in the mid 1960's. The increase of yields from 1967 prevented land use change to agriculture, with estimates that without an increase in yields a 20% increase in production would have required three times as much cultivated land (Cassman, 1999). Hectares of cultivated land only increased by 2.4% between 1985 and 2005, indicating that production increases were related to a 25% increase in global crop yields (Foley et al., 2011; Gliessman & Engles, 2014). The increase in production lowered the relative cost of food, with Americans spending half as much of their income on food in 2022 as in the 1950's (USDA ERS, 2022). Increased food security and crop production from the Green Revolution contributed to the rise in the global population which is estimated to reach 10 billion people in 2050 (Tilman et al., 2002).

The Green Revolution spurred specialization among farmers to a few crops, including wheat, rice, corn, and soybeans, with lucrative markets and high demand. This specialization led to monocultures dominating the agricultural landscape, creating economies of scale, and allowing farmers to manage fields of increasing sizes. While monocultures have increased profits and utilize machinery more efficiently, they have decreased diversity and resilience of agroecosystems to disease and economic perturbations (Gliessman & Engles, 2014). Despite covering more ground, fields are typically treated uniformly with synthetic fertilizers to ensure crop nutrient needs are met, akin to purchasing insurance. These cheap fertilizers are fossil fuel intensive in terms undesired losses of mobile chemical forms to the surrounding ecosystem, contributing to environmental damage, and decreases in net returns when applied fertilizer does not contribute to crop growth (Gliessman & Engles, 2014; Tilman et al., 2002).

Modern agricultural management practices, such as the application of fertilizer, stemming from the Green Revolution have been beneficial to avoiding more acute food crises over the past few decades, yet ironically have induced a greater food crisis. Despite the increases in production, the nature of industrialization in agriculture threatens the gains reaped from the process. The heavy reliance on machinery and chemicals has begun to affect the environment, limiting production due to ecological impacts of industrial farming (Foley et al., 2011). Additionally, the cost of machinery and chemicals requires large amounts of capital, spawning the formation of rural credit institutions to lend money to farmers, and consequently fueling a cycle of debt (Gliessman & Engles, 2014). The Green Revolution has ultimately impacted rural livelihood and society, in addition to greenhouse gas emissions, soil and water acidification, pesticide poisonings, and biodiversity loss, degrading arable land (Carlisle et al., 2019).

A third of the all land on earth is already moderately to highly degraded, with millions of hectares further damaged each year (Gliessman, 2016). Urbanization and aridification further reduce the amount of land available on which we can grow food (Gliessman, 2016). Additionally, the industrialization of agriculture has had untold impacts on rural communities and global inequality as farm work becomes more mechanized (DeLonge et al., 2016). Decisions are made based on advice from chemical companies and consultants, and farmers’ profits have declined as processors, distributers, and retailers take a bigger share. Of every dollar spent on food, farmers receive 16 cents, decreasing the motivation for young people to farm independently and increasing economic hardships on those who attempt to keep family farms operational (Gliessman & Engles, 2014).

As the tenants of agroecosystems, farmers and ranchers determine the way in which 37.5% of the total land area on the Earth is managed (Lynch et al., 2021). However, despite the vast amount of land controlled by farmers, rural populations are declining (White, 2012). The amount of cultivated land in the United States has remained constant for the last 100 years, but the number of farms has decreased from 6.5 million to just over 2 million, resulting in farm size increasing as fewer farmers tend the land (Carlisle et al., 2019). Increased industry presence in agriculture and the widespread vacancies in agricultural jobs have resulted in the “deskilling” of the rural workforce as automated machinery copes with larger farm sizes and lack of available labor (Carlisle et al., 2019). However, the mechanization and automation of agriculture in recent decades has lent itself well to the historic goals of agriculture in developed nations; maximization of production and the maximization of profits (Coleman, 1989; Gliessman & Engles, 2014; Paul & Robertson, 1989).

However, goals of agriculture need to reflect the current physical and cultural climates. With the global population increasing without a proportional increase in the amount of land available to cultivate, future farmers and ranchers will be faced with the dilemma of feeding more people without increasing the acreage on which to grow crops, while natural resources crop production requires are threatened by the legacy of the Green Revolution (Tomich et al., 2011). Agroecology is a science, practice, and social movement centered on effecting sustainable change at all levels of the food system. The scientific discipline and practice of agroecology are the study of and application of ecological concepts and principles to agroecosystems, respectively (Duff et al., 2022). The movement applies agroecological principles of resilience, equity, and responsible governance to the elements of food systems at the farm level and beyond. Gliessman (2016) identifies five levels of food system change; (1) increasing the efficiency of practices on farms, (2) substituting alternative practices for industrial practices, (3) redesigning systems to function based on ecological, (principles, 4) reconnecting consumers and producers, and (5) reinventing the global food system based on equity, participation, democracy, and justice.

The majority of inefficient agricultural practices occur in a handful of nations, where most have the affluence required to leverage modern tools and technology for sustainable production of food, fuel, fiber, and feed (West et al., 2014). Agriculture has steadily modernized throughout the centuries by improvements in equipment and technology, recently by precision agriculture: a management strategy that utilizes temporal and spatial data to inform agricultural production (Gebbers & Adamchuck, 2010). Industrial agriculture in developed nations contributes the most to agricultural inefficiency and pollution, but has also been the primary adopter of precision agriculture tools and technology due to the size of industrial farm operations and ability to afford the expense required for precision agriculture equipment (Schimmelpfennig & Lowenberg-DeBoer, 2020).

The advent of precision agriculture led to the expectation that agricultural inputs would be used more efficiently and crop productivity would be maximized, assuming that chemical applications of inputs like nitrogen fertilizer would be reduced (McBratney et al., 2005). The adoption of precision agriculture led to the substitution of technology for local farmer knowledge and is feared for its potential to lead agriculture towards “ecological dystopias” (Altieri & Nicholls, 2020; Daum, 2021). These criticisms are not unfounded; but transitioning industrial agriculture towards an agroecological future will not be achieved by a great leap from current conventional practices to a complete sustainability. This change will require a stepwise progression where the tools and technology of precision agriculture can be co-opted for agroecological management, a strategy defined as precision agroecology (Duff et al., 2022).

As resource efficiency remains limited in areas with high technology adoption, greater access to the data returned from day-to-day farming operations can be exploited to increase resource efficiency and produce crops in a manner that reduces negative environmental impacts and increases the ecological and economic resilience of agroecosystems (Bucci et al., 2018). This dissertation focuses on the first step towards transitioning food systems towards greater sustainability by increasing the efficiency of agricultural inputs, specifically synthetic nitrogen-based fertilizer, through precision agroecological practices.

Justification and Research Goals

The industrialization of agriculture created an input-centric approach to farming that has resulted in significant non-point pollution through excess applications of synthetic chemicals like nitrogen fertilizer. Nitrogen fertilizer use has been linked to elevated nitrate levels in drinking water and acidification of agricultural soils (Canfield et al., 2010; Fowler et al., 2013; Jones & Olson-Rutz, 2020). Improving nitrogen use efficiency, defined as the proportion of available nitrogen, including applied fertilizer, that is taken up by a crop or retained in the root zone and used by future crops was hypothesized as an avenue by which to decrease nitrogen pollution.

Nitrogen that is lost from an agroecosystem due to volatilization, denitrification, surface runoff, or leaching are all sources of potential nitrogen pollution.

The state of Montana is affected by nitrate contamination of drinking water and soil acidification (Engel, 2012; John et al., 2017; Sigler et al., 2018). To reduce the contribution of Montana's dryland small grain agroecosystems to non-point nitrogen pollution, the pathways of nitrogen loss from these systems should be investigated. Site-specific, within-field variable application of nitrogen may reduce loss by applying fertilizer only where net-return, calculated from crop responses, is maximized at a point sufficient to justify the costs of applying the fertilizer. Predicting crop responses and nitrogen use efficiency at the time of nitrogen fertilizer application introduces the research challenge.

To address the challenge, on-farm experimentation was used where farmers participating in the research applied experiments in their production fields. On-farm experiments overcome the problem of translating results from plots at research centers to farms many miles away (Bullock et al., 2019; Maxwell & Luschei, 2005). Experimental variation of nitrogen fertilizer rates across a field, followed by growing season and harvest monitoring of grain yield and protein allow for local quantification of the parameters in the crop response and nitrogen loss functions. The experiments can be stratified according to factors known to vary within the field to account for within-field variation. To account for the temporal variation in crop response and nitrogen loss experiments were repeated over a number of years (Lawrence et al., 2015a). Furthermore, optimization that seeks to simultaneously maximize the probability of producer profits and minimize the probability of input pollution can identify economically and environmentally sustainable within-field site-specific nitrogen fertilizer rates.

The **goal of this dissertation** is to demonstrate and justify how, using precision agriculture techniques for nitrogen application and interpreting the data collected in an ecological context, can provide farmers with data-driven recommendations on site-specific management that is optimized to both maximize their profits and improve their nitrogen use efficiency. Achievement of this goal coincided with three themes: identifying, understanding, and optimizing. First, the variability of crop responses over space and time must be identified by asking how on-field experimention, applied with precision agriculture technology, inform farmers and crop managers about variation of winter-wheat production and quality in Montana small-grain agroecosystems? Answering this question requires understanding the degree that crop responses vary within fields, across fields, and across time, and what the drivers of production, quality, and net-returns are within fields for a given year.

After justifying the value of on-farm experimentation and potential for site-specific nitrogen fertilizer management, my next step was understanding how winter-wheat responds to experimentally varied nitrogen fertilizer rates to identify the methods for informing decisions. Mathematical models are required to understand crop responses to nitrogen fertilizer, but raise the question: what type of model is effective for predicting management outcomes? Accordingly, is one model that performed best for crop production or quality appropriate for all fields and for all years, or are different models were more appropriate for certain fields and years?

With the amount of data available on farms and being generated from farms, such as from remote sensors on satellites, producers now have substantial theoretically useful data available at the critical times when they need to make decisions on nitrogen fertilizer management. The status quo of modeling crop responses uses data from previous years to generate relationships between wheat yield or protein to nitrogen fertilizer which are applied to future years. Understanding the value of data up to the decision point, rather than constrained only to previous years, was addressed by asking how the period from which data were collected influences model performance and simulated management outcomes?

The conditions influencing nitrogen loss in dryland conventional Montana agroecosystems has been well identified by Sigler et al. (2018, 2020, 2022). However, nitrogen use efficiency (NUE) remains to be evaluated in these systems with experimentally varied nitrogen fertilizer rates. Gaining an understanding of how NUE varied spatially, and the factors influencing it, allowed asking whether it can be adequately modeled to inform site-specific nitrogen fertilizer decisions?

Finally, the broad question of the dissertation was how experimentally derived N fertilizer application rates at the meter scale, conducted with data informed technologies and novel sensing techniques, influence nitrogen efficiency and net return at the subfield scale in dryland small grain agroecosystems? The objective of the final chapter was to synthesize the information from the previous steps into an algorithm that optimized site-specific nitrogen application rates based on maximized net return and considering the cost of lower NUE that inherently comes with higher application rates. This final chapter wrapped all the previous work into an agroecological framework where the crop response models to experimentally variable nitrogen fertilizer rates were used to identify profit maximizing rates while understanding of how NUE responds to varied nitrogen fertilizer rates was used to generate site-specific optimum recommendations for farmers.

Research Objectives

1. Develop the conceptual basis for performing on-farm precision experimentation to aid in agroecological management.
2. Understand the field-specificity in crop responses to variable nitrogen fertilizer rates.
3. Identify how statistical forms of crop response models describe variation between fields and across years.
4. Assess data constraints related to crop response modeling.
5. Evaluate nitrogen use efficiency (NUE) and develop a subfield NUE model for dryland winter-wheat agroecosystems in Montana.
6. Develop optimized nitrogen fertilizer management on a subfield scale based on maximization of farmer net-returns and nitrogen use efficiency.

Acknowledgement of Bias

Beginning with bias in personal decisions behind my work, I must acknowledge the fact that I am an environmentalist and support the scientific evidence highlighting the anthropogenic effects on climate and the environment. I am also deeply committed scientifically to agriculture and it's sustainability from a multidisciplinary perspective I perceive a need in agriculture to sustain the resources it relies on for future production of food, feed, fuel, and fiber. The over-application of nitrogen fertilizer has been linked with numerous environmental issues in the scientific literature and has been recognized by governments around the world. In recognition of my preconceived notions of the systems I study; I have an implicit bias that we need to develop decision support systems for farmers that aid in maximizing the efficiency of nitrogen fertilizer.

However, there is the possibility that some of my sites aren't "leaky" enough regarding nitrogen to demonstrate the usefulness of the approach, and I might be biased in looking for trends and patterns in the data that satisfy that bias.

REFERENCES CITED

Altieri, M. A., & Nicholls, C. I. (2020). Agroecology and the reconstruction of a post-COVID-19 agriculture. *Journal of Peasant Studies*, 881–898. https://doi.org/10.1080/03066150.2020.1782891

Bucci, G., Bentivoglio, D., & Finco, A. (2018). Precision agriculture as a driver for sustainable farming systems: State of art in litterature and research. *Quality - Access to Success*, *19*(S1), 114–121.

Bullock, D. S., Boerngen, M., Tao, H., Maxwell, B., Luck, J. D., Shiratsuchi, L., Puntel, L., & Martin, N. F. (2019). The data-intensive farm management project: Changing agronomic research through on-farm precision experimentation. *Agronomy Journal*, *111*(6), 2736–2746. https://doi.org/10.2134/agronj2019.03.0165

Canfield, D. E., Glazer, A. N., & Falkowski, P. G. (2010). The evolution and future of earth's nitrogen cycle. *Science*, *330*(6001), 192–196. https://doi.org/10.1126/science.1186120

Carlisle, L., Montenegro de Wit, M., DeLonge, M. S., Iles, A., Calo, A., Getz, C., Ory, J., Munden-Dixon, K., Galt, R., Melone, B., Knox, R., & Press, D. (2019). Transitioning to Sustainable Agriculture Requires Growing and Sustaining an Ecologically Skilled Workforce. *Frontiers in Sustainable Food Systems*, *3*(November), 96. https://doi.org/10.3389/fsufs.2019.00096

Cassman, K. G. (1999). Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(11), 5952–5959. https://doi.org/10.1073/pnas.96.11.5952

Coleman, D. C., & Dec, N. (1989). *Special Feature : Ecology , Agroecosystems , and Sustainable Agriculture*. *70*(6), 12–13.

Daum, T. (2021). Farm robots: ecological utopia or dystopia? *Trends in Ecology and Evolution*.

DeLonge, M. S., Miles, A., & Carlisle, L. (2016). Investing in the transition to sustainable agriculture. *Environmental Science and Policy*, *55*, 266–273. https://doi.org/10.1016/j.envsci.2015.09.013

Duff, H., Hegedus, P. B., Loewen, S., Bass, T., & Maxwell, B. D. (2022). Precision Agroecology. *Sustainability*, *14*(106), 1–18. https://doi.org/https://doi.org/10.3390/su14010106

Engel, R. (2012). Volatilization losses from surface-applied urea during cold weather months. *Montana State University - Dept. LRES*, 1–10. https://doi.org/10.1017/CBO9781107415324.004

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., … Zaks, D. P. M. (2011). Solutions for a cultivated planet. *Nature*, *478*(7369), 337–342. https://doi.org/10.1038/nature10452

Fowler, D., Coyle, M., Skiba, U., Sutton, M. A., Cape, J. N., Sheppard, L. J., Jenkins, A., Grizzetti, B., Galloway, J. N., Leach, A., Bouwman, A. F., Butterbach-bahl, K., Stevenson, D., Amann, M., Voss, M., Fowler, D., Coyle, M., Skiba, U., Reis, S., … Voss, M. (2013). *The global nitrogen cycle in the twenty-first century*.

Gebbers, R., & Adamchuck, V. I. (2010). Precision Agriculture and Food Security. *Science*, *February*, 828–830.

Gliessman, S. (2016). Transforming food systems with agroecology. *Agroecology and Sustainable Food Systems*, *40*(3), 187–189. https://doi.org/10.1080/21683565.2015.1130765

Gliessman, S. R., & Engles, E. W. (2014). *Agroecology: The Ecology of Sustainable Food Systems* (3rd ed.).

Gross, M. (2013). The paradoxical evolution of agriculture. *CURBIO*, *23*(16), R667–R670. https://doi.org/10.1016/j.cub.2013.08.001

John, A. A., Jones, C. A., Ewing, S. A., Sigler, W. A., Bekkerman, A., & Miller, P. R. (2017). Fallow replacement and alternative nitrogen management for reducing nitrate leaching in a semiarid region. *Nutrient Cycling in Agroecosystems*, *108*(3), 279–296. https://doi.org/10.1007/s10705-017-9855-9

Jones, C., & Olson-Rutz, K. (2020). Soil Acidification: Management. *The Soil Scoop*, *198*(August), 1–2. https://doi.org/10.1016/j.still.2019.104534

Lawrence, P. G., Rew, L. J., & Maxwell, B. D. (2015). A probabilistic Bayesian framework for progressively updating site-specific recommendations. *Precision Agriculture*, *16*(3), 275–296. https://doi.org/10.1007/s11119-014-9375-4

Lynch, J., Cain, M., Frame, D., & Pierrehumbert, R. (2021). Agriculture's Contribution to Climate Change and Role in Mitigation Is Distinct From Predominantly Fossil CO2-Emitting Sectors. *Frontiers in Sustainable Food Systems*, *4*(February), 1–9. https://doi.org/10.3389/fsufs.2020.518039

Maxwell, B. D., & Luschei, E. C. (2005). Justification for site-specific weed management based on ecology and economics. *Weed Science*, *53*(2), 221–227. https://doi.org/10.1614/ws-04-071r2

McBratney, A., Whelan, B., Ancev, T., & Bouma, J. (2005). Future directions of precision agriculture. *Precision Agriculture*, *6*(1), 7–23. https://doi.org/10.1007/s11119-005-0681-8

Paul, E. A., & Robertson, G. P. (1989). Ecology and the Agricultural Sciences: A False Dichotomy? *Ecology*, *70*(6), 1594–1597. https://doi.org/10.2307/1938091

Schimmelpfennig, D., & Lowenberg-DeBoer, J. (2020). Farm Types and Precision Agriculture Adoption: Crops, Regions, Soil Variability, and Farm Size. *SSRN Electronic Journal*, 1–38. https://doi.org/10.2139/ssrn.3689311

Sigler, W. A., Ewing, S. A., Jones, C. A., Payn, R. A., Brookshire, E. N. J., Klassen, J. K., Jackson-Smith, D., & Weissmann, G. S. (2018). Connections among soil, ground, and surface water chemistries characterize nitrogen loss from an agricultural landscape in the upper Missouri River Basin. *Journal of Hydrology*, *556*, 247–261. https://doi.org/10.1016/j.jhydrol.2017.10.018

Struik, P. C. (2017). *Sustainable intensification in agriculture : the richer shade of green . A review*. https://doi.org/10.1007/s13593-017-0445-7

Talukder, B., Blay-Palmer, A., vanLoon, G. W., & Hipel, K. W. (2020). Towards complexity of agricultural sustainability assessment: Main issues and concerns. *Environmental and Sustainability Indicators*, *6*(March), 100038. https://doi.org/10.1016/j.indic.2020.100038

Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., & Polasky, S. (2002). *Agricultural sustainability and intensive production practices*. 671–677.

Tomich, T. P., Brodt, S., Ferris, H., Galt, R., Horwath, W. R., Kebreab, E., Leveau, J. H. J., Liptzin, D., Lubell, M., Merel, P., Michelmore, R., Rosenstock, T., Scow, K., Six, J., Williams, N., & Yang, L. (2011). Agroecology: A review from a global-change perspective. *Annual Review of Environment and Resources*, *36*, 193–222. https://doi.org/10.1146/annurev-environ-012110-121302

USDA ERS. (2022, February 7). *Interactive charts: Food Expenditures*. USDA ERS - Interactive Charts: Food Expenditures. Retrieved April 13, 2022, from https://www.ers.usda.gov/data-products/food-expenditure-series/interactive-charts-food-expenditures/#household

West, P. C., Gerber, J. S., Engstrom, P. M., Mueller, N. D., Brauman, K. A., Carlson, K. M., Cassidy, E. S., Johnston, M., Macdonald, G. K., Ray, D. K., & Siebert, S. (2014). Leverage points for improving global food security and the environment. *Food Security*, *345*(6194), 1–5.

White, B. (2012). Agriculture and the Generation Problem: Rural Youth, Employment and the Future of Farming. *IDS Bulletin*, *43*(6), 9–19. https://doi.org/10.1111/j.1759-5436.2012.00375.x

Wills, M. (2016, May 4). *Are We Entering a New Golden Age of Guano?* JSTOR Daily. Retrieved February 8, 2022, from https://daily.jstor.org/golden-age-of-guano/

CHAPTER TWO

ON-FARM PRECISION EXPERIMENTS (OFPE) FRAMEWORK: TAPPING LOCAL DATA TO OPTIMIZE FIELD SCALE DECISIONS

Abstract

Precision agriculture and open-source databases produce a plethora of field-specific crop information about the agroecosystem of crop production, but few mechanisms have been developed to turn that information into management recommendations. The On-Farm Precision Experiments (OFPE) framework is a methodology that aims to improve a crop managers' ability to make agronomic input decisions that account for variation within individual fields. This work evaluates the use of field-specific experiments that can tap the abundance of open-source data and the data emanating from precision agriculture technologies to gain knowledge of the spatial and temporal variability in agroeconomic performance at the field scale. Quantification of the variability in response to inputs (e.g., fertilizers and other soil amendments) then allows estimation of the probability of future management outcomes (crop yield, quality, and economic return). The challenge is to integrate OFPE into management with minimal disruption of common practices while drawing on historic knowledge about the field and economic constraints. OFPE is a decision support system that includes a six-step cyclical process that harnesses precision agriculture technology to apply experiments and gather field-specific data, incorporates modern data management and analytical approaches, and generates probabilistic management recommendations. The OFPE framework allows field managers to assess the

tradeoffs in agronomic input management between the maximization of crop production, quality and profits from production while considering environmental effects.

Introduction

While crop productivity has increased due to the Green Revolution, the rate of yield increases have slowed since 1985 and yet, with a projected ca. two-fold increase in crop demand from 2005 to 2050, agriculture will need to increase productivity yet again, albeit with a higher priority on sustainability (Foley et al., 2011; Tilman et al., 2011). The challenge of moving toward more sustainable agriculture includes identification of practices that will simultaneously increase farm profits, promote environmental stewardship, enhance the quality of life of farmers and rural communities, and increase agricultural production (NIFA, 2022). Crop production gains have primarily been the result of application of inputs like fertilizer. To achieve sustainability, agricultural practices must recognize the tradeoffs, in regard to agronomic inputs, that are required to increase production while maintaining the resource base on which agriculture relies(Antle & Capalbo, 2002; Antle & Valdivia, 2021). Transitioning industrial agriculture towards a sustainable future will not be achieved by a single leap from conventional practices to complete sustainability but will require incremental development of knowledge of the interacting factors that contribute to variability in crop response and the resulting tradeoffs between production, profitability, and environmental impact. The tools and technology of precision agriculture are suggested as having the potential of providing for sustainable management through quantifying the tradeoffs of agronomic inputs at the field scale where input management decisions are made (Duff et al., 2022; Kanter et al., 2018).

The majority of inefficient agricultural practices occur in a handful of nations, most of which practice industrial agriculture and have the affluence required to leverage modern tools and technology for sustainable production of food, fuel, fiber, and feed (West et al., 2014). Agriculture has steadily modernized throughout the centuries by improvements in equipment and technology, recently with precision agriculture: a management strategy that utilizes temporal and spatial data to inform agricultural production. Industrial agriculture in developed nations contributes the most to agricultural inefficiency and pollution, but has also been the primary adopter of precision agriculture tools and technology due to the economy of scale in industrial farm operations (Schimmelpfennig & Lowenberg-DeBoer, 2020). The adoption of precision agriculture based on poorly-formed agroecological models has led to the perception of the substitution of baseless technology for local farmer knowledge and is feared for its potential to lead agriculture towards "ecological dystopias" (Altieri & Nicholls, 2020; Daum, 2021). OFPE is provided as a mechanism to gain local knowledge, use locally parameterized agroecological models, and incorporate and augment the farmers' knowledge necessary for locally relevant decision making, rather than replacing traditional knowledge (Duff et al. 2022).

On-farm experimentation (OFE) brings farmer research to the fields on which decisions about agronomic inputs are made and approaches the complexities of agronomic management on a field specific basis (Lacoste et al., 2022). Crop metrics within a field, such as yield, vary spatially due to soil and climate variability and a range of and management practices even when meant to be held constant. In addition, crop response varies over time due to factors such as weather (Hegedus & Maxwell, 2022a), yet the response of crops to varying agronomic input rates also varies, indicating the potential for site-specific agronomic management to increase profitability and sustainability when informed by on-farm experimentation (Trevisan et al., 2021). On-farm experimentation shifts the lens of agronomic research to “operational research”, with an applied focus for informing management decisions by using observations from farms rather than from research centers (Cook et al., 2004; Maxwell & Luschei, 2005). The benefits of OFPE have been well studied and include increased productivity, profits, and adoption of sustainable practices (Kyveryga, 2019). Although designed to inform spatially varying agronomic input rates, even a single year of experimentation potentially provides information that increases the profitability over spatially uniform management (Bullock et al., 2020).

The objectives of this paper are to detail the OFPE methodology and outline the generation and evaluation of management options for sustainable agronomic inputs. This paper is a proof of concept from rain-fed winter wheat production in Montana, where OFPE has been applied to quantify within-field spatial and temporal crop response variability to top-dress nitrogen fertilizer application in crop-fallow or crop-cover crop rotations over the last six years on twenty-eight fields. This system has the unique aspect of including crop quality (grain protein concentration) as well as grain yield as economically significant response variables.

Methodology

Overview

The goal of the OFPE framework is to provide farmers a method for generating management recommendations that aids in decision making that increases sustainability as well as production and profit. Management recommendations for farmers using the OFPE framework are derived and evaluated through a six-step process (Fig. 2.1). First, on-farm experimentation is implemented to assess the relationship between the crop and the agronomic input of interest.

Precision agriculture equipment and technology are utilized for the application of OFE and collection of field-specific data (Duff et al., 2022; Hegedus & Maxwell, 2022b). Statistical models that characterize the ecological interactions among crop responses, the environment (weather and edaphic features), and experimentally varied agronomic inputs are trained with data aggregated from on-farm and internet-available open sources, such as remote sensing data from satellites. Simulations are used to predict the probability of outcomes under variable conditions to generate agronomic input recommendations while considering uncertainty in future weather and economic conditions, where years from the past are sampled to emulate potentially anomalous future years (Hegedus & Maxwell, 2022c). Based on predetermined goals and modeled crop responses across simulations, optimized input rates are identified on a site-specific basis. Finally, management outcomes, ranging from farmers' status quo rates to site-specific optimized rates, are evaluated and presented to farmers and crop managers in a probabilistic framework, leaving decisions about future management in the hands of the farmers. All aspects of the OFPE framework can be facilitated through the *OFPE* R package, which contains automated algorithms for implementing all steps of the OFPE framework (Hegedus, 2020).

The OFPE framework is demonstrated where it was developed – in dryland winter-wheat systems in Montana – but is easily adapted to other systems by substituting response or spatially varying explanatory variables. The OFPE framework is demonstrated with the objective to optimize top-dress nitrogen fertilizer rates based on maximizing farmer profits and nitrogen use efficiency. The OFPE framework has also been applied in certified organic fields to identify optimized cash crop and cover crop seeding rates based on maximizing wheat grain yields. Adoption of the OFPE framework by the Data Intensive Farm Management (DIFM) project's trials in eight states and multiple countries (Bullock et al., 2019) demonstrates the flexibility and adaptability of the approach, as the field-specific nature of the methodology relies only on data from a specific field and performs model selection to identify the form that best characterizes crop responses in a specific field (Hegedus & Maxwell, 2022b).

Five principles distinguish the OFPE framework from other approaches towards agronomic decision making: First, experiments are intended to inform management on the field where conducted, not other fields, under the assumption that field history can have a significant impact on response to inputs (Hegedus & Maxwell, 2022a). Second, all variables used to predict yield and protein are farmer- or open-sourced data that are available up to the time of application decision (Hegedus & Maxwell, 2022c). Third, predictive equations evolve, and parameterization is updated each year as new data comes in and response functions are built specifically for each field (Hegedus & Maxwell, 2022b). Fourth, the manager can simulate net return outcomes given different weather and price conditions that most closely match the current year when a decision is made, and/or explore a possible range of outcomes under different assumed conditions. And finally, after predicted conditions are identified, the manager can compare the different profit maximizing management approaches (site-specific variable rate application, uniform rate application, no application or the field manager selected uniform rate) and determine the probability (based on uncertainty in outcomes) that site-specific management will produce the higher return on investment than the other approaches.

1A. Preparation & Initial Steps

Deployment and development of the OFPE framework centers around a data management system that stores and facilitates organization of the spatiotemporal data collected by precision agriculture equipment. Prior to implementation of the OFPE framework, a database needs to be created to contain farm and field information, such as: boundaries, grain yield, grain protein, agronomic input, and remotely sensed data. Development of the OFPE framework in this work used a PostgreSQL spatial database housed on a cloud-based virtual machine to securely store farm and field information from farmer collaborators. A key component of the OFPE framework is gathering open-source data that is expected to have an ecological relationship with yield or grain protein concentration. In our case, vegetation index data, weather variables, and soil characteristics were gathered from freely available repositories and used as predictors in crop response functions and for simulating outcomes in different years. This database also contained all the open-source, on-farm, and as-applied data gathered from experimental fields. The data management system serves as the keystone of the "ecosystem" needed for digitally informed agriculture (Bullock et al., 2019).

1B. Experimentation

The first step in the OFPE framework is the generation of field-specific experiments. Experiments can be generated in any design required by the user in the OFPE framework, to be flexible to the needs of the research and restrictions due to equipment. To identify optimal input rates, experiments should attempt to apply each experimental rate representatively across the entire field to capture potential subfield variation of crop responses. Experimentally varying agronomic inputs, such as nitrogen fertilizer, underpins the OFPE framework because it generates the datasets required for modeling field-specific crop responses (Bullock et al., 2019). On-farm experimentation using precision agriculture technology can take many forms, ranging from random assignment of agronomic input rates stratified on previous data, such as yield, to more structured experimental designs. Best practices of precision agriculture based OFE include trial plots that exceed 125 feet to allow equipment to change rates between plots and widths equivalent to at least two equipment passes to generate statistically qualified designs (Rzewnicki et al., 1988). Designing experiments in the OFPE framework is available online as an open-source web application (http://trialdesign.difm-cig.org/home).

Prior research indicates that repeated precision agriculture based experimentation in dryland Montana systems for 6-8 years is crucial to fully parameterize empirical models of grain yield and grain protein concentration responses to experimentally varied nitrogen fertilizer rates (Lawrence et al., 2015b). However, the number of years that experimentation is required may vary by field, depending on the uncertainty introduced by climate change induced increases in the probability of extreme weather events and shifting temporal trends in precipitation and temperature (Whitlock et al., 2017). A backlog of experimental data from years with a diverse range of weather conditions increases the likelihood that a given model can accurately simulate crop responses in unknown conditions of future years (Hegedus & Maxwell, 2022b).

Initial experiments in a field with the OFPE framework require experiments that cover an adequate range of agronomic input rates from which a relationship between crop responses and the input can be identified. Beyond representation of rates, initial experiments should also spatially represent the entire field (Fig. 2.2). As experimentation is repeated on a given field, the range of rates represented each year, as well as the amount of the field covered in experimental rates, can be minimized to diminish the influence of experimentation on management while retaining statistical relevance. After experiments are designed for a given field, they are given to and reviewed by the farmer to incorporate their input into the design of their on-field experiments and input rates are often adjusted to bracket their intuitive rate.

2. Data Collection

The OFPE framework process was demonstrated by using data specific to a field to make recommendations for that specific field. The field experimental design map is given to the farmer for application typically using variable rate application (VRA) technology on their input machine that will read the map and apply the rates accordingly. The application data from the equipment used is gathered and imported into the data management system because practical limits on machinery result in differences between the prescription and actual application. After harvest, crop response data are gathered from monitors mounted on combines. Crop response data are typically yield measured from on-combine yield monitors but could also include metrics of crop quality like grain protein concentration.

With the aim of supporting a low-cost decision support system, the OFPE framework only utilizes data collected from normal farm operations and data available from open sources, excluding the cost of data not used for other purposes (e.g., soil samples, weed surveys, etc.). Thus, in addition to data collected on location, data from open-source data repositories are utilized to provide field specific information not gathered from farmer's equipment. These data include information about the crop or environment which can be useful prior to input application as they can indicate the crop condition.

Multiple open source data repositories exist, where environmental data can be downloaded from, including Google Earth Engine (Gorelick et al., 2017). Examples of supplementary data collected from Google Earth Engine include vegetation index, water index, topographic, weather, and soil characteristic data (Table 2.1).

An important aspect of the OFPE framework is the constraints placed on covariate data. All farmers face a point in time at which decisions about inputs must be made. With modern satellite data collected at a weekly or daily resolution, farmers now have access to more up-to-date data than ever at the time of an application decision. Thus, temporal data used in crop response models of the OFPE framework were collected up to the decision point to maximize the amount of information available at the time of decision making (Hegedus & Maxwell, 2022c).

3. Data Aggregation

Aggregation of data consists of georeferencing all the data from different monitors and sensors to common locations within fields. Yet, the variety of data introduces a variety of resolutions that need to be rectified as using data from disparate sources introduces uncertainties associated with the resolution measurements are gathered (Blöschl & Sivapalan, 1995; Fritsch et al., 2020). The resolution of collected covariate data can range from 10 m for some vegetation index data to 1 km for weather data such as precipitation and growing degree days. On the other hand, the temporal resolution of grain yield and protein measurements are typically 3 to 10 seconds, respectively, with spatial resolution across fields thus depending on the velocity and cutter bar width of the harvester. Development of the OFPE framework was conducted using a scale of 10 m for data aggregation, though users of the framework have the power to decide the scale appropriate for their system based on their data. The 10 m scale was selected to minimize the amount of variation in crop responses smoothed over when taking the median of multiple observations in one 10 m grid cell. Additionally, the 10m scale is the smallest resolution at which open-source data in the development process were gathered, meaning that no upscaling of open-source data was required. Many of the open-source datasets used were collected or calculated at resolutions greater than 10 m yet no attempts at downscaling were made. Thus, uncertainty is introduced when using coarse resolution data at the 10 m scale because information about fine scale variation is missing. Downscaling represents an area in which future research can benefit the OFPE project yet is beyond the scope of this dissertation.

Grids are created to overlay each field where experiments were performed, and all data collected on-farms and from remotely sensed information are aggregated to the centroid of each 10-m grid cell. The scale selected for aggregation determined the degree to which variation in measurements were averaged. For example, only one grain protein measurement and a couple yield measurements typically fell within each 10-m cell of a grid, so smoothing of variation of crop response measurements were minimized. As-applied or as-planted data are also aggregated to the centroid of grid cells via a spatial intersection of the grid points and the experimental data. The aggregation process results in datasets at a user defined scale, for each field, that contained the crop responses, experimental rates, and all remotely sensed covariates (Fig. 2.3).

4. Data Analysis

Crucial to the OFPE framework is development of ecological models that are used to characterize the relationships between the observed crop responses, experimentally varied agronomic inputs, and the remotely sensed environmental variables. Modeling the response of crops to agronomic inputs has long been a subject of agronomic research, which has not yielded a consensus on one approach that adequately captures the variability that occurs across space and time for a given crop response (Anselin et al., 2004; Paccioretti et al., 2021). Minimization of uncertainty related to the characterization of crop responses to agronomic inputs requires model selection performed on a trial by trial basis, as the form of a model appropriate for one field is not always consistent with neighboring fields, or even the same field in different years (Thöle, Richter, & Ehlert, 2013). Additionally, crop responses to varying agronomic input rates do not always present a clear pattern, meaning that assumptions of a specific form or shape increase uncertainty and lead to models with less ability to accurately predict future crop responses than models that do not assume data takes a specific shape (Hegedus & Maxwell, 2022b). A greater ability to predict crop responses translates into a higher confidence in management recommendations and evaluation of optimized input rates (Hegedus & Maxwell, 2022b; Hegedus & Maxwell, 2022c).

At the current stage of development, the OFPE project has considered six crop response model types for experimental nitrogen fertilizer, which are programmed into the *OFPE* R package. These include a non-linear model assuming a logistic form, a non-linear model assuming a beta function (X. Yin et al., 2003), a generalized additive model, random forest regression, a Bayesian multiple linear regression model, and a Bayesian non-linear model (Lawrence et al., 2015b). This set of models are used in a selection process to determine the model type that best predicted crop responses in each field. However, depending on the crop and system, different models can and likely should be incorporated into the OFPE framework.

5A. Optimization

Crop response models are necessary to find optimum agronomic input rates because optimums cannot be found directly via experimentation. At any given point in a field, only one experimental rate can be applied, meaning that the true “optimum” is unobservable. Optimums thus must be identified via models in which crop responses are predicted under a range of experimental rates. The definition of an optimum rate varies based on the user of the OFPE framework and the agronomic system. For example, if a barley farmer using nitrogen fertilizer has a contract with a brewer that requires a low protein content to minimize the haziness of their beer, the optimum fertilizer rate for a given location in a field would be the rate that minimizes grain protein concentration. On the other hand, a wheat farmer having a contract with a baker that requires high protein content for their dough would identify the optimum fertilizer rate for a given location in the field as the rate that maximizes grain protein concentration.

In most systems, however, the motivation for farmers is not quality, but maximization of production relative to costs, which leads to profit. In corn or soybean systems this would require identifying optimum rates that simply maximize net-return, or the price received for their yield minus expenses. In the dryland winter-wheat systems of Montana, farmers receive a premium or dockage based on grain protein that is added or subtracted from the base price received for wheat, making profit maximization more complex. When farmer profits are of interest, an optimum nitrogen fertilizer rate would be the rate at which farmer net-return is maximized and further increase in revenue does not cover the cost of additional fertilizer.

Yet the future of agriculture requires an emphasis on sustainability, and promotion of sustainability in a numerical optimization requires the costs of environmental degradation to be included in the calculation of net return. The OFPE framework creates the infrastructure for farmers to move beyond considerations of simply profits and consider sustainable environmental stewardship as well. While nitrogen is a major macronutrient for crops, contributing to yield and protein, Montana is subject to soil acidification and nitrate loss to both leaching and denitrification due to excess nitrogen fertilizer use (Jones, 2018; Sigler et al., 2020). Thus, sustainable agriculture requires defining an optimum nitrogen fertilizer rate based on the tradeoff between profit and environmental quality. One approach for assessing the environmental impacts of nitrogen fertilizer requires models that estimate nitrogen use efficiency on subfield basis to inform crop managers on the potential and the cost of nitrogen loss (Hegedus & Ewing, 2022). In this case, an optimum would be considered the nitrogen fertilizer rate at which the distance between net-return and the value of nitrogen lost from a system (contributing to acidification or leaching) is maximized (Fig. 2.4).

5B. Simulation

Simulation is an important aspect of decision support systems because it allows users assess the potential for variation in the efficacy of management strategies in a probabilistic format. Unpredictable weather and economic conditions induce uncertainty to any management recommendations, as the optimal agronomic input rate at a given point in a field may change depending on any variation in weather conditions (Hegedus & Maxwell, 2022c). The optimum recommendation under one weather condition may not be appropriate for another, and a field manager can only take a best guess at what the weather will be like in a future year. Additionally, many farmers do not have control or information on the economic conditions at a future harvest date when their crop is sold, further introducing uncertainty into management recommendations, as optimum rates assuming a farmer will receive one price for their crop may not be consistent with an optimum rate under a different price scenario.

The OFPE framework uses a bootstrap Monte Carlo simulation approach to propagate these uncertainties through to predicted net returns. This error propagation provides report management outcomes in terms of ensemble results that allow a probabilistic perspective on risks in decision making (Fig. 2.5). The OFPE framework can be adapted to assess a range of different management strategies to compare against site-specific management, such as applying none of the agronomic input (like fertilizer or herbicide), a uniform rate selected by the farmer, or a full-field uniform optimized rate. Again, due to the flexibility of the OFPE framework, the optimization scheme from which rates are selected can be varied.

To address the uncertainty of management recommendations due to weather, the OFPE framework requires users to select a year from the past that they expect would be representative of an upcoming year. A year from the past can also be randomly selected to emulate a future anomalous year that is more likely due to climate change, for example. After selecting a year, data for covariates from the selected field specific model are used to predict crop responses under the simulated conditions. Forecasted crop responses in the new conditions are made at every location in the field for a range of experimental rates. This leads to the second layer of the simulation, when net-return is assumed in the optimization scheme, where another year from the past is randomly selected and the economic parameters are used to calculate net-return. Based on net-return, optimized rates at each location in the field are identified, as well as the net-return from the other strategies. Random sampling of different years is repeated for a given number of iterations, where the optimums and management outcomes are tracked and evaluated after the simulation of different management scenarios has completed.

The flexibility of the OFPE framework allows users to define the degree to which uncertainty is incorporated in the simulation. For example, if a farmer knows the cost of their input and price received for the crop prior to deciding input management strategy, they could run the simulation with a fixed economic scenario rather than drawing randomly from years of the past in the Monte Carlo simulation. Additionally, farmers have control over what and how many weather conditions they execute in a simulation. For example, if a farmer is confident in the consistency of the weather in their fields, they could run the simulation using the same conditions as the prior year, but if they are more uncertain, they can use simulations to compare management outcomes from years with varying weather conditions. The simulation aspect of the OFPE framework also enables farmers and crop managers to assess how management would change under climate change, by selecting a set or a single anomalous year from the past that may have not occurred in sequence but would represent a new condition. The simulation step is thus a crucial juncture in the OFPE framework where farmer involvement and knowledge of their system is required to inform data driven management.

5C. Evaluation

As optimum agronomic inputs cannot be empirically evaluated because only one input rate can be applied at any given point in the field at any given time, the crop response models fit in the fourth step of the OFPE framework are crucial for evaluating the profitability and sustainability of optimal rates and rates of different management strategies. Site-specific optimum rates are found by using crop response models to make site-specific predictions under varying agronomic input rates, but the model is also required to predict what site-specific crop responses would have been under a farmer's status quo management, or any other selected management strategy. The OFPE framework evaluates different strategies in probabilistic terms, where at each iteration of the simulation, the mean net-return of each strategy are compared. Given a strategy of interest, such as a site-specific approach, the number of times that the strategy yielded a higher net-return, or other metric of interest, compared to the other strategies is recorded and divided by the number of iterations in the simulation. In this regard farmers are given the chance that a given management strategy outcompetes another management strategy. The farmer then has the choice to select from any of the management strategies, leaving the decisions about management on farms up to the farmer. While presented with a site-specific and full-field optimized strategy, the farmer has free will to elect to continue with business as usual if they like the odds of that strategy outperforming the optimized approaches.

6. Decision Making

The final steps of the OFPE framework occur after the farmer is confronted with probabilistic outcomes from the simulation. After the farmer selects their strategy, three routes can be taken. First, a farmer can adopt and apply their selected management strategy. Second, the farmer can begin the OFPE process again with another experiment. In this approach a farmer has two further sub-options, incorporating a full-field uniform rate as the rate surrounding experimental trial plots, or continuing to use their status quo input rate surrounding experimental trial plots. The third approach is that a farmer can elect to adopt and apply the selected strategy in combination with further experimentation. In this case experimental rates are inserted into the optimum rates background map. While the first option is available to farmers, continued experimentation through the second or third approaches is crucial to increasing the statistical power of the field specific crop response models and for updating management recommendations as more data are available. In Montana, where fields are large, we often have several precipitation levels with a field in a year. However, where fields are smaller there may not be variation within a field and several years of experimentation would be required to estimate the impact of weather variables.

Conclusion

The OFPE framework provides farmers with a methodology for harnessing the data and tools from precision agriculture to make management decisions. While the flexibility and adaptability of the framework means it can be used to optimize agronomic inputs based on any user defined criteria, it was developed for evaluating the tradeoffs between farmer profits and environmental quality. Utilized by the DIFM project across the world, the framework has been employed to generate site-specific agronomic input recommendations, demonstrating the plug and play capability of the framework to work in varying agroecosystems (Bullock et al., 2019).

Only around 10% of farmers with precision agriculture technology utilize variable rate applications, with barriers to adoption including the lack of decision support aids (McBratney et al., 2005; Pierpaoli et al., 2013). A farmer with precision agriculture technology benefits from decision support systems that facilitate the organization, storage, and translation of data to management recommendations (Anwar et al., 2013; Foley et al., 2011; A. McBratney et al., 2005). Decision support systems are central to making management recommendations for agronomic inputs by facilitating collection and analysis of crop response and remote sensing information from farms and open-source datasets. As a big-data industry, precision agriculture is overcoming prior barriers to adoption surrounding the management and analysis of data by attracting profound investment and interest in agronomic data analytics and software development (Pham & Stack, 2018; Sykuta, 2016). Open-source decisions support systems that use data farmers can obtain for free or on their farms will prevent farmers having to pay to make informed decisions.

The OFPE framework alone is not feasibly adopted by farmers due to the data infrastructure and ecosystem requirements but provides the logical underpinning of a decision support system that can be managed by academic or research institutions and provided to farmers as an open source or low-cost software (Sawers, 2021). Harnessing this data, combined with on-farm experimentation, also facilitates adaptive management, where field-specific agronomic input decisions are generated and updated from iterative analysis of experimental data gathered on the given field (Haas et al., 2009; J. P. Mueller et al., 2009; Wyeth, 2009). While increasing adoption of variable rate technology remains a challenge, developing, providing, and training farmers and crop consultants on a low-cost decision support system can remove barriers to adoption surrounding the price associated with managing and exploiting field-specific data.

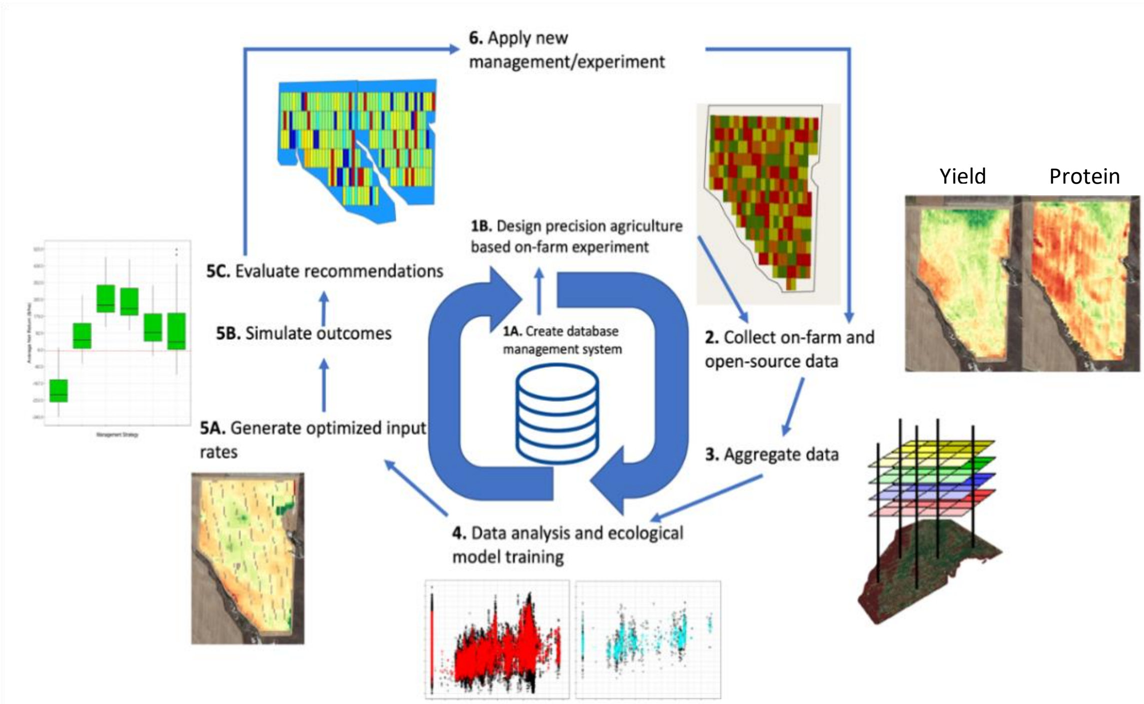


Figure 2.1. Overview of the OFPE methodology for generating optimized nitrogen fertilizer rates, beginning with 1A) Creation of a database management system, 1B) the creation of an initial field experiment, 2) collection of field-specific data from sensors mounted on farm equipment, grided internet available weather data and from open-source satellite imagery, 3) aggregation of data from the disparate datasets, 4) parameterization of ecological crop response models, 5A) generation of optimized (profit maximized and pollution minimized) agronomic input rates, 5B) simulation of management outcomes under varying weather and economic conditions, 5C) evaluation of recommendations, and 6) selection of future management based on a management strategy and/or continued experimentation. Each year begins with either step 1B or 6 which cyclically feed into step 2.

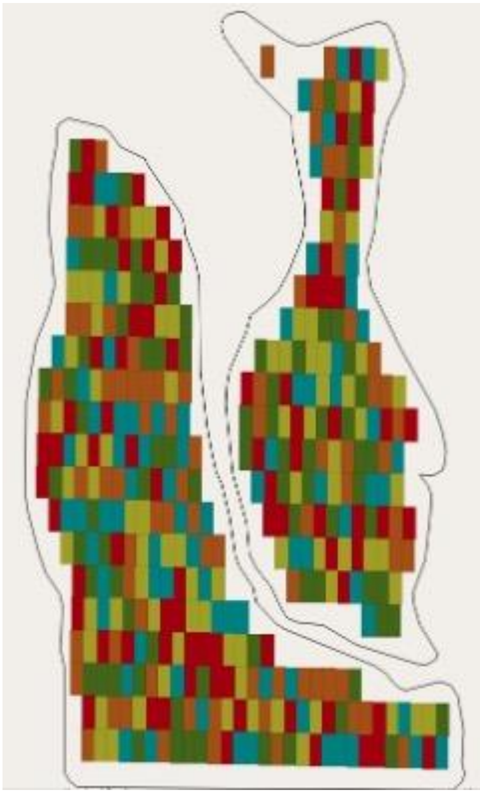


Figure. 2.2 Example of a typical experimental layout in the OFPE framework. Different colors represent different rates of the agronomic input.

Table 2.1. Table of covariate data types gathered from Google Earth Engine to enrich the crop yield and protein datasets gathered from on-farms. In some cases, multiple sources are used, however only one data source is used when aggregating to yield and protein.

Data Type	Data Sources	Resolution	Years Collected	Description
Normalized Difference Vegetation Index (NDVI)	Landsat 5/7/8	30m	L5: 1999-2011 L7: 2012-2013 L8: 2014 - present	Landsat is an ongoing USGS and NASA collaboration. Bands (NIR, red) L5/L7: B4 and B3 L8: B5 and B4
Normalized Difference Water Index (NDWI)	Landsat 5/7/8	30m	L5: 1999-2011 L7: 2012-2013 L8: 2014 - present	Landsat is an ongoing USGS and NASA collaboration. Bands (NIR, red) L5/L7: B2 and B4 L8: B2 and B5
Elevation	USGS NED	~10m (1/3 arc second), ~23m (3/4 arc second)	1999-present	USGS National Elevation Dataset. Measured in meters.
Aspect	USGS NED	~10m (1/3 arc second), 30m	1999-present	Direction the surface faces, function of neighboring elevations, in radians. Also calculated for each E/W and N/S direction as cosine and sine.
Slope	USGS NED	~10m (1/3 arc second), 30m	1999-present	Rate of change of height from neighboring cells, in degrees. Measured in degrees.
Topographic Position Index (TPI)	USGS NED	~10m (1/3 arc second), 30m	1999-present	Measure of divots and low spots as a function of neighboring elevation.
Precipitation	DaymetV3	1km	1999-present	Estimates from the NASA Oak Ridge National Laboratory (ORNL). Measured in mm.
Growing Degree Days (GDD)	DaymetV3	1km	1999-present	Estimates from the NASA Oak Ridge National Laboratory (ORNL).
OpenLandMap	bulkdensity	250m	1999-present	Soil bulk density (fine earth) 10 x kg / m3 averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
OpenLandMap	claycontent	250m	1999-present	Clay content in % (kg / kg) averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
OpenLandMap	sandcontent	250m	1999-present	Sand content in % (kg / kg) averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
OpenLandMap	pH (phw)	250m	1999-present	Soil pH in H_2O averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
OpenLandMap	watercontent	250m	1999-present	Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted and averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
OpenLandMap	carboncontent	250m	1999-present	Soil organic carbon content in x 5 g / kg averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).

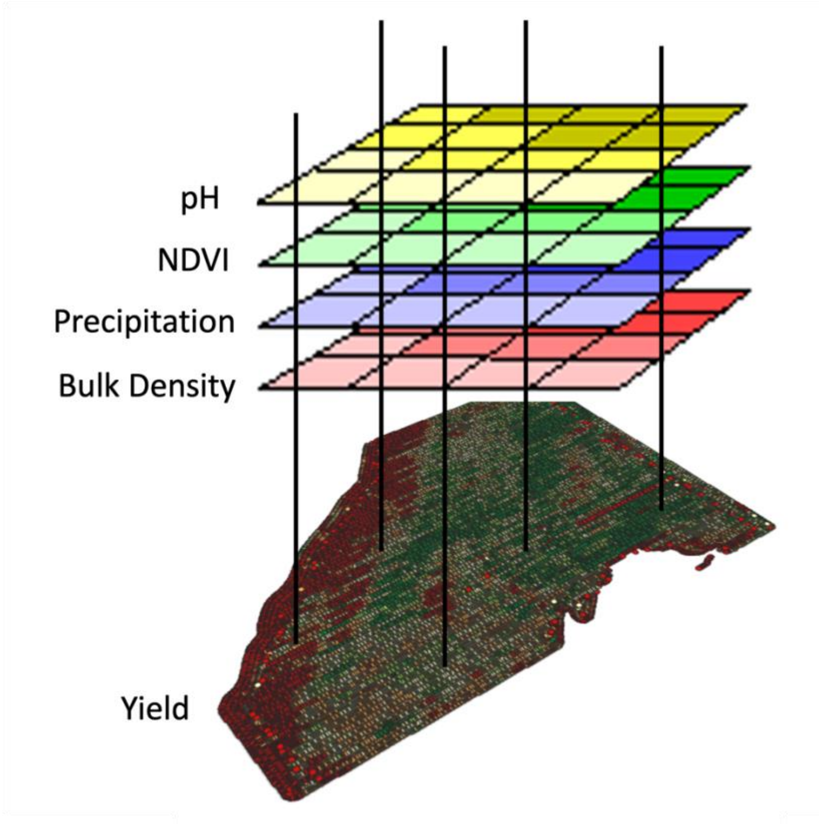


Figure 2.3. Conceptual diagram of the aggregation process. In this example, measurement values for a sample set of cells of remotely sensed raster data (pH, NDVI, precipitation, and soil bulk density) are georeferenced to grain yield points via spatial intersection.

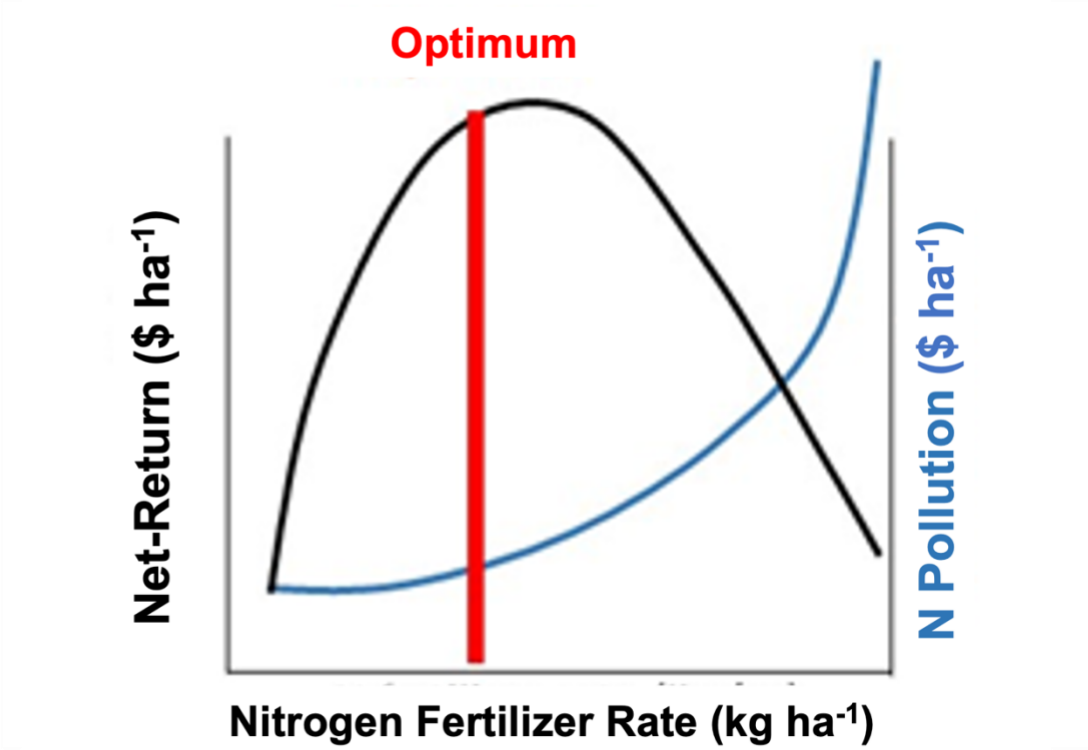


Figure 2.4. Conceptual diagram of the optimization process for finding the nitrogen fertilizer rate at each point that maximizes net-return and minimizes the potential value of nitrogen pollution.

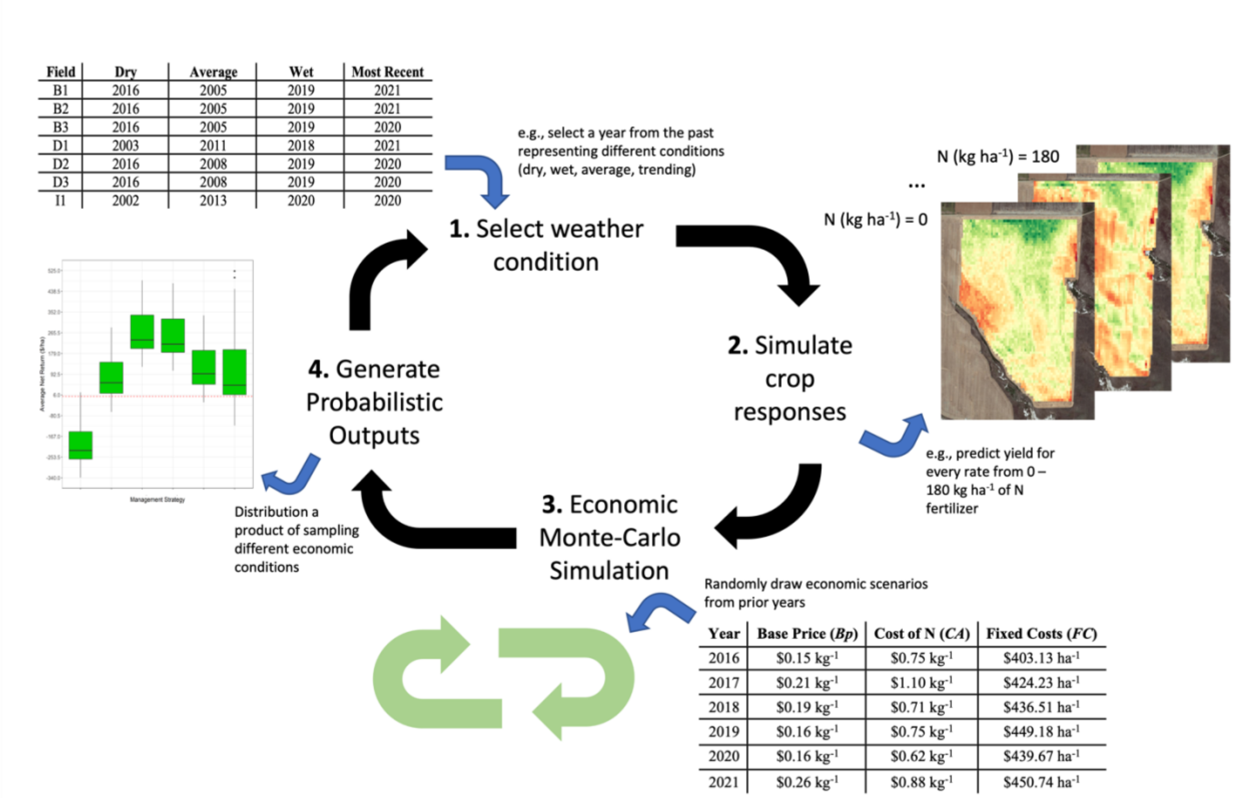


Figure 2.5. Overview of the OFPE framework simulation and optimization process. Simulations begin with the selection of weather conditions to forecast management outcomes in, typically as a selection from a year in the past. Then the data from that year are used to predict crop responses at every location in the field for a range of agronomic input rates, for example nitrogen fertilizer. The third step is a Monte Carlo simulation where economic conditions from previous years are used to account for uncertainty in future economic conditions. Finally, the repeated sampling of different economic scenarios enables the evaluation of management in a probabilistic format and the identification of optimum input rates based on economic uncertainty.

REFERENCES CITED

Altieri, M. A., & Nicholls, C. I. (2020). Agroecology and the reconstruction of a post-COVID-19 agriculture. Journal of Peasant Studies, 881–898. https://doi.org/10.1080/03066150.2020.1782891

Anselin, L., Bongiovanni, R., & Lowenberg-DeBoer, J. (2004). A Spatial Econometric Approach To The Economics of Site-Specific Nitrogen Management. Amer. J. Agr. Econ., 86(August), 675–687.

Antle, J. M., & Capalbo, S. M. (2002). Agriculture as a Managed Ecosystem: Policy Implications. Journal of Agricultural and Resource Economics, 27(1), 1–15.

Antle, J. M., & Valdivia, R. O. (2021). Trade-off analysis of agri-food systems for sustainable research and development. Q Open, 1(1), 1–34. https://doi.org/10.1093/qopen/qoaa005

Anwar, M. R., Liu, D. L., Macadam, I., & Kelly, G. (2013). Adapting agriculture to climate change : a review. 225–245. https://doi.org/10.1007/s00704-012-0780-1

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. Hydrological Processes, 9(3–4), 251–290. https://doi.org/10.1002/hyp.3360090305

Bullock, D. S., Boerngen, M., Tao, H., Maxwell, B., Luck, J. D., Shiratsuchi, L., … Martin, N. F. (2019). The data-intensive farm management project: Changing agronomic research through on-farm precision experimentation. Agronomy Journal, 111(6), 2736–2746. https://doi.org/10.2134/agronj2019.03.0165

Bullock, D. S., Mieno, T., & Hwang, J. (2020). The value of conducting on-farm field trials using precision agriculture technology: a theory and simulations. Precision Agriculture, 21(5), 1027–1044. https://doi.org/10.1007/s11119-019-09706-1

Cook, S., Cock, J., Oberthür, T., & Fisher, M. (2004). On-Farm Experimentation. Better Crops, 97, 17–20.

Daum, T. (2021). Farm robots: ecological utopia or dystopia? Trends in Ecology and Evolution.

Duff, H., Hegedus, P. B., Loewen, S., Bass, T., & Maxwell, B. D. (2022). Precision Agroecology. Sustainability, 14(106), 1–18. https://doi.org/https://doi.org/10.3390/su14010106

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., … Zaks, D. P. M. (2011). Solutions for a cultivated planet. Nature, 478(7369), 337–342. https://doi.org/10.1038/nature10452

Fritsch, M., Lischke, H., & Meyer, K. M. (2020). Scaling methods in ecological modelling. Methods in Ecology and Evolution, 11(11), 1368–1378. https://doi.org/10.1111/2041-210X.13466

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Haas, C. A., Frimpong, E. A., & Karpanty, S. M. (2009). Ecosystems and Ecosystem-Based Management. In K. M. Moore (Ed.), The Sciences and Art of Adaptive Management: Innovating for Sustainable Agriculture and Natural Resources Management (pp. 106–142). Soil and Water Conservation Society.

Hegedus, P.B. (2020). OFPE: On-Farm Precision Experiments (OFPE) Data Management and Analysis Tools. R package version 1.8.11. https://github.com/paulhegedus/OFPE.git

Hegedus, P.B & Ewing, S.E. (2022). Evaluation of nitrogen use efficiency in dryland winter-wheat agroecosystems for predictive modeling. Manuscript in preparation.

Hegedus, P.B & Maxwell, B.D. (2022a). Rationale for field specific on-farm precision experimentation. Agriculture, Ecosystems & Environment. Manuscript submitted for publication, in review.

Hegedus, P.B. & Maxwell, B.D. (2022b). Assessing performance of empirical models for forecasting crop responses to variable fertilizer rates using on-farm precision experimentation. Precision Agriculture. Manuscript submitted for publication, in review.

Hegedus, P.B. & Maxwell, B.D. (2022c). Best use of modern data for field-specific decision support. Precision Agriculture. Manuscript submitted for publication, in review.

Jones, C. (2018). *Soil Acidification: A growing concern for Montana farmers*. United States Department of Agriculture, Natural Resource Conservation Service.

Kanter, D. R., Musumba, M., Wood, S. L. R., Palm, C., Antle, J., Balvanera, P., … Andelman, S. (2018). Evaluating agricultural trade-offs in the age of sustainable development. Agricultural Systems, 163, 73–88. https://doi.org/10.1016/j.agsy.2016.09.010

Kyveryga, P. M. (2019). On-farm research: Experimental approaches, analytical frameworks, case studies, and impact. Agronomy Journal, 111(6), 2633–2635. https://doi.org/10.2134/agronj2019.11.0001

Lacoste, M., Cook, S., Mcnee, M., Gale, D., Ingram, J., Bellon-maurel, V., … Hall, A. (2022). On-Farm Experimentation to transform global agriculture. Nature Food. https://doi.org/https://doi.org/10.1038/s43016-021-00424-4 On-Farm

Lawrence, P. G., Rew, L. J., & Maxwell, B. D. (2015). A probabilistic Bayesian framework for progressively updating site-specific recommendations. Precision Agriculture, 16(3), 275–296. https://doi.org/10.1007/s11119-014-9375-4

Maxwell, B. D., & Luschei, E. C. (2005). Justification for site-specific weed management based on ecology and economics. Weed Science, 53(2), 221–227. https://doi.org/10.1614/ws-04-071r2

McBratney, A., Whelan, B., Ancev, T., & Bouma, J. (2005). Future directions of precision agriculture. Precision Agriculture, 6(1), 7–23. https://doi.org/10.1007/s11119-005-0681-8

Mueller, J. P., Finney, D., & Hepperly, P. (2009). The Field System. In K. M. Moore (Ed.), The Sciences and Art of Adaptive Management: Innovating for Sustainable Agriculture and Natural Resources Management (pp. 25–59). Soil and Water Conservation Society.

National Institute of Food and Agriculture (NIFA). (2022). Sustainable Agriculture. Sustainable Agriculture | National Institute of Food and Agriculture. Retrieved February 8, 2022, from https://nifa.usda.gov/topic/sustainable-agriculture

Paccioretti, P., Bruno, C., Gianinni Kurina, F., Córdoba, M., Bullock, D. S., & Balzarini, M. (2021). Statistical models of yield in on-farm precision experimentation. Agronomy Journal, 113(6), 4916–4929. https://doi.org/10.1002/agj2.20833

Pham, X., & Stack, M. (2018). How data analytics is transforming agriculture. Business Horizons, 61(1), 125–133. https://doi.org/10.1016/j.bushor.2017.09.011

Pierpaoli, E., Carli, G., Pignatti, E., & Canavari, M. (2013). Drivers of Precision Agriculture Technologies Adoption: A Literature Review. Procedia Technology, 8(Haicta), 61–69. https://doi.org/10.1016/j.protcy.2013.11.010

Rzewnicki, P. E., Thompson, R., Lesoing, G. W., Elmore, R. W., Francis, C. A., Parkhurst, A. M., & Moomaw, R. S. (1988). On-farm experiment designs and implications for locating research sites. American Journal of Alternative Agriculture, 3(4), 168–173. https://doi.org/10.1017/S0889189300002459

Sawers, P. (2021, May 5). Linux Foundation Launches Open Source Agriculture Infrastructure Project. VentureBeat. Retrieved March 27, 2022, from https://venturebeat.com/2021/05/05/linux-foundation-launches-open-source-agriculture-infrastructure-project/

Schimmelpfennig, D., & Lowenberg-DeBoer, J. (2020). Farm Types and Precision Agriculture Adoption: Crops, Regions, Soil Variability, and Farm Size. SSRN Electronic Journal, 1–38. https://doi.org/10.2139/ssrn.3689311

Sigler, W. A., Ewing, S. A., Jones, C. A., Payn, R. A., Miller, P., & Maneta, M. (2020). Water and nitrate loss from dryland agricultural soils is controlled by management, soils, and weather. Agriculture, Ecosystems and Environment, 304(July), 107158. https://doi.org/10.1016/j.agee.2020.107158

Sykuta, M. E. (2016). Big Data in Agriculture: Property Rights, Privacy and Competition in Ag Data Services. International Food and Agribusiness Management Review, 19, 57–74.

Thöle, H., Richter, C., & Ehlert, D. (2013). Strategy of statistical model selection for precision farming on-farm experiments. Precision Agriculture, 14(4), 434–449. https://doi.org/10.1007/s11119-013-9306-9

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences of the United States of America, 108(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108

Trevisan, R. G., Bullock, D. S., & Martin, N. F. (2021). Spatial variability of crop responses to agronomic inputs in on-farm precision experimentation. Precision Agriculture, 22(2), 342–363. https://doi.org/10.1007/s11119-020-09720-8

West, P. C., Gerber, J. S., Engstrom, P. M., Mueller, N. D., Brauman, K. A., Carlson, K. M., … Siebert, S. (2014). Leverage points for improving global food security and the environment. Food Security, 345(6194), 1–5.

Whitlock, C., Cross, W., Maxwell, B., Silverman, N., & Wade, A. (2017). 2017 Montana Climate Assessment. Montana Institute on Ecosystems, 269. Retrieved from http://montanaclimate.org/chapter/title-page%0Ahttp://ieeexplore.ieee.org/document/6340318/

Wyeth, P. (2009). Sustainable Agriculture and Natural Resource Manageemnt in Farm Enterprise Systems. In K. M. Moore (Ed.), The Sciences and Art of Adaptive Management: Innovating for Sustainable Agriculture and Natural Resources Management (pp. 60–77). Soil and Water Conservation Society.

Yin, X., Goudriaan, J., Lantinga, E. A., Vos, J., & Spiertz, H. J. (2003). A flexible sigmoid function of determinate growth. Annals of Botany, 91(3), 361–371. https://doi.org/10.1093/aob/mcg029

CHAPTER THREE

RATIONALE FOR FIELD-SPECIFIC ON-FARM PRECISION EXPERIMENTATION

Contribution of Authors and Co-Authors

Manuscript in Chapter Three

Author: Paul B. Hegedus

Contributions: Conceived the study design and methodology, compiled data, executed statistical analysis, and wrote initial manuscript.

Co-Author: Bruce D. Maxwell

Contributions: Assisted with study design, analysis, and manuscript preparation at all stages.

Manuscript Information

Paul B. Hegedus, Bruce D. Maxwell

Agriculture, Ecosystems, and Environment

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
__x__ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Published by Elsevier

Abstract

Uncertainties in farming necessitate detailed knowledge of the production efficiencies to maintain sustainability. To accomplish ecologically based agriculture, with the goal of intensification by maximizing production and profit as well as minimizing environmental impact, we hypothesized that a site-specific knowledge base can be efficiently achieved through modern precision agriculture (PA) technologies at the field scale. The two goals of this study were to quantify the spatiotemporal variation of crop responses and the variables driving crop production, crop quality, and field-scale farmer net-return. We conducted on-farm experiments (OFE) on several fields for three years where we varied nitrogen fertilizer rate as a management input, to induce changes in crop response. Using a Monte Carlo approach, we assessed the probability that crop responses varied across fields and between years. To determine the drivers of crop production, quality, and net-return, we performed sensitivity analyses to assess the impact of variation in the environment with the most influence on crop responses and farmer profits. Our analysis provided evidence that the degree of the response of winter wheat yield and protein content to variable nitrogen fertilizer rates are not homogenous across time and space. Elevation as a covariate to nitrogen fertilizer rate was the primary influence on predicted yields and protein across most fields, yet not among all fields and across years in fields. The drivers of net-return varied among fields and across years primarily between yield and protein. However, in some cases the most influential factor was the base price received, controlled by the grain elevators that growers sell to, indicating that in some fields and years, farmer's net-returns are dictated by variables outside of a farmer's control or ability to manage. These results provide basic evidence justifying the use of OFE for farm management and suggest that management

needs to be specific to each field and point in time, with recommendations being made specifically for a field based on information gathered from that field. On-farm experimentation will enable farmers to identify these drivers and understand how their inputs influence yield and protein within fields. Using information provided by OFE with decision support systems can enable farmers to make informed management decisions that maximize their profits and increase the efficiency of chemical inputs, such as nitrogen fertilizer.

Introduction

In developed nations, the goal of agriculture has become agricultural intensification, synonymous with two objectives; maximization of crop production and the maximization of profits to accomplish economic sustainability (Elliot & Cole, 1989; Pretty, 2008; Talukder et al., 2020). Although harmful impacts of agriculture on neighboring ecosystems and human communities have been documented, little change in practices have been developed (Foley et al., 2011; Gliessman & Engles, 2014; Pretty, 2008; Tilman et al., 2011; Weiner, 2017). In addition, the uncertainties in agricultural production due to climate change and economics all necessitate more detailed knowledge of the efficiency of production within each farm to maintain sustainability (Cambouris et al., 2014; McBratney et al., 2005). We hypothesize, based on results from other site-specific management studies, that a site-specific knowledge base will facilitate ecological intensification, defined as increased production through management based on ecological interactions, by maximizing production and profit as well as minimizing environmental impact (Bronson et al., 2006; Cambouris et al., 2014; Farid et al., 2016; Flowers et al., 2004; Huggins, 2010; Khosla et al., 2008; Koch et al., 2004; Link et al., 2008; Moshia et al., 2014). In addition, the intensification goal can be efficiently achieved through modern technologies at the field scale and collection of field specific data to inform management decisions (Cao et al., 2018). We begin by assessing the underlying assumption that crop field responses (e.g., grain yield or grain protein concentration) are site, time, and history of land use specific. Drawing on modern technologies to incorporate site-specific experimentation into adaptive management principles (Hecht & Crowley, 2020; Jackson et al., 2010; Lin, 2011) would overcome major drivers of variation in agroecological processes (e.g. crop responses to inputs) and increase resiliency of these systems. Our assessment started by determining if variation in crop responses were significant among fields and among years in the same field when crop species was held constant, and when nitrogen fertilizer rates were experimentally varied on each field.

We tested the rationale for empirically driven field-scale adaptive management based on within-field information. The two goals of this study were to quantify the spatiotemporal variation of crop responses and the variables driving crop production, quality, and farmer net-return. Variation in time and space of a model system of dryland winter-wheat production (yield) and quality (grain protein concentration) were assessed. Harvester-mounted yield and protein monitors provided high-resolution data from a suite of fields where strategically designed experimental nitrogen fertilizer rates were applied. These experiments were implemented to determine and compare the variation in production and quality 1) within a field for a given year, 2) across years, and 3) between fields of the same farm and geographically distinct other farms in a given year. Additionally, the drivers of crop production, quality, and net-return of the farmers within and across years were evaluated by performing a sensitivity analysis on all covariates of response functions. Determining the generalizability of spatial and temporal variation in crop response to a primary management factor (in this case nitrogen fertilizer) allowed us to test the hypothesis that every field requires field-specific experimentation over several years to identify optimized management for that field. We define optimization as maximized crop yield, quality, farmer net returns or a combination of these responses.

We hypothesize that crop yield and protein response relationships with nitrogen fertilizer rates will spatially vary within and between fields each year. In addition, we hypothesize that yield and protein responses to nitrogen fertilizer rates will vary across years in the same field, between fields on the same farm, and between different farms located in different regions. We further hypothesize that weather related variables, such as precipitation and temperature, will account for most of the variation in relationships between crop response and nitrogen fertilizer rates between years and farms in different regions. Additionally, we hypothesized that weather related variables will be constant drivers of the variation in crop response between fields on the same farm. With this assessment, we provide empirical evidence for the value of on-field experiments which take advantage of modern farm equipment technologies to make experimentation logistically and economically efficient and acceptable to farmers.

Methods

Study Sites and Data

Empirically driven adaptive management requires iterative experimentation and data collection from fields as well as testing of our hypotheses on the spatial and temporal extent of crop response variability. Our research was conducted using data from farmers with a 10-year participation agreement to collaborate and learn with the On-Field Precision Experiments (OFPE) group at Montana State University. Farmers were chosen based on their experience with

PA technologies (yield monitor and protein monitoring data), and ability to perform variable fertilizer/seed rate application (VRA) in dryland winter wheat production. Three conventional farms, with varying numbers of fields per farm were selected for this study based on their levels of data availability and quality (Fig. 3.1, Table 3.1).

Each field selected has been subjected to OFPE variable rate N fertilizer experiments in at least two years and has yield data available for at least one year prior to experimentation. Fertilizer rates were experimentally applied by randomly placing treatments of various rates across each field with replication, stratified based on previous year crop yield (hypothesized best predictor of future yield). Experiments were thus a two-factor factorial assessment by application rate and previous yield (Fig. 3.2).

The response variables of interest are crop productivity (grain yield in kg ha^{-1}) and quality (% protein of grain), both of which are gathered from monitors mounted on farmers' harvesters. Data from yield monitors are gathered on average every three seconds and have become standard equipment on most modern combines. All yield monitors are calibrated every spring by the farmers, according to their respective manufacturing instructions. Grain protein concentration (%) was measured with a CropScan 3000H near-infrared monitor (Clancy, 2019). These data are subjected to cleaning before analysis, including adjusting for the lag between measurements and the location the measurement characterizes. Additionally, observations within 30m of the field boundary were removed, and outliers were removed where the yield/protein values or distance from the next point in time were above or below 4 standard deviations from the mean value or distance (Blackmore & Moore, 1999; Sudduth et al., 2012; Sudduth & Drummond, 2007).

For both yield and protein datasets, as-applied N data from farmers' application equipment was georeferenced to the observed yield and protein data. Beyond data collected from the machines on the field, open-source data that characterizes crop productivity, weather, and soil characteristics were gathered (Table 3.2). These data were obtained or derived from Google Earth Engine (Gorelick et al., 2017) and aggregated to the locations of the yield and protein observations. Aggregating data to the location of the yield and protein observations prevented any loss of information about crop responses due to averaging and upscaling. However, the open-source data gathered and hypothesized to characterize ecological relationships with the responses (Table 3.2) were gathered at a coarser resolution, inducing uncertainty because of the lack of fine scale variation in covariate measurements at the resolution yield and protein observations were gathered at.

Assessment of Spatiotemporal Differences in Crop Responses

All analysis was conducted in R using version 4.1.0 (R Core Team, 2021). The first objective of the study was to assess the spatiotemporal differences in crop responses. We assessed the sub-field scale spatiotemporal variation in crop responses for each field each year by comparing the crop response distribution of a given field and year 1) to the distribution of the crop response for the given field from all other years, and 2) to the distribution of the crop response for all other fields in the given year. Comparing the probability that crop responses were greater for a given field and year than crop responses from the same field in all other years indicates variation across time, while variation across space was assessed by comparing the probability that crop responses were greater for a given field than all other fields cropped in the same year.

A probabilistic approach was used to compare yields and protein between individual fields and years for two reasons. First, compared to pooled distributions, the distributions of crop responses of fields and years were not consistently normally distributed, and second, we have data on the actual observed distributions for each field and year and can avoid assumptions of normality and central tendency in our comparisons. We assessed the probability that crop responses were greater for a given field and year compared to crop responses in 1) each of the other years for the given field, and 2) each of the other OFPE fields for the same year. For comparisons across years for a given field, we used spatially explicit sampling across the field to be inclusive of in-field response variation. We compared the crop response values from a target field and year of interest to the nearest measurement in the comparison year for the same field, with a maximum sampling distance of 10m to assess the probability of difference between crop responses. For a given field, the target year was the most recent year for which data was available, while comparison years corresponded to all previous years' data. If the crop response for a given measurement within the target field and year was greater in the year of interest compared to the closest measurement in the comparison year, a value of 1 was assigned to the test in the year of interest. All values across the field for the year of interest were summed and divided by the total number of measurement locations in the year of interest to derive the probability that the crop responses were higher in the year of interest compared to the comparison year for a given field.

$$\Pr(A > B) = \frac{\sum_{i=1}^{n_A} \begin{cases} 1 & A_i > B_i \\ 0 & A_i < B_i \end{cases}}{n_A}$$

(Equation 3.1)

Where *Pr(A > B)* was the probability that the crop response (yield or protein) in a given target field and year (A) was greater than the crop response (yield or protein) in the same field and comparison year (B) is the response. Every observation in A was compared to the nearest neighbor within 10m from B, and if the crop response in A was greater than B, the observation was given a value of 1. The probability was derived as the number of observations where the crop response in A was greater than in B divided by the total number of observations in A (n_a).

We used Monte Carlo sampling to compare crop responses between different fields for a given year. For each target field for a given year, we took a random set of n samples with replacement from each crop response distribution and compared it to a random set of n samples with replacement taken from the distribution from the comparison field. If the crop response value from the field and year of interest were greater than a random draw from the distribution of the comparison field then the test was scored a 1, else scored 0. The scores from tests were added and divided by n to obtain the probability of difference. The number of tests (n) were equal to the number of crop response observations multiplied by 100 if the crop response was yield and 1000 if the crop response was protein. The calculation of the probability followed similar logic as in Equation 1, except that tests (n) with replacement were randomly drawn from the target field and year (A) and the comparison field for the given year (B) and compared without being spatially explicit because A and B correspond to different fields cropped in the same year (Equation 3.2);

$$\Pr(A > B) = \frac{\sum_{i=1}^{n} \begin{cases} 1 & A_i > B_i \\ 0 & A_i < B_i \end{cases}}{n}$$

(Equation 3.2)

Where *Pr(A > B)* was the probability that the crop response (yield or protein) in a given target field and year (A) was greater than the crop response (yield or protein) in a comparison field from the same year (B). Random samples with replacement were drawn from the crop response distribution from A, and compared to a randomly drawn crop response from the distribution of B. If the crop response in A was greater than B the test was given a value of 1. The probability is derived as the number of draws where the crop response in A was greater than in B divided by the total number of random samples taken with replacement from the data (n).

Probability of difference provided a continuous scale to judge field and yearly differences in crop response. Thus, the closer the probability was to 1.0 (target field was always greater than compared field or time) or 0 (target field was always less than compared field or time) indicated high confidence in difference between fields or years, versus probabilities near 0.5 indicated that there was no discernable difference between crop responses between the target and comparison field.

Sensitivity Analysis

For every field, a generalized additive model (GAM) using thin plate shrinkage splines was fit individually to observed yield and protein data from all years (Wood et al., 2016; Wood, 2003). We pooled field data across years to fit one model for each field for each crop response and used these models to perform the sensitivity analysis on each crop response in each year for the field. Bidirectional AIC based model selection was performed from the initial full model for both yield and protein (Equation 3.3). Covariates included in the initial model were limited to those that did not show any correlation with other predictors and had no singularities across fields and years. Diagnostics were evaluated for each model before settling on a final form and all predictors were centered, defined as subtracting the value of the variable from the mean value of the variable, prior to model fitting to limit the influence of scale differences. As crop responses differed across fields, not all fit GAMs for yield and protein for a given field had the same predictors. All final models for each field were selected from the following initial full model;

$$
\begin{aligned}
R \sim f_1(x,y) \; &+ f_2(N) + f_3(aspectCos, aspectSin) + f_4(slope) + f_5(elev) + f_6(tpi) \\
&+ f_7(prec_{cy}) + f_8(prec_{py}) + f_9(gdd_{cy}) + f_{10}(gdd_{py}) + f_{11}(ndvi_{cy}) \\
&+ f_{12}(ndvi_{py}) + f_{13}(ndvi_{2py}) + f_{14}(bulkdensity) + f_{15}(claycontent) \\
&+ f_{16}(sandcontent) + f_{17}(phw) + f_{18}(watercontent) + \epsilon
\end{aligned}
$$

(Equation 3.3)

Where *R* was yield (kg/ha) or grain protein content (%) observations, f_1 – f_{18} were smoothing functions fit with maximum likelihood on the tensor product of the x and y coordinate (*x*, *y*) and the rest of the parameters. The main explanatory variable was as-applied nitrogen (*N*), along with topographic variables including the tensor product of aspect in the E/W direction (*aspectCos*) and aspect in the N/S direction (*aspectSin*). Additional topographic parameters were slope in degrees (*slope*), elevation in meters (*elev*), and topographic position index (*tpi*). Precipitation accumulated from November 1st of the year prior to the observation year to March 31st of the observation year ($prec_{cy}$) was also used, along with precipitation accumulated from November 1st of two years prior to the observation year to October 31st of the year prior to the observation year ($prec_{py}$). Additional weather variables included growing degree days accumulated from January 1st to March 31st of the observation year (gdd_{cy}), and growing degree days accumulated from January 1st to December 31st of the year prior to the observation year (gdd_{py}). To provide information on crop productivity of the current year up to the decision point, normalized difference vegetation index (NDVI) data aggregated from January 1st to March 31st of the observation year ($ndvi_{cy}$) was used. To provide information on past crop productivity, NDVI aggregated from January 1st to December 31st of the year prior to the observation year ($ndvi_{py}$)

and NDVI aggregated from January 1st to December 31st two years prior to the observation year (*ndvi*$_{2py}$) were used. Soil characteristics including bulk density in 10 x kg/m^3 (*bulkdensity*), percent clay averaged over 0 – 200cm (*claycontent*), percent sand averaged over 0 – 200cm (*sandcontent*), the pH (x10) of soil water averaged over 0 – 200cm (*phw*), and the volumetric water content averaged across field capacity and wilting point and over 0 – 200cm (*watercontent*) were used. Finally, random error (ϵ) was incorporated.

Generalized additive models were chosen due to their flexibility and ability to fit highly scattered observed crop response data. Thin plate shrinkage splines were used for all variables to allow the estimated degrees of freedom (EDF's) of parameters to shrink to zero, combining the process of model fitting and selection. The parameter exceptions to thin plate shrinkage splines were the use of a gaussian process for the interaction of latitude and longitude to account for spatial autocorrelation between observed points within the field.

To determine the site-to-site and annual differences in covariates with high relative impact on crop responses and net-return, sensitivity/elasticity analyses were performed on each of the covariates in GAM's developed on a field-specific basis. Sensitivity analyses were performed for each fit predictor variable by systematically increasing and decreasing the fit parameter value for each parameter by 10% while holding all other variables with default fit values. The change in estimated response for each point was recorded for the comparison of each change to the observed values to determine which variable contributed to the largest change in observed crop response. This was repeated for every field and year combination. The sensitivity of responses for each field was evaluated by calculating the elasticity compared to the default observed conditions. Elasticity is defined as the absolute value of the difference between the average crop response in the default condition to the average crop response in the altered condition (with a covariate increased or decreased by 10%). The most elastic variables were grouped for each field across years and between fields by farms to assess spatiotemporal trends in drivers of crop productivity and quality.

To perform a sensitivity/elasticity analysis on net-return, protein and yield were combined by using kriging to spatially interpolate protein observations to coincide with the locations of yield. Correlation coefficients were created for each field and year to quantify the direction and magnitude of the relationship between yield and grain protein. Net-return was calculated as a function of yield, protein, the cost of nitrogen, and other fixed costs associated with producing the crop (Equation 3.4) for each point in the field.

$$P = Bp + (B0pd + B1pd * protein + B2pd * protein^2)$$
$$NR = yield * P - CA * AA - FC - ssAC$$

(Equation 3.4)

Where *P* was the final price received ($/bushel), *Bp* was the base price received ($/bushel), *B0pd* was the intercept of the protein premium/dockage function set by the grain elevator, *B1pd* was the coefficient on the grain protein content (*protein*, %) and *B2pd* was the coefficient on the squared protein term. *NR* was the net-return received and a function of the product of the *yield* (kg/ha) and *P*, minus the cost of the applied input (*CA*) multiplied by the as-applied rate of the input (*AA*), the fixed costs (*FC*) associated with production ($/ac) that do not include the input, and the cost per hectare of the site-specific application (*ssAC*).

The protein premium/dockage equation was fit using the data from one elevator in 2016. Base prices were gathered from Montana State University Extension Service (https://econtools.msuextension.org/mt_ag_prices/GrainPrices.html) and the cost of the N fertilizer, the fixed costs, and the cost of site-specific application were obtained from each farmer. As with yield and protein, sensitivity analysis for net-return was performed over all distinct fields and years, where all variables in Equation 3.4 were increased or decreased by 10%. Results of the most elastic parameter were compiled to assess variation in drivers of net-return across fields and years.

Results

Assessment of Spatiotemporal Differences in Crop Responses

We observed a large spread across field and year combinations in the probability of the difference between crop responses for the same field across years. While we saw a trend that the probability that yields tended to be greater in more recent years for given fields compared to prior years across all farms, some fields showed no discernable difference in yields between a given target year and a comparison year (e.g., comparing B4 yields in 2020 to 2016), while other fields showed that yields were greater in prior years compared to the target (the most recent) years (Table 3.3). Trends of the probability that grain protein content in target years were greater than prior comparison years were less clear, indicating that there is variability between fields in the similarity of protein responses across years for a given field (Table 3.4). While in most fields there is a chance that grain protein content varies across years, only B4 showed a consistent lack of differentiation between protein across time (Table 3.4).

The results also provide evidence that crop responses differ among other fields during a given year. When comparing the yields of a field and year of interest to yields of different fields for the same year, we did not detect a clear pattern across time that the probability that yields for fields of the same farm were more similar than yields between fields from different farms, indicating that the chance that yields differed on fields of the same farm can be similar to the chance that yields differed between fields of different farms for a given year (Table 3.5). However, for some fields and years, we saw high probabilities that yields were greater in the target field compared to fields of different farms, particularly between fields of Farm B and Farm I in 2019 and 2017 (Table 3.5). While we expected that fields of farms that were geographically closest (Farms D and I) would be more likely to be similar, we observed that despite the chance that yields from Farm B were greater than Farm I, the chance that yields from Farm B were greater than fields in Farm D tended to be lower (Table 3.5).

This demonstrates the variability in the chance that yields differ between fields of different farms because while Farm D and B had a large distance between them, they were more similar in yields than yields between fields of Farm I and B, despite the similar distance between farms. When comparing the chance that grain protein contents of a target field is greater than those in a comparison field for a given year there was no evidence of any pattern of similarities in the chance that grain protein contents were similar between fields of the same or different farms (Table 3.6). The chance that protein differed between fields of the same farm and fields of different farms were equally variable. The only trend is the high degree of variation across time, as there were no years where protein seemed more similar between fields than in other years. These results indicate that not only do crop responses for a given field vary across time, but crop responses also vary across fields within farms and across regions between farms.

Sensitivity Analysis for Drivers of Crop Responses and Net-Return

The resolution of measurements for precipitation and growing degree days were coarser than the size of the field, so the influence of these parameters on the response for a given field and year shrunk to zero during fitting of the GAM's due to the lack of variation in these parameters across the field. However, we maintain the assumption that weather influences rainfed crop responses. Nitrogen fertilizer rate was not as influential as we expected, even though it was experimentally varied. We had yield data from 28 field and year combinations, protein data from 27 field and year combinations, and 28 field and year combinations for net-return, albeit omitting protein as a parameter in Equation 3.4 for I2 in 2017.

We observed that across fields, varying the elevation parameter had the greatest influence on variation in predicted yields across years, with only B1 indicating that aspect had a higher influence on yield than elevation. This is important to note because it differed with the adjacent field B2, which has similar geography (Figure 3.3). There was much more variability in the most influential parameter on predictions of grain protein content across fields and years (Figure 3.4). While *elevation* was the most common driver of protein, we saw that in some fields and years nitrogen, pH of soil water, sand content, and the *watercontent* parameter had the most influence on protein in some fields and years (Figure 3.4). Drivers of grain protein content were relatively consistent over time, with only I4 and I1 showing any variation in the most influential parameter on protein across years where, in some years for these fields, protein was most influenced by nitrogen, while in other years, protein was most influenced by elevation or the pH of soil water.

There was also variability in the most influential parameter on net-returns (Figure 3.5). The most common parameter that influenced net-returns was the protein concentration in the field. Farmers either receive a premium or dockage on their base price received above or below a protein concentration that varies annually. As expected, yield also had a large influence on net-returns. The third most common driver of net-return for a farmer in a given field and year was crop base price, a factor set by the grain elevators that farmers sell to. Drivers of net-returns not only showed variation across space, but time as well, with some fields experiencing different drivers of net-return in different years. In D2, D3, and I4 some years were most influenced by base price, while in others it was most influenced by grain protein concentration. Considering these cases individually indicated commonality across space, as base price was the biggest driver in each of these fields in 2018. On Farm B, the major driver of net-return in B2, B3, and B4 varied between yield and protein. In 2019 we see variation across space in terms of the major drivers of net-return in adjacent fields of B1 and B2, where yield had the most influence on net-return in B1, while *protein* had the most influence on net-return in B2 that year. These results indicate that drivers of net-return are also not consistent across time and space, and that the metric most directly applicable to farmer's livelihoods (net-return) is not always a factor that they can manage (i.e. base price or protein price premium/dockage are largely out of their control).

Discussion

The fields and farms we analyzed for spatial and temporal agroecosystem variability are typical of conventional dryland small grain farms in practices and management approaches and therefore our results are likely indicative across rainfed cropland farms elsewhere in the Northern Great Plains of North America where winter wheat is a commonly grown crop. Our analysis provides evidence that crop responses in fields are variable across time and space. Thus, crop response functions to inform management that are independently parameterized on each field by varying input rates in field-specific experiments need calibration and adaptability over time to properly predict crop responses and net-returns. We did observe cases where there was little variation in crop response within a farm. Thus, we must acknowledge that there may be cases where fields on the same farm may not require independent experiments. However, one would not be able to arrive at such a conclusion without conducting field-specific experiments. Similarly, we acknowledge that although we saw evidence of differences in crop responses over years, the variables that logically contribute to this variation, like precipitation received, are not measured at an appropriate spatial scale to examine variation within a field. However, over a few years of experimentation, one might assume that temporal variation in crop response functions might be adequately captured by a sample of years (weather conditions) covering the historic or expected distribution to assess variation over time without considering the unknowns of climate change. Thus, field-specific experimentation might not be required for an indefinite period and we can expect that parameterization of models will stabilize after a certain number of years of data collection (Lawrence et al., 2015). However, with uncertainty in climate change, future work will not only require accumulation of data from different weather conditions (years), but will require assessment of differences in weather patterns, such as the distribution of precipitation across years, rather than assessment of differences in total accumulation of precipitation. Continuous monitoring of crop response to an input, albeit at reduced intensity (fewer plots across the field), will allow one to evaluate how differences in the pattern of weather conditions that vary across time influence crop responses.

There were differences in crop responses and their drivers in fields adjacent to each other, as well as differences across years for some fields. Not only did crop responses vary, but the covariates for characterizing crop responses differed in adjacent fields and for the same field across time. This implies that information from small-plot research centers could be disseminated more usefully by targeting farms with matching soil and weather characteristics rather than based on geographic proximity to the research center, which have been associated with farmer distrust in recommendations (Luschei et al., 2001). The implications of the lack of consistency between crop response models parameterized from experiments in fields that share a border are an unnecessary increase in uncertainty to input management recommendations. Thus, with ease of conducting OFPE, we conclude that all fields should receive experiments to make the most accurate and subsequent efficient recommendations.

Our hypothesis that crop responses would be driven primarily by the management of variable N fertilizer rate was not supported, as elevation was the primary driver of yields and protein. This indicates that variable application of fertilizer within fields must be managed with consideration of the ecological relationship between topography and crop responses. The topography of a field influences the direction and speed at which water flows, and as fertilizer and organic nitrogen are converted to highly soluble nitrate influencing availability of plant available N across fields (Sebilo et al., 2013a; Trudgill et al., 1991). Aspect influences water and subsequently N in a similar fashion, as north facing slopes in in the Northern Hemisphere receive less solar insolation and retain more moisture compared to south facing slopes. Weather covariates were also expected to have a high influence on crop responses, but because of the resolution at which they were collected, they had little to no variation across fields and often were singular values, and rarely were present in the final forms of GAM’s used for the sensitivity analysis.

The significant covariates in crop yield and protein response models contribute to the variation across years within fields. Ecological processes are complex, local at the sub-field scale and change over time. Thus, it is imprecise to apply management with a broad stroke across farms or fields, or from recommendations taken from sources that have never considered the field-specific conditions and history. Different N fertilizer rates need to be applied spatially across fields, let alone fields or farms with recognition of the ecological parameters that may influence crop responses more so than fertilizer. However, the time-specificity of responses indicate that the levels of rates applied in spatially distinct portions of the field need to vary over time and according to potentially varying factors like precipitation received, which can only be achieved by repeating the in-field experiment over years. Variation in the relationship between crop production and quality within fields and across years suggest that managers need to consider the competing interests of maximizing yield and protein and identify the rate that optimizes net-return and not simply yield or protein. We see further evidence for variation in relationship between crop yield and quality in the sensitivity/elasticity analysis for net-returns, where there was variability among fields in whether yield or protein contribute a greater effect on net-return. In some fields and years, the base price received for the crop was the most important driver of net-return received by the farmer and is a variable that producers cannot typically control or manage apart from contracts or on-farm grain storage capability.

These results provide evidence that optimum management that maximizes net returns to the farmer needs to be specific to each field and growing season, with decisions being made specifically for a field based on information gathered from that field. Variation occurs at spatiotemporal scales beyond a farmer’s control yet understanding this variation and crop responses at the finest scale possible allow one to scale management up without a loss of information. This variation is important to consider when developing decision support systems by recognizing that most fields are different from other fields and from themselves over time.

Understanding and managing based on spatiotemporal variation requires embracing technology and utilizing it as the empirical basis for adaptive management and modern objective decision making. The automatic transfer of data from machine monitors and field sensors is becoming increasingly easy through automated transfer to cloud-based software. The data available includes as-applied maps of variable rate application (VRA) of inputs such as seeds, fertilizers, pesticides and crop production and quality (e.g., crop yield and grain protein content). The additional prevalence of remotely sensed satellite data provides a trove of data available to understand and amend management of inputs specific to fields. Remotely sensed productivity data can be collected at 2- to 4-week intervals from open-source imagery or as often as desired with private drone flights. Increasing the number of weather stations on farms will provide spatially accurate and temporally dense measurements of variables such as precipitation and temperature. The combination of these temporally dense measurements provides information for adaptive management at a finer scale than was historically possible.

Extensive access to data enables rapid response to weeds, insects, disease, and informs site-specific recommendations of fertilizer, seeding rates, and other agricultural inputs. Resource efficiency remains limited even in areas with high technology adoption because experimentally applied inputs have not been employed with the specific purpose to improve crop response models and subsequent recommended management (Schimmelpfennig & Lowenberg-DeBoer, 2020). Greater adoption of the data returned from day-to-day farming operations can be exploited to increase resource efficiency and produce crops in a manner that reduces the environmental impacts and increase the ecological and economic resilience of agroecosystems (Bucci et al., 2018).

Industrial farms in the developed world are typically the largest, wealthiest, and the most automated. The wealth typically associated with larger farms suggests that these producers have the means to adopt new technology and tools, yet conventional input-centric farms are managed according to generalized input prescriptions. The farms with increasing sizes are primed for an automated adaptive management utilizing OFPE. Utilizing VRA technology to create in-field experiments repeated over years in the same field and crop with automated data retrieval and analysis can result in optimized input decisions. Adaptive management embraces simulation of outcomes under a range of weather and economic conditions the farmer will expect in the upcoming year and provides producers with the probability of outcomes based on quantified uncertainty from past outcomes. Decision making remains in the hands of the farmer/field manager who is responsible for the selection of management strategies for a specific field based on the information gathered from that specific field. Adaptive management creates a new level of science-based information between the traditional research center paradigm and the field-specific knowledge of the producer. Thus, adaptive management augments the field-manager's site-specific knowledge of input management and allows understanding the result of complex interaction between, physical, biological, and economic factors that determine optimum management given multiple objectives of profit maximization, pollution minimization and ultimately resilience to factors causing variability.

Conclusion

Ecological principles tell us that variability and crop responses are site, time, and field history of management-specific, and thus it is inappropriate to apply management with a broad stroke across farms or fields, or from recommendations taken from sources that have never considered field specific conditions and history. Our results suggest the time-specificity of crop responses indicate that the levels of nitrogen fertilizer rates applied in spatially distinct portions of the field need to vary over time and according to potential for different weather and varying economic factors.

These results provide basic evidence justifying the use of OFE for farm management and that management needs to be specific to each field and point in time, with management prescriptions being made specifically for a field based on information gathered specifically from that field. Development of management decisions requires OFE to determine the site-specific responses of crops within farmers' fields. Variation occurs at different spatiotemporal scales, something that farmers can control most effectively with field specific knowledge. Understanding the variation in crop responses at the finest scale possible allows one to scale management up without a loss of information and reveals the uncertainty most relevant to management. Typical treatment of data is to aggregate crop responses and as-applied rates to the scale of management, averaging over spatial areas and removing observed variation within fields. Aggregation and averaging results in information loss when quantifying uncertainty in recommendations because the response of crops to variable rates of input are generalized to the management scale with no recognition of the true variation in crop response within a treatment area.

Embracing the digital age of agriculture makes adaptive management empirically based and thus directly applied. Adaptive management involves iterative application of OFE and the vast stream of data from farms to inform management decisions (Lacoste et al., 2021). Automatic collection of data from farms and about farms is becoming increasingly easier through cloud-based software and open-source data repositories. These findings are relevant to farmers because it shows that drivers influencing crop responses and net-returns vary within farms and between adjacent fields of the same year. OFE will enable farmers to identify these drivers and understand how their inputs influence yield and protein within fields. Using information provided by OFE with decision support systems can enable farmers to make informed management decisions that maximize their profits and increase the efficiency of chemical inputs, such as nitrogen fertilizer.

Acknowledgements

The authors wish to thank the team members of the On-Field Precision Experiment (OFPE) project and the Agroecology Lab at Montana State University for their comments and edits through the development of this work. This research was supported by a USDA-NIFA-AFRI Food Security Program Coordinated Agricultural Project, titled “Using Precision Technology in On-farm Field Trials to Enable Data-Intensive Fertilizer Management,” (Accession Number 2016-68004-24769), the USDA-NRCS Conservation Innovation Grant from the On-farm Trials Program, titled “Improving the Economic and Ecological Sustainability of US Crop Production through On-Farm Precision Experimentation” (Award Number NR213A7500013G021), and the Montana Fertilizer Advisory Council from 2016 - 2021.

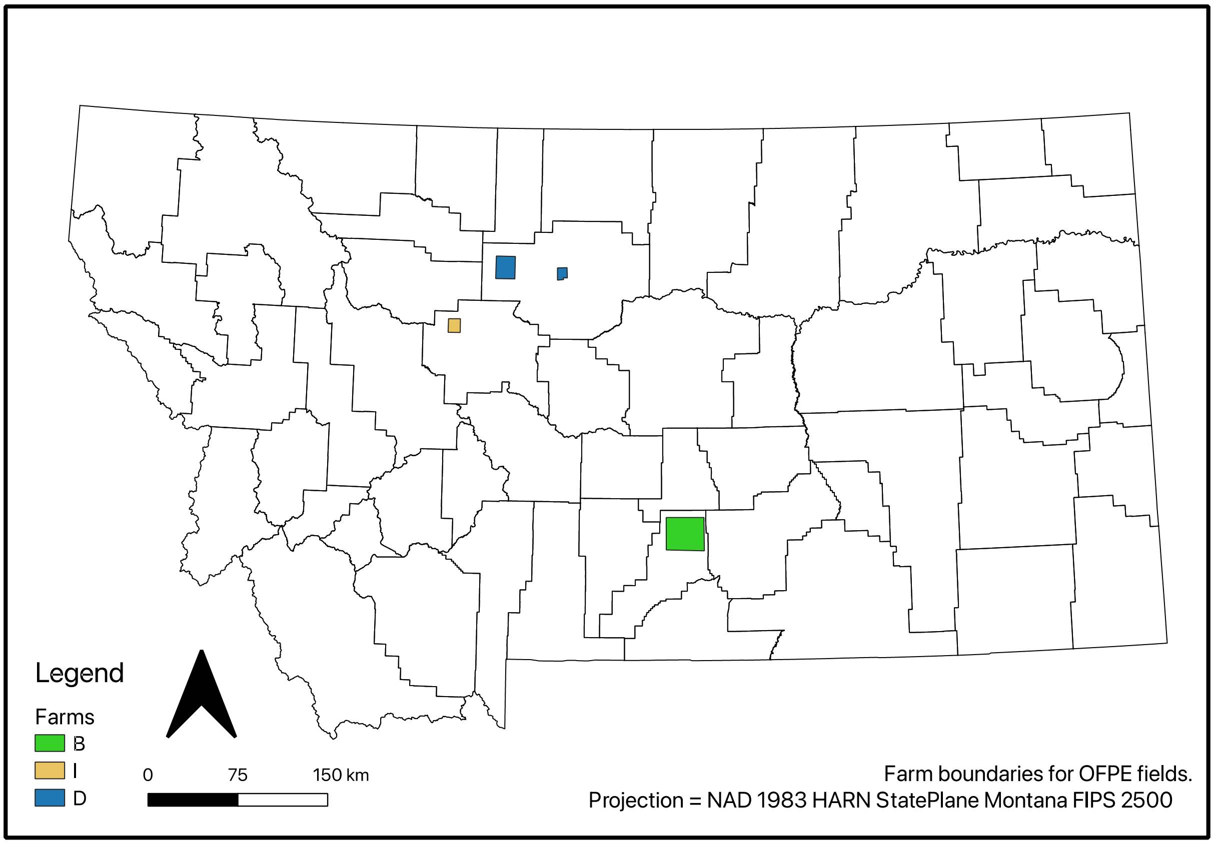


Figure 3.1. Map of OFPE farm boundaries for the three selected farmer collaborators with Montana State University. Colors represent different farmers, while shapes represent general areas in which their respective fields are located.

Table 3.1. Crop histories, field sizes, and years in VRA treatment for each field for given farmers. The farm identifier corresponds to the map above and is used instead of the farmer's name for privacy.

Farm	**Field**	**Field size (ha)**	**Crop History[1]:** *2014 / 2015 / 2016 / 2017 / 2018 / 2019 / 2020*	**Years N rate treatment**
B	B1	79	SF / WW / CF / WW / CF / WW / CF	2017, 2019
	B2	94	WW / CF / WW / CF / SF / WW / CF	2016, 2019
	B3	90	CF / SW / WW / CF / WW / CF / SW	2016, 2018
	B4	64	SW / CF / WW / CF / WW / CF / WW	2016, 2018, 2020
	B5	45	CF / WW / CF / WW / CF / SF / WW	2017, 2020
D	D1	46	CF / WW / CF / WW / CF / WW / CF	2017, 2019
	D2	48	WW / SW / WW / CF / WW / CF / WW	2016, 2018, 2020
	D3	20	SW / SW / WW / CF / WW / CF / WW	2016, 2018, 2020
	D4	191	WW / CF / WW / CF / WW /CF / WW	2018
I	I1	64	NA / WW / CF / WW / CF / WW / CF	2017, 2019
	I2	128	P / WW / CF / SW / CF / WW / CF	2017, 2019
	I3	64	CF / NA / NA / CF / WW / WW / WW	2019
	I4	94	SW / CF / WW / CF / WW / CF / WW	2016, 2018, 2020

[1] WW = winter wheat, CF = chemical fallow, P = peas, SW = spring wheat, SF = safflower, NA = Not Available

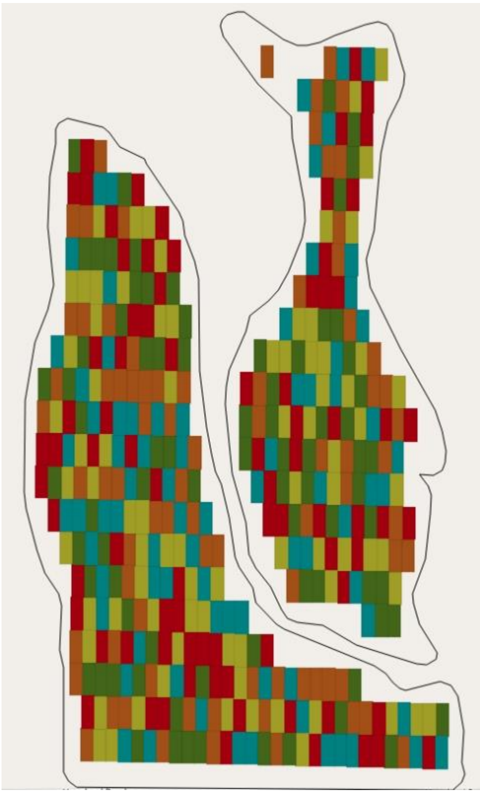


Figure 3.2. Conceptual example of a N fertilizer experiment where each color represents a distinct stratified randomly applied nitrogen fertilizer rate on a field split by an ephemeral water way. Treatment lengths vary by farmer preference and treatment widths are dictated by farmer's equipment width.

Table 3.2. Table of covariates gathered from Google Earth Engine to enrich the crop yield and protein datasets gathered from on-farms. In some cases, multiple sources are used, however only one data source is used when aggregating to yield and protein.

Data Type	Data Sources	Resolution	Years Collected	Description
Normalized Difference Vegetation Index (NDVI)	Sentinel 2, Landsat 5/7/8	10m, 30m	S2: 2016-present L5: 1999-2011 L7: 2012-2013 L8: 2014 - present	Sentinel 2 is from the European Space Agency as part of the Copernicus program. Landsat is an ongoing USGS and NASA collaboration. Bands (NIR, red) S2: B8 and B4 L5/L7: B4 and B3 L8: B5 and B4
Normalized Difference Red Edge (NDRE)	Sentinel 2	20m	S2: 2016-present	Bands B5 and B6
Red Edge Chlorophyll Index (CIRE)	Sentinel 2	20m	S2: 2016-present	Bands B7 and B5
Elevation	USGS NED	~10m (1/3 arc second), ~23m (3/4 arc second)	1999-present	USGS National Elevation Dataset. Measured in meters.
Aspect	USGS NED	~10m (1/3 arc second), 30m	1999-present	Direction the surface faces, function of neighboring elevations, in radians. Also calculated for each E/W and N/S direction as cosine and sine.
Slope	USGS NED	~10m (1/3 arc second), 30m	1999-present	Rate of change of height from neighboring cells, in degrees. Measured in degrees.
Topographic Position Index (TPI)	USGS NED	~10m (1/3 arc second), 30m	1999-present	Measure of divots and low spots as a function of neighboring elevation.
Precipitation	DaymetV3	1km	1999-present	Estimates from the NASA Oak Ridge National Laboratory (ORNL). Measured in mm.
GDD	DaymetV3	1km	1999-present	Estimates from the NASA Oak Ridge National Laboratory (ORNL).
susm	SMAP	10km	2016-present	Surface (0-5cm) and sub-surface (5-100cm) soil moisture content.
grtgroup	OpenLandMap	250m	1999-2018	Predicted USDA soil taxonomy great group probabilities.
texture	OpenLandMap	250m	1999-2018	Soil texture classes (USDA system) averaged over 6 soil depths (0, 10, 30, 60, 100 and 200 cm).
bulkdensity	OpenLandMap	250m	1999-2018	Soil bulk density (fine earth) 10 x kg / m3 averaged over 6 standard depths (0, 10, 30, 60, 100 and 200 cm).
claycontent	OpenLandMap	250m	1999-2018	Clay content in % (kg / kg) averaged over 6 standard depths (0, 10, 30, 60, 100 and 200 cm).
sandcontent	OpenLandMap	250m	1999-2018	Sand content in % (kg / kg) averaged over 6 standard depths (0, 10, 30, 60, 100 and 200 cm).
pH (phw)	OpenLandMap	250m	1999-2018	Soil pH in H_2O averaged over 6 standard depths (0, 10, 30, 60, 100 and 200 cm).
watercontent	OpenLandMap	250m	1999-2018	Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted and averaged over 6 standard depths (0, 10, 30, 60, 100 and 200 cm).
carboncontent	OpenLandMap	250m	1999-2018	Soil organic carbon content in x 5 g / kg averaged over 6 standard depths (0, 10, 30, 60, 100 and 200 cm).

Table 3.3. Probabilities that the yields for the given field in the target year (corresponding to A in Equation 3.1) were greater than yields for the given field in the comparison year (corresponding to B in Equation 3.1). D4 and I3 had only one year of data so no comparisons were possible.

Target (A)		**Comparison Year (B)**		
Field	*Year*	*2018*	*2017*	*2016*
B1	2019		0.687	
B2	2019			0.875
B3	2018			0.899
B4	2020	0.208		0.502
B5	2020		0.399	
D1	2019		0.885	
D2	2020	0.233		0.761
D3	2020	0.445		0.842
I1	2019		0.728	
I2	2019		0.749	
I4	2020	0.6		0.594

Table 3.4. Probabilities that the grain protein content for the given field in the target year (corresponding to A in Equation 3.1) were greater than grain protein content for the given field in the comparison year (corresponding to B in Equation 3.1). D4 and I3 had only one year of data so no comparisons were possible. No comparisons were made for I2 because of a lack of protein data.

Target (A)		**Comparison Year (B)**		
Field	*Year*	*2018*	*2017*	*2016*
B1	2019		0.857	
B2	2019		-	0.403
B3	2018			0.025
B4	2020	0.532		0.527
B5	2020		0.386	
D1	2019		0.355	
D2	2020	0.284		0.355
D3	2020	0.369		0.781
I1	2019		0.033	
I4	2020	0.402		0.073

Table 3.5. Probabilities that the yields of the target field for a given year (corresponding to A in Equation 3.2) are greater than yields from comparison fields for the given year (corresponding to B in Equation 3.2). Due to the stochastic nature of the bootstrapped sampling, the diagonal does not always correspond to a probability of 0.5.

2020

Target Field (A)	*Comparison Fields (B)*				
	B4	B5	D2	D3	I4
B4	0.500	0.364	0.523	0.353	0.366
B5		0.500	0.673	0.522	0.528
D2			0.500	0.285	0.307
D3				0.500	0.510
I4			-		0.500

2019

Target Field (A)	*Comparison Fields (B)*					
	B1	B2	D1	I1	I2	I3
B1	0.500	0.623	0.862	0.832	0.990	0.996
B2		0.500	0.729	0.728	0.986	0.997
D1			0.500	0.546	0.971	0.988
I1				0.501	0.917	0.984
I2					0.500	0.835
I3						0.500

2018

Target Field (A)	*Comparison Fields (B)*					
	B3	B4	D2	D3	D4	I4
B3	0.500	0.575	0.774	0.754	0.835	0.777
B4		0.500	0.721	0.696	0.794	0.732
D2			0.500	0.458	0.620	0.567
D3				0.501	0.670	0.597
D4					0.501	0.507
I4						0.500

2017

Target Field (A)	*Comparison Fields (B)*				
	B1	B5	D1	I1	I2
B1	0.500	0.540	0.908	0.908	0.995
B5		0.500	0.876	0.876	0.988
D1			0.500	0.463	0.961
I1				0.475	0.982
I2					0.497

2016

Target Field (A)	*Comparison Fields (B)*					
	B2	B3	B4	D2	D3	I4
B2	0.500	0.517	0.397	0.597	0.575	0.307
B3		0.500	0.369	0.572	0.551	0.273
B4			0.500	0.732	0.704	0.399
D2				0.500	0.476	0.148
D3					0.500	0.182
I4						0.499

Table 3.6. Probabilities that the grain protein content of the target field for a given year (corresponding to A in Equation 3.2) are greater than protein contents from comparison fields for the given year (corresponding to B in Equation 3.2). Due to the nature of the Monte Carlo sampling, the diagonal does not always correspond to a probability of 0.5.

2020

Target Field (A)	*Comparison Fields (B)*				
	B4	B5	D2	D3	I4
B4	0.499	0.813	0.724	0.630	0.744
B5		0.499	0.303	0.211	0.394
D2			0.499	0.341	0.575
D3				0.497	0.680
I4					0.500

2019

Target Field (A)	*Comparison Fields (B)*					
	B1	B2	D1	I1	I2	I3
B1	0.500	0.317	0.261	0.897	0.289	0.866
B2		0.499	0.420	0.973	0.462	0.966
D1			0.500	0.988	0.539	0.986
I1				0.500	0.014	0.347
I2					0.500	0.982
I3						0.474

2018

Target Field (A)	*Comparison Fields (B)*					
	B3	B4	D2	D3	D4	I4
B3	0.499	0.278	0.435	0.389	0.069	0.605
B4		0.499	0.698	0.671	0.108	0.826
D2			0.498	0.449	0.048	0.684
D3				0.498	0.044	0.737
D4					0.499	0.961
I4						0.501

2017

Target Field (A)	*Comparison Fields(B)*			
	B1	B5	D1	I1
B1	0.499	0.347	0.024	0.146
B5		0.499	0.033	0.219
D1			0.499	0.876
I1				0.498

2016

Target Field (A)	*Comparison Fields (B)*					
	B2	B3	B4	D2	D3	I4
B2	0.499	0.229	0.507	0.647	0.822	0.306
B3		0.498	0.855	0.966	0.994	0.705
B4			0.498	0.674	0.854	0.228
D2				0.498	0.758	0.063
D3					0.499	0.012
I4						0.498

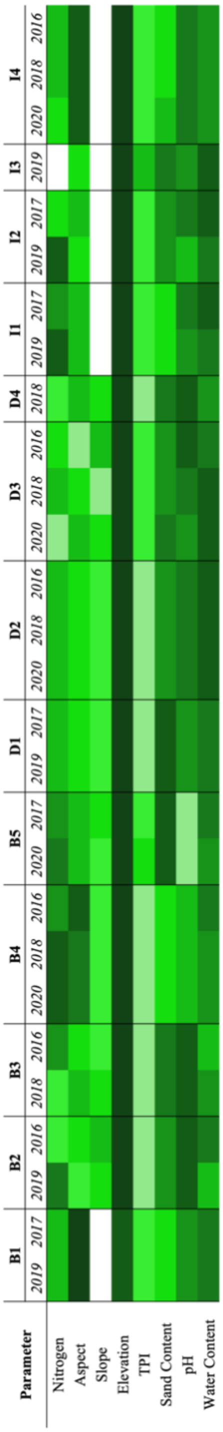


Figure 3.3. Results for sensitivity analyses of yield performed based on parameters of Equation 3.3 for each field and year. The darkest color represents the parameter with the most influence on yields with increasingly lighter colors corresponding to less influence on yields. Boxes left white indicate parameters that did not come into the model. For all models for all fields and years, precipitation, GDD, NDVI, bulk density, and clay content were all omitted due to model selection.

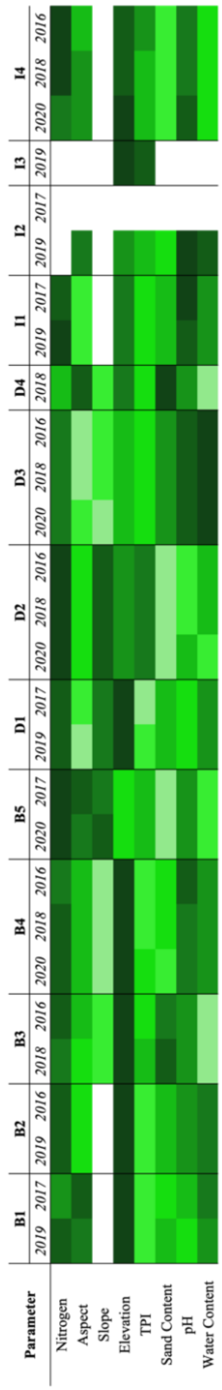


Figure 3.4. Results for sensitivity analyses of grain protein content performed based on parameters of Equation 3.3 for each field and year. The darkest color represents the parameter with the most influence on protein with increasingly lighter colors corresponding to less influence on protein. Boxes left white indicate parameters that did not come into the model. For all models for all fields and years, precipitation, GDD, NDVI, bulk density, and clay content were all omitted due to model selection. No sensitivity analysis was performed for I2 in 2017 because of a lack of protein data.

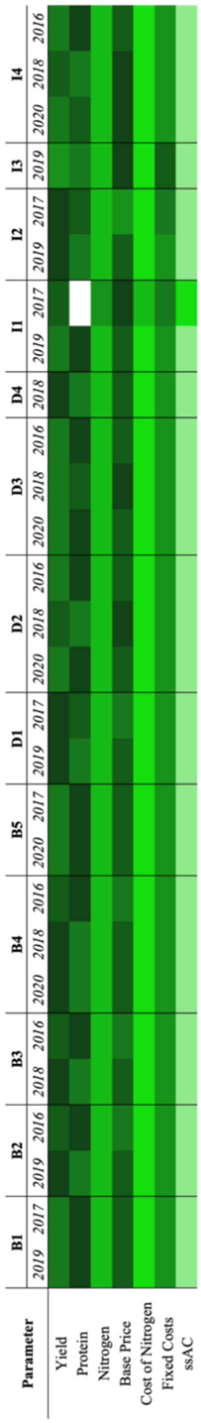


Figure 3.5. Results for sensitivity analyses of net-return performed based on parameters of Equation 3.4 for each field and year. The darkest color represents the parameter with the most influence on net-return with increasingly lighter colors corresponding to less influence on net-return. Boxes left white indicate parameters that did not come into the model. Grain protein content was not included in the sensitivity analysis for I2 in 2017 because of a lack of protein data.

REFERENCES CITED

Blackmore, S., & Moore, M. (1999). Remedial Correction of Yield Map Data. *Precision Agriculture*, *1*(1), 53–66. https://doi.org/10.1023/A:1009969601387

Bronson, K. F., Booker, J. D., Bordovsky, J. P., Keeling, J. W., Wheeler, T. A., Boman, R. K., Parajulee, M. N., Segarra, E., & Nichols, R. L. (2006). Site-specific irrigation and nitrogen management for cotton production in the Southern High Plains. *Agronomy Journal*, *98*(1), 212–219. https://doi.org/10.2134/agronj2005.0149

Bucci, G., Bentivoglio, D., & Finco, A. (2018). Precision agriculture as a driver for sustainable farming systems: State of art in litterature and research. *Quality - Access to Success*, *19*(S1), 114–121.

Cambouris, A. N., Zebarth, B. J., Ziadi, N., & Perron, I. (2014). Precision Agriculture in Potato Production. *Potato Research*, *57*(3–4), 249–262. https://doi.org/10.1007/s11540-014-9266-0

Cao, P., Lu, C., & Yu, Z. (2018). Historical nitrogen fertilizer use in agricultural ecosystems of the contiguous United States during 1850-2015: Application rate, timing, and fertilizer types. *Earth System Science Data*, *10*(2), 969–984. https://doi.org/10.5194/essd-10-969-2018

Clancy, M. (2019). Real-time protein monitoring. *Next Instruments*, *June*.

Elliot, E. T., & Cole, V. (1989). A Perspective on Agroecosystem Science. *Ecology*, *70*(6), 1597–1602.

Farid, H. U., Bakhsh, A., Ahmad, N., Ahmad, A., & Mahmood-Khan, Z. (2016). Delineating site-specific management zones for precision agriculture. *Journal of Agricultural Science*, *154*(2), 273–286. https://doi.org/10.1017/S0021859615000143

Flowers, M., Weisz, R., Heiniger, R., Osmond, D., & Crozier, C. (2004). In-Season Optimization and Site-Specific Nitrogen Management for Soft Red Winter Wheat. *Agronomy Journal*, *96*(1), 124–134. https://doi.org/10.2134/agronj2004.0124

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., … Zaks, D. P. M. (2011). Solutions for a cultivated planet. *Nature*, *478*(7369), 337–342. https://doi.org/10.1038/nature10452

Gliessman, S. R., & Engles, E. W. (2014). *Agroecology: The Ecology of Sustainable Food Systems* (3rd ed.).

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Hecht, M., & Crowley, K. (2020). Unpacking the Learning Ecosystems Framework: Lessons from the Adaptive Management of Biological Ecosystems. *Journal of the Learning Sciences*, *29*(2), 264–284. https://doi.org/10.1080/10508406.2019.1693381

Huggins, D. R. (2010). Site-Specific N Management for Direct-Seed Cropping Systems. *CSANR Research Report 2010 – 001*, Chp. 16. http://csanr.wsu.edu/wp-content/uploads/2013/02/CSANR2010-001.Ch16.pdf

Jackson, L., van Noordwijk, M., Bengtsson, J., Foster, W., Lipper, L., Pulleman, M., Said, M., Snaddon, J., & Vodouhe, R. (2010). Biodiversity and agricultural sustainagility: From assessment to adaptive management. *Current Opinion in Environmental Sustainability*, *2*(1–2), 80–87. https://doi.org/10.1016/j.cosust.2010.02.007

Khosla, R., Inman, D., Westfall, D. G., Reich, R. M., Frasier, M., Mzuku, M., Koch, B., & Hornung, A. (2008). A synthesis of multi-disciplinary research in precision agriculture: Site-specific management zones in the semi-arid western Great Plains of the USA. *Precision Agriculture*, *9*(1–2), 85–100. https://doi.org/10.1007/s11119-008-9057-1

Koch, B., Khosla, R., Frasier, W. M., Westfall, D. G., & Inman, D. (2004). Economic feasibility of variable-rate nitrogen application utilizing site-specific management zones. *Agronomy Journal*, *96*(6), 1572–1580. https://doi.org/10.2134/agronj2004.1572

Lawrence, P. G., Rew, L. J., & Maxwell, B. D. (2015). A probabilistic Bayesian framework for progressively updating site-specific recommendations. *Precision Agriculture*, *16*(3), 275–296. https://doi.org/10.1007/s11119-014-9375-4

Lin, B. B. (2011). Resilience in agriculture through crop diversification: Adaptive management for environmental change. *BioScience*, *61*(3), 183–193. https://doi.org/10.1525/bio.2011.61.3.4

Link, J., Batchelor, W. D., Graeff, S., & Claupein, W. (2008). Evaluation of current and model-based site-specific nitrogen applications on wheat (Triticum aestivum L.) yield and environmental quality. *Precision Agriculture*, *9*(5), 251–267. https://doi.org/10.1007/s11119-008-9068-y

Luschei, E. C., Van Wychen, L. R., Maxwell, B. D., Bussan, A. J., Buschena, D., & Goodman, D. (2001). Implementing and conducting on-farm weed research with the use of GPS. *Weed Science*, *49*(4), 536–542. https://doi.org/10.1614/0043-1745(2001)049[0536:iacofw]2.0.co;2

McBratney, A., Whelan, B., Ancev, T., & Bouma, J. (2005). Future directions of precision agriculture. *Precision Agriculture*, *6*(1), 7–23. https://doi.org/10.1007/s11119-005-0681-8

Moshia, M. E., Khosla, R., Longchamps, L., Reich, R., Davis, J. G., & Westfall, D. G. (2014). Precision manure management across site-specific management zones: Grain yield and economic analysis. *Agronomy Journal*, *106*(6), 2146–2156. https://doi.org/10.2134/agronj13.0400

Pretty, J. (2008). Agricultural sustainability: Concepts, principles and evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1491), 447–465. https://doi.org/10.1098/rstb.2007.2163

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Schimmelpfennig, D., & Lowenberg-DeBoer, J. (2020). Farm Types and Precision Agriculture Adoption: Crops, Regions, Soil Variability, and Farm Size. *SSRN Electronic Journal*, 1–38. https://doi.org/10.2139/ssrn.3689311

Sebilo, M., Mayer, B., Nicolardot, B., Pinay, G., & Mariotti, A. (2013). Long-term fate of nitrate fertilizer in agricultural soils. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18185–18189. https://doi.org/10.1073/pnas.1305372110

Sudduth, K. A., & Drummond, S. T. (2007). Yield editor: Software for removing errors from crop yield maps. *Agronomy Journal*, *99*(6), 1471–1482. https://doi.org/10.2134/agronj2006.0326

Sudduth, K. A., Drummond, S. T., & Myers, D. B. (2012). Yield editor 2.0: Software for automated removal of yield map errors. *American Society of Agricultural and Biological Engineers Annual International Meeting 2012, ASABE 2012*, *4*(12), 3378–3391. https://doi.org/10.13031/2013.41893

Talukder, B., Blay-Palmer, A., vanLoon, G. W., & Hipel, K. W. (2020). Towards complexity of agricultural sustainability assessment: Main issues and concerns. *Environmental and Sustainability Indicators*, *6*(March), 100038. https://doi.org/10.1016/j.indic.2020.100038

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108

Trudgill, S. T., Burt, T. P., Heathwaite, A. L., & Arkell, B. P. (1991). Soil nitrate sources and nitrate leaching losses, Slapton, South Devon. *Soil Use and Management*, *7*(4), 200–206. https://doi.org/10.1111/j.1475-2743.1991.tb00875.x

Weiner, J. (2017). *Applying plant ecological knowledge to increase agricultural sustainability*. 865–870. https://doi.org/10.1111/1365-2745.12792

Wood S.N., N. Pya and B. Saefken (2016) Smoothing parameter and model selection for general smooth models (with discussion). *Journal of the American Statistical Association* 111:1548-1575.

Wood, S.N. (2003) Thin-plate regression splines. *Journal of the Royal Statistical Society* (B) 65(1):95-114.

CHAPTER FOUR

ASSESSING PERFORMANCE OF EMPIRICAL MODELS FOR FORECASTING CROP RESPONSES TO VARIABLE FERTILIZER RATES USING ON-FARM PRECISION EXPERIMENTATION

Contribution of Authors and Co-Authors

Manuscript in Chapter Four

Author: Paul B. Hegedus

Contributions: Conceived the study design and methodology, compiled data, executed statistical analysis, and wrote initial manuscript.

Co-Author: Bruce D. Maxwell

Contributions: Assisted with study design, analysis, and manuscript preparation at all stages.

Co-Author: Taro Mieno

Contributions: Provided guidance on economics and assisted in manuscript editing.

Manuscript Information

Paul B. Hegedus, Bruce D. Maxwell, Taro Mieno

Precision Agriculture

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
__x__ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Published by Springer

Abstract

Data-driven decision making in agriculture can be augmented by utilizing the data gathered from precision agriculture technologies to make informed decisions that consider spatiotemporal specificity. Decision support systems utilize underlying models of crop responses to generate management recommendations, yet there is uncertainty in the literature on the best model forms to characterize crop responses to agricultural inputs, likely due for the most part to the variability in crop responses to input rates between fields and across years. Seven fields with at least three years of on-farm experimentation, in which nitrogen fertilizer rates were varied across the fields, were used to compare the ability of five different model types to forecast crop responses and net-returns in a year unseen by the model. All five model types were fit for each field using all permutations of the three years of data where two years were used for training and a third was held out to represent a "future" year. The five models tested were a frequentist based non-linear sigmoid function, a generalized additive model, a non-linear Bayesian regression model, a Bayesian multiple linear regression model, and a random forest regression model. The random forest regression typically resulted in the most accurate forecasts of crop responses and net-returns across most fields. However, in some cases the model type that produced the most accurate forecast of grain yield was not the same as the model producing the most accurate forecast of grain protein concentration. Models performed best when the data used for training models was collected from years with similar weather conditions to the forecasted year. The results are important to developers of decision support tools because the underlying models used to simulate management outcomes and calculate net-returns need to be selected with consideration for the spatiotemporal specificity of the data available.

Introduction

Farmers and scientists alike have recognized that crop responses to inputs vary between fields and across time, and even within fields (Hegedus & Maxwell, 2022). The digital data revolution in agriculture opens the door for providing insights and understanding of the ecological complexity of agroecosystems (Basso & Antle, 2020). Advances in remote sensing and the advent of big data gathered by precision agriculture technology allow farmers, managing ever greater land areas (Fig. 4.1), the ability to account for in-field variability and provide information that is otherwise infeasible for farmers to gather from day-to-day observations (Hatfield et al., 2002). Precision agriculture data makes information previously inaccessible to farmers accessible, enabling its use in decision making (Carolan, 2017). Although most farmers use precision agriculture technology, such as GPS and autosteer, more advanced technology and data utilization such as site-specific application of agricultural inputs have received far less adoption, stagnating around 10% in wheat systems (Schimmelpfennig & Lowenberg-DeBoer, 2020). Large farms are adopting precision agriculture technology at a faster rate than small farms (Schimmelpfennig & Lowenberg-DeBoer, 2020). This introduces the opportunity for policy incentives to encourage small farms to adopt data driven agriculture to address sustainability and efficiency issues related to the application of agricultural inputs. Even using precision agriculture data such as yield and soil maps increases the efficiency of agricultural inputs by 7.2 – 8.5% (McFadden et al., 2021). One aspect of precision agriculture technology and innovation, which represents a first step towards increasing sustainability on farms that have the capacity for it, is data-driven site-specific application of agricultural inputs that increase the efficiency of chemical usage and minimize potential degradation of ecological resources (Gebbers & Adamchuck, 2010).

Although the farm implement industry has wholly embraced big data, there still remains a disconnect between the collection and aggregation of data and utilization into decision making (Carolan, 2017). Agricultural businesses, academia, non-government organizations, and governments all recognize the issues faced in agriculture and are working toward providing solutions to food security in a sustainable manner, yet the approaches and methods for achieving these goals vary (West et al., 2014). The need for decision support systems that integrate data from farms and provide information for farmers to make decisions has long been recognized (McBratney et al., 2005; Wajid et al., 2021); even with industry led efforts into decision support systems and software, adoption of precision agriculture for efficient application of agricultural inputs remains limited (Schimmelpfennig & Lowenberg-DeBoer, 2020). Sole reliance on industry to provide the technological advances for generating prescription maps that increase the efficiency of agricultural inputs leads to yet another cost associated with farming, wherein producers are required to purchase software from companies that have obfuscated the process in which they generate maps for farmers (Gardner et al., 2021). Thus, farmers are removed from the decision-making matrix and the cost of accessing data intensive prescriptions reduce their net-returns. Even academic techniques for characterizing crop responses remain shrouded behind “black box” methods that obscure the way in which decisions are made, most commonly by machine learning and artificial intelligence approaches.

To make informed management decisions at the time when input decisions need to be made requires forecasting the response of future crops to inputs by simulating across unknown weather and economic conditions. The simulations not only enable farm managers to improve efficiencies of agricultural inputs, but enables understanding of the complexity and nonintuitive outcomes in their agroecological systems (Houlahan et al., 2017). Depending on the forecast model used, the difference in model selection at the field scale has important ramifications. Variation in model selection for enhancing agricultural input efficiency can increase yields by 5.6 to 11.9%, (McFadden et al., 2021). Additionally, management recommendations such as for nitrogen (N) fertilizer, vary between model types, highlighting the importance of understanding the implications and differences between models for decision making (de Lara et al., 2022).

Techniques for characterizing crop responses vary widely, and include detailed process-based crop models (Basso et al., 2011; Sela et al., 2016), parametric and non-parametric frequentist models (Bolton & Friedl, 2013; Johnson, 2014), traditional agronomic based models (Mueller et al., 2014), Bayesian approaches (Lawrence et al., 2015; McFadden et al., 2017), and machine learning methods (Peerlinck et al., 2018; 2019) all utilized in the literature and across industry. This paper addresses the knowledge gap around crop forecast model performance (Thöle et al., 2013) and, as a proof of concept, characterizes the response of rain-fed winter wheat to variable rates of N fertilizer for forecasting crop yield and protein content. A range of approaches to characterize responses (yield and grain protein concentration) to variable N fertilizer management for forecasting, and how this varies across space and time, were assessed. Additionally, how the different models influenced forecasts of net-returns (return on N fertilizer investment) was explored. Our objectives were to compare parametric, semi-parametric, machine learning, and Bayesian models to 1) assess the accuracy of winter wheat crop response (grain yield and grain protein concentration) forecasts in a hold-out year (representing a future year) from five different types of field specific models, 2) to investigate the influence the sequence of years used for training field specific models has on forecast accuracy, and 3) to evaluate the efficacy of the five types of models to predict observed net-returns in a hold-out year.

Methods

Study Location

The fields used in this study were from three farms geographically distributed across Montana that have been participating in the On-Farm Precision Experiments (OFPE) project at Montana State University for the past five years (Fig. 4.2). All fields selected were rain-fed conventional crop-fallow winter wheat systems that have available data from three years of variable rate N fertilizer rate experiments (Table 4.1). Variable rates were randomly applied, stratified by yield and grain protein concentration. Experiments in the first two years contained experimental plots where no N was applied, however on farm B in 2021, an agreement with the farmer meant that the lowest experimental N fertilizer rate (N-rate) would be 56 kg ha^{-1} instead of the statistically preferred control of zero.

Data Collection

Grain yield (kg ha^{-1}) was collected every three seconds directly from farmer's on-combine yield monitors that were calibrated prior to the harvest of each experimental field. Grain protein concentration (%) was collected every ten seconds with a combine mounted CropScan 3000H near infrared monitor (Clancy, 2019). Thus, the response metrics were gathered on different scales, and initially collected as two datasets. As-applied N fertilizer data were collected directly from each farmer's applicator equipment and all response and experimental data were cleaned using standard agronomic practices (Blackmore & Moore, 1999; Sudduth & Drummond, 2007; Sudduth et al., 2012). As-applied N fertilizer rates were georeferenced to the locations of each yield and protein observation in their respective datasets. Additional remotely sensed satellite covariate data were collected from the Google Earth Engine data repository (Gorelick et al., 2017) and georeferenced to the locations of each grain yield and protein concentration observation (Table 4.2).

Topographic data that do not change over time, such as slope, elevation, topographic position index (TPI), and soil characteristics were derived from datasets collected in 2015, an arbitrary year across the years collected (Table 4.2). Temporally varying covariate data were collected only through the year prior to the observed harvest to simulate realistic data constraints farmers would have when forecasting future crop responses. Vegetation indices (NDVI, NDWI) were collected up to December 31st of the prior year and two years prior harvest. For example, observed harvest data from 2021 would be modeled with the maximum vegetation index values (within field pixels) between 01/01/2020 - 12/31/2021 (py = previous year) and between 01/01/2019 and 12/31/2020 (2py = two years prior). Growing degree day (GDD) data was collected across the same timeframe for only the prior year (py). Precipitation data was collected up to October 31st of the prior year. For example, observed harvest data from 2021 would be modeled with the sum of precipitation between 11/01/2019 – 10/31/2020 (py = previous year).

The influence of the magnitude of covariate measurements on model fitting was reduced by centering covariate data, where centering is the subtraction of each covariate observation from the mean of the covariate observations. Soil characteristic and topographic data were centered because there was no expectation that the distribution of the predictor would vary across years.

Temporally varying data such as precipitation, GDD, and vegetation index (NDVI, NDWI) data were not centered because the distribution of these predictors could vary for a field in different years, meaning that the mean used for centering these data would be different than the mean required for centering in a new dataset from a year not used or not like a year used to train the model.

Experiment Design & Model Assessment

To make decision support systems useful for farmers, the crop response models they use must be accurate for forecasting future crop responses and allow for quantification of the uncertainty in the forecast. To mimic reality, in which models need to forecast crop responses in a future year not represented in training data, a k-fold cross validation type design was used. Crop grain yield, grain protein concentration, as applied fertilizer, and covariate data for each of the three years that every field underwent N fertilizer experimentation were collected as described above. The crop-fallow system practiced in all the fields means harvest data from experimental years are available every other year on a 2016, 2018, 2020 or 2017, 2019, 2021 schedule. For each field, three split cases with different data configurations were created where two years of data were used for training while data from one year were held out as a test set to evaluate the ability of the model to forecast crop responses in a year with observations unseen by the model (Fig. 4.3).

Split configuration A represents a scenario where the two oldest years of experimental data (2016 or 2017 and 2018 or 2019, respectively) are used to train models that predicted crop responses in the most recent year (2020 or 2021). Split configuration B uses the first (2016 or 2017) and most recent years (2020 or 2021) of experimental data for training models that predicted the crop responses in the intermediary year (2018 or 2019). Split configuration uses the most recent year (2020 or 2021) and the intermediary year (2018 or 2019) for training models that predicted crop responses in the first year of experimentation (2016 or 2017).

For each field and crop response (grain yield and grain protein concentration), the design matrix with the three data splits in Fig. 4.3 was applied to each of the five model types. For each of the 210 models (three-split case design across seven fields, with five model types, for two crop responses), the root-mean square error (RMSE) was used as the metric of performance. For each of the models tested, RMSE was calculated from the observed crop responses in the test year (yield or protein) and forecasts (predictions) of crop responses in the test year made by the model fit on the training years.

Models

Five parametric to non-parametric models that range from using frequentist, Bayesian, and machine learning approaches were fit to crop responses on each field. The models used were a modified version of a non-linear sigmoid model, generalized additive model (GAM), linear and non-linear Bayesian models, and a random forest regression model. For each field, each model type was fit individually using grain yield (kg ha^{-1}) or grain protein concentration (%) as response variables with variable as-applied N fertilizer rates (kg ha^{-1}) and covariates as predictor variables (Table 4.3). All analysis was performed in R (R Core Team, 2021), where the *OFPE* package was used for all data management and analysis (Hegedus, 2020). All covariate data in training datasets were tested to avoid aliasing and covariance between predictors. No interactions were included in the models, except for one, because analysis outline in this paper serves as the starting point for identifying the most appropriate model form for crop responses in each field.

After identification of the most appropriate model form, further testing and evaluation should be performed where model variations, refinements, and further feature selection are tested to generate the most accurate field specific forecasting model.

1) Modified Non-Linear Sigmoid Model (Beta Function)

The first model type was parametric and modified from a non-linear sigmoidal model called the beta function (Yin et al., 2003). Following the literature, a model that assumes a saturating asymptotic nature of crop response curves to additional N fertilizer was tested here (Anselin et al., 2004; Reynolds et al., 2021; Yin et al., 2003). Initial tests compared the beta function to multiple forms of non-linear models including hyperbolic, Richards, Gompertz, Weibull, and logistic forms. Selection of the beta function (Yin et al., 2003) over these other non-linear models was based on three factors; 1) the beta function yielded the lowest mean RMSE across fields for both yield and protein (Table 4.4), 2) the beta function required the least locking of parameters to facilitate convergence, and 3) the beta function captured the asymptotic response and a downturn of responses at high experimental N fertilizer rates.

The beta model was modified by incorporating an intercept term to represent the response of the crop at zero N-rate. Additionally, a Gaussian spatial correlation structure on the UTM coordinates of observations to account for the spatial dependence of responses was included (De Bastiani et al., 2015; Diggle & Tawn, 1998; Mardia & Marshall, 1984; Xia et al., 2008). The beta function was fit using generalized non-linear least squares regression through the *nlme* package in R (Pinheiro et al., 2021). Bottom-up feature selection was performed where predictors that did not contribute to a decrease in AIC by two units were omitted from the model. As crop responses differed across fields, not all beta functions fit for yield or protein for a given field and set of training years have the same predictors. All final models for each field were selected from the following beta function used for yield (kg ha^{-1}) or grain protein concentration (%) forecasting, with a Gaussian process spatial correlation structure, and the predictors that influence the intercept term (α);

$$R \sim \alpha + \left[(\beta - \alpha) * \left(1 + \frac{\delta_2 - N}{\delta_2 - \delta_1}\right) * \left(\frac{N}{\delta_2}\right)^{\frac{\delta_2}{\delta_2 - \delta_1}} \right] + \ Gaussian(x, y) + \ \epsilon$$

(Equation 4.1)

Where R is yield (kg ha^{-1}) or grain protein concentration (%), α is the fit minimum yield at a zero N-rate, β is the fit maximum yield (asymptote) at high N-rates, N is the as-applied N-rate in kg ha^{-1}, δ_2 is the fit N-rate where β is approached, δ_1 is the fit N-rate at the inflection point (center of the upturn of the curve), $Gaussian(x, y)$ is the Gaussian process spatial correlation structure, and ϵ is random error assumed to be normal and independent $N(0, \sigma^2)$. The intercept term, α, was assumed to be influenced by all topographic variables and information from prior years;

$$\begin{aligned} \alpha \sim \alpha 0 + \alpha_1 aspect_{cos}\ X\ aspect_{sin} + \alpha_2 slope + \ \alpha_3 elev + \ \alpha_4 tpi + \ \alpha_5 prec_{py} + \alpha_6 gdd_{py} \\ + \ \alpha_7 ndvi_{py} + \alpha_8 ndvi_{2py} + \ \alpha_9 ndwi_{py} + \alpha_{10} ndwi_{2py} + \alpha_{11} bulkdensity \\ + \ \alpha_{12} claycontent + \ \alpha_{13} phw + \ \alpha_{14} watercontent + \ \alpha_{15} carboncontent \end{aligned}$$

(Equation 4.2)

Where $\alpha_0 - \alpha_{15}$ were coefficients for the intercept and covariate data listed in Table 4.3.

2) Generalized Additive Model (GAM)

Staying in the non-linear realm but relaxing assumptions on the shape of the model for crop responses, a generalized additive model (GAM) was tested. Generalized additive models are a blend of semi-parametric generalized linear models with additive smoothing terms for modeling non-linear relationships between covariates and the response (Hastie & Tibshirani, 1986; Wood, 2017; Pinilla & Negrín, 2021). Despite their flexibility, the use of GAMs for modeling crop responses has been limited (Chen et al., 2019; Estes et al., 2013; Joshi et al., 2021; Liew & Forkman, 2015). The GAMs for grain yield and grain protein concentration used gamma family distributions and thin plate splines with shrinkage for the basis functions on most of the covariates in Table 4.3. A gamma family distribution was selected to ensure that estimates and predictions were above zero in accordance with the reality of grain yield and grain protein concentration measurements. Shrinkage allowed estimated degrees of freedom to fall to zero, combining the model fitting and feature selection process. The only covariates that did not use a thin plate basis function were the spatial UTM coordinates, which were included as smoothing terms with a Gaussian process basis function to account for spatial autocorrelation (Gotway & Stroup, 1997; Guisan et al., 2002; Holland et al., 2000; Zuur et al., 2012). To account for non-constant variance found in initial model fits, a log link function was used. Models were fit using the *mgcv* package in R (Wood, 2003; Wood et al., 2016). As crop responses differed across fields, not all GAMs for yield or protein for a given field and set of training years have the same predictors but selected from the following initial GAM with a gamma family distribution and log link function;

$$\begin{aligned} Gamma(\mathrm{R}, \mathrm{link} = \log) \sim I + \; & f_1(x,y) + f_2(aspect_{cos}, aspect_{sin}) + f_3(N) + f_4(slope) \\ & + f_5(elev) + f_6(tpi) + f_7(prec_{py}) + f_8(gdd_{py}) + f_9(ndvi_{py}) + f_{10}(ndvi_{2py}) \\ & + f_{11}(ndwi_{py}) + f_{12}(ndwi_{2py}) + f_{13}(bulkdensity) + f_{14}(claycontent) \\ & + f_{15}(phw) + f_{16}(watercontent) + f_{17}(carboncontent) + \epsilon \end{aligned}$$

(Equation 4.3)

Where R is the expected yield (kg ha^{-1}) or grain protein concentration (%) with a gamma log link function, I is the intercept, f_1 is the tensor product of the x and y coordinate (x, y) with a Gaussian process basis function, f_2 is the tensor product of centered aspect in the E/W direction ($aspect_{cos}$) and aspect in the N/S direction ($aspect_{sin}$) with a thin plate shrinkage spline basis function, $f_3 - f_{17}$ are smoothing functions fit with maximum likelihood and thin plate shrinkage splines as the basis function for covariates listed in Table 4.3, and ϵ is the random error, assumed to be normal and independent $N(0, \sigma^2)$.

3) <u>Non-Linear Bayesian Regression Model (Bayes NLR)</u>

Two Bayesian models were also tested. Bayesian approaches are commonly applied for modeling crop responses (Besag & Higdon, 1999; Huang et al., 2017; Jiang et al., 2009; Lawrence et al., 2015; Nandram et al., 2014; Ramsey, 2020; Wang et al., 2012) but are also used for models that influence policy such as yield gap and potential (Prost et al., 2008) and commodity prices (Drachal, 2019). The first Bayesian model used was adapted from Lawrence et al. (2015). Two modifications were made to fit this dataset and scenario, including replacing electrical conductivity data with clay content, based on the correlation between the two properties (McBratney et al., 2005). Second, the precipitation term used as a predictor for crop responses was constrained in this model to precipitation summed across the previous year, as only this data would be available when forecasting is required to make management decisions in an upcoming year. Finally, because this data represented scattered points across a field rather than a lattice grid structure in Lawrence et al., simultaneous spatial auto regression (SAR) was used rather than conditional auto regression (Ver Hoef et al., 2017; Wall, 2004). The base of this model was identical to Lawrence et al. (2015) where grain yield and protein were assumed to follow a normal distribution;

$$R \sim N(\mu, \sigma_e^2)$$

(Equation 4.4)

Where R was the grain yield (kg ha^{-1}) or grain protein concentration (%) that followed a normal distribution with a variance, σ_e^2, that followed an inverse gamma distribution;

$$\sigma_e^2 \sim IG(\alpha, \beta)$$

(Equation 4.5)

And the mean crop response, μ, was defined by a non-linear crop model;

$$\mu \sim \frac{\beta_{max}\, prec_{py}}{1 + \exp\,(\beta_{shp} - \beta_1 N - \beta_2 claycontent - \beta_3 claycontent\ X\ N)} + z$$

(Equation 4.6)

Where N was the as-applied N-rate (kg ha^{-1}), $prec_{py}$ and $claycontent$ are defined in Table 3, and z was the SAR spatial random effect;

$$z = (I - B)^{-1}\eta$$

(Equation 4.7)

Where η had a mean $E(\eta) \equiv 0$ and $Cov(\eta) \equiv \sigma_z^2 I$. I was the identity matrix, and B was the n x n spatial dependence matrix between n_i and n_j where $i \neq j$ and $i, j = 1, \dots, n$ (Hooten et al., 2019).

All priors used here were the same as in Lawrence et al., where the prior for σ_e^2 followed a gamma distribution with 0.01 used for both shape parameters in Equation 4.5. For β terms in (6), truncated normal distributions, $TN(0, 1000)$, were used for all terms except where $\beta_{max} =$

$N(0, 1000)$. All Bayesian methods were executed in R using the Stan based *brms* package (Bürkner, 2021; 2018; 2017).

4) Bayesian Multiple Linear Regression Model (Bayes MLR)

The second type of Bayesian approach tested was a multiple linear regression (MLR) model. Based on initial exploration of the data (Hegedus & Maxwell, 2022), a normal distribution as the base of a SAR MLR model was used (Anselin et al., 2004; Hooten et al., 2019; Kissling & Carl, 2008; Ver Hoef et al., 2018);

$$R \sim N(\mu, \sigma_e^2)$$

(Equation 4.8)

Where *R* was the grain yield (kg ha^{-1}) or grain protein concentration (%), σ_e^2 was identical in formation and priors as Equation 4.5, and;

$$\mu = X\beta + z$$

(Equation 4.9)

where β was the coefficient vector ($\beta_0, \ldots, \beta_{p-1}$) with *p* the number of predictors from Table 4.3, *X* is the *n x p* covariate matrix of covariate vectors, and *z* was the SAR random effect. To avoid overfitting, top-down feature selection was applied and omitted covariates where dropping them from the model reduced the AIC by more than two units. Thus, Bayesian MLR models vary in their covariates between fields and splits but are all fit from the same initial model;

$$\begin{aligned}\mu \sim \beta_0 + \beta_1 N + \beta_2 aspect_{cos}\, X\, aspect_{sin} + \beta_3 slope + \beta_4 elev + \beta_5 tpi + \beta_6 prec_{py} \\ + \beta_7 gdd_{py} + \beta_8 ndvi_{py} + \beta_9 ndvi_{2py} + \beta_{10} ndwi_{py} + \beta_{11} ndwi_{2py} \\ + \beta_{12} bulkdensity + \beta_{13} claycontent + \beta_{14} pw + \beta_{15} watercontent \\ + \beta_{16} carboncontent + z\end{aligned}$$

(Equation 4.10)

Where predictor variables are covariates from Table 4.3 with uniform priors on coefficients $\beta_0 - \beta_{16}$ (Gelman, 2006), and the SAR spatial random error, z, defined in Equation 4.7. As with the non-linear modeling, the Stan based *brms* package in R was used for Bayesian analyses (Bürkner, 2021; 2018; 2017).

5) Random Forest Regression (RF)

Random forest (RF) regression is an increasingly popular non-parametric machine learning method that has been used to model spatial data (Georganos et al., 2021; Jing et al., 2016; Li et al., 2011; Shi et al., 2015), specifically crop response data (Everingham et al., 2016; Jeong et al., 2016; Lawes et al., 2019; Mariano & Mónica, 2021; Marques Ramos et al., 2020; Paccioretti et al., 2021; Wang et al., 2016), where crop responses are estimated by the ensemble of regression trees using binary splits of observations based on covariate data. The *ranger* package in R was used for fitting and generating predictions (Wright & Ziegler, 2017). The predictors included in the RF models for yield and protein are in Table 4.3. To avoid over-fitting, top-down feature selection was performed where covariates were omitted if dropping them from the model resulted in an AIC decrease of over two units. The number of trees (*ntrees*) and the number of covariates sampled at each node (*mtry*) were optimized during the fitting process. To account for spatial effects, UTM coordinates were included as a variable definition following spatial random forest regression practices in the literature (Janatian et al., 2017; Langella et al., 2010; Walsh et al., 2017; Wang et al., 2017).

Net-Return Calculation and Prediction

While predicting yield and grain protein concentration are important, the metric of interest for most farmers is net-return, as management decisions are often ultimately based on the expected profit of the given management action. However, a net-return dataset had to be created for each field and year as this data was unavailable. In Montana's wheat systems, grain yield and protein concentration influence net-returns. A nearest neighbor analysis was performed for each split case, field, and year where yield and protein were co-located into a test dataset from which the ability of models to predict net-return could be evaluated. To reduce the distance between neighbors of the differing datasets, the nearest yield measurement was georeferenced to each protein measurement because yield observations were collected at a higher density than grain protein concentration observations. To assess the ability of each model type to forecast net-returns, yield and protein were predicted from their respective models to the same locations as in the "observed" net-return dataset, predicted net-return was calculated using the modeled values, and performance was measured by the RMSE between the "observed" and predicted net-returns.

As mentioned above, protein concentration is used to calculate wheat farmer's net-return in Montana. Farmers receive premiums or dockages to their base price based on the protein concentration of their wheat. Due to data constraints, the protein premium/dockage schedule from 2021 from one elevator in Billings, Montana was assumed for all fields and years. This schedule dictated that farmers received an additional two cents per half percent protein above 11.5% to 14% and lost eight cents per half percent protein below 11.5% to 9.5% off their received base price. Other economic data used to generate values for the economic parameters that influence net-return was gathered for each year from a combination of surveys from the

United States Department of Agriculture (USDA) Economic Research Service (ERS) and the farmers in this study;

$$P = Bp + (B0pd + B1pd * protein + B2pd * protein^2)$$

(Equation 4.11)

$$NR = yield * P - CA * AA - FC - ssAC$$

(Equation 4.12)

Where *P* was the final price received ($ kg^{-1}) after imposing the protein premium/dockage on *Bp*, the base price received ($ kg^{-1}), *B0pd* was the intercept of the protein premium/dockage function set by the grain elevator, *B1pd* was the coefficient on the grain protein concentration (%), and *B2pd* was the coefficient on the squared protein term. The net-return, *NR* ($ ha^{-1}), received was a function of the product of the *yield* (kg ha^{-1}) and *P*, minus the cost of the applied input (*CA*) multiplied by the as-applied N-rate of the input (*AA*), the fixed costs (*FC*) associated with production ($ ha^{-1}) that do not include the input, and the cost per hectare of the site-specific application (*ssAC*).

No cost of site-specific application (*ssAC*) was included since each farmer in this study used their own already available equipment to enact the varied rates. The parameters for the protein premium and dockage, *B0pd*, *B1pd*, and *B2pd*, were derived from fitting a smoothed step function to the protein premium schedule. The *Bp*, *CA*, and *FC* values used for each year are in Table 4.5.

Results

Decision support systems require models for forecasting the response of crops to variable agricultural inputs to assess management decisions when the conditions in the future year are unknown at the decision point in time. Thus, the ability of models to predict crop responses in datasets that have been held out from the model training process are vital for forecasting crop responses and expected net-return in future years. The RMSE values for grain yield, grain protein concentration, and net-return were reported for each of the five model types, for each field, and for each data split case (Fig. 4.3; Appendix A).

The two objectives of first identifying the model that best forecasts crop responses and then identifying the influence of training data (years) on making accurate forecasts of crop responses are linked, so they cannot be tested independently or in a logical order. This is because it is quite possible that different forecast models and different data configurations perform differently on different fields. It was assumed a priori that the field-specific model type that performed best for a given field would be identified by averaging forecast performance across data configurations. The influence of the data configuration of training data on prediction performance for each model type was assessed by comparing the mean RMSE across fields from each model type and data configuration.

The first objective was to assess the accuracy of five different types of winter wheat crop response (grain yield and grain protein concentration) forecast models using all combinations of training and holdout years (Fig. 4.3). The mean of the RMSE of observed and predicted grain yields were taken across the three data configurations for an average RMSE per model for each field (Table 4.6). These results indicate that in six out of seven fields, across any year forecasted, the RF regression resulted in the most accurate forecast of grain yield in a given test dataset with one year of data held out. Bayesian approaches were the next most accurate, where in one out of seven fields the Bayesian NLR model had the lowest RMSE out of all the model types. In two out of seven fields the Bayesian NLR had the second lowest RMSE out of all model types. Additionally, in three out of seven fields the Bayesian MLR model had the second best RMSE. The GAM was the poorest predictive model type, with the highest RMSE out of all models in five out of seven fields.

The mean RMSE of observed and predicted grain protein concentration were calculated across the three data cases in Fig. 4.3 for each field (Table 4.7). In four out of seven fields, predictions of grain protein concentration from the RF regression had the lowest mean RMSE across data configurations and in three out of seven fields had the second lowest RMSE. The non-linear beta function had the lowest RMSE for one out of seven fields and second lowest RMSE in four out of seven fields. These results indicate that, like the RF regression models for grain yield, the RF models generated the most accurate forecasts of grain protein concentration, averaged across data configurations. On average across fields, the second-best model for forecasting grain protein concentration was the beta function. In contrast to results of grain yield models, the GAM was the best predictive model type for grain protein concentration in two out of seven fields, while the Bayesian approaches were the least accurate at forecasting protein.

The second objective was to investigate the influence of the sequence of years used for training field specific models on forecast accuracy. The mean RMSE across fields from each crop response model for each split data case in Fig. 4.3 was calculated where; A = first two years of experimental data were used for training and the most recent year of experimental data was the test set, B = the first and last years of experimental data were used for training and the intermediary year was the test set, C = the most recent years of experimental data were used for training and the first year of experimentation was used for testing.

The mean RMSE of grain yield for each model using each data split case were taken across fields to assess which set of training years (split cases) resulted in the most accurate forecasts (Table 4.8). For the 2016, 2018, 2020 data collection schedules, split case C had the lowest mean RMSE of grain yield across fields for four out of the five models, with the exception being that split case B resulted in the lowest RMSE across fields using a beta function model. In the 2017, 2019, 2021 data schedule, split case C yielded the lowest mean RMSE of grain yield across fields for three out of the five models, while split case A yielded the lowest mean RMSE across fields when the GAM or Bayesian MLR model were used (Table 4.8).

The same assessment across data splits (Fig. 4.3) was performed where the models were used to forecast grain protein concentration and the mean RMSE was taken across fields (Table 4.9). For the 2016, 2018, 2020 data collection schedules, split case B had the lowest mean RMSE of protein across fields for four out of the five models with the exception being that split case C resulted in the lowest RMSE across fields when the Bayesian MLR was used. In the 2017, 2019, 2021 data schedule, split case C yielded the lowest mean RMSE for protein across fields for four out of the five models, while split case B yielded the lowest mean RMSE across fields when the RF regression model was used.

Accurate predictions of grain yield and protein concentration are necessary for calculating net-return, which is the metric of interest for most farmers when making decisions. Both the grain yield and protein models fit for a given field with a given model type were used to predict yield and protein in a dataset with calculated net-returns from the test year. The same tables as above were generated for summarizing the RMSE of net-returns.

The mean RMSE values for net-return were calculated across the three data split cases for each model type in each field (Table 4.10). The RF regression model produced the lowest mean RMSE of net-return across data split cases in four out of seven fields, and the second lowest mean RMSE in two out of seven fields. The second-best model for forecasting net-returns was the beta function, which produced the lowest RMSE of net-return across data split cases in two out of seven fields and produced the second lowest mean RMSE across data types in three out of seven fields.

As done for grain yield and protein above, the mean RMSE across fields was calculated for each data split case from Fig. 4.3 and model type (Table 4.11). For both data schedules there was no consistent split case for training and testing data across models based on the mean of net-return RMSEs across fields. In both data schedules the split test case where the intermediary year was used as the test scenario resulted in the lowest RMSE for two out of five models. Split case A and C generated the most accurate forecasts from Bayesian NLR models in the 2016-2020 and 2017-2021 data schedules, respectively. Split case C and A generated the most accurate forecasts from Bayesian MLR and RF regression models in the 2016-2020 and 2017-2021 data schedules, respectively, in juxtaposition to that of the Bayesian MLR and RF regression models.

Discussion

Variable rate application technology along with crop yield and quality monitors has allowed on-field experimentation to produce data that can refine input recommendations specific to each field (Hegedus & Maxwell, 2022). Consistently accurate models to forecast crop responses to agricultural inputs and the return on investment of the inputs are essential for building trusted input recommendations.

Due to the variation in crop responses across fields, and the drivers of responses and net-returns, between fields and across years (Hegedus & Maxwell, 2022), it was expected that the forecast model type most appropriate for characterizing the crop response to variable N fertilizer rates would vary across fields and depend on the years used for training the model. For both grain yield and grain protein concentration, the results indicate that the RF regression most consistently reduced the uncertainty of forecasts, based on the mean RMSE across the three test years. This contrasts with the assumption that the most appropriate model for predicting crop responses to variable N rates would vary across fields and crop responses. Yet the RF regression model should not be considered a silver bullet, despite performing best on average across fields. Despite performing best on average across fields, this simplifies field-specific variation in the most predictive model type and variation in the most predictive model for different crop responses in a given field. A RF model seems a reasonable assumption for a default model that a selection of other model types were compared against, yet the importance of testing multiple model types is demonstrated in field B1, where the Bayesian NLR regression model performed better than the RF at forecasting grain yield (Table 4.6). Additionally, the Beta model performed better than the RF model at predicting grain protein concentration in field B1 (Table 4.7). This is important for two reasons; the first being that the RF was not the most predictive model type for this field, despite performing best on average across fields, and the second that the most predictive model type for field B1 varied between the crop responses. The GAM outperformed the RF at predicting grain protein concentration in fields B3 and D3 (Table 4.7), indicating two other cases where, without conducting model selection on a field-specific basis, assuming a default (e.g., RF) model type would not result in the most accurate forecasts of grain protein concentration and subsequent net-return.

Variation in the most predictive model type for different crop responses was further illustrated when assessing the second-best model types for each field. The second-best model at predicting grain yield was typically one of the Bayesian methods (Table 4.6) while the beta function was typically the second-best model type at predicting grain protein concentration (Table 4.7).

Based on the influence of grain yield and grain protein on net-return and the results from objective 1, it was unsurprising that using RF regression for modeling both crop responses to calculate net-return tended to result in the most accurate net-return forecasts across field (Table 4.10). However, like the results for grain protein concentration, the second-best model across fields for generating an accurate forecast of net-returns was the beta function. In fields B1 and D3, the model type that produced the best forecasts of grain protein also produced the best forecasts of net-return. This indicates that when using models to forecast net-return there is a higher importance of generating accurate predictions of grain protein concentration compared to grain yield, though dependent on the protein premium/dockage set by the grain elevator. If accurate grain yield predictions had a higher influence on forecasting net-return, it would be expected that the second-best model type for net-return would be the Bayesian methods. Field I1 represents an outlier to this assessment however, as the RF regression produced the most accurate forecasts of grain yield and grain protein independently (Table 4.6; Table 4.7), yet when the beta function model was used for predicting both yield and protein to calculate net-return resulted in a more accurate forecast compared to when the RF regression was used for predicting yield and protein to calculate net-return (Table 4.10).

Net-return forecasts are important for simulating management recommendations and in forming the basis of farmer decision making. Despite the RF regression model's dominance in forecasting, these results should not detract from the importance of the field-specificity in model selection and the difference in the most appropriate model for the different crop responses, which require confronting data from on-farm experimentation with multiple models. For example, in a hypothetical situation for a field where a given model type produces the best forecasts for yield and an alternate model type produces the best forecasts for protein, forecasting net-returns and simulating management recommendations in a decision support tool should be based on separate models for grain yield and grain protein concentration. In the data shown in this paper, an example of this concept arises in fields B1, B3, and D3. Highlighting field B1, the Bayesian NLR model produced the best forecast for grain yield while the beta function model produced the best forecast for grain protein concentration, indicating not only that the RF regression should not be used at all in this situation, but that two separate model forms should be used for yield and protein. Thus, mixing and matching models for grain yield and grain protein might produce more accurate forecasts of net-return.

When selecting models, consideration must be given to the weather conditions in the years that data are available for training, as some models produce better predictions in certain weather conditions than others. It was predicted that if weather conditions of the years used for training differed from the conditions of the year in the test dataset forecasts of crop responses would be less accurate compared to when the conditions in the years used for training resembled the conditions in the year of the test dataset. Since the data used in this study came from rain-fed winter wheat production fields, amount of precipitation received was judged to be the best weather condition to assess year suitability for predictions. The wettest years in our data set included 2018 and 2019, while 2020 and 2021, were on average the driest (Table 4.12).

Averaged across fields, most of the five model types generated the most accurate forecasts of grain yield in split configuration C, where the most recent two years of experimental data were used to train models that made predictions on yields in the first year of experimentation (Table 4.8). The out-of-sequence result is interpreted as that the most accurate forecasts of grain yield were made when training data contained a wet and dry year, which meant that the relatively dry year in the test set was reasonably represented in the training data. The least accurate forecasts of yield were made in split configuration B, where the two drier years were used for training and forecasts were made for the wet year. Having a range of weather conditions (wet and dry years) in the training dataset meant that forecasts in an unknown year were stronger compared to when data from similar weather conditions were used to train a model.

The fields analyzed were collected on two crop-fallow schedules, with the 2016 schedule containing data collected in 2016, 2018, and 2020 and the 2017 schedule containing data collected in 2017, 2019, and 2021. The influence of the configuration of training and testing data on the predictive performance of grain protein models depended on whether training and testing data came from the 2016 or 2017 data schedule (Table 4.9). When data came from the 2017 schedule, split configuration C resulted in the most accurate forecasts across most models, like for grain yield. However, when the data came from the 2016 schedule the most accurate forecasts for most models came from split configuration B, where 2016 and 2020 were used for training to predict protein concentration in 2018. The 2017 schedule was the most extreme, with both the wettest and driest precipitation years (Table 4.12). Additionally, the difference in precipitation between the two driest years in the 2017 schedule, 2021 and 2017, was 21.6 mm which indicates that conditions in 2021 and 2017 were very similar. Thus, if the training dataset contained either 2021 or 2017 and the test set was either 2017 or 2021, the conditions of the test set would be represented in the training data (configuration A or C). On the other hand, precipitation in the 2016 schedule had a smaller range of precipitation, but precipitation varied between years more so than in the 2017 schedule where the difference in precipitation between two driest years of the 2016 schedule, 2020 and 2016, was 47.1 mm (Table 4.12). This indicates that in the 2016 schedule, there is not a pair of years that are as similar as 2021 and 2017, so no data configuration existed where conditions in the test year were almost identical to conditions in a year from the training dataset.

Analysis on Montana dryland conventionally managed agroecosystems has shown that variable rate experimentation of N fertilizer inputs requires 6-8 crop years of data for model coefficients to converge and enable accurate predictions of outcomes (Lawrence et al., 2015). This further highlights the need for public research funding of long-term agricultural studies on how precision agriculture technologies can improve management recommendations that promote sustainability (DeLonge et al., 2020). The results in this paper were based off three years of crop data, below the recommended threshold, and further illustrate the importance of a larger set of data from which to train models. More accurate forecasts in unknown years are expected with more than three years of data, because more weather conditions in the training datasets will increase the likelihood that conditions in the next year will be represented in the model fitting process. The importance of representation of conditions in an upcoming year in the training data is demonstrated using the situation where the best forecast for grain yield was observed: field D1 where the RF model was used with data configuration C (Table 4.13). Assuming perfect representation of conditions in the next year, achieved by training the model with a subset of data containing all three years and testing in the holdout set containing all three years, the forecast accuracy was almost halved compared to when 2019 and 2021 data were used to train the model and tested on 2017 data (Fig. 4.4). As a backlog of data from varying weather years is collected, it is expected that forecasts will iteratively improve year after year until uncertainty in crop response forecasts is constrained to model structure and selection rather than limitations in the conditions represented in the training data.

A further concern over interpretation of our results, especially in selection of model training data, is the influence of the fallow year or rotation crop inclusion. The interaction of these factors with weather will take more years to refine the models and characterize the certainty in recommendations. As more data are gathered and synthesized into decision support systems via modeling efforts, researchers and scientists need to recognize potential pitfalls of their modeling approaches beyond just model selection, such as issues of scaling model results and the data used in models (Fritsch et al., 2020).

Crop response models used in decision support systems directly influence management recommendations. The certainty of input management recommendations has been sharply improved with on-farm experimentation facilitated by precision technologies, site-specific data availability, and modern analytics. These results suggest continued use of on-farm experimentation to produce a backlog of data for training models in various weather conditions, crop rotations, and practices so the crop response signal can emerge from the noise caused by within field crop response and weather variation. It must also be recognized that field-specific data might require confrontation with a suite of models to make the best assessment of which model to base input recommendations. These results demonstrate that modern farms have the equipment and technology to inform adaptive management through on-farm experimentation, and that adoption will rely on development of decision support systems, demonstration of the profitability of variable rate application, and possible policy incentives.

Conclusion

Data generated from on-farm experimentation on seven fields were used to evaluate five different types of crop response models. The RF regression typically resulted in the most accurate forecasts of crop responses and net-returns across most fields, but not all. In some cases, the model type that produced the most accurate forecast of grain yield was not the same as the model producing the most accurate forecast of grain protein concentration. When the data used for training models was collected in years with varying weather conditions, forecasts of crop responses in a new year tended to be more accurate than when models were trained on data from similar weather conditions and used to forecast crop responses in a year with dissimilar weather conditions. When selecting crop response models for generating forecasts of net-returns and simulating management outcomes in a decision support system, field specificity in the most appropriate model type for each individual crop response should be considered. Additionally, the most appropriate model for a given crop response in each field may depend on the range of weather conditions in the years that data is available to train that model. These results are important to developers of decision support tools because selection of the underlying models used to simulate management outcomes and calculate net-returns need to consider the spatiotemporal specificity of the data available.

Acknowledgements

The authors wish to thank the team members of the On-Field Precision Experiment (OFPE) project and the Agroecology Lab at Montana State University. This research was supported by a USDA-NIFA-AFRI Food Security Program Coordinated Agricultural Project, titled “Using Precision Technology in On-farm Field Trials to Enable Data-Intensive Fertilizer Management,” (Accession Number 2016-68004-24769), the USDA-NRCS Conservation Innovation Grant from the On-farm Trials Program, titled “Improving the Economic and Ecological Sustainability of US Crop Production through On-Farm Precision Experimentation” (Award Number NR213A7500013G021), and the Montana Fertilizer Advisory Council from 2016 - 2021. Additionally, the authors would like to thank the Montana INBRE funded Statistical Consulting & Research Services at Montana State University, specifically Dr. Mark Greenwood and Caroline Hardy.

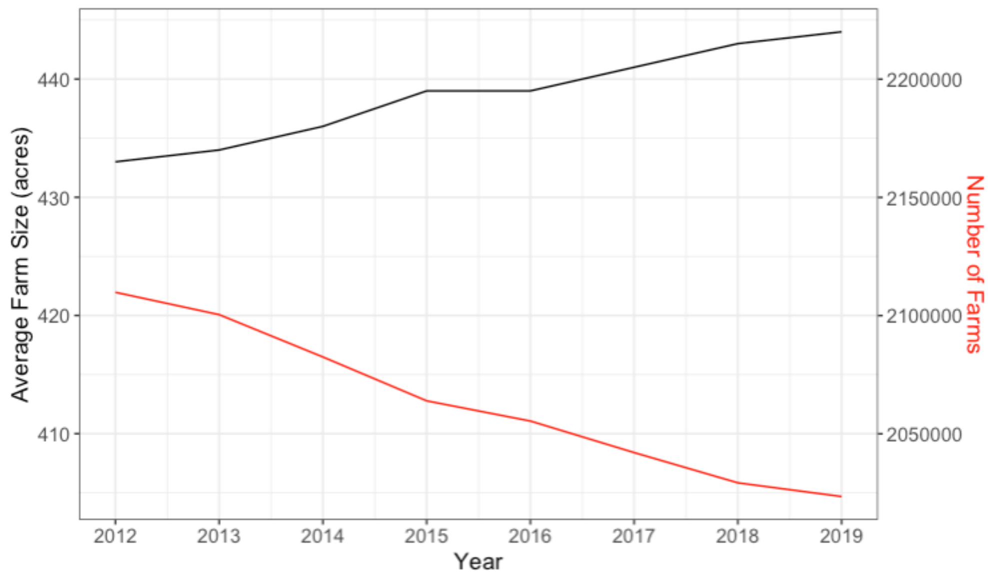


Figure 4.1. Average farm sizes in the United States (left y-axis; black line) and number of farms (right y-axis; red line) from 2012 to 2019. Data from USDA NASS, 2019.

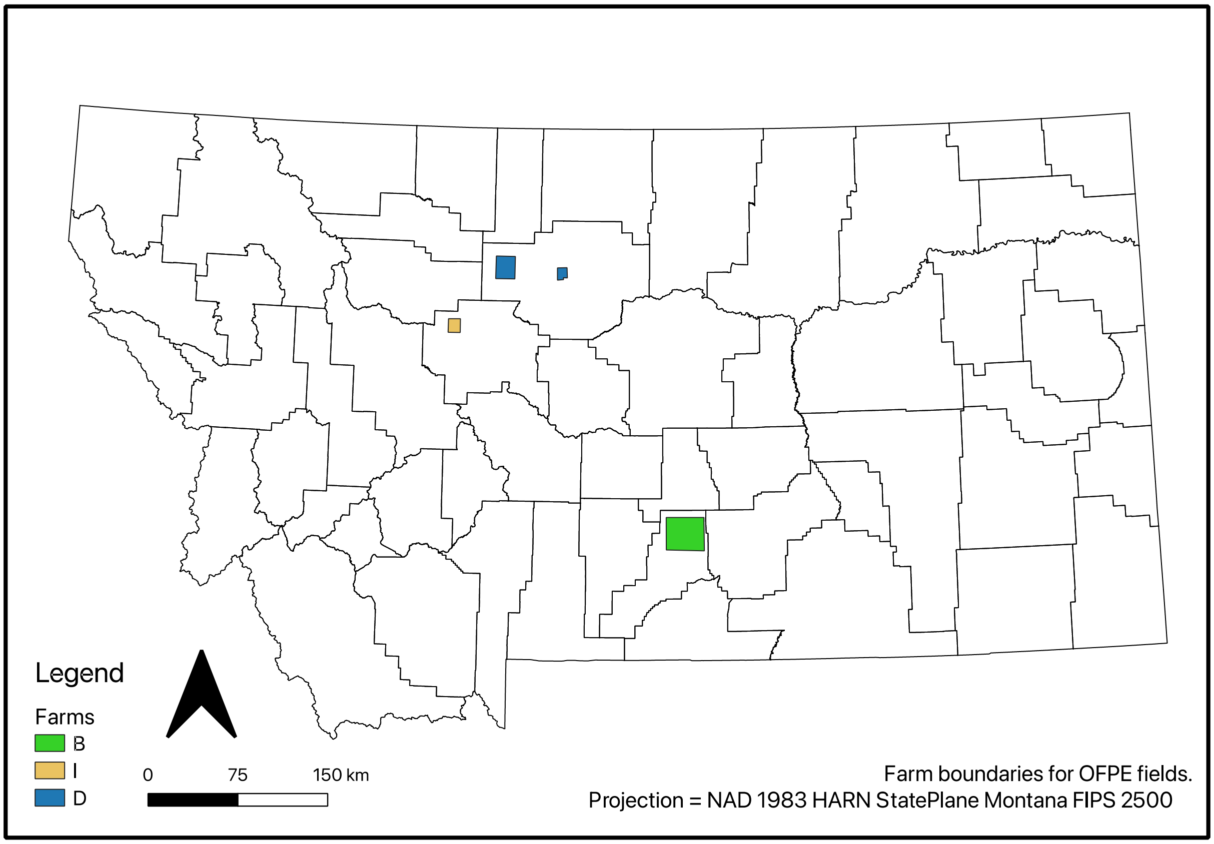


Figure. 4.2. Map of OFPE farm boundaries for the three selected farmer collaborators with Montana State University. Colors represent different farmers.

Table 4.1. Crop histories, field sizes, and years in VRA treatment for each field.

Farm	Field	Field size (ha)	**Crop History[1]:** *2014 / 2015 / 2016 / 2017 / 2018 / 2019 / 2020 / 2021*	Years N rate treatment
B	B1	79	SF / WW / CF / WW / CF / WW / CF / WW	2017, 2019, 2021
	B2	94	WW / CF / WW / CF / SF / WW / CF / WW	2016, 2019, 2021
	B3	64	SW / CF / WW / CF / WW / CF / WW / CF	2016, 2018, 2020
D	D1	46	CF / WW / CF / WW / CF / WW / CF / WW	2017, 2019, 2021
	D2	48	WW / SW / WW / CF / WW / CF / WW / CF	2016, 2018, 2020
	D3	20	SW / SW / WW / CF / WW / CF / WW / CF	2016, 2018, 2020
I	I1	94	SW / CF / WW / CF / WW / CF / WW / CF	2016, 2018, 2020

[1] WW = winter wheat, CF = chemical fallow, SW = spring wheat, SF = safflower

Table 4.2. Table of covariate data types gathered from Google Earth Engine to enrich the crop yield and protein datasets gathered from on-farms.

Data Type	Data Sources	Resolution	Years Collected	Description
Normalized Difference Vegetation Index (NDVI)	Landsat 5/7/8	30m	L5: 1999-2011 L7: 2012-2013 L8: 2014 - present	Landsat is an ongoing USGS and NASA collaboration. Bands (NIR, red) L5/L7: B4 and B3 L8: B5 and B4
Normalized Difference Water Index (NDWI)	Landsat 5/7/8	30m	L5: 1999-2011 L7: 2012-2013 L8: 2014 - present	Landsat is an ongoing USGS and NASA collaboration. Bands (NIR, red) L5/L7: B2 and B4 L8: B2 and B5
Elevation	USGS NED	~10m (1/3 arc second), ~23m (3/4 arc second)	1999-present	USGS National Elevation Dataset. Measured in meters.
Aspect	USGS NED	~10m (1/3 arc second), 30m	1999-present	Direction the surface faces, function of neighboring elevations, in radians. Also calculated for each E/W and N/S direction as cosine and sine.
Slope	USGS NED	~10m (1/3 arc second), 30m	1999-present	Rate of change of height from neighboring cells, in degrees. Measured in degrees.
Topographic Position Index (TPI)	USGS NED	~10m (1/3 arc second), 30m	1999-present	Measure of divots and low spots as a function of neighboring elevation.
Precipitation	DaymetV3	1km	1999-present	Estimates from the NASA Oak Ridge National Laboratory (ORNL). Measured in mm.
Growing Degree Days (GDD)	DaymetV3	1km	1999-present	Estimates from the NASA Oak Ridge National Laboratory (ORNL).
bulkdensity	OpenLandMap	250m	1999-present	Soil bulk density (fine earth) 10 x kg / m3 averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
claycontent	OpenLandMap	250m	1999-present	Clay content in % (kg / kg) averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
sandcontent	OpenLandMap	250m	1999-present	Sand content in % (kg / kg) averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
pH (phw)	OpenLandMap	250m	1999-present	Soil pH in H_2O averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
watercontent	OpenLandMap	250m	1999-present	Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted and averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
carboncontent	OpenLandMap	250m	1999-present	Soil organic carbon content in x 5 g / kg averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).

Split	Year(s)		
A	2016/2017	2018/2019	2020/2021
B	2016/2017	2018/2019	2020/2021
C	2016/2017	2018/2019	2020/2021

Train

Test

Figure 4.3. Design matrix for data configurations applied to each field for all 5 model types. Orange cells represent the years used for training in each case and blue cells represent the year used for testing in each case.

Table 4.3. Identifiers for covariates in the models below, the covariate's data type and a description of the covariate.

Covariate ID	Data Type	Description
$aspect_{cos}$	Aspect	Aspect in the E/W direction
$aspect_{sin}$	Aspect	Aspect in the N/S direction
slope	Slope	Slope in degrees
elev	Elevation	Elevation in meters
tpi	TPI	Topographic Position Index
$prec_{py}$	Precipitation	Precipitation summed from the water year two years prior to observed data (mm)
gdd_{py}	GDD	Growing degree days summed from the year prior
$ndvi_{py}$	NDVI	Maximum Normalized Difference Vegetation Index from across the prior year
$ndvi_{2py}$	NDVI	Maximum Normalized Difference Vegetation Index from across two years prior
$ndwi_{py}$	NDWI	Maximum Normalized Difference Water Index from across the prior year
$ndwi_{2py}$	NDWI	Maximum Normalized Difference Water Index from across two years prior
bulkdensity	Bulk Density	Soil bulk density (fine earth) 10 x kg / m3 averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
claycontent	Clay Content	Clay content in % (kg / kg) averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
phw	pH of Water	Soil pH in H_2O averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
watercontent	Water Content	Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted and averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
carboncontent	Carbon Content	Soil organic carbon content in x 5 g / kg averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).

Table 4.4. Results of the mean RMSE across the seven fields of yield and protein from each of the non-linear models tested with a randomly split training dataset with 70% of observations across three years for model fitting and a hold-out dataset with the other 30% of observations for predicting responses with new data and calculating RMSE.

Model	**Mean RMSE for Yield (kg ha^{-1})**	**Mean RMSE for Protein (%)**
Beta Function	**1193.133**	**1.6214**
Gompertz	1219.139	1.6741
Hyperbolic	1216.698	1.6771
Logistic 1	1205.729	1.6317
Logistic 2	1194.276	1.6236
Richards	1205.265	1.6256
Weibull 1	1199.535	1.6389
Weibull 2	1252.273	1.7610

Table 4.5. Economic parameters used for the calculation of net-return from Equation 4.6. Derived from USDA ERS and farmer survey information. Parameters vary across years but did not differ across fields for a given year.

Year	**Base Price (*Bp*)**	**Cost of N (*CA*)**	**Fixed Costs (*FC*)**
2016	\$0.15 kg^{-1}	\$0.75 kg^{-1}	\$403.13 ha^{-1}
2017	\$0.21 kg^{-1}	\$1.10 kg^{-1}	\$424.23 ha^{-1}
2018	\$0.19 kg^{-1}	\$0.71 kg^{-1}	\$436.51 ha^{-1}
2019	\$0.16 kg^{-1}	\$0.75 kg^{-1}	\$449.18 ha^{-1}
2020	\$0.16 kg^{-1}	\$0.62 kg^{-1}	\$439.67 ha^{-1}
2021	\$0.26 kg^{-1}	\$0.88 kg^{-1}	\$450.74 ha^{-1}

Table 4.6. The average RMSE of grain yield (kg ha^{-1}) across data splits from Fig. 4.3. Darker shading and bolded values indicate the model type with the lowest RMSE for each field, and intermediate shading represents the second most accurate model type for forecasting crop responses in each field.

Field	Beta Function	GAM	Bayes NLR	Bayes MLR	RF
B1	2440.9090	5015.0961	**2122.4399**	2138.5884	2412.0508
B2	2645.2881	2149.6369	1883.1525	2068.4823	**1797.6458**
B3	3382.3902	3700.7943	2660.5830	2606.1046	**1351.0698**
D1	1300.2283	1139.2400	1406.4796	1085.3439	**1023.4495**
D2	1464.8009	3726.0322	1856.8694	2004.0815	**1233.9528**
D3	1333.0457	2123.1780	1521.8785	1537.6187	**1187.4798**
I1	3146.7093	6568.3756	1837.1322	6727.2450	**1002.2410**

Table 4.7. The average RMSE of grain protein concentration (%) across data splits from Fig. 4.3. Darker shading and bolded values indicate the model type with the lowest RMSE for each field, and intermediate shading represents the second most accurate model type for forecasting crop responses in each field.

Field	Beta Function	GAM	Bayes NLR	Bayes MLR	RF
B1	**2.5005**	3.5369	6.4875	3.7920	3.2084
B2	3.0625	3.7834	7.8040	3.1288	**2.7156**
B3	3.8652	**1.7155**	4.4366	4.8908	1.7377
D1	1.5658	2.0148	1.7419	2.1491	**1.4782**
D2	3.5321	4.2154	4.5993	7.3038	**1.3496**
D3	1.9023	**1.3381**	3.7526	2.0791	1.7089
I1	3.7393	11.2870	7.4033	9.5703	**3.1114**

Table 4.8. Mean RMSE of grain yield (kg ha^{-1}) across fields for each model type, where the given model type was fit for each field using the split case scenario in Fig. 4.3. The mean RMSE across fields for each model and split case were divided by the crop-fallow data schedule to assess differences between a 2016, 2018, and 2020 schedule and a 2017, 2019, and 2021 schedule. Shaded and bolded cells indicate the split data case (A, B, or C) that had the lowest mean RMSE across fields for each model type.

Model	***Data Schedule 2016-2020***			***Data Schedule 2017-2021***		
	A	*B*	*C*	*A*	*B*	*C*
Beta Function	1524.7320	**1324.2424**	4146.2352	2462.2306	2605.4652	**1318.7295**
GAM	7987.6525	2533.4528	**1567.6798**	**1386.2126**	3788.2681	3129.4923
Bayes NLR	2073.4396	2182.0752	**1651.8325**	1789.5136	1876.7739	**1745.7845**
Bayes MLR	2869.1192	5036.1426	**1751.0255**	**1285.1462**	2220.8420	1786.4264
RF	1205.8547	1418.3604	**956.8424**	1635.3190	2379.0422	**1218.7850**

Table 4.9. Mean RMSE of grain protein concentration (%) across fields for each model type, where the given model type was fit for each field using the split case scenario in Fig. 4.3. The mean RMSE across fields for each model and split case were divided by the crop-fallow data schedule to assess differences between a 2016, 2018, and 2020 schedule and a 2017, 2019, and 2021 schedule. Shaded and bolded cells indicate the split data case (A, B, or C) that had the lowest mean RMSE across fields for each model type.

Model	***Data Schedule 2016-2020***			***Data Schedule 2017-2021***		
	A	*B*	*C*	*A*	*B*	*C*
Beta Function	5.9346	**1.4285**	2.4161	2.6705	2.2597	**2.1985**
GAM	6.85385	**3.482275**	3.580825	3.3854	3.7547	**2.1951**
Bayes NLR	6.1162	**4.2991**	4.7285	5.9406	6.1742	**3.9186**
Bayes MLR	6.8080	6.4687	**4.6063**	3.0865	3.7888	**2.1946**
RF	1.9789	**1.7655**	2.1863	2.4966	**2.1858**	2.7199

Table 4.10. The average RMSE of net-return ($ ha^{-1}) across data splits from Fig. 4.3. Darker shading and bolded values indicate the model type with the lowest RMSE for each field, and intermediate shading represents the second most accurate model type for forecasting crop responses in each field.

Field	Beta Function	GAM	Bayes NLR	Bayes MLR	RF
B1	**1551.8917**	1787.8861	3509.5598	1785.5684	1585.8252
B2	1102.0123	1354.9238	3672.1688	1399.2662	**816.4497**
B3	1391.4319	1450.8455	2032.7059	3283.4635	**545.1199**
D1	439.4707	443.2503	456.4161	430.3751	**340.1714**
D2	3058.1792	2814.7075	1884.1768	3501.2414	**561.1614**
D3	770.0615	**644.0324**	1604.6263	950.4935	731.6449
I1	**1193.0592**	5667.1530	3414.2282	2478.4261	1307.3384

Table 4.11. Mean RMSE of net-return (\$ ha^{-1}) across fields for each model type, where the given model type was fit for each field using the split case scenario in Fig. 4.3. The mean RMSE across fields for each model and split case were divided by the crop-fallow data schedule to assess differences between a 2016, 2018, and 2020 schedule and a 2017, 2019, and 2021 schedule. Shaded and bolded cells indicate the split data case (A, B, or C) that had the lowest mean RMSE across fields for each model type.

Model	***Data Schedule 2016-2020***			***Data Schedule 2017-2021***		
	A	*B*	*C*	*A*	*B*	*C*
Beta Function	3438.3808	**632.6774**	738.4906	943.4489	**913.6267**	1236.2992
GAM	5095.3797	**1192.5409**	1644.6331	1180.5406	**1053.4916**	1352.0279
Bayes NLR	**1716.6650**	2425.1159	2560.0221	2340.6239	3249.3770	**2048.1439**
Bayes MLR	4117.9830	2125.0722	**1417.1633**	**1065.6500**	1440.2512	1109.3086
RF	935.3906	728.9077	**694.6501**	**676.4806**	915.6078	1150.3580

Table 4.12. The total precipitation from the AgriMet station setup by the Bureau of Land Reclamation at the Central Ag Research Station in Conrad, MT.

Year	**Total Precipitation (mm)**	**Mean Precipitation (mm)**
2016	339.7	286.8
2017	233.9	
2018	473.7	496.65
2019	519.6	
2020	292.6	274.05
2021	255.5	

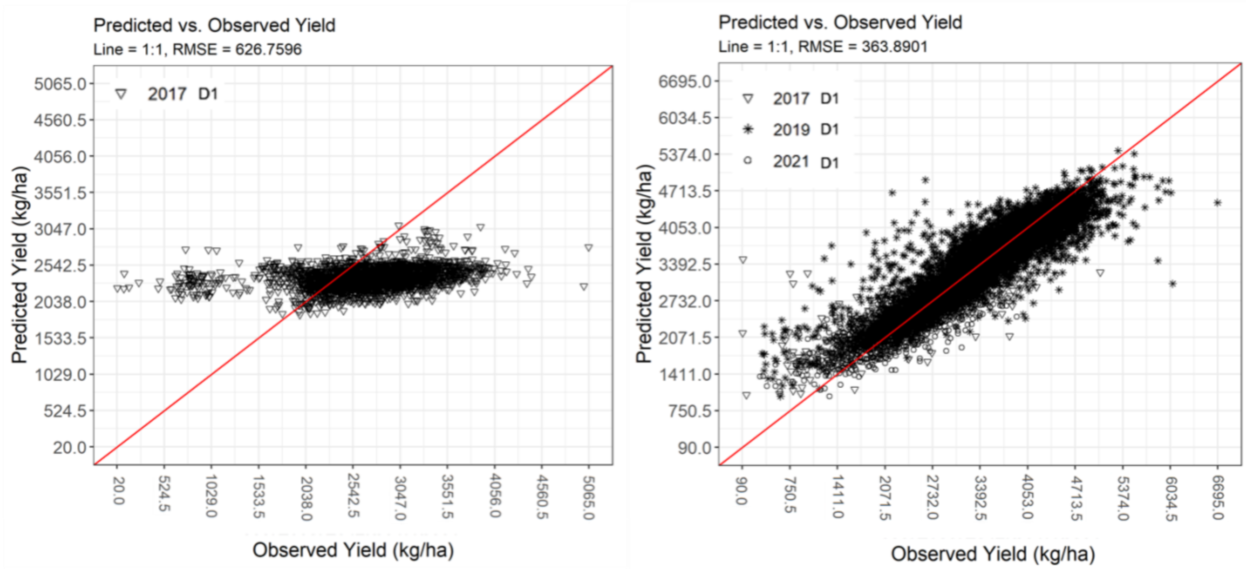


Fig. 4.4. Predicted vs. observed grain yields (kg ha^{-1}) from field D1 using the RF regression model where 2019 and 2021 data were used for training and tested in 2017 (left) and training and test datasets were derived from all three years (right). RMSE reported in kg ha^{-1}.

REFERENCES CITED

Anselin, L., Bongiovanni, R., & Lowenberg-DeBoer, J. (2004). A Spatial Econometric Approach To The Economics of Site-Specific Nitrogen Management. *Amer. J. Agr. Econ.*, *86*(August), 675–687.

Basso, B., & Antle, J. (2020). Digital agriculture to design sustainable agricultural systems. *Nature Sustainability*, *3*(4), 254–256. https://doi.org/10.1038/s41893-020-0510-0

Basso, B., Ritchie, J. T., Cammarano, D., & Sartori, L. (2011). A strategic and tactical management approach to select optimal N fertilizer rates for wheat in a spatially variable field. *European Journal of Agronomy*, *35*(4), 215–222. https://doi.org/10.1016/j.eja.2011.06.004

Besag, J., & Higdon, D. (1999). Bayesian analysis of agricultural field experiments. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *61*(4), 691–746. https://doi.org/10.1111/1467-9868.00201

Blackmore, S., & Moore, M. (1999). Remedial Correction of Yield Map Data. *Precision Agriculture*, *1*(1), 53–66. https://doi.org/10.1023/A:1009969601387

Bolton, D. K., & Friedl, M. A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*, *173*, 74–84. https://doi.org/10.1016/j.agrformet.2013.01.007

Bürkner, P.C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80*(1), 1-28. doi:10.18637/jss.v080.i01

Bürkner, P.C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. T*he R Journal, 10*(1), 395-411. doi:10.32614/RJ-2018-017

Bürkner, P.C. (2021). Bayesian Item Response Modeling in R with brms and Stan. J*ournal of Statistical Software, 100*(5), 1-54. doi:10.18637/jss.v100.i05

Carolan, M. (2017). Publicising Food: Big Data, Precision Agriculture, and Co-Experimental

Techniques of Addition. *Sociologia Ruralis*, *57*(2), 135–154. https://doi.org/10.1111/soru.12120

Chen, K., O'Leary, R. A., & Evans, F. H. (2019). A simple and parsimonious generalised additive model for predicting wheat yield in a decision support tool. *Agricultural Systems*, *173*(December 2018), 140–150. https://doi.org/10.1016/j.agsy.2019.02.009

Clancy, M. (2019). Real-time protein monitoring. *Next Instruments*, (June).

De Bastiani, F., Mariz de Aquino Cysneiros, A. H., Uribe-Opazo, M. A., & Galea, M. (2015). Influence diagnostics in elliptical spatial linear models. *Test*, *24*(2), 322–340. https://doi.org/10.1007/s11749-014-0409-z

DeLonge, M., Robbins, T., Basche, A., & Haynes-Mawlow, L. (2020). The State of Sustainable Agriculture and Agroecology Research and Impacts: A Survey of U.S. Scientists. *Journal of Agriculture, Food Systems, and Community Development*, *9*(2), 1–26. https://doi.org/10.5304/jafscd.2020.092.009

Diggle, P. J., & Tawn, J. A. (1998). Model-based geostatistics. *Applied Statistics*, *47*(3), 299–350.

Drachal, K. (2019). Analysis of agricultural commodities prices with new Bayesian model combination schemes. *Sustainability (Switzerland)*, *11*(19). https://doi.org/10.3390/su11195305

Estes, L. D., Bradley, B. A., Beukes, H., Hole, D. G., Lau, M., Oppenheimer, M. G., … Turner, W. R. (2013). Comparing mechanistic and empirical model projections of crop suitability and productivity: Implications for ecological forecasting. *Global Ecology and Biogeography*, *22*(8), 1007–1018. https://doi.org/10.1111/geb.12034

Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development*, *36*(2). https://doi.org/10.1007/s13593-016-0364-z

Fritsch, M., Lischke, H., & Meyer, K. M. (2020). Scaling methods in ecological modelling.

Methods in Ecology and Evolution, *11*(11), 1368–1378. https://doi.org/10.1111/2041-210X.13466

Gardner, G., Mieno, T., & Bullock, D. S. (2021). An economic evaluation of site-specific input application Rx maps: evaluation framework and case study. *Precision Agriculture*, *22*(4), 1304–1316. https://doi.org/10.1007/s11119-021-09785-z

Gebbers, R., & Adamchuck, V. I. (2010). Precision Agriculture and Food Security. *Science*, (February), 828–830.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534. https://doi.org/10.1214/06-BA117A

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., … Kalogirou, S. (2021). Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, *36*(2), 121–136. https://doi.org/10.1080/10106049.2019.1595177

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gotway, C. A., & Stroup, W. W. (1997). A Generalized Linear Model Approach to Spatial Data Analysis and Prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, *2*(2), 157–178.

Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling*, *157*(2–3), 89–100. https://doi.org/10.1016/S0304-3800(02)00204-1

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, *1*(3), 297–310. https://doi.org/10.1214/ss/1177013604

Hatfield, J. L., Cryder, M., & Basso, B. (2020). Remote Sensing: Advancing the Science and the

Applications to Transform Agriculture. *IT Professional*, *22*(3), 42–45. https://doi.org/10.1109/MITP.2020.2986102

Hegedus, P.B. (2020). OFPE: On-Farm Precision Experiments (OFPE) Data Management and Analysis Tools. R package version 1.8.11. https://github.com/paulhegedus/OFPE.git

Hegedus, P.B & Maxwell, B.D. (2022). Rationale for field specific on-farm precision experimentation. *Agriculture, Ecosystems & Environment.* Manuscript submitted for publication, in review.

Holland, D. M., De Oliveira, V., Cox, L. H., & Smith, R. L. (2000). Estimation of regional trends in sulfur dioxide over the eastern United States. *Environmetrics*, *11*(4), 373–393. https://doi.org/10.1002/1099-095X(200007/08)11:4<373::AID-ENV419>3.0.CO;2-2

Hooten, M. B., Ver Hoef, J. M., & Hanks, E. M. (2019). Simultaneous Autoregressive Model (SAR). *Wiley StatsRef: Statistics Reference Online*, 1–10. https://doi.org/10.1007/978-0-387-35973-1_1217

Houlahan, J. E., McKinney, S. T., Anderson, T. M., & McGill, B. J. (2017). The priority of prediction in ecological understanding. *Oikos*, *126*(1), 1–7. https://doi.org/10.1111/oik.03726

Huang, X., Huang, G., Yu, C., Ni, S., & Yu, L. (2017). A multiple crop model ensemble for improving broad-scale yield prediction using Bayesian model averaging. *Field Crops Research*, *211*(June), 114–124. https://doi.org/10.1016/j.fcr.2017.06.011

Janatian, N., Sadeghi, M., Sanaeinejad, S. H., Bakhshian, E., Farid, A., Hasheminia, S. M., & Ghazanfari, S. (2017). A statistical framework for estimating air temperature using MODIS land surface temperature data. *International Journal of Climatology*, *37*(3), 1181–1194. https://doi.org/10.1002/joc.4766

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., … Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *PLoS ONE*, *11*(6), 1–15. https://doi.org/10.1371/journal.pone.0156571

Jiang, P., He, Z., Kitchen, N. R., & Sudduth, K. A. (2009). Bayesian analysis of within-field variability of corn yield using a spatial hierarchical model. *Precision Agriculture*, *10*(2), 111–127. https://doi.org/10.1007/s11119-008-9070-4

Jing, W., Yang, Y., Yue, X., & Zhao, X. (2016). A comparison of different regression algorithms for downscaling monthly satellite-based precipitation over North China. *Remote Sensing*, *8*(10), 1–17. https://doi.org/10.3390/rs8100835

Johnson, D. M. (2014). An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment*, *141*, 116–128. https://doi.org/10.1016/j.rse.2013.10.027

Joshi, V. R., Kazula, M. J., Coulter, J. A., Naeve, S. L., & Garcia y Garcia, A. (2021). In-season weather data provide reliable yield estimates of maize and soybean in the US central Corn Belt. *International Journal of Biometeorology*, *65*(4), 489–502. https://doi.org/10.1007/s00484-020-02039-z

Kissling, W. D., & Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, *17*(1), 59–71. https://doi.org/10.1111/j.1466-8238.2007.00334.x

Langella, G., Basile, A., Bonfante, A., & Terribile, F. (2010). High-resolution space-time rainfall analysis using integrated ANN inference systems. *Journal of Hydrology*, *387*(3–4), 328–342. https://doi.org/10.1016/j.jhydrol.2010.04.027

de Lara, A., Mieno, T., Luck, J., & Puntel, L. (2022). Site-specific economic optimum nitrogen rate using machine learning and on-farm experimentation. *Precision Agriculture.* Manuscript submitted for publication, in review.

Lawes, R. A., Oliver, Y. M., & Huth, N. I. (2019). Optimal nitrogen rate can be predicted using average yield and estimates of soil water and leaf nitrogen with infield experimentation. *Agronomy Journal*, *111*(3), 1155–1164. https://doi.org/10.2134/agronj2018.09.0607

Lawrence, P. G., Rew, L. J., & Maxwell, B. D. (2015). A probabilistic Bayesian framework for progressively updating site-specific recommendations. *Precision Agriculture*, *16*(3), 275–296. https://doi.org/10.1007/s11119-014-9375-4

Li, J., Heap, A. D., Potter, A., & Daniell, J. J. (2011). Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling and Software*, *26*(12), 1647–1659. https://doi.org/10.1016/j.envsoft.2011.07.004

Liew, J., & Forkman, J. (2015). A guide to generalized additive models in crop science using SAS and R. *Communications in Biometry and Crop Science*, *10*(1), 41–57.

Mardia, A. K. V, & Marshall, R. J. (1984). Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression. *Biometrika*, *71*(1), 135–146.

Mariano, C., & Mónica, B. (2021). A random forest-based algorithm for data-intensive spatial interpolation in crop yield mapping. *Computers and Electronics in Agriculture*, *184*(March). https://doi.org/10.1016/j.compag.2021.106094

Marques Ramos, A. P., Prado Osco, L., Elis Garcia Furuya, D., Nunes Gonçalves, W., Cordeiro Santana, D., Pereira Ribeiro Teodoro, L., … Pistori, H. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, *178*(July). https://doi.org/10.1016/j.compag.2020.105791

McBratney, A. B., Minasny, B., & Whelan, B. M. (2005). Obtaining Useful High-resolution Soil Data from Proximally - Sensed Electrical Conductivity / Resistivity (PSEC / R) Surveys. *Precision Agriculture 2005, ECPA 2005*.

McBratney, A., Whelan, B., Ancev, T., & Bouma, J. (2005). Future directions of precision agriculture. *Precision Agriculture*, *6*(1), 7–23. https://doi.org/10.1007/s11119-005-0681-8

McFadden, B. R., Brorsen, B. W., & Raun, W. R. (2017). Nitrogen fertilizer recommendations based on plant sensing and Bayesian updating. *Precision Agriculture*, *19*(1), 1–14. https://doi.org/10.1007/s11119-017-9499-4

McFadden, J. R., Rosburg, A., & Njuki, E. (2021). Information Inputs and Technical Efficiency in Midwest Corn Production: Evidence from Farmers' Use of Yield and Soil Maps. *American Journal of Agricultural Economics*, 1–24. https://doi.org/10.1111/ajae.12251

Mueller, N. D., West, P. C., Gerber, J. S., Macdonald, G. K., Polasky, S., & Foley, J. A. (2014). A tradeoff frontier for global nitrogen use and cereal production. *Environmental Research Letters*, *9*(5). https://doi.org/10.1088/1748-9326/9/5/054002

Nandram, B., Berg, E., & Barboza, W. (2014). A hierarchical Bayesian model for forecasting state-level corn yield. *Environmental and Ecological Statistics*, *21*(3), 507–530. https://doi.org/10.1007/s10651-013-0266-z

Paccioretti, P., Bruno, C., Gianinni Kurina, F., Córdoba, M., Bullock, D. S., & Balzarini, M. (2021). Statistical models of yield in on-farm precision experimentation. *Agronomy Journal*, *113*(6), 4916–4929. https://doi.org/10.1002/agj2.20833

Peerlinck, A., Sheppard, J., & Maxwell, B. (2018). Using Deep Learning in Yield and Protein Prediction of Winter Wheat Based on Fertilization Prescriptions in Precision Agriculture. *14th International Conference on Precision Agriculture*, 1–13.

Peerlinck, A., Sheppard, J., & Senecal, J. (2019). AdaBoost with Neural Networks for Yield and Protein Prediction in Precision Agriculture. *Proceedings of the International Joint Conference on Neural Networks*, *2019-July*. https://doi.org/10.1109/IJCNN.2019.8851976

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2021). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-153, https://www.CRAN.R-project.org/package=nlme

Pinilla, J., & Negrín, M. (2021). Non-parametric generalized additive models as a tool for evaluating policy interventions. *Mathematics*, *9*(4), 1–12. https://doi.org/10.3390/math9040299

Prost, L., Makowski, D., & Jeuffroy, M. H. (2008). Comparison of stepwise selection and Bayesian model averaging for yield gap analysis. *Ecological Modelling*, *219*(1–2), 66–76. https://doi.org/10.1016/j.ecolmodel.2008.07.026

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramsey, A. F. (2020). Probability Distributions of Crop Yields: A Bayesian Spatial Quantile Regression Approach. *American Journal of Agricultural Economics*, *102*(1), 220–239. https://doi.org/10.1093/ajae/aaz029

Reynolds, W. D., Drury, C. F., Phillips, L. A., Yang, X., & Agomoh, I. V. (2021). An adapted weibull function for agricultural applications. *Canadian Journal of Soil Science*, *101*(4), 680–702. https://doi.org/10.1139/cjss-2021-0046

Schimmelpfennig, D., & Lowenberg-DeBoer, J. (2020). Farm Types and Precision Agriculture Adoption: Crops, Regions, Soil Variability, and Farm Size. *SSRN Electronic Journal*, 1–38. https://doi.org/10.2139/ssrn.3689311

Sela, S., van Es, H. M., Moebius-Clune, B. N., Marjerison, R., Melkonian, J., Moebius-Clune, D., … Gomes, S. (2016). Adapt-N outperforms grower-selected nitrogen rates in northeast and midwestern united states strip trials. *Agronomy Journal*, *108*(4), 1726–1734. https://doi.org/10.2134/agronj2015.0606

Shi, Y., Song, L., Xia, Z., Lin, Y., Myneni, R. B., Choi, S., … Yang, F. (2015). Mapping annual precipitation across Mainland China in the period 2001-2010 from TRMM3B43 product using spatial downscaling approach. *Remote Sensing*, *7*(5), 5849–5878. https://doi.org/10.3390/rs70505849

Sudduth, K. A., & Drummond, S. T. (2007). Yield editor: Software for removing errors from crop yield maps. *Agronomy Journal*, *99*(6), 1471–1482. https://doi.org/10.2134/agronj2006.0326

Sudduth, K. A., Drummond, S. T., & Myers, D. B. (2012). Yield editor 2.0: Software for automated removal of yield map errors. *American Society of Agricultural and Biological Engineers Annual International Meeting 2012, ASABE 2012*, *4*(12), 3378–3391. https://doi.org/10.13031/2013.41893

Thöle, H., Richter, C., & Ehlert, D. (2013). Strategy of statistical model selection for precision farming on-farm experiments. *Precision Agriculture*, *14*(4), 434–449. https://doi.org/10.1007/s11119-013-9306-9

Ver Hoef, J. M., Hanks, E. M., & Hooten, M. B. (2017). On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. *Spatial Statistics*, *25*, 68–85. https://doi.org/10.1016/j.spasta.2018.04.006

Ver Hoef, J. M., Petersen, E. E., Hooten, M. B., Hanks, E. M., & Fortin, M.-J. (2018). Spatial autoregressive models for statistical inference from ecological data. *Ecological Monographs*, *88*(1), 36–59.

Wajid, A., Hussain, K., Ilyas, A., Habib-Ur-rahman, M., Shakil, Q., & Hoogenboom, G. (2021). Crop Models: Important Tools in Decision Support System to Manage Wheat Production under Vulnerable Environments. *Agriculture (Switzerland)*, *11*(11). https://doi.org/10.3390/AGRICULTURE11111166

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, *121*(2), 311–324. https://doi.org/10.1016/S0378-3758(03)00111-3

Walsh, E. S., Kreakie, B. J., Cantwell, M. G., & Nacci, D. (2017). A Random Forest approach to predict the spatial distribution of sediment pollution in an estuarine system. *PLoS ONE*, *12*(7), 1–18. https://doi.org/10.1371/journal.pone.0179473

Wang, J. C., Holan, S. H., Nandram, B., Barboza, W., Toto, C., & Anderson, E. (2012). A Bayesian Approach to Estimating Agricultural Yield Based on Multiple Repeated Surveys. *Journal of Agricultural, Biological, and Environmental Statistics*, *17*(1), 84–106. https://doi.org/10.1007/s13253-011-0067-5

Wang, L., Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Crop Journal*, *4*(3), 212–219. https://doi.org/10.1016/j.cj.2016.01.008

Wang, Y., Wu, G., Deng, L., Tang, Z., Wang, K., Sun, W., & Shangguan, Z. (2017). Prediction of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-07197-6

West, P. C., Gerber, J. S., Engstrom, P. M., Mueller, N. D., Brauman, K. A., Carlson, K. M., … Siebert, S. (2014). Leverage points for improving global food security and the environment.

Food Security, *345*(6194), 1–5.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *65*(1), 95–114. https://doi.org/10.1111/1467-9868.00374

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, *111*(516), 1548–1563. https://doi.org/10.1080/01621459.2016.1180986

Wright, M.N. & A. Ziegler (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 77(1), 1-17. doi:10.18637/jss.v077.i01

Xia, H., Ding, Y., & Wang, J. (2008). Gaussian process method for form error assessment using coordinate measurements. *IIE Transactions (Institute of Industrial Engineers)*, *40*(10), 931–946. https://doi.org/10.1080/07408170801971502

Yin, X., Goudriaan, J., Lantinga, E. A., Vos, J., & Spiertz, H. J. (2003). A flexible sigmoid function of determinate growth. *Annals of Botany*, *91*(3), 361–371. https://doi.org/10.1093/aob/mcg029

Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture - a worldwide overview. *Computers and Electronics in Agricultre*, *36*, 113–132. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S002216941401066X

Zuur, A. F., & Camphuysen, Kees, C. (2012). 6 Generalized Additive Models applied on northern gannets. In *A Beginners Guide to Generalized Additive Models with R* (p. 15). Highland Statistics Ltd.

CHAPTER FIVE

BEST USE OF MODERN DATA FOR FIELD-SPECIFIC DECISION SUPPORT

Contribution of Authors and Co-Authors

Manuscript in Chapter Five

Author: Paul B. Hegedus

Contributions: Conceived the study design and methodology, compiled data, executed statistical analysis, and wrote initial manuscript.

Co-Author: Bruce D. Maxwell

Contributions: Assisted with study design, analysis, and manuscript preparation at all stages.

Manuscript Information

Paul B. Hegedus, Bruce D. Maxwell

Precision Agriculture

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
__x__ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Published by Springer

Abstract

Many agronomic input decisions are based on data from previous production years. Data from the gap in time between a previous harvest and the point that an input (e.g. fertilizer) decision is made may improve crop response predictions. This research was designed to study the potential value of using the extra data in decision making, by comparing the uncertainty in predictions from agronomic models derived solely from data from previous production years (PY) versus data collected until immediately before the need for top-dress nitrogen fertilizer application decisions (DP). As there is additional uncertainty in the appropriate functional form of agronomic models related to crop responses to fertilizer, this research utilized two models: generalized additive models (GAMs) and random forest regression (RF). The ultimate objective was to evaluate differences in management recommendations and ability to accurately predict rain-fed winter wheat responses to nitrogen fertilizer rates under the two data constraints (PY vs. DP). To evaluate differences in management recommendations, a wet and dry year were selected from the historical record in which site-specific (precision) rate recommendations for nitrogen fertilizer were simulated. The accuracy of predictions with the DP data constraint was observed in 5 out of 28 cases across both model types. Functional form of the model used to predict responses to variable N rate fertilizer played a more important role in characterizing uncertainty surrounding recommendations than did the PY or DP data constraint. Inclusion of more recent data only benefitted model accuracy, thus, given the potential value of more recent data in decision making and its inclusion in decision support systems, practitioners should consider using models fit with all available data up to the point a crop manager must make decisions.

Introduction

Agricultural innovation is stuck in a paradox between finding ways to increase food production for a growing population while doing so with more efficient and sustainable practices. General productivity has increased from the Green Revolution to present, but the rate of yield increase has slowed since 1985, meaning agriculture will need to increase productivity while minimizing chemical input demands that threaten the environmental resources upon which sustained agronomic production relies (Foley et al., 2011; Tilman et al., 2011). Precision agriculture technologies provide extensive data collected from machines applying inputs and harvesting crops as well as remote sensing and weather data that can be automatically scraped from the internet. Greater adoption of decision support systems using data returned from day-to-day farming operations and public internet resources can increase input use efficiency while increasing or maintaining yields, which ultimately increases the ecological and economic resilience of agroecosystems (Bucci et al., 2018; Duff et al., 2022; McBratney et al., 2005).

Reliable predictions of the relationship between crop yields, crop quality, and input rates result in more informed input management and provide the foundation for maximizing farmer net-returns. A wide array of research has suggested that site-specific input management may reduce excess input application by ensuring application rates are justified by increased net financial returns for farmers (Biermacher et al., 2009; Ewing & Runck, 2015; Flowers et al., 2004; Khosla et al., 2002; Koch et al., 2004; Lawrence et al., 2015b; Rütting et al., 2018; Weersink et al., 2018). The major benefits of site-specific management are increasing the field-specific knowledge necessary to transfer applied inputs from low profit potential areas into high profit potential areas, and reducing expenditures by producers for costly inputs (Boyer et al., 2013; Córdova et al., 2018; Hurley et al., 2004, 2005; Khosla et al., 2008; Koch et al., 2004).

Decision support for site-specific inputs is typically based on empirical models that predict crop responses such as crop yield, or quality, and covariate data collected from sensors and satellites. These models are commonly parameterized based on fits to data across experiments and/or farms that are pooled to generate a generalized model for informing field-specific management. However, crop responses vary across space and time, within and between fields, even those fields that share a border (Hegedus & Maxwell, 2022a). This variation also influences differences in the most appropriate functional form of crop response models for a given crop (Hegedus & Maxwell, 2022b). Thus, field-specific management should logically be informed by the now available field specific data (Cook et al., 2004; Hegedus & Maxwell, 2022a; Lacoste et al., 2022).

Field-specific experiments have potential to locally optimize input rates and to reveal the role of field-specific variables influencing crop production and farmer net-returns (Luschei et al., 2001; Hegedus & Maxwell, 2022a). Uncertainty associated with crop response variability over time becomes the primary concern for using empirical models to convert site-specific information into decision recommendations. A common approach for developing empirical predictive models is to calibrate a model with data from previous years, quantify the quality of that calibrated model by comparing its predictions with observations gathered in subsequent years, and to retrain and amend the model as necessary (Hair, 2007). This process is then repeated over time, with each passing year providing a new dataset to test and improve a model. For example, a range of crop responses in each field over different weather and economic conditions is required to predict the most profitable input level before the next year, assuming the next year is average for these variables and the past years sampled represent the range of possible conditions. The backlog of conditions and responses can be used to simulate the probability of an outcome given a particular treatment if the past responses and economic variability (i.e., training data) are applied to making a current year decision (i.e., validation data). Simulation models have determined that high degrees of certainty in probabilities of net return can be achieved by conducting experiments on a field with a given crop for about 6-8 years (Lawrence et al., 2015). Future anomalous weather conditions due to climate change may introduce further uncertainty in decision support recommendations based on any assumptions of stationarity in this empirical approach. However, incorporating strategic sampling of past years in which to simulate projected climate changes increases the power of decision support systems by allowing farmers to examine outcomes in years of the past that may represent anomalous future conditions of an upcoming year.

Technology has given producers data sources not conventionally used in agronomic decision making (given the recency of the agricultural data revolution) that allow the application of data gathered immediately before the crop manager decides on the amount of input to buy and apply (decision point). Simulating management outcomes in a decision support system requires empirical predictive models that utilize weather data, however one element of uncertainty for constructing field-specific models is determining what data to include to make the best prediction of crop response. More specifically, uncertainty surrounds the question of what portion of past weather and crop reflectance data to include in analysis, e.g., whether covariate data is constrained to only past years, includes data from past years plus weather data into the growing season up to the decision point, or whether covariate data from further into the growing season is included. Many studies have been conducted to predict crop responses, such as yield, using in-season covariate data. However, estimates of future yield from within-season measurements after a decision point are not appropriate for use in decision support systems (Croft et al., 2020; Gaso et al., 2019; Gonzalez-Sanchez et al., 2014; Stepanov et al., 2020).

In a rain-fed winter wheat (*Triticum aestivum*) production system, response models of grain yield and protein concentration to nitrogen (N) fertilizer are hypothesized to predict crop responses more accurately when the cumulative influence of weather and early-season growth prior to a decision point in mid-growing season are included in the model, compared to just using information from previous years (Fig. 5.1). A rainfed crop would logically be influenced by weather conditions during the growing season as well as previous seasons, thereby providing increased predictive power if included in the crop response models. Mid-growing season application of N fertilizer is thought to boost yields, but especially increase grain protein concentrations that can receive a price premium (Jaynes & Colvin, 2006; Jones, 2013; Qianqian et al., 2021; Zeng et al., 2012).

This study attempts to address a critical component of the agronomic data revolution by addressing how constraints on data used to fit models of crop production and quality influence the ability of a model to support effective application decisions. Multiple models for characterizing crop responses to N fertilizer and covariate data were used in recognition of the uncertainty surrounding the appropriate functional form of crop response models between fields (Hegedus & Maxwell, 2022b). Two models were selected, based on prior investigations into the adequacy of modeling crop responses: a generalized additive model (GAM) and a random forest regression (RF) model. Both models were fit to characterize winter wheat yield and grain protein concentration response to variable N fertilizer rates with climate, environmental, and vegetation index covariates to assess the influence of data constraints on predictions of crop responses. Model performance when data were constrained to previous calendar years was compared to performance with data collected up to a March 30th (current growing season) decision point for top-dress N fertilizer (Fig. 5.1). Assessing performance between the same type of model fit under the two data constraints was the main objective of this paper, where performance was defined as the ability of the model to predict responses in test datasets and measured by root mean square error (RMSE). In addition, assessments of the crop model performance were repeated under simulated wet and dry weather conditions to assess how nitrogen application recommendations are influenced by interannual uncertainty in weather.

Methods

The objectives were evaluated using seven dryland fields from three farms distributed across Montana that had at least three years of on-field N fertilizer experimentation (Fig. 5.2, Table 5.1). The response variables of interest were crop productivity (yield in kg ha^{-1}) and quality (grain protein concentration), both of which are gathered from monitors mounted on farmers' combine harvesters. Data from yield monitors were gathered on average every three seconds. All yield monitors were calibrated every spring by the farmers, according to their respective manufacturing instructions. Grain protein concentration (%) was measured with a CropScan 3000H near infrared monitor (Clancy, 2019). Beyond data collected from the machines on the field, open-source covariate data relating to the environment were gathered (Table 5.2). These data were obtained or derived from Google Earth Engine (Gorelick et al., 2017). Note that while elevation, aspect, slope, TPI, and OpenLandMap data were collected every year, they did not vary from year to year. Temporal covariate data included Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), precipitation, and growing degree days (GDD). Yield and protein datasets were created by aggregating the response variable data, as-applied N, and remote sensing data to the centroids of a 10m grid laid across each field.

All temporal covariates from Google Earth Engine were collected under two constraints, 1) the first, hereafter labeled PY, included all previous years of information, and 2) hereafter labeled DP, included all previous years plus the current year to the mid growing season when the decision point for top-dress fertilizer rates is determined for winter wheat. The PY data were collected up to December 31st of the year prior to upcoming harvest for vegetation index and GDD data and October 31st of the year prior to upcoming harvest for precipitation to follow the water year. Data collected to the decision point included the extra three months from January 1st to March 30th of the harvest year for vegetation index and GDD data and the extra five months from November 1st of the previous year to March 30th of the harvest year for precipitation (Fig. 1). A rainfed crop would logically be influenced by weather conditions (precipitation and temperature) during the winter portion of the growing season and therefore provide increased predictive power if included in the crop response models

The first model type fit to the yield and protein datasets was a GAM. Generalized additive models have been used in the literature for estimating crop responses to variable rates of N fertilizer and were chosen in our case because of their flexibility and ability to characterize the response of observed data (Chen et al., 2019; Estes et al., 2013; Joshi et al., 2021; Liew &

Forkman, 2015). A gamma family distribution was used due to the realistic constraint that yield and protein of a crop cannot be negative. To account for non-constant variance found in initial model fits, a log link function was used. Thin plate shrinkage splines were used for all variables to allow the estimated degrees of freedom of parameters to shrink to zero, combining the process of model fitting and selection. To account for spatial autocorrelation, coordinates were included as smoothing terms with a Gaussian process basis function (Gotway & Stroup, 1997; Guisan et al., 2002; Holland et al., 2000; Zuur et al., 2012). The exception to using thin plate shrinkage splines was the use of a Gaussian process for the interaction of coordinates accounting for spatial autocorrelation between observations within the field. While time is not explicitly included in the model, the singularities within years for predictors that vary across time, such as precipitation and growing degree days, mean these variables serve as proxies for the effect of time. Models were fit using the *mgcv* package in R (Wood, 2003; Wood et al., 2016).

The second model used was a RF where the number of trees and the number of covariates sampled at each node were optimized during the fitting process. To account for spatial autocorrelation, the coordinates were included as covariates (Janatian et al., 2017; Langella et al., 2010; Walsh et al., 2017; Y. Wang et al., 2017). Random forest regression has also been used to fit crop responses to agricultural inputs (Everingham et al., 2016; Jeong et al., 2016; Lawes et al., 2019; Mariano & Mónica, 2021; Marques Ramos et al., 2020; Paccioretti et al., 2021; Wang et al., 2016), and initial testing of various model forms indicated that the predictive ability of the random forests outcompeted parametric and Bayesian empirical models (Hegedus & Maxwell, 2022b). The *ranger* package in R was used for fitting and generating predictions (Wright & Ziegler, 2017).

Predictors that did not vary across time were centered by taking the difference between each value and the mean value for the predictor to reduce discrepancies in scale between covariates. Covariates that varied across time (vegetation indices and weather data) were left uncentered because the distribution and subsequent mean of these predictors could vary across time. The covariates used for both models for both crop responses (yield and protein) under the two data constraints are shown in Table 5.3.

Data analyses was performed using R (R Core Team, 2021) where data management, aggregation, cleaning, and analysis with the two model types used the *OFPE* package (Hegedus, 2020). Comparison of the performance of a given model using the PY or DP data was assessed by 5x2 cross validation (CV) developed by Dietterich (1998) for each field, crop response (yield and protein), and for both model types (GAM and RF). The dataset was split 50/50 into a training and test set, and the model using both the PY and DP constraints were fit on the training set and RMSE was calculated on the test set. Then the training and test sets were swapped and both models were fit on the new training set (previously the test set) with RMSE calculated on the test set (previously the training set). This was repeated five times, and a 5x2 CV t statistic was calculated (Dietterich, 1998). Thus, for each field, response, and model type, a t statistic and corresponding two tailed p-value were computed to compare the predictive accuracy (RMSE) of the model fit using data solely from previous years (PY) and data using covariates up to the decision point of the harvest year (DP).

After fitting each model type for both grain yield and protein concentration for each field, crop responses at every location in the field under varying weather conditions were simulated to derive a site-specific N fertilizer recommendation optimized to maximize net-return. For each field, two weather conditions were simulated based on data for a given field: the wettest year from 1999-2021, and the driest year from 1999-2021. Temporal covariate data from the years selected were supplanted into each dataset to perform the simulation under different weather conditions. Each type of model (GAM and RF) for each field and each crop response (yield and grain protein concentration) were trained under both data constraints using the observed three years of data. Crop yield and grain protein concentration were predicted at the centroid of a 10m grid across the field for N rates ranging from 0 to 168 kg ha^{-1}. For every point and N rate, the predicted yield and protein concentration were used to calculate a net-return using the economic conditions from the simulated weather year, beginning with a calculation of the amended price received based on base price and the protein concentration of a given observation;

$$P = Bp + (B0pd + B1pd * protein + B2pd * protein^2)$$

(Equation 5.1)

where *P* is the final price received (\$ kg^{-1}), *Bp* is the base price received (\$ kg^{-1}), *B0pd* is the intercept of the protein premium/dockage function set by the grain elevator, *B1pd* is the coefficient on the grain protein concentration (protein, %) and *B2pd* is the coefficient on the squared protein term. Using this price, net-return was calculated as;

$$NR = yield * P - CA * AA - FC - ssAC$$

(Equation 5.2)

where *NR* is the net-return (\$ ha^{-1}) received and a function of the product of the *yield* (kg ha^{-1}) and *P*, minus the cost of the applied input (*CA*) multiplied by the as-applied rate of the input (*AA*), the fixed costs (*FC*) associated with production (\$ ha^{-1}) that do not include the input, and the cost per hectare of the site-specific application (*ssAC*). The net-return at each point was used to identify the N rates across the field that maximized farmer net-returns and generate a N fertilizer recommendation on a 10m scale.

Results

Model Performance Comparison

The 5x2 CV was performed first using the GAM to simulate yield responses for each field to compare the error estimates relative to observed data between the GAM fit with PY rather than the GAM fit with the DP, measured by RMSE. In two (B1, D2) out of seven fields there was strong evidence against the null hypothesis that there was no difference in the predictive ability between a GAM fit with PY or DP data ($\alpha < 0.05$), with the GAM for yield responses resulting in a higher predictive accuracy under the DP constraint than the PY constraint (Table 5.4). In two (B2, I1) out of seven fields, there was moderate evidence ($\alpha < 0.15$) against the null hypothesis, and weak to no evidence in three (B3, D1, D3) out of seven fields against the null hypothesis ($\alpha \geq 0.15$; Table 5.4). The 5x2 CV was then performed for each field using the RF to simulate yield responses in the same manner as the GAM to compare RMSE between a RF fit with the PY and DP. In contrast to the results of the GAM, there was no evidence against the null hypothesis for any field where either of the data constraints resulted in a difference in performance of the RF for yield responses (Table 5.5).

Assessing the comparative model performance of protein response models, only one (D1) out of seven fields showed a difference in the predictive ability between a GAM fit under the PY or DP ($\alpha < 0.05$), where a higher predictive accuracy under the DP was observed (Table 5.6). In three (B2, B3, I1) out of seven fields, there was moderate evidence ($\alpha < 0.15$) against the null hypothesis of no difference between data constraints, and weak to no evidence ($\alpha > 0.15$) in three (B1, D2, D3) out of seven fields (Table 5.6). While there was no evidence in any field that data constraints played a role in model performance of the RF for yield responses, two (D1, D3) out of seven fields were identified as having a difference in the predictive ability between a RF fit with PY or DP data ($\alpha < 0.05$), with the RF for protein responses under the DP constraint resulting in greater predictive accuracy compared to the RF for protein responses under the PY constraint (Table 5.7). In one (B3) out of seven fields, there was moderate evidence ($\alpha = 0.15$) against the null hypothesis that there was no difference in the data constraints, and weak to no evidence in four (B1, B2, D2, I1) out of seven fields (Table 5.7).

Data Constraint Impact on N Fertilizer Recommendations

For each field, crop yield and protein responses were simulated in the driest and wettest year from 1999-2021 for the field's geographic area, using both model types with both data constraints to compare how the data constraint influenced the pattern of site-specific N fertilizer recommendations and the total N fertilizer applied when fertilizer rates were optimized on maximizing net-returns.

Across all of Montana, 2021 was the driest year on record from 1999-present and used across farms as the dry weather year. For all other farms, the wettest year varied based on the geographic location. The wettest year was 2018 for farm B, 2016 for farm D, and 2006 for farm I. Due to constraints on data availability, the protein premium dockage scheme used for all simulated years was from 2021 and equaled a \$0.02 increase in the base price for every half percent protein above 11.5% up to 14% and an \$0.08 dockage to the base price for every half percent protein below 11.5% down to 9.5%. All the farmers that manage the fields used in the study were able to apply variable rate fertilizer with their own equipment, so the *ssAC* parameter in equation 2 was zero. The fixed costs (*FC*), base price (*Bp*), and cost of N (*CA*) from equation 1 were derived from a survey of our farmers and the USDA Economic Research Service (ERS) using data from the USDA Agricultural Resource Management Survey (Table 5.8). All fields used economic conditions from 2021 to calculate net-return in the simulation of a dry year, while economic conditions from 2018 were used to calculate the net-return in the simulations of a wet year in farm B, 2016 was used in the simulation of a wet year for Farm D, and 2006 was used in the simulation of a wet year for farm I.

While there was evidence of a difference in the predictive ability of the GAM between the PY and DP data constraint for yield responses in field B1 (Table 5.4), this did not appear to play a role in generating a difference in the pattern of site-specific profit maximizing N fertilizer rates and total N applied across the field in either weather year (Appendix Fig. B1). In general, across all fields of farm B and both weather conditions, there were minimal differences between the pattern of profit maximizing rates and total N applied between the data constraints for simulations using the GAM or RF (Appendix Fig. B1-B3).

In field D1, there was a pronounced difference in the pattern of N fertilizer rates produced from the GAM between the DP and PY in the wet year for D1 yet no apparent differences between management outcomes with the GAM in the dry year between DP and PY (Appendix Fig. B4). While there were minimal differences in the pattern of N rates between the DP and PY for the RF model, there were slight differences between DP and PY in the amount of total N recommended for both the wet and dry years (Appendix Fig. B4). For both models fit under the DP constraint there was evidence that it produced more accurate predictions of protein

(Table 5.6, Table 5.7), however no conclusive evidence is shown that a greater predictive ability for protein responses with the DP constraint contributed to the differences in management recommendations for either model in the wet or dry year. While the scaling is deceptive for field D2, there were no drastic differences in the pattern of N recommendations or total amount of N between DP and PY for either model in the wet or dry year, with the largest difference in recommendations between DP and PY occurring with the GAM in the dry year between PY and DP (Appendix Fig. B5). However, in field D3, we observed stark differences in the pattern of N fertilizer recommendations and total amount of N applied between DP and PY in both the wet and dry year with the GAM (Appendix Fig. B6). This occurred despite any difference in the ability of the GAM to predict yield or protein responses between DP and PY, furthering the conclusion that the ability of a model to predict responses is not conclusively correlated to differences in management recommendations between data constraints.

In field I1, comparisons between DP and PY of recommendations in the dry year with either model or the wet year with the RF model showed little to no variation in recommendations (Appendix Fig. B7). There was a slight difference in the pattern of profit maximizing N rates and total N applied in the simulated wet year between the GAM fit with PY data compared to the GAM fit with DP data (Appendix Fig. B7). However, there was no difference in the ability of either model to predict crop responses between data constraints so any difference in recommendations cannot be contributed to differences in the predictive ability of a model between DP and PY.

In most fields, there were only slight differences in the pattern of profit maximizing N fertilizer rates and total N applied between simulations generated from the two data constraints for a given model in a given weather year, albeit with a few exceptions. Field D3 showed the most significant differences in the pattern of N fertilizer rates and total fertilizer applied between data constraints, however this was only apparent between a GAM fit with PY compared to a GAM fit with DP data, but not with the RF model (Appendix Fig. B6). Across most fields, simulations using GAMs appeared to be most sensitive to the different data constraints, resulting in slightly different patterns of N fertilizer rates and total N fertilizer applied, though mostly in wet years, while dry years tended to show less of a difference. The RF seemed to be more robust to differences in the data constraints with minimal variation in either wet or dry years between management recommendations from the PY or DP data constraint. While not the intended comparison, differences in the pattern of profit maximizing N fertilizer rates and total N applied were most prevalent between simulated outcomes from the different model forms (GAM or RF) compared to between DP and PY data constraints for a given model.

Discussion

Grain yield and protein response models to variable top-dress N fertilizer rates were hypothesized to benefit from utilizing covariate information into the crop growing season (DP) rather than relying only on data from the previous years (PY) because the DP contains more information on the current growing season. Using a glimpse of the conditions in the current growing season to predict crop responses at harvest was expected to increase the predictive ability of a model because uncertainty in what the growing conditions were at the beginning of the season would be reduced. It may seem obvious that training models with data up to a decision point would result in better predictions of crop responses at harvest compared to models fit with data from past years. Our analysis explicitly tested the value of utilizing previous year as well as data closer to the decision point for the prediction of crop responses and generation of management recommendations, filling a gap in the literature.

Contrary to our assumptions, there were only a few cases where using data up to a decision point in either a GAM or RF model improved model predictive performance. In field B1, one of the few cases where utilizing data up to the mid-season reduced uncertainty in yield predictions, the field sharing a border (B2) did not share the same result, indicating a difference in data constraints in adjacent fields (Table 5.4). These results highlight the site-specificity of crop responses and justifies the use of on-field experimentation and field specific modeling to make decisions for a field (Hegedus & Maxwell, 2022a).

While in two of seven fields using data up to the decision point improved yield predictions from a GAM compared to a GAM fit with data only from past years, there was no evidence across any fields that yield predictions from RF models were improved by including data up to the decision point (Table 5.4). The RF model resulted in an average 251 kg ha^{-1} reduction in RMSE of yield predictions compared to the GAM, and no observed differences in the predictive ability of the RF between data constraints (Table 5.5). This indicates that the greater general ability to predict crop responses of the RF resulted in less sensitivity of predictive accuracy when the RF was fit under different data constraints.

Further evidence of the insensitivity of the RF to data constraints was found when simulating management outcomes in dry and wet years (Appendix Fig. B1-B7). Even when there was no difference in the RMSE of a GAM (for either crop response) between the PY or DP data constraint, there were more differences between simulated management outcomes of a GAM fit with the PY and DP compared to between RF models fit with either the PY or DP constraint.

Similar to how the RF was more robust to differences in raw predictive ability between data constraints than the GAM, the low degree of differences in N fertilizer recommendations from the RF between DP and PY further indicates how the RF is robust to differences in data constraints, even when simulating management outcomes in extreme weather conditions (Appendix Fig. B1-B7).

Despite the sensitivity of the GAM, there did not seem to be a clear pattern between observing a difference in the predictive ability of either model fit to the two data constraints from the 5x2 CV and observing a difference in predicted outcomes in simulated weather conditions. For example, in field B1, there was strong evidence that the predictive ability of a GAM was greater using data up to the decision point compared to data from past years (Table 5.4), yet in both the wet and dry year simulations, there was no discernable difference in the pattern of profit maximizing N rates or the recommended total N applied to the field between a GAM fit with DP or PY data (Appendix Fig. B1). On the other hand, the 5x2 CV analysis of field I1 showed no indication of a difference in predictive ability for the GAM between DP and PY (Table 5.4), yet in the wet year there were some differences in the recommended pattern of profit maximizing N rates and total N applied between a GAM fit with data up to the decision point and a GAM fit with data from past years (Appendix Fig. B7).

Differences in the predictive ability of a given model between PY and DP constraints did not tend to translate into differences in management recommendations. Across all fields, the greatest difference in the pattern of recommended profit maximizing N rates and total N applied for either a wet or dry year was between the model forms, rather than between data constraints for a given model.

Management decisions for N fertilizer management require high quality data that is appropriate for usage under realistic data constraints. Using models under the DP constraint compared to the PY constraint resulted in greater predictive ability of the model in 5/28 cases. Thus, because of the ease of obtaining weather information, crop response models used to make N fertilizer management recommendations should be fit with data constrained to the point in time that decisions need to be made. Using models with better predictive abilities results in less uncertainty in predictions of yield and protein, an important objective for increasing the resiliency of farmer livelihoods, as their profits are predominately dictated by yield and protein (Hegedus & Maxwell, 2022a). In the cases where there was no difference in model prediction accuracy, such as for RF models fit to yield data, there are only potential benefits to fitting models with all the data that the farmer will have available, and models should still be fit with data up to the point where the farmer must make management decisions.

Open source remotely sensed satellite data was used for this study (Table 5.2) but using data up to a decision point rather than past years may have a greater benefit when collecting sensor data, likely to come available in the future on farms. Data collected on soil moisture or weeds, for example, should be gathered up to the point that producers need to make decisions, so that information can be obtained in the year where decisions in that year affect outcomes. Only covariate data where measurements from the remote sensing sources varied over time (precipitation, GDD, NDVI, and NDWI) differed between the data constraints, while intransient edaphic and topographic variables that did not vary over time were held constant (Table 5.3). However, some of these edaphic variables realistically change over time, for example soil water content, which is sensitive not only to precipitation up until the decision point but throughout the growing season and has a large impact on crop yield and grain protein. This represents a limitation in the nature of the open-source data used in this study. Further limitations of the open-source data beyond not capturing realistic temporal changes are accuracy and scale. This study used the best open-source data available on hand and trusted that the developers of these datasets did due diligence in verifying the accuracy of measurements. Downscaling of remote sensing datasets from satellites constitutes a career of research itself, and was beyond the scope of this paper, though it must be recognized that imperfections and averaging across spatial scales could have contributed to the inability in discerning differences between the predictive ability of models between data constraints. However, despite these limitations, these tools still need to be applied towards the goal of improving efficiency of agricultural inputs. Collaboration between scientific disciplines is needed for improving the spatial scale and accuracy of measurements by ground truthing open-source satellite data to aid in informing farmer decision making at a low-cost.

There will still be uncertainty in predicted yield and protein responses when making management recommendations no matter the data constraint applied to training crop response models. The functional form of the model, rather than the data used to constrain a model, has a greater effect on uncertainty in predicted outcomes, highlighting the need for evaluating the most appropriate crop response model form for a given field in a given year. However, no matter the functional form of crop response models used, accumulating data over time increases producers' understanding of their fields and generates a rich backlog of data to use in precision agriculture decision support systems. These decisions support systems will be critical to sustaining the resources that agriculture relies on for production by harnessing the data from modern farms to increase producer net-returns and reduce pollution from inputs such as N fertilizer.

Conclusion

With increasing uncertainty faced by producers, decision support systems will need to be developed that reduce as much uncertainty in crop response predictions as possible for farmers to make the most informed management decisions. In all cases where there was evidence of a difference between data constraints, models using data up to the decision point that a farmer needs to make management decisions resulted in higher accuracies of crop response predictions. The increased ease of scraping site-specific data from the internet within analysis code relaxes much of the concern of analysis efficiency and thus data inclusions is of little concern, so there is no reason not to include data up to a decision point. Regardless of the data constraints used to fit a model, the greatest uncertainty in predicted management recommendations resulted from the functional form of the model itself, highlighting the need for decision support systems to assess various model types when providing field specific farmers management recommendations.

Acknowledgements

The authors wish to thank the team members of the On-Field Precision Experiment (OFPE) project and the Agroecology Lab at Montana State University. This research was supported by a USDA-NIFA-AFRI Food Security Program Coordinated Agricultural Project, titled “Using Precision Technology in On-farm Field Trials to Enable Data-Intensive Fertilizer Management,” (Accession Number 2016-68004-24769), the USDA-NRCS Conservation Innovation Grant from the On-farm Trials Program, titled “Improving the Economic and

Ecological Sustainability of US Crop Production through On-Farm Precision Experimentation" (Award Number NR213A7500013G021), and the Montana Fertilizer Advisory Council from 2016 - 2021.

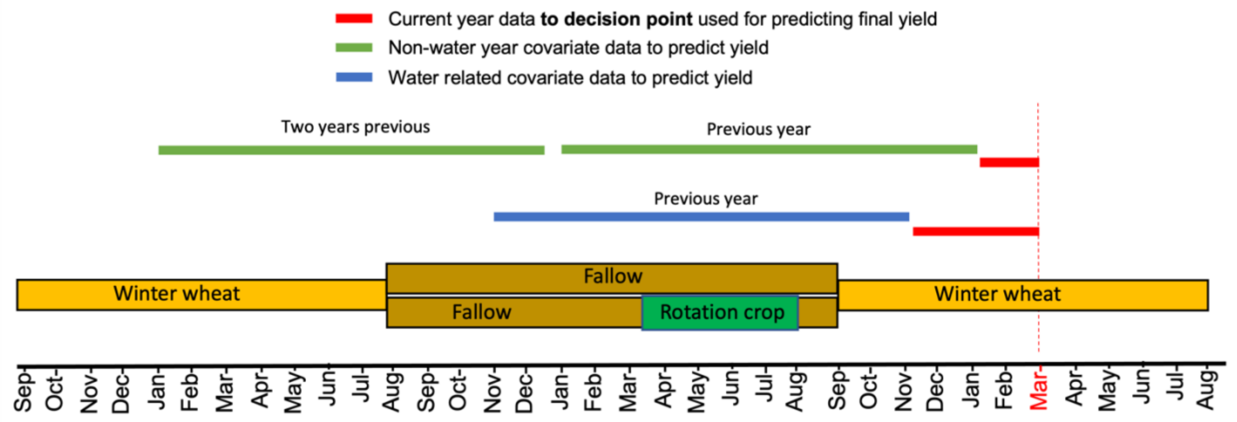


Figure 5.1. Timeline of the data constraints for non-water and water related covariates. Water related covariates are based on water years, while non-water related covariates follow calendar years. The two situations for collecting data correspond to when in relation to harvest the data are collected, where the difference between the approaches is indicated by the red bars.

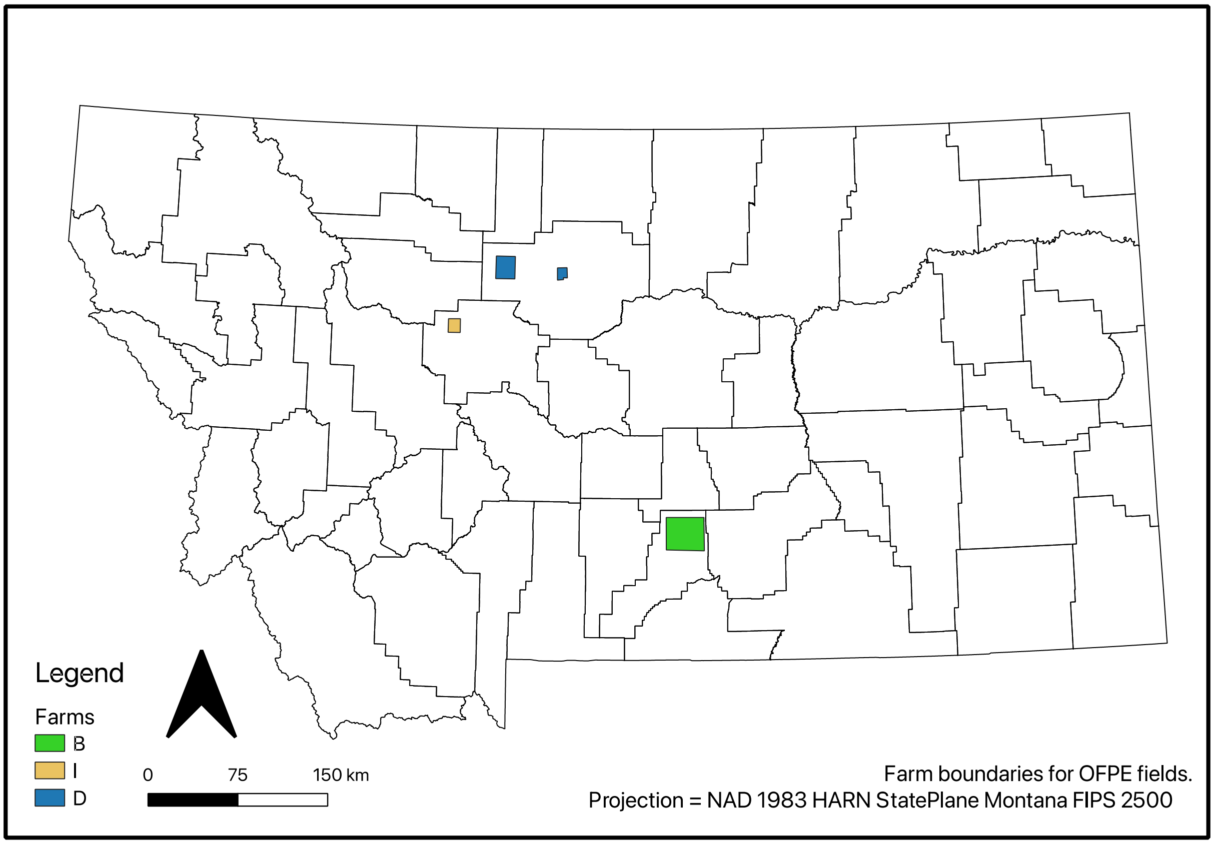


Figure 5.2. Map of OFPE farm boundaries for the three selected farmer collaborators with Montana State University. Colors represent different farmers, while shapes represent general areas in which their respective fields are located.

Table 5.1. Crop histories, field sizes, and years in VRA treatment for each field for given farmers. The farm identifier corresponds to the map above and is used instead of the farmer's name for privacy.

Farm	**Field**	**Field size (ha)**	**Crop History[1]:** *2014 / 2015 / 2016 / 2017 / 2018 / 2019 / 2020 / 2021*	**Years N rate treatment**
B	B1	79	SF / WW / CF / WW / CF / WW / CF / WW	2017, 2019, 2021
	B2	94	WW / CF / WW / CF / SF / WW / CF / WW	2016, 2019, 2021
	B3	64	SW / CF / WW / CF / WW / CF / WW / CF	2016, 2018, 2020
D	D1	46	CF / WW / CF / WW / CF / WW / CF / WW	2017, 2019, 2021
	D2	48	WW / SW / WW / CF / WW / CF / WW / CF	2016, 2018, 2020
	D3	20	SW / SW / WW / CF / WW / CF / WW / CF	2016, 2018, 2020
I	I1	94	SW / CF / WW / CF / WW / CF / WW / CF	2016, 2018, 2020

[1] WW = winter wheat, CF = chemical fallow, SW = spring wheat, SF = safflower

Table 5.2. Table of covariates gathered from Google Earth Engine to enrich the crop yield and protein datasets gathered from farms. In some cases, multiple sources are used, however only one data source is used when aggregating to yield and protein.

Data Type	Data Sources	Resolution	Years Collected	Description
Normalized Difference Vegetation Index (NDVI)	Sentinel 2, Landsat 5/7/8	10m, 30m	S2: 2016-present L5: 1999-2011 L7: 2012-2013 L8: 2014 - present	Sentinel 2 is from the European Space Agency as part of the Copernicus program. Landsat is from USGS and NASA. Bands (NIR, red) S2: B8 and B4 L5/L7: B4 and B3 L8: B5 and B4
Normalized Difference Water Index (NDWI)	Sentinel 2, Landsat 5/7/8	10m, 30m	S2: 2016-present L5: 1999-2011 L7: 2012-2013 L8: 2014 - present	Sentinel 2 is from the European Space Agency as part of the Copernicus program. Landsat is from USGS and NASA. Bands (NIR, red) S2: B3 and B5 L5/L7: B2 and B4 L8: B2 and B5
Normalized Difference Red Edge (NDRE)	Sentinel 2	20m	S2: 2016-present	Bands B5 and B6
Red Edge Chlorophyll Index (CIRE)	Sentinel 2	20m	S2: 2016-present	Bands B7 and B5
Elevation	USGS NED	~10m (1/3 arc second), ~23m (3/4 arc second)	1999-present	USGS National Elevation Dataset. Measured in meters.
Aspect	USGS NED	~10m (1/3 arc second), 30m	1999-present	Direction the surface faces, function of neighboring elevations, in radians. Also calculated for each E/W and N/S direction as cosine and sine.
Slope	USGS NED	~10m (1/3 arc second), 30m	1999-present	Rate of change of height from neighboring cells, in degrees. Measured in degrees.
Topographic Position Index (TPI)	USGS NED	~10m (1/3 arc second), 30m	1999-present	Measure of divots and low spots as a function of neighboring elevation.
Precipitation	DaymetV3	1km	1999-present	Estimates from the NASA Oak Ridge National Laboratory (ORNL). Measured in mm.
Growing Degree Days (GDD)	DaymetV3	1km	1999-present	Estimates from the NASA Oak Ridge National Laboratory (ORNL).
susm	SMAP	10km	2016-present	Surface (0-5cm) and sub-surface (5-100cm) soil moisture content.
grtgroup	OpenLandMap	250m	1999-present	Predicted USDA soil taxonomy great group probabilities.
texture	OpenLandMap	250m	1999-present	Soil texture classes (USDA system) averaged over 6 soil depths (0, 10, 30, 60, 100 and 200 cm).
bulkdensity	OpenLandMap	250m	1999-present	Soil bulk density (fine earth) 10 x kg / m3 averaged over 6 standard depths (0, 10, 30, 60, 100 and 200 cm).
claycontent	OpenLandMap	250m	1999-present	Clay content in % (kg / kg) averaged over 6 standard depths (0, 10, 30, 60, 100 and 200 cm).
sandcontent	OpenLandMap	250m	1999-present	Sand content in % (kg / kg) averaged over 6 standard depths (0, 10, 30, 60, 100 and 200 cm).
pH (phw)	OpenLandMap	250m	1999-present	Soil pH in H_2O averaged over 6 standard depths (0, 10, 30, 60, 100 and 200 cm).
watercontent	OpenLandMap	250m	1999-present	Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted and averaged over 6 standard depths (0, 10, 30, 60, 100 and 200 cm).
carboncontent	OpenLandMap	250m	1999-present	Soil organic carbon content in x 5 g / kg averaged over 6 standard depths (0, 10, 30, 60, 100 and 200 cm).

Table 5.3. Covariates used under the two data constraints. Units and sources are in Table 5.2. CY indicates data from the harvest year, PY indicates data from the year prior to the harvest year, 2PY indicates data from two years prior to the harvest year.

PY	**DP**
As-Applied N	As-Applied N
UTM Coordinates	UTM Coordinates
Aspect	Aspect
Slope	Slope
Elevation	Elevation
Topographic Position Index	Topographic Position Index
Precip. from November 1st (2PY) to October 31st (PY)	Precip. from November 1st (2PY) to October 31st (PY)
	Precip. from November 1st (PY) to March 30th (CY)
GDD from January 1st (PY) to December 31st (PY)	GDD from January 1st (PY) to December 31st (PY)
	GDD from January 1st (CY) to March 30th (CY)
NDVI from January 1st (PY) to December 31st (PY)	NDVI from January 1st (PY) to December 31st (PY)
NDVI from January 1st (2PY) to December 31st (2PY)	NDVI from January 1st (2PY) to December 31st (2PY)
	NDVI from January 1st (CY) to March 30th (CY)
NDWI from January 1st (PY) to December 31st (PY)	NDWI from January 1st (PY) to December 31st (PY)
NDWI from January 1st (2PY) to December 31st (2PY)	NDWI from January 1st (2PY) to December 31st (2PY)
	NDWI from January 1st (CY) to March 30th (CY)
Bulk Density averaged over 0cm – 200cm	Bulk Density averaged over 0cm – 200cm
Clay Content averaged over 0cm – 200cm	Clay Content averaged over 0cm – 200cm
pH of water averaged over 0cm – 200cm	pH of water averaged over 0cm – 200cm
Water Content averaged over 0cm – 200cm	Water Content averaged over 0cm – 200cm
Carbon Content averaged over 0cm – 200cm	Carbon Content averaged over 0cm – 200cm

Table 5.4. Results from 5x2 CV for the GAM fit to yield responses. The mean RMSE, in kg ha^{-1}, were calculated across each split and the 5 folds. An asterisk indicates significance at an alpha level of 0.05 and bolded RMSE values indicate the data constraint with a lower RMSE.

Field	5x2 t statistic	p-value	Mean RMSE PY	Mean RMSE DP
B1	4.3416	0.0074*	858.8653	**848.3403**
B2	2.1672	0.0824	788.9887	783.4343
B3	0.4362	0.6809	876.8696	899.8229
D1	-0.8908	0.4139	538.0681	536.7981
D2	4.8024	0.0049*	719.2763	**714.4325**
D3	0.1869	0.8591	762.5487	763.6518
I1	1.9276	0.1118	812.1172	805.8488

Table 5.5. Results from 5x2 CV for the RF fit to yield responses. The mean RMSE, in units of kg ha^{-1}, were calculated across each split and the 5 folds.

Field	5x2 t statistic	p-value	Mean RMSE PY	Mean RMSE DP
B1	-1.5857	0.1737	572.5701	578.1144
B2	-0.5464	0.6083	533.3079	534.0513
B3	-0.5831	0.5852	590.6175	592.1910
D1	1.2008	0.2836	387.3872	385.2688
D2	0.7394	0.4929	398.1208	397.2827
D3	0.1423	0.8924	548.1305	546.6204
I1	-0.1685	0.8728	556.0405	561.4883

Table 5.6. Results from 5x2 CV for the GAM fit to grain protein concentration responses. The mean RMSE, in % protein, were calculated across each split and the 5 folds. An asterisk indicates significance at an alpha level of 0.05 and bolded RMSE values indicate the data constraint with a lower RMSE.

Field	**5x2 t statistic**	**p-value**	**Mean RMSE PY**	**Mean RMSE DP**
B1	1.5650	0.1784	1.6953	1.6963
B2	-2.3742	0.0636	1.3179	1.3212
B3	1.9562	0.1078	1.3187	1.3116
D1	4.5167	0.0063*	1.1303	**1.1026**
D2	1.3191	0.2443	1.2219	1.2160
D3	0.2611	0.8045	1.2785	1.2815
I1	-2.4841	0.0556	1.5910	1.6007

Table 5.7. Results from 5x2 CV for the RF fit to grain protein concentration responses. The mean RMSE, in % protein, were calculated across each split and the 5 folds. An asterisk indicates significance at an alpha level of 0.05 and bolded RMSE values indicate the data constraint with a lower RMSE.

Field	5x2 t statistic	p-value	Mean RMSE PY	Mean RMSE DP
B1	1.5946	0.1717	1.6697	1.6683
B2	0.1305	0.9013	1.2754	1.2780
B3	1.7632	0.1382	1.2871	1.2830
D1	2.7191	0.0418*	1.0866	**1.0786**
D2	-1.3101	0.2471	1.1660	1.1684
D3	2.9727	0.0311*	1.2420	**1.2394**
I1	-0.4395	0.6786	1.4567	1.4566

Table 5.8. Economic parameters across the years used in the dry (2021) and wet (2018, 2016, 2006) simulations that were used to calculate net-return in Equation 5.1.

Economic Parameter	**2021**	**2018**	**2016**	**2006**
Fixed Costs (FC)	\$181.87 ha^{-1}	\$181.94 ha^{-1}	\$169.61 ha^{-1}	\$136.08 ha^{-1}
Base Price (Bp)	\$0.27 kg^{-1}	\$0.18 kg^{-1}	\$0.13 kg^{-1}	\$0.15 kg^{-1}
Cost of N (CA)	\$0.89 kg^{-1}	\$0.71 kg^{-1}	\$0.75 kg^{-1}	\$0.95 kg^{-1}

REFERENCES CITED

Biermacher, J. T., Brorsen, B. W., Epplin, F. M., Solie, J. B., & Raun, W. R. (2009). The economic potential of precision nitrogen application with wheat based on plant sensing. *Agricultural Economics*, *40*(4), 397–407. https://doi.org/10.1111/j.1574-0862.2009.00387.x

Boyer, C. N., Larson, J. A., Roberts, R. K., McClure, A. T., Tyler, D. D., & Zhou, V. (2013). Stochastic Corn Yield Response Functions to Nitrogen for Corn after Corn, Corn after Cotton, and Corn after Soybeans. *Journal of Agricultural and Applied Economics*, *45*(4), 669–681. https://doi.org/10.1017/s1074070800005198

Bucci, G., Bentivoglio, D., & Finco, A. (2018). Precision agriculture as a driver for sustainable farming systems: State of art in litterature and research. *Quality - Access to Success*, *19*(S1), 114–121.

Chen, K., O'Leary, R. A., & Evans, F. H. (2019). A simple and parsimonious generalised additive model for predicting wheat yield in a decision support tool. *Agricultural Systems*, *173*(December 2018), 140–150. https://doi.org/10.1016/j.agsy.2019.02.009

Clancy, M. (2019). Real-time protein monitoring. *Next Instruments*, (June).

Cook, S., Cock, J., Oberthür, T., & Fisher, M. (2004). On-Farm Experimentation. *Better Crops*, *97*, 17–20.

Córdova, C., Barrera, J. A., & Magna, C. (2018). Spatial variation of nitrogen mineralization as a guide for variable application of nitrogen fertilizer to cereal crops. *Nutrient Cycling in Agroecosystems*, *110*(1), 83–88. https://doi.org/10.1007/s10705-017-9886-2

Croft, H., Arabian, J., Chen, J. M., Shang, J., & Liu, J. (2020). Mapping within-field leaf chlorophyll content in agricultural crops for nitrogen management using Landsat-8 imagery. *Precision Agriculture*, *21*(4), 856–880. https://doi.org/10.1007/s11119-019-09698-y

Dietterich, T. G. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, *10*(7), 1895–1923. https://doi.org/10.1162/089976698300017197

Duff, H., Hegedus, P. B., Loewen, S., Bass, T., & Maxwell, B. D. (2022). Precision Agroecology. *Sustainability*, *14*(106), 1–18. https://doi.org/https://doi.org/10.3390/su14010106

Estes, L. D., Bradley, B. A., Beukes, H., Hole, D. G., Lau, M., Oppenheimer, M. G., … Turner, W. R. (2013). Comparing mechanistic and empirical model projections of crop suitability and productivity: Implications for ecological forecasting. *Global Ecology and Biogeography*, *22*(8), 1007–1018. https://doi.org/10.1111/geb.12034

Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development*, *36*(2). https://doi.org/10.1007/s13593-016-0364-z

Ewing, P. M., & Runck, B. C. (2015). Optimizing nitrogen rates in the midwestern United States for maximum ecosystem value. *Ecology and Society*, *20*(1). https://doi.org/10.5751/ES-06767-200118

Flowers, M., Weisz, R., Heiniger, R., Osmond, D., & Crozier, C. (2004). In-Season Optimization and Site-Specific Nitrogen Management for Soft Red Winter Wheat. *Agronomy Journal*, *96*(1), 124–134. https://doi.org/10.2134/agronj2004.0124

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., … Zaks, D. P. M. (2011). Solutions for a cultivated planet. *Nature*, *478*(7369), 337–342. https://doi.org/10.1038/nature10452

Gaso, D. V., Berger, A. G., & Ciganda, V. S. (2019). Predicting wheat grain yield and spatial variability at field scale using a simple regression or a crop model in conjunction with Landsat images. *Computers and Electronics in Agriculture*, *159*(February 2018), 75–83. https://doi.org/10.1016/j.compag.2019.02.026

Gonzalez-Sanchez, A., Frausto-Solis, J., & Ojeda-Bustamante, W. (2014). Predictive ability of machine learning methods for massive crop yield prediction. *Spanish Journal of Agricultural Research*, *12*(2), 313–328. https://doi.org/10.5424/sjar/2014122-4439

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google

Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gotway, C. A., & Stroup, W. W. (1997). A Generalized Linear Model Approach to Spatial Data Analysis and Prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, *2*(2), 157–178.

Gross, M. (2013). The paradoxical evolution of agriculture. *CURBIO*, *23*(16), R667–R670. https://doi.org/10.1016/j.cub.2013.08.001

Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling*, *157*(2–3), 89–100. https://doi.org/10.1016/S0304-3800(02)00204-1

Hair, J. F. (2007). Knowledge creation in marketing: The role of predictive analytics. *European Business Review*, *19*(4), 303–315. https://doi.org/10.1108/09555340710760134

Hegedus, P.B. (2020). OFPE: On-Farm Precision Experiments (OFPE) Data Management and Analysis Tools. R package version 1.8.11. https://github.com/paulhegedus/OFPE.git

Hegedus, P.B & Maxwell, B.D. (2022a). Rationale for field specific on-farm precision experimentation. *Agriculture, Ecosystems & Environment.* Manuscript submitted for publication, in review.

Hegedus, P.B & Maxwell, B.D. (2022b). Assessing performance of empirical models for forecasting crop responses to variable fertilizer rates using on-farm precision experimentation. Manuscript in preparation.

Holland, D. M., De Oliveira, V., Cox, L. H., & Smith, R. L. (2000). Estimation of regional trends in sulfur dioxide over the eastern United States. *Environmetrics*, *11*(4), 373–393. https://doi.org/10.1002/1099-095X(200007/08)11:4<373::AID-ENV419>3.0.CO;2-2

Hurley, T. M., Malzer, G. L., & Kilian, B. (2004). Estimating site-specific nitrogen crop response functions: A conceptual framework and geostatistical model. *Agronomy Journal*, *96*(5), 1331–1343. https://doi.org/10.2134/agronj2004.1331

Hurley, Terrance M., Oishi, K., & Malzer, G. L. (2005). Estimating the potential value of variable rate nitrogen applications: A comparison of spatial econometric and geostatistical models. *Journal of Agricultural and Resource Economics*, *30*(2), 231–249. https://doi.org/10.2307/40988069

Janatian, N., Sadeghi, M., Sanaeinejad, S. H., Bakhshian, E., Farid, A., Hasheminia, S. M., & Ghazanfari, S. (2017). A statistical framework for estimating air temperature using MODIS land surface temperature data. *International Journal of Climatology*, *37*(3), 1181–1194. https://doi.org/10.1002/joc.4766

Jaynes, D. B., & Colvin, T. S. (2006). Corn yield and nitrate loss in subsurface drainage from midseason nitrogen fertilizer application. *Agronomy Journal*, *98*(6), 1479–1487. https://doi.org/10.2134/agronj2006.0046

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., … Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *PLoS ONE*, *11*(6), 1–15. https://doi.org/10.1371/journal.pone.0156571

Jones, C. (2013). Mid- to late-season nitrogen fertilizer: a viable option to increase wheat yield and grain protein. Retrieved from MSU News Service website: https://www.montana.edu/news/11934/msu-extension-mid-to-late-season-nitrogen-fertilizing-can-boost-yield-protein-in-wheat-crops

Joshi, V. R., Kazula, M. J., Coulter, J. A., Naeve, S. L., & Garcia y Garcia, A. (2021). In-season weather data provide reliable yield estimates of maize and soybean in the US central Corn Belt. *International Journal of Biometeorology*, *65*(4), 489–502. https://doi.org/10.1007/s00484-020-02039-z

Khosla, R., Inman, D., Westfall, D. G., Reich, R. M., Frasier, M., Mzuku, M., … Hornung, A. (2008). A synthesis of multi-disciplinary research in precision agriculture: Site-specific management zones in the semi-arid western Great Plains of the USA. *Precision Agriculture*, *9*(1–2), 85–100. https://doi.org/10.1007/s11119-008-9057-1

Khosla, Rajiv, Fleming, K., Delgado, J. A., Shaver, T. M., & Westfall, D. G. (2002). Use of site-specific management zones to improve nitrogen management for precision agriculture.

Journal of Soil and Water Conservation, *57*(6), 513–518.

Koch, B., Khosla, R., Frasier, W. M., Westfall, D. G., & Inman, D. (2004). Economic feasibility of variable-rate nitrogen application utilizing site-specific management zones. *Agronomy Journal*, *96*(6), 1572–1580. https://doi.org/10.2134/agronj2004.1572

Lacoste, M., Cook, S., Mcnee, M., Gale, D., Ingram, J., Bellon-maurel, V., … Hall, A. (2022). On-Farm Experimentation to transform global agriculture. *Nature Food*. https://doi.org/https://doi.org/10.1038/s43016-021-00424-4 On-Farm

Langella, G., Basile, A., Bonfante, A., & Terribile, F. (2010). High-resolution space-time rainfall analysis using integrated ANN inference systems. *Journal of Hydrology*, *387*(3–4), 328–342. https://doi.org/10.1016/j.jhydrol.2010.04.027

Lawes, R. A., Oliver, Y. M., & Huth, N. I. (2019). Optimal nitrogen rate can be predicted using average yield and estimates of soil water and leaf nitrogen with infield experimentation. *Agronomy Journal*, *111*(3), 1155–1164. https://doi.org/10.2134/agronj2018.09.0607

Lawrence, P. G., Rew, L. J., & Maxwell, B. D. (2015). A probabilistic Bayesian framework for progressively updating site-specific recommendations. *Precision Agriculture*, *16*(3), 275–296. https://doi.org/10.1007/s11119-014-9375-4

Liew, J., & Forkman, J. (2015). A guide to generalized additive models in crop science using SAS and R. *Communications in Biometry and Crop Science*, *10*(1), 41–57.

Luschei, E. C., Van Wychen, L. R., Maxwell, B. D., Bussan, A. J., Buschena, D., & Goodman, D. (2001). Implementing and conducting on-farm weed research with the use of GPS. *Weed Science*, *49*(4), 536–542. https://doi.org/10.1614/0043-1745(2001)049[0536:iacofw]2.0.co;2

Mariano, C., & Mónica, B. (2021). A random forest-based algorithm for data-intensive spatial interpolation in crop yield mapping. *Computers and Electronics in Agriculture*, *184*(March). https://doi.org/10.1016/j.compag.2021.106094

Marques Ramos, A. P., Prado Osco, L., Elis Garcia Furuya, D., Nunes Gonçalves, W., Cordeiro

Santana, D., Pereira Ribeiro Teodoro, L., … Pistori, H. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, *178*(July). https://doi.org/10.1016/j.compag.2020.105791

McBratney, A., Whelan, B., Ancev, T., & Bouma, J. (2005). Future directions of precision agriculture. *Precision Agriculture*, *6*(1), 7–23. https://doi.org/10.1007/s11119-005-0681-8

Paccioretti, P., Bruno, C., Gianinni Kurina, F., Córdoba, M., Bullock, D. S., & Balzarini, M. (2021). Statistical models of yield in on-farm precision experimentation. *Agronomy Journal*, *113*(6), 4916–4929. https://doi.org/10.1002/agj2.20833

Qianqian, D., Mieno, T., Bullock, D. S., & Edge, B. (2021). Economically Optimal Nitrogen Side-dressing Based on Vegetation Indices from Satellite Images Through On-farm Experiments. In Harper Adams University (Ed.), *Proceedings of the 4th Symposium on Agri-Tech Economics for Sustainable Futures* (pp. 51–62). Newport, United Kingdom: HAU Publications.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rütting, T., Aronsson, H., & Delin, S. (2018). Efficient use of nitrogen in agriculture. *Nutrient Cycling in Agroecosystems*, *110*(1), 1–5. https://doi.org/10.1007/s10705-017-9900-8

Stepanov, A., Dubrovin, K., Sorokin, A., & Aseeva, T. (2020). Predicting soybean yield at the regional scale using remote sensing and climatic data. *Remote Sensing*, *12*(12). https://doi.org/10.3390/rs12121936

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108

Walsh, E. S., Kreakie, B. J., Cantwell, M. G., & Nacci, D. (2017). A Random Forest approach to predict the spatial distribution of sediment pollution in an estuarine system. *PLoS ONE*, *12*(7), 1–18. https://doi.org/10.1371/journal.pone.0179473

Wang, L., Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Crop Journal*, *4*(3), 212–219. https://doi.org/10.1016/j.cj.2016.01.008

Wang, Y., Wu, G., Deng, L., Tang, Z., Wang, K., Sun, W., & Shangguan, Z. (2017). Prediction of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-07197-6

Weersink, A., Fraser, E., Pannell, D., Duncan, E., & Rotz, S. (2018). Opportunities and Challenges for Big Data in Agricultural and Environmental Analysis. *Annual Review of Resource Economics*, *10*(August), 19–37. https://doi.org/10.1146/annurev-resource-100516-053654

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *65*(1), 95–114. https://doi.org/10.1111/1467-9868.00374

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, *111*(516), 1548–1563. https://doi.org/10.1080/01621459.2016.1180986

Wright, M.N. & A. Ziegler (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 77(1), 1-17. doi:10.18637/jss.v077.i01

Zeng, X., Han, B., Xu, F., Huang, J., Cai, H., & Shi, L. (2012). Effects of modified fertilization technology on the grain yield and nitrogen use efficiency of midseason rice. *Field Crops Research*, *137*, 203–212. https://doi.org/10.1016/j.fcr.2012.08.012

Zuur, A. F., & Camphuysen, Kees, C. (2012). 6 Generalized Additive Models applied on northern gannets. In *A Beginners Guide to Generalized Additive Models with R* (p. 15). Highland Statistics Ltd.

CHAPTER SIX

EVALUATION OF NITROGEN USE EFFICIENCY IN DRYLAND WINTER-WHEAT AGROECOSYSTEMS FOR PREDICTIVE MODELING

Contribution of Authors and Co-Authors

Manuscript in Chapter Six

Author: Paul B. Hegedus

Contributions: Conceived the study design and methodology, collected data, executed statistical analysis, and wrote initial manuscript.

Co-Author: Stephanie A. Ewing

Contributions: Assisted with study design, analysis, and manuscript preparation at all stages.

Manuscript Information

Paul B. Hegedus, Stephanie A. Ewing

Nutrient Cycling in Agroecosystems

Status of Manuscript:
__x__ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Published by Springer

Abstract

Low nitrogen use efficiency (NUE) is ubiquitous in agricultural systems, with mounting global scale consequences for both atmospheric aspects of climate and downstream ecosystems. In dryland cereal production systems, NUE is driven by soil water holding capacity (WHC), weather, and management, and reflects internal soil production of available nitrogen (N) as well as fertilizer application. Fallow fields in wet years results in high N loss, yet in any conditions, soils with low WHC within fields have the greatest N loss to leaching. Since NUE related soil characteristics are likely to vary at small scales (~30 m pixel) understanding the influence of soil structure on NUE at the subfield scale represents an opportunity to increase the efficiency of applied N fertilizer. We evaluated NUE in four conventionally managed winter-wheat fields in Montana following multiple years of sub-field scale variation in experimental N fertilizer applications. Soil and plant biomass sampling were conducted on each field at sample locations randomly stratified based on estimates of WHC and experimental N rates. Nitrogen use efficiency was calculated from soil and plant biomass samples using soil total N, total N in aboveground winter wheat biomass, and estimated N mineralization rates. While NUE varied across fields and years, efficiency was highest in areas of fields with low N availability from both fertilizer and estimated mineralization of soil organic N (SON). Results suggested that to maximize NUE, site-specific applied N fertilizer rates should account for available N supplied by SON. A predictive model of subfield scale NUE was developed using data gathered from open-source data repositories and from normal farm operations. Ecological and biogeochemical parameters important to NUE included topographic and soil characteristics, normalized difference vegetation index (NDVI), normalized difference water index (NDWI), and

precipitation. A support vector regression model of NUE that included these parameters provided predictions with the least error relative to observed NUE. Developing subfield scale accurate predictive models of NUE for fertilizer application in dryland agriculture has potential to increase efficiency in agronomic systems through integration with a decision support system, where optimized application prescriptions are based on maximizing farmer net profit and NUE. This research also contributes a general approach to developing prescriptive models of inputs vs. input use efficiency for any agronomic system.

Introduction

The Green Revolution has increased crop production via an altered the trajectory of agriculture, including the development of high yielding crop varieties, sensor and GIS technology, chemical inputs, and agricultural infrastructure (Foley et al., 2011; Gliessman & Engles, 2014). This trajectory has led to the industrialization of farms in developed nations, where maximization of production and profit guide management decisions that impact neighboring agricultural and natural ecosystems and communities (Coleman, 1989; Gliessman & Engles, 2014; Paul & Robertson, 1989). Haber's 1909 discovery of an artificial nitrogen (N) fixation process catalyzed a century of booming agricultural production and development, feeding the rapidly growing world population and fueling consumption of N fertilizer. Between 1940 and 1980, the amount of fertilizer applied in the United States increased from 9 to 47 million metric tons and reached 12 Tg N yr^{-1} in 2015, furthering long-held concerns about the ecological and environmental implications of intensive fertilizer use (Gliessman & Engles, 2014; Vitousek et al., 1997). Studies to date suggest that an amount of reactive N equal to at least half of fertilizer N applied is now lost each year to N leaching and denitrification from soils

(Bouwman et al., 2013). Overapplication of inorganic fertilizers in the quest for maximizing current productivity degrades soil through acidification, degrades water resources through pollution or eutrophication, contributes to biodiversity loss and greenhouse gas emissions (Allaire et al., 2018; Capel et al., 2008; DeLonge et al., 2016; Weiner, 2017).

Increasing the efficiency of agricultural inputs, such as N fertilizer, is a key step in transitioning modern agriculture towards sustainability (Gliessman, 2016). Nitrogen use efficiency (NUE) is typically measured as the amount of nitrogen in crop biomass compared to available N (Ping & Dobermann, 2003; Prey et al., 2019; L. Yin et al., 2020). The relationship between crop N and available N over time depends on how available N is assessed, and here we distinguish between NUE and fertilizer use efficiency (FUE). Fertilizer use efficiency is a ratio of the amount of N taken up by a crop to the amount of N fertilizer applied, while NUE is the ratio of N taken up by a crop to the total amount of available N in the soil. Depending on the method, FUE estimates range from 18-65% while typical NUE estimates are around 30% (Guttieri et al., 2017; Macnack et al., 2014; Sebilo et al., 2013b).

Improving NUE is of interest because excess N from fertilizers has been linked to elevated nitrate levels in drinking water, acidification of agricultural soils in dryland small-grain agroecosystems, and substantial loss to denitrification (Engel, 2012; John et al., 2017; Sigler et al., 2018; Sigler et al. 2022), arguing against the sustainability of agronomic systems reliant on conventional N fertilizer management practices. Generally, the degree of NUE from conventionally managed fields in dryland Montana agroecosystems is dictated by soil character, weather, and agronomic management (Sigler et al., 2020). Efficiency is lowest when nitrogen loss is greatest, which occurs with deep percolation of water following high precipitation events and post fallow conditions (Sigler et al. 2020, 2022). In contrast to fallow systems, crop rotations including crops that required reduced N additions reduced N leaching, indicating the influence of management on N loss from Montana's dryland agroecosystems (Sigler et al., 2018). In a study from central Montana conducted in dryland agroecosystems, nitrate leaching was estimated to be three times higher post-fallow than post-pea (John et al. 2017). Not only does fallow increase the potential of deep percolation of soil water to the aquifer, but evidence suggests it stimulates mineralization, increasing the risk of nitrate leaching (Sigler et al. 2020). While heavy precipitation induced hot moments of nitrate leaching, hot spots of nitrate leaching within agricultural fields were dictated by soil character.

Soil texture within the rooting zone and the resulting WHC are a primary control on NUE that interacts with variable precipitation in dryland systems (Sigler et al. 2020). The transport of nitrate via water through the root zone was mitigated by the water holding capacity (WHC) of soils, and soils with fine-textured zones less than 25 cm thick were estimated to experience nitrate leaching loss 16 times higher than soils with fine-textured zones greater than 100 cm thick over a 15-year period of variable precipitation (Sigler et al. 2020). If precipitation inputs exceed crop water uptake and soil water storage, nitrate leaching is hypothesized to be a dominant pathway of N loss because nitrate that has not been taken up by the crop or retained in the soil can be transported below the crop rooting depth in well-drained soil. Nitrate lost to leaching augments loss via denitrification, the combination of which can be up to 50% of total available nitrate in Montana dryland agroecosystems (Sigler et al., 2022). Denitrification occurs when water storage limits oxygen, facilitating this anaerobic process. If N limitation of dryland small-grain agroecosystems is a function of leaching and denitrification loss with excess N application, then soil water storage dynamics that occur as a function of WHC will dictate yield capacity when N fertilizer application exceeds crop requirements. Thus, soils with high WHC will generally contribute to higher yields and less nitrate leaching of excess fertilizer because the water is retained in the root zone, although greater susceptibility to denitrification with high WHC may compromise efficiency depending on rainfall patterns and crop rotation (Sigler et al. 2022).

Farmers who understand the controls and drivers of NUE can alter decisions based on conditions and practices that expose them to potential economic loss due to low NUE. Understanding the variation in soil character within fields represents an opportunity to inform site-specific N fertilizer management that maximizes NUE within fields. While farmers can control cropping systems choices, they cannot control the weather; yet advances in remote sensing and modeling have improved the accuracy and spatial scale of weather predictions, which, compounded with the affordability of personal weather stations, provide farmers with weather forecasts to inform N fertilizer management decisions. Importantly, crop rotations and weather influence N loss at the field scale, while soil texture and structure, and the resulting WHC, vary at the sub-field scale.

The subfield variation in crop response to variable N fertilizer (Hegedus & Maxwell, 2022a) suggests that NUE varies substantially within fields, at a resolution higher than most farmers sampling practices can characterize. Soil mapping and soil sampling provide information on the WHC of soils yet current soil maps may not provide the resolution needed for site-specific N fertilizer management (Kamilaris et al., 2017; Lokers et al., 2016; Wolfert et al., 2017). Additionally, soil sampling and analysis are expensive and typically conducted in a sparse manner across fields, representing the difficulty and feasibility for farmers to acquire and analyze soil data at spatial resolutions adequate for making site-specific N management decisions. In response to this challenge, models of the relationships between NUE and spectral imagery and landscape characteristics have been developed (Oelofse et al., 2015; Pavuluri et al., 2015; Semenov et al., 2007). These attempts at modeling NUE have typically involved simple linear models that require site and time specific parameters, and represent a first step towards developing tools to estimate efficiency at subfield scales for informing N fertilizer management decisions (Arnall et al., 2009; Macnack et al., 2014; Van Sanford & MacKown, 1986).

Precision agriculture has progressed over the last few decades into a management approach for inputs like N fertilizer that accounts for spatial variation at relatively high resolutions across fields (Bullock et al., 2019). At the same time, the recent revolution in agricultural data acquisition has coincided with the introduction of data science approaches into agronomic research and development (Coble et al., 2018; Pham & Stack, 2018; Vinila Kumari et al., 2016). The amount of data available on and about farms has introduced the opportunity for data science to aid in the management and analysis of agronomic data and for informing local within-field decision making (Gibert et al., 2018; Provost & Fawcett, 2013). Investigating the benefit of generalizable data science tools for modeling the relationship of NUE with soil parameters, N fertilizer rates, and other open-source data has implications for informing precision agroecological approaches to N fertilizer management that harness precision agriculture to apply management conscious of economic and ecological outcomes (Duff et al., 2022).

The goal of this study was to introduce protocols for utilizing empirical measurements to estimate NUE and develop tools and information that can aid farmers in the application of N fertilizer that maximizes profit by reducing the risk of the loss of applied N when NUE is low. The first objective of this study was to evaluate the relationships of causal variables derived from low-cost open-source and on-farm data sources and NUE in conventional dryland Montana agroecosystems cropped with winter wheat (*Triticum aestivum* L.) To allow this understanding to be used in a hypothetical low-cost decision support system, the second objective was exploration of a suite of potential generalizable models to assess their capacity to predict NUE in similar dryland winter-wheat systems without soil sampling. The model was intended to estimate subfield NUE for a given year without requiring producers to take detailed soil samples or provide parameters for complex N transport models. While developed in dryland winter-wheat systems of Montana, the methods outlined in this paper are applicable to other crops and systems when similar procedures are used to generate information for increasing agronomic input efficiency with simultaneous consideration for profitability.

Methods

Study Sites

Four fields from farmers collaborating with the On-Farm Precision Experiments (OFPE) project (https://sites.google.com/site/ofpeframework/home) at Montana State University (MSU) were used to evaluate and develop a general NUE model for dryland small-grain agroecosystems. Fields selected from the OFPE project were based on previous data quality, site availability, and geographic representation in different growing conditions and regions of Montana. Two fields from each of two distinct climatic and soil growing regions of Montana

(Fig. 6.1) were sampled in 2020 and 2021(Table 6.1). Each of the fields were subjected to spatially variable N applications in the seasons since 2016 when winter wheat was grown, and N fertilizer rates were randomly stratified on previous yield, protein, and as-applied N (Fig. 6.2).

Mean annual precipitation from 2000 - 2020 was 367 mm for Farm B, 315 mm for Farm M, and 322 mm for Farm W (Montana Climate Office, 2020). The mean annual temperature f from 2000 - 2020 for all three locations was 7 °C (Montana Climate Office, 2020). Soils in BSE primarily were well-drained Aridic Haplocambids, characteristic of alluvial fans and stream terraces (Soil Survey Staff, 2020). The predominant soil type in BS was a well-drained Lambeth silt loam (calcareous, frigid Aridic Ustorthents) with deep Yawdim clay (frigid, shallow Aridic Ustorthents), while in the eastern and western portions of the field a shallow and well-drained Rentsac loam (superactive, frigid Lithic Calciustepts) was most prevalent. Tanna clay loam (Fine, smectitic, frigid Aridic Argiustolls) dominated MS, a well-drained soil developed from alluvium. In the southern portion of WH, Scobey clay loam (smectitic, frigid Aridic Argiustolls) dominated, while Ethridge silty clay loam (smectitic, frigid Torrertic Argiustolls) was the most abundant in the northern half of the field. Both soils are well drained and formed in alluvium and glaciofluvial deposits. Fields with differing soil characteristics were chosen to maximize the representation of variation in soils across Montana, in the interest of maximizing the generality of the model for estimating NUE within dryland small-grain agroecosystems. While in an ideal world, field specific measurements to generate field specific models would be available, offering a general NUE for dryland small-grain agroecosystems gives farmers access to NUE information at a low-cost.

The fields observed in this study were similar in management and environment to the fields observed by Sigler et al. (2020) and John et al. (2017). All fields used in our study were cropped with winter wheat in alternate years with chemical fallow. While the detailed character of soils used in this study differed from those in Sigler et al. (2020) and John et al. (2017), these well-drained soils are on similar landforms with similar climate, and are similarly cropped under dryland conventional management, suggesting that inference on the drivers of N loss from Sigler et al. (2020) and John et al. (2017) pertain to the fields used in this study.

Classification of Water Holding Capacity

A high spatial density estimate of WHC was created based on remotely sensed Normalized Difference Vegetation Index (NDVI) data. Greater soil WHC is expected to result in consistency in productivity over time at a specific point in a field, while more variable productivity may reflect moderate WHC and risk of N loss. While no studies have linked historical maps of NDVI to the corresponding NUE or WHC, Araya et al. (2016) used NDVI to estimate available water content, indicating the promise NDVI has for estimating soil properties. Identifying areas of a field at risk of N loss using freely available remotely sensed data benefits farmers by providing information on where NUE is potentially low, which can inform site-specific N fertilizer management decisions.

The fields used in this study were classified into WHC zones that were used to stratify soil and plant biomass sampling. NDVI was derived from satellite imagery from 2000 – 2021. Landsat imagery was collected on 30 m x 30 m raster grids, a coarser resolution than satellites such as Sentinel-2. Landsat data, however, was chosen based on the longer history of data available (Table 6.2).

Imagery was downloaded from Google Earth Engine (GEE), where image processing was performed to correct images for clouds and other atmospheric anomalies (Gorelick et al., 2017). The cloud computing platform of GEE was used to calculate NDVI as:

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

(Equation 6.1)

where ‘NIR’ is reflectance at the near-infrared wavelength (~865 nm) and ‘Red’ correspond to reflectance at a visible red wavelength (~665 nm). Landsat data were downloaded for each of the boundaries of the individual farms, from which images were clipped to the boundary of each field. Although data were collected from 2000 – 2021, years in which average NDVI across each field fell below 0.3 were omitted because these represent fallow years in which no crops were grown, and any productivity was due to weeds growing in the fallow field (Fig. 6.3).

For each field, the rasterized NDVI for each year was stacked, and WHC was classified based on each pixel’s mean and coefficient of variation (CV) to produce a single map with classifications of high, medium, and low WHC. High, medium, and low categories for the mean and CV for each pixel were delineated via the Jenks optimization method, an algorithm for maximizing between class variance and minimizing within class variance (Jenks, 1967). A matrix of the mean and CV groups for NDVI were then derived to estimate hypothesized WHC zones (Table 6.3). If an area of a field is consistently productive over time, the crop is acquiring adequate water and N to maximize growth across varying annual precipitation.

Soil and Plant Biomass Sampling

Soil samples were taken at 75 locations in each field twice per year, in March and August, randomly selected with stratification based on WHC category and N application rate

(Fig. 6.4). Soil samples were taken at depths of 0-15, 15-30, 30-60, 60-90, and from 90cm to as deep as possible. The depth of the deepest sample corresponded to either gravel content, rock layer or the limit of the probe (110 cm). The maximum depth at each location was recorded during sampling. Soil samples were obtained by a hydraulic probe coring device in March (pre-fertilization) and August (post-harvest). Soil samples were taken within seven days prior to fertilizer application and within seven days post-harvest. Winter wheat biomass samples were hand harvested in June at plant maturity at each of the sampling locations. Samples were transported in coolers on ice from the fields back to MSU where soil and plant samples were weighed wet, oven dried at 50 degree C for 7-14 days, and then weighed dry to determine water content. Plant samples were ground using an UDY Cyclone Sample Mill (*UDY Corporation*, Fort Collins, CO, USA). Soil samples were milled to break up peds and aggregates, sieved to remove rocks and roots, and homogenized with a Dynacrush DC-5 soil crusher (*Custom Laboratory*, Hodlen, MO, USA). Aliquots were taken from each sample (plant and soil) and analyzed for total N with a combustion analyzer (*Costech analytical technologies Inc*., Valencia, CA, USA) in the Environmental Analytical Laboratory at MSU.

Derivation of Nitrogen Use Efficiency

A mass balance approach was used to determine an estimate for efficiency for each sample location. Measured soil mass percent total N (TN) and bulk density (g cm^{-3}) were multiplied to generate a bulk N concentration at each depth (g cm^{-3}) and multiplied by the depth (cm) of each interval to generate an areal density of soil TN (g cm^{-2}), which was converted to kg ha^{-1}. Crop N uptake (kg ha^{-1}) was derived by multiplying the percent TN in the plant aboveground biomass with the closest crop yield measurement measured by the harvester (kg ha-

^{1}), which was typically within 5 m of where plants were sampled. Nitrogen taken up by the plant was regarded as economically beneficial (efficient) use of N because N in the crop contributes to net-return in the season the crop was grown. While available N left in the root zone has the potential to be taken up by future crops and contribute to net-returns for farmers in upcoming years, it also has potential to be lost to leaching or atmosphere but was minimal post-harvest.

More importantly, knowledge of NUE requires knowledge of N mineralization from soil organic nitrogen (SON). Nitrogen use efficiency (NUE) was calculated using soil samples prior to fertilizer application and post-harvest, plant biomass samples at plant maturity, and estimates of other inputs to derive the NUE. Contrasting definitions of NUE in the literature, we defined NUE that included soil TN as NUE_{Soil}, defined here as outputs deemed efficient over the total amount of N inputs:

$$NUE_{Soil} = \frac{N_{crop} + N_t}{N_{inputs} + N_{t-1}}$$

(Equation 6.3)

where variables are defined in Table 6.4 and N_{inputs} were;

$$N_{inputs} = N_{dep} + N_{min} + N_{fert} + N_{fix}$$

(Equation 6.4)

where variables are defined in Table 6.4. Due to the nature of the crop-fallow rotation, N fixation is likely negligible, so N fixation was assumed to be zero. Total N deposition was estimated to be 2.92 kg N ha^{-1} year^{-1} based on the mean total N deposition from 2000 – 2018 recorded at the closest CASTNET station at Glacier National Park (U.S. EPA, 2018).

A key unknown in the equation of NUE_{Soil} is the N production via mineralization within fields. Mineralized N was modeled using a regression model derived from a review of multiple mineralization studies on the relationship between mineralization rate and percent total nitrogen (Vigil et al., 2002);

$$N_{min} = \ 10.8 + 304.6TN$$

(Equation 6.5)

where TN is the percent total nitrogen in the soil. Measured percent TN is close to SON and was used to estimate mineralized N (kg ha^{-1}) at each point, for every depth interval, and in both seasons. Mineralization of SON to plant available forms is a function of organic matter, moisture, and temperature (Booth et al., 2005); however, Equation 6.5 uses only percent soil TN to estimate mineralized N, introducing uncertainty into the mass balance approach (Vigil et al., 2002). Soil organic N is the dominant form of N in the upper soil profile, thus mineralization in most soils occurs predominately near the soil surface (Cassman & Munns, 1980). Crop N uptake is also greatest near the surface of soils due to mineralization of N and the surficial application of fertilizer N. Due to the likelihood that most processes that contribute to NUE occur predominately near the shallower depths, only soil samples from the top 30cm were used to calculate and model NUE.

Soil TN inventories (kg ha^{-1}) and mineralized N (kg ha^{-1}) from each season were summed over the top 30 cm at every point, respectively. The summed estimates of mineralized N from Equation 6.5 using percent TN in March and August were averaged to generate a mean estimate of total mineralized N in the top 30cm at each sample point. Crop N uptake measurements were co-located by sample point and as-applied nitrogen fertilizer rates were georeferenced to each sample point. Using the constant for N deposition defined above, NUE_{Soil} was calculated for every observed point in each field as:

$$NUE_{Soil} = \frac{N_{crop} + N_t}{N_{dep} + N_{min} + N_{fert} + N_{t-1}}$$

(Equation 6.6)

where all terms were in kg ha^{-1} and defined in Table 6.4, using summed soil TN inventories and mineralized N from the top 30 cm of soil from each sampling site during each season.

Based on uncertainty of soil TN measurements and the dominance of the soil TN inventories on the NUE_{Soil} calculation (Equation 6.6), NUE was redefined as;

$$NUE = \frac{N_{crop}}{N_{dep} + N_{min} + N_{fert}}$$

(Equation 6.7)

where all terms are defined above. This formulation calculates NUE as the ratio of N taken up by the crop to the amount of applied fertilizer N, N deposition, and estimated mineralized N. The calculated NUE from Equation 6.7 was the response variable used for developing a generalized model of NUE for conventionally managed rainfed winter wheat fields.

Nitrogen Use Efficiency Model Comparison

A collection of potential generalizable models for estimating NUE at a subfield scale were compared to explore the potential for low-cost methods of developing fertilizer application prescriptions. Development of a general model for predicting subfield NUE relied on environmental covariates gathered from Google Earth Engine (GEE), which were assumed to influence NUE (Table 6.5). All environmental covariate data were gathered from public domain GEE data repositories (Gorelick et al., 2017) and aggregated with soil sample data. Information from each covariate was extracted to each soil sample location to form the analysis dataset (Hegedus & Maxwell, 2022b; 2022c). Temporal data were gathered, delineated by year and constrained in range to the point in time that a farmer needs to make decisions on N fertilizer management of dryland winter-wheat in Montana (Hegedus & Maxwell, 2022c). Harvest data were collected for each field, but because they would not be available to make predictions at the point of a decision on N fertilizer management, were not included as covariate information for modeling NUE. The covariates selected for modeling were based on assumptions about environmental variables that influence the components of NUE and observed relationships from evaluation of observed NUE (Fig. 6.5). Grain yield and grain protein concentration were gathered from farms but were not included in the model to restrict variables used in the NUE model to data available in the public domain. The yield and protein circles are colored black because these data are not available, however light orange circles indicate remotely sensed covariate data that may benefit predictions of NUE and were included in the model selection process (Fig. 6.5).

A set of five candidate models were tested for their ability to predict NUE. Candidate models were chosen based on their potential for predictive and not inferential modeling because our objective was not to test direct causality between predictors and NUE. Model candidates were selected to represent simple to complex approaches with varying degrees of flexibility and assumptions. The models tested included (1) a baseline simple linear regression (SLR) model that only included as-applied nitrogen as a covariate, and (2) a non-linear model using an exponential decay function with as-applied N (NLR). The exponential decay function was selected based on exploratory analysis of the relationship between NUE and N sources and assumed that NUE could be predicted solely based on as-applied nitrogen, like the SLR model. Other models tested included all spatially variable covariates from Fig. 6.5 and were (3) a multiple linear regression model (MLR), (4) a generalized additive model (GAM), (5) a random forest regression machine-learning model (RF) and (6) a support vector regression machine-learning model (SVR). Random forest regression is an ensemble tree approach where covariates are randomly sampled in each tree and used to classify observations (Breiman, 2001). Support vector regression mapped observations to the feature space generated by the covariate data (Lau & Wu, 2008; Smola & Schölkopf, 2004).

Spatial covariance structures were incorporated into the models in varying ways to account for spatial autocorrelation of NUE observations, from a Gaussian process spatial correlation structure used in the SLR, MLR, and NLR, (De Bastiani et al., 2015; Diggle & Tawn, 1998; Mardia & Marshall, 1984; Xia et al., 2008), a Gaussian basis function on spatial coordinates in the GAM (Gotway & Stroup, 1997; Guisan et al., 2002; Holland et al., 2000; Zuur & Camphuysen, Kees, 2012), to common machine learning approaches for spatial data with the RF and SVR (Janatian et al., 2017; Langella et al., 2010; Walsh et al., 2017; Y. Wang et al., 2017).

All models except the SLR and NLR were subjected to features selection during the fitting process to optimize model performance. Top-down AIC based feature selection was performed for the MLR and GAM models, where a reduction of two AIC units with the removal of a covariate justified its omission. As AIC is not appropriate for machine learning methods, top-down RMSE based feature selection was performed for the RF and SVR models. Feature selection was based on a reduction in RMSE to determine whether withholding a covariate improved model performance. All models, except the SLR and NLR that only used as-applied N fertilizer, began feature selection with a full model that included the WHC classification, and the covariates illustrated in the yellow circles of Fig 6.5.

The initial candidate models were subjected to a hold-one observed farm field out cross validation approach to initially assess the ability of the model to predict NUE. Each model was independently trained on data from three out of four observed fields, and performance was tested by calculating the root-mean square error (RMSE) between the predicted and observed NUE from the field not used for training (Fig. 6.6). The RMSE across the four test datasets were averaged for each model and compared to identify the two candidate models that generated the lowest mean RMSE. The two models selected from the hold-one field out cross validation approach were then empirically assessed using 5x2 cross validation (Dietterich, 1998). All the data from the fields were pooled and randomly selected evenly into a training and test dataset. Both models were trained on the training dataset and RMSE was calculated based on predicted and observed NUE in the test dataset. Then the training and testing datasets were swapped (2 folds), and the process was repeated to generate four error metrics and calculate the variance of the difference in the predictions. Beginning with randomly splitting the data evenly, the process was repeated five times (5 folds), and the error estimates and variance from the five repetitions were used to calculate the 5x2 t-statistic (Dietterich, 1998). Assuming a t-distribution with five degrees of freedom, evidence against the null hypothesis that there was no difference in RMSE between observations and predictions from the two models was evaluated with a two-tailed significance test with a 95% confidence threshold ($\alpha = 0.05$).

Results

Nitrogen Use Efficiency

Crop N uptake. soil N, and NUE_{Soil} were highest on average in fields BS 2020 and BSE 2021, respectively (Table 6.6). BSE 2021 had the highest average mineralization, highest average soil N, and the highest mean experimental N fertilizer rates with N deposition assumed equal, (Table 6.6). The higher N fertilizer rates in BSE 2021 were due to an agreement made prior to 2021 with farmer B to discontinue the use of zero N rate experimental plots. The highest mean NUE_{Soil} was in MS 2021 (Table 6.6). However, NUE_{Soil} rates above 1 were unexpected, as efficiency is calculated as the proportion of N allocated to efficient uses over the total sources of N to the system, meaning NUE_{Soil} values greater than 1 indicate crop N uptake and post-harvest soil TN inventories exceeded all sources of N from the soil and pre-fertilization soil TN inventories.

Measured at the same georeferenced locations in March and August with a 1m resolution GPS, analyses for three fields provided little statistical evidence that soil TN varied at a given site between March and August in BS 2020 (t_{66} = -0.2516, p-value = 0.8022), WH 2020 (t_{59} = -0.2898, p-value = 0.773), and BSE 2021 (t_{68} = -0.2772, p-value = 0.7824). Field MS 2021, however, exhibited moderate evidence of variation in TN between the beginning and end of the production season (t_{74} = -2.0369, p-value = 0.0425). Most fields did not have statistically different soil TN between seasons and the total N measured in each field in March and August had a root mean square error of 58.08 Mg ha^{-1} (Appendix Fig. C1). A value of 58 Mg ha^{-1} is roughly 10-15% of measured soil TN, which is within the commonly expected uncertainty of soil total N measurements due to uncertainty in bulk density measures.

Two potential sources of variation in measures of soil TN from March and August are; 1) differences in percent TN measured by the Costech total combustion analyzer at corresponding points in the two seasons, or 2) differences in bulk density measurements at corresponding points in the two seasons due to measurement errors, possibly related to water content in the field (Appendix Fig. C1). The only field in which a statistical difference between percent TN or bulk density occurred was in BS 2020, where there was strong and very strong evidence that measures from March and August for percent TN (t_{66} = 2.488, p-value = 0.0154) and bulk density (t_{66} = -3.223, p-value = 0.0019) differed. Measured percent TN was greater in March compared to August while measured bulk density was greater in August compared to March in BS 2020. However, despite few differences observed in aggregated data across fields, most measurements varied between seasons, introducing the discrepancy between total soil N between March and August. The RMSE in percent TN and bulk density were 0.0137% and 0.1399 g cm^{-3}, which divided by the mean percent TN and bulk density indicated about 10% uncertainty between measures from the two seasons for both metrics (Appendix Fig. C2), consistent with expected measurement uncertainty for these parameters. Propagation of uncertainty to measures of soil TN in kg ha^{-1} yielded 14% relative uncertainty in soil TN measurements and indicated that the inventories during the two seasons were the same within fields uncertainty.

Despite the relatively small degree of uncertainty between soil TN inventories from the two seasons, small changes in soil TN had a large effect on NUE calculations. Converted to kg ha^{-1} as used in the NUE equation, a mean discrepancy between TN in March and August of up to 58,000 kg ha^{-1} had a large effect on calculated NUE because it obscures any difference between crop N uptake and mineralized or as-applied N, which occur on a magnitude rarely exceeding

100 kg ha^{-1} (Table 6.6). For example, if crop N uptake was 34 kg ha^{-1} and mineralized N, fertilizer N, and N deposition inputs were 70, 74, and 2.92 kg ha^{-1}, respectively, NUE using only those components would be 0.24. However, if the soil TN in the top 30cm was 328,160 kg ha^{-1} in March and 334,414 kg ha^{-1} in August, NUE_{Soil} defined by Equation 6.6 is 1.02. This calculation of NUE might be interpreted to mean that 6,254 kg ha^{-1} of soil TN were gained between March and August but is within uncertainty of the TN measures. This is consistent with understanding of SON dynamics, where detectable gains occur over timescales of decades to centuries due to rates of mineralization that yield amounts of available N that are small relative to SON, but on par with N fertilizer additions.

Nitrogen deposition was negligible relative to other nitrogen sources and sinks (Table 6.6), leaving NUE, defined in Equation 6.7, essentially a ratio of crop N to fertilizer and mineralized N, with the latter two of comparable magnitude. The distribution of NUE under the assumption that soil TN was invariant over the time scale of observations yielded more realistic estimates of NUE, where WH 2020 was observed to have the highest mean NUE compared to the other fields (Table 6.7). The discrepancies in percent TN measurements in March and August were less influential to calculation of NUE because estimated N mineralization was averaged across the two seasons at each point. Additional uncertainty of modeling N mineralization due to the interaction between soil microbiota, moisture and temperature was recognized, but due to the importance of mineralized N in small grain agroecosystems of Montana (John et al., 2017; Sigler et al., 2018), mineralized N was retained in the final derivation of NUE (Equation 6.7). Equation 6.7 was used to calculate NUE at all 75 points within each field for both years. Sample points that did not have corresponding measurements at each depth interval to 30cm in both March and

August were omitted to generate consistent estimates of mineralized N at each point, yielding observations at 67, 69, 60, and 75 out of 75 sample locations in BS 2020, BSE 2021, WH 2020, and MS 2021, respectively.

To make an initial assessment of efficiency common in the literature, FUE was calculated:

$$FUE = \frac{N_{crop}}{N_{fert}}$$

(Equation 6.8)

where only measured crop N uptake and as-applied N rates are used to define efficiency. Characterization of the relationship between crop N uptake and N fertilizer using a generalized additive model with a smoothing term using cubic shrinkage splines for as-applied nitrogen showed an asymptotic relationship (Fig. 6.7), indicating that crop N uptake is maximized as fertilizer rates increase and excess fertilizer has the potential to be lost from the system.

Efficiency was expected to decline with increasing N fertilizer rates due to the asymptotic nature of crop N uptake and fertilizer. A negative exponential relationship between FUE and as-applied N fertilizer was observed, where low N rate samples had the largest calculated FUE before plateauing, like crop N uptake, at higher N rates (Fig. 6.8).

Using the calculation of NUE defined in Equation 6.7, a decrease in NUE was observed with increasing as-applied N fertilizer, like the relationship between FUE and fertilizer albeit less distinct (Fig. 6.9). The relationship was more muted for NUE than for FUE because the denominator of the efficiency calculation decreases the proportion of crop N uptake to N inputs. When mineralized N is included in the calculation for efficiency, it compensates for a lack of N fertilizer at low rates (Fig. 6.10). Efficiency is lower in soils with more TN because estimated mineralization is higher.

Soil TN (or mineralized N which was directly calculated from soil TN) did not appear to be related to fertilizer N. We assumed that mineralization was not influenced by fertilization, though the influence of this interaction could be embedded in the observed yield data. However, similar to the relationship between NUE and fertilizer N, areas with low soil TN were also the most efficient due to lower estimated N mineralization and total N availability (Fig. 6.11). Areas of the field with low N fertilizer or soil N, but not necessarily both, indicate places where there were relatively limited amounts of available N, resulting in variable efficiency. As mineralized N and fertilizer N increase, the total amount of N inputs increases, but efficiency decreases because the crop N uptake does not increase at the same rate. While efficiency increased with a reduction of N availability due to some combination of lower N fertilizer rates or lower N mineralization, the location of lower fertilizer and mineralization rates within a field did not always coincide due to the randomization of variable N fertilizer rates.

Model Evaluation

The SVR had the lowest mean RMSE, and the RF had the second lowest mean RMSE (Table 6.8). However, for each field the model resulting in the most accurate predictions of NUE was different (Table 6.8). Due to the objective of creating a predictive model of subfield NUE that can be used across dryland winter-wheat agroecosystems in Montana, the two best models on average across the fields, the SVR and RF were compared by 5x2 CV.

Compared to the RF model, the SVR had a lower mean RMSE across the 5x2 CV of 0.0013 units, however there was little to no evidence against the null hypothesis that RMSE did not differ between the RF and SVR models (t_5 = -0.1983, p-val = 0.8506) and both modeled the relationship of NUE and as-applied fertilizer N well (Fig. 6.12). Although both the SVR and RF produced similar predictions and there was no statistical difference between them, but the conclusion was that SVR provided the best general model of NUE because the lower mean RMSE from the HOOCV analysis indicated that the SVR performed better than the RF when making predictions in fields unseen by the model during training. All the sampled data were compiled as a training dataset to fit a final SVR model. After top-down features selection, the SVR model that provided the best predictions of NUE using environmental and management (e.g., N fertilizer rates) covariate data was:

$$NUE \sim f(N, WHC, tpi, aspect_{sin}, aspect_{cos}, prec_{cy}, prec_{py}, ndvi_{cy}, ndvi_{py}, ndvi_{2py}, ndwi_{cy}, ndwi_{py}, ndwi_{2py}, bulkdensity, claycontent, watercontent\ phw, carboncontent) \quad \text{(Equation 6.9)}$$

where terms are defined in Table 6.9 with descriptions of the data sources in Table 6.5.

Discussion

This research contributes a methodology for evaluating the nature of within-field variation in NUE and developing a predictive model based on low-cost data to aid with N fertilizer management decisions. The approach is adaptable to any system with any input that can be spatially varied and where detailed sampling of representative fields across an area of interest is conducted. Despite the clear importance of other controls creating substantial scatter, measured uptake of N by a crop generally had a saturating asymptotic relationship with the amount of N fertilizer applied (Fig. 6.7), indicating that efficiency generally decreases with applied N fertilizer (Fig. 6.8, Fig. 6.9). Additionally, this implies that while the crop uptake was low at low N fertilizer rates, the crop exceeded fertilizer inputs, suggesting the presence of non-fertilizer N contributions to crop N. At these low levels of applied N, distinct responses among fields suggest distinct capacities to supply non-fertilizer N, supporting the assumption that mineralization supplies more available N in locations with higher TN, reducing efficiency for a given applied rate.

The relationship of efficiency across N rates was consistent for both NUE and FUE (Fig. 6.8, Fig. 6.9). A similar relationship between NUE and mean total soil N (Fig. 6.11) indicates that efficiency was greatest where N availability was lowest due to relatively low TN. The relationship between NUE and N sources (fertilizer or soil/mineralized N) mirrors the relationship between crop uptake and N fertilizer (Fig. 6.7) where low crop uptake at low N fertilizer rates corresponds to the highest efficiency. While farmers are not expected to cease applying N fertilizer rates, and have little control over mineralization of N, the relationship of NUE and crop uptake with as-applied N fertilizer provides an opportunity to elucidate N rates that balance the tradeoff between net-returns and efficiency, which save producers' money on fertilizer and prevents the overapplication of N.

The SVR model developed from samples taken from the four fields demonstrates a method for providing farmers with predictions of NUE that can be used to identify the N rates that optimize crop N uptake and NUE. The NUE of fields varied spatially and in magnitude (Table 6.7), indicating that management of NUE would be suited for OFPE, where data from a field is used to make decisions on that field. However, the cost and time expenditures to gather and analyze the data to develop field specific NUE models is not feasible for a farmer to accomplish in addition their normal operations. The similar relationships across fields between as-applied N fertilizer and crop N uptake, FUE, or NUE (Fig. 6.7, Fig. 6.8, Fig. 6.9) indicate the potential for a generalized model of NUE trained on these fields to be applied across similar fields.

In Montana agroecosystems, with adequate moisture and optimal temperatures, microbes mineralize organic N into plant available forms which increases the amount of available N for crop uptake (Weil & Brady, 2017). This is important to account for in calculations of NUE for these systems and is especially evident when low or zero N fertilizer rates are applied (Fig. 6.10). Moreover, N mineralization reflects a chronic loss of SON that should be considered in evaluating pollution potential of agroecosystems. Uncertainty in the evaluation and modeling of NUE could be reduced by improving estimates of N mineralization, or by using direct measures. Measuring mineralization of N requires intensive and expensive sampling and retains inherent uncertainty due to the interactions between microbes, moisture, temperature, and N availability. Thus, modeling mineralization from measured percent TN is also likely to be inaccurate due to variable soil moisture and temperature conditions not incorporated into Equation 6.5; however, this approach captures the likely overall variation in mineralization and is too important in Montana agroecosystems to ignore. Mineralization rates that account for soil temperature and moisture conditions, in addition to SON, would improve these estimates. Possible avenues here include use of remotely sensed soil moisture interpolated for depth effects, and upscaled modeling of soil climate based on rainfall and temperature patterns and calibrated using the growing soil climate observation network.

An additional limitation of this study is the sampling of N in above ground biomass without explicit regard to N in the roots. However, by mid-growing season, winter-wheat remobilizes N from stems, leaves, and roots to grain kernels, so the measurements of crop N in above ground biomass taken at the end of June reflected N that had been re-allocated from roots to the above ground winter wheat biomass (Kätterer et al., 1997). If the crop N is of a similar magnitude to the fertilizer N, one can assume that the total amount of available N is likely to exceed the crop nitrogen needs due to likely mineralization of organic N. When crop N is higher than fertilizer N, the crop has tapped into mineralized N sources to satisfy its N requirements, and greater efficiency is achieved. Conversely, if crop N is lower than fertilizer N, one can assume the presence of N inefficiencies resulting from N loss, or that there were constraints on N uptake by the crop, resulting in amplified inefficiency of N use.

Fields used in this study were not randomly selected from all conventional dryland farms in Montana yet were representative of conventionally managed winter-wheat agroecosystems in Montana, thus inference and predictions are relevant to fields not selected in the study. However, fields were only sampled in two weather conditions, either 2020 or 2021, which introduces increased uncertainty into predictions of NUE in conditions dissimilar to 2020 and 2021. This is important because 2021 was a major drought year in Montana. Increasing the number of fields sampled would increase confidence in the applicability of models trained on these data to other fields. Repeated sampling of more fields in more years would provide data representative of a range of weather conditions and field characteristics to further develop a NUE model. Future research to downscale and ground truth remote sensing covariate data is also needed to reduce uncertainty based on the accuracy and scale of the data used to train NUE models.

As the medium that supplies N to plants, soil characteristics play a role in influencing N dynamics (Sigler et al., 2018, 2020). Soil texture influences the rate and ease of movement of water in soil. Fine textured soils have smaller and less-connected pores than coarse textured soils, which limit water and nitrate movement in the profile (Weil & Brady, 2017). Fine textured soils to a point can increase nitrogen efficiency by retaining water and nitrogen in the profile for use by crops, subsequently reducing leaching rates, but can also reduce efficiency if denitrification is substantial. Fine textured soils, particularly those with high clay and soil organic carbon contents, could decrease NUE if drought makes nitrogen held in the soil matrix unavailable for crop uptake. The influence of fine textured soils on NUE was captured in the SVR model through the incorporation, and retention after feature selection, of clay and soil organic carbon content estimates from OpenLandMap (Equation 6.9). Additionally, because fine textured soils hold more water and N, but can limit oxygen, NUE further decreases when denitrification reduces nitrate to gaseous forms that are lost to the atmosphere (Sigler et al., 2022). The importance of soil water storage, dictated by soil texture, on N transport implies that WHC would describe NUE (Sigler et al., 2018, 2020). Although the estimate of WHC used to stratify sampling was only associated with a difference in NUE between high and medium zones, the WHC classification provided information relevant to the SVR for characterizing NUE (Equation 6.9) and could be refined in future efforts.

Information on soil characteristics related to soil structure and texture was also indirectly incorporated in the SVR through the inclusion of bulk density (Equation 6.9), as the density of the soil is influenced by soil structure and porosity. The SVR model also captured the influence of soil pH on NUE (Equation 6.9), and especially FUE, via influence on volatilization rates, which are greater in high pH soils.

Due to the linkage of water and temperature with N mineralization, weather plays a significant role in NUE (Thilakarathna et al., 2020). Soil water may increase efficiency by suppressing volatilization of fertilizer, yet excess moisture can increase nitrate leaching and denitrification, which may reduce efficiency (Engel, 2012; Sigler et al., 2022; Sigler et al., 2018). Unsurprisingly, precipitation and water content estimated from OpenLandMap was retained in the SVR model as a covariate that decreased variation between predicted and observed NUE (Equation 6.9). As a result of feature selection based on the contribution of covariates to model performance, the SVR model only utilized precipitation data collected up to March 1st from the year in which samples were collected and not precipitation from prior years, indicating that estimates of NUE were influenced by recent precipitation. The combination of temperature and precipitation influence photosynthesis of the crop, which in turn influences the N efficiency of crops (Michaletz et al., 2016; Pearcy et al., 2005). At high temperatures or when soil moisture is low due to a lack of precipitation, photosynthesis decreases as the crop balances water loss and carbon fixation (Reich, 2014). When stomata are closed, the soil-plant-atmosphere continuum is disconnected, and the rate that soil water containing nitrate flows to roots and into the crop decreases (Lambers & Oliveira, 2019). Thus, at high temperatures or during droughts when photosynthesis decreases, crop uptake of N also decreases, and NUE subsequently declines. Information on the water status of the crop were derived from NDWI data, which generated more accurate NUE predictions when included in the SVR model than when omitted (Equation 6.9). Unlike precipitation, NDWI data from the two years prior to when samples were collected were informative toward predicting NUE, while information on the water status of the crop in the year samples were collected was omitted from the SVR model based on feature selection.

Information on the productivity of the crop, measured by NDVI, was retained in the final SVR model post-feature selection only for years in which the crop was grown (Equation 6.9). Due to the crop fallow rotation, the crop was grown in the year samples were taken and two years prior. This indicates that the SVR model utilized the measure of the crop greenness and indications of fallow to predict NUE. The greenness of a crop is related to crop productivity and biomass and can indicate when a field was in fallow and provide information on the amount of stubble post-harvest. High in organic C, the stubble increases C:N ratios of the soil, influencing mineralization and subsequent NUE (Watson et al., 2002). Additional topographic parameters such as slope, aspect, and TPI were included in the final SVR model (Equation 6.9), likely due to the influence of site characteristics on fine scale water and N transport, nutrient availability, and microclimate conditions (Grant et al., 2016).

Recognized as imperfect, the SVR model demonstrates the development process of subfield NUE models that have the potential to be incorporated into decision support systems to provide farmers with N management recommendations that optimize fertilizer based on both efficiency and profit. The field of data science can contribute to developing NUE models by providing a suite of tools from which the most appropriate model for the right system can be selected.

Evaluation and modeling NUE not only inform spatial N fertilizer management but has potential for temporal N management as well. The proportion of mineralized N to fertilizer N taken up by the crop depends on the growth stage and the amount of supply by either mineralization or fertilization. This highlights the potential for optimizing fertilizer application timing to conserve either SON or N fertilizer. Efficiency was greatest with low available N (Fig. 6.8, Fig. 6.9, Fig. 6.11), so mineralization or fertilizer applications early in the growing season contribute to lower NUE or FUE because the supply exceeds demand, leaving N vulnerable to leaching as nitrate or denitrification (Ravier et al., 2017; Yin et al., 2020). Applying N fertilizer early in the growing season thus further reduces NUE due to the lack of demand by the crop. When fertilizer is applied at a rate equivalent to plant needs, any mineralized N thus contributes to a lower NUE as N saturation in the crop leads to excess N available in the soil, thus systems with significant mineralization require less supplemental N from fertilizer. The interaction between the crop growth stage and N availability also means that producers can optimize their sowing rates with fertilizer application rates, where fertilizer is applied at an adequate amount of time after sowing to ensure that the fertilizer is taken up by the crop (Yin et al., 2020).

Increased effort in the downscaling and ground truthing of remote sensing data can decrease uncertainty in NUE predictions. Knowledge of the spatial variation of NUE across fields can highlight areas of fields that require differing management strategies to maximize NUE, thus models that adequately predict the spatial variation of NUE are crucial to decision support systems. When applying NUE models in data-driven decision making, users must be willing to accept current uncertainties in N mineralization and remote sensing covariate data to begin shifting conventional N management towards greater sustainability in terms of both economic gains and efficiency. Improvements in these arenas warrant future research and updating of initial methods that incorporate efficiency into decision making, yet the pressing global need to increase production without detriment to the resource base of agronomic production requires increasing agronomic input efficiency with the tools and technology available now.

Conclusion

Nitrogen that is taken up by crops is an efficient use of managed N that benefits farmer's net-returns. Processes of crop nutrient uptake and those that remove N from the crop-soil system include denitrification, leaching, volatilization, and surface runoff, and can be considered an economic loss to farmers that contribute to N pollution, thus reducing efficiency. Evaluation of NUE in dryland winter-wheat systems of Montana indicated decreasing efficiency with increasing N fertilizer application rates. As N fertilizer is an input managed by farmers, variable N fertilizer management presents an opportunity to directly influence NUE on a subfield scale. The investigation of NUE in the four fields sampled also highlighted the importance of N mineralization in these systems, evident in calculated FUE values greater than one, which indicate crop uptake in excess of N fertilizer applications. Higher NUE was observed with low available N. Subfield information on NUE informs farmers about where applying N fertilizer decreases NUE and does not improve crop N uptake. A support vector regression model provided predictions of NUE most similar to NUE estimated from soil and plant biomass measurements. The predictive NUE model incorporated weather, soil, crop, and site characteristic data with ecological or biogeochemical links to NUE. The development of the SVR model for predicting NUE demonstrates a potential avenue for including information about efficiency into a farmer's decision matrix for variable N fertilizer rates.

Acknowledgements

This research was supported by a WSARE Graduate Student Fellowship titled “Nitrogen Fertilizer Management Based on Site-Specific Maximized Profit and Minimized Pollution” (Award Number GW19-190), a USDA-NIFA-AFRI Food Security Program Coordinated Agricultural Project, titled “Using Precision Technology in On-farm Field Trials to Enable Data-Intensive Fertilizer Management,” (Accession Number 2016-68004-24769), the USDA-NRCS Conservation Innovation Grant from the On-farm Trials Program, titled “Improving the Economic and Ecological Sustainability of US Crop Production through On-Farm Precision Experimentation” (Award Number NR213A7500013G021), and the Montana Fertilizer Advisory Council from 2016 - 2021 , and the Consortium for Research on Environmental Water Systems (CREWS) National Science Foundation EPSCoR Cooperative Agreement OIA-1757351.

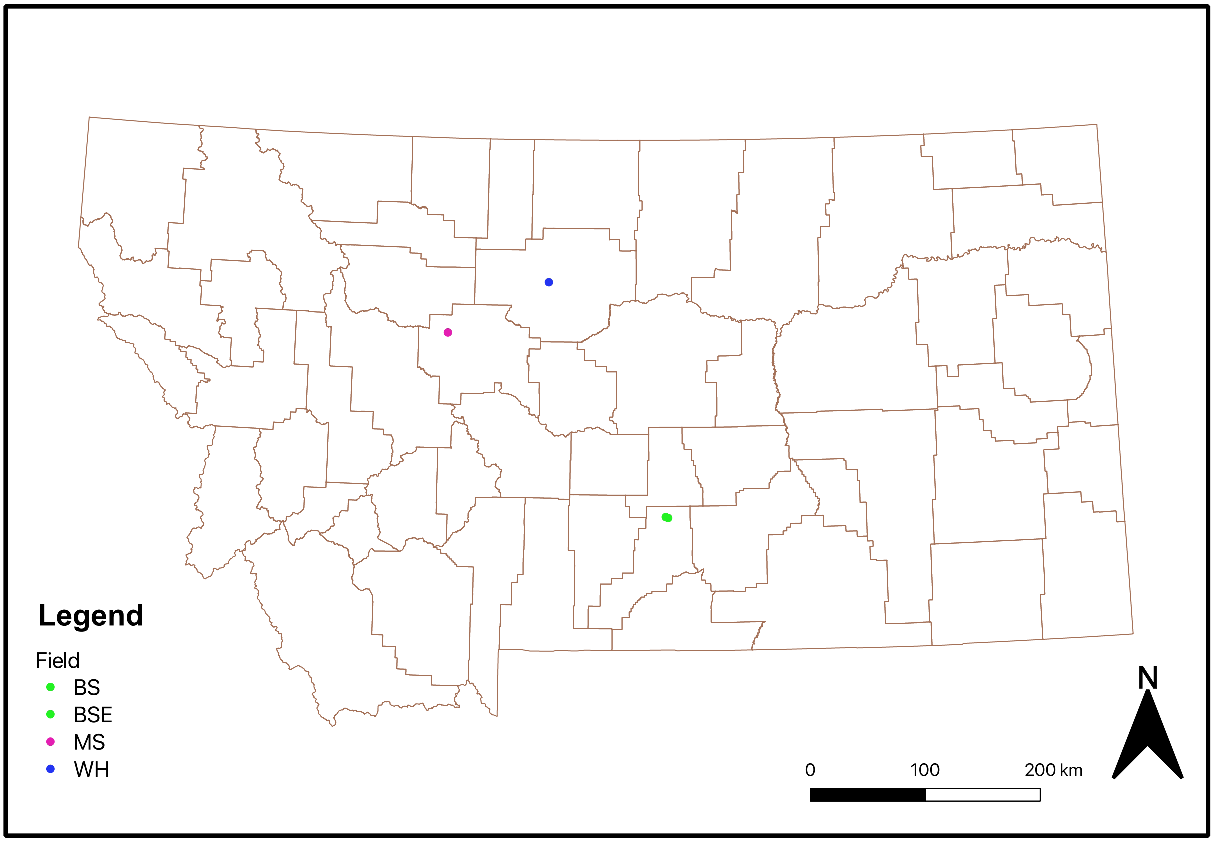


Figure 6.1. Fields sampled during 2020 and 2021. Two fields from Farm B were sampled (green), one in each year, with a corresponding field in Farm W (2020) or M (2021).

Table 6.1. Collaborator farm ID (Fig. 1), field names, field size (acres), crop harvest history, years of N experimentation.

Farm	Field	Field size (ac)	Crop History[1]: 2014 / 2015 / 2016 / 2017 / 2018 / 2019 / 2020 / 2021	Year Sampled
B	BSE	79	SF / WW / CF / WW / CF / WW / CF / WW	2021
B	BS	64	SW / CF / WW / CF / WW / CF / WW / CF	2020
M	MS	46	CF / WW / CF / WW / CF / WW / CF / WW	2021
W	WH	48	WW / SW / WW / CF / WW / CF / WW / CF	2020

[1] WW = winter wheat, CF = chemical fallow, P = peas, SW = spring wheat, SF = safflower, NA = Not Available

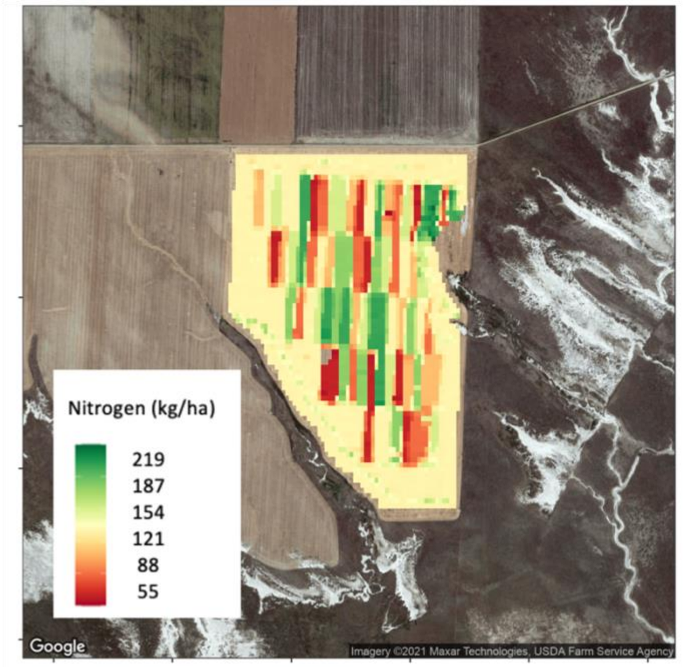


Figure 6.2. Example of a nitrogen fertilizer experiment applied by the OFPE project on field B1 in 2021, mapped as as-applied nitrogen fertilizer (kg ha^{-1}) rates from variable rate application equipment.

Table 6.2. Satellite programs used for collecting NDVI imagery.

Program	Resolution	Years Available	Years Collected
Landsat 5	30m	March 1984 – May 2012	1999 – 2012
Landsat 7	30m	January 1999 – present	2012 – 2013
Landsat 8	30m	April 2013 – present	2014 – 2021
Sentinel 2	10m	June 2015 – present	2016 – 2021

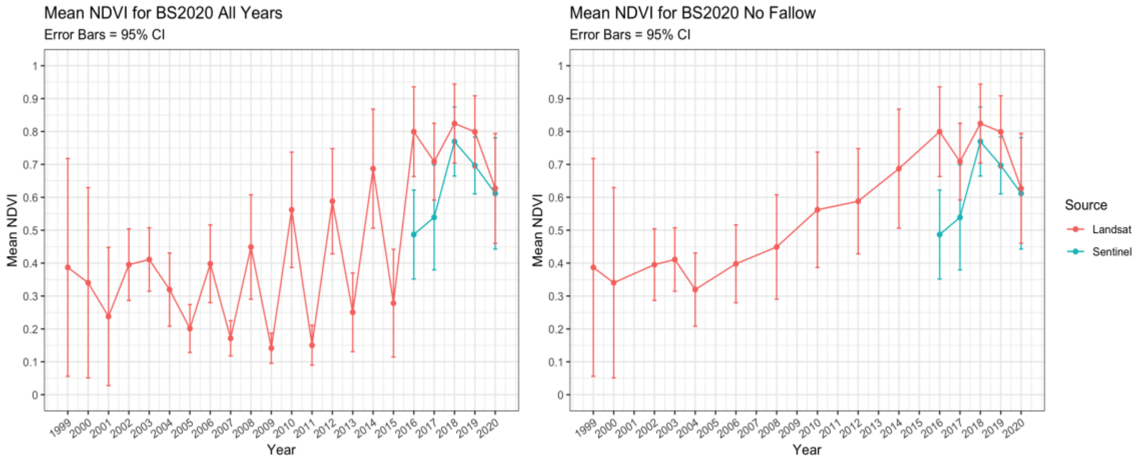


Figure 6.3. Mean NDVI on field BS every year including fallow or crop rotation years (a), and fallow years omitted (b) where error bars indicate spatial variation across the field each year.

Table 6.3. Classification scheme of water holding capacity (WHC) for each pixel (30 m by 30 m) based on the mean and coefficient of variation (CV) of normalized difference vegetation index (NDVI). Each pixel is classified based on the mean and CV across the 20 years of data collected.

	Low Mean NDVI	Medium Mean NDVI	High Mean NDVI
High CV	Medium WHC	Medium WHC	Medium WHC
Medium CV	Low WHC	Medium WHC	High WHC
Low CV	Low WHC	Medium WHC	High WHC

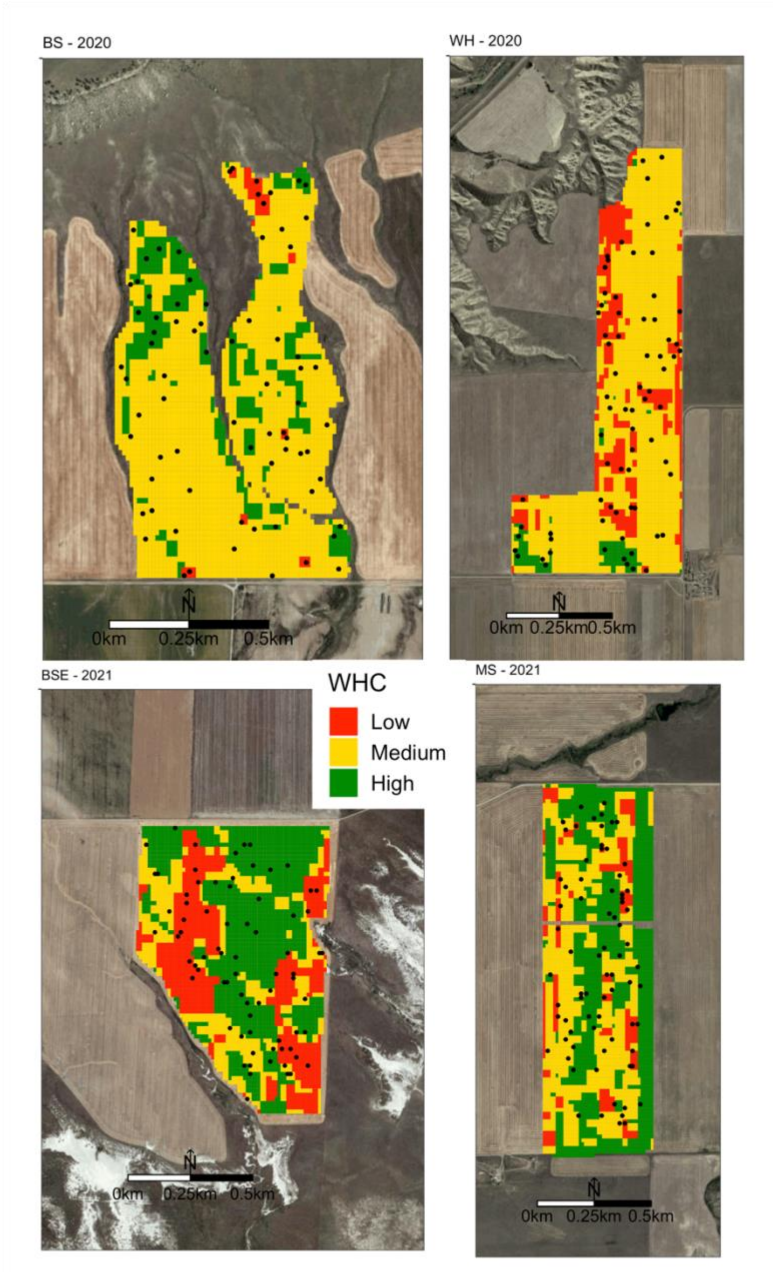


Figure 6.4. Water holding capacity classifications and sampling locations for each field.

Table 6.4. Definitions for parameter used in the calculation of nitrogen use efficiency.

Parameter	Definition
N_{crop}	N taken up by the crop measured from plant biomass samples
N_t	Amount of N in the soil after harvest measured by soil sampling
N_{t-1}	Amount of N in the soil prior to N fertilizer application in the spring measured by soil sampling
N_{dep}	N deposited from the atmosphere
N_{min}	Mineralized N
N_{fert}	Inputs of N fertilizer gathered from the farmer's as-applied N fertilizer map
N_{fix}	N fixed from the atmosphere

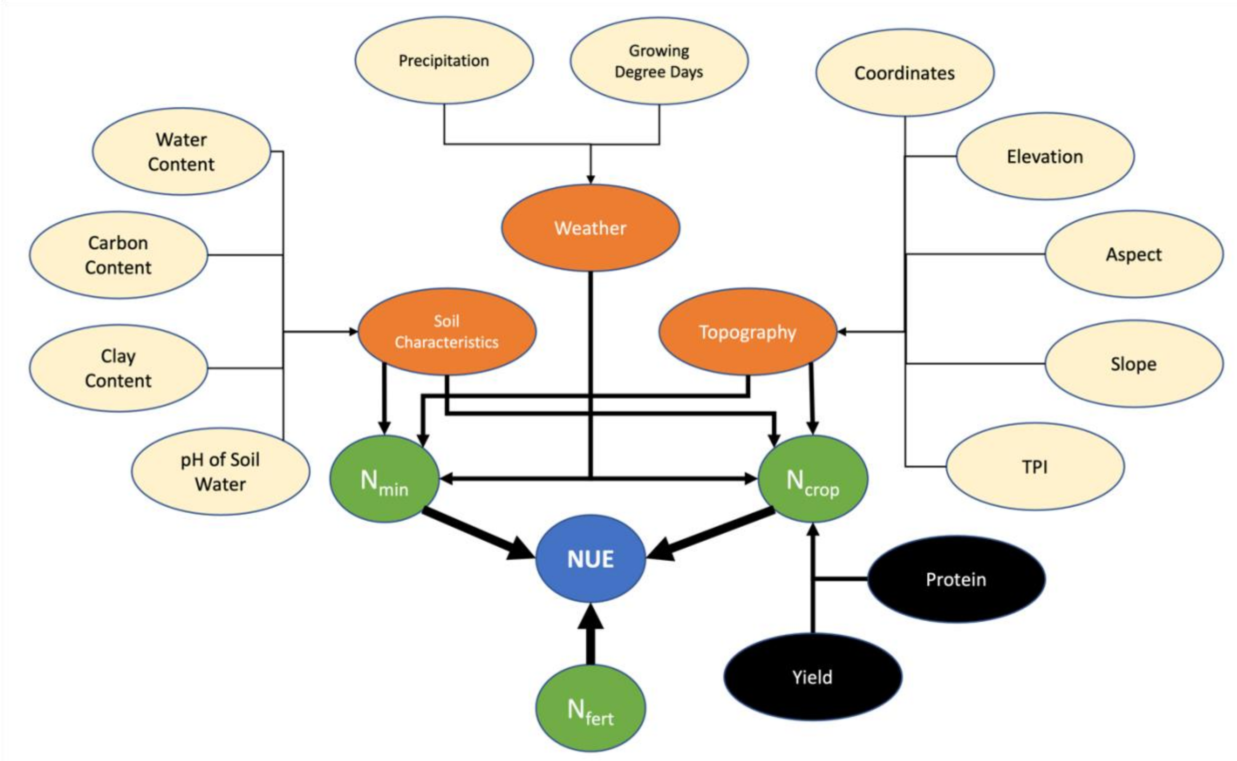


Figure 6.5. Conceptual model of the development of a general NUE model for Montana dryland agroecosystems. Green circles are variables that are used to directly calculate NUE, thus measuring these variables would provide the best characterization of true NUE. Orange circles are general data types that influence the components used to calculate NUE (green). Yellow circles are data types organized by general data type (orange), that are measured from remote sensing sources. The thickness of connecting arrows represents the strength of the relationship between the variable and NUE, where a variable is more directly related to NUE with thick arrows and more indirectly related with a thin arrow. The strength of a relationship between data and NUE decreases as distance between data and the blue NUE circle increases. Yield and protein would be beneficial data types for predicting crop N uptake and subsequent NUE but are not available data when a farmer models NUE to derive optimal N fertilizer rates.

Table 6.5. Table of covariate data types gathered from Google Earth Engine to enrich the crop yield and protein datasets gathered from on-farms. In some cases, multiple sources are used, however only one data source is used when aggregating to yield and protein.

Data Type	**Data Source(s)**	**Resolution**	**Years Collected**	**Description**
Normalized Difference Vegetation Index (NDVI)	Landsat 5/7/8	30m	L5: 1999-2011 L7: 2012-2013 L8: 2014 - present	Landsat is an ongoing USGS and NASA collaboration. Bands (NIR, red) L5/L7: B4 and B3 L8: B5 and B4
Normalized Difference Water Index (NDWI)	Landsat 5/7/8	30m	L5: 1999-2011 L7: 2012-2013 L8: 2014 - present	Landsat is an ongoing USGS and NASA collaboration. Bands (NIR, red) L5/L7: B2 and B4 L8: B2 and B5
Elevation	USGS NED	~10m (1/3 arc second), ~23m (3/4 arc second)	1999-present	USGS National Elevation Dataset. Measured in meters.
Aspect	USGS NED	~10m (1/3 arc second), 30m	1999-present	Direction the surface faces, function of neighboring elevations, in radians. Also calculated for each E/W and N/S direction as cosine and sine.
Slope	USGS NED	~10m (1/3 arc second), 30m	1999-present	Rate of change of height from neighboring cells, in degrees. Measured in degrees.
Topographic Position Index (TPI)	USGS NED	~10m (1/3 arc second), 30m	1999-present	Measure of divots and low spots as a function of neighboring elevation.
Precipitation	DaymetV3	1km	1999-present	Estimates from the NASA Oak Ridge National Laboratory (ORNL). Measured in mm.
Growing Degree Days (GDD)	DaymetV3	1km	1999-present	Estimates from the NASA Oak Ridge National Laboratory (ORNL).
Bulk Density	OpenLandMap	250m	1999-present	Soil bulk density (fine earth) 10 x kg / m3 averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
Clay Content	OpenLandMap	250m	1999-present	Clay content in % (kg / kg) averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
Sand Content	OpenLandMap	250m	1999-present	Sand content in % (kg / kg) averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
Soil Water pH	OpenLandMap	250m	1999-present	Soil pH in H_2O averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
Water Retention Against Drainage	OpenLandMap	250m	1999-present	Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted and averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
Soil Organic Carbon Content	OpenLandMap	250m	1999-present	Soil organic carbon content in x 5 g / kg averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).

Case	Fields/Years			
A	WH 2020	BS 2020	BSE 2021	MS 2021
B	WH 2020	BS 2020	BSE 2021	MS 2021
C	WH 2020	BS 2020	BSE 2021	MS 2021
D	WH 2020	BS 2020	BSE 2021	MS 2021

Train

Test

Figure 6.6. Hold-one field out cross validation scheme where one field is held out as a test dataset while three fields are used to train models in each of the four cases.

Table 6.6. Summary statistics for each component of NUE_{Soil} calculated in Equation 6.6. Bolded lines separate parameters representing efficient N fates, sources, and the response (NUE_{Soil}). Parameters are defined in Table 6.4. Note that units of N in the soil are in Mg ha^{-1} below, although NUE was calculated with all components in kg ha^{-1}.

	N_{crop} (kg ha^{-1})	**N_t (Mg ha^{-1})**	**N_{dep} (kg ha^{-1})**	**N_{fert} (kg ha^{-1})**	**N_{min} (kg ha^{-1})**	**N_{t-1} (Mg ha^{-1})**	**NUE_{Soil}**
				BS 2020			
Min.	13.73	94.84	2.92	0	27.55	102.6	0.59
1st Q.	29.31	305.12	2.92	53.46	74.91	297.3	0.91
Median	41.14	383.23	2.92	100.86	85.57	373.9	1.02
Mean	44.56	354.03	2.92	95.58	80.12	352.3	1.01
3rd Q.	55.28	419.92	2.92	134.23	93.18	423.6	1.12
Max.	139.94	554.01	2.92	163.06	109.93	508.9	1.50
				BSE 2021			
Min.	8.56	226.4	2.92	60.64	48.88	148.1	0.69
1st Q.	25.5	434.1	2.92	104.33	90.14	437.2	0.92
Median	34.01	466.6	2.92	131.76	93.18	468.4	0.99
Mean	35.26	459.2	2.92	129.43	91.31	457.1	1.02
3rd Q.	43.83	498.1	2.92	154.58	96.99	501.6	1.10
Max.	79.36	617.8	2.92	198.18	106.89	557.1	1.58
				MS 2021			
Min.	9.25	209.5	2.92	0	42.78	178.3	0.75
1st Q.	17.97	417.5	2.92	11.58	85.57	426.7	0.95
Median	23.05	483.2	2.92	23.79	91.66	460.5	1.03
Mean	23.08	458.2	2.92	31.7	85.71	444	1.04
3rd Q.	26.84	520.4	2.92	41.12	95.47	499.4	1.14
Max.	51.72	739.1	2.92	89.99	109.93	647.7	1.41
				WH 2020			
Min.	8.76	129.8	2.92	5.027	33.65	136.5	0.51
1st Q.	28.94	248	2.92	18.86	65.77	248.2	0.94
Median	34.74	307.9	2.92	40.78	73.38	298.9	1.02
Mean	37.26	295.1	2.92	54.93	70.6	293.4	1.02
3rd Q.	45.67	350.3	2.92	77.03	81.76	344.8	1.08
Max.	72.03	460.9	2.92	137.96	99.27	453.5	1.48

Table 6.7. Summary statistics of NUE calculated from Equation 6.7 without soil TN for each field.

Field	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
BS 2020	0.071	0.175	0.232	0.255	0.295	0.789
BSE 2021	0.029	0.122	0.152	0.167	0.203	0.470
MS 2021	0.080	0.156	0.184	0.200	0.234	0.431
WH 2020	0.066	0.224	0.288	0.312	0.354	0.739

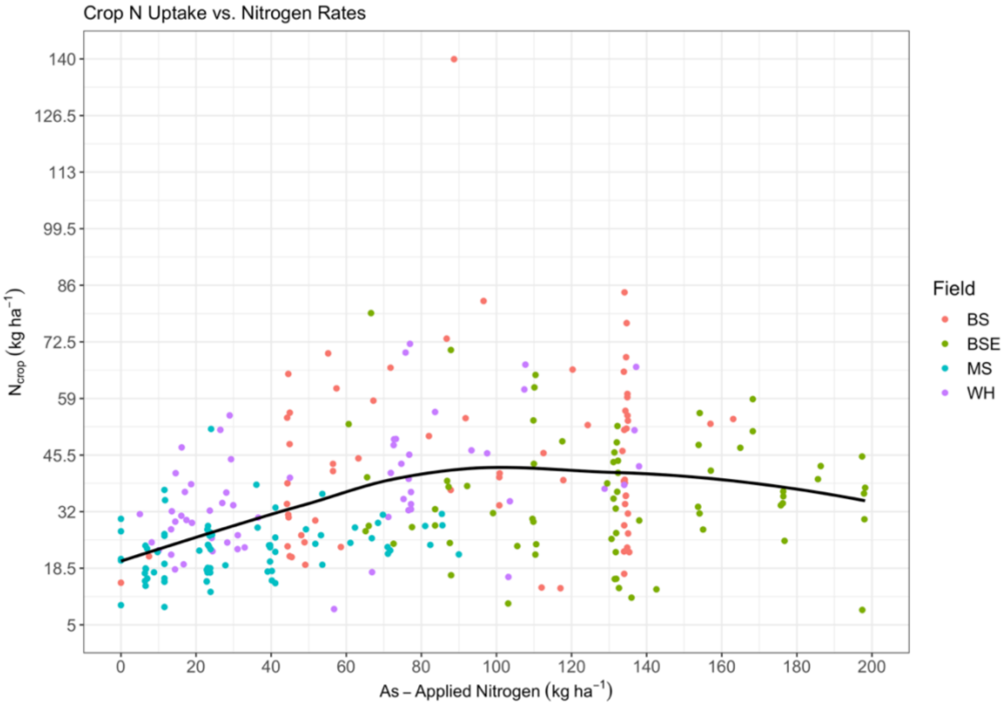


Figure 6.7. The relationship between crop N uptake measured from Costech total combustion analysis of above ground plant biomass samples and as-applied nitrogen measured from the farmer's VRA equipment. Colors represent measurements in each field and the black line shows the trend smoothed across all fields.

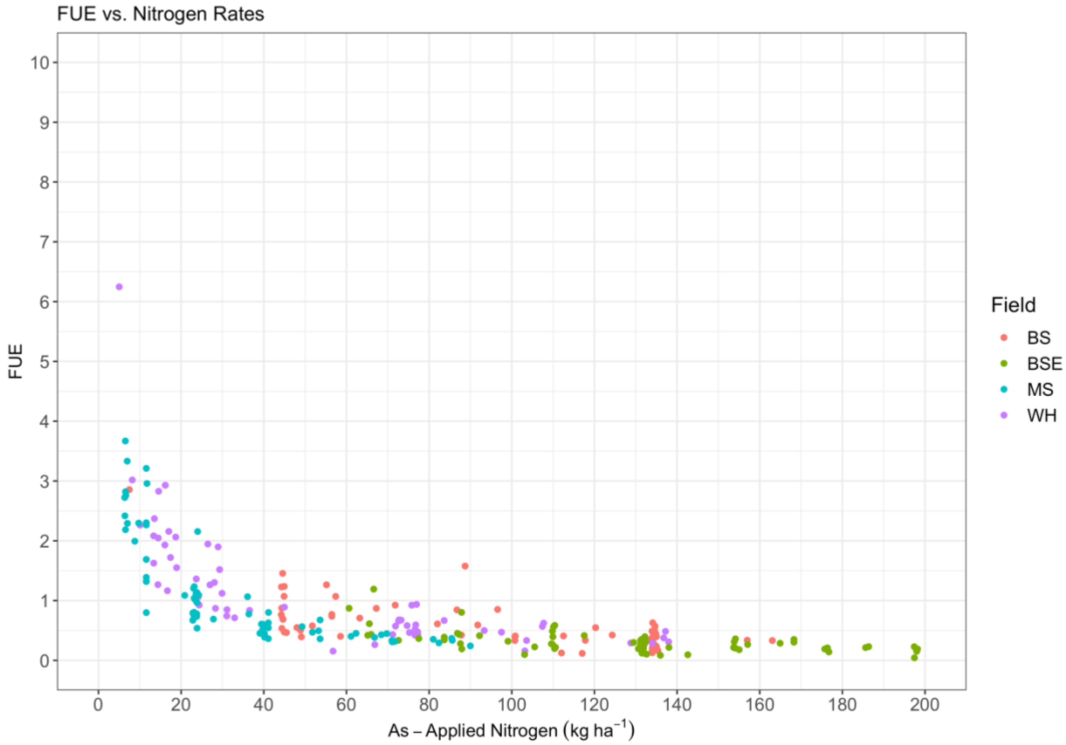


Figure 6.8. The relationships between fertilizer use efficiency (FUE) and variable rate as-applied N fertilizer. Colors represent measurements in each field.

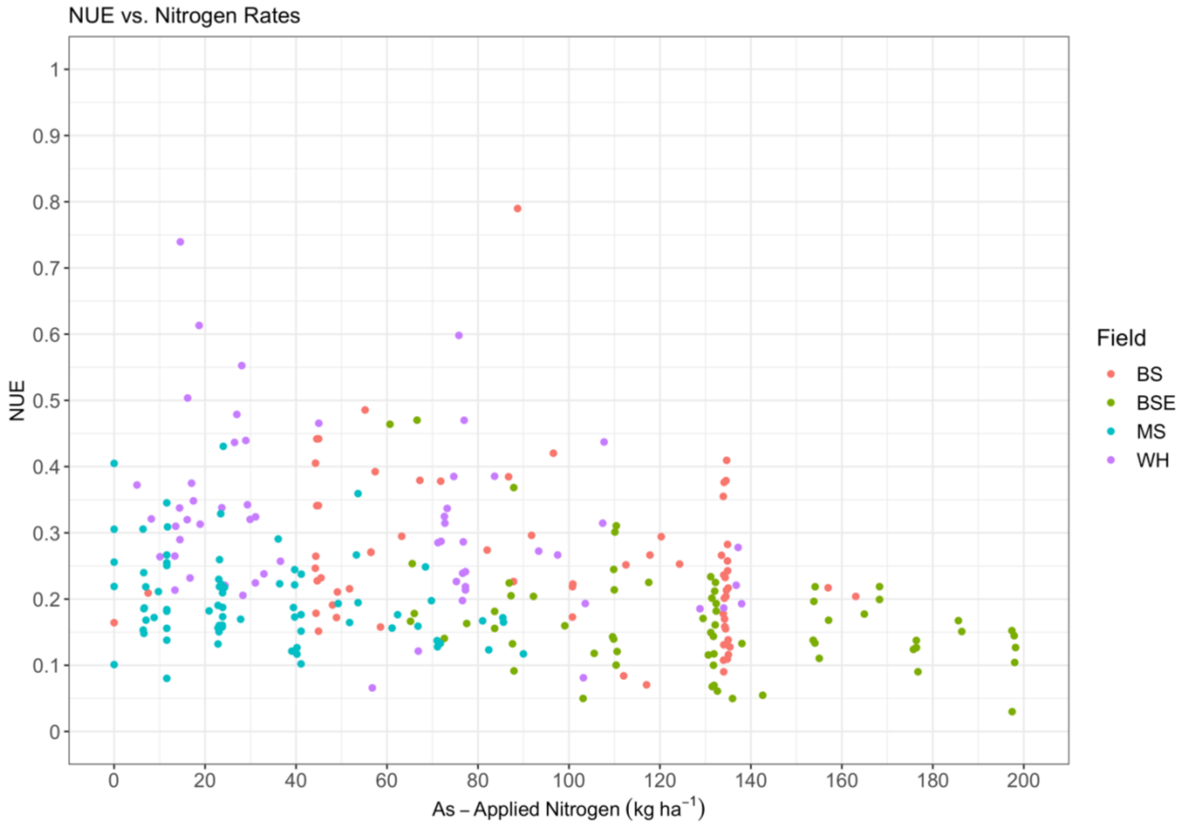


Figure 6.9. The relationship between NUE and variable rate as-applied N fertilizer. Colors represent measurements in each field.

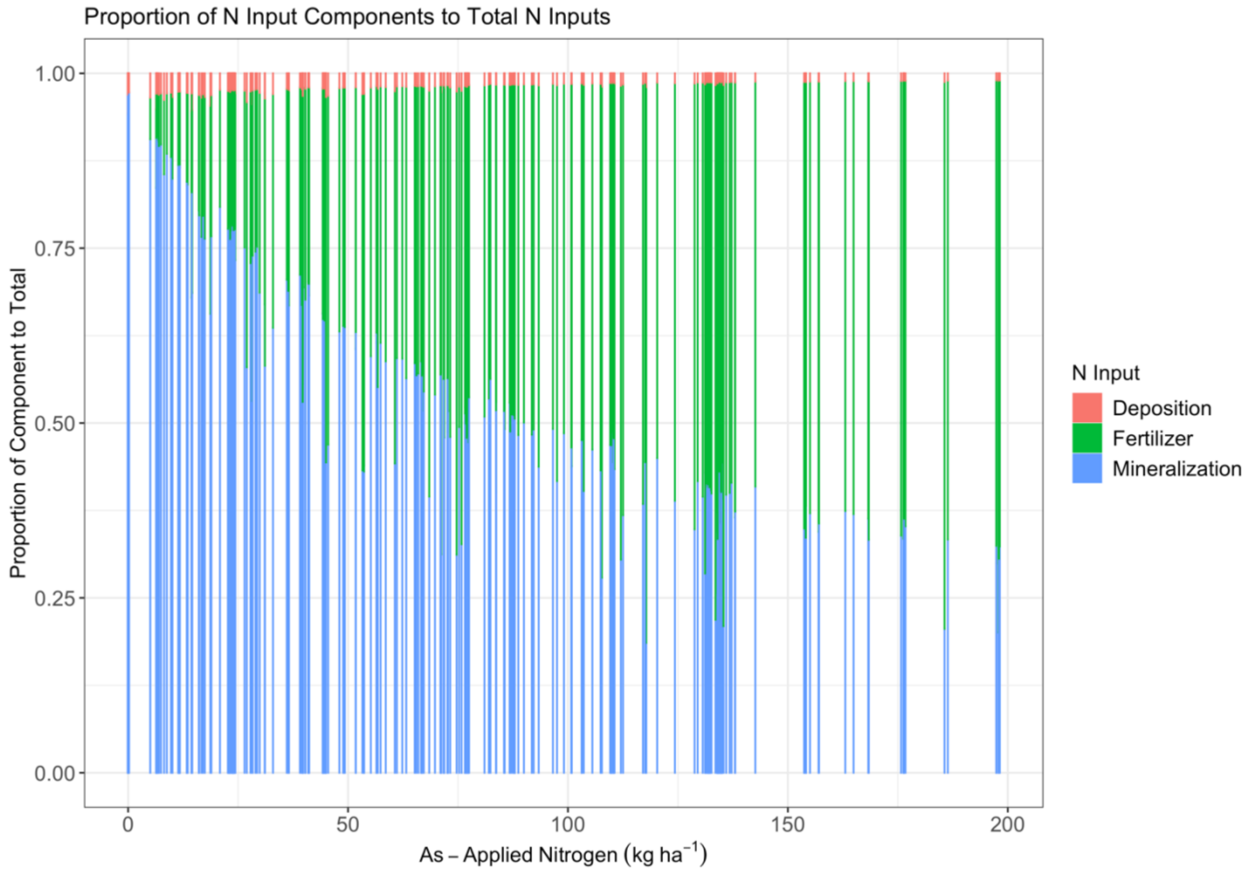


Figure 6.10. The proportion of each component of N inputs to the total N inputs across N fertilizer rates. Bars are from observations across all fields.

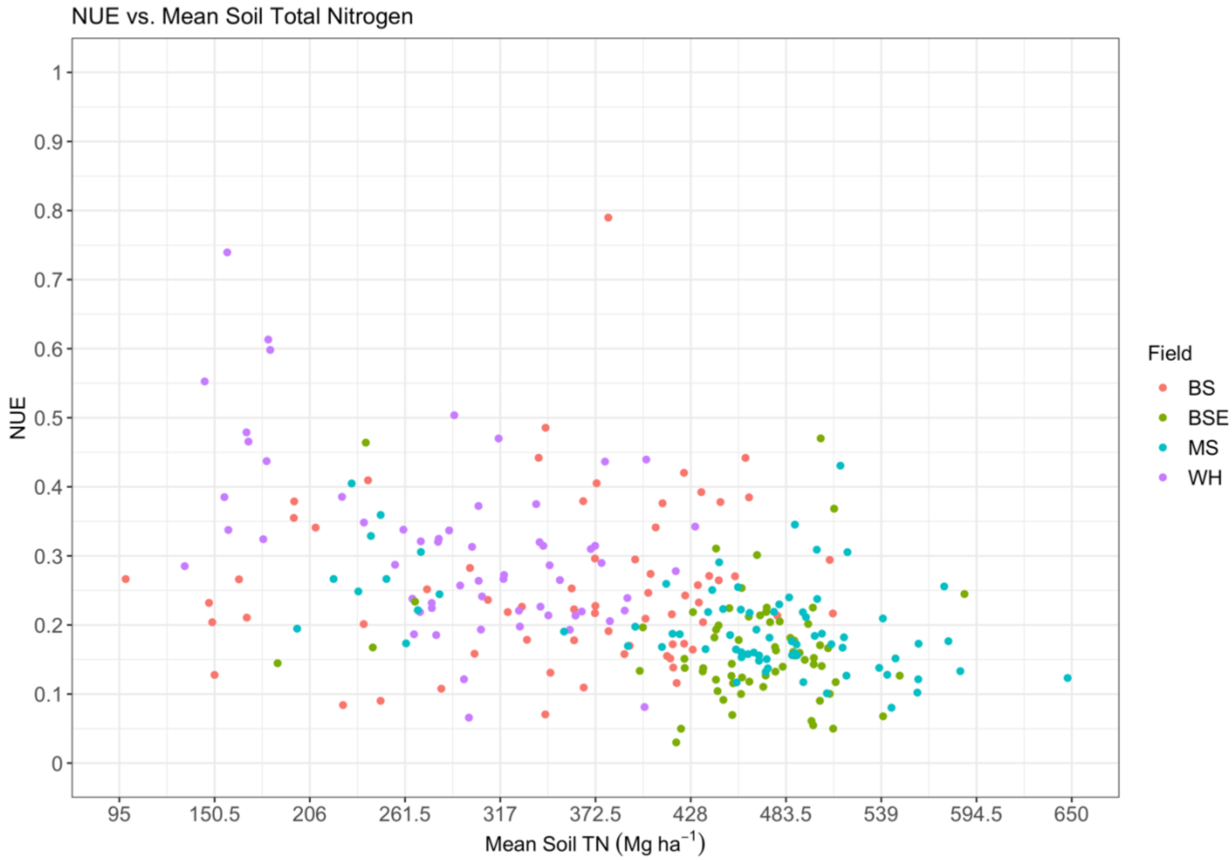


Figure 6.11. The relationship of nitrogen use efficiency, calculated via Equation 6.7, to the mean soil TN at each point. Colors indicate samples from differing fields.

Table 6.8. RMSE calculated from the observed and predicted NUE from each model during the hold-one field out cross validation. Bolded values indicate the model with the lowest RMSE for a field across models and italicized values indicate the model with the second lowest RMSE for a field.

Test Field	Model Tested					
	***SLR*[1]**	***MLR*[2]**	***NLR*[3]**	***GAM*[4]**	***RF*[5]**	***SVR*[6]**
BS 2020	0.1374	0.1704	*0.1274*	0.7585	0.1402	**0.1253**
WH 2020	*0.1584*	0.2718	**0.1462**	0.8322	0.1717	0.1673
BSE 2021	*0.0812*	**0.0781**	0.0879	0.1861	0.0986	0.1072
MS M21	0.1043	0.1145	0.1260	0.5503	**0.0688**	*0.0720*
Mean	0.1203	0.1587	0.1219	0.5818	*0.1198*	**0.1179**

[1]SLR: linear regression NUE predicted by as-applied nitrogen as a covariate.
[2]MLR: multiple linear regression model using available spatially variable potential covariates.
[3]NLR: non-linear model using an exponential decay function with as-applied N as a covariate.
[4]GAM: generalized additive model with as-applied N and available spatially variable potential covariates.
[5]RF: random forest regression model with as-applied N and available spatially variable potential covariates.
[6]SVR: support vector regression model with as-applied N and available spatially variable potential covariates.

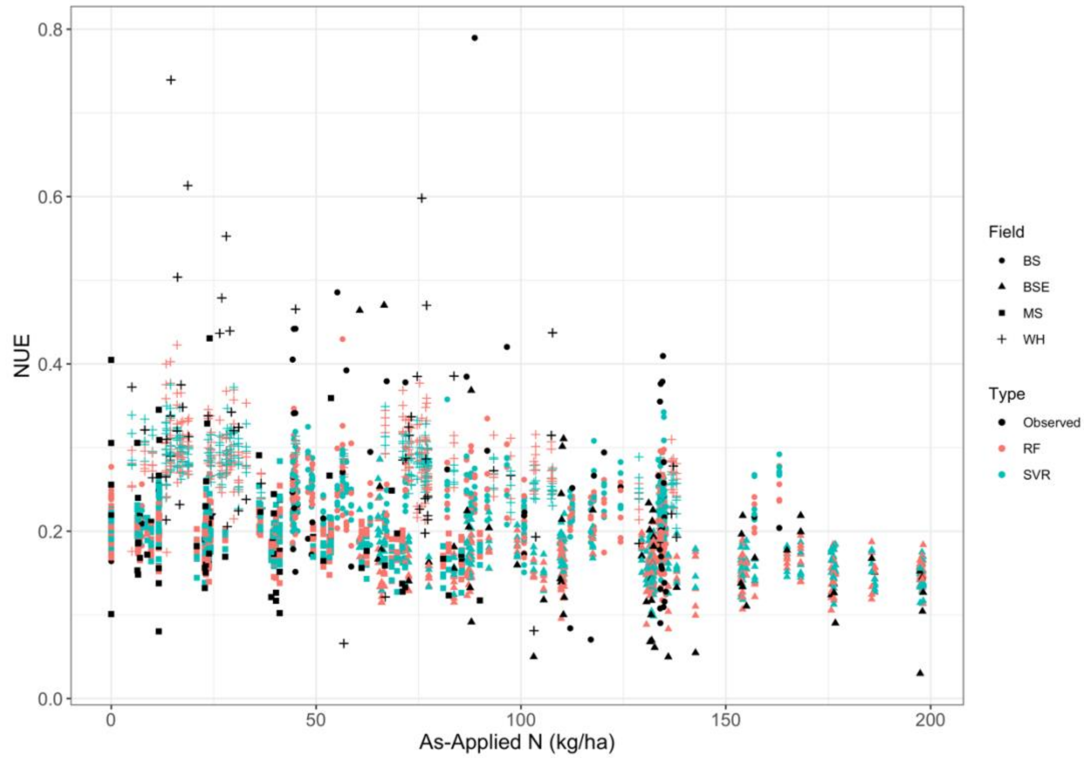


Figure 6.12. Observed NUE (black points) across variable rate as-applied N rates, where shapes indicate each field. Predicted NUE values are RF and SVR predictions from the test sets performed in the 5x2 CV.

Table 6.9. Parameter identifiers from Equation 6.9 with the corresponding data type and description.

Parameter ID	Data Type	Description
N	N Fertilizer	As-applied N fertilizer rate (kg ha^{-1})
WHC	Water Holding Capacity (WHC)	Estimated WHC zone
tpi	Topographic Position Index (TPI)	TPI calculated from USGS NED (Table 6.5)
$aspect_{sin}$	Aspect (Exposure)	Aspect in N/S direction calculated from the USGS NED (Table 6.5)
$aspect_{cos}$	Aspect (Exposure)	Aspect in E/W direction calculated from the USGS NED (Table 6.5)
$prec_{cy}$	Precipitation	Precipitation (mm) measured from Daymet V3 (Table 6.5) across November 1st of the previous year to March 30th of the year samples were taken
$precp_{y}$	Precipitation	Precipitation (mm) measured from Daymet V3 (Table 6.5) across November 1st of two years prior to October 31st of the year prior to when samples were taken
$ndvi_{cy}$	Normalized Difference Vegetation Index (NDVI)	NDVI from a greenest pixel composite across Sentinel 2 images (Table 6.5) from January 1st to March 30th of the year samples were taken
$ndvi_{py}$	Normalized Difference Vegetation Index (NDVI)	NDVI from a greenest pixel composite across Sentinel 2 images (Table 6.5) from January 1st to December 31st of the year prior to when samples were taken
$ndvi_{2py}$	Normalized Difference Vegetation Index (NDVI)	NDVI from a greenest pixel composite across Sentinel 2 images (Table 6.5) from January 1st to December 31st two years prior to when samples were taken
$ndwi_{cy}$	Normalized Difference Water Index (NDVI)	Median NDWI from Sentinel 2 (Table 6.5) from January 1st to March 30th of the year samples were taken
$ndwi_{py}$	Normalized Difference Water Index (NDWI)	Median NDWI from Sentinel 2 (Table 6.5) across January 1st to December 31st in the year prior to when samples were taken
$ndwi_{2py}$	Normalized Difference Water Index (NDWI)	Median NDWI from Sentinel 2 (Table 6.5) across January 1st to December 31st two years prior to when samples were taken
bulkdensity	Bulk Density	Bulk density (kg m^{-3}) from OpenLandMap data (Table 6.5)
claycontent	Clay Content	Clay content (%) from OpenLandMap data (Table 6.5)
watercontent	Water Content	Water content (%) from OpenLandMap data (Table 6.5)
phw	Soil Water pH	pH of soil water from OpenLandMap data (Table 6.5)
carboncontent	Soil Organic Carbon	Soil organic carbon content (%) from OpenLandMap data (Table 6.5)

REFERENCES CITED

Allaire, M., Wu, H., & Lall, U. (2018). National trends in drinking water quality violations. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(9), 2078–2083. https://doi.org/10.1073/pnas.1719805115

Arnall, D. B., Tubaña, B. S., Holtz, S. L., Girma, K., & Raun, W. R. (2009). Relationship between nitrogen use efficiency and response index in winter wheat. *Journal of Plant Nutrition*, *32*(3), 502–515. https://doi.org/10.1080/01904160802679974

Booth, M. S., Stark, J. M., & Rastetter, E. (2005). Controls on Nitrogen Cycling in Terrestrial Ecosystems: A Synthetic Analysis of Literature Data. *Ecological Monographs*, *75*(2), 139–157.

Bouwman, A. F., Beusen, A. H. W., Griffioen, J., Van Groenigen, J. W., Hefting, M. M., Oenema, O., Van Puijenbroek, P. J. T. M., Seitzinger, S., Slomp, C. P., & Stehfest, E. (2013). Global trends and uncertainties in terrestrial denitrification and N2O emissions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1621). https://doi.org/10.1098/rstb.2013.0112

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32.

Bullock, D. S., Boerngen, M., Tao, H., Maxwell, B., Luck, J. D., Shiratsuchi, L., Puntel, L., & Martin, N. F. (2019). The data-intensive farm management project: Changing agronomic research through on-farm precision experimentation. *Agronomy Journal*, *111*(6), 2736–2746. https://doi.org/10.2134/agronj2019.03.0165

Capel, P. D., McCarthy, K. A., & Barbash, J. E. (2008). National, holistic, watershed-scale approach to understand the sources, transport, and fate of agricultural chemicals. *Journal of Environmental Quality*, *37*(3), 983–993. https://doi.org/10.2134/jeq2007.0226

Cassman, K. G., & Munns, D. N. (1980). Nitrogen Mineralization as Affected by Soil Moisture, Temperature, and Depth. *Soil Science Society of America Journal*, *44*, 1233–1237.

Coble, K. H., Mishra, A. K., Ferrell, S., & Griffin, T. (2018). Big data in agriculture: A challenge for the future. *Applied Economic Perspectives and Policy*, *40*(1), 79–96.

https://doi.org/10.1093/aepp/ppx056

Coleman, D. C., & Dec, N. (1989). *Special Feature : Ecology , Agroecosystems , and Sustainable Agriculture*. *70*(6), 12–13.

De Bastiani, F., Mariz de Aquino Cysneiros, A. H., Uribe-Opazo, M. A., & Galea, M. (2015). Influence diagnostics in elliptical spatial linear models. *Test*, *24*(2), 322–340. https://doi.org/10.1007/s11749-014-0409-z

DeLonge, M. S., Miles, A., & Carlisle, L. (2016). Investing in the transition to sustainable agriculture. *Environmental Science and Policy*, *55*, 266–273. https://doi.org/10.1016/j.envsci.2015.09.013

Dietterich, T. G. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, *10*(7), 1895–1923. https://doi.org/10.1162/089976698300017197

Diggle, P. J., & Tawn, J. A. (1998). Model-based geostatistics. *Applied Statistics*, *47*(3), 299–350.

Duff, H., Hegedus, P. B., Loewen, S., Bass, T., & Maxwell, B. D. (2022). Precision Agroecology. *Sustainability*, *14*(106), 1–18. https://doi.org/https://doi.org/10.3390/su14010106

Engel, R. (2012). Volatilization losses from surface-applied urea during cold weather months. *Montana State University - Dept. LRES*, 1–10. https://doi.org/10.1017/CBO9781107415324.004

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., … Zaks, D. P. M. (2011). Solutions for a cultivated planet. *Nature*, *478*(7369), 337–342. https://doi.org/10.1038/nature10452

Gibert, K., Horsburgh, J. S., Athanasiadis, I. N., & Holmes, G. (2018). Environmental Data

Science. *Environmental Modelling and Software*, *106*, 4–12. https://doi.org/10.1016/j.envsoft.2018.04.005

Gliessman, S. (2016). Transforming food systems with agroecology. *Agroecology and Sustainable Food Systems*, *40*(3), 187–189. https://doi.org/10.1080/21683565.2015.1130765

Gliessman, S. R., & Engles, E. W. (2014). *Agroecology: The Ecology of Sustainable Food Systems* (3rd ed.).

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gotway, C. A., & Stroup, W. W. (1997). A Generalized Linear Model Approach to Spatial Data Analysis and Prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, *2*(2), 157–178.

Grant, C. A., Moulin, A. P., & Tremblay, N. (2016). Nitrogen management effects on spring wheat yield and protein concentration vary with seeding date and slope position. *Agronomy Journal*, *108*(3), 1246–1256. https://doi.org/10.2134/agronj2015.0510

Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling*, *157*(2–3), 89–100. https://doi.org/10.1016/S0304-3800(02)00204-1

Guttieri, M. J., Frels, K., Regassa, T., Waters, B. M., & Baenziger, P. S. (2017). Variation for nitrogen use efficiency traits in current and historical great plains hard winter wheat. *Euphytica*, *213*(4), 1–18. https://doi.org/10.1007/s10681-017-1869-5

Hatfield, J. L., & Stewart, B. A. (1992). Limitations to Plant Root Growth. In J. L. Hatfield & B. A. Stewart (Eds.), *Advances in Soil Science* (19th ed.). Springer-Verlag. https://doi.org/10.1007/978-1-4612-2894-3 e-ISBN:

Hegedus, P.B & Maxwell, B.D. (2022a). Rationale for field specific on-farm precision experimentation. *Agriculture, Ecosystems & Environment.* Manuscript submitted for publication, in review.

Hegedus, P.B. & Maxwell, B.D. (2022b). Assessing performance of empirical models for forecasting crop responses to variable fertilizer rates using on-farm precision experimentation. *Precision Agriculture.* Manuscript submitted for publication, in review.

Hegedus, P.B. & Maxwell, B.D. (2022c). Best use of modern data for field-specific decision support. *Precision Agriculture.* Manuscript submitted for publication, in review.

Holland, D. M., De Oliveira, V., Cox, L. H., & Smith, R. L. (2000). Estimation of regional trends in sulfur dioxide over the eastern United States. *Environmetrics*, *11*(4), 373–393. https://doi.org/10.1002/1099-095X(200007/08)11:4<373::AID-ENV419>3.0.CO;2-2

Janatian, N., Sadeghi, M., Sanaeinejad, S. H., Bakhshian, E., Farid, A., Hasheminia, S. M., & Ghazanfari, S. (2017). A statistical framework for estimating air temperature using MODIS land surface temperature data. *International Journal of Climatology*, *37*(3), 1181–1194. https://doi.org/10.1002/joc.4766

Jenks, George F. 1967. "The Data Model Concept in Statistical Mapping", International Yearbook of Cartography 7: 186-190.

John, A. A., Jones, C. A., Ewing, S. A., Sigler, W. A., Bekkerman, A., & Miller, P. R. (2017). Fallow replacement and alternative nitrogen management for reducing nitrate leaching in a semiarid region. *Nutrient Cycling in Agroecosystems*, *108*(3), 279–296. https://doi.org/10.1007/s10705-017-9855-9

Jones, C., Olson-Rutz, K., Miller, P., Zabinski, C., & Tallman, S. (2022). *Cover Crops : Management for Organic Matter and Nitrogen Cover Crops : Management for Organic Matter and Nitrogen*. https://apps.msuextension.org/publications/pub.html?sku=EB0237

Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, *143*(September), 23–37. https://doi.org/10.1016/j.compag.2017.09.037

Kätterer, T., Eckersten, H., Andrén, O., & Pettersson, R. (1997). Winter wheat biomass and nitrogen dynamics under different fertilization and water regimes: Application of a crop growth model. *Ecological Modelling*, *102*(2–3), 301–314. https://doi.org/10.1016/S0304-3800(97)00065-3

Lambers, H., & Oliveira, R. S. (2019). Plant physiological ecology. In *Plant Physiological Ecology*. 2`. https://doi.org/10.1007/978-3-030-29639-1

Langella, G., Basile, A., Bonfante, A., & Terribile, F. (2010). High-resolution space-time rainfall analysis using integrated ANN inference systems. *Journal of Hydrology*, *387*(3–4), 328–342. https://doi.org/10.1016/j.jhydrol.2010.04.027

Lau, K. W., & Wu, Q. H. (2008). Local prediction of non-linear time series using support vector regression. *Pattern Recognition*, *41*(5), 1539–1547. https://doi.org/10.1016/j.patcog.2007.08.013

Lokers, R., Knapen, R., Janssen, S., van Randen, Y., & Jansen, J. (2016). Analysis of Big Data technologies for use in agro-environmental science. *Environmental Modelling and Software*, *84*, 494–504. https://doi.org/10.1016/j.envsoft.2016.07.017

Macnack, N., Khim, B. C., Mullock, J., & Raun, W. (2014). In-Season Prediction of Nitrogen Use Efficiency and Grain Protein in Winter Wheat (Triticum aestivum L.). *Communications in Soil Science and Plant Analysis*, *45*(18), 2480–2494. https://doi.org/10.1080/00103624.2014.904337

Mardia, A. K. V, & Marshall, R. J. (1984). Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression. *Biometrika*, *71*(1), 135–146.

Michaletz, S. T., Weiser, M. D., McDowell, N. G., Zhou, J., Kaspari, M., Helliker, B. R., & Enquist, B. J. (2016). The energetic and carbon economic origins of leaf thermoregulation. *Nature Plants*, *2*(9). https://doi.org/10.1038/nplants.2016.129

Montana Climate Office, University of Montana. 2020. Montana Mesonet. The Montana Forest and Conservation Experiment Station. http://climate.umt.edu/mesonet/

Oelofse, M., Markussen, B., Knudsen, L., Schelde, K., Olesen, J. E., Jensen, L. S., & Bruun, S. (2015). Do soil organic carbon levels affect potential yields and nitrogen use efficiency? An analysis of winter wheat and spring barley field trials. *European Journal of Agronomy*, *66*, 62–73. https://doi.org/10.1016/j.eja.2015.02.009

Paul, E. A., & Robertson, G. P. (1989). Ecology and the Agricultural Sciences: A False Dichotomy? *Ecology*, *70*(6), 1594–1597. https://doi.org/10.2307/1938091

Pavuluri, K., Chim, B. K., Griffey, C. A., Reiter, M. S., Balota, M., & Thomason, W. E. (2015). Canopy spectral reflectance can predict grain nitrogen use efficiency in soft red winter wheat. *Precision Agriculture*, *16*(4), 405–424. https://doi.org/10.1007/s11119-014-9385-2

Pearcy, R. W., Muraoka, H., & Valladares, F. (2005). Crown architecture in sun and shade environments: Assessing function and trade-offs with a three-dimensional simulation model. *New Phytologist*, *166*(3), 791–800. https://doi.org/10.1111/j.1469-8137.2005.01328.x

Pham, X., & Stack, M. (2018). How data analytics is transforming agriculture. *Business Horizons*, *61*(1), 125–133. https://doi.org/10.1016/j.bushor.2017.09.011

Ping, J. L., & Dobermann, A. (2003). Creating Spatially Contiguous Yield Classes for Site-Specific Management. *Agronomy Journal*, *95*(5), 1121–1131. https://doi.org/10.2134/agronj2003.1121

Prey, L., Hu, Y., & Schmidhalter, U. (2019). Temporal Dynamics and the Contribution of Plant Organs in a Phenotypically Diverse Population of High-Yielding Winter Wheat: Evaluating Concepts for Disentangling Yield Formation and Nitrogen Use Efficiency. *Frontiers in Plant Science*, *10*(October), 1–14. https://doi.org/10.3389/fpls.2019.01295

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, *1*(1), 51–59. https://doi.org/10.1089/big.2013.1508

Ravier, C., Meynard, J., Cohan, J., Gate, P., & Jeu, M. (2017). *Early nitrogen de fi ciencies favor high yield , grain protein content and N use e ffi ciency in wheat*. *89*(November 2016), 16–24. https://doi.org/10.1016/j.eja.2017.06.002

Reich, P. B. (2014). The world-wide “fast-slow” plant economics spectrum: A traits manifesto. *Journal of Ecology*, *102*(2), 275–301. https://doi.org/10.1111/1365-2745.12211

Sebilo, M., Mayer, B., Nicolardot, B., Pinay, G., & Mariotti, A. (2013). Long-term fate of nitrate fertilizer in agricultural soils. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18185–18189. https://doi.org/10.1073/pnas.1305372110

Semenov, M. A., Jamieson, P. D., & Martre, P. (2007). Deconvoluting nitrogen use efficiency in wheat: A simulation study. *European Journal of Agronomy*, *26*(3), 283–294. https://doi.org/10.1016/j.eja.2006.10.009

Sigler, W. A., Ewing, S. A., Jones, C. A., Payn, R. A., Brookshire, E. N. J., Klassen, J. K., Jackson-Smith, D., & Weissmann, G. S. (2018). Connections among soil, ground, and surface water chemistries characterize nitrogen loss from an agricultural landscape in the upper Missouri River Basin. *Journal of Hydrology*, *556*, 247–261. https://doi.org/10.1016/j.jhydrol.2017.10.018

Sigler, W. A., Ewing, S. A., Jones, C. A., Payn, R. A., Miller, P., & Maneta, M. (2020). Water and nitrate loss from dryland agricultural soils is controlled by management, soils, and weather. *Agriculture, Ecosystems and Environment*, *304*(July), 107158. https://doi.org/10.1016/j.agee.2020.107158

Sigler, W. A., Ewing, S. A., Wankel, S. D., Jones, C. A., Leuthold, S., Brookshire, E. N. J., & Payn, R. A. (2022). Isotopic signals in an agricultural watershed suggest denitrification is locally intensive in riparian areas but extensive in upland soils. *Biogeochemistry*, *158*(2), 251–268. https://doi.org/10.1007/s10533-022-00898-9

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Soil Survey Staff, Natural Resources Conservation Service (NRCS), United States Department of Agriculture (USDA). 2020. Web Soil Survey. Available online at the following link: http://websoilsurvey.sc.egov.usda.gov/

Thilakarathna, S. K., Hernandez-Ramirez, G., Puurveen, D., Kryzanowski, L., Lohstraeter, G., Powers, L. A., Quan, N., & Tenuta, M. (2020). Nitrous oxide emissions and nitrogen use efficiency in wheat: Nitrogen fertilization timing and formulation, soil nitrogen, and weather effects. *Soil Science Society of America Journal*, *84*(6), 1910–1927. https://doi.org/10.1002/saj2.20145

U.S. Environmental Protection Agency. (2018). *Glacier NP (GLR468)*. EPA CASTNET. https://www3.epa.gov/castnet/site_pages/GLR468.html

Van Sanford, D. A., & MacKown, C. T. (1986). Variation in nitrogen use efficiency among soft red winter wheat genotypes. *Theoretical and Applied Genetics*, *72*(2), 158–163. https://doi.org/10.1007/BF00266987

Vigil, M. F., Eghball, B., Cabrera, M. L., Jakubowski, B. R., & Davis, J. G. (2002). Accounting for seasonaI nitrogen mineralization : An overview. *Journal of Soil and Water Conservation*, *57*(6), 464–469.

Vinila Kumari, S., Bargavi, P., & Subhashini, U. (2016). Role of Big Data Analytics in Agriculture. *Special Issue on Computational Science, Mathematics and Biology*, 110–113. https://doi.org/10.18645/IJCSME.SPC.0025

Vitousek, P. M., Aber, J. D., Howarth, R. W., Likens, G. E., Matson, P. A., Schindler, D. W., Schlesinger, W. H., & Tilman, D. G. (1997). Human alteration of the global nitrogen cycle: Sources and consequences. *Ecological Applications*, *7*(3), 737–750. https://doi.org/10.1890/1051-0761(1997)007[0737:HAOTGN]2.0.CO;2

Walsh, E. S., Kreakie, B. J., Cantwell, M. G., & Nacci, D. (2017). A Random Forest approach to predict the spatial distribution of sediment pollution in an estuarine system. *PLoS ONE*, *12*(7), 1–18. https://doi.org/10.1371/journal.pone.0179473

Wang, Y., Wu, G., Deng, L., Tang, Z., Wang, K., Sun, W., & Shangguan, Z. (2017). Prediction of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-07197-6

Watson C.A., Atkinson, D., Gosling, P., Jackson, L. R., & Rayns, F. W. (2002). Managing soil fertility in organic farming systems. *Soil Use and Management*, *18*(3), 239–247. https://doi.org/10.1079/sum2002131

Weil, R. R., & Brady, N. C. (2017). *The Nature and Properties of Soils* (D. Fox, A. Gilfillan, & L. Dimmick (eds.); 15th ed.). Pearson.

Weiner, J. (2017). *Applying plant ecological knowledge to increase agricultural sustainability*. 865–870. https://doi.org/10.1111/1365-2745.12792

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big Data in Smart Farming – A review. *Agricultural Systems*, *153*, 69–80. https://doi.org/10.1016/j.agsy.2017.01.023

Xia, H., Ding, Y., & Wang, J. (2008). Gaussian process method for form error assessment using coordinate measurements. *IIE Transactions (Institute of Industrial Engineers)*, *40*(10), 931–946. https://doi.org/10.1080/07408170801971502

Yin, L., Liu, K., Li, L., Wei, M., Yang, R., Xue, K., Cao, Z., Zhang, C., Li, Y., Wu, X., & Wang, X. (2020). Late-sown winter wheat requires less nitrogen input but maintains high grain yield. *Agronomy Journal*, *112*(3), 1992–2005. https://doi.org/10.1002/agj2.20171

Zuur, A. F., & Camphuysen, Kees, C. (2012). 6 Generalized Additive Models applied on northern gannets. In *A Beginners Guide to Generalized Additive Models with R* (Issue January, p. 15). Highland Statistics Ltd.

CHAPTER SEVEN

DEVELOPMENT AND EVALUATION OF SITE-SPECIFIC OPTIMIZED NITROGEN FERTILIZER MANAGEMENT BASED ON MAXIMIZATION OF PROFIT AND MINIMIZATION OF POLLUTION

Abstract

The excess application of nitrogen (N) fertilizer has led to degradation of soil and water which agriculture requires for crop production. A first step toward transitioning global agriculture to a sustainable future may be increasing the efficiency of chemical inputs, such as N fertilizer. Data gathered from the tools and technology of precision agriculture, augmented by on-farm precision experimentation (OFPE), were used to characterize crop responses to experimental variable N fertilizer rates with field-specific predictive models parameterized for seven rainfed winter wheat fields on three farms over three years. Modeling and simulation identified optimum within-field site-specific N fertilizer rates that maximized both farmer net-returns and N use efficiency (NUE). Optimized N fertilizer recommendations varied in magnitude between fields and, in a given field, varied in pattern and magnitude among years with differing weather conditions. Site-specific optimized N fertilizer rates were predicted to increase farmer profits in 100% of experimental fields across all simulated weather conditions compared to farmer-selected, uniform N-application rates. The total amount of N applied with site-specific optimized N fertilizer management was less than farmer selected uniform rates 43-57% of the time across all fields and weather conditions simulated. Simulation results indicate that optimized within-field N fertilizer variable rates were more profitable than current

management approaches and, in many fields, reduce total N load. These results demonstrate through a modeling exercise that OFPE can result in field-specific recommendations of N fertilizer based on net-return and efficiency that reduce N rates below those based only on net-return.

Introduction

The mechanization and automation of agriculture in the past sixty years has driven success in achieving agronomic goals in developed nations, based on maximization of production and the maximization of profits (Coleman, 1989; Gliessman & Engles, 2014; Paul & Robertson, 1989). Crop production consistently increased globally throughout the 1900's, notably after the Green Revolution, when production increased by 56% globally between 1965 and 1985, and 20% between 1985 and 2005 (Foley et al., 2011). With the area of cultivated land only increasing by 2.4% between 1985 and 2005, this increase in production has been associated with a 25% increase in global crop yields, due to the breeding of new crop varieties, technological progress, chemical inputs, and extensive agricultural infrastructure development (Foley et al., 2011; Gliessman & Engles, 2014). However, these objectives have been to the detriment of neighboring ecosystems, rural communities, and agricultural production itself. For example, the high demands of chemical inputs in industrial agriculture degrade soils and pollute air and water, threatening the resource pool upon which future crops rely and impacting natural ecosystems (Carlisle et al., 2019; Foley et al., 2011; Tilman et al., 2011). Agronomic practices must transition towards increasing efficiency in the effectiveness of agricultural inputs, such as nitrogen (N) fertilizer, to sustain the environmental resources that agricultural production requires (Duff et al., 2022; Gliessman, 2016).

Applying fertilizer on a site-specific basis within fields, based on the balance between crop needs and risk of pollution, may provide a more sustainable approach to N fertilizer management (Gebbers & Adamchuck, 2010; Pierpaoli et al., 2013). Traditionally, N fertilizers are commonly applied uniformly across fields, despite the fact that crop responses can vary with both space and time within fields (Hegedus & Maxwell, 2022a). Variability in crop responses to N fertilizer corresponds to nitrogen use efficiency (NUE), where the areas of the field that exceed a threshold of inefficiency that harm farmers profits will also contribute to environmental consequences like nitrate contamination of water resources, N_2O emissions, and soil acidification (Jones, 2018; Sigler et al., 2018, 2020). The advent of precision agriculture has spurred experimentation and development into site-specific N fertilizer management aimed at balancing the tradeoff between minimized inefficiency of N fertilizer use and farmer profitability(McBratney et al., 2005). Site-specific N management tailors fertilizer application rates to reduce fertilizer usage in areas within a field that will not generate a net return (Grzebisz, 2021; Koch et al., 2004; Meyer-Aurich et al., 2010; Wagner & Hank, 2013). Despite the important success in increasing farmer profits, incorporating NUE and not solely economic efficiency into site-specific N management decisions is necessary to use N fertilizer sustainably (Argento et al., 2021; Bucci et al., 2018; Link et al., 2008).

Sustainable N fertilizer use is a tradeoff between maximization of production and profit with minimization of the risk of pollution or other environmental harm (Antle & Valdivia, 2021; Kanter et al., 2018). Precision agriculture was created to apply agricultural inputs at the right place, in the right amount, and at the right time, presenting the opportunity to apply sustainable N rates that are optimized on the return on investment for a farmer and NUE (Gebbers &

Adamchuck, 2010; Pierpaoli et al., 2013). Investment and adoption of precision agriculture technology has grown in recent decades, from the use of GPS on agricultural equipment to remote sensing and sensors that measure yield, crop, and soil characteristics (Carolan, 2017; Schimmelpfennig & Lowenberg-DeBoer, 2020). Most farm machinery in the USA collects, or at least interacts, with data in one or many ways (Meola, 2021), and harnessing these data, combined with on-farm experimentation, facilitates data-driven decision making for individual fields (Haas et al., 2009; Mueller et al., 2009; Wyeth, 2009).

Founded from precision agriculture, precision agroecology harnesses the data and technology of modern agriculture to apply ecological principles that allow for more sustainably produced food, fuel, fiber, and feed needed for an increasing population (Duff et al., 2022). This study addressed the question: can Montana farmers achieve greater profitability and reduce potential for N pollution in their N application practices, compared to farmer's status quo management, by adopting a precision agroecological approach to dryland winter-wheat production? Our first objective was to develop site-specific N fertilizer rates based on the tradeoff between net-return and NUE using on-farm experimentation and precision agriculture technology. Our second objective was to evaluate the probability that site-specific management, accounting for net-return and cost of N inefficiency, was more profitable and reduced potential N pollution compared to conventional farmer-selected uniform rate management strategies.

Methods

Study Sites and Experimental Design

The precision agroecological On-Farm Precision Experiments (OFPE) framework, developed at Montana State University, was used in seven fields across Montana (Hegedus et al., 2022). The data used to develop and test the OFPE framework were from fields managed by farmers who have cooperated in the OFPE project at Montana State University from its inception in 2016 (https://sites.google.com/site/ofpeframework/). Farmers were selected based on their interest in precision agriculture management and their geographic and climatic diversity across Montana (Fig. 7.1). The seven fields, managed by three farmers, were selected based on the quality and specificity of data collected to address the specific tradeoff between within-field site-specific NUE and net return on investment (Table 7.1). All farmers owned or were outfitted with precision agriculture equipment, including GPS monitors, grain yield and protein sensors, and variable N rate application capabilities.

Variable N fertilizer rates were randomly applied with stratification on previous yields and N applications, across each field (Fig. 7.2). Experimental N rates were applied by March 30th each year and as-applied N fertilizer rate data were gathered from each farmer's equipment after the application date. Grain yield and grain protein concentration measurements from the on-combine sensors were gathered after harvest in August of each year. Grain yields (kg ha^{-1}) were measured every three seconds with a harvester-mounted yield monitor, and grain protein concentrations (%) were collected every ten seconds with a CropScan 3000H near-infrared monitor (Clancy, 2019). Typical agronomic cleaning and processing methods, that include lag corrections and outlier detection, were used for crop response and as-applied fertilizer data (Blackmore & Moore, 1999; Sudduth et al., 2012; Sudduth & Drummond, 2007).

In addition to data collected on-farms, open-source data measuring environmental variables were collected from Google Earth Engine data repositories were downloaded and imported into the OFPE database (Gorelick et al., 2017). For all open-source data types that the

OFPE framework uses, data were collected for every year from 2000 (or the first year available) to the present (Table 7.2). All remotely sensed data were collected and aggregated with on-farm data per the methods of the OFPE framework (Hegedus et al. 2022; Hegedus & Maxwell, 2022c).

Crop Response Models

After grain yield and grain protein concentration datasets were compiled, field-specific ecological models characterizing the response of grain yield or grain protein concentration to variable N fertilizer rates and environmental factors were trained. Model selection was conducted using the methods of the OFPE framework, where five model forms were assessed on their ability to minimize error of predictions related to observed crop responses (Hegedus et al., 2022; Hegedus & Maxwell, 2022b). In most fields, a random forest model was used for characterizing grain yield and grain protein concentration relationships to as-applied N rates (Table 7.3). In one field, grain yield was modeled using a modified version of a Bayesian non-linear function (Lawrence et al., 2015b) and grain protein concentration was modeled with a modified version of a non-linear sigmoid model, called the beta function (Yin et al., 2003). In two other fields, grain protein concentration was best predicted by a generalized additive model. Detailed description of the model types and the selection process for each field used can be found in Hegedus & Maxwell (2022b).

All models were trained with crop production or quality (grain yield or grain protein) as the response variable and as-applied N as the explanatory variable. All open-source covariate data were initially included in each model, but automated feature selection was performed during the model fitting process for each. Crop response models were then used to simulate crop responses in each field under varying weather conditions to compare management outcomes and inform identification and evaluation of site-specific N management.

Simulation

Agricultural management studies are limited by the reality that only one fertilizer rate can be applied in a given area of a field in a year. Thus, models are required to predict how crop responses will change in a given year under different management strategies with varying N fertilizer application schemes. Models are also required to derive optimum N fertilizer strategies, as multiple N fertilizer rates cannot be applied in the same location of a field and compared. Accordingly, a simulation approach using the OFPE framework at MSU and defined in Hegedus et al. (2022) was used to develop and evaluate site-specific N management that accounts for the tradeoff in net-return and cost of N inefficiency.

Weather plays a vital role in crop production, particularly in dryland systems, as precipitation and temperature are important drivers of crop growth. Farmers cannot control weather conditions; however, they can adjust management strategies based on previous weather conditions and expectations of future weather. For example, farmers tend to apply more N fertilizer when they expect to receive high precipitation and lower N fertilizer rates when they expect a year to be dry. Yet predicting variability of weather also presents a challenge for farmers, so simulation is important for assessing how the outcomes of N management vary in different weather conditions, particularly in years with varying precipitation. In this paper, the historical weather record for each year was used to identify the driest, wettest, and most average year from 2000 – 2021 (Table 7.4). These years were selected to assess how robust the profitability and N rate strategies were when accounting for uncertainty in future weather conditions. A 10-m grid was generated for each field, to which remotely sensed data from selected years were aggregated to the centroids of each cell to create the simulation datasets. The field-specific crop response models trained on observed data were then used to predict grain yield and grain protein responses across a field in the datasets that represent different weather conditions.

Optimization

To generate optimized N fertilizer rates based on maximized profit and minimized cost of N inefficiency, which relates to the potential for pollution, the field-specific crop response models were used to predict grain yield and grain protein concentration every 10m across each field in each simulated weather condition. At every point in a given field, crop responses were predicted for every N rate from 0 to 168 kg N ha^{-1}, from which the optimum N fertilizer rate was identified. After grain yield and grain protein concentration are predicted at a given rate for each N rate, net-return was calculated as;

$$NR = yield * P - CA * AA - FC - ssAC$$

(Equation 7.1)

where *NR* is the net-return (\$ ha^{-1}) received and a function of the product of the *yield* (kg ha^{-1}) and the price received (*P*, \$ kg^{-1}), minus the cost of the applied input (*CA*, \$ kg^{-1}) multiplied by the as-applied rate of the input (*AA*, kg ha^{-1}), the fixed costs (*FC*, \$ ha^{-1}) associated with production that do not include the input, and the cost per hectare of the site-specific application (*ssAC*, \$ ha^{-1}). As Montana farmers receive a premium/dockage to the base price received for wheat based on grain protein concentration, the price received for calculating net-return was (Hegedus & Maxwell, 2022b; 2022c);

$$P = Bp + (B0pd + B1pd * protein + B2pd * protein^2)$$

(Equation 7.2)

where *P* is the final price received ($ kg^{-1}), *Bp* is the base price received ($ kg^{-1}), *B0pd* is the intercept of the protein premium/dockage function set by the grain elevator, *B1pd* is the coefficient on the grain protein concentration (protein, %) and *B2pd* is the coefficient on the squared protein term to account for non-linearity in the relationship.

To simultaneously account for maximization of profit and minimization of the cost of N inefficiency, related to N pollution potential, NUE was predicted at every point for every N rate. Efficiency was modeled using a support vector regression model developed from intensive soil and plant tissue sampling conducted over two years in four of the fields (Hegedus & Ewing, 2022). Ecological and biogeochemical covariates important to NUE included topographic and edaphic variables, normalized difference vegetation index (NDVI), normalized difference water index (NDWI), and precipitation. The optimum N fertilizer rate for any given point in the field was then identified as the rate that maximized profit while considering the cost of N inefficiency. Under the assumption that NUE cannot exceed 100% (1.0), 1 – NUE represents the proportion of N not used by the crop, or nitrogen use inefficiency (NUI). Desiring maximization of net-return and minimization of NUI, the conceptual optimum N fertilizer rate at each point was the rate at which the distance between net-return and NUI was greatest (Fig. 7.3). Rather than solving for a multi-objective optimum, the tradeoff in net-return and cost of N inefficiency was incorporated into derivation of optimum N fertilizer rates by calculating the dollar cost of NUI. At every point, NUI was multiplied first by the amount of N fertilizer applied at each location in the field, and then by the cost of N fertilizer to generate a cost of nitrogen inefficiency. The cost of NUI at each point was then subtracted from the net-returns at each point and the optimum N fertilizer rate for each point was identified as the N rate at which net-return was maximized and the cost of adding one unit of N was greater than the increase in net-return, corresponding to the maximized distance between the two relationships (Fig. 7.3).

Unless a farmer has a contract for a pre-determined price, that farmer's uncertainty in weather is compounded by the unknown economic trajectories for an upcoming year that parallels uncertainty in the weather. This uncertainty was addressed through Monte Carlo simulations for each field and each weather condition, where economic conditions from 2000 – 2021 were randomly selected at each iteration and used to calculate net-return (Hegedus et al., 2022). This repetition and variation allow net-return and the cost of NUI to vary across the simulation and for the reporting of probabilistic outcomes.

Management Strategies

Simulations can identify optimized site-specific N fertilizer rates under varying weather and economic conditions, but also allow for the comparison of site-specific management to other management strategies. Site-specific optimized N fertilizer rates, herein referred to as the 'SS.Opt' strategy, were compared to five alternative management strategies. The first strategy was the application of the minimum amount of possible N fertilizer, zero, labeled as 'Min.Rate'' in figures. The second strategy was the application of a uniform farmer selected N fertilizer rate ('FS'), and the third was a full-field uniform N rate that balances the tradeoff between net-return and NUE ('FF.Opt'). The 'FF.Opt' rate was derived in a similar fashion to 'SS.Opt' rates, but was identified as a single rate across the entire field where the net-return was maximized when accounting for the cost of inefficiency. The fourth strategy was a scenario where the actual experimental rates from the most recent year of OFPE in each field were applied, labeled as 'Actual' below, and represents the potential additional cost of experimentation (i.e., non-optimized application for the purpose of better-informed models) when compared to other strategies. The fifth and final strategy was the application of zero N fertilizer with the assumption that a farmer would have been able to receive organic prices ('Org.Price') and where a uniform 15% reduction in yield due to weeds was assumed based on farmer feedback. The 'Min.Rate' and 'Org.Price' approaches take a drastic shift towards sustainable management by completely foregoing the application of N fertilizer management and represent the transformation of an agroecosystem from conventional management to organic management. The 'Min.Rate' approach characterizes the years during which farmers must cease conventional management yet cannot receive organic prices for crops, while the 'Org.Price' scenario describes management outcomes when the transition to organic management is completed, and they can receive the premium price.

At every realization of each simulation, the average net-return and total nitrogen applied from the 'SS.Opt' strategy was compared to each of the other strategies, and the number of realizations in which the 'SS.Opt' approach generated a higher net-return (profitability) or applied less nitrogen (potential for pollution) was recorded. The probability that the 'SS.Opt' approach outperformed other management strategies regarding net-return or total N applied was calculated as the number of realizations the 'SS.Opt' approach yielded higher net-returns or applied a lower amount of nitrogen compared to each other strategy, divided by the number of realizations in the simulation. Importantly, farmers do not currently receive a dockage to their net-returns based on the cost of NUI. The cost of NUI was subtracted from the net-return calculated at each point to identify the N rate that maximizes profit while considering efficiency; however, to make realistic comparisons to other strategies, the cost of NUI was added back to net-return to report and compare the profitability of optimized strategies with the other management scenarios.

For each combination of field and weather condition, the Monte Carlo simulation was performed with 100 iterations where base prices, costs of N fertilizer, and fixed costs from 2000 – 2021 were repeatedly sampled with replacement. The number of iterations was selected based on constraints related to the computation time required for running the simulations. The net-return and cost of NUI was calculated for every point under every N rate for each sampled economic condition to identify the N rate that maximized profit while considering the cost of N that has the potential to contribute to N pollution.

Results

Development of Optimized N Fertilizer Rates

While the relationship between net-return and as-applied N fertilizer varied between fields and across weather conditions, on average across realizations of the Monte Carlo simulation for each field and weather combination, the cost of NUI increased with increasing N fertilizer rates (Fig. 7.4). The net-return response to variable N fertilizer rates varied among fields, yet in most fields, a similar pattern in the response of net-return to variable N fertilizer rates was observed across weather conditions (Fig. 7.4). Note however, that the range and magnitude of net-returns predicted across N fertilizer rates varied among both weather conditions and fields (Fig. 7.4).

The optimum N rate at every point in the field grid was then averaged over the 100 simulations to generate a site-specific N fertilizer map for each weather condition simulated for a given field (Fig. 7.5). Areas where no rate were reported was due to a lack of data availability. Recommendations for N fertilizer rates that maximize profit and minimized the cost of N loss varied across simulated years, indicating that the optimum N fertilizer rates were sensitive to varying weather conditions and, in general, optimum N fertilizer rates were greatest in wet years (Fig. 7.5).

Variability in the pattern of site-specific N fertilizer recommendations across weather conditions occurred on a field specific basis In field B1, N rates were lower in the southwest portion of the field under dry conditions while in the average and wet conditions prescribed N rates were higher in the southwest portion of the field (Fig. 7.5). The range of N rates was greatest in dry conditions and least in average conditions for B1, however the highest N rates were observed across the field in the wet condition (Fig. 7.5). Differences in patterns of N fertilizer rate recommendations were more pronounced in field B2, which is adjacent to B1. In dry conditions, a uniform rate of N fertilizer was predicted to generate the highest net-returns when considering efficiency, indicating that a farmer would maximize profits with respect to costs of NUI by applying the full-field (‘FF.Opt’) uniform N fertilizer rate of zero (Fig. 7.5). In field B2, average years showed more areas of the field receiving lower N fertilizer rates than in wet years, particularly noticeable in the southeast portion of the field (Fig. 7.5). In field B3, the fractions of the field receiving low and high N rates tended to be similar, though with varying degrees of magnitude of prescribed application (Fig. 7.5). Recommendations when wet conditions were expected showed the highest N fertilizer rates (Fig. 7.5). Compared to the average conditions, a greater area of the field received high N fertilizer rates in the wet conditions.

Field D1 showed the least variation in the pattern of optimized N fertilizer rates across weather conditions. In wet conditions, the site-specific optimized strategy (‘SS.Opt’) was identical to the full-field uniform optimum rate strategy (‘FF.Opt’), where a uniform rate of zero kg N ha^{-1} was recommended across the entire field (Fig. 7.5). In average and dry conditions, a zero rate was recommended across most of the field, except for a few sites where N rates up to 22 and 11 kg N ha^{-1} were recommended, respectively (Fig. 7.5). In fields D2 and D3 the pattern of optimized N rates did not vary dramatically between average and dry conditions, yet in both fields the recommendations in wet conditions were of a higher magnitude and contained fewer low rates (Fig. 7.5).

The range of optimized N rates was smallest in dry conditions for field I1 due to the lack of any recommended N rates below 74 kg ha^{-1} (Fig. 7.5). In average and wet conditions, there were small sections of the field where a zero rate was recommended (Fig. 7.5). The maximum recommended N rate in average conditions was lower than the maximum in the dry and wet conditions, however in wet conditions the maximum rate was applied in a greater amount of the field compared to the other strategies (Fig. 7.5). In both dry and wet conditions, the southwest corner of the field was recommended to receive lower N rates compared to the rest of the field (Fig. 7.5).

Profitability of Site-Specific N Fertilizer Management

The simulated profitability of the site-specific optimized N fertilizer strategy (‘SS.Opt’) was compared to that of the four other fertilizer management strategies. Across all fields, the

'SS.Opt' approach yielded a higher net-return compared to applying no N fertilizer ('Min.Rate') 72% of the time in dry years, 87% of the time in wet years, and 100% of the time in an average year (Fig. 7.6). Averaging across all Monte Carlo ensembles for each field, the 'SS.Opt' approach in dry years was predicted to increase mean field net-returns by \$338.39 ha^{-1} compared to if no N fertilizer was applied (Fig. 7.7). In wet years, the 'SS.Opt' approach was predicted to increase net-returns by \$597.35 ha^{-1} compared to when no fertilizer was applied (Fig. 7.7). In a year simulating average weather conditions, the average difference in net-returns across Monte Carlo ensembles between the 'SS.Opt' and 'Min.Rate' approach was \$468.25 ha^{-1}, indicating the potential income across all fields resulting from fertilization with the 'SS.Opt' approach relative to zero application of fertilizer (Fig. 7.7).

The 'SS.Opt' approach increased net-returns from the farmer selected uniform rate ('FS') strategy 100% of the time, on average, across fields in all weather conditions (Fig. 7.6), indicating that the site-specific N management is expected to always be more profitable than the farmer-selected uniform rate. The 'SS.Opt' approach was predicted to increase net-returns by \$392.82 ha^{-1} on average across all fields compared to the 'FS' approach in dry years, by \$217.09 ha^{-1} in wet years, and \$241.78 ha^{-1} in an average year (Fig. 7.7). These values indicate the probable financial benefit of the site-specific approach for farmers relative to the conventional approach.

Compared to the 'FF.Opt' strategy, the 'SS.Opt' approach was estimated to generate a higher net-return 64% of the time in dry years, 74% of the time in wet years, and 94% of the time in an average year (Fig. 7.6). In dry years, the 'SS.Opt' approach increased net-returns by an average of \$172.07 ha^{-1} across all fields compared to the 'FF.Opt' approach (Fig. 7.7). In weather conditions simulating the wettest year from 2000 – 2021 for each field, the average difference in net-return was \$126.98 ha^{-1}, indicating farmers would have made more money with the 'SS.Opt' than the 'FF.Opt' approach (Fig. 7.7). In a year with average precipitation, the 'SS.Opt' was estimated to increase mean net-returns by \$161.96 ha^{-1} compared to the 'FF.Opt' approach (Fig. 7.7).

In order to assess the cost incurred by doing experimentation with different N fertilizer rates, the 'Actual' outcome with an experiment in the field was compared with the site-specific optimum application rates ('SS.Opt'). If the experimental design from the most recent year of OFPE was applied across the field in the simulated weather conditions ('Actual'), the 'SS.Opt' approach yielded a higher net-return, on average across fields, 100% of the time regardless of the weather conditions (Fig. 7.6). In a dry year, the 'SS.Opt' approach was predicted to increase net-returns by \$400.41 ha^{-1} compared to when the experiment was applied (Fig. 7.7). The difference between the 'SS.Opt' and 'Actual' approach where the experiment was applied was \$342.02 ha^{-1} on average across fields in a wet year, and \$313 ha^{-1} and on average across fields in an average year (Fig. 7.7).

Finally, when the 'SS.Opt' approach was compared to the scenario in which the farmer had completed a transition to organic management and applied zero N fertilizer with a 15% yield reduction, but received an organic price for their wheat ('Org.Price'), the 'SS.Opt' strategy was estimated to yield a higher net-return 59% of the time in dry years, 57% of the time in wet years, and 55% of the time in an average year (Fig. 7.6). These corresponded to average differences in net-returns between the 'SS.Opt' and 'Org.Price' strategies of \$150.34 ha^{-1} in dry years, \$48.58 ha^{-1} in wet years, and \$127.99 ha^{-1} in an average year (Fig. 7.7). However, in wet years, the difference in average net-return ranged from -\$29 ha^{-1} to \$116 ha^{-1}, indicating that in some economic conditions the ‘SS.Opt’ approach generated less money than if the farmer had transitioned to organic (Fig. 7.7).

The management strategy most likely to yield the highest profit depended on the field. While in most fields the ‘SS.Opt’ is expected to pay more than the ‘Min.Rate’ in a dry year, in field B2 and D1, a farmer was more likely to receive a higher net-return if they used the ‘Min.Rate’ or ‘FF.Opt’ strategies compared to the ‘SS.Opt’ approach (Fig. 7.8). In an average year, the ‘SS.Opt’ approach was expected to earn farmers a higher net-return than most of the other approaches, except if an organic premium was applied, regardless of the field (Fig. 7.9). The importance of taking a field-specific approach to selecting a management strategy was further demonstrated in the simulations of a wet year, where there was an 87% chance that the ‘SS.Opt’ approach would yield a higher net-return compared to no fertilizer strategy when averaged across fields (Fig. 7.6). If a farmer was selecting a management strategy based on results averaged across fields, they would conclude that the ‘SS.Opt’ approach was their best bet for making money in a wet year, with the exception in field D1, where there was only a 10% chance that the ‘SS.Opt’ approach would generate higher net-returns compared to if no fertilizer was applied (Fig. 7.10). Thus, in a wet year in field D1, a farmer would profit the most from applying no N fertilizer, even though in neighboring fields D2 and D3 there was a 100% chance a ‘SS.Opt’ approach would generate higher profits compared to applying no fertilizer (Fig. 7.10).

<u>Pollution Reduction of Site-Specific N Fertilizer Management</u>

The total amount of N fertilizer applied across fields was assessed to address whether the ‘SS.Opt’ approach provided an option for N fertilizer management that reduced the potential for

N pollution compared to the other strategies under the assumption that due to inherent NUI more N applications result in more potential for N pollution. As both the ‘Min.Rate’ and ‘Org.Price’ strategies do not apply N fertilizer across a field, they represent maximum pollution reduction regarding N fertilizer management because there is no risk of environmental damage due to N fertilizer. The scenario where the experiment from the most recent year of OFPE was applied in a field (‘Actual’) represents the cost of experimentation where the goal was to identify optimized N rates and doesn’t necessarily represent a competing strategy, as it is the path towards developing the ‘SS.Opt’ and ‘FF.Opt’ approaches. However, the ‘Actual’ scenario was included in the comparisons to the ‘SS.Opt’ approach to demonstrate how the pollution reduction potential of the optimized approach compared to the approach required to find the optimized rates.

Potential for pollution was assessed like the assessment of net returns. The fraction of realizations where the ‘SS.Opt’ approach applied less than or an equal amount of total N fertilizer across a field compared to each of the other strategies was recorded. This was used to derive the probability that the ‘SS.Opt’ was more effective for reducing pollution as the other strategies. Due to the difference in sizes of fields, differences in the total amount of N applied between the ‘SS.Opt’ and other strategies were assessed on a field specific basis.

In dry years, averaged across all fields, the ‘SS.Opt’ approach had a 14% chance of prescribing application of no fertilizer (‘Min.Rate’ or ‘Org.Price’ strategies) (Fig. 7.11). In no scenarios did the ‘SS.Opt’ approach prescribe no fertilizer in wet years, and 0% in an average year (Fig. 7.11). As the ‘Min.Rate’ and ‘Org.Price’ strategies applied zero fertilizer, the difference in total amount of N applied across the field between the ‘SS.Opt’ strategy and the ‘Min.Rate’ or ‘Org.Price’ strategy was simply the amount of N applied with the ‘SS.Opt’ strategy averaged across simulations. In the dry conditions, the difference in total N applied across field B2 and D1 between the ‘SS.Opt’ and both ‘Min.Rate’ and ’Org.Price’ was 0 kg N, indicating that the ‘SS.Opt’ recommendation for these fields in a dry year was to apply zero N fertilizer across the whole field (Fig. 7.12).

Compared to the ‘FS’ strategy, the average probability across fields that the ‘SS.Opt’ approach applied less N fertilizer was 57% in dry years, 43% in wet years, and 43% in an average year (Fig. 7.11). A 57% chance of applying less N fertilizer across a field with the ‘SS.Opt’ approach compared to the farmer selected (‘FS’) approach corresponded with less N applied in four out of seven fields in a dry year. The greatest reduction of N fertilizer was predicted in a dry year on B2 because the ‘SS.Opt’ approach recommended a uniform rate of zero N fertilizer across the field (Fig. 7.12, Fig. 7.5). In wet years, the ‘SS.Opt’ approach applied less N fertilizer than the ‘FS’ approach in three out of seven fields, albeit with a difference in total N applied of 135 kg across B1 (Fig. 7.13). This indicated that the ‘SS.Opt’ approach was almost as effective at reducing N as the ‘FS’ approach. The same pattern was seen in average years, where in three out of seven fields the ‘SS.Opt’ approach applied less than or equal the amount of total N applied with the ‘FS’ approach, with a slightly higher amount of total N applied with the ‘SS.Opt’ compared to the ‘FS’ approach in B1 (Fig. 7.14).

When comparing the ‘SS.Opt’ and ‘FF.Opt’ approaches in wet conditions, there was an average probability across fields (71%) that the ‘SS.Opt’ approach applied less than or an equal amount of total N across the field compared to the ‘FF.Opt’ approach (Fig. 7.11). In dry years, the average probability that the ‘SS.Opt’ approach applied less N fertilizer than the ‘FF.Opt’ approach was 57%, while in an average year the probability was 43% (Fig. 7.11). In wet years, five out of seven fields showed the ‘SS.Opt’ approach was predicted to apply less N fertilizer than the full-field optimized rate, although in fields B3 and D1 the difference in total N applied between the two strategies was 806 and 3 kg N across the entire field (Fig. 7.13).

The applications of OFPE is a critical step to derive field specific optimized management strategies (‘SS.Opt’ and ‘FF.Opt’). The comparisons between the N applied with the site-specific approach (‘SS.Opt’) and a scenario where the experiment from the most recent year of OFPE was applied (‘Actual’) was made to assess the increase in sustainability between the optimized rates and the experimentation required for developing them. There was a 57% probability that the ‘SS.Opt’ approach applied a less or equal amount of total N across fields compared to the ‘Actual’ scenario in dry years, 29% in wet years, and 43% in an average year (Fig. 7.11). In dry years the ‘SS.Opt’ approach was predicted to apply less N fertilizer across the field in four out of seven fields, although in field D3 there was a small predicted difference between the total N applied with the ‘SS.Opt’ strategy and ‘Actual’ scenario (Fig. 7.12).

Cost Of Inefficiency

Of interest to both policy makers and farmers is: how much would net-returns be affected if farmers were penalized for inefficiency? The cost of NUI using the price of N fertilizer was subtracted from net-returns as an estimate of the cost of inefficiency that a farmer would experience when optimizing N rates on maximized profit and minimized cost of N inefficiency. The total difference in net-return across fields between net-returns with and without a penalty were reported for each field and weather condition (Table 7.5). In fields and weather conditions, such as B2 in dry conditions and D1, where optimized N rates were zero or close to zero (Fig. 7.5) there was no penalty for inefficiency, as no or minimal nitrogen was applied with the potential to be inefficient. In fields where optimized N rates were high, the cost of inefficiency increased due to NUE decreasing with increasing N rates (Hegedus & Ewing, 2022). In most fields, the greatest amount of N was applied in wet conditions, hence wet conditions typically resulted in higher costs of inefficiency across fields.

Discussion

The cost of potential N loss increased (Fig. 7.4) with N fertilizer rates in all fields because of the negative relationship between NUE and N fertilizer (Hegedus & Ewing, 2022). The decrease in NUE with increasing N fertilizer rates reflects a decrease in crop N uptake relative to inputs of both fertilizer and mineralized N when crop N requirements are met. At a certain N fertilizer rate, crop N uptake will plateau, creating a threshold of NUE where further additions of N fertilizer decrease NUE with no benefit of crop N uptake (Hegedus & Ewing, 2022).

While the previously observed relationship between NUE and N fertilizer was relatively similar across fields (Hegedus and Ewing), variation in crop responses among fields translated to differences in the magnitude of net return, and in the shape and trend of the relationship between net-return and N fertilizer (Fig. 7.4). Field specific variation in the response of crops to experimental N fertilizer rates during years OFPE was conducted led to variation in the characterization of crop response models between fields (Table 7.3). The field specificity of crop responses ultimately led to differences in the modeled net-return response to N fertilizer between fields (Fig. 7.4) which corresponded to differences in the magnitude of optimized site-specific N fertilizer rates among fields (Fig. 7.5). Of note are differences in the magnitude of optimized N rates in B1 and B2, fields that share a border. This difference in adjacent fields indicates that experiments designed to understand the field-specific crop responses to management enable decision making that maximizes net-return on each field (Hegedus and Maxwell, 2022).

Field differences in the relationship between net-returns and total N fertilizer applied were clear, which is likely due to legacy effects from previous fertilizer use. Additional variation was observed, albeit to a slightly lesser extent, across weather conditions for a given field (Fig. 7.4). In some cases, such as in D1, the similarity in the response of net-return to N fertilizer across weather conditions (Fig. 7.4) translated to a lack of variation in the magnitude and pattern of optimized N fertilizer rates (Fig. 7.5). In field D1, no fertilizer was recommended for any weather condition, indicating that on-farm experimentation suggested that crops planted in that field do not benefit from N fertilizer applied above ambient conditions (Fig. 7.5). However, most fields also showed variation across simulation years in the magnitude and pattern of optimized N fertilizer rates (Fig. 7.5), indicating that weather conditions influence a farmer's optimized approach in specific fields. As upcoming weather conditions are not necessarily reliably predictable beforehand, a best guess at conditions must be taken at the time a decision on N fertilizer management is made. The potential to make effective optimized N fertilizer recommendations with unknown future weather conditions is first limited by a farmer or crop manager's ability to choose an appropriate proxy year from the past that might represent conditions in the coming season. The second limitation to profitable N fertilizer recommendations is the representation of differing weather conditions in the data used to train crop response models. This highlights the importance of Monte Carlo techniques to better constrain farmers facing uncertain conditions.

Generating an adequate backlog of data that represents different weather conditions is a process that is predicted to take 6-8 years of OFPE (Lawrence et al., 2015). Despite requiring a commitment to long-term on-farm experimentation, the site-specific optimized N fertilizer recommendations developed here are predicted to generate a higher mean net-return compared to use of a farmer selected uniform N rate (Fig. 7.6). The NUE associated with achieving higher net-returns with site-specific optimized rates compared to a farmer selected uniform rate is expected to vary between fields.

Site-specific management was not always predicted to be the best approach for applying optimized N fertilizer rates, as in some field and weather conditions an optimized full-field uniform rate was more likely to generate higher net-returns and reduce the potential for N pollution compared to site-specific management (Fig. 7.6, Fig. 7.11). Additionally, the difference between the two approaches was sometimes negligible in terms of profitability and amount of N applied, indicating that the optimized site-specific rates were not variable, and a full-field uniform application may be just as profitable and minimize pollution risk (Fig. 7.7, Fig. 7.14-7.17). This is a situation that would require farmer input to make a management decision. In situations where the site-specific optimized rate strategy was more profitable than the farmer selected uniform rate, but not different than the full-field optimized strategy, the full-field optimized strategy might be selected by the farmer because it is a simpler application or would mean that a variable rate applicator would not need to be rented or hired, an additional cost that would need to be incorporated into calculations of net-return. Although the optimum full-field uniform strategy would have the potential to increase adoption of more efficient N fertilizer management practices by removing the limitation of requiring variable rate applicator equipment, variable applications are still required to derive optimized rates, even if a uniform rate is selected.

In two fields, the optimized rates tended to be zero when the site-specific management was similar in profitability and pollution reduction potential compared to the full-field optimized approach. This result indicated that the farmer would have profited the most by not applying any N fertilizer (Fig. 7.5). This meant that the optimized N rates were similar in performance to the strategy where no N fertilizer was applied, and no organic premium was received.

Results of the OFPE process from this work suggest that development of optimized site-specific N fertilizer rates based on maximized net-returns and minimized cost of N inefficiency make more money compared to farmer selected uniform rates and offer a 50% chance that the risk of N pollution would be reduced. This combination demonstrates the possibility that balancing the tradeoff between profit and pollution when making N fertilizer management decisions is a competitive management strategy when OFPE is employed Increased net-returns of optimized management compared to a uniform N rate approach provide an incentive for conventional farmers to begin the process of increasing efficiency on their fields by reducing N fertilizer rates where less N is needed, and increasing rates where more N can be used by the crop . However, the cost of N inefficiency has the potential to be calculated in many ways, depending on the value that the person defining it has on NUE. This has implications for these results because raising the cost of N inefficiency would likely reduce the profitability of 'SS.Opt' management compared to farmer selected management because a larger penalty for potential N pollution is imposed. On the other hand, raising the cost of N inefficiency would likely reduce the probability that the 'SS.Opt' approach applies more N to fields compared to the farmer selected management strategy, because of the larger penalty for potential N pollution.

Although optimized rates were reported on a 10-m scale across fields, they are easily aggregated up to the management scale, dictated by farmer's equipment, with varying degrees of information loss depending on the scale of spatial variation between pixels. Although the optimized approaches, whether site-specific or full-field, do not always contribute to increased efficiency in dryland winter wheat production, in all cases they contributed to economic sustainability by increasing farmer net-returns, which benefit rural communities.

The seven fields examined in this work were not randomly selected from a larger pool of fields in Montana and were not randomly assigned to conventional management but selected from farmers in the OFPE project based on a survey of their interest and capability in precision agriculture technology and management. Thus, the inference for these results is technically limited to OFPE fields without the ability to make causal conclusions. However, the seven fields represent a range of geographic conditions experienced across Montana and are representative of dryland winter-wheat farmers in Montana that apply conventional N fertilizer management with variable rate equipped technology. This representation suggests that these findings apply to other dryland winter-wheat fields of Montana in conventional N fertilizer management. Furthermore, the economic loss to achieve maximum NUE could be used as field-specific incentive payments to reduce N pollution.

Conclusion

These results indicate that experimentation using precision agriculture tools and technology can be harnessed for the development of optimized N fertilizer recommendations that maximize net-returns and minimize the value of potential N loss. Site-specific optimized N fertilizer rates were predicted to be more profitable than applying a farmer selected uniform rate in all fields and across all weather conditions using conventional nitrogen fertilizer-based approaches. A site-specific optimized approach to determine maximized profitability and NUE requires a framework of experimentation and analysis like OFPE. The profitability and efficiency of optimized rates varies across fields and weather conditions, highlighting the importance of field-specific parameterization of models and simulation of outcomes to accomplish farmer objectives under uncertain conditions of weather variability, economic variability, and legacy effects of field history. Uncertainty in management outcomes characterized as probabilities are an effective way to provide crop input decision support. Finally, the analysis provided here describes a field-specific mechanism to incentivize minimization of nitrogen fertilizer pollution risk.

Acknowledgements

This research was supported by a WSARE Graduate Student Fellowship titled "Nitrogen Fertilizer Management Based on Site-Specific Maximized Profit and Minimized Pollution" (Award Number GW19-190), a USDA-NIFA-AFRI Food Security Program Coordinated Agricultural Project, titled "Using Precision Technology in On-farm Field Trials to Enable Data-Intensive Fertilizer Management," (Accession Number 2016-68004-24769), the USDA-NRCS Conservation Innovation Grant from the On-farm Trials Program, titled "Improving the Economic and Ecological Sustainability of US Crop Production through On-Farm Precision Experimentation" (Award Number NR213A7500013G021), and the Montana Fertilizer Advisory

Council from 2016 - 2021 , and the Consortium for Research on Environmental Water Systems (CREWS) National Science Foundation EPSCoR Cooperative Agreement OIA-1757351.

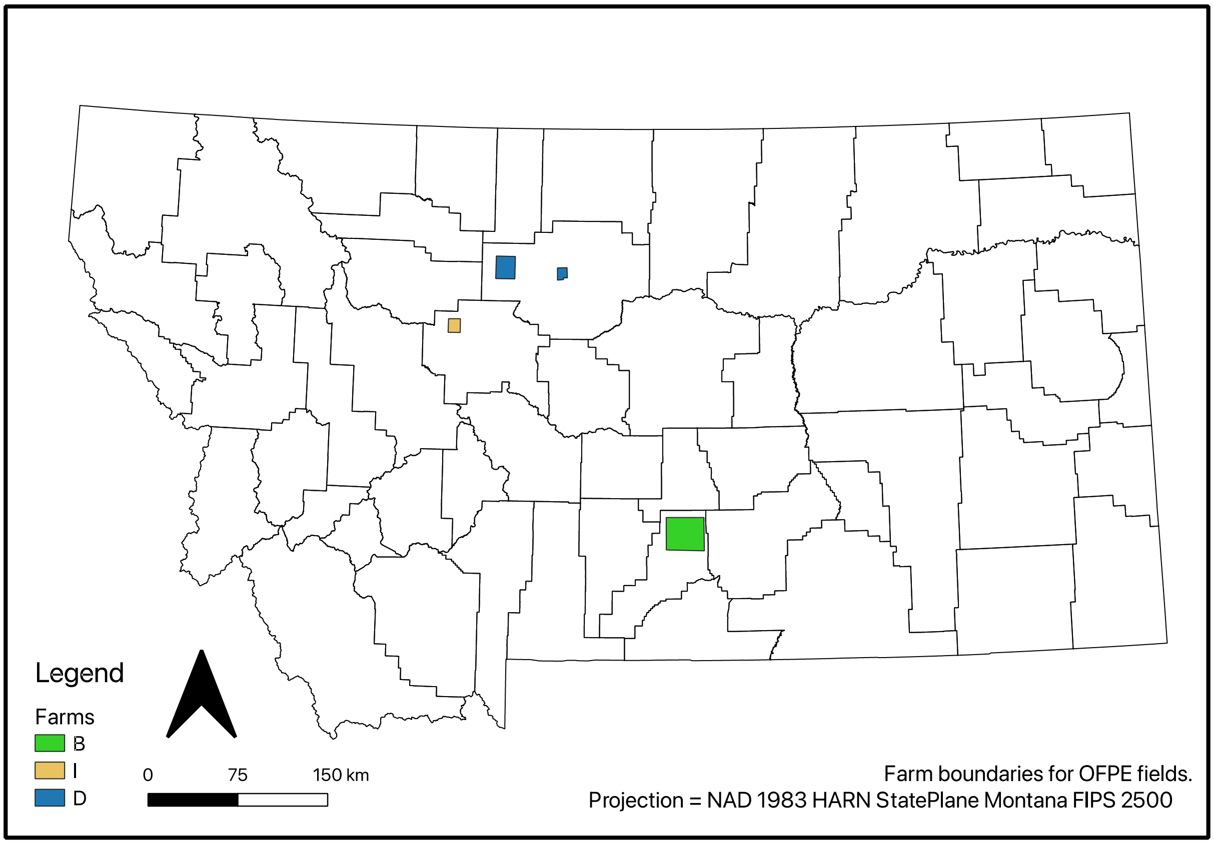


Figure 7.1. Map indicating the general geographic areas of Montana that include fields selected for On-Farm Precision Experimentation (OFPE).

Table 7.1. Crop histories, field sizes, and years in VRA treatment for each field for given farmers. The farm identifier corresponds to the map above and is used instead of the farmer's name for privacy.

Farm	**Field**	**Field size (ha)**	**Crop History[1]:** *2014 / 2015 / 2016 / 2017 / 2018 / 2019 / 2020 / 2021*	**Years N rate treatment**
B	B1	79	SF / WW / CF / WW / CF / WW / CF / WW	2017, 2019, 2021
	B2	94	WW / CF / WW / CF / SF / WW / CF / WW	2016, 2019, 2021
	B3	64	SW / CF / WW / CF / WW / CF / WW / CF	2016, 2018, 2020
D	D1	46	CF / WW / CF / WW / CF / WW / CF / WW	2017, 2019, 2021
	D2	48	WW / SW / WW / CF / WW / CF / WW / CF	2016, 2018, 2020
	D3	20	SW / SW / WW / CF / WW / CF / WW / CF	2016, 2018, 2020
I	I1	94	SW / CF / WW / CF / WW / CF / WW / CF	2016, 2018, 2020

[1] WW = winter wheat, CF = chemical fallow, SW = spring wheat, SF = safflower

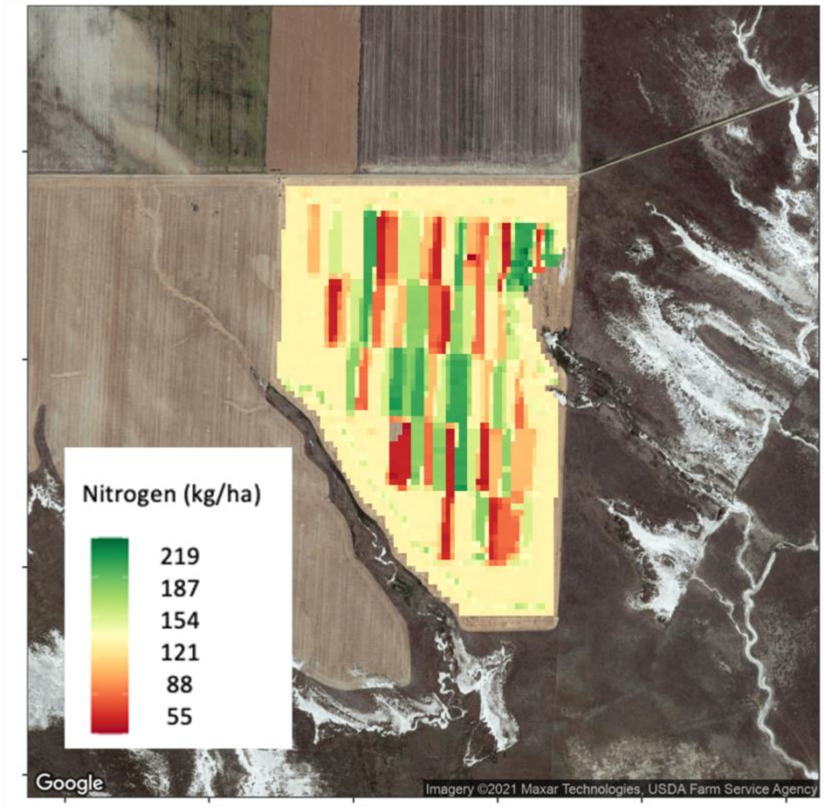


Figure 7.2. Example of an OFPE based experiment using as-applied N fertilizer (kg ha^{-1}) data from field B1.

Table 7.2. Table of covariate data types gathered from Google Earth Engine to enrich the crop yield and protein datasets gathered from on-farms. In some cases, multiple sources are used, however only one data source is used when aggregating to yield and protein.

Data Type	Data Sources	Resolution	Years Collected	Description
Normalized Difference Vegetation Index (NDVI)	Landsat 5/7/8	30m	L5: 1999-2011 L7: 2012-2013 L8: 2014 - present	Landsat is an ongoing USGS and NASA collaboration. Bands (NIR, red) L5/L7: B4 and B3 L8: B5 and B4
Normalized Difference Water Index (NDWI)	Landsat 5/7/8	30m	L5: 1999-2011 L7: 2012-2013 L8: 2014 - present	Landsat is an ongoing USGS and NASA collaboration. Bands (NIR, red) L5/L7: B2 and B4 L8: B2 and B5
Elevation	USGS NED	~10m (1/3 arc second), ~23m (3/4 arc second)	1999-present	USGS National Elevation Dataset. Measured in meters.
Aspect	USGS NED	~10m (1/3 arc second), 30m	1999-present	Direction the surface faces, function of neighboring elevations, in radians. Also calculated for each E/W and N/S direction as cosine and sine.
Slope	USGS NED	~10m (1/3 arc second), 30m	1999-present	Rate of change of height from neighboring cells, in degrees. Measured in degrees.
Topographic Position Index (TPI)	USGS NED	~10m (1/3 arc second), 30m	1999-present	Measure of divots and low spots as a function of neighboring elevation.
Precipitation	DaymetV3	1km	1999-present	Estimates from the NASA Oak Ridge National Laboratory (ORNL). Measured in mm.
Growing Degree Days (GDD)	DaymetV3	1km	1999-present	Estimates from the NASA Oak Ridge National Laboratory (ORNL).
bulkdensity	OpenLandMap	250m	1999-present	Soil bulk density (fine earth) 10 x kg / m3 averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
claycontent	OpenLandMap	250m	1999-present	Clay content in % (kg / kg) averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
sandcontent	OpenLandMap	250m	1999-present	Sand content in % (kg / kg) averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
pH (phw)	OpenLandMap	250m	1999-present	Soil pH in H_2O averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
watercontent	OpenLandMap	250m	1999-present	Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted and averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).
carboncontent	OpenLandMap	250m	1999-present	Soil organic carbon content in x 5 g / kg averaged over 6 standard depths (0, 0.1, 0.3, 0.6, 1 and 2 m).

Table 7.3. Table with the best performing model type for each field, where performance was measured using repeated hold one-year out cross validation (Hegedus & Maxwell, 2022b).

Field	Yield Model Type	Protein Model Type
B1	Bayesian Non-Linear Regression	Non-Linear Sigmoid Function
B2	Random Forest Regression	Random Forest Regression
B3	Random Forest Regression	Generalized Additive Model
D1	Random Forest Regression	Random Forest Regression
D2	Random Forest Regression	Random Forest Regression
D3	Random Forest Regression	Generalized Additive Model
I1	Random Forest Regression	Random Forest Regression

Table 7.4. Table of the years used to simulate management outcomes in different weather conditions for each field. Selected based on field-specific data from 2000 – 2021.

Field	**Average**	**Dry**	**Wet**
B1	2016	2001	2018
B2	2016	2001	2018
B3	2016	2001	2018
D1	2005	2001	2011
D2	2009	2001	2019
D3	2009	2001	2019
I1	2002	2003	2006

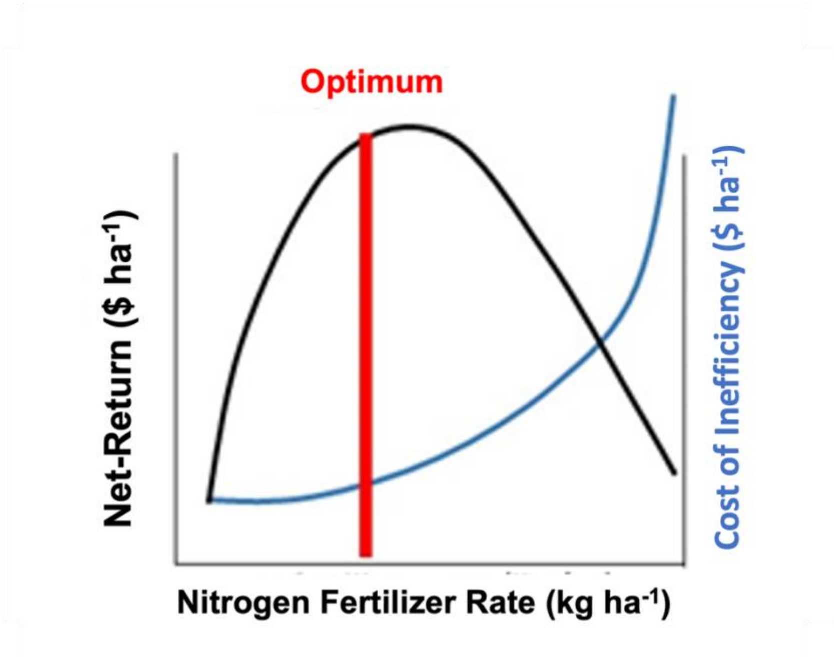


Figure 7.3. Conceptual diagram of the optimization process for finding the nitrogen fertilizer rate at each point that maximizes net-return and minimizes nitrogen use inefficiency.

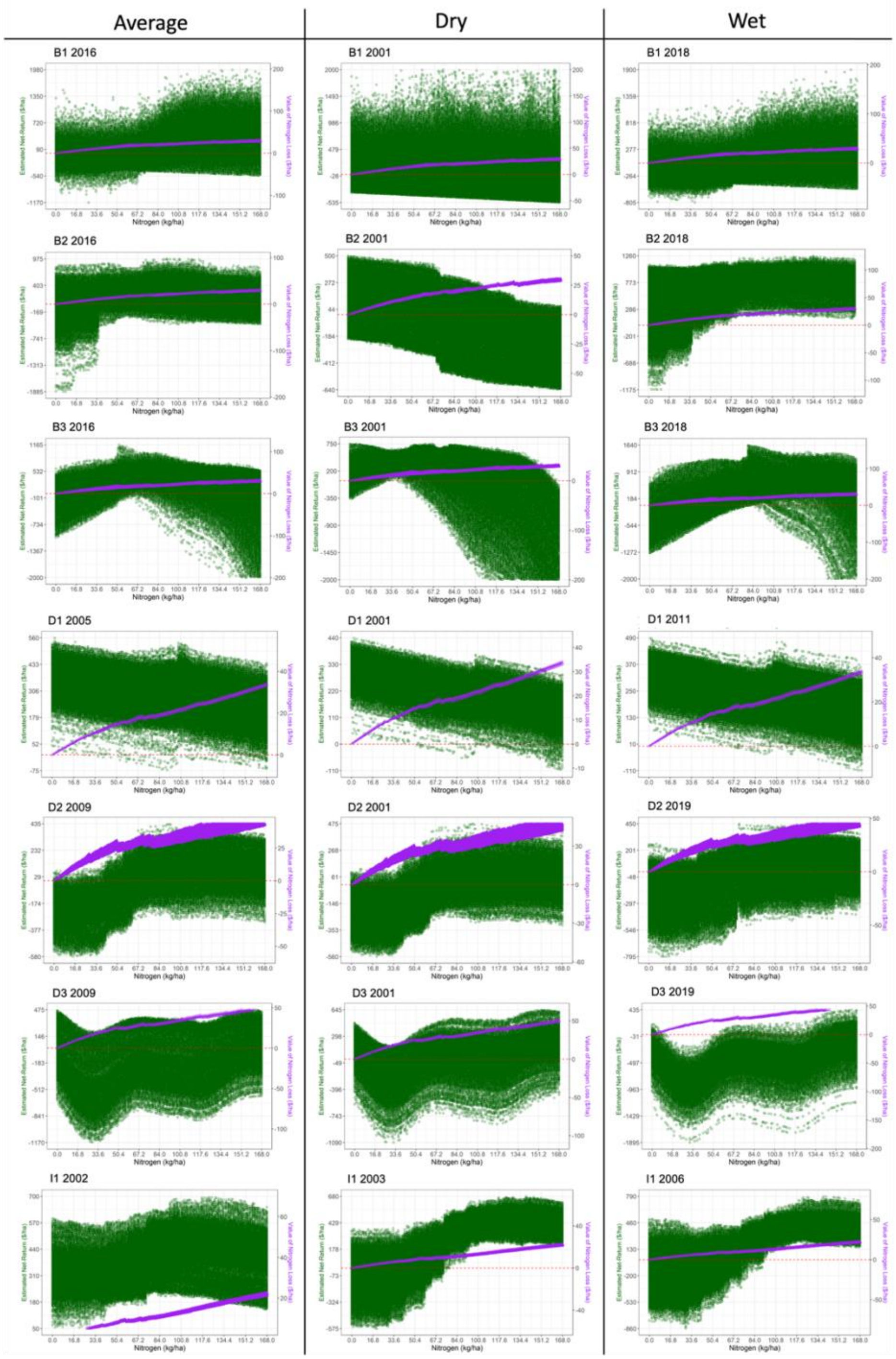


Figure 7.4. Predicted net-return (left axis, green) and predicted value of nitrogen loss calculated from the nitrogen use inefficiency (right axis, purple), averaged across Monte Carlo realizations, plotted against as-applied nitrogen rate for three weather conditions. The distribution of points at each nitrogen rate are from predictions at that rate from every point in the field. Note that scales are dependent on the field and weather condition to represent the variation across fields.

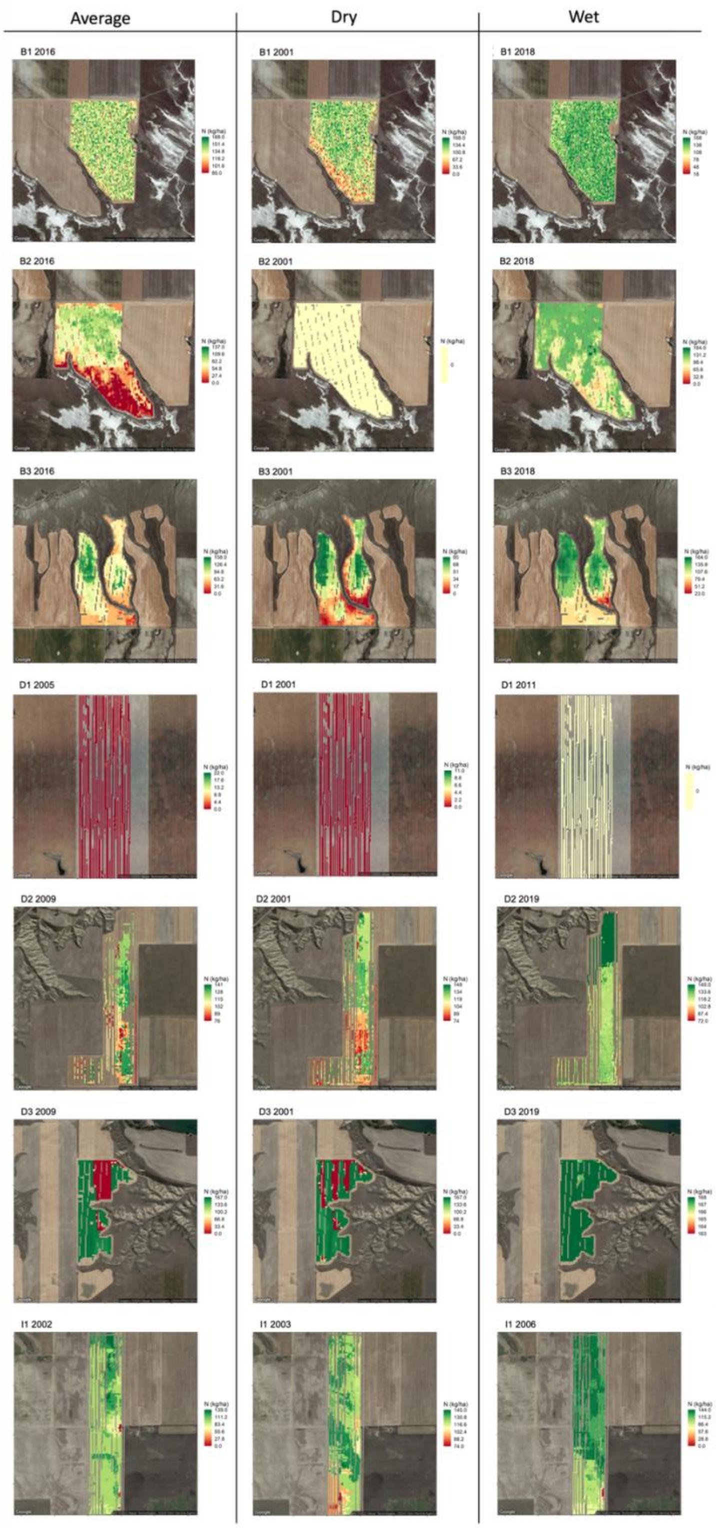


Figure 7.5. Site-specific optimized nitrogen rates (kg ha^{-1}) averaged across repetitions of the simulation for each field and weather condition. Note that scales are dependent on the field and weather condition to represent the variation across fields.

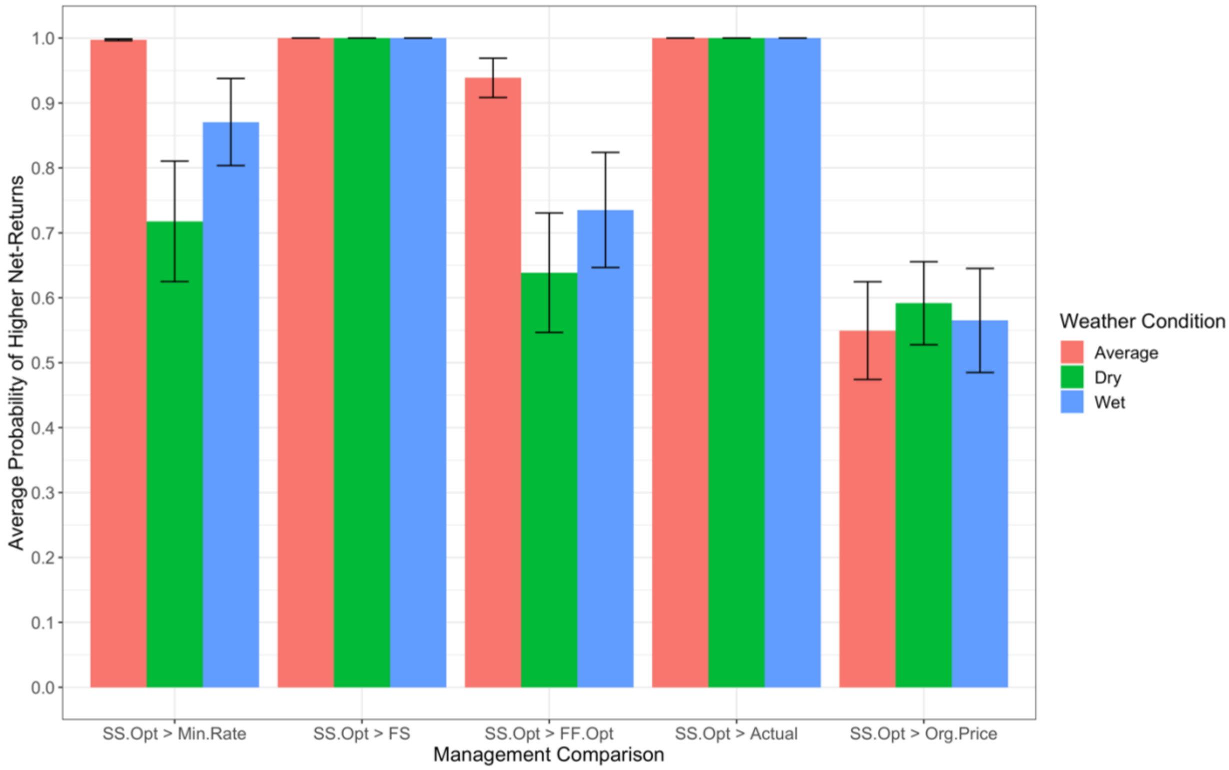


Figure 7.6. Average probability across fields that the 'SS.Opt' approach yielded a higher net-return compared to each of the other five strategies in the different weather conditions. The error bars represent the 95% confidence interval derived from the distribution of average probabilities across fields.

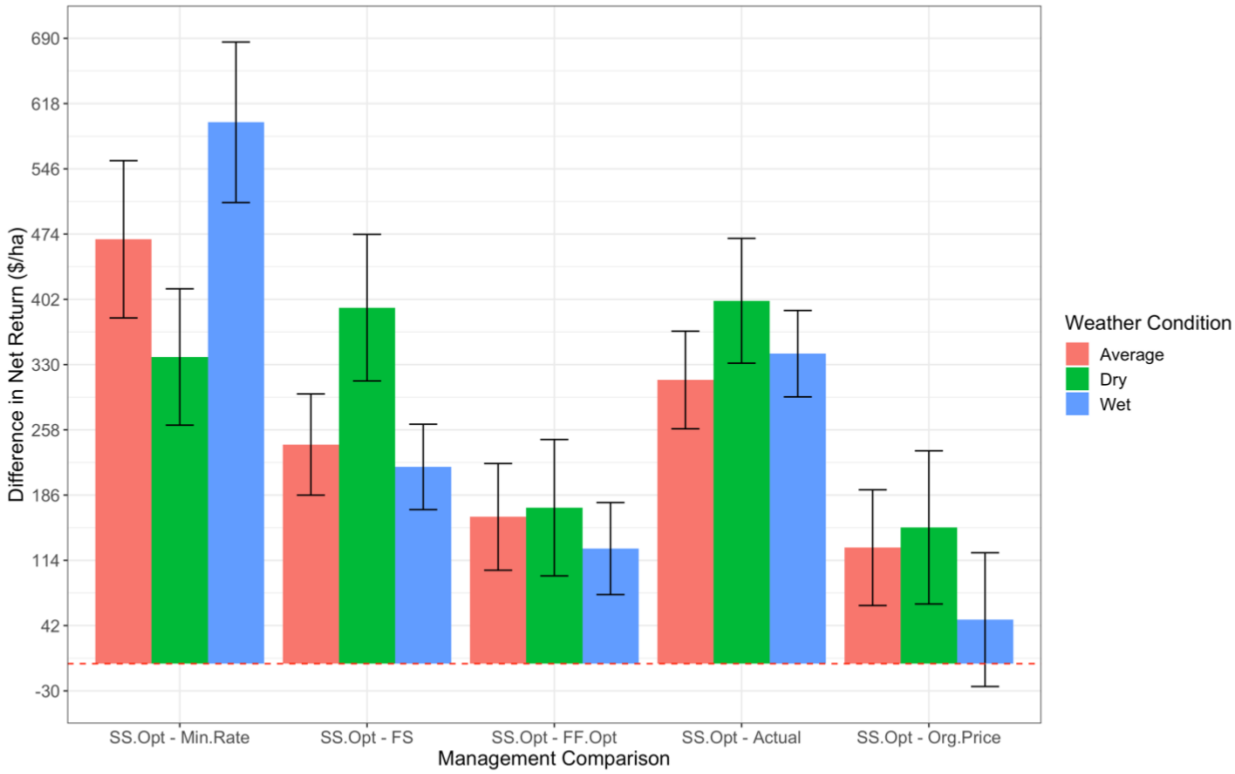


Figure 7.7. Average difference in net-returns across fields between the 'SS.Opt' approach and each of the other five strategies in the different weather conditions. The error bars represent the 95% confidence interval derived from the distribution of average differences across fields.

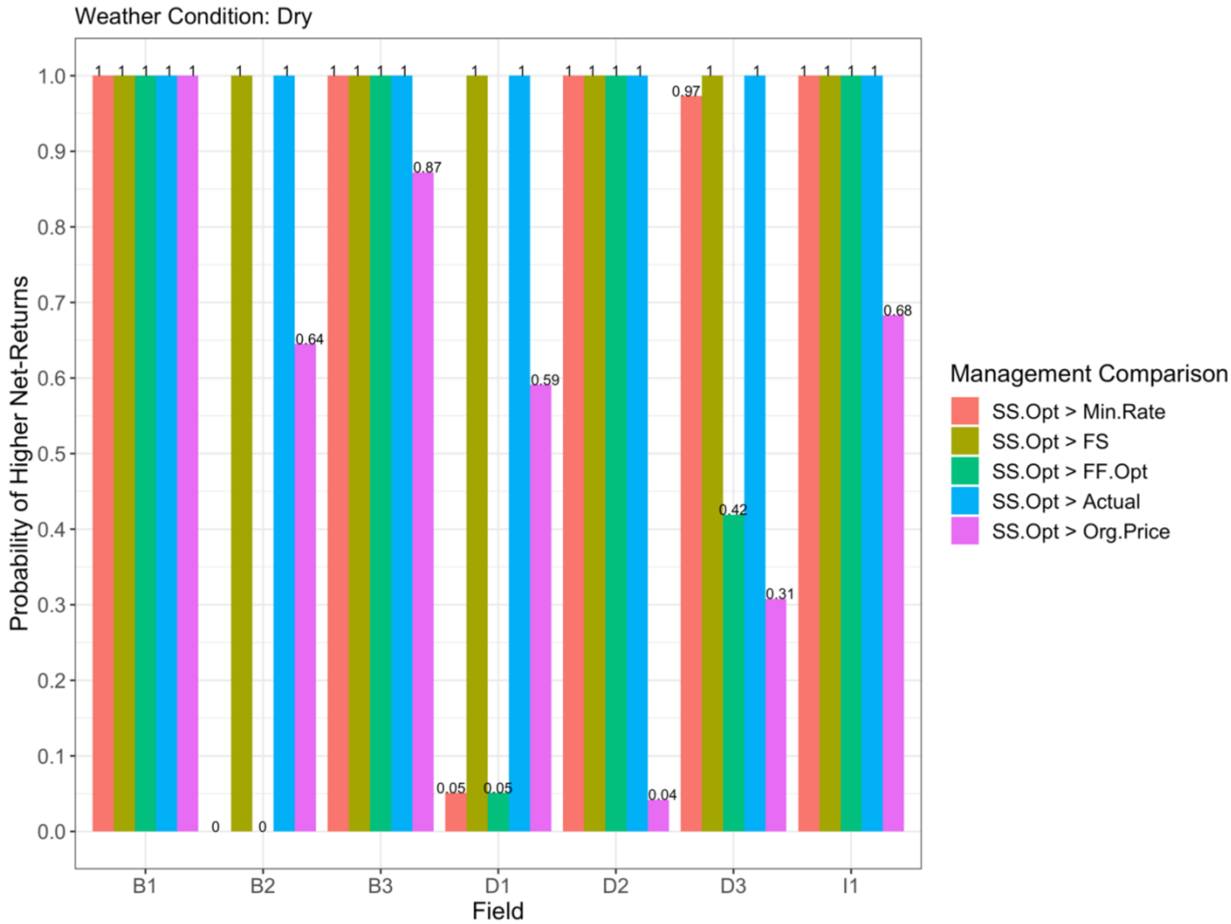


Figure 7.8. The probability that the 'SS.Opt' approach yielded a higher net-return compared to each of the other strategies on a field-specific basis when the driest precipitation year from 2000 – 2021 was simulated for each field. Probabilities are generated from the 100 iterations of each Monte Carlo simulation for a given field and weather condition.

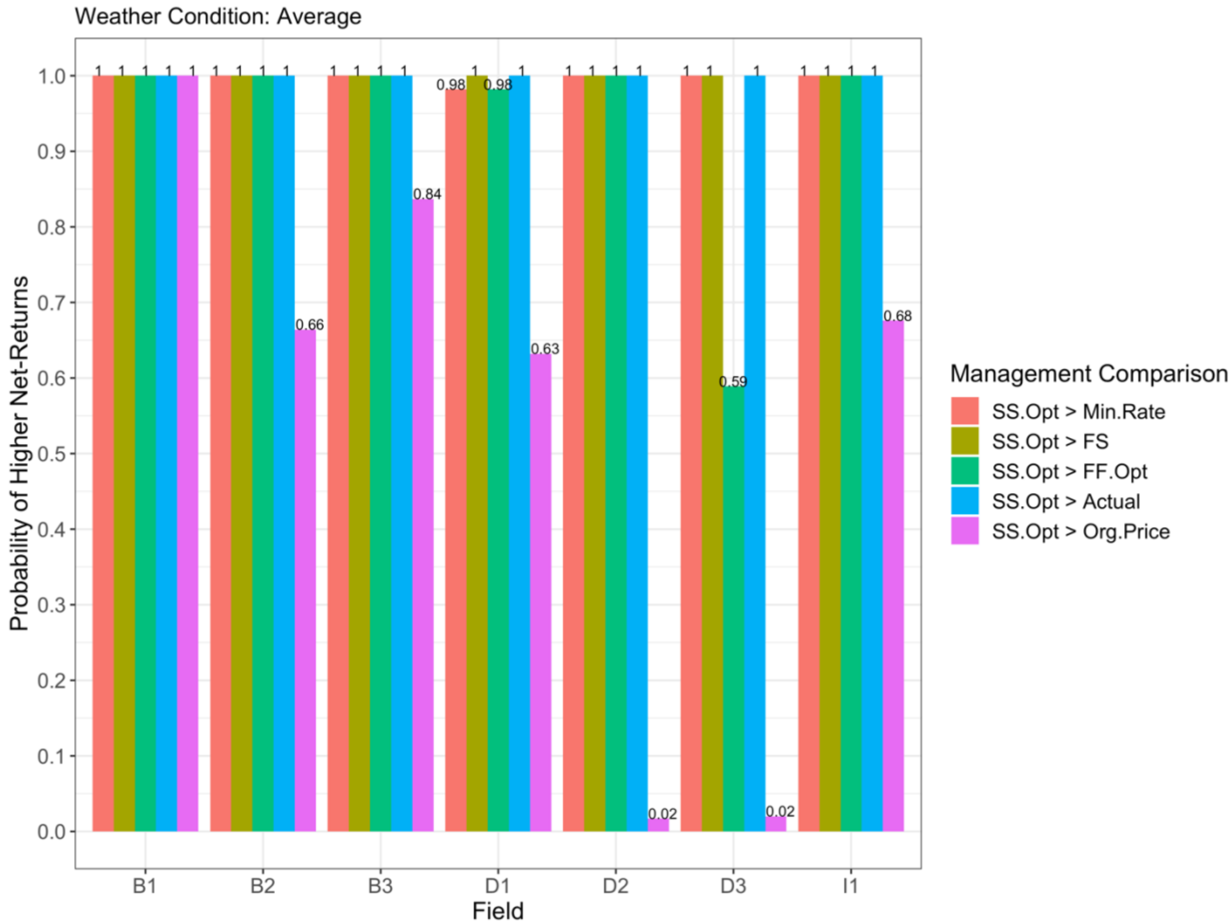


Figure 7.9. The probability that the 'SS.Opt' approach yielded a higher net-return compared to each of the other strategies on a field-specific basis when the most average precipitation year from 2000 – 2021 was simulated for each field. Probabilities are generated from the 100 realizations of each Monte Carlo ensemble for a given field and weather condition.

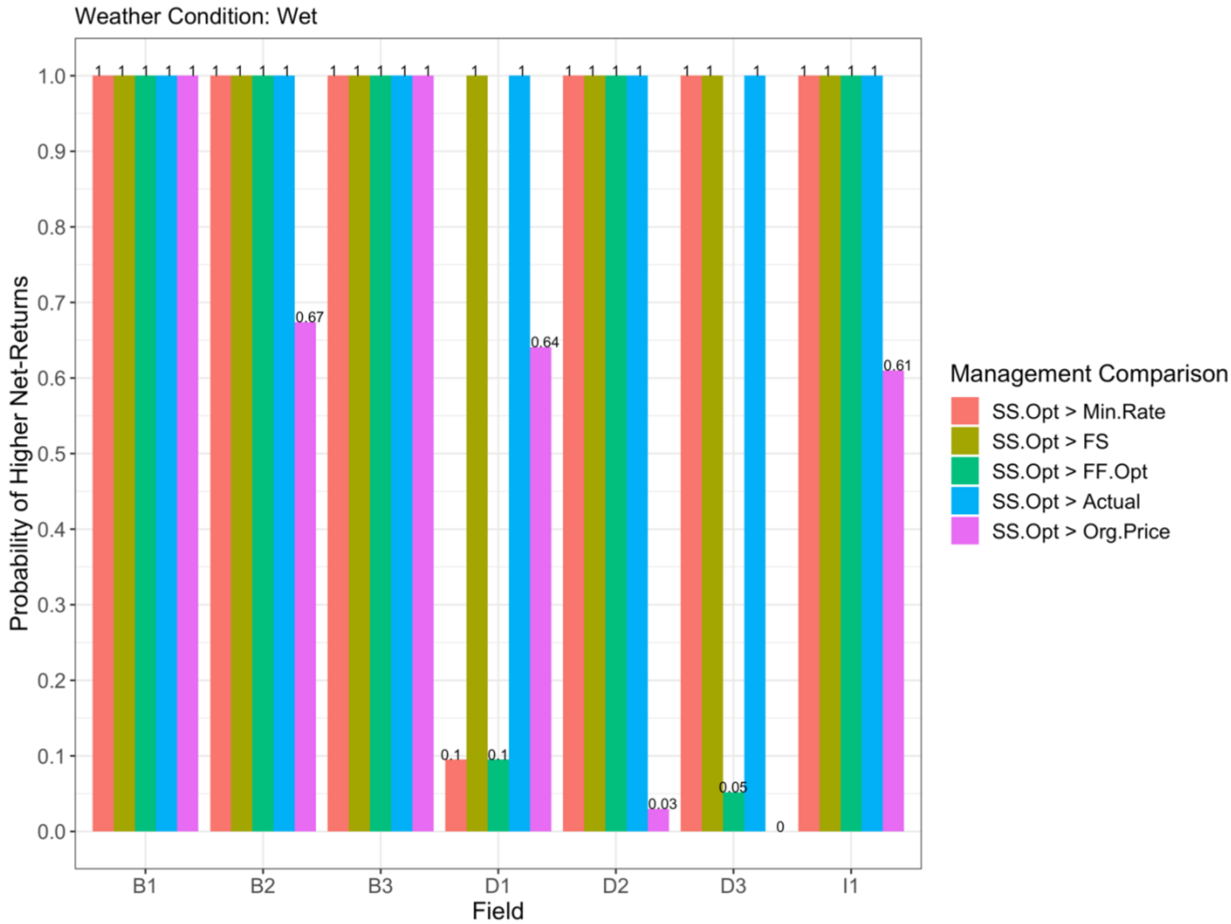


Figure 7.10. The probability that the 'SS.Opt' approach yielded a higher net-return compared to each of the other strategies on a field-specific basis when the wettest precipitation year from 2000 – 2021 was simulated for each field. Probabilities are generated from the 100 iterations of each Monte Carlo simulation for a given field and weather condition.

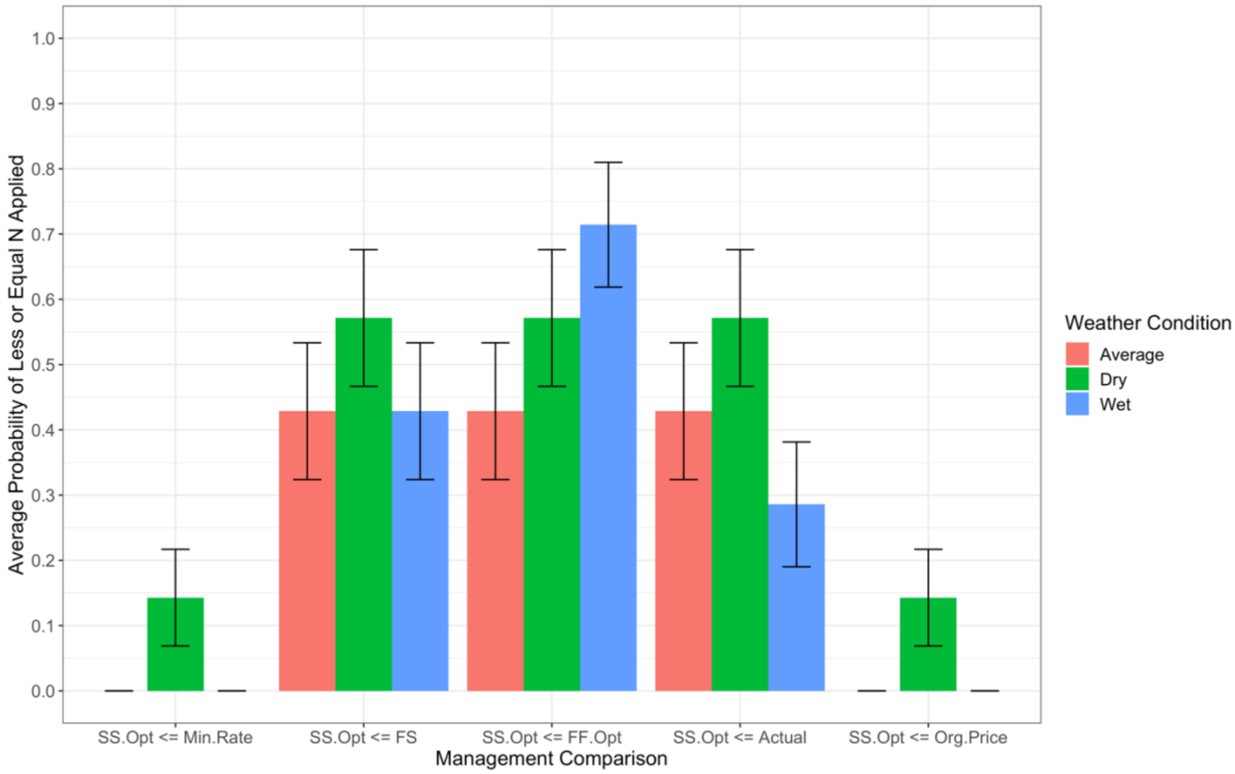


Figure 7.11. Average probability across fields that the 'SS.Opt' approach applied less or an equal amount of total N fertilizer across a field compared to each of the other five strategies. Colors represent comparisons from the simulations of three different weather conditions. The error bars represent the 95% confidence interval derived from the distribution of average probabilities across fields.

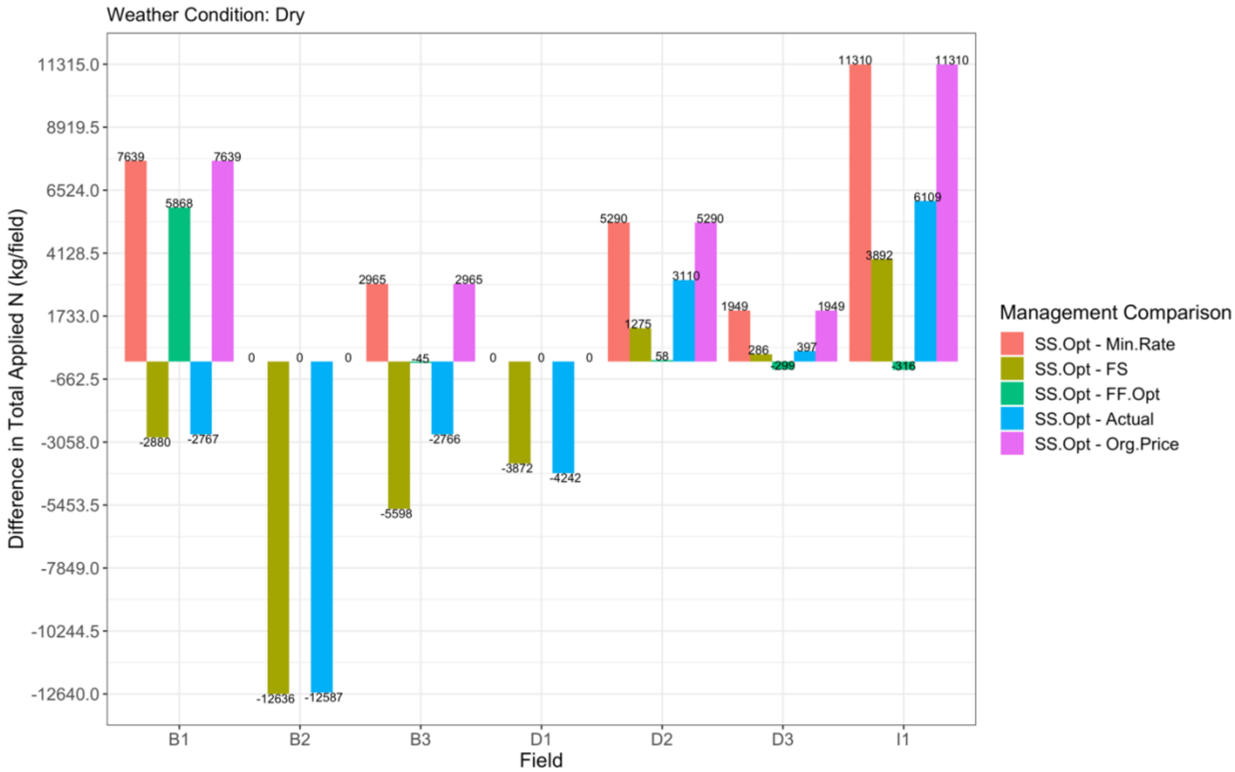


Figure 7.12. The difference in total N applied across each field with the 'SS.Opt' approach compared to each of the other five strategies in simulations where the driest year from 2000 – 2021 was selected for each field. Labels are rounded to the nearest whole number. Negative differences indicate that the 'SS.Opt' approach applied less N fertilizer than the comparison approach while positive differences indicate that the 'SS.Opt' approach applied more total N fertilizer across the field than the total N applied with the comparison approach.

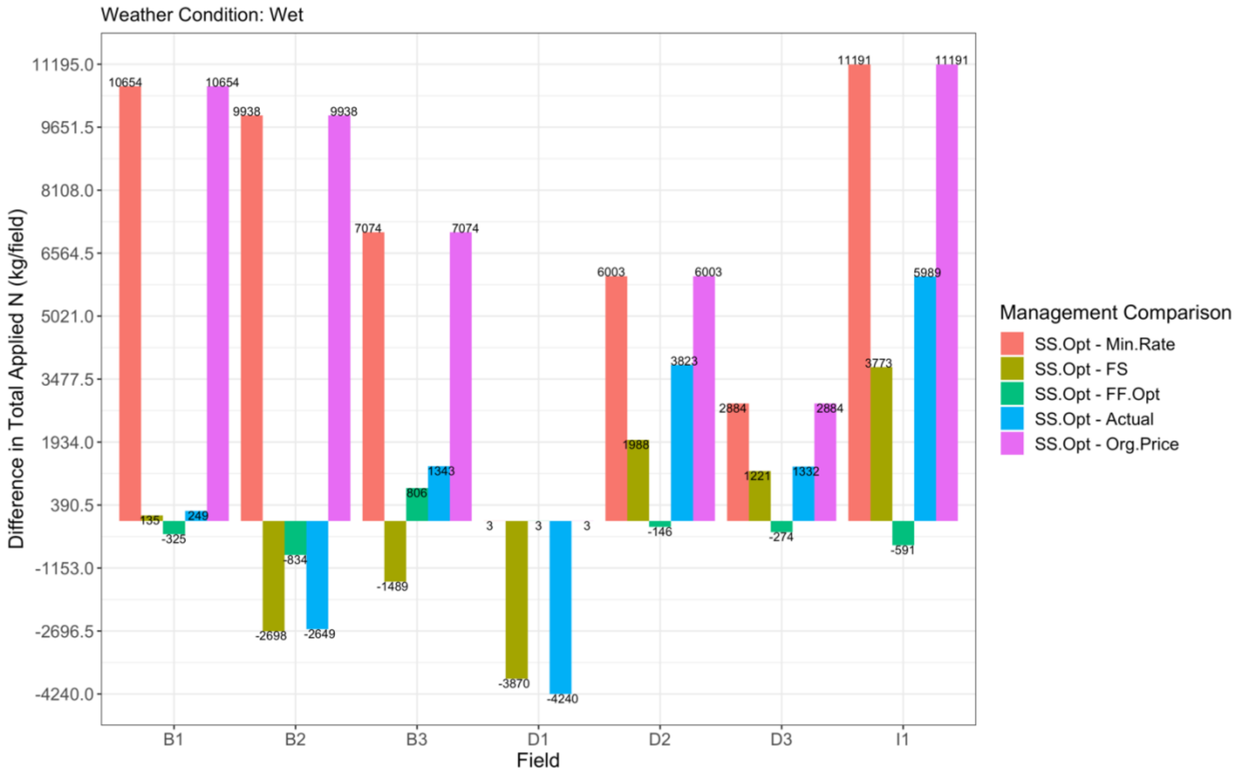


Figure 7.13. The difference in total N applied across each field with the 'SS.Opt' approach compared to each of the other five strategies in simulations where the wettest year from 2000 – 2021 was selected for each field. Labels are rounded to the nearest whole number. Negative differences indicate that the 'SS.Opt' approach applied less N fertilizer than the comparison approach while positive differences indicate that the 'SS.Opt' approach applied more total N fertilizer across the field than the total N applied with the comparison approach.

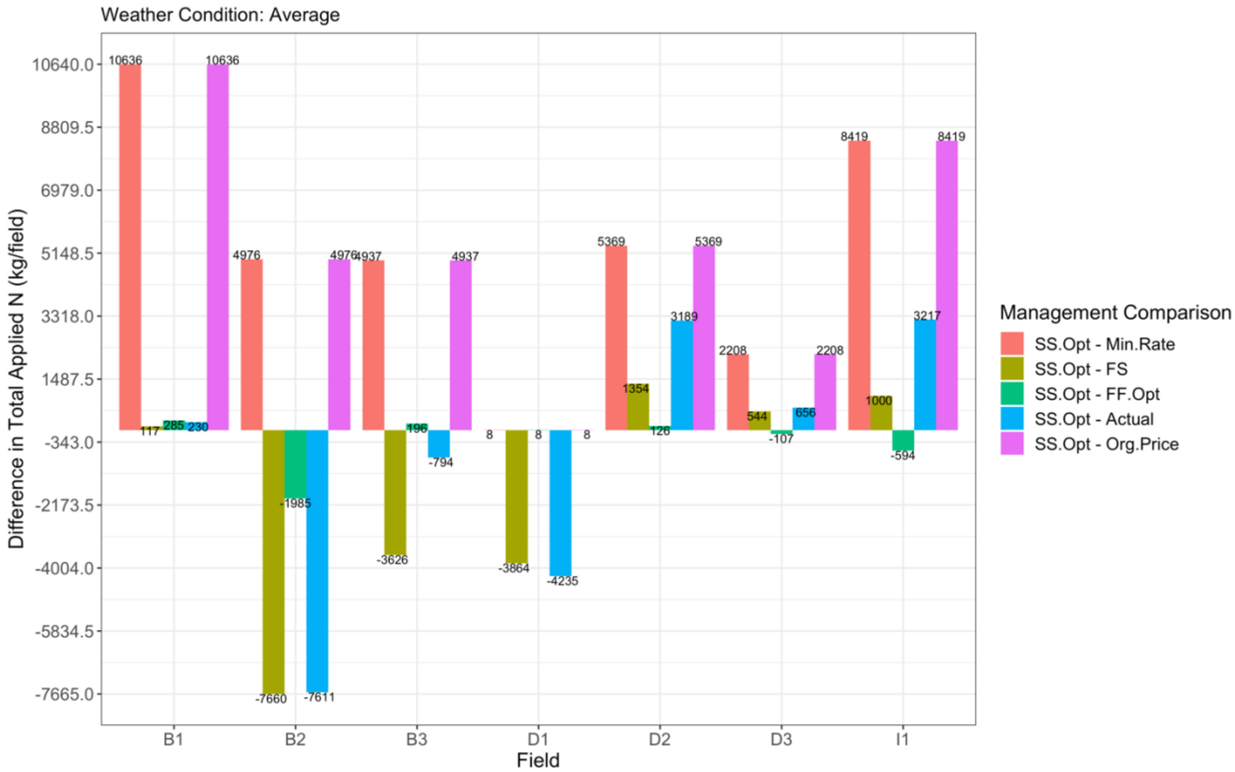


Figure 7.14. The difference in total N applied across each field with the 'SS.Opt' approach compared to each of the other five strategies in simulations where the most average year from 2000 – 2021 was selected for each field. Labels are rounded to the nearest whole number. Negative differences indicate that the 'SS.Opt' approach applied less N fertilizer than the comparison approach while positive differences indicate that the 'SS.Opt' approach applied more total N fertilizer across the field than the total N applied with the comparison approach.

Table 7.5. Total cost of nitrogen use inefficiency across fields in each weather condition. This represents the total amount of money that would potentially be lost by farmers on a field basis if they were penalized for inefficiency.

Field	Average	Dry	Wet
B1	$1991.76	$1592.12	$2010.96
B2	$1205.80	$0	$2030.79
B3	$1199.04	$798.02	$1469.12
D1	$1.66	$0.07	$0.35
D2	$1696.24	$1693.82	$1851.76
D3	$621.83	$503.62	$809.27
I1	$1063.97	$1465.36	$1433.28

REFERENCES CITED

Antle, J. M., & Valdivia, R. O. (2021). Trade-off analysis of agri-food systems for sustainable research and development. *Q Open*, *1*(1), 1–34. https://doi.org/10.1093/qopen/qoaa005

Argento, F., Anken, T., Abt, F., Vogelsanger, E., Walter, A., & Liebisch, F. (2021). Site-specific nitrogen management in winter wheat supported by low-altitude remote sensing and soil data. *Precision Agriculture*, *22*(2), 364–386. https://doi.org/10.1007/s11119-020-09733-3

Blackmore, S., & Moore, M. (1999). Remedial Correction of Yield Map Data. *Precision Agriculture*, *1*(1), 53–66. https://doi.org/10.1023/A:1009969601387

Bucci, G., Bentivoglio, D., & Finco, A. (2018). Precision agriculture as a driver for sustainable farming systems: State of art in litterature and research. *Quality - Access to Success*, *19*(S1), 114–121.

Carlisle, L., Montenegro de Wit, M., DeLonge, M. S., Iles, A., Calo, A., Getz, C., Ory, J., Munden-Dixon, K., Galt, R., Melone, B., Knox, R., & Press, D. (2019). Transitioning to Sustainable Agriculture Requires Growing and Sustaining an Ecologically Skilled Workforce. *Frontiers in Sustainable Food Systems*, *3*(November), 96. https://doi.org/10.3389/fsufs.2019.00096

Carolan, M. (2017). Publicising Food: Big Data, Precision Agriculture, and Co-Experimental Techniques of Addition. *Sociologia Ruralis*, *57*(2), 135–154. https://doi.org/10.1111/soru.12120

Clancy, M. (2019). Real-time protein monitoring. *Next Instruments*, *June*.

Coleman, D. C., & Dec, N. (1989). *Special Feature : Ecology , Agroecosystems , and Sustainable Agriculture*. *70*(6), 12–13.

Duff, H., Hegedus, P. B., Loewen, S., Bass, T., & Maxwell, B. D. (2022). Precision Agroecology. *Sustainability*, *14*(106), 1–18. https://doi.org/https://doi.org/10.3390/su14010106

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., … Zaks, D. P. M. (2011). Solutions for a cultivated planet. *Nature*, *478*(7369), 337–342. https://doi.org/10.1038/nature10452

Gebbers, R., & Adamchuck, V. I. (2010). Precision Agriculture and Food Security. *Science*, *February*, 828–830.

Gliessman, S. (2016). Transforming food systems with agroecology. *Agroecology and Sustainable Food Systems*, *40*(3), 187–189. https://doi.org/10.1080/21683565.2015.1130765

Gliessman, S. R., & Engles, E. W. (2014). *Agroecology: The Ecology of Sustainable Food Systems* (3rd ed.).

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Grzebisz, W. (2021). Site-Specific Nutrient Management. *Agronomy*, *11*(752), 1–6. https://doi.org/https://doi.org/ 10.3390/agronomy11040752

Haas, C. A., Frimpong, E. A., & Karpanty, S. M. (2009). Ecosystems and Ecosystem-Based Management. In K. M. Moore (Ed.), *The Sciences and Art of Adaptive Management: Innovating for Sustainable Agriculture and Natural Resources Management* (pp. 106–142). Soil and Water Conservation Society.

Hegedus, P.B & Ewing, S.E. (2022). Evaluation of nitrogen use efficiency in dryland winter-wheat agroecosystems for predictive modeling. Manuscript in preparation.

Hegedus, P.B & Maxwell, B.D. (2022a). Rationale for field specific on-farm precision experimentation. *Agriculture, Ecosystems & Environment.* Manuscript submitted for publication, in review.

Hegedus, P.B. & Maxwell, B.D. (2022b). Assessing performance of empirical models for forecasting crop responses to variable fertilizer rates using on-farm precision experimentation. *Precision Agriculture.* Manuscript submitted for publication, in review.

Hegedus, P.B. & Maxwell, B.D. (2022c). Best use of modern data for field-specific decision support. *Precision Agriculture.* Manuscript submitted for publication, in review.

Hegedus, P.B., Maxwell, B.D., Loewen, S., Sheppard J.W., Morales G., & Peerlinck, A. (2022). On-Farm Precision Experiments (OFPE) framework: tapping local data to optimize field scale decisions. Manuscript in preparation.

Jones, C. (2018). *Soil Acidification: A growing concern for Montana farmers*. United States Department of Agriculture, Natural Resource Conservation Service.

Kanter, D. R., Musumba, M., Wood, S. L. R., Palm, C., Antle, J., Balvanera, P., Dale, V. H., Havlik, P., Kline, K. L., Scholes, R. J., Thornton, P., Tittonell, P., & Andelman, S. (2018). Evaluating agricultural trade-offs in the age of sustainable development. *Agricultural Systems*, *163*, 73–88. https://doi.org/10.1016/j.agsy.2016.09.010

Koch, B., Khosla, R., Frasier, W. M., Westfall, D. G., & Inman, D. (2004). Economic feasibility of variable-rate nitrogen application utilizing site-specific management zones. *Agronomy Journal*, *96*(6), 1572–1580. https://doi.org/10.2134/agronj2004.1572

Lawrence, P. G., Rew, L. J., & Maxwell, B. D. (2015). A probabilistic Bayesian framework for progressively updating site-specific recommendations. *Precision Agriculture*, *16*(3), 275–296. https://doi.org/10.1007/s11119-014-9375-4

Link, J., Batchelor, W. D., Graeff, S., & Claupein, W. (2008). Evaluation of current and model-based site-specific nitrogen applications on wheat (Triticum aestivum L.) yield and environmental quality. *Precision Agriculture*, *9*(5), 251–267. https://doi.org/10.1007/s11119-008-9068-y

McBratney, A., Whelan, B., Ancev, T., & Bouma, J. (2005). Future directions of precision agriculture. *Precision Agriculture*, *6*(1), 7–23. https://doi.org/10.1007/s11119-005-0681-8

Meola, A. (2021). Smart Farming in 2020 : How IoT sensors are creating a more efficient precision agriculture industry. *Business Insider*.

Meyer-Aurich, A., Weersink, A., Gandorfer, M., & Wagner, P. (2010). Optimal site-specific fertilization and harvesting strategies with respect to crop yield and quality response to nitrogen. *Agricultural Systems*, *103*(7), 478–485. https://doi.org/10.1016/j.agsy.2010.05.001

Mueller, J. P., Finney, D., & Hepperly, P. (2009). The Field System. In K. M. Moore (Ed.), *The Sciences and Art of Adaptive Management: Innovating for Sustainable Agriculture and Natural Resources Management* (pp. 25–59). Soil and Water Conservation Society.

Paul, E. A., & Robertson, G. P. (1989). Ecology and the Agricultural Sciences: A False Dichotomy? *Ecology*, *70*(6), 1594–1597. https://doi.org/10.2307/1938091

Pierpaoli, E., Carli, G., Pignatti, E., & Canavari, M. (2013). Drivers of Precision Agriculture Technologies Adoption: A Literature Review. *Procedia Technology*, *8*(Haicta), 61–69. https://doi.org/10.1016/j.protcy.2013.11.010

Schimmelpfennig, D., & Lowenberg-DeBoer, J. (2020). Farm Types and Precision Agriculture Adoption: Crops, Regions, Soil Variability, and Farm Size. *SSRN Electronic Journal*, 1–38. https://doi.org/10.2139/ssrn.3689311

Sigler, W. A., Ewing, S. A., Jones, C. A., Payn, R. A., Brookshire, E. N. J., Klassen, J. K., Jackson-Smith, D., & Weissmann, G. S. (2018). Connections among soil, ground, and surface water chemistries characterize nitrogen loss from an agricultural landscape in the upper Missouri River Basin. *Journal of Hydrology*, *556*, 247–261. https://doi.org/10.1016/j.jhydrol.2017.10.018

Sigler, W. A., Ewing, S. A., Jones, C. A., Payn, R. A., Miller, P., & Maneta, M. (2020). Water and nitrate loss from dryland agricultural soils is controlled by management, soils, and weather. *Agriculture, Ecosystems and Environment*, *304*(July), 107158. https://doi.org/10.1016/j.agee.2020.107158

Sudduth, K. A., & Drummond, S. T. (2007). Yield editor: Software for removing errors from crop yield maps. *Agronomy Journal*, *99*(6), 1471–1482. https://doi.org/10.2134/agronj2006.0326

Sudduth, K. A., Drummond, S. T., & Myers, D. B. (2012). Yield editor 2.0: Software for automated removal of yield map errors. *American Society of Agricultural and Biological Engineers Annual International Meeting 2012, ASABE 2012*, *4*(12), 3378–3391. https://doi.org/10.13031/2013.41893

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108

Wagner, P., & Hank, K. (2013). Suitability of aerial and satellite data for calculation of site-specific nitrogen fertilisation compared to ground based sensor data. *Precision Agriculture*, *14*(2), 135–150. https://doi.org/10.1007/s11119-012-9278-1

Wyeth, P. (2009). Sustainable Agriculture and Natural Resource Manageemnt in Farm Enterprise Systems. In K. M. Moore (Ed.), *The Sciences and Art of Adaptive Management: Innovating for Sustainable Agriculture and Natural Resources Management* (pp. 60–77). Soil and Water Conservation Society.

Yin, X., Goudriaan, J., Lantinga, E. A., Vos, J., & Spiertz, H. J. (2003). A flexible sigmoid function of determinate growth. *Annals of Botany*, *91*(3), 361–371. https://doi.org/10.1093/aob/mcg029

CHAPTER EIGHT

EPILOGUE

Big data and agriculture have been used together more and more frequently in the last decade, as the amount of information available about farms (from machines, drones, weather stations, sensors, and from satellite based remote sensing) becomes available at a high frequency (Carolan, 2017). Precision agriculture was developed to harness this technology to farm "smarter" yet is a broad term that spans simply using GPS and satellite guidance, to targeted application of inputs such as fertilizers and pesticides. In recent decades, precision agriculture technologies have become commonplace across modern farms in the United States and the globe, with a market for precision agriculture exceeding US$ 2.5 billion in 2014 (Carolan, 2017; Michalopoulos 2015). Most farm machinery today collects, or at least interacts, with data in one or many ways, and data is a rich resource being generated on and for farms every day (Meola, 2021). Although vast amounts of data are and have been acquired on farms for several decades, agriculture has largely left these data unused to inform management decisions, with precision agriculture largely focused on GPS and GIS technologies (Schimmelpfennig & Lowenberg-DeBoer, 2020).

Advances in data availability associated with precision agriculture allow producers to manage at the subfield scale. As expected in a US $2.5+ billion market, there are many companies invested in precision agriculture technology. Despite a lack of adoption, small- to large-scale companies have created services to prescribe agriculture inputs, such as fertilizer rates. These services come in the form of software, such as Climate Corporation's Climate Field View (https://climate.com), Trimble Farmer Core (https://agriculture.trimble.com/product/farmer-core/) , Ag Leader SMS (https://www.agleader.com/farm-management/sms-software/), and Adapt–N (http://www.adapt-n.com). Most software provides recommendations for specific fields, though based on the analysis of data from many other fields or small plot studies (Carolan, 2017). While companies apply their models to the data from each specific field, they assume that the general average response model is appropriate for every field and year, and they neglect variation in crop responses within and between fields, as well as over time.

Models typically fail when extrapolated across scales, as the factors influencing processes at one scale may not coincide with factors influencing processes at another scale, a likely pitfall of the general NUE model developed. We observe phenomena at an observational scale above the scale at which natural phenomena and processes occur, and then model at yet larger scale (Blöschl & Sivapalan, 1995). A common limitation to most crop response models is applying them beyond the scope and scale for which they were developed (Blöschl & Sivapalan, 1995; Pilkey & Pilkey-Jarvis, 2008). If models are developed from data gathered at a specific place or place in time, there is no guarantee they will work in another place or at another time (Blöschl & Sivapalan, 1995; Fritsch et al., 2020). Conversely, developing generalized models from large diverse datasets does not guarantee that specific places or times will behave in a generalized or average way. Every model is a simplification of reality, and thus the outcome of natural processes can never be absolutely predicted because of the complexity of nature, natural variation, and limitations in our ability to observe the natural world (Pilkey-Jarvis & Pilkey, 2008). Pilkey and Pilkey-Jarvis argue, “nature does not use averages”, yet most models require a smoothing of the complexity of the system, reducing natural variation that is likely important, resulting in models that typically fail at generating accurate predictions of future phenomena (Blöschl & Sivapalan, 1995; Oreskes, 2000). (2008).

Farmers are managers of ecological systems, which contain the inescapable attribute of complex interactions among a host of biotic and abiotic factors, resulting in uncertainty in management outcomes (Walters & Hilborn, 1978). Site, time, and history specificity associated with agricultural fields and the decisions required to manage crops in those fields has challenged integration of ecological principles into agriculture (Maxwell & Luschei, 2005). Agroecosystem futures at any time scale are unpredictable and so management must be able to adapt and be flexible (Walters, 1986), yet agroecosystems are managed to homogenize the response of the crop with the goal of maximizing production and profit (Coleman, 1989; Gliessman & Engles, 2014; Paul & Robertson, 1989). However, homogenization requires a single state of equilibrium at which variation can be controlled through the addition of inputs, such as fertilizer or seeding rates, to attain maximum potential yield. Yet, the equilibrium of ecosystems is determined by destabilizing forces, and ecosystems potentially have zero to multiple equilibria (Holling, 1996). The fluctuation of an agroecosystem's equilibrium, or stable state, means that these systems vary in their responses to biotic and abiotic forces, including the addition of inputs. Management recommendations thus cannot neglect the site-specificity of crop responses to inputs, and the interaction with the localized environment.

Despite the limitations of contemporary precision agriculture, the promise of precision agriculture technology and data collection to inform small-scale management across fields should not be discounted (Cassman, 1999; Gebbers & Adamchuck, 2010). Reducing agricultural inputs like fertilizers and pesticides is perceived to increase uncertainty in management outcomes. Thus, there is a significant opportunity to employ modern data science in agriculture, one of the most data rich arenas in the world, to help understand the complexity underpinning these systems and inform ecologically-based management (Sykuta, 2016). Utilizing data and technology to understand and address the scales at which variation in net-returns, biology, and weather occur and interact is vital to recommending the application of the right input, at the right time, in the right place, and in the right amount. The development of variable rate technology and on-the-go crop quantity and quality monitoring devices allows managers to experimentally vary inputs and observe the response of crop quality and production within each field. The constant requirement of data collection and analysis sets up precision agriculture perfectly for adaptive management, where the tenets are monitor, experiment, repeat (Haas et al., 2009; Moore, 2009; Mueller et al., 2009; Wyeth, 2009). Adaptive management applies experimentation with variable input rates to understand the response of the crop in a specific field and to use that information to inform site-specific management within that field. Field specific data can be used to inform small-scale management within fields by reducing generalizations made across spatial scales (Lawrence, Rew, & Maxwell, 2015; Maxwell & Luschei, 2005; Gebbers & Adamchuk 2010; Luschei et al. 2001). Bringing field-specific data to managers reduces information degradation associated with production recommendations developed based on geographic proximity to research sites, or assumptions related to generalized models (Maxwell & Luschei, 2005).

Adaptive management in agroecosystems has the advantage of frequent manipulations, yielding increasingly large amounts of site-specific data that allow for rigorous empirical assessment of system performance at the field scale. Therefore, adaptive management applied to crop production can start with formalized experimentation within a field to quantify spatial variability and, when repeated, to quantify temporal variability. The experimentation accelerates system knowledge, removing the confounding effects stemming from trial and error that have traditionally driven farm management. Adaptive management recognizes and aims to understand uncertainty in the managed system and acknowledges that there is not a single management method to optimize producer goals. When repeatedly applied in agroecosystems, adaptive management feeds an iterative decision-making process that balances learning and producer goals in the face of uncertainty associated with crop production.

Pilkey and Pilkey-Jarvis (2008) advocate for adaptive management to avoid shortcomings of quantitative models. Our on-farm precision experiment (OFPE) approach advocates for adaptive management and is highly dependent on models. Blöschl & Sivapalan (1995) categorize models as either predictive, to obtain an answer to a specific problem, or investigative, to further our understanding of process. The models that we (and industry) are developing fit comfortably in the predictive category as our work has an applied focus of generating decision support systems.

But why believe our models? In terms of scale issues and scope of inference, we do not attempt to develop models that can be generalized beyond the specific field they are developed and restrict our inference to only the field from which the data came. We develop our models using data collected from each individual field to be used only for that field, in recognition that a model that may work in one field may not be the best model to use even in an adjacent field. However, while our models are restricted to usage in only the locale the data used to develop them is from (the field), we are subject to error when making predictions of what future net-return will be by extrapolating over time.

We recognize the shortcomings of predictions in the future, develop probabilistic outcomes, and avoid absolutes, when possible, via our Monte Carlo simulation approach. Our repetitive process of experimentation, collection of on-farm and remotely sensed data, and analysis is an iterative approach that sequentially updates our models as more information becomes available will steadily improve the ability to predict future crop responses . Based on this approach, we update, refine, and tune our models with variation in environmental processes that make our model more robust to predictions in the future. We also apply adaptive analysis, where we do not assume one model form that fits all fields or even the same field in different years. By experimenting with multiple models, we effectively test competing hypotheses for system behavior and, avoid the pitfall of assuming one model is best.

The tools and models that utilize precision agriculture technology and data to accurately forecast crop responses are crucial for making decisions on-farms and for drafting policy decisions. Agricultural innovation and technology have flourished over the past twenty years, yet sustainability and social issues plague food systems, necessitating policy to drive innovation and technology to focus not only on production, but on minimization of practices that degrade the environmental resources on which agriculture depends (Jones et al., 2021). Policy decisions trickle down into the decision-making matrix for farmers, where policy influences management decisions on-farms by giving carrots and sticks (incentives and disincentives) for management choices at the farm and field scale.

Data and artificial intelligence have the capacity to transform agriculture at all scales and provide opportunities to transform the agricultural landscape in a more sustainable direction (Egli et al., 2018; Meola, 2021; Peters et al., 2020). Due to the importance of agriculture to individual, regional, national, and global economics, policy must be informed by high quality data and disincentivize management practices that counteract agricultural sustainability (King et al., 2019). Integrating information from farms and from rural communities into policy decisions can provide policy makers the data required to govern future agricultural practices that balance the need for increased production and minimization of ecological detriments associated with agriculture. Impactful change on farms requires decision support systems for farmers that have accurate forecasts of future crop response to agricultural inputs, combined with data-driven policy that provides the appropriate incentive/disincentive in the right situations.

While decisions made on farms or fields may seem insignificant, the sum of the actions on farms and fields have regional and global effects that influence sustainability of agriculture beyond the farm scale (Gross, 2013). The advent of big data, the internet of things, artificial intelligence, and robotics provide huge benefits for farmer decision making, but also provide opportunities to promote sustainability across the entire food system. However maintenance of these opportunities resides with policy makers to enable utilization and application of this data (Wolfert et al., 2017).

With precision agriculture and the data revolution being adopted by large enterprises and companies, small and medium sized companies are falling behind in terms of the adoption and application of precision agriculture and utilization of, despite the promise of technology and data to increase sustainable management of agroecosystems (Bucci et al., 2018). Rural populations are declining, and farm sizes are increasing, due to the trend of agriculture towards technological replacements of farmer knowledge, thus making policy decisions that empower and support farmers with technical and ecological skills vital to utilizing the data stream from farms in a way that can improve the sustainability of agriculture. We must recognize the varying attitudes and rates of adoption of precision agriculture technology and data of farmers and find policy incentives for groups that have already adopted precision agriculture and those that have not (Pierpaoli et al., 2013). In regard to conservation and minimization of agricultural pollution, policy needs to recognize that farmers vary in the way they think about their systems, indicating that policy requires diversification to incentivize late adopters while rewarding early adopters of technology and sustainable practices (Church et al., 2020). Agricultural policy also needs to recognize the value of local farmer knowledge and understand the potential outcomes of governance to avoid ecological dystopias where technology and robotics drive rural communities into irrelevance by increasing farm sizes and excluding farmers from decisions (Daum, 2021).

The digital revolution in agriculture has resulted in low-cost data available to all farms via satellite and remote sensing information, which are recognized as potential avenues for increasing food production in a sustainable manner (Diacono et al., 2013). Yet availability and usability of this data remains a barrier to adoption on most farms (Weersink et al., 2018). One way to promote sustainable agriculture is by drafting policy that encourages academic researchers to develop low-cost decision support systems that empower farmers to use this open-source data in their decision making. However, the majority of scientists and researchers have indicated that a lack of public research funding is a major obstacle to researching sustainable agriculture (DeLonge et al., 2020; Miles et al., 2017). It must be recognized that precision agriculture technology remains unattainable for small and subsistence farmers; however, on-farm experimentation and participatory research provides an avenue for these farms to increase sustainability, learn the skills and techniques necessary for harnessing data from open-sources, and understand how to make informed management decision on their farms (Cook et al., 2004; MacMillan & Benton, 2014).

This dissertation and the OFPE framework reflect my commitment to contribute to increased sustainability in agricultural systems. An overarching theme of my dissertation was observing, understanding, and optimizing agroecosystems with respect to N fertilizer management in dryland winter-wheat systems of Montana. The experimental basis of the OFPE framework allows farmers and crop managers to understand the ecological relationship between crops in their fields to variable N applications and the surrounding environment. Work by Sigler et al. (2018, 2020, 2022) in the Judith River Watershed of central Montana provided the baseline understanding of N dynamics in dryland Montana agroecosystems upon which we created a model for NUE. Gathering environmental data from open-source repositories delivered the opportunity to understand how edaphic, weather, and vegetation index parameters relate to crop production and quality, as well as providing the covariates necessary for identifying the type of model most appropriate for 1) characterizing crop responses on specific fields and 2) predicting nitrogen use efficiency (NUE) on a subfield scale. Using the combination of crop response and NUE models, we were able to optimize nitrogen rates that maximize farmer profit and minimize the risk of pollution due to inefficient N use.

Farmers with precision agriculture technology have the tools to make data-driven management decisions that maintain production as well as the soil and water resources needed for future production. Conventional agriculture will not transition to sustainable practices overnight but will take a step-by-step process beginning with increasing the efficiency of N fertilizer and other chemical inputs. Increasing global agricultural efficiency and sustainability cannot occur unless efficiency and sustainability are prioritized on a field-by-field basis. The OFPE framework gives farmers the guidance to make more sustainable field-specific management decisions, while including them in the decision-making framework. With advancements in the technology, equipment, and spatiotemporal resolution and accuracy of sensor data, the OFPE framework can help to establish a sustainable future for modern agriculture.

REFERENCES CITED

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, *9*(3–4), 251–290. https://doi.org/10.1002/hyp.3360090305

Bucci, G., Bentivoglio, D., & Finco, A. (2018). Precision agriculture as a driver for sustainable farming systems: State of art in litterature and research. *Quality - Access to Success*, *19*(S1), 114–121.

Carolan, M. (2017). Publicising Food: Big Data, Precision Agriculture, and Co-Experimental Techniques of Addition. *Sociologia Ruralis*, *57*(2), 135–154. https://doi.org/10.1111/soru.12120

Cassman, K. G. (1999). Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(11), 5952–5959. https://doi.org/10.1073/pnas.96.11.5952

Church, S. P., Lu, J., Ranjan, P., Reimer, A. P., & Prokopy, L. S. (2020). The role of systems thinking in cover crop adoption: Implications for conservation communication. *Land Use Policy*, *94*(December 2019). https://doi.org/10.1016/j.landusepol.2020.104508

Coleman, D. C., & Dec, N. (1989). *Special Feature : Ecology , Agroecosystems , and Sustainable Agriculture*. *70*(6), 12–13.

Cook, S., Cock, J., Oberthür, T., & Fisher, M. (2004). On-Farm Experimentation. *Better Crops*, *97*, 17–20.

Daum, T. (2021). Farm robots: ecological utopia or dystopia? *Trends in Ecology and Evolution*.

DeLonge, M., Robbins, T., Basche, A., & Haynes-Mawlow, L. (2020). The State of Sustainable Agriculture and Agroecology Research and Impacts: A Survey of U.S. Scientists. *Journal of Agriculture, Food Systems, and Community Development*, *9*(2), 1–26. https://doi.org/10.5304/jafscd.2020.092.009

Diacono, M., Rubino, P., & Montemurro, F. (2013). Precision nitrogen management of wheat. A review. *Agronomy for Sustainable Development*, *33*(1), 219–241. https://doi.org/10.1007/s13593-012-0111-z

Egli, L., Meyer, C., Scherber, C., Kreft, H., & Tscharntke, T. (2018). Winners and losers of national and global efforts to reconcile agricultural intensification and biodiversity conservation. *Global Change Biology*, *24*(5), 2212–2228. https://doi.org/10.1111/gcb.14076

Fritsch, M., Lischke, H., & Meyer, K. M. (2020). Scaling methods in ecological modelling. *Methods in Ecology and Evolution*, *11*(11), 1368–1378. https://doi.org/10.1111/2041-210X.13466

Gliessman, S. R., & Engles, E. W. (2014). *Agroecology: The Ecology of Sustainable Food Systems* (3rd ed.).

Gross, M. (2013). The paradoxical evolution of agriculture. *CURBIO*, *23*(16), R667–R670. https://doi.org/10.1016/j.cub.2013.08.001

Haas, C. A., Frimpong, E. A., & Karpanty, S. M. (2009). Ecosystems and Ecosystem-Based Management. In K. M. Moore (Ed.), *The Sciences and Art of Adaptive Management: Innovating for Sustainable Agriculture and Natural Resources Management* (pp. 106–142). Soil and Water Conservation Society.

Holling, C. S. (1996). Surprise for science, resilience for ecosystems, and incentives for people. *Ecological Applications*, *6*(3), 733–735. https://doi.org/10.2307/2269475

Jones, J., Verma, B., Basso, B., Mohtar, R., & Matlock, M. (2021). Transforming food and agriculture to circular systems: A perspective for 2050. *Resource: Engineering and Technology for Sustainable World*, *28*(2), 7–9.

King, C. L. Mackie, C. (2019). Improving Data Collection and Measurement of Complex Farms. In *Improving Data Collection and Measurement of Complex Farms*. National Academies Press. https://doi.org/10.17226/25260

Lawrence, P. G., Rew, L. J., & Maxwell, B. D. (2015). A probabilistic Bayesian framework for progressively updating site-specific recommendations. *Precision Agriculture*, *16*(3), 275–

296. https://doi.org/10.1007/s11119-014-9375-4

MacMillan, T., & Benton, T. (2014). Engage farmers in research. *Nature, 509*, 25–27.

Maxwell, B. D., & Luschei, E. C. (2005). Justification for site-specific weed management based on ecology and economics. *Weed Science, 53*(2), 221–227. https://doi.org/10.1614/ws-04-071r2

Meola, A. (2021). Smart Farming in 2020 : How IoT sensors are creating a more efficient precision agriculture industry. *Business Insider*.

Miles, A., DeLonge, M. S., & Carlisle, L. (2017). Triggering a positive research and policy feedback cycle to support a transition to agroecology and sustainable food systems. *Agroecology and Sustainable Food Systems*, *41*(7), 855–879. https://doi.org/10.1080/21683565.2017.1331179

Moore, K. M. (2009). Landscape systems framework for adaptive management. In K. M. Moore (Ed.), *The sciences and art of adaptive management: innovating for sustainable agriculture and natural resource management* (pp. 3–22). Soil and Water Conservation Society.

Mueller, J. P., Finney, D., & Hepperly, P. (2009). The Field System. In K. M. Moore (Ed.), *The Sciences and Art of Adaptive Management: Innovating for Sustainable Agriculture and Natural Resources Management* (pp. 25–59). Soil and Water Conservation Society.

Oreskes. (2000). *Why believe a Computer, Models, Measures and Meaning in a Natural World.pdf.*

Paul, E. A., & Robertson, G. P. (1989). Ecology and the Agricultural Sciences: A False Dichotomy? *Ecology*, *70*(6), 1594–1597. https://doi.org/10.2307/1938091

Peters, D. P. C., Rivers, A., Hatfield, J. L., Lemay, D. G., Liu, S., & Basso, B. (2020). Harnessing AI to Transform Agriculture and Inform Agricultural Research. *IT Professional, 22*(3), 16–21. https://doi.org/10.1109/MITP.2020.2986124

Pierpaoli, E., Carli, G., Pignatti, E., & Canavari, M. (2013). Drivers of Precision Agriculture Technologies Adoption: A Literature Review. *Procedia Technology*, *8*(Haicta), 61–69. https://doi.org/10.1016/j.protcy.2013.11.010

Pilkey-Jarvis, L., & Pilkey, O. H. (2008). Useless arithmetic: Ten points to ponder when using mathematical models in environmental decision making. *Public Administration Review*, *68*(3), 470–479. https://doi.org/10.1111/j.1540-6210.2008.00883_2.x

Schimmelpfennig, D., & Lowenberg-DeBoer, J. (2020). Farm Types and Precision Agriculture Adoption: Crops, Regions, Soil Variability, and Farm Size. *SSRN Electronic Journal*, 1–38. https://doi.org/10.2139/ssrn.3689311

Sykuta, M. E. (2016). Big Data in Agriculture: Property Rights, Privacy and Competition in Ag Data Services. *International Food and Agribusiness Management Review*, *19*, 57–74.

Walters, C. J., & Hilborn, R. (1978). Ecological Optimization and Adaptive Management. *Annual Review of Ecology and Systematics*, *9*(1), 157–188. https://doi.org/10.1146/annurev.es.09.110178.001105

Weersink, A., Fraser, E., Pannell, D., Duncan, E., & Rotz, S. (2018). Opportunities and Challenges for Big Data in Agricultural and Environmental Analysis. *Annual Review of Resource Economics*, *10*(August), 19–37. https://doi.org/10.1146/annurev-resource-100516-053654

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big Data in Smart Farming – A review. *Agricultural Systems*, *153*, 69–80. https://doi.org/10.1016/j.agsy.2017.01.023

Wyeth, P. (2009). Sustainable Agriculture and Natural Resource Manageemnt in Farm Enterprise Systems. In K. M. Moore (Ed.), *The Sciences and Art of Adaptive Management: Innovating for Sustainable Agriculture and Natural Resources Management* (pp. 60–77). Soil and Water Conservation Society.

CUMULATIVE REFERENCES CITED

Allaire, M., Wu, H., & Lall, U. (2018). National trends in drinking water quality violations. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(9), 2078–2083. https://doi.org/10.1073/pnas.1719805115

Altieri, M. A., & Nicholls, C. I. (2020). Agroecology and the reconstruction of a post-COVID-19 agriculture. *Journal of Peasant Studies*, 881–898. https://doi.org/10.1080/03066150.2020.1782891

Anselin, L., Bongiovanni, R., & Lowenberg-DeBoer, J. (2004). A Spatial Econometric Approach To The Economics of Site-Specific Nitrogen Management. *Amer. J. Agr. Econ.*, *86*(August), 675–687.

Antle, J. M., & Capalbo, S. M. (2002). Agriculture as a Managed Ecosystem: Policy Implications. *Journal of Agricultural and Resource Economics*, *27*(1), 1–15.

Antle, J. M., & Valdivia, R. O. (2021). Trade-off analysis of agri-food systems for sustainable research and development. *Q Open*, *1*(1), 1–34. https://doi.org/10.1093/qopen/qoaa005

Anwar, M. R., Liu, D. L., Macadam, I., & Kelly, G. (2013). *Adapting agriculture to climate change : a review*. 225–245. https://doi.org/10.1007/s00704-012-0780-1

Argento, F., Anken, T., Abt, F., Vogelsanger, E., Walter, A., & Liebisch, F. (2021). Site-specific nitrogen management in winter wheat supported by low-altitude remote sensing and soil data. *Precision Agriculture*, *22*(2), 364–386. https://doi.org/10.1007/s11119-020-09733-3

Arnall, D. B., Tubaña, B. S., Holtz, S. L., Girma, K., & Raun, W. R. (2009). Relationship between nitrogen use efficiency and response index in winter wheat. *Journal of Plant Nutrition*, *32*(3), 502–515. https://doi.org/10.1080/01904160802679974

Basso, B., & Antle, J. (2020). Digital agriculture to design sustainable agricultural systems. *Nature Sustainability*, *3*(4), 254–256. https://doi.org/10.1038/s41893-020-0510-0

Basso, B., Ritchie, J. T., Cammarano, D., & Sartori, L. (2011). A strategic and tactical management approach to select optimal N fertilizer rates for wheat in a spatially variable field. *European Journal of Agronomy*, *35*(4), 215–222.

https://doi.org/10.1016/j.eja.2011.06.004

Besag, J., & Higdon, D. (1999). Bayesian analysis of agricultural field experiments. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *61*(4), 691–746. https://doi.org/10.1111/1467-9868.00201

Biermacher, J. T., Brorsen, B. W., Epplin, F. M., Solie, J. B., & Raun, W. R. (2009). The economic potential of precision nitrogen application with wheat based on plant sensing. *Agricultural Economics*, *40*(4), 397–407. https://doi.org/10.1111/j.1574-0862.2009.00387.x

Blackmore, S., & Moore, M. (1999). Remedial Correction of Yield Map Data. *Precision Agriculture*, *1*(1), 53–66. https://doi.org/10.1023/A:1009969601387

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, *9*(3–4), 251–290. https://doi.org/10.1002/hyp.3360090305

Bolton, D. K., & Friedl, M. A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*, *173*, 74–84. https://doi.org/10.1016/j.agrformet.2013.01.007

Booth, M. S., Stark, J. M., & Rastetter, E. (2005). Controls on Nitrogen Cycling in Terrestrial Ecosystems: A Synthetic Analysis of Literature Data. *Ecological Monographs*, *75*(2), 139–157.

Bouwman, A. F., Beusen, A. H. W., Griffioen, J., Van Groenigen, J. W., Hefting, M. M., Oenema, O., Van Puijenbroek, P. J. T. M., Seitzinger, S., Slomp, C. P., & Stehfest, E. (2013). Global trends and uncertainties in terrestrial denitrification and N2O emissions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1621). https://doi.org/10.1098/rstb.2013.0112

Boyer, C. N., Larson, J. A., Roberts, R. K., McClure, A. T., Tyler, D. D., & Zhou, V. (2013). Stochastic Corn Yield Response Functions to Nitrogen for Corn after Corn, Corn after Cotton, and Corn after Soybeans. *Journal of Agricultural and Applied Economics*, *45*(4), 669–681. https://doi.org/10.1017/s1074070800005198

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32.

Bronson, K. F., Booker, J. D., Bordovsky, J. P., Keeling, J. W., Wheeler, T. A., Boman, R. K., Parajulee, M. N., Segarra, E., & Nichols, R. L. (2006). Site-specific irrigation and nitrogen management for cotton production in the Southern High Plains. *Agronomy Journal*, *98*(1), 212–219. https://doi.org/10.2134/agronj2005.0149

Bucci, G., Bentivoglio, D., & Finco, A. (2018). Precision agriculture as a driver for sustainable farming systems: State of art in litterature and research. *Quality - Access to Success*, *19*(S1), 114–121.

Bullock, D. S., Boerngen, M., Tao, H., Maxwell, B., Luck, J. D., Shiratsuchi, L., Puntel, L., & Martin, N. F. (2019). The data-intensive farm management project: Changing agronomic research through on-farm precision experimentation. *Agronomy Journal*, *111*(6), 2736–2746. https://doi.org/10.2134/agronj2019.03.0165

Bullock, D. S., Mieno, T., & Hwang, J. (2020). The value of conducting on-farm field trials using precision agriculture technology: a theory and simulations. *Precision Agriculture*, *21*(5), 1027–1044. https://doi.org/10.1007/s11119-019-09706-1

Bürkner, P.C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80*(1), 1-28. doi:10.18637/jss.v080.i01

Bürkner, P.C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. T*he R Journal*, *10*(1), 395-411. doi:10.32614/RJ-2018-017

Bürkner, P.C. (2021). Bayesian Item Response Modeling in R with brms and Stan. J*ournal of Statistical Software, 100*(5), 1-54. doi:10.18637/jss.v100.i05

Cambouris, A. N., Zebarth, B. J., Ziadi, N., & Perron, I. (2014). Precision Agriculture in Potato Production. *Potato Research*, *57*(3–4), 249–262. https://doi.org/10.1007/s11540-014-9266-0

Canfield, D. E., Glazer, A. N., & Falkowski, P. G. (2010). The evolution and future of earth's nitrogen cycle. *Science*, *330*(6001), 192–196. https://doi.org/10.1126/science.1186120

Cao, P., Lu, C., & Yu, Z. (2018). Historical nitrogen fertilizer use in agricultural ecosystems of the contiguous United States during 1850-2015: Application rate, timing, and fertilizer types. *Earth System Science Data*, *10*(2), 969–984. https://doi.org/10.5194/essd-10-969-2018

Capel, P. D., McCarthy, K. A., & Barbash, J. E. (2008). National, holistic, watershed-scale approach to understand the sources, transport, and fate of agricultural chemicals. *Journal of Environmental Quality*, *37*(3), 983–993. https://doi.org/10.2134/jeq2007.0226

Carlisle, L., Montenegro de Wit, M., DeLonge, M. S., Iles, A., Calo, A., Getz, C., Ory, J., Munden-Dixon, K., Galt, R., Melone, B., Knox, R., & Press, D. (2019). Transitioning to Sustainable Agriculture Requires Growing and Sustaining an Ecologically Skilled Workforce. *Frontiers in Sustainable Food Systems*, *3*(November), 96. https://doi.org/10.3389/fsufs.2019.00096

Carolan, M. (2017). Publicising Food: Big Data, Precision Agriculture, and Co-Experimental Techniques of Addition. *Sociologia Ruralis*, *57*(2), 135–154. https://doi.org/10.1111/soru.12120

Cassman, K. G. (1999). Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(11), 5952–5959. https://doi.org/10.1073/pnas.96.11.5952

Cassman, K. G., & Munns, D. N. (1980). Nitrogen Mineralization as Affected by Soil Moisture, Temperature, and Depth. *Soil Science Society of America Journal*, *44*, 1233–1237.

Chen, K., O'Leary, R. A., & Evans, F. H. (2019). A simple and parsimonious generalised additive model for predicting wheat yield in a decision support tool. *Agricultural Systems*, *173*(December 2018), 140–150. https://doi.org/10.1016/j.agsy.2019.02.009

Church, S. P., Lu, J., Ranjan, P., Reimer, A. P., & Prokopy, L. S. (2020). The role of systems thinking in cover crop adoption: Implications for conservation communication. *Land Use Policy*, *94*(December 2019). https://doi.org/10.1016/j.landusepol.2020.104508

Clancy, M. (2019). Real-time protein monitoring. *Next Instruments*, *June*.

Coble, K. H., Mishra, A. K., Ferrell, S., & Griffin, T. (2018). Big data in agriculture: A challenge for the future. *Applied Economic Perspectives and Policy*, *40*(1), 79–96. https://doi.org/10.1093/aepp/ppx056

Coleman, D. C., & Dec, N. (1989). *Special Feature : Ecology , Agroecosystems , and Sustainable Agriculture*. *70*(6), 12–13.

Cook, S., Cock, J., Oberthür, T., & Fisher, M. (2004). On-Farm Experimentation. *Better Crops*, *97*, 17–20.

Córdova, C., Barrera, J. A., & Magna, C. (2018). Spatial variation of nitrogen mineralization as a guide for variable application of nitrogen fertilizer to cereal crops. *Nutrient Cycling in Agroecosystems*, *110*(1), 83–88. https://doi.org/10.1007/s10705-017-9886-2

Croft, H., Arabian, J., Chen, J. M., Shang, J., & Liu, J. (2020). Mapping within-field leaf chlorophyll content in agricultural crops for nitrogen management using Landsat-8 imagery. *Precision Agriculture*, *21*(4), 856–880. https://doi.org/10.1007/s11119-019-09698-y

Daum, T. (2021). Farm robots: ecological utopia or dystopia? *Trends in Ecology and Evolution*.

De Bastiani, F., Mariz de Aquino Cysneiros, A. H., Uribe-Opazo, M. A., & Galea, M. (2015). Influence diagnostics in elliptical spatial linear models. *Test*, *24*(2), 322–340. https://doi.org/10.1007/s11749-014-0409-z

DeLonge, M., Robbins, T., Basche, A., & Haynes-Mawlow, L. (2020). The State of Sustainable Agriculture and Agroecology Research and Impacts: A Survey of U.S. Scientists. *Journal of Agriculture, Food Systems, and Community Development*, *9*(2), 1–26. https://doi.org/10.5304/jafscd.2020.092.009

DeLonge, M. S., Miles, A., & Carlisle, L. (2016). Investing in the transition to sustainable agriculture. *Environmental Science and Policy*, *55*, 266–273. https://doi.org/10.1016/j.envsci.2015.09.013

Diacono, M., Rubino, P., & Montemurro, F. (2013). Precision nitrogen management of wheat. A review. *Agronomy for Sustainable Development*, *33*(1), 219–241. https://doi.org/10.1007/s13593-012-0111-z

Dietterich, T. G. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, *10*(7), 1895–1923. https://doi.org/10.1162/089976698300017197

Diggle, P. J., & Tawn, J. A. (1998). Model-based geostatistics. *Applied Statistics*, *47*(3), 299–350.

Drachal, K. (2019). Analysis of agricultural commodities prices with new Bayesian model combination schemes. *Sustainability (Switzerland)*, *11*(19). https://doi.org/10.3390/su11195305

Duff, H., Hegedus, P. B., Loewen, S., Bass, T., & Maxwell, B. D. (2022). Precision Agroecology. *Sustainability*, *14*(106), 1–18. https://doi.org/https://doi.org/10.3390/su14010106

Egli, L., Meyer, C., Scherber, C., Kreft, H., & Tscharntke, T. (2018). Winners and losers of national and global efforts to reconcile agricultural intensification and biodiversity conservation. *Global Change Biology*, *24*(5), 2212–2228. https://doi.org/10.1111/gcb.14076

Elliot, E. T., & Cole, V. (1989). A Perspective on Agroecosystem Science. *Ecology*, *70*(6), 1597–1602.

Engel, R. (2012). Volatilization losses from surface-applied urea during cold weather months. *Montana State University - Dept. LRES*, 1–10. https://doi.org/10.1017/CBO9781107415324.004

Estes, L. D., Bradley, B. A., Beukes, H., Hole, D. G., Lau, M., Oppenheimer, M. G., Schulze, R., Tadross, M. A., & Turner, W. R. (2013). Comparing mechanistic and empirical model projections of crop suitability and productivity: Implications for ecological forecasting. *Global Ecology and Biogeography*, *22*(8), 1007–1018. https://doi.org/10.1111/geb.12034

Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development*, *36*(2). https://doi.org/10.1007/s13593-016-0364-z

Ewing, P. M., & Runck, B. C. (2015). Optimizing nitrogen rates in the midwestern United States for maximum ecosystem value. *Ecology and Society*, *20*(1). https://doi.org/10.5751/ES-06767-200118

Farid, H. U., Bakhsh, A., Ahmad, N., Ahmad, A., & Mahmood-Khan, Z. (2016). Delineating site-specific management zones for precision agriculture. *Journal of Agricultural Science*, *154*(2), 273–286. https://doi.org/10.1017/S0021859615000143

Flowers, M., Weisz, R., Heiniger, R., Osmond, D., & Crozier, C. (2004). In-Season Optimization and Site-Specific Nitrogen Management for Soft Red Winter Wheat. *Agronomy Journal*, *96*(1), 124–134. https://doi.org/10.2134/agronj2004.0124

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., … Zaks, D. P. M. (2011). Solutions for a cultivated planet. *Nature*, *478*(7369), 337–342. https://doi.org/10.1038/nature10452

Fowler, D., Coyle, M., Skiba, U., Sutton, M. A., Cape, J. N., Sheppard, L. J., Jenkins, A., Grizzetti, B., Galloway, J. N., Leach, A., Bouwman, A. F., Butterbach-bahl, K., Stevenson, D., Amann, M., Voss, M., Fowler, D., Coyle, M., Skiba, U., Reis, S., … Voss, M. (2013). *The global nitrogen cycle in the twenty-first century*.

Fritsch, M., Lischke, H., & Meyer, K. M. (2020). Scaling methods in ecological modelling. *Methods in Ecology and Evolution*, *11*(11), 1368–1378. https://doi.org/10.1111/2041-210X.13466

Gardner, G., Mieno, T., & Bullock, D. S. (2021). An economic evaluation of site-specific input application Rx maps: evaluation framework and case study. *Precision Agriculture*, *22*(4), 1304–1316. https://doi.org/10.1007/s11119-021-09785-z

Gaso, D. V., Berger, A. G., & Ciganda, V. S. (2019). Predicting wheat grain yield and spatial variability at field scale using a simple regression or a crop model in conjunction with Landsat images. *Computers and Electronics in Agriculture*, *159*(February 2018), 75–83. https://doi.org/10.1016/j.compag.2019.02.026

Gebbers, R., & Adamchuck, V. I. (2010). Precision Agriculture and Food Security. *Science*, *February*, 828–830.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534. https://doi.org/10.1214/06-BA117A

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., & Kalogirou, S. (2021). Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, *36*(2), 121–136. https://doi.org/10.1080/10106049.2019.1595177

Gibert, K., Horsburgh, J. S., Athanasiadis, I. N., & Holmes, G. (2018). Environmental Data Science. *Environmental Modelling and Software*, *106*, 4–12. https://doi.org/10.1016/j.envsoft.2018.04.005

Gliessman, S. (2016). Transforming food systems with agroecology. *Agroecology and Sustainable Food Systems*, *40*(3), 187–189. https://doi.org/10.1080/21683565.2015.1130765

Gliessman, S. R., & Engles, E. W. (2014). *Agroecology: The Ecology of Sustainable Food Systems* (3rd ed.).

Gonzalez-Sanchez, A., Frausto-Solis, J., & Ojeda-Bustamante, W. (2014). Predictive ability of machine learning methods for massive crop yield prediction. *Spanish Journal of Agricultural Research*, *12*(2), 313–328. https://doi.org/10.5424/sjar/2014122-4439

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gotway, C. A., & Stroup, W. W. (1997). A Generalized Linear Model Approach to Spatial Data Analysis and Prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, *2*(2), 157–178.

Grant, C. A., Moulin, A. P., & Tremblay, N. (2016). Nitrogen management effects on spring wheat yield and protein concentration vary with seeding date and slope position. *Agronomy Journal*, *108*(3), 1246–1256. https://doi.org/10.2134/agronj2015.0510

Gross, M. (2013). The paradoxical evolution of agriculture. *CURBIO*, *23*(16), R667–R670. https://doi.org/10.1016/j.cub.2013.08.001

Grzebisz, W. (2021). Site-Specific Nutrient Management. *Agronomy*, *11*(752), 1–6. https://doi.org/https://doi.org/ 10.3390/agronomy11040752

Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling*, *157*(2–3), 89–100. https://doi.org/10.1016/S0304-3800(02)00204-1

Guttieri, M. J., Frels, K., Regassa, T., Waters, B. M., & Baenziger, P. S. (2017). Variation for nitrogen use efficiency traits in current and historical great plains hard winter wheat. *Euphytica*, *213*(4), 1–18. https://doi.org/10.1007/s10681-017-1869-5

Haas, C. A., Frimpong, E. A., & Karpanty, S. M. (2009). Ecosystems and Ecosystem-Based Management. In K. M. Moore (Ed.), *The Sciences and Art of Adaptive Management: Innovating for Sustainable Agriculture and Natural Resources Management* (pp. 106–142). Soil and Water Conservation Society.

Hair, J. F. (2007). Knowledge creation in marketing: The role of predictive analytics. *European Business Review*, *19*(4), 303–315. https://doi.org/10.1108/09555340710760134

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, *1*(3), 297–310. https://doi.org/10.1214/ss/1177013604

Hatfield, J. L., Cryder, M., & Basso, B. (2020). Remote Sensing: Advancing the Science and the Applications to Transform Agriculture. *IT Professional*, *22*(3), 42–45. https://doi.org/10.1109/MITP.2020.2986102

Hatfield, J. L., & Stewart, B. A. (1992). Limitations to Plant Root Growth. In J. L. Hatfield & B. A. Stewart (Eds.), *Advances in Soil Science* (19th ed.). Springer-Verlag. https://doi.org/10.1007/978-1-4612-2894-3 e-ISBN:

Hecht, M., & Crowley, K. (2020). Unpacking the Learning Ecosystems Framework: Lessons from the Adaptive Management of Biological Ecosystems. *Journal of the Learning Sciences*, *29*(2), 264–284. https://doi.org/10.1080/10508406.2019.1693381

Hegedus, P.B. (2020). OFPE: On-Farm Precision Experiments (OFPE) Data Management and Analysis Tools. R package version 1.8.11. https://github.com/paulhegedus/OFPE.git

Hegedus, P.B & Ewing, S.E. (2022). Evaluation of nitrogen use efficiency in dryland winter-wheat agroecosystems for predictive modeling. Manuscript in preparation.

Hegedus, P.B., Maxwell, B.D., Loewen, S., Sheppard J.W., Morales G., & Peerlinck, A. (2022). On-Farm Precision Experiments (OFPE) framework: tapping local data to optimize field scale decisions. Manuscript in preparation.

Hegedus, P.B & Maxwell, B.D. (2022a). Rationale for field specific on-farm precision experimentation. Agriculture, Ecosystems & Environment. Manuscript submitted for publication, in review.

Hegedus, P.B. & Maxwell, B.D. (2022b). Assessing performance of empirical models for forecasting crop responses to variable fertilizer rates using on-farm precision experimentation. Precision Agriculture. Manuscript submitted for publication, in review.

Hegedus, P.B. & Maxwell, B.D. (2022c). Best use of modern data for field-specific decision support. Precision Agriculture. Manuscript submitted for publication, in review.

Holland, D. M., De Oliveira, V., Cox, L. H., & Smith, R. L. (2000). Estimation of regional trends in sulfur dioxide over the eastern United States. *Environmetrics*, *11*(4), 373–393.

https://doi.org/10.1002/1099-095X(200007/08)11:4<373::AID-ENV419>3.0.CO;2-2

Holling, C. S. (1996). Surprise for science, resilience for ecosystems, and incentives for people. *Ecological Applications*, *6*(3), 733–735. https://doi.org/10.2307/2269475

Hooten, M. B., Ver Hoef, J. M., & Hanks, E. M. (2019). Simultaneous Autoregressive Model (SAR). *Wiley StatsRef: Statistics Reference Online*, 1–10. https://doi.org/10.1007/978-0-387-35973-1_1217

Houlahan, J. E., McKinney, S. T., Anderson, T. M., & McGill, B. J. (2017). The priority of prediction in ecological understanding. *Oikos*, *126*(1), 1–7. https://doi.org/10.1111/oik.03726

Huang, X., Huang, G., Yu, C., Ni, S., & Yu, L. (2017). A multiple crop model ensemble for improving broad-scale yield prediction using Bayesian model averaging. *Field Crops Research*, *211*(June), 114–124. https://doi.org/10.1016/j.fcr.2017.06.011

Huggins, D. R. (2010). Site-Specific N Management for Direct-Seed Cropping Systems. *CSANR Research Report 2010 – 001*, Chp. 16. http://csanr.wsu.edu/wp-content/uploads/2013/02/CSANR2010-001.Ch16.pdf

Hurley, T. M., Malzer, G. L., & Kilian, B. (2004). Estimating site-specific nitrogen crop response functions: A conceptual framework and geostatistical model. *Agronomy Journal*, *96*(5), 1331–1343. https://doi.org/10.2134/agronj2004.1331

Hurley, T. M., Oishi, K., & Malzer, G. L. (2005). Estimating the potential value of variable rate nitrogen applications: A comparison of spatial econometric and geostatistical models. *Journal of Agricultural and Resource Economics*, *30*(2), 231–249. https://doi.org/10.2307/40988069

Jackson, L., van Noordwijk, M., Bengtsson, J., Foster, W., Lipper, L., Pulleman, M., Said, M., Snaddon, J., & Vodouhe, R. (2010). Biodiversity and agricultural sustainagility: From assessment to adaptive management. *Current Opinion in Environmental Sustainability*, *2*(1–2), 80–87. https://doi.org/10.1016/j.cosust.2010.02.007

Janatian, N., Sadeghi, M., Sanaeinejad, S. H., Bakhshian, E., Farid, A., Hasheminia, S. M., & Ghazanfari, S. (2017). A statistical framework for estimating air temperature using MODIS land surface temperature data. *International Journal of Climatology*, *37*(3), 1181–1194. https://doi.org/10.1002/joc.4766

Jaynes, D. B., & Colvin, T. S. (2006). Corn yield and nitrate loss in subsurface drainage from midseason nitrogen fertilizer application. *Agronomy Journal*, *98*(6), 1479–1487. https://doi.org/10.2134/agronj2006.0046

Jenks, George F. 1967. "The Data Model Concept in Statistical Mapping", International Yearbook of Cartography 7: 186-190.

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K. M., Gerber, J. S., Reddy, V. R., & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *PLoS ONE*, *11*(6), 1–15. https://doi.org/10.1371/journal.pone.0156571

Jiang, P., He, Z., Kitchen, N. R., & Sudduth, K. A. (2009). Bayesian analysis of within-field variability of corn yield using a spatial hierarchical model. *Precision Agriculture*, *10*(2), 111–127. https://doi.org/10.1007/s11119-008-9070-4

Jing, W., Yang, Y., Yue, X., & Zhao, X. (2016). A comparison of different regression algorithms for downscaling monthly satellite-based precipitation over North China. *Remote Sensing*, *8*(10), 1–17. https://doi.org/10.3390/rs8100835

John, A. A., Jones, C. A., Ewing, S. A., Sigler, W. A., Bekkerman, A., & Miller, P. R. (2017). Fallow replacement and alternative nitrogen management for reducing nitrate leaching in a semiarid region. *Nutrient Cycling in Agroecosystems*, *108*(3), 279–296. https://doi.org/10.1007/s10705-017-9855-9

Johnson, D. M. (2014). An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment*, *141*, 116–128. https://doi.org/10.1016/j.rse.2013.10.027

Jones, C. (2013). *Mid- to late-season nitrogen fertilizer: a viable option to increase wheat yield and grain protein*. MSU News Service. https://www.montana.edu/news/11934/msuextension-mid-to-late-season-nitrogen-fertilizing-can-boost-yield-protein-in-wheat-crops

Jones, C. (2018). *Soil Acidification: A growing concern for Montana farmers*. United States Department of Agriculture, Natural Resource Conservation Service.

Jones, C., & Olson-Rutz, K. (2020). Soil Acidification: Management. *The Soil Scoop*, *198*(August), 1–2. https://doi.org/10.1016/j.still.2019.104534

Jones, C., Olson-Rutz, K., Miller, P., Zabinski, C., & Tallman, S. (2022). *Cover Crops : Management for Organic Matter and Nitrogen Cover Crops : Management for Organic Matter and Nitrogen*. https://apps.msuextension.org/publications/pub.html?sku=EB0237

Jones, J., Verma, B., Basso, B., Mohtar, R., & Matlock, M. (2021). Transforming food and agriculture to circular systems: A perspective for 2050. *Resource: Engineering and Technology for Sustainable World*, *28*(2), 7–9.

Joshi, V. R., Kazula, M. J., Coulter, J. A., Naeve, S. L., & Garcia y Garcia, A. (2021). In-season weather data provide reliable yield estimates of maize and soybean in the US central Corn Belt. *International Journal of Biometeorology*, *65*(4), 489–502. https://doi.org/10.1007/s00484-020-02039-z

Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, *143*(September), 23–37. https://doi.org/10.1016/j.compag.2017.09.037

Kanter, D. R., Musumba, M., Wood, S. L. R., Palm, C., Antle, J., Balvanera, P., Dale, V. H., Havlik, P., Kline, K. L., Scholes, R. J., Thornton, P., Tittonell, P., & Andelman, S. (2018). Evaluating agricultural trade-offs in the age of sustainable development. *Agricultural Systems*, *163*, 73–88. https://doi.org/10.1016/j.agsy.2016.09.010

Kätterer, T., Eckersten, H., Andrén, O., & Pettersson, R. (1997). Winter wheat biomass and nitrogen dynamics under different fertilization and water regimes: Application of a crop growth model. *Ecological Modelling*, *102*(2–3), 301–314. https://doi.org/10.1016/S0304-3800(97)00065-3

Khosla, R., Fleming, K., Delgado, J. A., Shaver, T. M., & Westfall, D. G. (2002). Use of site-specific management zones to improve nitrogen management for precision agriculture. *Journal of Soil and Water Conservation*, *57*(6), 513–518.

Khosla, R., Inman, D., Westfall, D. G., Reich, R. M., Frasier, M., Mzuku, M., Koch, B., & Hornung, A. (2008). A synthesis of multi-disciplinary research in precision agriculture: Site-specific management zones in the semi-arid western Great Plains of the USA. *Precision Agriculture*, *9*(1–2), 85–100. https://doi.org/10.1007/s11119-008-9057-1

King, Catherine L. Mackie, C. (2019). Improving Data Collection and Measurement of Complex Farms. In *Improving Data Collection and Measurement of Complex Farms*. National Academies Press. https://doi.org/10.17226/25260

Kissling, W. D., & Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, *17*(1), 59–71. https://doi.org/10.1111/j.1466-8238.2007.00334.x

Koch, B., Khosla, R., Frasier, W. M., Westfall, D. G., & Inman, D. (2004). Economic feasibility of variable-rate nitrogen application utilizing site-specific management zones. *Agronomy Journal*, *96*(6), 1572–1580. https://doi.org/10.2134/agronj2004.1572

Kyveryga, P. M. (2019). On-farm research: Experimental approaches, analytical frameworks, case studies, and impact. *Agronomy Journal*, *111*(6), 2633–2635. https://doi.org/10.2134/agronj2019.11.0001

Lacoste, M., Cook, S., Mcnee, M., Gale, D., Ingram, J., Bellon-maurel, V., Macmillan, T., Sylvester-bradley, R., Kindred, D., Bramley, R., Tremblay, N., Longchamps, L., Thompson, L., Ruiz, J., Oscar Garcia, F., Maxwell, B., Griffin, T., Oberthür, T., Huyghe, C., … Hall, A. (2022). On-Farm Experimentation to transform global agriculture. *Nature Food*. https://doi.org/https://doi.org/10.1038/s43016-021-00424-4 On-Farm

Lambers, H., & Oliveira, R. S. (2019). Plant physiological ecology. In *Plant Physiological Ecology*. 2`. https://doi.org/10.1007/978-3-030-29639-1

Langella, G., Basile, A., Bonfante, A., & Terribile, F. (2010). High-resolution space-time rainfall analysis using integrated ANN inference systems. *Journal of Hydrology*, *387*(3–4), 328–

342. https://doi.org/10.1016/j.jhydrol.2010.04.027

Lau, K. W., & Wu, Q. H. (2008). Local prediction of non-linear time series using support vector regression. *Pattern Recognition*, *41*(5), 1539–1547. https://doi.org/10.1016/j.patcog.2007.08.013

Lawes, R. A., Oliver, Y. M., & Huth, N. I. (2019). Optimal nitrogen rate can be predicted using average yield and estimates of soil water and leaf nitrogen with infield experimentation. *Agronomy Journal*, *111*(3), 1155–1164. https://doi.org/10.2134/agronj2018.09.0607

Lawrence, P. G., Rew, L. J., & Maxwell, B. D. (2015a). A probabilistic Bayesian framework for progressively updating site-specific recommendations. *Precision Agriculture*, *16*(3), 275–296. https://doi.org/10.1007/s11119-014-9375-4

Lawrence, P. G., Rew, L. J., & Maxwell, B. D. (2015b). A probabilistic Bayesian framework for progressively updating site-specific recommendations. *Precision Agriculture*, *16*(3), 275–296. https://doi.org/10.1007/s11119-014-9375-4

Li, J., Heap, A. D., Potter, A., & Daniell, J. J. (2011). Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling and Software*, *26*(12), 1647–1659. https://doi.org/10.1016/j.envsoft.2011.07.004

Liew, J., & Forkman, J. (2015). A guide to generalized additive models in crop science using SAS and R. *Communications in Biometry and Crop Science*, *10*(1), 41–57.

Lin, B. B. (2011). Resilience in agriculture through crop diversification: Adaptive management for environmental change. *BioScience*, *61*(3), 183–193. https://doi.org/10.1525/bio.2011.61.3.4

Link, J., Batchelor, W. D., Graeff, S., & Claupein, W. (2008). Evaluation of current and model-based site-specific nitrogen applications on wheat (Triticum aestivum L.) yield and environmental quality. *Precision Agriculture*, *9*(5), 251–267. https://doi.org/10.1007/s11119-008-9068-y

Lokers, R., Knapen, R., Janssen, S., van Randen, Y., & Jansen, J. (2016). Analysis of Big Data technologies for use in agro-environmental science. *Environmental Modelling and Software*, *84*, 494–504. https://doi.org/10.1016/j.envsoft.2016.07.017

Luschei, E. C., Van Wychen, L. R., Maxwell, B. D., Bussan, A. J., Buschena, D., & Goodman, D. (2001). Implementing and conducting on-farm weed research with the use of GPS. *Weed Science*, *49*(4), 536–542. https://doi.org/10.1614/0043-1745(2001)049[0536:iacofw]2.0.co;2

Lynch, J., Cain, M., Frame, D., & Pierrehumbert, R. (2021). Agriculture's Contribution to Climate Change and Role in Mitigation Is Distinct From Predominantly Fossil CO2-Emitting Sectors. *Frontiers in Sustainable Food Systems*, *4*(February), 1–9. https://doi.org/10.3389/fsufs.2020.518039

MacMillan, T., & Benton, T. (2014). Engage farmers in research. *Nature, 509*, 25–27.

Macnack, N., Khim, B. C., Mullock, J., & Raun, W. (2014). In-Season Prediction of Nitrogen Use Efficiency and Grain Protein in Winter Wheat (Triticum aestivum L.). *Communications in Soil Science and Plant Analysis*, *45*(18), 2480–2494. https://doi.org/10.1080/00103624.2014.904337

Mardia, A. K. V, & Marshall, R. J. (1984). Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression. *Biometrika*, *71*(1), 135–146.

Mariano, C., & Mónica, B. (2021). A random forest-based algorithm for data-intensive spatial interpolation in crop yield mapping. *Computers and Electronics in Agriculture*, *184*(March). https://doi.org/10.1016/j.compag.2021.106094

Marques Ramos, A. P., Prado Osco, L., Elis Garcia Furuya, D., Nunes Gonçalves, W., Cordeiro Santana, D., Pereira Ribeiro Teodoro, L., Antonio da Silva Junior, C., Fernando Capristo-Silva, G., Li, J., Henrique Rojo Baio, F., Marcato Junior, J., Eduardo Teodoro, P., & Pistori, H. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, *178*(July). https://doi.org/10.1016/j.compag.2020.105791

Maxwell, B. D., & Luschei, E. C. (2005). Justification for site-specific weed management based on ecology and economics. *Weed Science*, *53*(2), 221–227. https://doi.org/10.1614/ws-04-

071r2

McBratney, A. B., Minasny, B., & Whelan, B. M. (2005). Obtaining Useful High-resolution Soil Data from Proximally - Sensed Electrical Conductivity / Resistivity (PSEC / R) Surveys. *Precision Agriculture 2005, ECPA 2005*.

McBratney, A., Whelan, B., Ancev, T., & Bouma, J. (2005). Future directions of precision agriculture. *Precision Agriculture*, *6*(1), 7–23. https://doi.org/10.1007/s11119-005-0681-8

McFadden, B. R., Brorsen, B. W., & Raun, W. R. (2017). Nitrogen fertilizer recommendations based on plant sensing and Bayesian updating. *Precision Agriculture*, *19*(1), 1–14. https://doi.org/10.1007/s11119-017-9499-4

McFadden, J. R., Rosburg, A., & Njuki, E. (2021). Information Inputs and Technical Efficiency in Midwest Corn Production: Evidence from Farmers' Use of Yield and Soil Maps. *American Journal of Agricultural Economics*, 1–24. https://doi.org/10.1111/ajae.12251

Meola, A. (2021). Smart Farming in 2020 : How IoT sensors are creating a more efficient precision agriculture industry. *Business Insider*.

Meyer-Aurich, A., Weersink, A., Gandorfer, M., & Wagner, P. (2010). Optimal site-specific fertilization and harvesting strategies with respect to crop yield and quality response to nitrogen. *Agricultural Systems*, *103*(7), 478–485. https://doi.org/10.1016/j.agsy.2010.05.001

Michaletz, S. T., Weiser, M. D., McDowell, N. G., Zhou, J., Kaspari, M., Helliker, B. R., & Enquist, B. J. (2016). The energetic and carbon economic origins of leaf thermoregulation. *Nature Plants*, *2*(9). https://doi.org/10.1038/nplants.2016.129

Miles, A., DeLonge, M. S., & Carlisle, L. (2017). Triggering a positive research and policy feedback cycle to support a transition to agroecology and sustainable food systems. *Agroecology and Sustainable Food Systems*, *41*(7), 855–879. https://doi.org/10.1080/21683565.2017.1331179

Montana Climate Office, University of Montana. 2020. Montana Mesonet. The Montana Forest and Conservation Experiment Station. http://climate.umt.edu/mesonet/

Moore, K. M. (2009). Landscape systems framework for adaptive management. In K. M. Moore (Ed.), *The sciences and art of adaptive management: innovating for sustainable agriculture and natural resource management* (pp. 3–22). Soil and Water Conservation Society.

Moshia, M. E., Khosla, R., Longchamps, L., Reich, R., Davis, J. G., & Westfall, D. G. (2014). Precision manure management across site-specific management zones: Grain yield and economic analysis. *Agronomy Journal*, *106*(6), 2146–2156. https://doi.org/10.2134/agronj13.0400

Mueller, J. P., Finney, D., & Hepperly, P. (2009). The Field System. In K. M. Moore (Ed.), *The Sciences and Art of Adaptive Management: Innovating for Sustainable Agriculture and Natural Resources Management* (pp. 25–59). Soil and Water Conservation Society.

Mueller, N. D., West, P. C., Gerber, J. S., Macdonald, G. K., Polasky, S., & Foley, J. A. (2014). A tradeoff frontier for global nitrogen use and cereal production. *Environmental Research Letters*, *9*(5). https://doi.org/10.1088/1748-9326/9/5/054002

Nandram, B., Berg, E., & Barboza, W. (2014). A hierarchical Bayesian model for forecasting state-level corn yield. *Environmental and Ecological Statistics*, *21*(3), 507–530. https://doi.org/10.1007/s10651-013-0266-z

Oelofse, M., Markussen, B., Knudsen, L., Schelde, K., Olesen, J. E., Jensen, L. S., & Bruun, S. (2015). Do soil organic carbon levels affect potential yields and nitrogen use efficiency? An analysis of winter wheat and spring barley field trials. *European Journal of Agronomy*, *66*, 62–73. https://doi.org/10.1016/j.eja.2015.02.009

Oreskes. (2000). *Why believe a Computer, Models, Measures and Meaning in a Natural World.pdf*.

Paccioretti, P., Bruno, C., Gianinni Kurina, F., Córdoba, M., Bullock, D. S., & Balzarini, M. (2021). Statistical models of yield in on-farm precision experimentation. *Agronomy Journal*, *113*(6), 4916–4929. https://doi.org/10.1002/agj2.20833

Paul, E. A., & Robertson, G. P. (1989). Ecology and the Agricultural Sciences: A False

Dichotomy? *Ecology*, *70*(6), 1594–1597. https://doi.org/10.2307/1938091

Pavuluri, K., Chim, B. K., Griffey, C. A., Reiter, M. S., Balota, M., & Thomason, W. E. (2015). Canopy spectral reflectance can predict grain nitrogen use efficiency in soft red winter wheat. *Precision Agriculture*, *16*(4), 405–424. https://doi.org/10.1007/s11119-014-9385-2

Pearcy, R. W., Muraoka, H., & Valladares, F. (2005). Crown architecture in sun and shade environments: Assessing function and trade-offs with a three-dimensional simulation model. *New Phytologist*, *166*(3), 791–800. https://doi.org/10.1111/j.1469-8137.2005.01328.x

Peerlinck, A., Sheppard, J., & Maxwell, B. (2018). Using Deep Learning in Yield and Protein Prediction of Winter Wheat Based on Fertilization Prescriptions in Precision Agriculture. *14th International Conference on Precision Agriculture*, 1–13.

Peerlinck, A., Sheppard, J., & Senecal, J. (2019). AdaBoost with Neural Networks for Yield and Protein Prediction in Precision Agriculture. *Proceedings of the International Joint Conference on Neural Networks*, *2019-July*. https://doi.org/10.1109/IJCNN.2019.8851976

Peters, D. P. C., Rivers, A., Hatfield, J. L., Lemay, D. G., Liu, S., & Basso, B. (2020). Harnessing AI to Transform Agriculture and Inform Agricultural Research. *IT Professional*, *22*(3), 16–21. https://doi.org/10.1109/MITP.2020.2986124

Pham, X., & Stack, M. (2018). How data analytics is transforming agriculture. *Business Horizons*, *61*(1), 125–133. https://doi.org/10.1016/j.bushor.2017.09.011

Pierpaoli, E., Carli, G., Pignatti, E., & Canavari, M. (2013). Drivers of Precision Agriculture Technologies Adoption: A Literature Review. *Procedia Technology*, *8*(Haicta), 61–69. https://doi.org/10.1016/j.protcy.2013.11.010

Pilkey-Jarvis, L., & Pilkey, O. H. (2008). Useless arithmetic: Ten points to ponder when using mathematical models in environmental decision making. *Public Administration Review*, *68*(3), 470–479. https://doi.org/10.1111/j.1540-6210.2008.00883_2.x

Ping, J. L., & Dobermann, A. (2003). Creating Spatially Contiguous Yield Classes for Site-

Specific Management. *Agronomy Journal*, *95*(5), 1121–1131. https://doi.org/10.2134/agronj2003.1121

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2021). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-153, https://www.CRAN.R-project.org/package=nlme

Pinilla, J., & Negrín, M. (2021). Non-parametric generalized additive models as a tool for evaluating policy interventions. *Mathematics*, *9*(4), 1–12. https://doi.org/10.3390/math9040299

Pretty, J. (2008). Agricultural sustainability: Concepts, principles and evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1491), 447–465. https://doi.org/10.1098/rstb.2007.2163

Prey, L., Hu, Y., & Schmidhalter, U. (2019). Temporal Dynamics and the Contribution of Plant Organs in a Phenotypically Diverse Population of High-Yielding Winter Wheat: Evaluating Concepts for Disentangling Yield Formation and Nitrogen Use Efficiency. *Frontiers in Plant Science*, *10*(October), 1–14. https://doi.org/10.3389/fpls.2019.01295

Prost, L., Makowski, D., & Jeuffroy, M. H. (2008). Comparison of stepwise selection and Bayesian model averaging for yield gap analysis. *Ecological Modelling*, *219*(1–2), 66–76. https://doi.org/10.1016/j.ecolmodel.2008.07.026

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, *1*(1), 51–59. https://doi.org/10.1089/big.2013.1508

Qianqian, D., Mieno, T., Bullock, D. S., & Edge, B. (2021). Economically Optimal Nitrogen Side-dressing Based on Vegetation Indices from Satellite Images Through On-farm Experiments. In Harper Adams University (Ed.), *Proceedings of the 4th Symposium on Agri-Tech Economics for Sustainable Futures* (Issue September, pp. 51–62). HAU Publications.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramsey, A. F. (2020). Probability Distributions of Crop Yields: A Bayesian Spatial Quantile Regression Approach. *American Journal of Agricultural Economics*, *102*(1), 220–239. https://doi.org/10.1093/ajae/aaz029

Ravier, C., Meynard, J., Cohan, J., Gate, P., & Jeu, M. (2017). *Early nitrogen de fi ciencies favor high yield , grain protein content and N use e ffi ciency in wheat*. *89*(November 2016), 16–24. https://doi.org/10.1016/j.eja.2017.06.002

Reich, P. B. (2014). The world-wide "fast-slow" plant economics spectrum: A traits manifesto. *Journal of Ecology*, *102*(2), 275–301. https://doi.org/10.1111/1365-2745.12211

Reynolds, W. D., Drury, C. F., Phillips, L. A., Yang, X., & Agomoh, I. V. (2021). An adapted weibull function for agricultural applications. *Canadian Journal of Soil Science*, *101*(4), 680–702. https://doi.org/10.1139/cjss-2021-0046

Rütting, T., Aronsson, H., & Delin, S. (2018). Efficient use of nitrogen in agriculture. *Nutrient Cycling in Agroecosystems*, *110*(1), 1–5. https://doi.org/10.1007/s10705-017-9900-8

Rzewnicki, P. E., Thompson, R., Lesoing, G. W., Elmore, R. W., Francis, C. A., Parkhurst, A. M., & Moomaw, R. S. (1988). On-farm experiment designs and implications for locating research sites. *American Journal of Alternative Agriculture*, *3*(4), 168–173. https://doi.org/10.1017/S0889189300002459

Sawers, P. (2021, May 5). Linux Foundation Launches Open Source Agriculture Infrastructure Project. VentureBeat. Retrieved March 27, 2022, from https://venturebeat.com/2021/05/05/linux-foundation-launches-open-source-agriculture-infrastructure-project/

Schimmelpfennig, D., & Lowenberg-DeBoer, J. (2020). Farm Types and Precision Agriculture Adoption: Crops, Regions, Soil Variability, and Farm Size. *SSRN Electronic Journal*, 1–38. https://doi.org/10.2139/ssrn.3689311

Sebilo, M., Mayer, B., Nicolardot, B., Pinay, G., & Mariotti, A. (2013a). Long-term fate of nitrate fertilizer in agricultural soils. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18185–18189.

https://doi.org/10.1073/pnas.1305372110

Sebilo, M., Mayer, B., Nicolardot, B., Pinay, G., & Mariotti, A. (2013b). Long-term fate of nitrate fertilizer in agricultural soils. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18185–18189. https://doi.org/10.1073/pnas.1305372110

Sela, S., van Es, H. M., Moebius-Clune, B. N., Marjerison, R., Melkonian, J., Moebius-Clune, D., Schindelbeck, R., & Gomes, S. (2016). Adapt-N outperforms grower-selected nitrogen rates in northeast and midwestern united states strip trials. *Agronomy Journal*, *108*(4), 1726–1734. https://doi.org/10.2134/agronj2015.0606

Semenov, M. A., Jamieson, P. D., & Martre, P. (2007). Deconvoluting nitrogen use efficiency in wheat: A simulation study. *European Journal of Agronomy*, *26*(3), 283–294. https://doi.org/10.1016/j.eja.2006.10.009

Shi, Y., Song, L., Xia, Z., Lin, Y., Myneni, R. B., Choi, S., Wang, L., Ni, X., Lao, C., & Yang, F. (2015). Mapping annual precipitation across Mainland China in the period 2001-2010 from TRMM3B43 product using spatial downscaling approach. *Remote Sensing*, *7*(5), 5849–5878. https://doi.org/10.3390/rs70505849

Sigler, W. A., Ewing, S. A., Jones, C. A., Payn, R. A., Brookshire, E. N. J., Klassen, J. K., Jackson-Smith, D., & Weissmann, G. S. (2018). Connections among soil, ground, and surface water chemistries characterize nitrogen loss from an agricultural landscape in the upper Missouri River Basin. *Journal of Hydrology*, *556*, 247–261. https://doi.org/10.1016/j.jhydrol.2017.10.018

Sigler, W. A., Ewing, S. A., Jones, C. A., Payn, R. A., Miller, P., & Maneta, M. (2020). Water and nitrate loss from dryland agricultural soils is controlled by management, soils, and weather. *Agriculture, Ecosystems and Environment*, *304*(July), 107158. https://doi.org/10.1016/j.agee.2020.107158

Sigler, W. A., Ewing, S. A., Wankel, S. D., Jones, C. A., Leuthold, S., Brookshire, E. N. J., & Payn, R. A. (2022). Isotopic signals in an agricultural watershed suggest denitrification is locally intensive in riparian areas but extensive in upland soils. *Biogeochemistry*, *158*(2), 251–268. https://doi.org/10.1007/s10533-022-00898-9

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Soil Survey Staff, Natural Resources Conservation Service (NRCS), United States Department of Agriculture (USDA). 2020. Web Soil Survey. Available online at the following link: http://websoilsurvey.sc.egov.usda.gov/

Stepanov, A., Dubrovin, K., Sorokin, A., & Aseeva, T. (2020). Predicting soybean yield at the regional scale using remote sensing and climatic data. *Remote Sensing*, *12*(12). https://doi.org/10.3390/rs12121936

Struik, P. C. (2017). *Sustainable intensification in agriculture : the richer shade of green . A review*. https://doi.org/10.1007/s13593-017-0445-7

Sudduth, K. A., & Drummond, S. T. (2007). Yield editor: Software for removing errors from crop yield maps. *Agronomy Journal*, *99*(6), 1471–1482. https://doi.org/10.2134/agronj2006.0326

Sudduth, K. A., Drummond, S. T., & Myers, D. B. (2012). Yield editor 2.0: Software for automated removal of yield map errors. *American Society of Agricultural and Biological Engineers Annual International Meeting 2012, ASABE 2012*, *4*(12), 3378–3391. https://doi.org/10.13031/2013.41893

Sykuta, M. E. (2016). Big Data in Agriculture: Property Rights, Privacy and Competition in Ag Data Services. *International Food and Agribusiness Management Review*, *19*, 57–74.

Talukder, B., Blay-Palmer, A., vanLoon, G. W., & Hipel, K. W. (2020). Towards complexity of agricultural sustainability assessment: Main issues and concerns. *Environmental and Sustainability Indicators*, *6*(March), 100038. https://doi.org/10.1016/j.indic.2020.100038

Thilakarathna, S. K., Hernandez-Ramirez, G., Puurveen, D., Kryzanowski, L., Lohstraeter, G., Powers, L. A., Quan, N., & Tenuta, M. (2020). Nitrous oxide emissions and nitrogen use efficiency in wheat: Nitrogen fertilization timing and formulation, soil nitrogen, and weather effects. *Soil Science Society of America Journal*, *84*(6), 1910–1927. https://doi.org/10.1002/saj2.20145

Thöle, H., Richter, C., & Ehlert, D. (2013). Strategy of statistical model selection for precision farming on-farm experiments. *Precision Agriculture*, *14*(4), 434–449. https://doi.org/10.1007/s11119-013-9306-9

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108

Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., & Polasky, S. (2002). *Agricultural sustainability and intensive production practices*. 671–677.

Tomich, T. P., Brodt, S., Ferris, H., Galt, R., Horwath, W. R., Kebreab, E., Leveau, J. H. J., Liptzin, D., Lubell, M., Merel, P., Michelmore, R., Rosenstock, T., Scow, K., Six, J., Williams, N., & Yang, L. (2011). Agroecology: A review from a global-change perspective. *Annual Review of Environment and Resources*, *36*, 193–222. https://doi.org/10.1146/annurev-environ-012110-121302

Trevisan, R. G., Bullock, D. S., & Martin, N. F. (2021). Spatial variability of crop responses to agronomic inputs in on-farm precision experimentation. *Precision Agriculture*, *22*(2), 342–363. https://doi.org/10.1007/s11119-020-09720-8

Trudgill, S. T., Burt, T. P., Heathwaite, A. L., & Arkell, B. P. (1991). Soil nitrate sources and nitrate leaching losses, Slapton, South Devon. *Soil Use and Management*, *7*(4), 200–206. https://doi.org/10.1111/j.1475-2743.1991.tb00875.x

USDA ERS. (2022, February 7). *Interactive charts: Food Expenditures*. USDA ERS - Interactive Charts: Food Expenditures. Retrieved April 13, 2022, from https://www.ers.usda.gov/data-products/food-expenditure-series/interactive-charts-food-expenditures/#household

U.S. Environmental Protection Agency. (2018). *Glacier NP (GLR468)*. EPA CASTNET. https://www3.epa.gov/castnet/site_pages/GLR468.html

Van Sanford, D. A., & MacKown, C. T. (1986). Variation in nitrogen use efficiency among soft red winter wheat genotypes. *Theoretical and Applied Genetics*, *72*(2), 158–163. https://doi.org/10.1007/BF00266987

Ver Hoef, J. M., Hanks, E. M., & Hooten, M. B. (2017). On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. *Spatial Statistics*, *25*, 68–85. https://doi.org/10.1016/j.spasta.2018.04.006

Ver Hoef, J. M., Petersen, E. E., Hooten, M. B., Hanks, E. M., & Fortin, M.-J. (2018). Spatial autoregressive models for statistical inference from ecological data. *Ecological Monographs*, *88*(1), 36–59.

Vigil, M. F., Eghball, B., Cabrera, M. L., Jakubowski, B. R., & Davis, J. G. (2002). Accounting for seasonaI nitrogen mineralization : An overview. *Journal of Soil and Water Conservation*, *57*(6), 464–469.

Vinila Kumari, S., Bargavi, P., & Subhashini, U. (2016). Role of Big Data Analytics in Agriculture. *Special Issue on Computational Science, Mathematics and Biology*, 110–113. https://doi.org/10.18645/IJCSME.SPC.0025

Vitousek, P. M., Aber, J. D., Howarth, R. W., Likens, G. E., Matson, P. A., Schindler, D. W., Schlesinger, W. H., & Tilman, D. G. (1997). Human alteration of the global nitrogen cycle: Sources and consequences. *Ecological Applications*, *7*(3), 737–750. https://doi.org/10.1890/1051-0761(1997)007[0737:HAOTGN]2.0.CO;2

Wagner, P., & Hank, K. (2013). Suitability of aerial and satellite data for calculation of site-specific nitrogen fertilisation compared to ground based sensor data. *Precision Agriculture*, *14*(2), 135–150. https://doi.org/10.1007/s11119-012-9278-1

Wajid, A., Hussain, K., Ilyas, A., Habib-Ur-rahman, M., Shakil, Q., & Hoogenboom, G. (2021). Crop Models: Important Tools in Decision Support System to Manage Wheat Production under Vulnerable Environments. *Agriculture (Switzerland)*, *11*(11). https://doi.org/10.3390/AGRICULTURE11111166

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, *121*(2), 311–324. https://doi.org/10.1016/S0378-3758(03)00111-3

Walsh, E. S., Kreakie, B. J., Cantwell, M. G., & Nacci, D. (2017). A Random Forest approach to predict the spatial distribution of sediment pollution in an estuarine system. *PLoS ONE, 12*(7), 1–18. https://doi.org/10.1371/journal.pone.0179473

Walters, C. J., & Hilborn, R. (1978). Ecological Optimization and Adaptive Management. *Annual Review of Ecology and Systematics*, *9*(1), 157–188. https://doi.org/10.1146/annurev.es.09.110178.001105

Wang, J. C., Holan, S. H., Nandram, B., Barboza, W., Toto, C., & Anderson, E. (2012). A Bayesian Approach to Estimating Agricultural Yield Based on Multiple Repeated Surveys. *Journal of Agricultural, Biological, and Environmental Statistics*, *17*(1), 84–106. https://doi.org/10.1007/s13253-011-0067-5

Wang, L., Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Crop Journal*, *4*(3), 212–219. https://doi.org/10.1016/j.cj.2016.01.008

Wang, Y., Wu, G., Deng, L., Tang, Z., Wang, K., Sun, W., & Shangguan, Z. (2017). Prediction of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-07197-6

Watson C.A., Atkinson, D., Gosling, P., Jackson, L. R., & Rayns, F. W. (2002). Managing soil fertility in organic farming systems. *Soil Use and Management*, *18*(3), 239–247. https://doi.org/10.1079/sum2002131

Weersink, A., Fraser, E., Pannell, D., Duncan, E., & Rotz, S. (2018). Opportunities and Challenges for Big Data in Agricultural and Environmental Analysis. *Annual Review of Resource Economics*, *10*(August), 19–37. https://doi.org/10.1146/annurev-resource-100516-053654

Weil, R. R., & Brady, N. C. (2017). *The Nature and Properties of Soils* (D. Fox, A. Gilfillan, & L. Dimmick (eds.); 15th ed.). Pearson.

Weiner, J. (2017). *Applying plant ecological knowledge to increase agricultural sustainability*.

865–870. https://doi.org/10.1111/1365-2745.12792

West, P. C., Gerber, J. S., Engstrom, P. M., Mueller, N. D., Brauman, K. A., Carlson, K. M., Cassidy, E. S., Johnston, M., Macdonald, G. K., Ray, D. K., & Siebert, S. (2014). Leverage points for improving global food security and the environment. *Food Security*, *345*(6194), 1–5.

White, B. (2012). Agriculture and the Generation Problem: Rural Youth, Employment and the Future of Farming. *IDS Bulletin*, *43*(6), 9–19. https://doi.org/10.1111/j.1759-5436.2012.00375.x

Whitlock, C., Cross, W., Maxwell, B., Silverman, N., & Wade, A. (2017). 2017 Montana Climate Assessment. *Montana Institute on Ecosystems*, 269. http://montanaclimate.org/chapter/title-page%0Ahttp://ieeexplore.ieee.org/document/6340318/

Wills, M. (2016, May 4). *Are We Entering a New Golden Age of Guano?* JSTOR Daily. Retrieved February 8, 2022, from https://daily.jstor.org/golden-age-of-guano/

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big Data in Smart Farming – A review. *Agricultural Systems*, *153*, 69–80. https://doi.org/10.1016/j.agsy.2017.01.023

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *65*(1), 95–114. https://doi.org/10.1111/1467-9868.00374

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, *111*(516), 1548–1563. https://doi.org/10.1080/01621459.2016.1180986

Wright, M.N. & A. Ziegler (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 77(1), 1-17. doi:10.18637/jss.v077.i01

Wyeth, P. (2009). Sustainable Agriculture and Natural Resource Manageemnt in Farm Enterprise Systems. In K. M. Moore (Ed.), *The Sciences and Art of Adaptive Management: Innovating* *for Sustainable Agriculture and Natural Resources Management* (pp. 60–77). Soil and Water Conservation Society.

Xia, H., Ding, Y., & Wang, J. (2008). Gaussian process method for form error assessment using coordinate measurements. *IIE Transactions (Institute of Industrial Engineers)*, *40*(10), 931–946. https://doi.org/10.1080/07408170801971502

Yin, L., Liu, K., Li, L., Wei, M., Yang, R., Xue, K., Cao, Z., Zhang, C., Li, Y., Wu, X., & Wang, X. (2020). Late-sown winter wheat requires less nitrogen input but maintains high grain yield. *Agronomy Journal*, *112*(3), 1992–2005. https://doi.org/10.1002/agj2.20171

Yin, X., Goudriaan, J., Lantinga, E. A., Vos, J., & Spiertz, H. J. (2003). A flexible sigmoid function of determinate growth. *Annals of Botany*, *91*(3), 361–371. https://doi.org/10.1093/aob/mcg029

Zeng, X., Han, B., Xu, F., Huang, J., Cai, H., & Shi, L. (2012). Effects of modified fertilization technology on the grain yield and nitrogen use efficiency of midseason rice. *Field Crops Research*, *137*, 203–212. https://doi.org/10.1016/j.fcr.2012.08.012

Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture - a worldwide overview. *Computers and Electronics in Agricultre*, *36*, 113–132. http://linkinghub.elsevier.com/retrieve/pii/S002216941401066X

Zuur, A. F., & Camphuysen, Kees, C. (2012). 6 Generalized Additive Models applied on northern gannets. In *A Beginners Guide to Generalized Additive Models with R* (Issue January, p. 15). Highland Statistics Ltd.

APPENDICES

APPENDIX A

FULL RMSE TABLE

Appendix A. Table of RMSE values for grain yield, grain protein concentration, and net-return for each of the five model forms and for each of the seven fields.

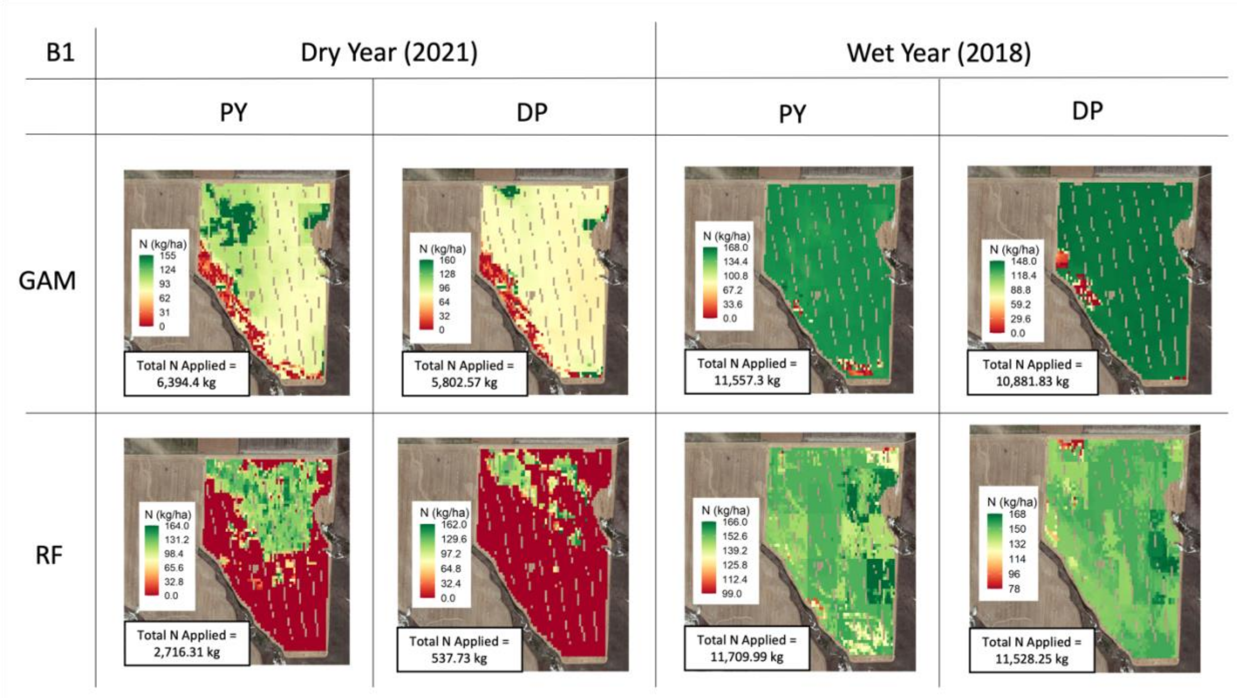
RMSE for Grain Yield (kg ha^{-1})				RMSE for Protein Concentration %				RMSE for Net-Return (\$ ha^{-1})			
Beta Function				***Beta Function***				***Beta Function***			
Field	*A*	*B*	*C*	*Field*	*A*	*B*	*C*	*Field*	*A*	*B*	*C*
B1	3148.90	2935.09	1238.74	B1	3.19	2.36	1.95	B1	977.28	1345.10	2333.30
B2	2940.79	3334.18	1660.90	B2	3.37	3.14	2.67	B2	1250.88	1040.66	1014.49
B3	1584.93	1690.53	6871.70	B3	8.06	1.72	1.82	B3	2587.74	541.11	1045.44
D1	1297.01	1547.13	1056.54	D1	1.45	1.27	1.98	D1	602.19	355.12	361.10
D2	1622.90	1567.17	1204.33	D2	8.10	1.02	1.47	D2	8157.14	553.68	463.72
D3	1744.54	926.49	1328.11	D3	2.04	1.45	2.22	D3	581.17	587.50	1141.51
I1	1146.55	1112.78	7180.80	I1	5.54	1.52	4.16	I1	2427.47	848.42	303.29
GAM				***GAM***				***GAM***			
Field	*A*	*B*	*C*	*Field*	*A*	*B*	*C*	*Field*	*A*	*B*	*C*
B1	2418.55	5080.02	7546.72	B1	2.84	6.10	1.67	B1	862.55	1511.10	2990.01
B2	981.91	4583.06	883.94	B2	5.79	3.41	2.15	B2	2315.97	1160.35	588.45
B3	7771.55	2362.69	968.14	B3	2.02	1.60	1.53	B3	2760.57	1150.34	441.62
D1	758.17	1701.73	957.82	D1	1.53	1.75	2.76	D1	363.11	489.02	477.62
D2	7681.69	2653.18	843.23	D2	7.28	3.83	1.53	D2	6523.76	1461.54	458.82
D3	1390.75	2172.43	2806.35	D3	1.52	0.96	1.53	D3	533.45	564.66	833.98
I1	15106.63	2945.50	1653.00	I1	16.60	7.53	9.73	I1	10563.73	1593.62	4844.12
Bayes NLR				***Bayes NLR***				***Bayes NLR***			
Field	*A*	*B*	*C*	*Field*	*A*	*B*	*C*	*Field*	*A*	*B*	*C*
B1	2147.47	1976.54	2243.31	B1	7.71	7.44	4.31	B1	3699.99	3725.88	3102.81
B2	1848.52	2153.62	1647.32	B2	8.40	9.34	5.67	B2	2810.97	5602.23	2603.31
B3	2974.07	2823.95	2183.72	B3	5.39	3.93	3.99	B3	1718.17	2244.68	2135.27
D1	1372.56	1500.15	1346.72	D1	1.70	1.74	1.78	D1	510.91	420.02	438.31
D2	1816.12	2219.94	1534.55	D2	5.58	4.27	3.95	D2	1024.08	2314.66	2313.80
D3	1374.78	1919.44	1271.42	D3	4.07	3.85	3.33	D3	802.84	2040.95	1970.09
I1	2128.79	1764.97	1617.64	I1	9.42	5.15	7.64	I1	3321.58	3100.18	3820.93
Bayes MLR				***Bayes MLR***				***Bayes MLR***			
Field	*A*	*B*	*C*	*Field*	*A*	*B*	*C*	*Field*	*A*	*B*	*C*
B1	2010.19	2578.19	1827.39	B1	3.40	5.68	2.29	B1	1034.45	2355.51	1966.75
B2	1019.07	2506.86	2679.51	B2	3.79	3.24	2.36	B2	1789.97	1379.80	1028.03

APPENDIX A CONTINUED

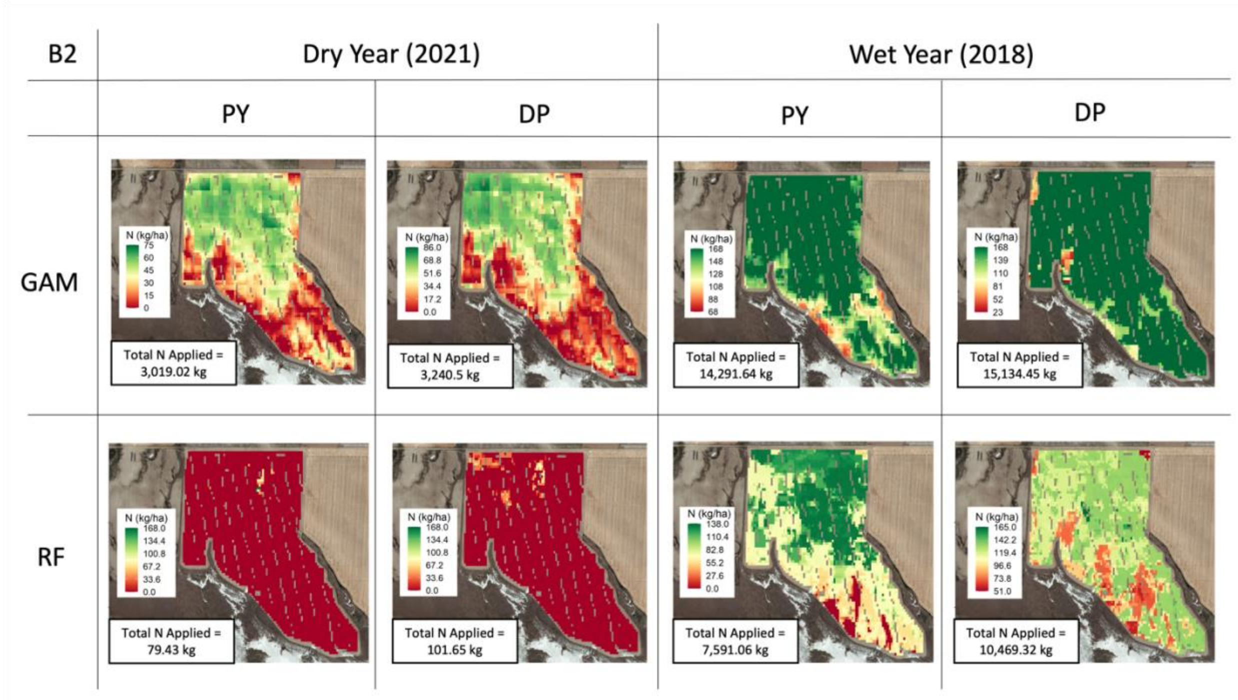
B3	4280.07	2239.78	1298.46	B3	10.06	2.32	2.29	B3	8177.02	946.31	727.07
D1	826.17	1577.48	852.38	D1	2.07	2.45	1.93	D1	372.53	585.44	333.15
D2	2591.85	2246.79	1173.60	D2	7.59	4.17	10.15	D2	5912.85	1657.33	2933.55
D3	1518.81	1575.44	1518.61	D3	2.00	1.86	2.38	D3	721.41	905.55	1224.52
I1	3085.74	14082.57	3013.43	I1	7.58	17.52	3.61	I1	1660.66	4991.10	783.52
RF				***RF***				***RF***			
Field	*A*	*B*	*C*	*Field*	*A*	*B*	*C*	*Field*	*A*	*B*	*C*
B1	2693.10	2313.15	2229.90	B1	3.42	2.39	3.82	B1	809.90	1417.41	2530.17
B2	1425.97	3167.28	799.69	B2	2.64	2.86	2.65	B2	812.64	939.62	697.09
B3	1350.20	1676.90	1026.12	B3	1.69	1.74	1.78	B3	639.70	540.88	454.77
D1	786.89	1656.70	626.76	D1	1.43	1.31	1.70	D1	406.90	389.80	223.81
D2	822.63	1851.63	1027.60	D2	1.53	1.20	1.32	D2	639.13	637.55	406.80
D3	1450.04	989.93	1122.47	D3	1.57	1.35	2.21	D3	584.27	510.87	1099.79
I1	1200.55	1154.99	651.18	I1	3.12	2.77	3.44	I1	1878.46	1226.32	817.23

APPENDIX B

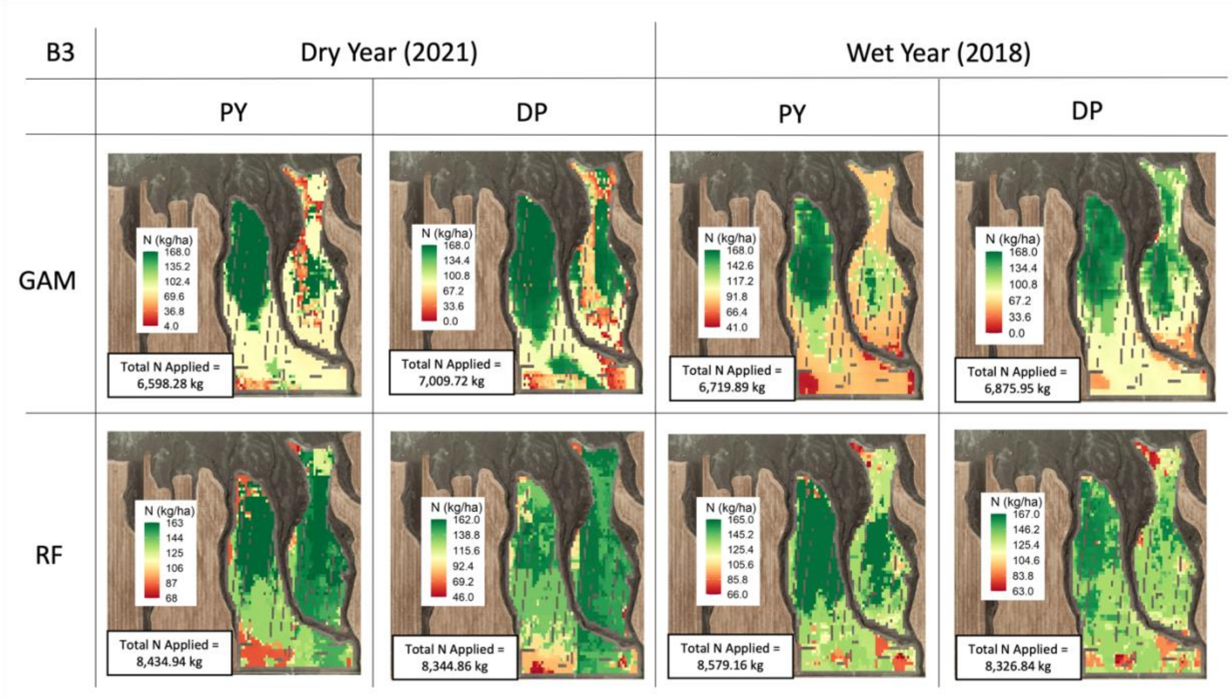
PROFIT MAXIMIZING NITROGEN FERTILIZER MAPS



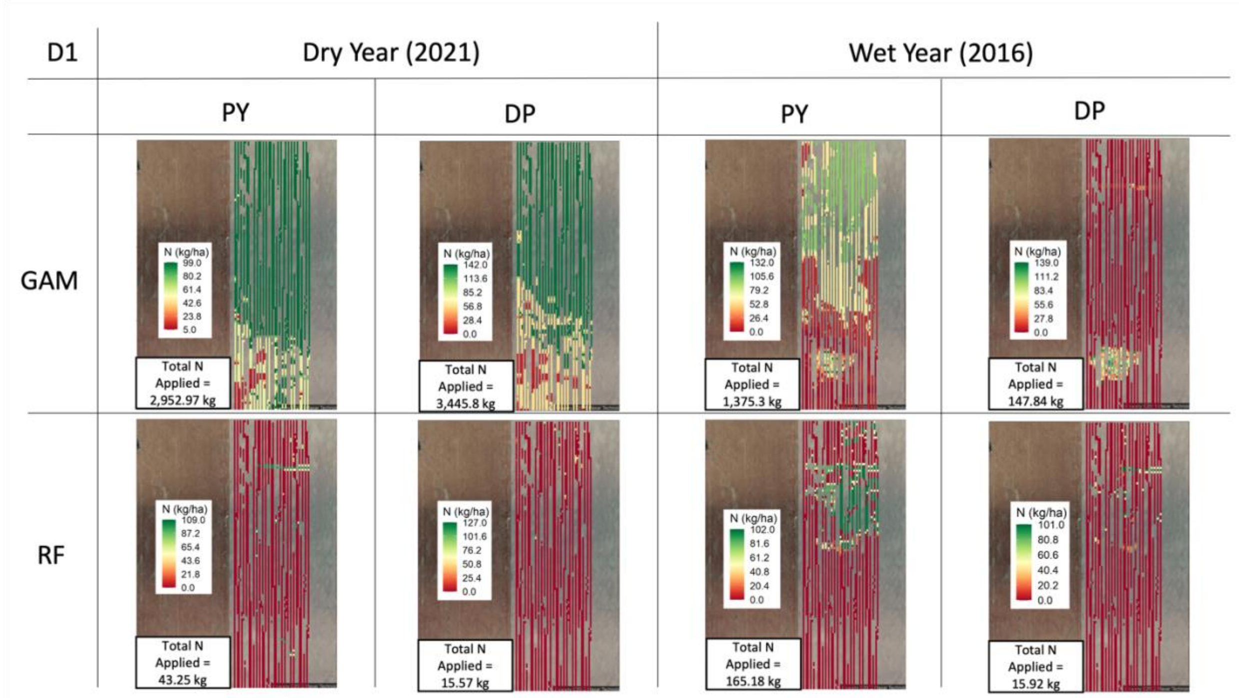
Appendix Figure B1. Site-specific profit maximizing nitrogen fertilizer rates for each model type and data constraint in a simulated dry and wet year for field B1.



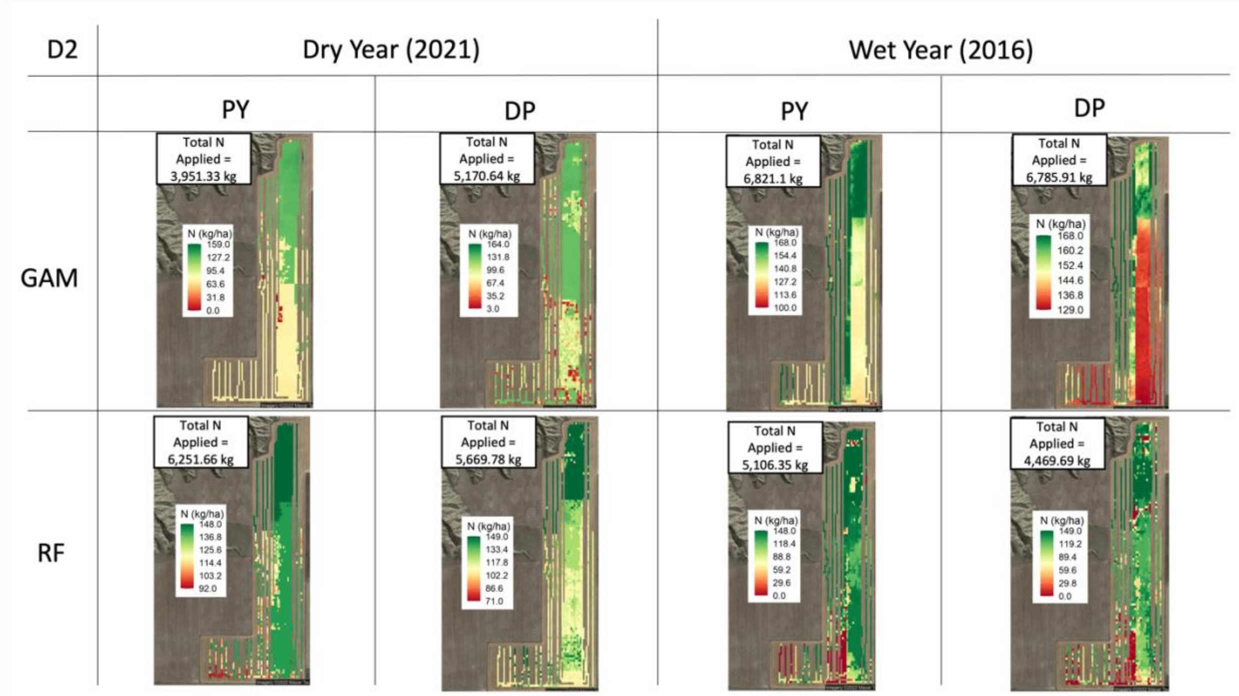
Appendix Figure B2. Site-specific profit maximizing nitrogen fertilizer rates for each model type and data constraint in a simulated dry and wet year for field B2.



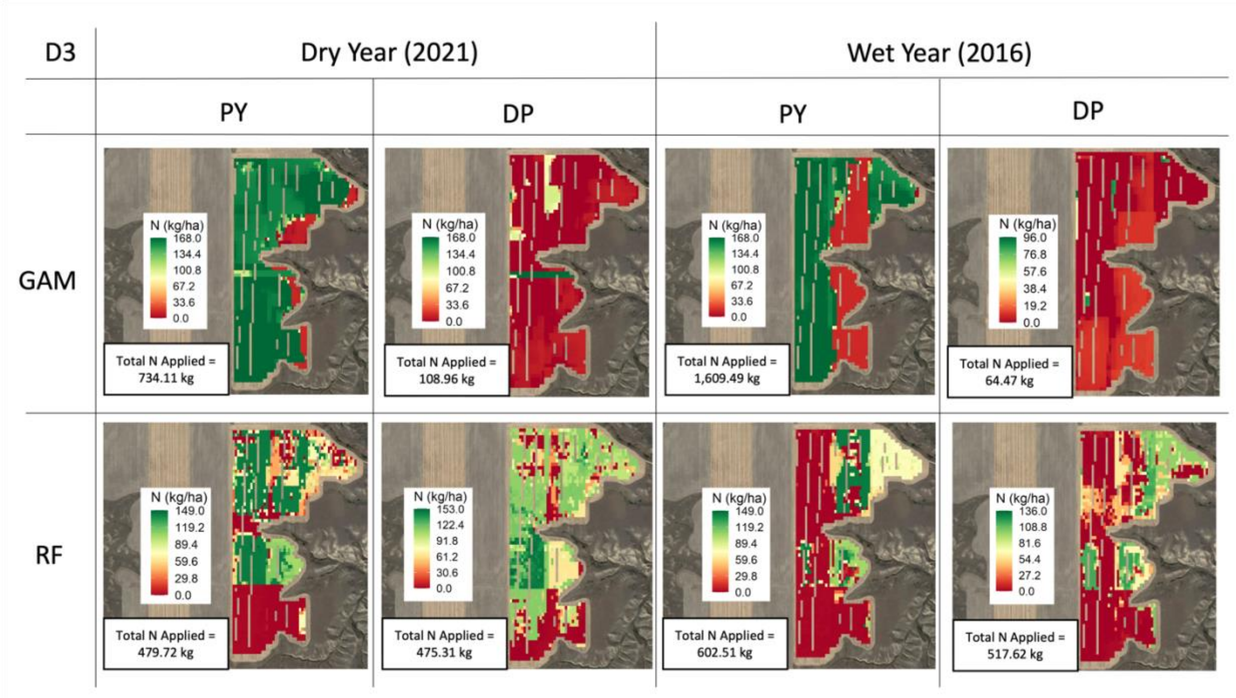
Appendix Figure B3. Site-specific profit maximizing nitrogen fertilizer rates for each model type and data constraint in a simulated dry and wet year for field B3.



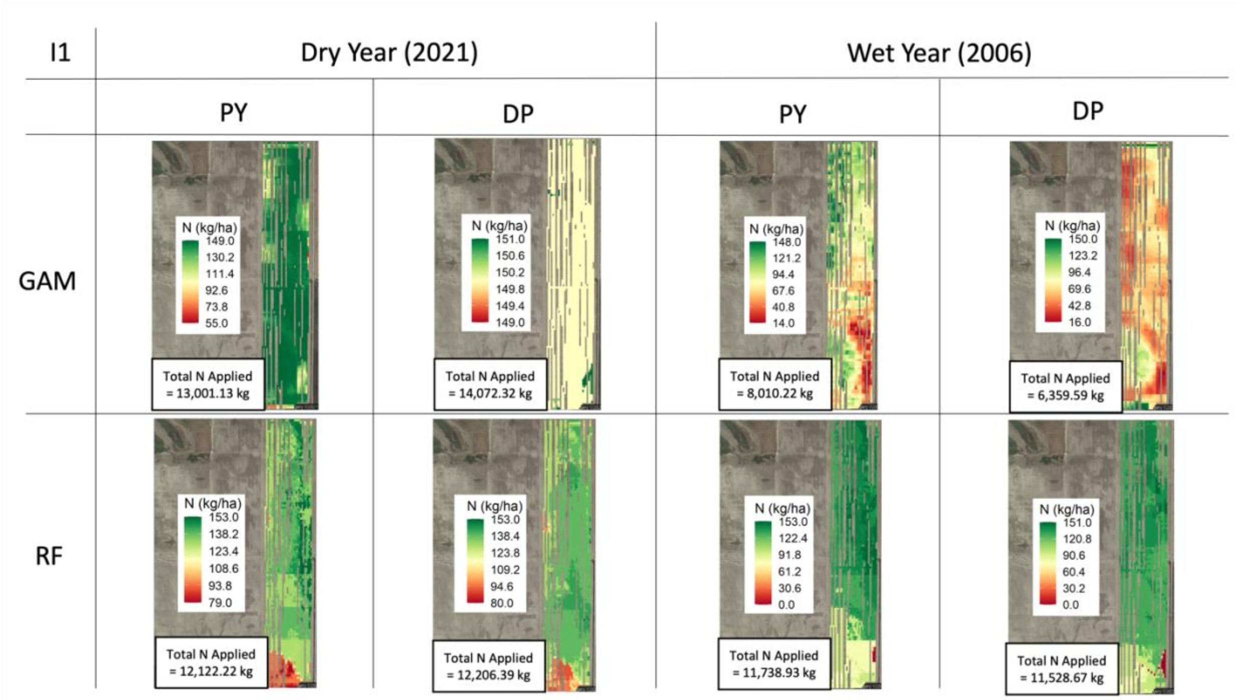
Appendix Figure B4. Site-specific profit maximizing nitrogen fertilizer rates for each model type and data constraint in a simulated dry and wet year for field D1.



Appendix Figure B5. Site-specific profit maximizing nitrogen fertilizer rates for each model type and data constraint in a simulated dry and wet year for field D2.



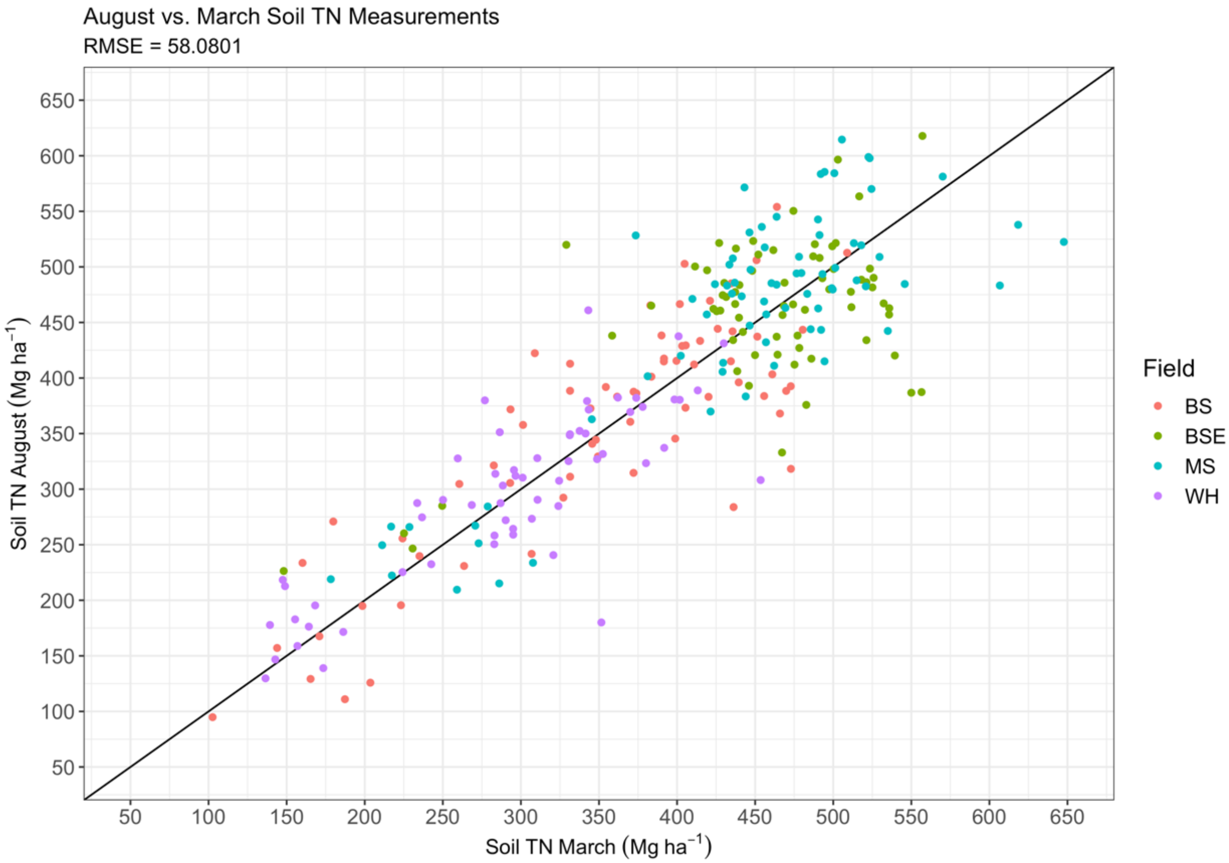
Appendix Figure B6. Site-specific profit maximizing nitrogen fertilizer rates for each model type and data constraint in a simulated dry and wet year for field D3.



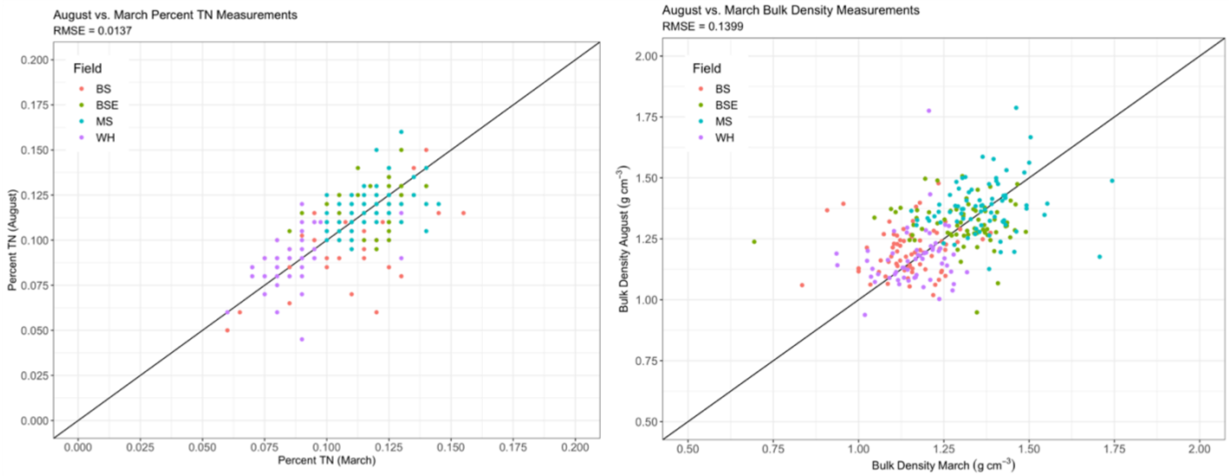
Appendix Figure B7. Site-specific profit maximizing nitrogen fertilizer rates for each model type and data constraint in a simulated dry and wet year for field I1.

APPENDIX C

SOIL MEASUREMENT UNCERTAINTIES



Appendix Figure C1. Scatterplot of August vs. March soil TN measurements. The RMSE is calculated in units of Mg ha^{-1}, and the 1:1 line indicates a perfect match between measurements.



Appendix Figure C2. Scatterplot of measured percent TN in August vs. March (left) and measured bulk density in August vs. March (right). Colors represent different fields, the RMSE is in the units of the axes, and the 1:1 line indicates a perfect match between measurements.