A quantitative assessment of soybean yield and nitrogen economy

by

Luiz Felipe Antunes de Almeida

B.S., State University of Maringá, 2020

AN ABSTRACT OF DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agronomy
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Abstract

Improving the nitrogen (N) economy in soybeans (*Glycine max* L.) production is essential for developing more efficient, profitable, and environmentally sustainable agricultural systems. While soybeans can meet much of their N demand through biological N_2 fixation, uncertainty remains around when and where fertilizer inputs are needed and how environmental conditions influence N uptake, seed yield, and quality. This dissertation combines field data, crop modeling, and machine learning to provide data-driven insights into soybean N dynamics across diverse U.S. environments. Chapter 1 provides a summary and presentation of the background and objectives of this dissertation. Chapter 2 uses a standardized protocol to evaluate soybean yield responses to N and sulfur (S) fertilization across 26 field trials located across the Midwest region of the United States (US). Results showed that N fertilization rarely increased yield and often introduced uncertainty. In contrast, S fertilization improved N uptake, yield, and N status in some environments. An apparent N dilution curve was developed to help identify in-season plant N limitations, improve nutrient diagnosis, and provide more precise recommendations. Chapter 3 explores the effect of weather and soil factors that influence seed yield, N_2 fixation, and their uncertainty using data from 35 sites. Precipitation, vapor pressure deficit, and soil texture emerged as key factors explaining variability in both yield and N_2 fixation. Small additions of S helped improve N uptake and yield stability in environments with limited organic matter or water availability.

Chapter 4 focuses on characterizing seasonal patterns of N_2 fixation using plant samples collected at reproductive stages and analyzed through Bayesian modeling. Peak N_2 fixation typically occurred between full pod and seed fill stages but varied widely depending on water availability and atmospheric conditions. These results highlight the limitations of relying on

single-timepoint measurements and the value of time-series data for modeling complex biological processes. Together, the findings of this dissertation support more targeted, data-informed recommendations for improving yield and N use efficiency while delivering practical tools to help farmers, advisors, and researchers advance soybean productivity and long-term sustainability.

A quantitative assessment of soybean yield and nitrogen economy

by

Luiz Felipe Antunes de Almeida

B.S., State University of Maringá, 2020

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agronomy
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Approved by:

Co-Major Professor
Dr. P. V. Vara Prasad

Approved by:

Co-Major Professor
Dr. Ignacio A. Ciampitti

Copyright

Abstract

Improving the nitrogen (N) economy in soybeans (*Glycine max* L.) production is essential for developing more efficient, profitable, and environmentally sustainable agricultural systems. While soybeans can meet much of their N demand through biological N_2 fixation, uncertainty remains around when and where fertilizer inputs are needed and how environmental conditions influence N uptake, seed yield, and quality. This dissertation combines field data, crop modeling, and machine learning approach to provide data-driven insights into soybean N dynamics across diverse U.S. environments. Chapter 1 provides a summary, presentation of background, and objectives of this dissertation. Chapter 2 evaluates soybean yield responses to N and sulfur (S) fertilization across 26 field trials located across the Midwest region of the United States (US) using a standardized protocol. Results showed that N fertilization rarely increased yield and often introduced uncertainty. In contrast, S fertilization improved N uptake, yield, and N status in some environments. An apparent N dilution curve was developed to help identify in-season plant N limitations, improve nutrient diagnosis, and provide more precise recommendations.

Chapter 3 explores the effect of weather and soil factors influence seed yield, N_2 fixation, and their uncertainty using data from 35 sites. Precipitation, vapor pressure deficit, and soil texture emerged as key factors explaining variability in both yield and N_2 fixation. Small additions of S helped improve N uptake and yield stability in environments with limited organic matter or water availability. Chapter 4 focuses on characterizing seasonal patterns of N_2 fixation using plant samples collected at reproductive stages and analyzed through Bayesian modeling. Peak N_2 fixation typically occurred between full pod and seed fill stages but varied widely depending on water availability and atmospheric conditions. These results highlight the limitations of relying

on single-timepoint measurements and the value of time-series data for modeling complex biological processes.

Together, the findings of this dissertation support more targeted, data-informed recommendations for improving yield and N use efficiency, and seed quality, while delivering practical information to help farmers, advisors, and researchers advance soybean productivity and long-term sustainability.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This dissertation represents more than the completion of a degree. It reflects the support, trust, and encouragement I've received from many people along the way. I am deeply grateful to Dr. Ignacio Ciampitti for his mentorship and for setting an example through his leadership, dedication, and generosity. Thank you for pushing me to think critically, for trusting me with responsibilities that helped me grow, and for all the opportunities that shaped my path.

Thank you as well to my committee members, Dr. Adrian Correndo, Dr. Trevor Hefley, and Dr. Vara Prasad for your invaluable feedback, encouragement, and belief in my work. Your guidance and diverse perspectives have been instrumental in helping me grow as a professional and develop a more critical mindset, constantly challenging me to push boundaries.

Special thanks to my colleagues and friends: Luiz Moro, Andre Reis, Adrian Correndo, Dennis Boller, Federico Gomez, Gabriel Hintz, Valentina Pereyra, and Francisco Palmero. It was a privilege to learn and work with you. The long field days, tough moments, and good memories we shared made this journey meaningful. I also want to thank all the other students, visiting scholars, interns, and postdocs from the Ciampitti Lab who helped make this work possible.

I am also grateful for the institutional and collaborative support that made this research possible. Financial support was provided by the Kansas Soybean Commission and Corteva Agriscience. I extend my sincere appreciation to AdvanSix and Agrauxine North America for their contribution of materials for our experiments. My deepest thanks go to the collaborating principal investigators, technicians, and research scholars from other universities who contributed in some way to this dissertation. Your contributions to this chapter were invaluable.

To my wife, Larissa, thank you for your love, care, and the strength you bring into my life each day. Your unwavering support, encouragement, companionship, and kindness have shaped this journey more than anything else. There will never be words big enough to express how grateful I am to you. To my parents, Rosana and Valdir, my sister, Amanda, and my grandparents, thank you for always being there for me and for supporting my decision to move so far from home to pursue my dreams. Finally, to Alexei and Helena, I feel incredibly lucky and thankful to have you as my in-laws. This achievement is not mine alone, it belongs to all of you!

Dedication

I would like to dedicate this dissertation to my wife, Larissa, whose love, support, and belief in me have been my greatest source of strength and purpose. To my parents, Rosana and Valdir, thank you for always standing by me, loving me, and encouraging me to pursue this journey. I deeply appreciate all that you have done for me.

Chapter 1 - General Introduction

Soybean [Glycine max *(L.)* Merr.] requires large amounts of nitrogen (N) to achieve higher yields (Sinclair and de Wit, 1975), often raising concerns about long-term sustainability. The spatio-temporal factors influencing plant growth and soil processes add a layer of uncertainty and complexity to the development of environmentally and economically tailored management practices to optimize the N economy in soybeans (biological N_2 fixation, N uptake, and N budget). Concerning internal use efficiency, N has implications not only for seed yield but also for seed composition. Soybean needs 1 g of glucose to produce 0.33 g of lipid, 0.40 g of protein, or 0.83 g of carbohydrates (De Vries et al., 1974), implying that a possible energetic cost of N_2 fixation can shift protein and oil concentration (Tamagno et al., 2018). Notably, N fertilization has reportedly increased the total productivity of protein and oil, often due to a decreased seed protein-to-oil ratio, favoring greater oil accumulation in seeds. Therefore, understanding N dynamics in soybeans is critical for selecting plants with improved N utilization efficiency, leading to higher yields while maintaining biomass allocation to seeds (Assefa et al., 2019).

Historically, soybean has shown consistent annual improvements in seed yield (0.058 Mg ha^{-1}), plant biomass (0.025 Mg ha^{-1}), and N uptake (1.57 kg ha^{-1}). At the same time, increments in the N harvest index have minimally increased N use efficiency (Balboa et al., 2018). Thus, enhancing N uptake in soybeans remains challenging, particularly due to the limited understanding and predictability of biological N_2 fixation dynamics and the inhibitory effect of high N fertilization rates on this process (Ciampitti et al., 2021). Studies simultaneously addressing the impact of N fertilization on yield, plant N status, and biological N_2 fixation are scarce. Additionally, the typically limited geographical distribution of experiments, lack of standardized methodologies, and superficial analyses have constrained the generalizability of outcomes and overlooked intrinsic uncertainties underlying these processes.

Although soybeans rarely respond to N fertilization (Almeida et al., 2023), there may be specific environmental circumstances under which it could occur. Identifying these low-probability scenarios is challenging due to the restricted quantification of key elements such as N_2 fixation and plant nutritional status. Recent findings demonstrating that other nutrients, such as sulfur (S), enhance N uptake in cereal crops (Salvagiotti et al., 2009) open new opportunities to explore nutrient interactions as potential co-limiting factors (Borja Reis et al., 2021). However, a shift in the soybean research paradigm toward more comprehensive and quantitative approaches is necessary to effectively capture complex interactions and deliver tailored recommendations. We often face a paradox of pursuing more sustainable farming systems while overlooking fundamental processes such as N_2 fixation in soybeans. Despite its central role in the crop's N economy, the seasonal dynamics of N_2 fixation have received limited attention in modern research (Ciampitti et al., 2021; Córdova et al., 2019; Moro Rosso et al., 2023). To address complex biological questions in soybean systems, it is often necessary to bridge multiple advanced analytical tools, including Bayesian modeling, machine learning, and crop simulation, that can generate novel insights and inform future research and decision-making.

To better understand N limitations in soybeans and clarify the role of S on N_2 fixation, N uptake, and yield, Chapters 2, 3, and 4 present a quantitative assessment of these components using integrated modeling approaches. These chapters also provide a comprehensive framework for characterizing the seasonal dynamics of N_2 fixation, uncovering opportunities to enhance environmental and economic sustainability, contribute to food security, and align production systems with evolving global demands.

Chapter 2 - Soybean yield response to nitrogen and sulfur fertilization in the United States: contribution of soil N and N fixation processes

Published in European Jornal of Agronomy

Almeida, L.F.A., Correndo, A., Ross, J., Licht, M., Casteel, S., Singh, M., Naeve, S., Vann, R., Bais, J., Kandel, H. and Lindsey, L., 2023. Soybean yield response to nitrogen and sulfur fertilization in the United States: contribution of soil N and N fixation processes. European Journal of Agronomy, 145, p.126791. https://doi.org/10.1016/j.eja.2023.126791

Abstract

Soybean [Glycine max (L.) Merr.] is the most important legume grown worldwide. The effect of nitrogen (N) and sulfur (S) fertilization on seed yield is commonly studied in the United States (US). However, soybean yield response to fertilization remains inconsistent, partly due to the lack of standardized field designs and a better understanding of the plant nutrition processes underpinning yield formation. The aims of this study were to assess the i) seed yield, (ii) plant N status (as N nutrition index, NNI), (iii) the contribution of N fixation, and (iv) the uncertainties on i), ii), and iii) in response to N-S fertilization using a uniform protocol across environments.

Twenty-six trials in twelve US states tested five fertilization strategies that combined N and S at varying rates and timings. Using Bayesian statistics, seed yield response to fertilizer, NNI, and contribution of N fixation were analyzed at site and treatment levels providing both magnitude of responses and estimation of their uncertainties. From the significance of responses on seed yield, sites were split into two groups: non-responsive (18 sites) and responsive (8 sites). The NNI, ratio of the actual to the critical plant N concentration, was calculated to diagnose soybean N deficiency, and the N derived from the atmosphere (Ndfa, %) as N fixation contribution were investigated to better understand the source of plant N across all sites.

Roughly for three-fourths of the sites, fertilization resulted in an unlikely (non-responsive) yield effect, with uncertainties ranging from 0.09 to 2.62 Mg ha^{-1}. The other one-third of the sites were mainly responsive to S or both N+S, with the yield responses ranging from -0.42 to 1.1 Mg ha^{-1} and uncertainties varying from 0.47 to 1.36 Mg ha^{-1}. For the yield responsive sites, NNI presented a high proportion of deficiency (NNI<1) for most of the treatments, except for the "Full" signaling to a potential for yield response. Likewise, only 6% of the changes in Ndfa were not related to the treatment "Full", and regardless of the seed yield response to fertilization, within the same site, soil and N fixation showed similar contributions to plant N demand.

Due to the high uncertainty in treatment response and contribution of N fixation, N fertilization is unlikely to increase yields, leading to non-profitable recommendations. Sulfur deficiency, on the other hand, should be explored under site-specific conditions. A decision support system should include appropriate diagnosis methods for identifying N and S deficiencies, such as NNI in soybean. Attainable maximum Ndfa did not appear to be affected by fertilization but largely varying depending on the site. Future research should assess the role of soil and meteorological variables underpinning N fixation and soil N, along with the impact on seed quality composition, as a critical trait for this crop.

2.1 Introduction

Soybean [*Glycine max* (L.) Merr.] is a major field crop in the United States (US), reaching a total of 35 million hectares (USDA NASS, 2021a) and with production exceeding 119 million Mg during the 2021 season. Brazil, Argentina, and the US are the top producers, contributing to over 75% of global production (USDA NASS, 2021b). Soybean has significant relevance for animal consumption, but it is commonly enriched with sulfur (S) for livestock feeding (Krishnan, 2015). Furthermore, it accounts for more than half of total oilseed production (USDA, 2022).

Seed production, plant biomass, and nitrogen (N) uptake have demonstrated an overall annual increase by 0.058 Mg yield ha^{-1}, 0.025 Mg biomass ha^{-1}, and 1.57 kg N ha^{-1}, respectively, followed by enhanced nutrient internal efficiency due to a greater N harvest index (Balboa et al., 2018). Due to its high protein, the soybean crop is highly dependent on the process of N fixation to support maximum crop growth and satisfy plant N demand (Herridge and Bergensen, 1988; Sinclair and de Wit, 1975; Tamagno et al., 2018, 2017). The contribution of N fixation has been reported to range from 40% to 80% of the total N uptake (Salvagiotti et al., 2008; Zapata et al., 1987), depending on plant growth, soil, and weather factors (Ciampitti et al., 2021).

Soybean presents an erratic response in yield to N fertilization due to the contribution of the N fixation process, which often will sustain about half the seasonal crop N requirement, but this process is highly influenced by climatic and edaphic factors (Hungria and Vargas, 2000; Purcell and King, 1996; Santachiara et al., 2017). Overall, if crop growth is impaired, the contribution of N fixation may also decrease relative to the activity under optimum conditions (Raper, et al., 1977). The effect of fertilization is described as site-year specific and interacts with management practices (Mourtzinis et al., 2018). Recent US studies investigated the effect of sulfur on seed quality (e.g., protein, oil) and its relationship of this factor with seed yield (Borja Reis et al., 2021). The sulfur response is usually associated with low soil organic matter (SOM) levels where the soil supply may not satisfy the crop demand (Divito et al., 2015; Mahal et al., 2022). Moreover, for specific site conditions, S fertilization may also help to increase the N utilization by increasing N recovery from soil N supply, as documented for cereals (Salvagiotti et al., 2009).

A better understanding of the crop nutritional status is essential for more robust fertilizer recommendations (Tilman et al., 2011). Seasonal plant N concentration (Nc) declines with crop age (N dilution process) as plant biomass increases over time (Lemaire and Ciampitti, 2020). This process is represented by a mathematical power function model relating plant N concentration (Nc, %) and dry biomass (W, Mg ha^{-1}) (Lemaire and Gastal, 1997). Given the importance of this physiological process, average N dilution curves were proposed for several field crops (C3 and C4) (Lemaire et al., 2021), including other grain legumes such as pea and lucerne (Ney et al., 1997). This critical curve defines the minimum Nc needed to maximize crop growth, herein termed as the critical Nc. The association of the actual Nc to the critical Nc defines the Nitrogen Nutrition Index (NNI), calculated at given plant biomass.

Recent literature has addressed the effects of crop management practices on soybean performance based on experimental networks and synthesis reviews (Andrade et al., 2019; Borja Reis et al., 2021; Santos et al., 2019; Froes et al., 2022; Rotundo and Westgate, 2009; Tamagno et al., 2017). However, the lack of standardized methodologies across sites is still a common limitation to study the effect of management on seed yield, plant nutritional status, and their connection with N fixation contribution to plant N demand. In this regard, establishing a multi-state trial network under a standard experimental design will enable straightforward data analysis and make inferences unambiguous (Tremblay et al., 2012). For example, Matcham et al. (2021) studied the effect of foliar fertilization on soybean yields across 46 site-years with a common protocol, providing a more robust conclusion on the topic. To the extent of our knowledge, no trial large spatial scale networks have been formally established in the U.S. studying the soybean response to N and S fertilization following a standardized protocol and focusing on providing a determination of plant N and N fixation dynamics. This work aims to assess soybean response to

N and S fertilization strategies across different US environments under a standardized protocol. The specific goals were to assess the factorial effects of N-S fertilization on i) seed yield, (ii) plant N status (as N nutrition index, NNI), (iii) the contribution of N fixation, and (iv) the uncertainties on i), ii), and iii) in response to N-S fertilization using a uniform protocol across environments.

2.2 Material and Methods

2.2.1. Geographical distribution

During the 2021 growing season, 26 studies were carried out across twelve states in the U.S.: Arkansas (AR1 and AR2), Iowa (IA1 and IA2), Indiana (IN1, IN2, and IN3), Kansas (KS1, KS2, KS3, and KS4), Michigan (MI1 and MI2), Minnesota (MN1 and MN2), North Carolina (NC1 and NC2), North Dakota (ND1 and ND2), Ohio (OH1 and OH2), South Dakota (SD1 and SD2), Virginia (VA1 and VA2), and Wisconsin (WI1 and WI2) (Figure 2.1A).

2.2.2. Environmental conditions

Seasonal weather data for all sites was gathered from Daymet at a spatial resolution of 1 km^2 using *daymetr* package (Correndo et al., 2021a; Hufkens et al., 2018). Most of the sites were under rainfed conditions, but when an irrigation system was available (6 out of 26 sites), the crop was maintained out of water stress **(Table A.1)**. During the cropping season, precipitation levels ranged from 255 mm to 643 mm across studies, and mean air temperature ranged from 18°C to 25°C **(Figure 2.1C)**.

Soils under study included 22 soil series, ranging from loamy sand to clay soil textures **(Figure 2.1B)**. Prior to planting, compound (6 cores) soil samples at 0 to 0.15 m depth were collected at each location to describe their soil attributes such as pH, S.O.M., phosphorus (P), potassium (K), soil cation exchange capacity (C.E.C.), and soil texture (sand, silt, and clay percentage). In addition, 0 to 0.2 m and 0.2 to 0.61 m soil samples were taken to describe initial soil nitrate (NO3--N), soil ammonium (NH4+-N), and soil sulfate (SO4-2-S) fertility **(Figure 2.1B; Table A.2)**.

2.2.3. Experimental design

Across all sites, the experimental design followed a randomized complete block design with four replications. Plot dimensions were 12.2 to 15.2 m in length and 6.1 m wide. Soybean row spacing varied between 0.38 m to 0.76 m, with wider as the most frequent row width. Using standard agronomic practices, pests, diseases, and weeds were managed at each site. The treatments applied on the soybean crop consisted of five fertilization managements combining N and S applied at different rates and timings during the cropping season **(Table 2.1)**. Treatments "Check", "N", "S", and "NS" were defined considering potential fertilization strategies implemented by US soybean growers, with potential for return of investment. The treatment "Full" was introduced and split at planting and at R3 stage to provide a non-limiting N supply at the period of highest crop N demand. By including these 5 treatments, we are able to more accurately assess the impacts of N and S not only on seed yield, but also on the process of N fixation and on plant N status under varying treatments.

2.2.4. Plant sampling

Plant and soil sampling consisted of two phenological stages (Fehr et al., 1971) during the growing season: full bloom (R2) and full pod (R4) phenological stages (Fehr and Caviness, 1977). At all sites, aboveground samples were collected by cutting a 1.5 m-length row segment at ground level. The number of rows to be sampled was adjusted according to row spacing to equal the collected areas. A sub-sample of five plants was dried in an air-forced oven at 65□C until constant weight for dry biomass (Mg ha^{-1}). For laboratory determinations, all dried tissue samples were ground and sieved (250-µm) for determination of N derived from the atmosphere (Ndfa, %) and total nitrogen concentration (Nc, %) in plant tissue.

The Ndfa estimation was carried out to assess the percentage of N derived from the N_2 fixation process in each sampling time. The Ndfa was calculated by the 15N natural abundance method measured with the isotope mass spectrometer, which gives the excess of the 15N isotope in the soybean and the reference plant. For this purpose, adjacent to the soybean plots, a strip of unfertilized corn was planted to serve as the reference plant, which only relies on soil N (Ndfa was not assessed at R4 in Michigan due to the lack of reference unfertilized corn samples).Therefore, the following equation **(Equation 2.1)** was used (Unkovich et al., 2008):

$$\text{Ndfa} = \frac{\delta 15\text{N of corn} - \delta 15\text{N of soybean}}{\delta 15\text{N of corn} - \text{Bvalue}} \times 100 \quad \textbf{(Equation 2.1)}$$

where we assume "Bvalue" = -2.54, obtained from Balboa and Ciampitti (2020).

The Ndfa sampling time was done at the R4 crop growth stage, as documented by Ciampitti et al. (2021), to be the point for maximum Ndfa, remaining relatively constant until the end of the crop season.

At crop maturity, 1.5 m width of each plot was harvested by a combine (yield was not recorded at VA1 site), and seed yield was adjusted to 130 g kg^{-1} moisture basis and later extrapolated to hectares (Mg ha^{-1}).

2.2.5. Statistical analysis

2.2.5.1. Seed yield

Soybean seed yield response to treatments by site was analyzed following a hierarchical Bayesian framework. The top level of the model (data layer) describes the seed yield response (y_{ijs}, Mg ha^{-1}) for the i^{th} block within the j^{th} treatment, which is nested within the s^{th} site

(Equation 2.2). Soybean seed yield response $y_{ij(s)}$ was assumed to follow a normal distribution with expected mean $\mu_{ij(s)}$, and variance σ_s^2:

$y_{ij(s)} \sim N(\mu_{ij(s)}, \sigma_s^2)$, **(Equation 2.2)**

$i = 1, 2, \ldots, 5$,

$j = 1, 2, 3, 4, 5$,

$s = 1, 2, \ldots, 26$.

For the secondary level that describes the underlying process model, $\mu_{ij(s)}$ was defined as a linear effect, where $B_{1(s)}$ is the linear parameter, $X_{ij(s)}$ the observed seed yields for the i^{th} block and j^{th} treatment nested within the s^{th} site. Informative and uniformly distributed sample prior was defined to B_1 with a minimum value of 0 Mg ha^{-1} and a maximum of 10 Mg ha^{-1}, and for σ_s was assigned an uninformative Half-student-t (υ , θ, σ) prior, where υ , θ, σ represent the shape, the location, and the scale parameters, respectively.

$$\mu_{ij(s)} = \begin{cases} B_{1(1)} . X_{ij(1)} \\ \vdots \\ B_{1(26)} . X_{ij(26)} \end{cases},$$

$$\sigma_s^2 = \begin{cases} \sigma_1^2 \; for \; site \; 1 \\ \vdots \\ \sigma_{26}^2 \; for \; site \; 26 \end{cases},$$

$B_1 \sim Uniform(0,10)$

$\sigma_s \sim Half - student - t\ (3, 0, 2.5)$.

2.2.5.2. Crop nitrogen status

The crop N status was described by a relationship between Nc (%) and biomass (Mg ha^{-1}) (Lemaire et al., 1984). Following a hierarchical structure, the Nc data (y_i) was assumed to follow a normal distribution with expected mean μ_i, and variance σ^2 **(Equation 2.3)**, which describes the top level of the model (data layer):

$y_i \sim N(\mu_i, \sigma^2)$, **(Equation 2.3)**

$i = 1, 2, \ldots, 939.$

For the secondary level that describes the underlying process model, μ_i was defined as a non-linear effect model, were B_1 and B_2 are two non-linear parameters, and W_i the plant biomass measurements. The B_1 represents Nc at W = 1 Mg ha^{-1}, and B_2 (dimensionless) the ratio between the decline in Nc and relative crop growth. Informative and normally distributed sample priors were defined to $B_1$and based on possible values found in previous literature (Santachiara et al., 2018). For σ_s was assigned an uninformative *Half-student-t* (υ , θ, σ) prior, where υ , θ, σ represent the shape, the location, and the scale parameters, respectively.

From the average N dilution curve fitted, NNI values were calculated following **Equation 2.4** (Lemaire and Gastal, 1997):

$\mathrm{NNI} = \frac{\mathrm{Nc_{obs}}}{\mathrm{N_{curve}}}$ **(Equation 2.4)**

with the observed nitrogen as a percentage (N_{obs}, %) divided by the critical N content percentage (N_{curve}, %) of the average curve on the same aboveground biomass level.

Besides the calculation of the NNI based on the fitted curve, the NNI response to treatments by site was also analyzed following a hierarchical Bayesian framework. The top level of the model describes the NNI response ($y_{ij(s)}$) for the i^{th} block and j^{th} treatment nested within the s^{th} site **(Equation 2.5)**. The NNI response $y_{ij(s)}$ was assumed to follow a normal distribution with expected mean $\mu_{ij(s)}$, and variance σ_s^2:

$y_{ij(s)} \sim N(\mu_{ij(s)},\ \sigma_s^2)$, **(Equation 2.5)**

$i = 1, 2, \ldots, 5,$

$j = 1, 2, 3, 4, 5,$

$s = 1, 2, \ldots, 25.$

For the secondary level that describes the underlying process model, μ_{js} was defined as a linear effect model, were B_1 is the linear parameter for each of the 25 sites, and X_i the observed NNI for j[th] treatment. Informative and uniformly distributed sample prior was defined to B_1 with a minimum value of 0 and a maximum of 2, allowing the model to cover all the possible variability of NNI by the studied site. For σ_s was assigned an uninformative *Half-student-t (*υ , θ, σ) prior, where υ , θ, σ represent the shape, the location, and the scale parameters, respectively.

$$\mu_{ij(s)} = \begin{cases} B_{1(1)}.X_{ij(1)} \\ \vdots \\ B_{1(26)}.X_{ij(25)} \end{cases},$$

$$\sigma_s^2 = \begin{cases} \sigma_1^2 \, for \; site \; 1 \\ \vdots \\ \sigma_{26}^2 \, for \; site \; 26 \end{cases},$$

$$B_1 \sim Uniform(0,2)$$

$$\sigma_s \sim Half - student - t \; (3, 0, 2.5).$$

2.2.5.3. Nitrogen derived from the atmosphere

A hierarchical Bayesian framework was used to explore the response of soybean Ndfa to treatments by site. The Ndfa ($y_{ij(s)}$) for the i^{th} block within the j^{th} treatment, which is nested within the s^{th} site, is described at the model's top level (data layer) **(Equation 2.6)**. The first layer of the model was defined by a parametrization of the beta distribution (Lacasa et al., 2021) with an expected value $\mu_{ij(s)}$, and dispersion parameter δ_s:

$$y_{ij(s)} \sim Beta(\mu_{ij(s)}, \delta_s), \qquad \textbf{(Equation 2.6)}$$

$$i = 1, 2, \dots, 5 \, ,$$

$$j = 1, 2, 3, 4, 5,$$

$$s = 1, 2, \dots, 25.$$

For the secondary level, which represents the underlying process model, $\mu_{ij(s)}$ was specified as a linear effect model, where B_1 is the linear parameter for each site and $X_{ij(s)}$ is the observed Ndfa for the i^{th} block and j^{th} treatment at the s^{th} site. To B_1 was defined as an informative and normally distributed sample prior with an expected mean of 0.6 (60%) and a standard deviation of 0.3 (30%). For k_s was assigned an uninformative *Half-student-t* (υ , θ, σ) prior, where υ , θ, σ represent the shape, the location, and the scale parameters, respectively.

$$\mu_{ij(s)} = \begin{cases} B_{1(1)} . X_{ij(1)} \\ \vdots \\ B_{1(25)} . X_{ij(25)} \end{cases},$$

$$\delta_s = \begin{cases} \delta_1 \ for \ site \ 1 \\ \vdots \\ \delta_{25} \ for \ site \ 25 \end{cases},$$

$$B_1 \sim Normal(0.6, 0.3)$$

$$k_s \sim Half - student - t \ (3, 0, 2.5).$$

2.2.5.4. Model-based visualization estimation

The 95% credible intervals of posterior distributions of means at the treatment level were utilized to visualize model-based estimations for seed yield, NNI, and Ndfa analysis. For the comparison of treatments within each site, the 95% credible intervals were estimated from posterior distributions of the differences between pairs of treatments (9 comparisons), and subsequently tested to be different from zero (Supplementary Figures 2, 3, and 4). The magnitude of the effect size and uncertainty of each treatment comparison within each site was assessed from the estimate of the posterior distribution, obtained from the 95% credible intervals and their interquartile.

Data wrangling was performed using the *tidyverse* meta-package (Wickham et al., 2019), and data visualization were produced using the *ggplot2* package (Wickham, 2016). All the analyses were conducted using R software (R Core Team, 2020). The Bayesian analysis was performed with the *brms* package (Bürkner, 2017), which applies Bayesian regression models using stan language (Stan Team, 2022). For all evaluations, 5000 posterior samples were drawn as a warm-up, followed by 20,000 iterations with a thinning rate of 10 to minimize correlation between subsequent posterior samples and enhance four chain mixing (Hooten and Hefley, 2019).

2.2.6. Nutritional gap classification

Sites were split into **responsive and non-responsive,** accounting for the response in seed yield for the "Full" relative to the "Check", herein termed the seed yield nutrition gap. A principal component analysis was performed, and groups were distinguished by soil-sulfur-S availability at planting in the first dimension and environmental yield potential in the second **(Figure A.1)**. A total of 8 sites were identified as **responsive**. The proposed classification on yield responsiveness was used to study the response to the other factors, NNI and Ndfa.

2.3. Results

2.3.1 Seed yield

Average seed yield across sites ranged from 2.0 to 5.5 Mg ha^{-1}, with a mean of 4.0 Mg ha^{-1} with first and third interquartile ranges (IQR_{25-75}) of 3.4 to 4.8 Mg ha^{-1} **(Figure 2.2)**. Posterior distributions obtained from treatment pairwise comparisons were evaluated, and considering the 95% credible interval, 8 out of 26 sites (IA1, IA2, IN1, IN2, IN3, MN1, ND2, and WI1) resulted in a positive response for the comparison between treatments "Full" and "Check", which is associated to the nutritional gap. The remaining 18 sites revealed non-responsive behavior for the same treatment comparison. From a combined response to the application of both N+S, 7 out of 8 responsive sites reported a positive yield response (IA2, IN1, IN2, IN3, MN1, ND2, and WI1) while IA1 to S only **(Figure 2.2; Figure A.2).**

2.3.2 Crop nitrogen status

The final model of the N dilution curve was described as $\mathrm{Nc} = 3.91\ \mathrm{W}^{-0.09}$. Observations were proportionally distributed along the sufficiency curve, with different variability between the two response groups. The non-responsive sites reported a greater median Nc of 3.5% ($IQR_{25\text{-}75}$ = 2.97 to 4.04%) compared to the responsive sites, which documented median Nc of 3.4% ($IQR_{25\text{-}75}$ = 2.92 to 3.79%). The observed plant Nc averaged 3.45% ($IQR_{25\text{-}75}$ = 2.95 to 3.94%) across sites and treatments through R2 and R4 stages. Non-responsive sites reflected 50% of calculated NNI over a median value of 1.0 ($IQR_{25\text{-}75}$ = 0.85 to 1.13). Responsive sites documented 43% of the NNI above the sufficiency level, resulting in a median of 0.96 ($IQR_{25\text{-}75}$ = 0.84 to 1.08) (**Figure 2.3A**).

Non-responsive sites showed similar NNI for treatments "Check" ($IQR_{25\text{-}75}$ = 0.84 to 1.08), "N" ($IQR_{25\text{-}75}$ = 0.86 to 1.12), "S" ($IQR_{25\text{-}75}$ = 0.83 to 1.09), and "NS" ($IQR_{25\text{-}75}$ = 0.84 to 1.09) with roughly 46% of the NNI values at a sufficiency level. With a median NNI of 1.08 ($IQR_{25\text{-}75}$ = 0.98 to 1.23), the treatment "Full" resulted in the highest N status for non-responsive sites, with 71% of the NNI over 1. On the other hand, responsive sites revealed contrasting plant N status. The treatments "Check" and "N" (without sulfur) resulted in 31% of the NNI at the sufficiency level, with median NNI values of 0.89 ($IQR_{25\text{-}75}$ = 0.78 to 1.07) and 0.91 ($IQR_{25\text{-}75}$ = 0.79 to 1.03). Next, treatments "S" and "NS" resulted in similar 45% and 47%, respectively, of the NNI above 1, with median values of 0.97 and 0.96 with the same variation ($IQR_{25\text{-}75}$ = 0.86 to 1.07). Lastly, the treatment Full had 63% of the NNI above 1, with a median NNI of 1.09 ($IQR_{25\text{-}75}$ = 0.97 to 1.21). Thus, lower NNI values at responsive sites provided a signal for potential need for improvement of plant N nutrition and seed yield response to S or NS fertilization **(Figure 2.3B-C; Figure A.3**).

2.3.3 Ndfa contribution

The median Ndfa values ranged from 2 to 89% across sites and treatments, with a median of 44% (IQR_{25-75} = 35 to 52%). Treatments comparisons **(Figure A.4)** revealed that only 15 out of the 73 significant differences in Ndfa were not related to the effect of the Full treatment (336 kg fertilizer N ha^{-1}), occurring in seven sites (AR1, KS2, KS4, MN1, MN2, NC2, and VA1).

When comparing Ndfa of “Check” and “Full” in non-responsive sites, relatively similar soil and N fixation contributions were observed at R4, with an overall median N fixation of 41% (IQR_{25-75} = 28 to 48%) for the 'Check” and 40% (IQR_{25-75} = 25 to 59%) for the “Full”. Meanwhile, the difference in N fixation between “Check” and “Full” treatments was slightly larger on responsive sites, with a median N fixation of 35% (IQR_{25-75} = 30 to 55%) for the “Check” and 44% (IQR_{25-75} = 39 to 61%) for “Full”. Regardless of the response to fertilization in seed yield, contributions of N fixation and soil for plant N demand were similar within site. However, the variation across sites was high, with N fixation contribution varying from 10 to 89% for the “Check” and from 4 to 86% for the “Full” treatments **(Figure 2.4).**

2.3.4 Effect size and uncertainty of responses

For non-responsive sites, fertilization resulted in seed yield response between -0.93 and 0.83 Mg ha^{-1}, but with a greater magnitude on their uncertainty that ranged from 0.09 to 2.62 Mg ha^{-1} **(Figure 2.5A)**. In parallel, yield responsive sites presented yield response ranging from -0.42 to 1.1 Mg ha^{-1} with a reduced level of uncertainty, from 0.47 to 1.36 Mg ha^{-1}. Effect size ranged from 0.45 to 1.1 Mg ha^{-1}, while uncertainty varied from 0.49 and 1.34 Mg ha^{-1}. On the other hand, the largest uncertainty over the response was observed in the comparison between “NS” and “S” (N effect), with effect size ranging between -0.18 and 0.26 Mg ha^{-1} and uncertainty from 0.48 to 1.3 Mg ha^{-1}. Given the considerable magnitude of the nutritional gap on responsive sites, response to fertilization strongly correlated to S **(Figure 2.5B)**, which resulted in seed yield response ranging from 0.12 to 0.9 Mg ha^{-1}, whereas N resulted in a small (and erratic) yield response from 0.05 to 0.26 Mg ha^{-1}.

Regarding the NNI, non-responsive sites resulted in greater differences ranging between -0.2 and 0.37. In responsive sites, NNI differences ranged between -0.09 and 0.29. Uncertainties were consistent for non-responsive and responsive sites, ranging from 0.13 to 0.7 and from 0.13 to 0.31, respectively **(Figure 2.5C).** It is noteworthy the positive correlation between the NNI response to S fertilizer (NS - N) and the seed yield responses to S fertilization **(Figure 2.5D)**. In agreement with the previously reported seed yield gap, S application resulted in a favorable NNI increase in 7 of the 8 responsive sites between 0.02 and 0.17. At the same time, the response to N ("NS - S") was less meaningful, with only 4 sites showing positive but small responses between 0.007 and 0.1 for the NNI variable.

Responses in terms of Ndfa were generally connected to a contrast involving the “Full” treatment. Nonetheless, these responses were highly variable and uncertain, resulting in effect size ranging between -53% to 58% and –43% to 34% and uncertainties from 23% to 75% and 7% to 41% in non-responsive and yield responsive sites, respectively. Conversely, comparing among “Check”, “N”, “S”, and “NS” treatments, the average response resulted negative, ranging between -32% to 0% and –7% to 6% for non-responsive and responsive sites, respectively. Large uncertainties were associated with N (“NS – S”) and S (“NS – N”) responses, ranging from 7 to 45% **(Figure 2.5E)**. At responsive sites, as expected, there was no correlation between Ndfa and seed yield responses to N or S **(Figure 2.5F)**. The Ndfa response to N ("NS – S") varied from -7% to 1%, and similarly, the response to S ("NS – N") ranged between -11% and 7%.

2.4. Discussion

To the extent of our knowledge, this multi-state study is the first attempt in the US to determine plant N sources (soil N and N fixation) to diagnose yield response to N-S fertilization under a uniform protocol. In general, N fertilization had no effect on soybean yields, regardless of rate and/or time during the season. This research remarks on the challenge of accurately identifying responses to N fertilization due to improbable and uncertain seed yield, NNI, and Ndfa responses. In this regard, no previous literature has reported the magnitude of uncertainty for soybean responses to fertilization linked to seed yield, NNI, or Ndfa. A similar concept was stated for N management in corn (*Zea mays* L.), developing response models that account for uncertainties (Correndo et al., 2021b) to produce more informed fertilizer decisions and contribute with prior information to guide future research studies.

This work suggests that the seed yield nutritional gap is linked to S responsiveness rather than N. Although S deficiency is yet difficult to diagnose via soil testing (Mahal et al., 2022), specific conditions such as low organic matter and low soil sulfate-S levels are likely to reduce the S availability to the crop and increase the probability of response to S fertilizer (Borja Reis et al., 2021; Casteel et al., 2019; Divito et al., 2015). Most of the yield response to S was connected to an increase in NNI, depicting the intertwined link between these nutrients and the impact of S on crop N status, which may have a positive impact on yields (Brooks et al., 2022; Salvagiotti et al., 2009). The NNI may be an effective instrument for diagnosing not only N but also NS colimitation at the site level (Divito et al., 2016a).

On the other hand, increases in seed yield due to N fertilization are rare and inconsistent (Córdova and Licht, 2020; Mourtzinis et al., 2018; Osborne and Riedell, 2006), and commonly related to high-yielding potential fields where soil N and N fixation are insufficient to meet crop N demand (Cafaro La Menza et al., 2020, 2017; Ciampitti and Salvagiotti, 2018). The pursuit of more sustainable agriculture and greater efficiency in fertilizer utilization brought the development of better plant diagnostic tools, such as the NNI, to better comprehend the balance between crop N supply and demand (Lemaire and Ciampitti, 2020). The proposed average N dilution curve for soybeans in this study is an attempt to identify the N deficiency in soybeans across the US, which is very unlikely to occur. In agreement with our results, previous studies have shown that soybeans relying more on N fixation can still attain high seed yields without incurring significant penalties, compensating due to low soil N contribution (Santachiara et al., 2018). A few exceptions to this general rule were reported by a review study (Ciampitti and Salvagiotti, 2018), highlighting a potential N gap for N fixation to satisfy crop N demand for high yielding environments (>7 Mg ha^{-1}), situation that was not reported in this study. However, if N fixation is impacted due to inadequate nodulation, biotic/abiotic stressors, or another nutrient limitation such as S, then potential response to external N application might occur due to a potential reduction in crop N status (Divito et al., 2016b).

Considering fertilization practices potentially adopted by farmers (discarding the "Full" treatment due to economical unsuitability), the observed median N fixation contribution of 40% to crop N demand resulted lower than the 60% previously reported for the US (Herridge et al., 2008), although with a considerable variation range (IQR_{25-75} = 28 - 44%). Spatio-temporal variation in the soil-plant system may be the most critical variable controlling N fixation levels in the U.S. agro-ecosystems (Purcell et al., 2004; Salvagiotti et al., 2008); hence, the highest yield levels in our study were not necessarily associated with the largest N fixation contributions. Looking ahead to an improved understanding of the processes dictating fluctuations in N fixation contribution, the most likely candidates to aid in explaining the uncertainty surrounding soybean N sources, N status, and Ndfa are soil and meteorological variables (Ciampitti et al., 2021; Collino et al., 2015).

Future research should further investigate sources of spatio-temporal variability of crop yield (and other traits) responses to management practices that assist farmers in evaluating the risks for more informed decision-making. As this research shows, the estimated uncertainties of the effect sizes of fertilization on the variables under study resulted considerably high to overlook. In addition, to produce more accurate assessments of soybean nutritional status, breeding techniques must invest in non-N-fixing varieties and developing methodologies to provide controlled conditions to separate management effects on crop growth and N fixation. Investigating potential N-S co-limitation (including plant %S/%N ratio, and calculation of S nutrition index), due to their differences in nutrient dilution curves and allometry association with biomass, is a critical future research step to assist on providing better interpretation of this nutrient interaction on the main plant traits (e.g., yield, N fixation) and to develop more holistic (multi-nutrient) diagnostic approaches.

2.5. Conclusion

This research revealed that N fertilization is unlikely to improve soybean yields due to remarkably high uncertainty on treatment response, which may lead to unprofitable and not sustainable recommendations. On the other hand, S deficiencies are worth further exploring under site-specific conditions. Although challenging, using diagnostic methods such as the NNI in soybean to detect N and S deficits may be feasible but still needs further research development. Prospective studies should explore the uncertainties around fertilization effects and integrate weather and soil factors to obtain a more holistic understanding of the main drivers for plant N sources and yield formation, as well as how changes in partial N budget might impact seed composition, mainly crop protein levels. Finally, future research endeavors using Bayesian analysis could be benefited from the provided estimations of effect sizes and uncertainty magnitudes of seed yield, NNI, and Ndfa responses to fertilization.

Table 2.1. Fertilizer nutrient rate (kg ha^{-1}) and their respective fertilizer sources applied to the different treatments during the 2021 growing season.

Treatment	Timing				Total Applied	
	Planting		R3 Stage			
	N	S	N	S	N	S
	kg ha^{-1}		*kg ha^{-1}*		*kg ha^{-1}*	
Check	-	-	-	-	-	-
N *	30	-	-	-	30	-
S $^{+}$	-	34	-	-	-	34
NS ^	30	34	-	-	30	34
Full $^{\pm}$	168	17	168	17	336	34

Symbols indicate fertilizer source: * (Urea, 46-0-0); $^{+}$ (Gypsum, 0-0-0-23.5S); ^ (Ammonium Sulfate, 21-0-0-24S); and $^{\pm}$ (Urea and Ammonium Sulfate).

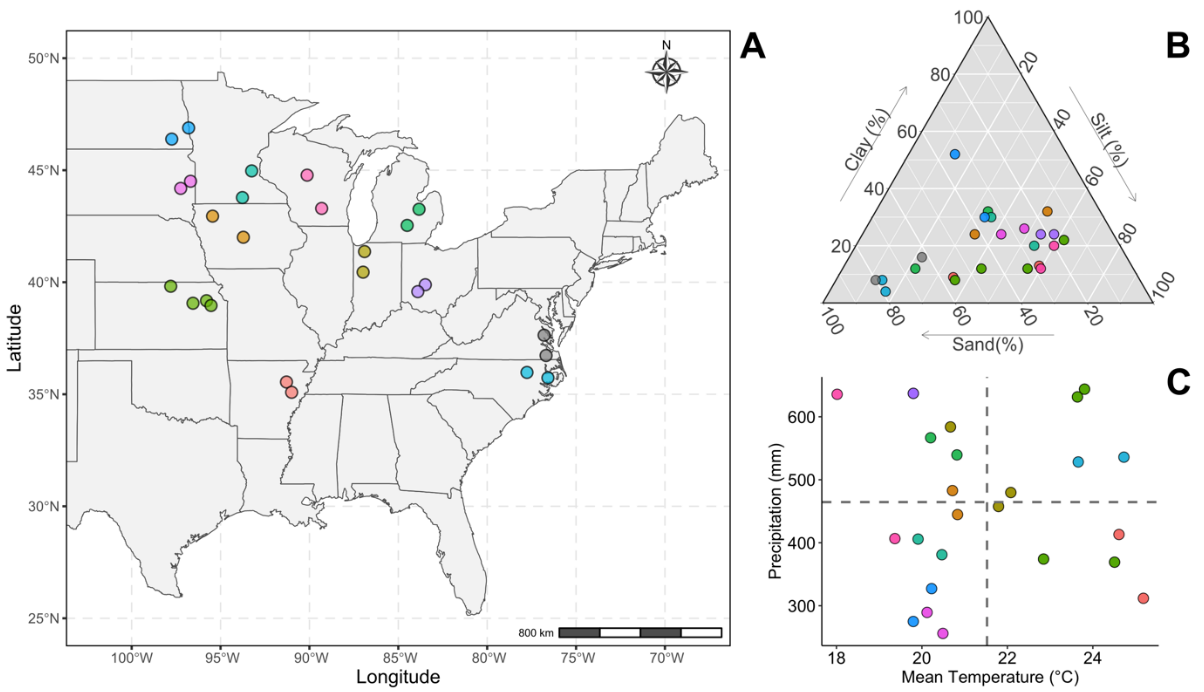


Figure 2.1. (A) Geographical distribution of field trials during the 2021 growing season (26 sites); (B): soil texture variability (0 to 0.15 m); and (C): distribution of mean temperature (°C) and precipitation (mm) during the crop development period, with dashed lines representing overall mean.

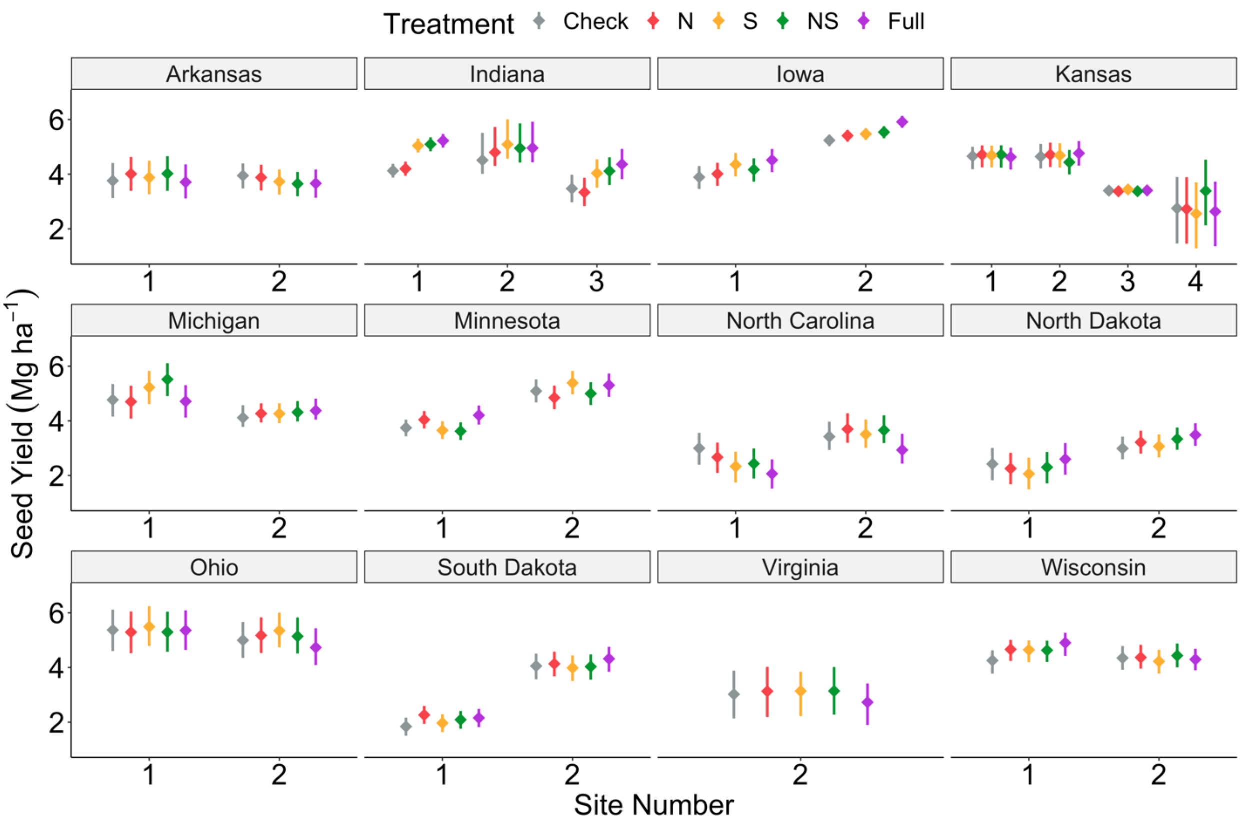


Figure 2.2. Fertilization treatments effect on seed yield by site. The error bars indicate a 95% credible interval for each of the treatments within the 26 sites covering a total of twelve states of the United States (US).

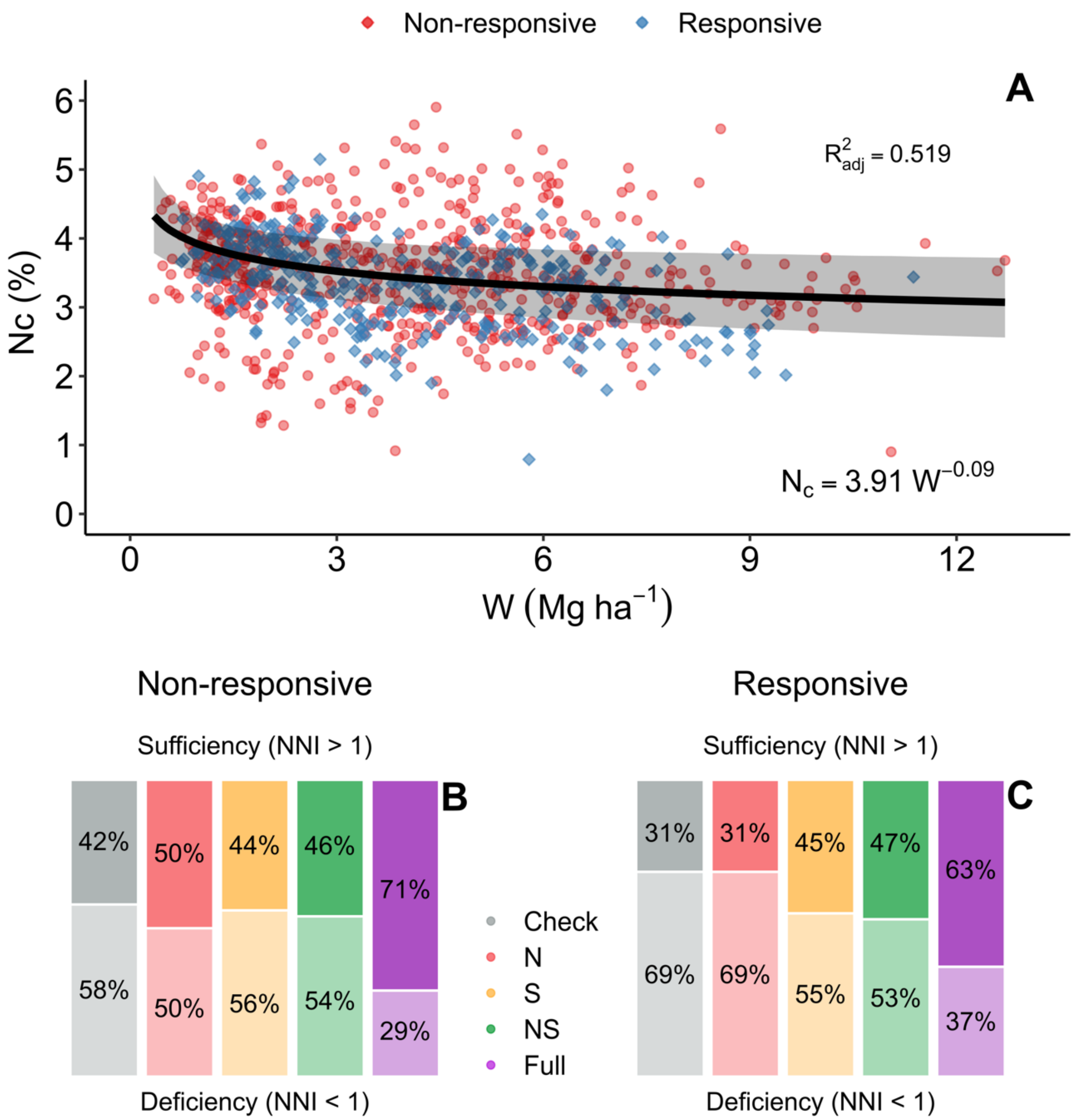


Figure 2.3. Top panel (A) displays the average plant nitrogen (N) dilution curve for all observations (within sites, treatments, and stages), represented by the relationship of Nc (%) and dry biomass (W, Mg ha^{-1}). The shaded area represents the overall 95% credible interval. Bottom panels (B-C) show the proportion of NNI observations above and below the average sufficiency curve by treatment at responsive and non-responsive sites throughout the season.

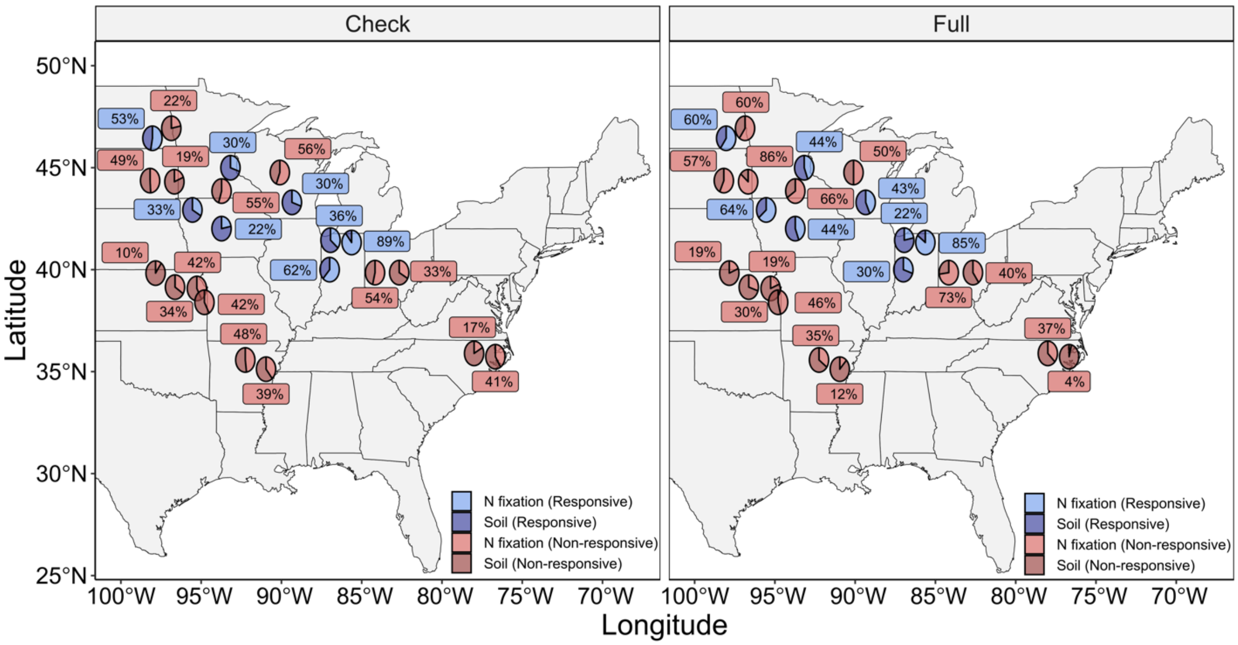


Figure 2.4. Contribution of soil N and N fixation to the total soybean N demand at the R4 stage (full pod formation) across 12 states of the United States, 26 studies. The proportion (%) shown inside the labels refers to the contribution of N fixation to the total soybean plant N demand.

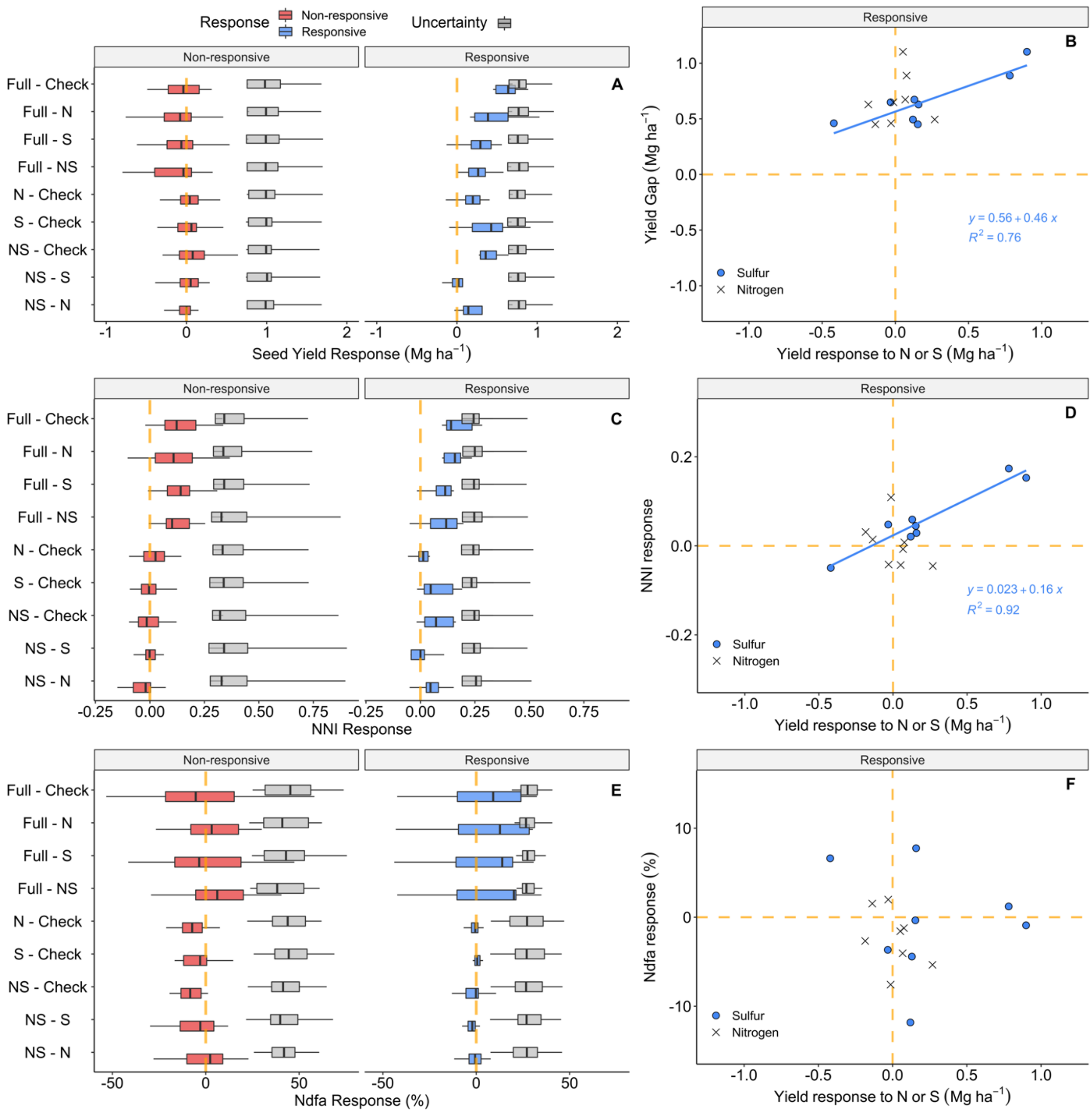


Figure 2.5. Variability across sites of effect size and uncertainty on yield, nitrogen nutritional index (NNI), and nitrogen derived from the atmosphere (Ndfa) in response to N-S fertilization strategies. Panels A, C, and E show the average effect size and uncertainty of the treatment comparisons of all sites, for non-responsive and responsive sites, assessed from the seed yield, NNI, and Ndfa posterior distributions model. The error bars indicate 95% credible intervals for each treatment comparison, and the orange dashed line is the magnitude of the effect size and uncertainty equal to zero. Panels B, D, and F show the magnitude of the seed yield, NNI, and Ndfa gap (calculated as the difference between Full – Check) and the relationship with N and S fertilization response at non-responsive and responsive sites; the vertical and horizontal dashed lines represent effect size = 0.

Chapter 3 - Assessing the influence of environmental drivers on soybean seed yield and nitrogen fixation estimates and uncertainties in the United States

Published in European Jornal of Agronomy

Almeida, L.F., Correndo, A.A., Hefley, T., Hintz, G., Prasad, P.V., Licht, M., Casteel, S., Singh, M., Naeve, S., Bais, J. and Lindsay, L., 2025. Assessing the influence of environmental drivers on soybean seed yield and nitrogen fixation estimates and uncertainties in the United States. European Journal of Agronomy, 162, p.127428. https://doi.org/10.1016/j.eja.2024.127428

Abstract

Soybean [*Glycine max* (L.) Merr.] is one of the major crops worldwide. Identification of environmental factors that improve both yield and N_2 fixation remain of high importance. The study aimed to i) assess the effect (estimate and uncertainty) of sulfur (S) fertilization on seed yield and N_2 fixation (as N derived from the atmosphere, Ndfa), and (ii) evaluate the influence of soil and weather variables on these estimates and uncertainties. Thirty-five studies from nine US states were analyzed, comparing no fertilization (Check) with S fertilization at planting (S), using a regression tree approach to assess environmental effects on yield and Ndfa. For both treatments, precipitation from full-pod to full-seed explained 40% of the yield variation. For the Check, [soil organic matter, SOM/(clay+silt)] was a secondary factor. For the S, seasonal precipitation above 73 mm resulted in the highest yield (4.9 Mg ha^{-1}), with 51% Ndfa and 135 kg ha^{-1} of fixed-N. Yield uncertainty, averaging 1.2 Mg ha^{-1}, was associated with soil clay content below 11%. Vapor-pressure-deficit from full-bloom to full-pod influenced Ndfa, accounting for 40% of its variation between treatments. For both treatments, the highest Ndfa (~65%) required vapor-pressure-deficit below 0.92 kPa. Soil clay was pivotal to the uncertainty in Ndfa,

explaining 34% and 40% of the variation for Check and S, but with a reduction in uncertainty when soil clay was above 26%.The main regulators of yield and Ndfa were precipitation, temperature, SOM, and soil texture. Sulfur fertilization moderately increased yield and Ndfa, especially in environments with high plant N-demand. Ndfa uncertainty was more related to crop growth factors, with high seed yield correlating with high Ndfa. Future research should focus on controlled studies to improve the knowledge of the identified soil and weather factors and their interplay with seed yield and Ndfa.

3.1 Introduction

Soybean [*Glycine Max* (L.) Merr.] is grown worldwide due to its high nutritional seed value. The United States (US) is a leading nation in soybean production, with an output of approximately 122 million Mg, ranking it second in the world behind Brazil, 163 million Mg, and Argentina, 48 million Mg (USDA, 2023). However, the expected annual increase in soybean yield in the US is approximately 1.2%, which is lower than the projected global rate of 2.4% required to meet anticipated global demand by 2050 (Ray et al., 2013). Therefore, it is still vital to enhance our understanding of yield-limiting factors to facilitate more effective adaptation of soybean production strategies.

Crop nutrition is a key factor influencing soybean productivity (Balboa et al., 2018). Although nitrogen (N) is the most limiting nutrient in soybean production (Sinclair and de Wit, 1975), the trade-offs between plant N sources (soil and atmosphere) present a major challenge for consistent responses to N fertilization (Ortez et al., 2018). Despite significant N requirements, numerous studies have demonstrated that N fertilization is not economically viable or environmentally desirable for soybeans (Salvagiotti et al., 2008), with a large proportion of this nutrient supplied by the process of N_2 fixation (Ciampitti and Salvagiotti, 2018). Although the effects of sulfur (S) fertilization are erratic, they have been shown to have recurrent positive effects, leading to greater seed yields (Almeida et al., 2023; de Borja Reis et al., 2021; Salvagiotti et al., 2009).

In terms of N nutrition, the biological N_2 fixation represents the most important process in soybean production, converting atmospheric nitrogen (N_2) to ammonia (NH_3) via a symbiotic association with soil bacteria. The reduction of N_2 is catalyzed by the nitrogenase enzyme, which is known for its high S content. A sufficient supply of S can enhance the contribution of N_2

fixation by improving nodulation and nitrogenase activity (Anderson and Spencer, 1950). As a source of S, gypsum ($CaSO_4 \cdot 2H_2O$) has been shown to improve water infiltration and reduce nutrient loss (Murphy and Stevens, 2010). Additionally, it has also been shown to act as a booster for N_2 fixation in soybean (Hussain et al., 2011), as well as other legumes, thereby contributing to reduced soil N depletion. Over 40% of the N requirements of soybeans are met via fixation during the growing season, with N_2 fixation activity peaking around the pod setting (Ciampitti et al., 2021b). The process of N_2 fixation exhibits a wide range of variation, and several factors, including the activity of *Rhizobia*, among others, can influence its contribution. Previous research has identified key environmental factors that affect N_2 fixation contribution to plant N demand such as soil variables including pH (Ferguson et al., 2013), phosphorus (Collino et al., 2015), S (de Borja Reis et al., 2021), soil organic matter (SOM) (Lawson et al., 1995), and soil texture (de Borja Reis et al., 2021), and weather variables including extreme temperatures (Lindemann and Ham, 1979), vapor pressure deficit, VPD (Ciampitti et al., 2021b; Monnens et al., 2023), and drought/water deficit (Purcell, 2009).

Generalized linear models (GLMs) combined with biological data are a powerful tool for assuming distributions that extend beyond the normal. This approach can yield reliable estimates, facilitate uncertainty quantification, and enable the formulation of statistical inferences (McCullagh and Nelder, 1989). The exploration of complex interactions among genotype, environmental, and management (G × E × M) variables often involves the use of classification and regression tree (CART) models (Mourtzinis et al., 2018). The combination of GLM and CART models can facilitate the identification of the factors influencing the estimation and uncertainty associated with soybean seed yield and nitrogen derived from the atmosphere (Ndfa). The utilization of CART models can contribute by segmenting data into distinct subsets through a data-driven process, thereby revealing distinct patterns. This integration of data analysis methods improves the understandability of the results, especially for those not well versed in statistical methods (Breiman et al., 1984).

The main objectives of this study were to: i) assess the effect of S fertilization on seed yield and N_2 fixation (as N derived from atmosphere; Ndfa), and (ii) evaluate the influence of soil and weather variables on both estimates and uncertainties of seed yield and N_2 fixation. This study provides new estimates and uncertainties on the effect of S fertilization management for both yield and N_2 fixation processes and identifies the main drivers linked to soil, plant, and weather factors in soybeans grown under field conditions.

3.2 Material and methods

This study extends the work previously published by Almeida et al. (2023), documenting the positive effects of S fertilization on yield, N_2 fixation at the full-pod phenological stage, and plant N status. This current work presents a new set of locations (~50% new locations) and focused on the main environmental drivers underpinning yield formation and N_2 fixation processes.

3.2.1. Field trials

During the 2021 and 2022 growing seasons, a total of 35 site-year studies was conducted in nine states in the USA: Iowa, Indiana, Kansas, Michigan, Minnesota, North Dakota, Ohio, South Dakota, and Wisconsin **(Table 3.1, Figure 3.1)**. Soils in the present study included 23 soil series, from fine to coarse textures.

The experiments followed a randomized complete block design (RCBD) with four to five replications, depending on the location. The plots presented dimensions between 12.2 and 15.2 m in length and 6.1 m in width. The spacing between soybean rows varied between 0.38 m and 0.76 m, and soybean varieties were defined according to the region. Pests, diseases, and weeds were managed in all studies. Prior to sowing, liquid inoculant (*Bradyrhizobium japonicum*) were applied to seeds. Across all the sites, two different strategies were assessed. The first, referred to as the "Check" treatment, did not involve any S fertilization. The second treatment, labeled as the "S" treatment, consisted of 34 kg S ha^{-1} applied at planting, using pelletized gypsum ($CaSO_4 \cdot 2H_2O$; Calcium Products, Ames, IA) as the source. The Check represents the standard practice in soybean production, while the S treatment was implemented based on prior research that identified S as a limiting nutrient for soybean yields (Almeida et al., 2023).

3.2.2. N_2 fixation and seed yield

The contribution of biological N_2 fixation was determined at the R6 phenological stage (full-seed) across all sites (Fehr et al., 1971). To estimate this variable, aboveground plant sample were collected from a 1.50 m length of row, representing an area of 1.16 m², and dried in a forced-air oven at 65 °C until reaching a constant weight for dry matter determination (Mg ha^{-1}). Then, sub-samples of 5 plants were finely ground (250 µm) and analyzed for %N and ^{15}N natural abundance contents. The %N and the isotopic ratio of ^{15}N, were measured using a PyroCube elemental analyzer (Elementar Americas, Ronkonkoma, NY), which was connected to a visION isotope ratio mass spectrometer (Elementar Americas). These analyses were conducted at the Stable Isotope Mass Spectrometry Laboratory at Kansas State University (Manhattan, KS). The estimation of Ndfa (%) was conducted using the ^{15}N natural abundance method, which involves measuring the excess of the ^{15}N isotope by comparing soybean plants to a reference non-nodulating plant. Adjacent to the soybean plots, a strip of unfertilized corn was planted, which served as the reference plant because it solely relied on N from the soil. Following Unkovich et al. (2008), Ndfa was estimated as follows:

$$Ndfa = \frac{15N\ of\ corn - 15N\ of\ soybean}{15N\ of\ corn - B} \times 100. \qquad \textbf{(Equation 3.1)}$$

Where Ndfa stands for the percentage of total N derived from the atmosphere, ^{15}N represents the abundance of the ^{15}N isotope on soybean and corn plants, and a *B* = -2.54‰ (Balboa and Ciampitti, 2020). The *B* value compensates for any isotopic bias during the allocation of symbiotically fixed N. The fixed-N (kg ha^{-1}) was estimated as the product of aboveground plant biomass and %N, and then multiplied by the Ndfa. After crop maturity, plots were mechanically harvested by collecting the central two rows using a combine. The harvested area ranged from 9.3 to 11.6 m^2. Seed yields were recorded and adjusted to a moisture basis of 130 g kg^{-1}.

3.2.3 Environmental covariates

Prior to planting, soil samples were collected at a depth of 0.15 m for soil for characterization purposes at the field level. This included analyzing soil texture, pH level, soil organic matter (SOM), cation exchange capacity (CEC), macronutrients, and the calculation of the SOM mineralization resilience (Hassink, 1994) as the ratio between SOM over clay and silt [SOM/(clay+silt)] **(Table 3.1 and Table B.1)**. Variables with missing values were excluded from the statistical analysis to avoid misinterpretation. Daily weather data including maximum and minimum temperatures (°C), precipitation and irrigation when available (mm), and solar radiation (MJ/m^2) were obtained from the Iowa Environmental Mesonet database (ISU, 2024) and processed in R software. Daily relative humidity data were gathered from *Nasapower* (Sparks, 2018) at a spatial resolution of 50 km^2, and along with temperature readings, used to calculate vapor pressure deficit (VPD, kPa) which is closely correlated to temperature and regulates plant transpiration rates (Sinclair et al., 2017). Additionally, growing degree days

(GDD) were calculated to measure the accumulated heat energy necessary for crop growth considering 10 °C as the base temperature. The seasonal weather was summarized into four different phenological periods: 20 days prior to planting, planting to R2 (full-bloom), R2 to R4 (full-pod), and R4 to R6 (full-seed). Precipitation and GDD were summarized cumulatively for each of the periods, and VPD was represented by the average values in a given period. In the statistical analysis, the set of covariates included accumulated precipitation for all the four periods, and accumulated GDD and average VPD for three periods **(Table 3.2)**, along with soil pH, SOM, soil texture, and the ratio of SOM/(clay+silt). It is noteworthy that the analysis included 35 sites, none of which presented missing data for the 16 evaluated covariates.

3.2.4. Statistical analysis

3.2.4.1. Treatment estimate and uncertainty: seed yield, Ndfa, and fixed-N

To assess the effects and uncertainties on seed yield of the two treatments (Check and S) at the plot level for each site, generalized linear model was employed. To accommodate the data distribution that ranges from near zero to potentially infinity, it was assumed a gamma distribution with the log link function. More specifically, the generalized linear model used was:

$$y_{ij(s)} \sim Gamma(\mu_{ij(s)}, \psi),$$

$$\log(\mu_{ij(s)}) = \boldsymbol{x}_{ij}\boldsymbol{\beta}_{i(s)}.$$

In the equations above, $y_{ij(s)}$ describes the seed yield (Mg ha^{-1}) response for the i^{th} treatment, nested within the j^{th} block and s^{th} site-year. $\boldsymbol{x}_{ij}$ represents the vector of predictor variables for each plot that is subjected to the i^{th} fertilization treatment within the j^{th} block. Specifically, this vector comprises variables that identify the fertilization treatment (either 'Check' or 'S') and each block. $\boldsymbol{\beta}_{i(s)}$ corresponds to the coefficient vector of the i^{th} treatment for the s^{th} site-year. The parameter ψ is the dispersion parameter of the gamma distribution where $\text{Var}(y_{ij(s)}) = \mu_{ij(s)}\psi^2$.

For Ndfa, a generalized linear model with a beta distribution and the logit link function were utilized. In the equations below, $y_{ij(s)}$ describes the Ndfa (%) response of the i^{th} treatment in the j^{th} block, nested within the s^{th} site-year. $\boldsymbol{x}_{ij}$ is the vector that comprises the predictor variables for every plot receiving the i^{th} fertilization treatment within the j^{th} block. This vector is composed of variables that distinctly categorize the treatment (with or without S) and block. $\boldsymbol{\beta}_{i(s)}$ corresponds to the coefficient vector of the i^{th} treatment within the s^{th} site-year. The models can be written as follows:

$$y_{ij(s)} \sim Beta\left(\mu_{ij(s)}, \phi\right),$$

$$\text{logit}\left(\mu_{ij(s)}\right) = \boldsymbol{x}_{ij}\boldsymbol{\beta}_{i(s)},$$

where ϕ is the dispersion parameter with $\text{Var}\left(y_{ij(s)}\right) = \mu_{ij(s)}\left(1 - \mu_{ij(s)}\right)/(1 + \phi)$.

Lastly, to evaluate the amount of fixed-N, a generalized linear model was also applied. To handle the data distribution from near zero to positive infinity, a gamma distribution with the log link function was adopted. In the equations below, $y_{j(s)}$ describes the fixed-N (kg ha^{-1}) response of the i^{th} treatment, within the j^{th} block nested within the s^{th} site-year. $\boldsymbol{x}_{ij}$ denotes the vector of predictors for each plot under the i^{th} treatment in the j^{th} block, encompassing variables for the treatment type ('Check' or 'S') and block identification. $\boldsymbol{\beta}_{i(s)}$ stands for the i^{th} treatment within the s^{th} site-year. The model can be written out as follows:

$$y_{ij(s)} \sim Gamma\left(\mu_{ij(s)}, \psi\right),$$

$$\log\left(\mu_{ij(s)}\right) = \boldsymbol{x}_{ij}\boldsymbol{\beta}_{i(s)}.$$

From the models, treatment effects were summarized for each response variable by calculating the expectation as the mean and computing the length of the 95% confidence intervals as the uncertainty measurement. Analyses were performed in R software (R Core Team, 2022) using the *mgcv* package (Wood, 2017).

3.2.4.2. Regression tree analysis

To explore the environmental factors that impact soybean seed yield and Ndfa along with their associated uncertainties in the two tested fertilizer managements (Check and S), a regression tree approach was employed. For each treatment, the average estimates (expected value) and confidence interval lengths (uncertainty) from previous GLM modeling were used as input data for a tree-building algorithm implemented with the *rpart* package (Therneau et al., 2023) in R. The complexity parameter was set at 0.1, indicating that a split would be considered only if it increased the model R^2 by a minimum of 10%. A minimum of two observations was set as a requirement for each terminal node. Finally, the maximum tree depth was limited to three to prevent excessive complexity and partitioning, prioritizing the identification of the most significant patterns in the data.

3.3 Results

3.2.1. Seed yield and N_2 fixation

Seed yield ranged from 1.84 to 5.95 Mg ha^{-1} for the Check and from 1.97 to 5.60 Mg ha^{-1} for the S treatment across all sites. The estimated uncertainty in seed yield ranged from 0.08 to 1.63 Mg ha^{-1} for the Check and up to 1.52 Mg ha^{-1} for the S treatment **(Figure 3.2A)**. The Ndfa exhibited a similar mean for both treatments (44%), with a range from 2 % to 88 %. The Ndfa uncertainty ranged from 4 to 58 % for the Check and between 4 and 47 % for the S, with a narrower range **(Figure 3.2B)**.

In terms of the amount of fixed-N, the Check treatment ranged from 8 to 220 kg ha^{-1} with an average of 94 kg ha^{-1}, and uncertainties from 10 to 212 kg ha^{-1}. The S fertilization resulted in an average of 107 kg ha^{-1} of fixed-N, and with a greater degree of variation for both estimates, ranging from 3 to 304 kg ha^{-1}, and uncertainties, from 12 to 254 kg ha^{-1} **(Figure 3.2C)**.

3.2.2. Environmental covariates

The regression tree for expected seed yield **(Figure 3.3)** accounted for 67 % of the variation in the Check and 59 % for the S, while for yield uncertainty **(Figure 3.4)** around 52 % of the variation was explained for both treatments. For expected Ndfa **(Figure 3.5)**, the environmental variables explained 65 % of variation for the Check and 63 % for the S. For the Ndfa uncertainty **(Figure 6)**, covariates accounted for 52 % of variation for the Check and 61 % for the S.

For both treatments, precipitation was identified as the primary factor governing seed yield differences. Specifically, precipitation from R2 to R4 and from planting to R2 explained 38 % and 18 % of the variation for the Check **(Figure 3.3A)**, and 39 % and 19 % for the S **(Figure 3.3D)**, respectively. For S, in addition to precipitation, SOM/(clay+silt) accounted for an extra 11 % of the yield variation **(Figure 3.3A)**. Elevated yields were associated with higher Ndfa and fixed-N.

Precipitation levels below 75 mm for Check and below 73 mm for S from R4 to R6, associated with less than 123 mm from planting to R2, resulted in the lowest average yields of 3.0 Mg ha^{-1} for Check and 3.1 Mg ha^{-1} for S (Node 3) **(Figures 3.3A-D)**. Hence, for both Check and S, precipitation from planting to R2 exceeding 123 mm (Node 4) resulted in yields that were 36 % greater, of average 4.1 Mg ha^{-1} and 4.25 Mg ha^{-1}, respectively. Along with the increases in yield, the average Ndfa contribution rose from 26% to 47% for the Check **(Figure 3.3B)** and from 24 % to 45 % for the S **(Figure 3.3E).** In the Check, precipitation from R4 to R6 above 75 mm and SOM/(clay+silt) ratios below 7 resulted in the highest yields (Node 7), with an average of 4.95 Mg ha^{-1}. For SOM/(clay+silt) values of the ratio greater than 7 (Node 6) were linked to yields 25 % lower (3.7 Mg ha^{-1}). The higher-yielding sites (Node 7) reported an average Ndfa of

47 % and fixed-N of 110 kg ha^{-1} **(Figures 3.3B-C)**. For the S, precipitation from R4 to R6 above 73 mm (Node 5) led to the highest average yields of 4.94 Mg ha^{-1}, along with 51 % Ndfa and 135 kg ha^{-1} of fixed-N **(Figures 3.3E-F)**.

For both Check and S, the main factor driving seed yield uncertainty was the clay content in the soil, which explained about 31 % of the variation **(Figures 3.4A-E)**. Additionally, the accumulated GDD from planting to R2 **(Figure 3.4A)**, and from R2 to R4 **(Figure 3.4E)**, accounted for another 21 % of the variation in yield uncertainty for S and Check, respectively. The soil clay content and accumulated growing degree days (GDD) yielded an equivalent grouping effect on sites for both treatments, resulting in a comparable magnitude of yield uncertainty across terminal nodes **(Figures 3.4A-E)**.

For instance, the highest seed yield uncertainty for both Check and S of average 1.2 Mg ha^{-1} (Node 5), resulted from soils with lower clay content (< 11 %) **(Figure 3.4 A-E)**. This high uncertainty magnitude was linked to the lowest average seed yield of 3.29 Mg ha^{-1} for Check and 3.43 Mg ha^{-1} for S. Sites with the highest yield uncertainty (Node 5) for both treatments showed the largest average Ndfa contribution of 57 % **(Figures 3.4 C-G)**. However, Check resulted 110 kg ha^{-1} of fixed-N, compared to 156 kg ha^{-1} for S **(Figures 3.4 D-H)**. For both treatments, a clay content above 11 %, combined with different thresholds of GDD, resulted in the lowest yield uncertainty of average 0.5 Mg ha^{-1}. More specifically, for the Check, this was associated with GDD from planting to R2 below 1907 **(Figure 3.4A)**, and for the S, with GDD from R2 to R4 below 2390 **(Figure 3.4E)**. The lowest yield uncertainty (Node 3) was associated with the lowest Ndfa contributions of 40 % **(Figures 3.4C-G)** and the lowest average fixed nitrogen of 90 kg ha^{-1} for Check and 98 kg ha^{-1} for S **(Figures 3.4D-H)**.

In the analysis of the expected Ndfa, VPD from R2 to R4 accounted for 41 % of the variation in both the Check and S **(Figure 3.5A-D)**. For the Check treatment, an additional 24 % of the variation in Ndfa was explained by precipitation prior to planting and from planting to R2 **(Figure 3.5A)**. In contrast, for the S, soil sand content and precipitation from R2 to R4 accounted for an additional 22 % of the variation **(Figure 3.5D)**. In the case of S, increases in Ndfa were associated with corresponding increases in both yield and fixed-N **(Figure 3.5D-E-F)**.

In the Check, the highest average Ndfa of 67 % (Node 7) was observed at sites with an average VPD between planting and R2 below 0.92 kPa, and precipitation from planting to R2 exceeding 282 mm **(Figure 3.5A)**. This level of Ndfa was, on average, 18 % higher compared to sites where precipitation from planting to R2 was below 282 mm (Node 6). Notwithstanding the discrepancies in Ndfa, a comparable mean seed yield of 4.6 Mg ha^{-1} was observed in sites with VPD below 0.92 kPa determined for the crop period from R2 to R4 **(Figure 3.5B)**. Additionally, sites with precipitation from planting to R2 below 282 mm resulted in an average fixed-N of 127 kg ha^{-1} (Node 6), which was 10 % higher compared to sites with higher precipitation **(Figure 3.5C)**.

For the S, the highest Ndfa contribution of average 62 % (Node 7) resulted from sites with a VPD below 0.92 kPa and precipitation over 76 mm from R2 to R4 **(Figure 3.5D)**, along with an average yield of 5.15 Mg ha^{-1} and a fixed-N amount of 165 kg ha^{-1} **(Figures 3.5E-F)**. When precipitation declined to less than 76 mm during the same period (Node 6), the average Ndfa decreased to 47 %, as well as the yield (4.6 Mg ha^{-1} and a fixed-N of 127 kg ha^{-1}) **(Figures 3.5E-F)**.

In the Ndfa uncertainty regression tree, soil clay content accounted for 34 % of the variation for the Check and 40 % for the S **(Figures 3.6A-E)**. For the Check, precipitation prior to planting explained an additional 18 % of the variation **(Figure 3.6A)**. Meanwhile, for the S, VPD from planting to R2 explained an additional 21 % of the variation in Ndfa uncertainty **(Figure 3.6E)**. In general, the magnitude of the uncertainty for the Ndfa showed a negative correlation with both estimates for the Ndfa and fixed-N.

For the Check, sites with clay content below 13 % and precipitation prior to planting under 96 mm (Node 5) **(Figure 3.6A)** resulted in the lowest Ndfa uncertainty, averaging 10 %, along with an average yield of 4.05 Mg ha^{-1}, and average Ndfa and fixed-N amounts of 47% and 104 kg ha^{-1}, respectively **(Figures 3.6B-D-C)**. Similarly, for the S, an average Ndfa uncertainty of 10 % corresponded to clay content below 19 % and VPD from planting to R2 under 0.9 kPa (Node 5) **(Figure 3.6E)**, along with an average yield of 4.33 Mg ha^{-1}, and average Ndfa and fixed-N amounts of 50 % and 120 kg ha^{-1}, respectively **(Figures 3.6F-G-H)**. When clay content was found over 26 % (Node 2) for both Check and S, the average Ndfa uncertainty was 12 %.

The greatest uncertainty was observed for both the Check and the S treatment when the clay content was below 26% and when specific environmental conditions were present. For the Check, this condition was characterized by precipitation exceeding 96 mm before planting (Node 7), while for the S treatment, it was marked by a VPD from planting to R2 above 0.9 kPa (Node 7). In the case of Check, the highest average Ndfa uncertainty, at 40 %, was associated with the lowest average fixed-N of 74 kg ha^{-1}, an Ndfa of 40 %, and a yield of 3.67 Mg ha^{-1} **(Figures 3.6B-C-D)**. Similarly, for S, the greatest average Ndfa uncertainty, at 33 %, corresponded to the lowest averages of fixed-N at 60 kg ha^{-1}, Ndfa contributions at 32 %, and yields of 3.85 Mg ha^{-1} **(Figures 3.6F-G-H)**.

3.3. Discussion

This study offers a comprehensive assessment that integrates model-based estimates and regression tree models to untangle the influence of soil and weather factors on seed yield, Ndfa, and their associated uncertainties for soybean crops. The primary regulators of both seed yield and Ndfa estimates and uncertainties were linked to weather and soil. It is notable that S, which is often overlooked, has emerged as a significant factor in mitigating soil-related seed yield limitations (de Borja Reis et al., 2021). By partitioning the sites into more homogeneous and representative groups, this study reinforces the evidence of S co-limitation (Cossani and Sadras, 2018) affecting the overall plant N status, N_2 fixation, and their respective uncertainty. The average contribution of N_2 fixation (45 %) was lower compared to recent published literature in the US (Ciampitti and Salvagiotti, 2018; Peoples et al., 2021). It is noteworthy that the lowest Ndfa values (less than 15%) observed in this study were attributed to environmental conditions that were detrimental to the establishment and effectiveness of N_2 fixation and crop growth. This serves to illustrate the environmental dependencies inherent to N_2 fixation process.

Under conditions of reduced crop growth stress, soybean yields correlated to increased Ndfa and superior fixed-N uptake. Similar to Ciampitti et al., (2021b), soybeans were affected by the VPD as a critical variable (Monnens et al., 2023) modulating fixed-N amounts, primarily via crop growth. This effect is likely due to a pronounced increase in transpiration rates under high temperatures (de Borja Reis et al., 2021; Sinclair et al., 1987) and lower air relative humidity. It is noteworthy that the interaction between SOM and soil texture emerged as a significant factor influencing yields in the absence of S fertilization. Furthermore, soil texture serves as an important regulator of Ndfa and its associated uncertainty. A potential hypothesis to explain this interaction, and the influence of clay content is that it relates to the stabilization of SOM in fine-textured soils (Hassink, 1994), or to the impediments imposed by suboptimal soil root oxygenation, which adversely affects the process of N_2 fixation (Linn and Doran, 1984). This stabilization process can result in reduced availability of both N and S (Castellano et al., 2012), which in turn affects the dynamics of soil NO_3-N and the contribution of Ndfa. These findings emphasize the interplay of environmental factors and crop growth influencing the Ndfa. Notably, when supplied with S, sites with greater Ndfa showed improved yields and fixed-N amounts, supporting the theory of N and S co-limitation (Almeida et al., 2023; Salvagiotti et al., 2009).

In general, sites with a higher sand content have a lower water-holding capacity and, which renders them more susceptible to physiological stresses. Furthermore, these stresses can be further exacerbated by high temperatures during the growing season (Doss et al., 1974). In specific situations, extreme precipitation events prior to planting may delay planting date from optimal as well as impair plant establishment, impacting early crop growth and Ndfa potential (Ciampitti et al., 2021a). Environments with larger seed yield uncertainty suggest that Ndfa significantly contributes to the variability of resource allocation, ultimately impacting yields (Tamagno et al., 2018). Soils with reduced clay content correlated with increased yield uncertainty and a tendency for reduced yields (Collino et al., 2015). This relationship may be attributed to indirect interactions involving changes in soil N availability due to SOM protection and weather.

Although the present study has elucidated a significant portion of the variance with a select number of covariates, the research scope did not examine the influence of soybean genotypes and planting dates within an environment (Purcell et al., 1997), as relevant factors impacting both crop growth and the potential for Ndfa. The examination of these variables has the potential to elucidate effective strategies for overcoming co-limitations, such as those related to nitrogen and water availability, a concept supported by the findings of Sadras et al. (2016).

Investigating potential co-limitations is crucial for advancing multi-nutrient diagnosis toward more efficient and sustainable soybean production. In addition, exploring a range of S fertilizer application rates could enhance our understanding of varying levels of limitation or co-limitation across different environments. Future research should prioritize controlled studies to examine weather and soil factors influencing N uptake, Ndfa, and yield processes in soybeans. Such studies should not only assess these factors, but also explore the associated uncertainties and their implications for decision-making. Additionally, future endeavors should be approached from both a production standpoint and with a look toward environmental and sustainability concerns.

3.4. Conclusion

The present findings highlight the critical role of weather factors such as water and temperature regulating both yield and N_2 fixation processes in soybeans. Sulfur might play a major role in reducing soil limitation for both yield and Ndfa. The uncertainties on yield and N_2 fixation were closely linked to soil characteristics, particularly texture, and SOM content, and were found to increase in magnitude with lower yields. In addition, a positive association between greater yields with superior Ndfa was evident. Some limitations of this study and potential opportunities for future research may focus on screening for more genotypic variation and other management strategies (e.g. planting date), in addition to exploring other potential nutrient interactions. Lastly, it is essential to consider the uncertainties when accounting for the variation of key biological processes, rather than focusing solely on their estimates.

Table 3.1 Study characterization by state, corresponding soil series, soybean variety, sowing date, and harvest date.

State	Soil Series	Soybean Variety	Maturity group	Planting mm/dd/yy	Harvesting
IA	Clarion	P28T14E	2.8	5/6/2021	10/13/2021

	Galva	P23A15	2.3	5/10/2021	10/6/2021
	Galva	P22T18E	2.2	5/13/2022	9/27/2022
	Drummer	P34T21SE	3.4	5/25/2021	10/11/2021
	Palms	P30T99	3.0	5/20/2021	9/30/2021
IN	Maumee	P26T23E	2.6	5/5/2021	9/28/2021
	Drummer	P34T21	3.4	5/12/2022	10/10/2022
	Maumee	P30T99E	3.0	5/20/2022	10/4/2022
	Eudora	AG40X70	4.0	5/12/2021	10/4/2021
	Rossville	AG40X70	4.0	5/12/2021	10/8/2021
	Crete	P39A45X	3.9	5/12/2021	10/12/2021
KS	Belvue	P39A45X	3.9	5/30/2021	10/8/2021
	Eudora	AG43XF2	4.3	5/9/2022	10/18/2022
	Eudora	AG43XF2	4.3	6/13/2022	10/18/2022
	Geary	P39A45X	3.9	6/20/2022	10/26/2022
MI	Riddles	GH2011E3	2.0	5/25/2021	10/1/2021
	Waukegan	XO2181E	0.2	4/30/2021	9/22/2021
MN	Marna	XO2181E	0.2	5/5/2021	10/10/2021
	Waukegan	AG18XF1	1.8	5/5/2022	9/30/2022
	Marna	AG18XF1	1.8	5/31/22	10/4/2022
	Barnes	0819R2X	1.9	5/7/2021	9/28/2021
ND	Fargo	AG07XF2	0.7	5/25/2022	10/3/2022
	Barnes	AG09XF0	0.9	5/24/2022	10/8/2022
	Kokomo	P34T21SE	3.4	5/21/2021	11/9/2021
OH	Crosby	P34T21SE	3.4	5/21/2021	11/9/2021
	Kokomo	P34T21SE	3.4	5/12/2022	10/20/2022
	Crosby	P34T21SE	3.4	5/12/2022	10/20/2022
	Brandt	P18T26R2	1.8	5/21/2021	9/30/2021
SD	Vienna	P18T26R2	1.8	5/21/2021	10/12/2021
	Brandt	P13T16E	1.3	5/21/2022	10/14/2022
	Vienna	P13T16E	1.3	5/27/2022	10/13/2022
	Plano	Renk RS219NX	2.1	4/29/2021	9/28/2021
WI	Loyal	AG14X8	1.4	5/7/2021	10/19/2021
	Plano	Renk RS219NX	2.1	5/16/2022	10/18/2022
	Mendota	Renk RS219NX	2.1	5/6/2022	10/7/2022

Table 3.2 Summary of soil and weather covariates (units or labels, minimum, mean, and maximum).

Parameter	**Unit**	**Min.**	**Mean**	**Max.**

$\text{Precipitation}_{\text{Planting-R2}}$	mm	53.58	185.85	391.41
$\text{Precipitation}_{\text{Prior-planting}}$	mm	4.32	56.13	152.67
$\text{Precipitation}_{\text{R2-R4}}$	mm	2.28	61.72	145.04
$\text{Precipitation}_{\text{R4-R6}}$	mm	1.53	72.33	165.59
$\text{VPD}_{\text{Planting-R2}}$	kPa	0.42	0.82	1.41
$\text{VPD}_{\text{R2-R4}}$	kPa	0.50	0.97	1.43
$\text{VPD}_{\text{R4-R6}}$	kPa	0.43	0.95	1.57
$\text{GDD}_{\text{Planting-R2}}$	°C	105.05	1171.39	3970.90
$\text{GDD}_{\text{R2-R4}}$	°C	314.90	1436.61	4117.70
$\text{GDD}_{\text{R4-R6}}$	°C	694.85	1766.97	4505.35
pH	-	5.10	6.47	7.90
Clay	%	8.00	22.57	45.00
Sand	%	10.00	29.45	71.31
Silt	%	17.70	47.39	67.00
SOM	%	1.20	3.38	5.20
SOM/(clay+silt)	-	2.22	4.95	8.10

SOM: soil organic matter; GDD: growing degree days; VPD: vapor pressure deficit.

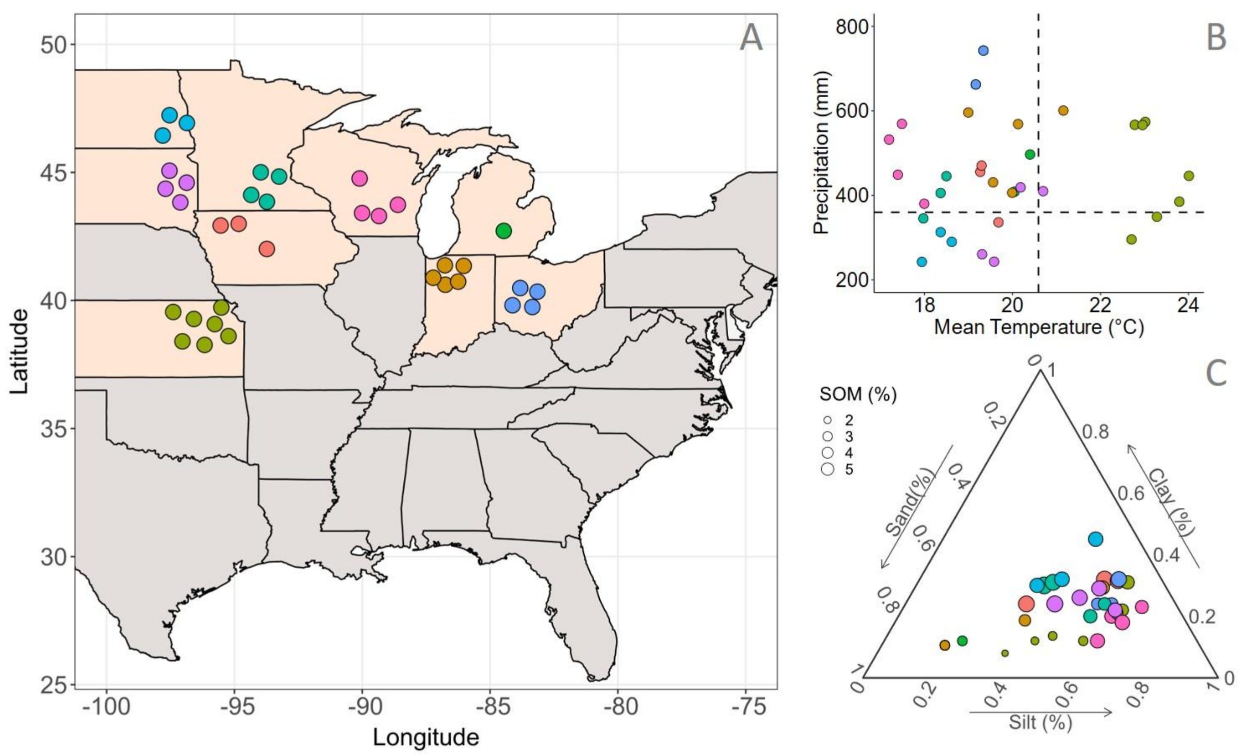


Figure 3.1 Map showcasing the geographical distribution of 35 site-year studies across the 2021 and 2022 soybean growing seasons (A); Graphical representation of the prevailing temperature and precipitation patterns observed during the study period (dashed lines represent the average precipitation and temperature) (B); Soil texture triangle and soil organic matter (SOM) patterns for all site-year studies (C).

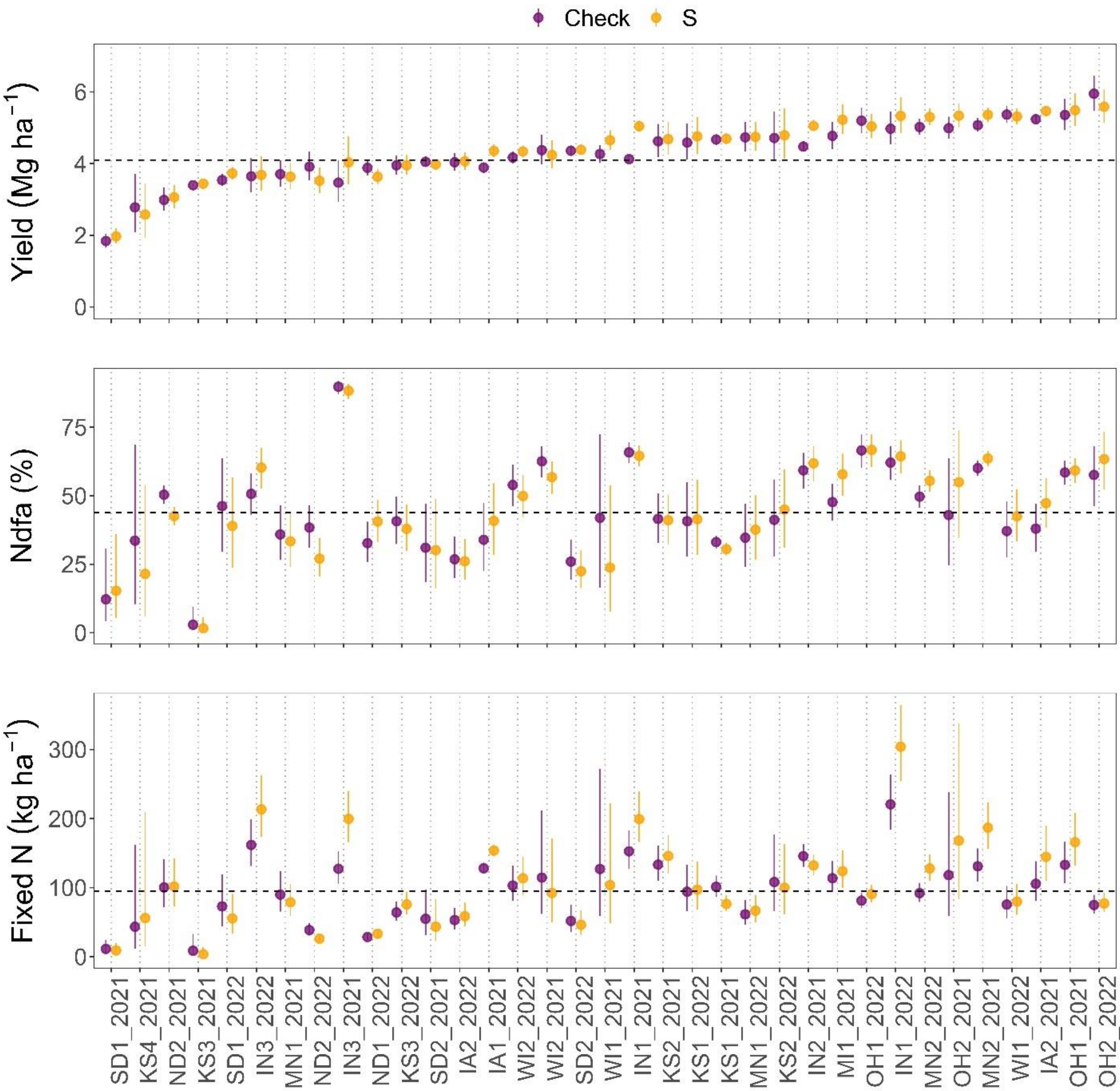


Figure 3.2. (A) Seed yield (Mg ha^{-1}), (B) percentage of N_2 fixation (Ndfa, %), and (C) amount of fixed-N (kg ha^{-1}) across all the 35 site-year studies. Site-year studies are ordered based on the average seed yield of treatments Check (purple) and S (orange), respectively in (A), (B), and (C). Data points represent the expected values (median from bootstrapping) and the uncertainty (length of the 95% confidence interval) for each treatment within each site-year. The horizontal dashed lines represent the average seed yield across sites (4.3 Mg ha^{-1}), the average N_2 fixation (Ndfa, %) (43 %), and the average amount of fixed-N (98 kg ha^{-1}).

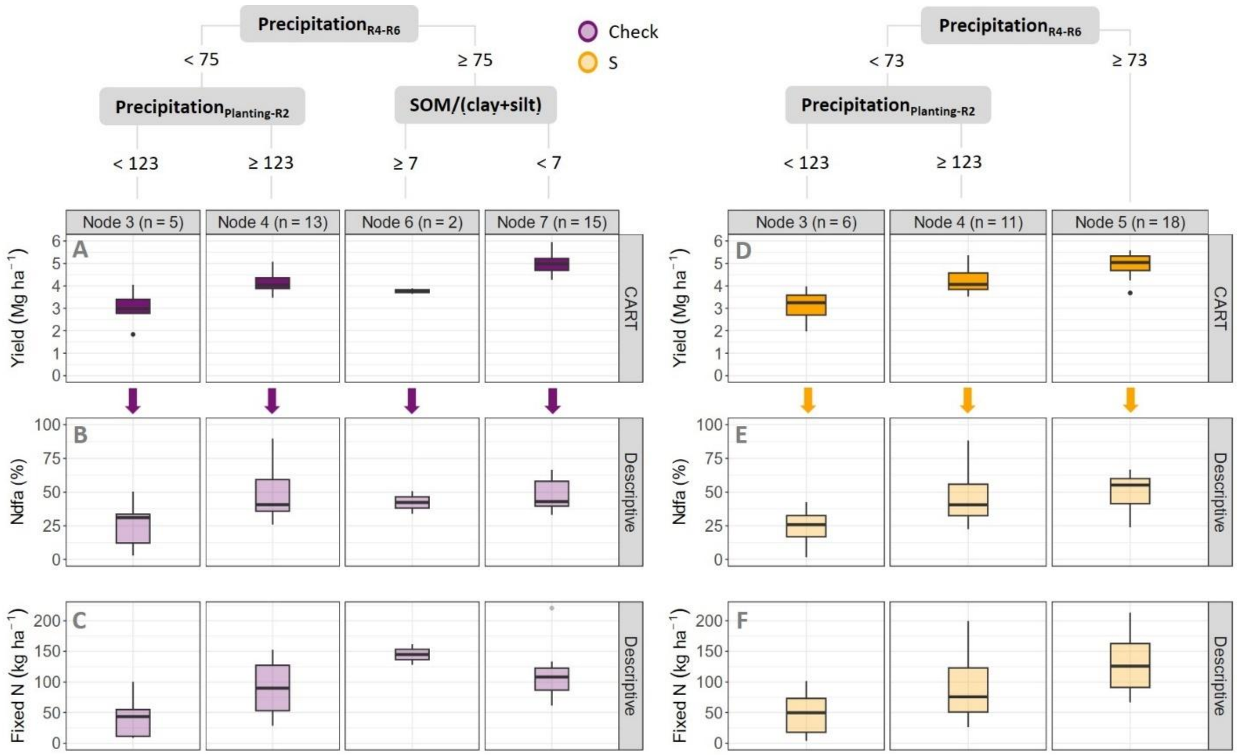


Figure 3.3 Regression tree depicting soybean yield expected value (Mg ha^{-1}; dark color) across 35 sites explained by the most relevant environmental covariables, grouped into four terminal nodes for the (A) Check treatment (purple) and three terminal nodes for the (D) S treatment (orange). Sites within each terminal node of the Check and S CART are described by their respective Ndfa (%) (B-E) and amount of fixed-N (kg ha^{-1}) (C-F) (light color). Precipitation is expressed in mm and SOM/(clay+silt) is dimensionless. The boxplots show the median and interquartile range (IQR: 25th–75th percentiles), with whiskers extending to 1.5 × IQR. Data points beyond the whiskers are plotted individually.

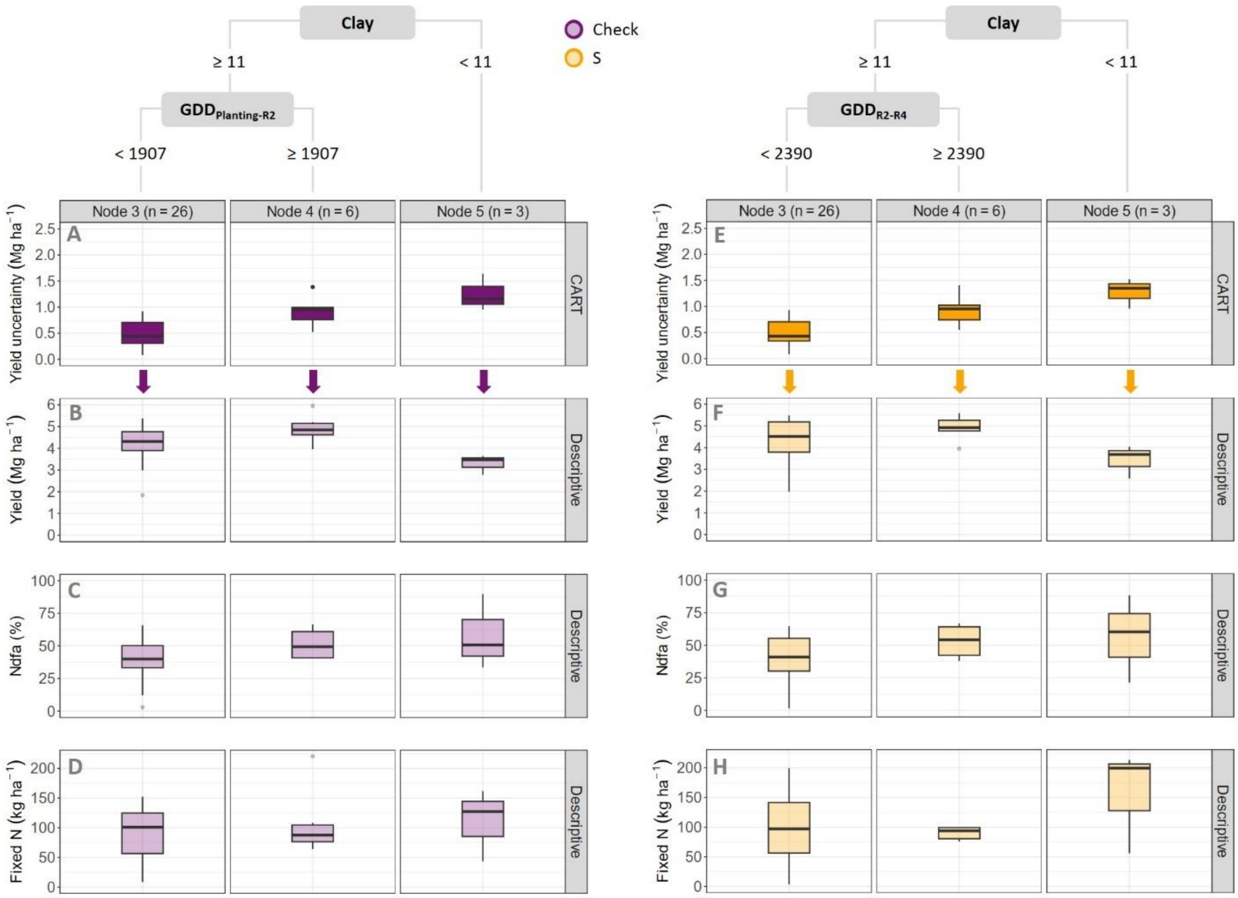


Figure 3.4 Regression tree depicting soybean yield uncertainty (Mg ha^{-1}; dark color) across 35 sites explained by the most relevant environmental covariables, grouped into three terminal nodes for the (A) Check treatment (purple) and (E) S treatments (orange). Sites within each terminal node of the Check and S CART are described by their respective seed yield (Mg ha^{-1}) (B-F), Ndfa (%) (C-G), and amount of fixed-N (kg ha^{-1}) (D-H) (light color). Soil clay content is expressed as percentage, and GDD in cumulative °C. The boxplots show the median and interquartile range (IQR: 25th–75th percentiles), with whiskers extending to 1.5 × IQR. Data points beyond the whiskers are plotted individually.

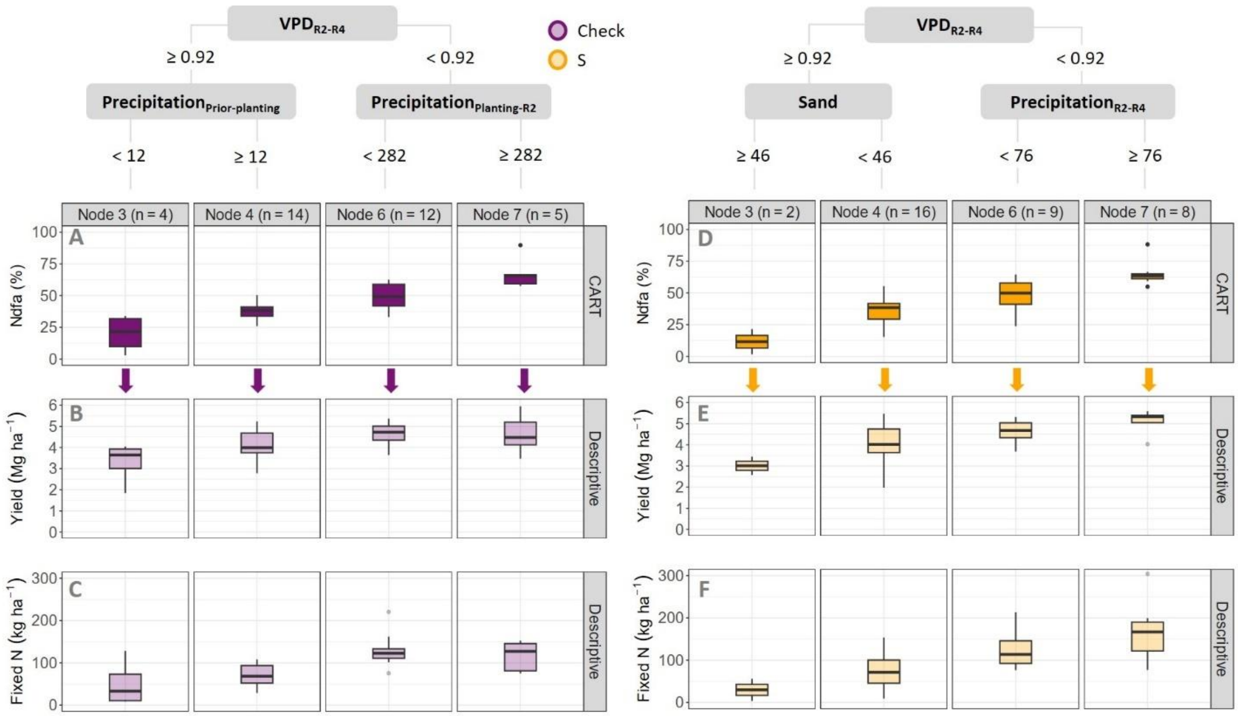


Figure 3.5 Regression tree depicting soybean Ndfa expected value (%; dark color) across 35 sites explained by the most relevant covariables, grouped into four terminal nodes for the (A) Check (purple) and (D) 5 terminal nodes for the S (orange). Sites within each terminal node of the Check and S CART are described by their respective seed yield (Mg ha^{-1}) (B-E) and amount of fixed-N (kg ha^{-1}) (C-F) values (light color). Vapor pressure deficit (VPD) is expressed in kPa, precipitation as mm, and soil sand content as a percentage. The boxplots show the median and interquartile range (IQR: 25th–75th percentiles), with whiskers extending to 1.5 × IQR. Data points beyond the whiskers are plotted individually.

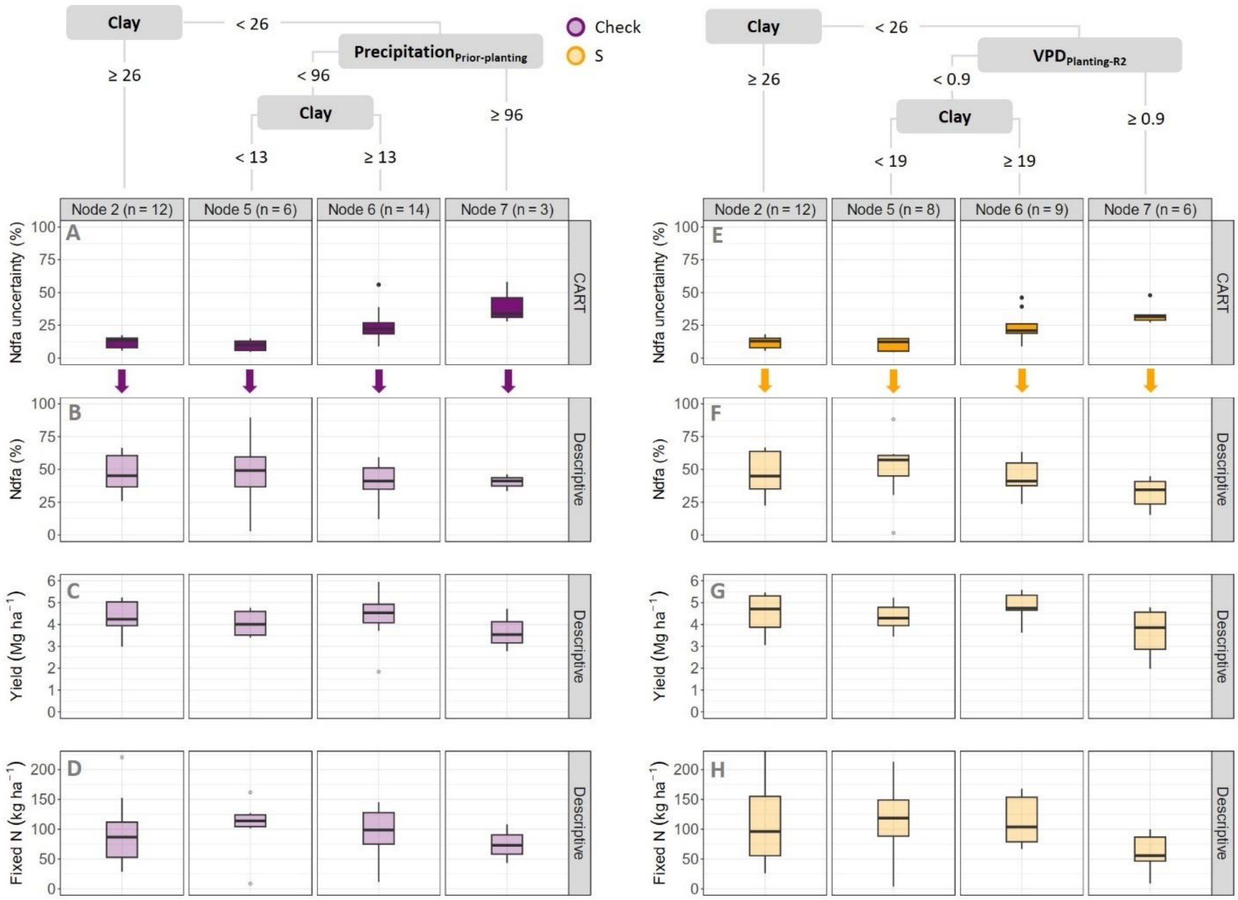


Figure 3.6 Regression tree depicting Ndfa uncertainty (%; dark color) across 35 sites explained by the most relevant environmental covariables, grouped into four terminal nodes for the (A) Check (purple) and (E) S (orange) treatments. Sites within each terminal node of the Check and S CART are described by their respective Ndfa (%) (B-F), seed yield (Mg ha^{-1}) (C-G), and amount of fixed-N (kg ha^{-1}) (D-H) (light color). Vapor pressure deficit (VPD) is expressed in kPa, precipitation as mm, and soil clay content as a percentage. The boxplots show the median and interquartile range (IQR: 25th–75th percentiles), with whiskers extending to 1.5 × IQR. Data points beyond the whiskers are plotted individually.

Chapter 4 - Soybean seasonal nitrogen fixation patterns and drivers

Submitted to European Journal of Agronomy

Luiz Felipe Almeida, Gabriel Hintz, Federico Gomez, Francisco Palmero, Adrian Correndo, Jeremy Ross, Mark Licht, Shaun Casteel, Manninder Singh, Seth Naeve, Rachel Vann, Jose Bais, Laura Lindsey, Shawn Conley, Jonathan Kleinjan, Péter Kovács, Mark Reiter, David Houlshouser, Trevor Hefley, P.V. Vara Prasad, and Ignacio Ciampitti

Abstract

Soybean [(*Glycine max* (L.) Merr)] production occupies approximately 35 million hectares in the United States (US) and contributes 29% of the global production. Despite its importance, the seasonal dynamics of N derived from the atmosphere (Ndfa, %) herein referred to as the N_2 fixation contribution, remain poorly characterized and a single-point measurement may misrepresent its true contribution to plant N nutrition. This study aimed to i) characterize in-season dynamics of Ndfa (peak, timing of the peak, and growth and decay rate of Ndfa), ii) investigate the association of mean seasonal Ndfa with seed yield and protein concentration, and iii) identify the main environmental variables driving the seasonal Ndfa.

The dataset comprises Ndfa quantified by the ^{15}N natural abundance method at 38 experiments across ten US states, with plant samples collected at full-bloom (R2), full-pod (R4), and full-seed (R6) developmental stages. Soybean phenology from emergence to beginning of maturity (R7) was simulated using ensembled models, APSIM and CSM-CROPGO. The environmental variables, including the initial conditions and in-season weather data, were grouped according to three seasonal crop phases: emergence to R2, R2 to R4, and R4 to R7. Seasonal Ndfa varied from 3% to 64%, with peaks ranging from 7% to 90%, typically occurring between R4 and R6 stages. Main factors influencing seasonal Ndfa include vapor pressure deficit (VPD), radiation, and precipitation. Higher VPD reduced peak Ndfa by up to 45%, while

intermediate organic matter levels and adequate precipitation enhanced this process. Increased seasonal Ndfa showed a slight positive effect on seed yield, but it was negatively linked to seed protein concentration. The primary challenge in advancing Ndfa models is linked to the scarcity of seasonal data. This study demonstrates that seasonal Ndfa patterns are environmentally driven. Proper characterization of seasonal Ndfa has broad applications, including crop improvement efforts and determination of precise nutrient budgets.

4.1. Introduction

Soybean [*Glycine max* (L.) Merr] is one of the most extensively cultivated crops in the United States (US), occupying 35 million hectares and contributing to 29% of global soybean production (USDA, 2024). Soybeans obtain nitrogen (N) from soil-N and atmospheric dinitrogen (N_2) fixation (Harper, 1974). The latter occurs via a symbiotic relationship, which promotes the differentiation of soybean root hair cells into nodules, enabling bacteroid to assimilate N_2 (Smith and Atkins, 2002). Annually, this N_2 fixation process is estimated to contribute 7.8 million Mg of N in the US (Herridge et al., 2022).

Although the symbiotic relationship is established early in the season (Tian et al., 2012), the N_2 fixation process persists during the seed filling (Bils and Howell, 1963), and thus, plays a critical role in defining the attainable the target seed protein concentration (Fabre and Planchon, 2000; Leffel et al., 1992). Seasonal N_2 fixation in soybean is often assessed or estimated based on a single seasonal measurement, which overlooks its temporal variation and responsiveness to environmental conditions (Ciampitti and Salvagiotti, 2018; de Borja Reis, 2021; Herridge et al., 2022). In addition, plant stressors such as low water availability (Serraj et al., 1999), reduced oxygenation (del Castillo et al., 1994), high soil temperatures (Lindemann and Ham, 1979), and nutritional co-limitations (Divito and Sadras, 2014), not only suppress seed yield and biomass allocation (Tamagno et al., 2018), but they also affect the overall contribution of N_2 fixation (Bethlenfalvay and Phillips, 1977). Therefore, relying on a single measurement does not provide the full extent of seasonal variation or a reliable estimate of mean crop N_2 fixation. A comprehensive understanding of seasonal N_2 fixation is therefore essential for advancing strategies that enhance its efficiency and maximize its agronomic benefits. Despite its relevance, the temporal dynamics of the N_2 fixation process remains poorly understood, as only a few studies have quantified N_2 fixation throughout the growing season (Córdova et al., 2019; Correndo et al., 2022; Tamagno et al., 2018).

Studying seasonal N_2 fixation dynamics requires the collection of multiple samples, a process that is inherently more costly and labor-intensive than a single measurement. One potential solution is the development of predictive models to estimate seasonal N_2 fixation. However, one of the major challenges in developing predictive models for N_2 fixation is the limited understanding of seasonal dynamics and how they are influenced by key environmental factors (Collino et al., 2015; Schipanski et al., 2010). For instance, de Borja Reis et al. (2021) recently developed a predictive model using a single measure of N_2 fixation. While promising, this approach did not incorporate integrated measurements and relied on a less reliable approach (relative abundance of ureides) under rainfed conditions (Purcell et al., 2004). By integrating temporal and spatial dimensions, advancements in this area could enable more accurate representation of seasonal N_2 fixation patterns, ultimately contributing to more precise determination of nutrient budgets for the sustainability of soybean agroecosystems.

Following this rationale, the aims of this research were to i) characterize seasonal N_2 fixation process, focusing on its timing, peak, and growth and decay rates; ii) evaluate the association of mean seasonal N_2 fixation with seed yield and protein concentration, and iii) identify key environmental variables driving the seasonal N_2 fixation.

4.2. Material and methods

4.2.1. Data description and experimental design

A total of 38 field studies were conducted during the 2021 and 2022 growing seasons across 10 US states: Arkansas, Iowa, Indiana, Kansas, Minnesota, North Carolina, North Dakota, Ohio, South Dakota, and Wisconsin **(Table C.1)**. These studies created 38 unique experiments (site ×

year combinations) with varied soil and weather conditions **(Table C.2)**. At all experiments, soybeans were grown without N fertilizer and inoculated with *Bradyrhizobium japonicum*, following the most common farming practices. The planting dates spanned from April 29th to June 20th over the course of the two years. Experiments included 3 to 5 replications, with plots ranging from 12.2 to 15.5 m in length and 6.1 m in width, and row spacing between 0.38 and 0.76 m.

4.2.2. N_2 fixation, seed yield, and seed protein measurements

Next to the soybean plots, strips of unfertilized corn were sown as a reference non-nodulating crop for determining the N_2 fixation contribution, quantified as the percentage of N derived from the atmosphere (Ndfa, %), or the proportion of fixed N relative to the total plant N uptake in kg N ha^{-1} (Peoples and Herridge, 1990). Soybean plants were collected at three stages of the growing season: full-bloom (R2), full-pod (R4), and full-seed (R6) (Fehr et al., 1971). References corn samples were collected at the same time. At each of these soybean phenological stages, aboveground biomass samples were collected from an area of 1.16 m^2, air-dried at 65°C and then ground to a particle size of 250 µm. Mass spectrometry analysis was conducted to measure the excess of isotope ^{15}N in the plant tissues (visION, Elementar Americas, Ronkonkoma, NY, US), through the ^{15}N natural abundance methodology. This is the most cost-effective method for N_2 fixation estimation (Shearer and Kohl, 1988). Additionally, this method accounts for different characteristics of the isotopic composition of the plant tissue, providing estimates of the proportion of Ndfa (%) (Unkovich et al., 1994), calculated using the following equation (Unkovich et al., 2008) **(Equation. 4.1)**:

$$Ndfa = \frac{15N\ of\ corn - 15N\ of\ soybean}{15N\ of\ corn - B} \times 100, \qquad \textbf{(Equation 4.1)}$$

where the *"δ^{15N} unfertilized corn"* is the ^{15}N natural abundance of the reference plant, *"$\delta 15^{N}$ soybean"* is the ^{15}N natural abundance of the N_2 fixing plant, and the "B" value is the ^{15}N natural abundance of a soybean grown in N free medium, assumed at -2.54 (Balboa and Ciampitti, 2020). At soybean maturity, an area between 18.3 and 23.2 m^2 per plot was mechanically harvested with a combine harvester to determine seed yield. Seed yield was adjusted to a standard moisture content of 130 g kg^{-1}. Seed samples were collected from the harvested area in each plot and seed protein content was determined by near-infrared (NIR) spectroscopy technique using a diode array analyzer (DA 7250, Perten Instruments, Sweden) and expressed on a dry basis. Seed protein data were not available for experiments IA2 (2022), IN1 (2022), and IN3 (2022).

4.2.3 Estimation of soybean phenology and collection of environmental variables

For better characterization of the crop growth period, an ensemble of the APSIM Next-Generation (Holzworth et al., 2014) and CSM-CROPGRO (Boote et al., 1998) in DSSAT v4.8.2 (Hoogenboom et al., 2019) models were used to simulate plant emergence (VE) and physiological maturity (R7), expressed in days after emergence (DAE) at each experiment, as these stages were not measured. Both crop growth models were run with generic coefficients according to the maturity groups of the cultivars used in the experiments and the simulated crop phenological stages were evaluated with observed data (stages R2, R4, and R6) **(Figure C.1)**.

Soil samples were collected at each site prior to planting to characterize soil properties (soil texture, pH, soil organic matter, N-nitrate, **Table C.1**). Weather variables were summarized across three growth periods: VE to R2, R2 to R4, and R4 to R7. These weather data were obtained from the nearest weather station using the Iowa Environmental Mesonet Reanalysis database (IEM, 2024). Specifically, vapor pressure deficit (VPD, kPa) was represented by average values for each period, while precipitation (PP, mm) and solar radiation (Rad, MJ m^{-2}) were summarized as cumulative values for the same periods.

4.2.4 Statistical analysis

A Bayesian model was fitted to the data to characterize seasonal Ndfa, and it is fully represented in the supplementary material **(Equation C.1)**. The likelihood followed a beta distribution, where Ndfa observations depend on a non-linear function of days after emergence (DAE). Parameters of the non-linear model include the peak (β_1), peak timing (β_2), and growth/decay rate (β_3), and followed weakly informative priors to capture a broad range of seasonal Ndfa patterns (Ciampitti et al., 2021; Córdova et al., 2019; Zapata et al., 1987). The accumulated seasonal Ndfa was calculated as the area under the curve (AUC) for the nonlinear model fitted to the relationship between Ndfa and time (DAE). The AUC was approximated using the trapezoidal rule, spanning growth stages VE to R7 for each experiment. In addition, *mean seasonal Ndfa* was estimated by dividing the AUC by the length of the growing season (in days) as a proxy for seasonal N_2 fixation efficiency. To investigate the relationship between seed yield and seed protein with the seasonal Ndfa, Bayesian linear regression was fitted to the data **(Equation C.2)**. In Bayesian, when the 95% credible interval of the regression slope includes zero, the relationship is considered undeterminable.

The relative importance of soil properties or meteorological variables on seasonal Ndfa (β1, β2, and β3) was determined using Bayesian Additive Regression Trees (BART) models (Chipman et al., 2010) **(Equation C.3)**. BART provides a flexible and robust framework for modeling complex relationships through a sum of regression trees (Chipman et al., 2010), capturing both additive and non-additive effects, including interactions and non-linear relationships. This analysis was performed using complete-case environmental variables to ensure that no experiments were excluded. Since initial soil N data (N-nitrate and N-ammonium) were not available for all experiments **(Table C.1)**, the same analysis was performed on a subset of data that included initial soil N to assess its relevance to seasonal Ndfa. Results showed that initial soil N was among the least important variables, so all experiments were included in the final analysis, excluding initial soil N from this step.

Following Tadesse and Vannucci (2021), 20 trees were used to improve predictor selection. Uniform priors were assigned to the variables chosen for splitting at each interior node, ensuring equal opportunity for all predictors to influence the model. To enhance robustness and compute variable inclusion proportions with standard errors, BART models were fitted using 5-fold cross-validation. In-sample predictions were then used to construct relationship plots with loess regression (Cleveland, 1979), examining the trends between parameters β_1, β_2, β_3 and environmental variables.

All statistical analyses were conducted within a Bayesian framework using R software (R Core Team, 2020). Non-linear seasonal Ndfa model and linear regression analysis to study relationship between seasonal N_2 fixation with yield and protein were fitted using Just another Gibbs sampler (JAGS) (Plummer, 2003). For all evaluations, 1,000 posterior samples were drawn as a warm-up, followed by 5,000 iterations per chain across 4 chains, with a thinning rate of 5. The MCMC convergence was monitored following Gelman and Rubin (1992) and R found to be 1 for all sampled parameters. The BART model was implemented using the BART package in R (McCulloch et al., 2024), with 10,000 burn-in iterations and 10,000 posterior draws.

4.3. Results

4.3.1. Seasonal N_2 fixation

Mean seasonal Ndfa varied significantly across locations, ranging from 3 to 64% **(Figure 1; Figure C.2 and Table C.3)**. The Ndfa peak (β_1) ranged from 7% to 90%, averaging 45%. These peaks typically occurred between 59 and 128 DAE, with 50% between R4 and R6 and 21% between R2 and R4, averaging 96 DAE (β_2). From sampled data, half of the average peak times coincided between R4 and R6 crop phenological stages. Considering the 95% and 60% credible intervals, the uncertainty in estimating the Ndfa peaks averaged 24% and 10%, respectively, while the uncertainty for the timing of these peaks was 80 and 37 days, respectively. Furthermore, the mean Ndfa growth and decay rate (β_3) ranged from 0.4 to 1.7, with an average of 1.32. The growth and decay rate uncertainty was 1.25 and 0.66 for the 95% and 60% credible intervals, respectively. The posterior distribution of the three model parameters describing seasonal Ndfa were detailed in **Table C.3 and Figure C.2** offering a comprehensive summary for future applications.

4.3.2. Seed yield and protein *vs* seasonal Ndfa

The mean seed yield across experiments ranged from 1.85 to 5.95 Mg ha^{-1} and protein content varied from 35.5 to 42.2%. A higher peak (β_1) of seasonal Ndfa was associated with increased yields, with an estimated gain of approximately 0.02 Mg ha^{-1} per unit increase in β_1 **(Figure 4.2A)**. However, the relationship between β_1 and protein content was unlikely, as the 95% credible interval for the slope included zero **(Figure 4.2B)**. Among the three parameters, β_1 exhibited the strongest association with mean seasonal Ndfa, with its peak consistently exceeding the seasonal mean **(Figure 4.2C)**. However, the timing of this peak showed no relationship with mean seasonal Ndfa **(Figure C.3).**

4.3.3. Seasonal N_2 fixation drivers

The most relevant variable explaining changes in the peak of Ndfa (β_1) was VPD from R2 to R4, which had an inclusion proportion of 13% in the total number of ensembled decision trees in the BART model. Following in importance were radiation from R2 to R4 and precipitation from plant emergence to R2, each with an inclusion proportion of 10% **(Figure 4.3A)**. For the timing of the peak (β_2), precipitation from R2 to R4 emerged as the most relevant variable, with an average inclusion proportion of 12%, followed by OM and clay, each with 11% **(Figure 4.3B)**. Initial soil N-nitrate and N-ammonium had low relevance for seasonal Ndfa patterns **(Figure C.3)**. Finally, the growth and decay rates of Ndfa (β_3) were most influenced by VPD between R4 and R7, which had an inclusion proportion of 11% **(Figure 4.3C)**.

In-sample model predictions indicated a general association between greater VPD at any point during the growing season and decreases of up to 45% in peak Ndfa. In contrast, greater precipitation, mainly before R2 stage, was associated with probable increases in peak Ndfa. The earliest peaks in Ndfa were primarily observed in soils with higher OM levels and clay content exceeding 20%. Additionally, greater precipitation volumes during the R2 to R4 stages tended to delay the peaks of Ndfa. Similarly, increased VPD up to a potential threshold before the R4 stage was also associated with a delayed Ndfa peak. The growth and decay rates of seasonal Ndfa exhibited more pronounced fluctuations (lower β_3) with increased VPD, particularly from R2 to R7 stages. Furthermore, higher accumulated radiation until R2 was linked to a decline in the growth and decay rates. Lastly, OM levels demonstrated an increasing plateau response, thereby attenuating the seasonal Ndfa curves **(Figure 4.4)**.

4.4. Discussion

This multi-environment study provides a comprehensive framework for assessing site-specific patterns of seasonal Ndfa, incorporating uncertainty estimation and laying the foundation for the development of more precise future predictive models. The high uncertainty on the timing of the Ndfa peak raises concerns about the reliability of single-sampling approaches and the assumption of a universal sampling time (Ciampitti et al., 2021). Seasonal Ndfa peaks showed high plasticity and frequently occurred between the pod and seed formation crop stages, aligning with the period of highest plant N demand (Zapata et al., 1987; Bergersen et al., 1992; Ciampitti et al., 2021). The positive relationship between the peak of Ndfa and seed yield suggests that soybeans fixing greater amounts of N may achieve a physiological cost-benefit advantage by increasing photosynthetic rate (Holland et al., 2023) rather than penalizing photosynthate allocation to seeds (Tamagno, 2018). This topic requires further investigation from a more physiological standpoint.

Ndfa dynamics are influenced by multiple stressors, with water stress consistently reducing its contribution (King and Purcell, 2005). In combination with high evapotranspiration, driven by elevated VPD and lower rainfall, water stress accelerates crop development, shortening the period of active N_2 fixation (Lindemann and Ham, 1979; Santachiara et al., 2019). Drought stress, particularly during the seed filling, exacerbates the negative impacts on Ndfa, making recovery unlikely and potentially penalizing yield (Mastrodomenico et al., 2013). This highlights the importance of investigating genetic mechanisms for water deficit tolerance or delayed leaf senescence, helping to sustain photosynthate allocation to nodules and prolong Ndfa activity later into the season (Abu-Shakra et al., 1978; Purcell, 1997). Additionally, Ndfa suppression can result from ureide-N accumulation in root nodules, a physiological response triggered by drought and heat stress (Serraj and Sinclair, 1996; Hungria and Kaschuk, 2014). Soybeans increasingly rely on N remobilization from storage proteins under stress conditions, which might reduce carbon supply to nodules (Sinclair and de Wit, 1975). The synchronization of N sources is another key factor contributing to the high uncertainty in single measurements of Ndfa or for defining its peak timing. Soils with higher OM content or conditions that accelerate OM mineralization, such as high radiation periods or late planting, can increase soil N availability (Almeida et al., 2025; Moro Rosso et al., 2023), reducing Ndfa. The peaks in soil N supply can depress N_2 fixation levels, complicating the development of an effective diagnosis method based only on initial soil N testing (Latimore et al.,1977; Correndo et al., 2022).

Modeling seasonal Ndfa might serve as a proxy for the identification of more resilient soybean varieties (Patterson and LaRue, 1983; Herridge, 1990) with traits that enhance an effective use of N for yield formation (Ladha et al., 2022). Crop management strategies that minimize crop stress can improve soybean N utilization and remobilization efficiency (Barat-Carnino et al., 2022) by optimizing growing conditions that regulate plant biomass and N demand (de Borja Reis et al., 2021; Eck et al., 1987; Tamagno et al., 2018). Despite limited spatial and temporal data challenges, integrating Bayesian and non-parametric modeling (Cook et al., 2023) provided a robust framework for quantifying key biological parameters of seasonal Ndfa. These approaches help address fixation timing and magnitude uncertainties, offering insights into the environmental and physiological factors shaping its dynamics.

Future research could further enhance seasonal Ndfa modeling by incorporating high-resolution temporal data on crop development, soil conditions, and weather variability. Bayesian statistical methods offer a powerful and accessible tool for this purpose (Ghosh, 2018), integrating prior knowledge with observed data to generate more precise and reliable inferences (Wakefield, 2013). Expanding data integration across diverse environments could extend seasonal Ndfa predictions beyond experimental locations, enabling a more comprehensive assessment of the potential of N_2 fixation process in soybean agroecosystems. This, in turn, would support the development of realistic N budgets, improving management strategies, and long-term sustainability.

4.5. Conclusion

Single-season estimates may misrepresent N_2 fixation contribution and fail to capture the influence of environmental stressors. A more precise seasonal Ndfa characterization is essential for crop improvement to optimize resource use efficiency and for obtaining more precise quantification of N nutrient budget. The main challenge remains the scarcity of long-term N_2 fixation data. This study addressed this limitation by incorporating biological support into model parameters, allowing updates as new data emerge and reducing uncertainty in seasonal estimates. Key variables, such as VPD, precipitation, soil texture, and OM, provide a foundation for refining predictive models, improving N_2 fixation estimates at experimental and on-farm scales, and improving agroecosystem N budgets.

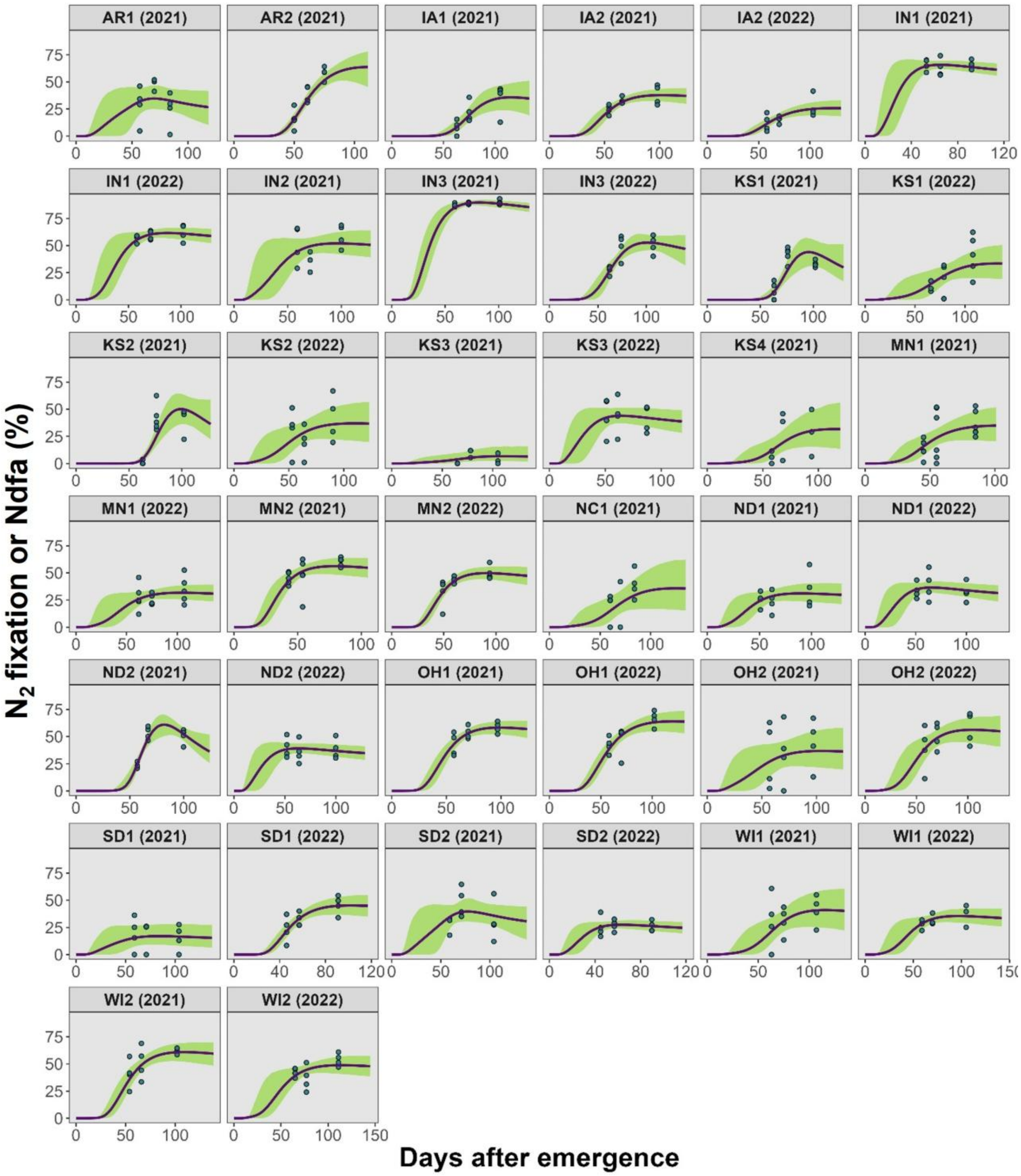


Figure 4.1 The N_2 fixation (expressed as the proportion of the N derived from the atmosphere, Ndfa; %) over time (days after emergence) across 38 experiments. The solid purple line represents the mean expected prediction, and the green ribbon the 95% credible interval of the predictions. Blue circles correspond to the N_2 fixation in each plot at soybean full-bloom (R2), full-pod (R4), and full-seed (R6) crop phenological stages.

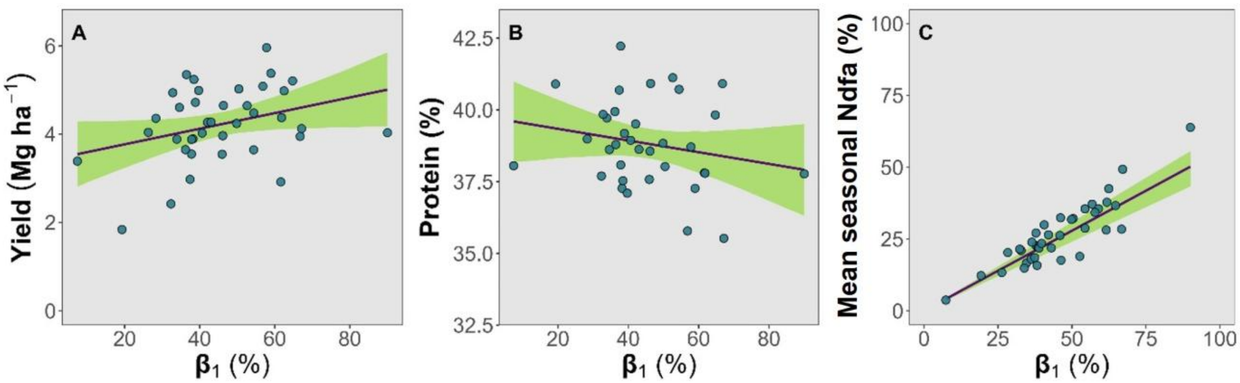


Figure 4.2 Relationship between seasonal N_2 fixation (expressed as a proportion of the N derived from the atmosphere, Ndfa; %) peak (β_1) with seed yield (Mg ha^{-1}) (A), seed protein content (%) (B), and mean seasonal Ndfa (seasonal Ndfa area under the curve over the season length) (C). The solid line represents the mean expected prediction, and the green ribbon the 95% credible interval of the predictions. Blue cirles correspond to the mean values of seed yield (A), protein (B), mean seasonal N_2 fixation (C) of each experiment. The fitted models were Yield = 3.41 + 0.018 β_1 ($R^2 = 0.09$), Protein = 30.74 - 0.02 β_1 ($R^2 = 0.04$), and Mean seasonal Ndfa = 3.406 + 0.018 β_1 ($R^2 = 0.28$).

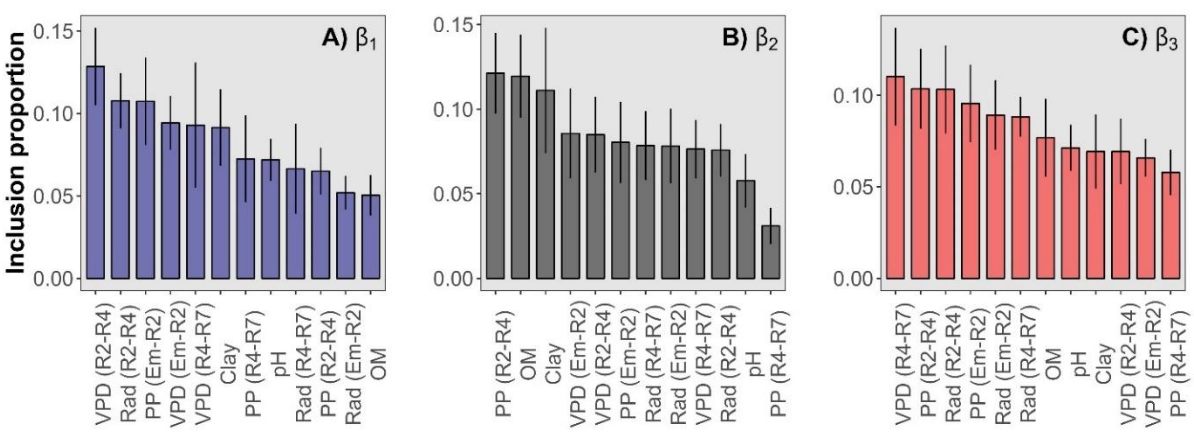


Figure 4.3 Posterior inclusion proportions for N_2 fixation (expressed as a proportion of the N derived from the atmosphere, Ndfa; %) parameters estimated by the BART model: (A) β_1, the peak level of Ndfa; (B) β_2, the timing of Ndfa peak; (C) β_3, the Ndfa growth and decay rate. Bars represent the mean posterior inclusion proportions for each variable, while the vertical lines indicate the standard error of these proportions based on a 5-fold cross-validation scheme. Variables are ordered according to their importance for each parameter.

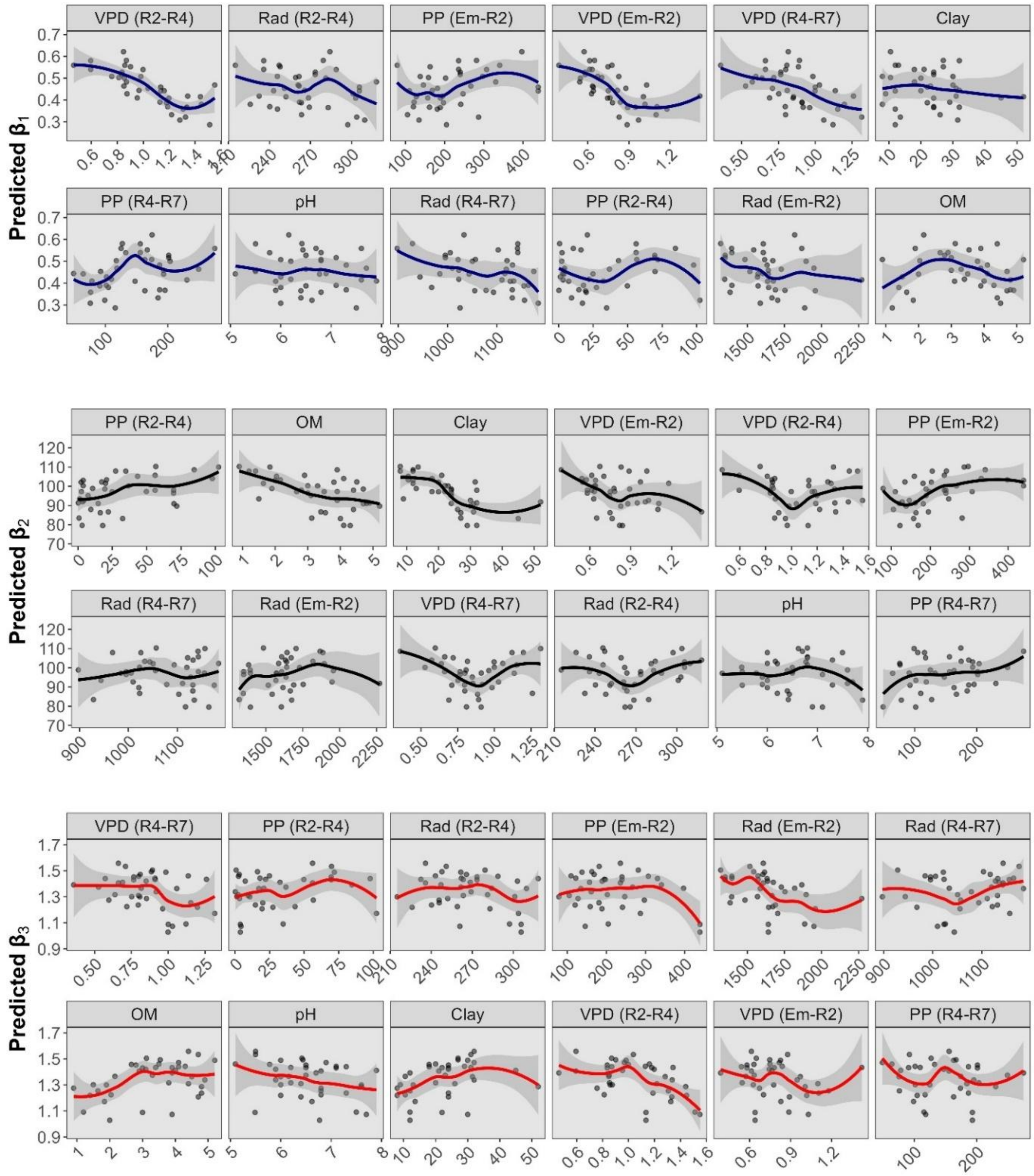


Figure 4.4 Relationship between the predicted coefficient values of β_1 (the peak level of Ndfa), β_2 (the timing of Ndfa peak), and β_3 (the Ndfa growth and decay rate) and environmental variables (VPD: vapor-pressure deficit, kPa; PP: precipitation, mm; Rad: solar radiation, MJ m^{-2}; OM: soil organic matter, %) from the in-sample BART model average from 5-fold cross-validation. Circles represent the predicted values, with loess smooth regressions indicating trends and shaded areas representing the 95% credible intervals.

Chapter 5 - Final remarks

This dissertation addresses critical questions related to the N economy in soybeans through a collaborative effort among multiple universities. We compiled a large dataset that connects various components of soybean production, including N_2 fixation footprints, providing a strong foundation for N management and future efforts. Using advanced data analysis tools such as Bayesian statistics, machine learning, generalized additive models and crop modelling, we provide quantitative and probabilistic assessments to improve soybean N management guidelines, optimize N use efficiency and refine N budgeting in soybean, considering environmental effects on seed yield, plant nutrition and seed composition.

The primary objective of Chapter 2 was to assess the impact of N and S fertilization on soybean yield, N status, and biological N_2 fixation. Strong evidence suggests that N fertilization is highly unlikely to increase soybean yields, mainly due to the previously unexplored uncertainty in treatment responses. Conversely, S was found to be a major co-limiting factor influencing plant N uptake and status, thereby increasing yields in certain environments. The strong relationship between N and S implies that in-season diagnostic tools may help to identify N deficiencies and co-limitations. The proposed apparent N dilution curve provides an opportunity to develop a more efficient methodology for benchmarking potential in-season N limitations that, while not necessarily resulting in yield penalties, may negatively impact seed quality, N balance and overall system sustainability.

Expanding on the influence of environmental and soil factors, Chapter 3 further explored the primary drivers of soybean yield and N_2 fixation, along with their associated uncertainties. The results confirm the critical influence of weather on plant growth, with vapor pressure deficit (VPD) and precipitation emerging as the primary regulators of yield and N_2 fixation, followed by soil texture and organic matter. The increase in N uptake with S fertilization further supports the hypothesis of nutrient synergy, particularly in low organic matter environments or specific soil textures that may limit nutrient availability. Even low rates of S fertilization can help to alleviate soil limitations and improve plant N economy. Environments with higher N_2 fixation showed greater yield stability, emphasizing the compensatory effect of N_2 fixation in environments with lower in-season soil N availability. Future research should explore genetic variation, multiple S application rates and timings, and management practices such as planting date and crop rotation to better understand scenarios where N limitation can be mitigated to improve N use efficiency.

Building on the findings from previous chapters, Chapter 4 characterized the seasonal dynamics of N_2 fixation at the local level using Bayesian modeling, revealing high variability in fixation patterns. Relying on single time point measurements to assess N_2 fixation is likely to misrepresent its true contribution due to uncertainty in the timing of peak fixation, despite its frequent occurrence between early pod and seed formation stages. Water stress proved to be a significant constraint on N_2 fixation, particularly under high VPD and low rainfall conditions, which accelerated crop development and shortened the contribution of N_2 fixation. Our results highlight the importance of integrating limited temporal biological data with Bayesian statistical methods to refine models for complex and chaotic processes such as N_2 fixation. In addition, increased N_2 fixation had minimal effects on seed protein, and its C cost may be relevant under circumstances not investigated in the current study. Overall, optimizing management recommendations based on environmental characteristics will ensure economically and environmentally sustainable soybean production for decades to come, while meeting global food security and industrial demands.

References

Abu-Shakra, S.S., et al., 1978. Nitrogen fixation and delayed leaf senescence in soybeans. Science, 199(4332), pp.973-975.

Almeida, L.F, et al., 2023. Soybean yield response to nitrogen and sulfur fertilization in the United States: contribution of soil N and N fixation processes. European Journal of Agronomy 145, 126791. https://doi.org/10.1016/j.eja.2023.126791

Almeida, L.F., et al., 2025. Assessing the influence of environmental drivers on soybean seed yield and nitrogen fixation estimates and uncertainties in the United States. European Journal of Agronomy, 162, p.127428.

Anderson, A.J., Spencer, D.1950. Sulphur in nitrogen metabolism of legumes and non-legumes. Aust J Sci Res B 3, 431–449. https://doi.org/10.1071/bi9500431

Andrade, J.F., Rattalino, J.I., Mourtzinis, S., Conley, S.P., Ciampitti, I.A., Dunphy, J.E., Gaska, J.M., Glewen, K., Holshouser, D.L., Kandel, H.J., Kyveryga, P., Lee, C.D., Licht, M.A., Lindsey, L.E., Mcclure, M.A., Naeve, S., Nafziger, E.D., Orlowski, J.M., Ross, J., Staton, M.J., Thompson, L., Specht, J.E., Grassini, P., 2019. Assessing the influence of row spacing on soybean yield using experimental and producer survey data. F. Crop. Res. 230, 98–106. https://doi.org/10.1016/j.fcr.2018.10.014

Balboa, G.R., Ciampitti, I.A., 2020. Estimating biological nitrogen fixation in field-grown soybeans: impact of B value. Plant Soil 446, 195–210. https://doi.org/10.1007/s11104-019-04317-1

Balboa, G.R., et al., 2018. Shifts in Soybean Yield, Nutrient Uptake, and Nutrient Stoichiometry: A Historical Synthesis-Analysis. Crop Science 58, 43–54. https://doi.org/10.2135/cropsci2017.06.0349

Barat-Carnino, M., et al., 2022. Sowing date affects soybean biological nitrogen fixation. Crop Science 62, 2428–2438. https://doi.org/10.1002/csc2.20812

Bergersen, F.J., et al., 1992. Nitrogen fixation during vegetative and reproductive growth of irrigated soybeans in the field: application of δ 15N methods. *Australian Journal of Agricultural Research*, *43*(1), pp.145-153.

Bethlenfalvay, G.J., Phillips, D.A., 1977. Ontogenetic Interactions between Photosynthesis and Symbiotic Nitrogen Fixation in Legumes 1. Plant Physiology 60, 419–421. https://doi.org/10.1104/pp.60.3.419

Bils, R.F., Howell, R.W., 1963. Biochemical and Cytological Changes in Developing Soybean Cotyledons 1. Crop Science 3, 304–308. https://doi.org/10.2135/cropsci1963.0011183X000300040008x

Boote, K.J., Jones, J.W., Hoogenboom, G., 1998. Simulation of Crop Growth: CROPGRO Model, in: Agricultural Systems Modeling and Simulation. CRC Press.

Borja Reis, A.F. de, Rosso, L.H.M., Davidson, D., Kovács, P., Purcell, L.C., Below, F.E., Casteel, S.N., Knott, C., Kandel, H., Naeve, S.L., Carciochi, W., Ross, W.J., Favoretto, V.R., Archontoulis, S., Ciampitti, I.A., 2021. Sulfur fertilization in soybean: A meta-analysis on yield and seed composition. Eur. J. Agron. 127, 126285. https://doi.org/10.1016/j.eja.2021.126285

Breiman, L., et al., 1984. Classification and Regression Trees. https://doi.org/10.1201/9781315139470

Brooks, K., Mourtzinis, S., Conley, S.P., Reiter, M.S., Gaska, J., Holshouser, D.L., Irby, T., Kleinjan, J., Knott, C., Lee, C., Lindsey, L., Naeve, S., Ross, J., Singh, M.P., Vann, R., Matcham, E., 2022. Soybean yield response to sulfur and nitrogen additions across diverse US environments. Agron. J. 1–14. https://doi.org/10.1002/agj2.21216

Bürkner, P.-C., 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. J. Stat. 80, 1–28. https://doi.org/10.18637/jss.v080.i01

Cafaro La Menza, N., Monzon, J.P., Lindquist, J.L., Arkebauer, T.J., Knops, J.M.H., Unkovich, M., Specht, J.E., Grassini, P., 2020. Insufficient nitrogen supply from symbiotic fixation reduces seasonal crop growth and nitrogen mobilization to seed in highly productive soybean crops. Plant. Cell Environ. 43, 1958–1972. https://doi.org/10.1111/pce.13804

Cafaro La Menza, N., Monzon, J.P., Specht, J.E., Grassini, P., 2017. Is soybean yield limited by nitrogen supply? F. Crop. Res. 213, 204–212. https://doi.org/10.1016/j.fcr.2017.08.009

Casteel, S.N., Chien, S.H., Gearhart, M.M., 2019. Field Evaluation of Ammonium Sulfate versus Two Fertilizer Products Containing Ammonium Sulfate and Elemental Sulfur on Soybeans. Commun. Soil Sci. Plant Anal. 50, 2941–2947. https://doi.org/10.1080/00103624.2019.1689257

Castellano, M.J., et al., 2012. Linking Carbon Saturation Concepts to Nitrogen Saturation and Retention. Ecosystems 15, 175–187. https://doi.org/10.1007/s10021-011-9501-3

Chipman, H.A., et al., 2010. BART: Bayesian additive regression trees. The Annals of Applied Statistics 4, 266–298. https://doi.org/10.1214/09-AOAS285

Ciampitti et al., 2021a. Agronomy eUpdate May 27th, Issue 857. https://eupdate.agronomy.ksu.edu/article_new/soybean-response-to-standing-water-and-saturated-soils-445-2 (accessed 9 October 2023).

Ciampitti, I.A., et al., 2021b. Revisiting biological nitrogen fixation dynamics in soybeans. Frontiers in plant science 12, 727021. https://doi.org/10.3389/fpls.2021.727021

Ciampitti, I.A., Salvagiotti, F., 2018. New Insights into Soybean Biological Nitrogen Fixation. Agronomy Journal 110, 1185–1196. https://doi.org/10.2134/agronj2017.06.0348

Cleveland, W.S., 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association 74, 829–836. https://doi.org/10.1080/01621459.1979.10481038

Collino, D.J., et al., 2015. Biological nitrogen fixation in soybean in Argentina: relationships with crop, soil, and meteorological factors. Plant Soil 392, 239–252. https://doi.org/10.1007/s11104-015-2459-8

Collino, D.J., Salvagiotti, F., Perticari, A., Piccinetti, C., Ovando, G., Urquiaga, S., Racca, R.W., 2015. Biological nitrogen fixation in soybean in Argentina: relationships with crop, soil, and meteorological factors. Plant Soil 392, 239–252. https://doi.org/10.1007/s11104-015-2459-8

Cook, J.D., Williams, D.M., Walsh, D.P., Hefley, T.J., 2023. Bayesian forecasting of disease spread with little or no local data. Sci Rep 13, 8137. https://doi.org/10.1038/s41598-023-35177-6

Córdova, S.C., Castellano, M.J., Dietzel, R., Licht, M.A., Togliatti, K., Martinez-Feria, R., Archontoulis, S. V., 2019. Soybean nitrogen fixation dynamics in Iowa, USA. F. Crop. Res. 236, 165–176. https://doi.org/10.1016/j.fcr.2019.03.018

Córdova, S.C., Licht, M.A., 2020. Soybean profitability and yield component response to nitrogen fertilizer in Iowa 1–16. https://doi.org/10.1002/agg2.20092

Correndo, A.A., et al., 2022. Footprints of corn nitrogen management on the following soybean crop. Agronomy Journal, 114(2), pp.1475-1488.

Correndo, A.A., Moro Rosso, L.H., Ciampitti, I.A., 2021a. Retrieving and processing agro-meteorological data from API-client sources using R software. BMC Res. Notes 14, 205. https://doi.org/10.1186/s13104-021-05622-8

Correndo, A.A., Tremblay, N., Coulter, J.A., Ruiz-diaz, D., Franzen, D., Nafziger, E., Prasad, V., Moro, L.H., Steinke, K., Du, J., Messina, C.D., Ciampitti, I.A., 2021b. Unraveling uncertainty drivers of the maize yield response to nitrogen : A Bayesian and machine learning approach. Agric. For. Meteorol. 311, 108668. https://doi.org/10.1016/j.agrformet.2021.108668

Cossani, C.M., Sadras, V.O., 2018. Chapter Six - Water–Nitrogen Colimitation in Grain Crops, in: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 231–274. https://doi.org/10.1016/bs.agron.2018.02.004

de Borja Reis, A.F., et al., 2021. Environmental factors associated with nitrogen fixation prediction in soybean. Frontiers in Plant Science 12, 675410. https://doi.org/10.3389/fpls.2021.675410

de Borja Reis, A.F.D., et al., 2021. Sulfur fertilization in soybean: A meta-analysis on yield and seed composition. European Journal of Agronomy 127, 126285. https://doi.org/10.1016/j.eja.2021.126285 de Vries, et al., 2003. Uncertainties in the fate of nitrogen II: A quantitative assessment of the uncertainties in major nitrogen fluxes in the Netherlands. Nutrient Cycling in Agroecosystems 66, 71–102. https://doi.org/10.1023/A:1023354109910

del Castillo, L.D., et al., 1994. The Role of Oxygen in the Regulation of Nitrogenase Activity in Drought-Stressed Soybean Nodules. Plant Physiology 106, 949–955. https://doi.org/10.1104/pp.106.3.949

Divito, G.A., Echeverría, H.E., Andrade, F.H., Sadras, V.O., 2015. Diagnosis of S deficiency in soybean crops : Performance of S and N : S determinations in leaf , shoot and seed. F. Crop. Res. 180, 167–175. https://doi.org/10.1016/j.fcr.2015.06.006

Divito, G.A., Echeverría, H.E., Andrade, F.H., Sadras, V.O., 2016a. N and S concentration and stoichiometry in soybean during vegetative growth : Dynamics of indices for diagnosing the S status. F. Crop. Res. 198, 140–147. https://doi.org/10.1016/j.fcr.2016.08.018

Divito, G.A., Sadras, V.O., 2014. How do phosphorus, potassium and sulphur affect plant growth and biological nitrogen fixation in crop and pasture legumes? A meta-analysis. Field Crops Research 156, 161–171. https://doi.org/10.1016/j.fcr.2013.11.004

Doss, B.D., et. al, 1974. Effect of Soil Water Stress at Various Growth Stages on Soybean Yield. Agronomy Journal 66, 297–299. https://doi.org/10.2134/agronj1974.00021962006600020032x

Eck, H.V., et al., 1987. Plant water stress at various growth stages and growth and yield of soybeans. Field Crops Research 17, 1–16. https://doi.org/10.1016/0378-4290(87)90077-3

Fabre, F., Planchon, C., 2000. Nitrogen nutrition, yield and protein content in soybean. Plant Science 152, 51–58. https://doi.org/10.1016/S0168-9452(99)00221-6

Fehr, W., Caviness, C., Burmood, D., Pennington, J., 1971. Stage of Development Descriptions for Soybeans, Glycine Max (L.) Merrill. Crop Science 2, 929–930.

Fehr, W., Caviness, et .al, 1971. Stage of Development Descriptions for Soybeans, Glycine Max (L.) Merrill. Crop Science 2, 929–930.

Fehr, W.R., Caviness, C.E., Burmood, D.T., Pennington, J.S., 1971. Stage of Development Descriptions for Soybeans, Glycine Max (L.) Merrill 1. Crop Sci. 11, 929–931. https://doi.org/10.2135/cropsci1971.0011183X001100060051x

Fehr, Walter R. and Caviness, Charles E., 1977. Stages of soybean development. Special Report. 87. http://lib.dr.iastate.edu/specialreports/87

Ferguson, B.J., Lin, M.-H., Gresshoff, P.M., 2013. Regulation of legume nodulation by acidic growth conditions. Plant Signal Behav 8, e23426. https://doi.org/10.4161/psb.23426

Ferraro, D.O., et al., 2009. An analysis of the factors that influence sugarcane yield in Northern Argentina using classification and regression trees. Field Crops Research 112, 149–157. https://doi.org/10.1016/j.fcr.2009.02.014

Froes, A., Reis, D.B., Moro, L.H., Adee, E., Davidson, D., Purcell, L.C., Below, F.E., Casteel, S.N., Knott, C., Kandel, H., Naeve, S.L., Singh, M.P., Archontoulis, S., Ciampitti, I.A., 2022. Seed inoculation with Azospirillum brasilense in the U.S. soybean systems 283. https://doi.org/10.1016/j.fcr.2022.108537

Gaspar, A.P., et al., 2017. Dry Matter and Nitrogen Uptake, Partitioning, and Removal across a Wide Range of Soybean Seed Yield Levels. Crop Science 57, 2170–2182. https://doi.org/10.2135/cropsci2016.05.0322

Gelman, A., Rubin, D.B., 1992. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science 7, 457–472. https://doi.org/10.1214/ss/1177011136

Ghosh, I., 2018. Bayesian Methods, in: Handbook of Statistics. Elsevier, pp. 173–196. https://doi.org/10.1016/bs.host.2018.07.001

Harper, J.E., 1974. Soil and Symbiotic Nitrogen Requirements for Optimum Soybean Production1. Crop Science 14, cropsci1974.0011183X001400020026x. https://doi.org/10.2135/cropsci1974.0011183X001400020026x

Hassink, J., 1994. Effects of soil texture and grassland management on soil organic C and N and rates of C and N mineralization. Soil Biology and Biochemistry 26, 1221–1231. https://doi.org/10.1016/0038-0717(94)90147-3

Herridge, D., Rose, I., 2000. Breeding for enhanced nitrogen fixation in crop legumes. Field Crops Research 65, 229–248. https://doi.org/10.1016/S0378-4290(99)00089-1

Herridge, D.F., Bergensen, F.J., 1988. Symbiotic nitrogen fixation. In Advances in Nitrogen Cycling in Agricultural Ecosystems (J. R. Wilson, Ed.), pp. 46-65. C.A.B. International, Wallingford.

Herridge, D.F., et al., 1990. Measurement of nitrogen fixation by soybean in the field using the ureide and natural 15N abundance methods. Plant physiology, 93(2), pp.708-716.

Herridge, D.F., et al., 2008. Global inputs of biological nitrogen fixation in agricultural systems. Plant Soil 311, 1–18. https://doi.org/10.1007/s11104-008-9668-3

Herridge, D.F., et al., 2022. Quantifying country-to-global scale nitrogen fixation for grain legumes II. Coefficients, templates and estimates for soybean, groundnut and pulses. Plant Soil 474, 1–15. https://doi.org/10.1007/s11104-021-05166-7

Holland, B.L., et al., 2023. A genome-scale metabolic reconstruction of soybean and Bradyrhizobium diazoefficiens reveals the cost–benefit of nitrogen fixation. New Phytologist 240, 744–756. https://doi.org/10.1111/nph.19203

Holzworth, D.P., et al, 2014. APSIM – Evolution towards a new generation of agricultural systems simulation. Environmental Modelling & Software 62, 327–350. https://doi.org/10.1016/j.envsoft.2014.07.009

Hoogenboom, G., et al., 2019. The DSSAT crop modeling ecosystem, in: Advances in Crop Modelling for a Sustainable Agriculture. Burleigh Dodds Science Publishing.

Hooten, M.B., Hefley, T.J., 2019. Bringing Bayesian Models to Life. CRC Press (1st ed.). https://doi.org/10.1201/9780429243653

Hufkens, K., Basler, D., Milliman, T., Melaas, E.K., Richardson, A.D., 2018. An integrated phenology modelling framework in r. Methods Ecol. Evol. 9, 1276–1285. https://doi.org/10.1111/2041-210X.12970

Hungria, M. and Kaschuk, G., 2014. Regulation of N2 fixation and NO3−/NH4+ assimilation in nodulated and N-fertilized Phaseolus vulgaris L. exposed to high temperature stress. Environmental and Experimental Botany, 98, pp.32-39.

Hungria, M., Vargas, M., 2000. Environmental factors affecting N2 fixation in grain legumes in the tropics, with an emphasis on Brazil. F. Crop. Res. 65, 151–164. https://doi.org/10.1016/S0378-4290(99)00084-2

Hussain, K., et al., 2011. Soybean Growth and Nitrogen Fixation as Affected by Sulfur Fertilization and Inoculation under Rainfed Conditions in Pakistan. International Journal of Agriculture & Biology, 13(6).

ISU, 2024. Iowa Environmental Mesonet. https://mesonet.agron.iastate.edu/ (accessed 2.22.24).

King, C.A., Purcell, L.C., 2005. Inhibition of N2 Fixation in Soybean Is Associated with Elevated Ureides and Amino Acids. Plant Physiology 137, 1389–1396. https://doi.org/10.1104/pp.104.056317

Krishnan, H.B., 2015. Improving the Sulfur-Containing Amino Acids of Soybean to Enhance its Nutritional Value in Animal Feed. pp. 235–249. https://doi.org/10.2134/agronmonogr50.c15

La Menza, N.C., et al., 2017. Is soybean yield limited by nitrogen supply? Field Crops Research 213, 204–212. https://doi.org/10.1016/j.fcr.2017.08.009

Lacasa, J., Hefley, T.J., Otegui, M.E., Ciampitti, I.A., 2021. A practical guide to estimating the light extinction coefficient with nonlinear models—a case study on maize. Plant Methods 17, 1–11. https://doi.org/10.1186/s13007-021-00753-2

Ladha, J.K., et al., 2022. Biological nitrogen fixation and prospects for ecological intensification in cereal-based cropping systems. Field Crops Research 283, 108541. https://doi.org/10.1016/j.fcr.2022.108541

Latimore Jr, et al., 1977. Effect of Ammonium and Nitrate Nitrogen upon Photosynthate Supply and Nitrogen Fixation by Soybeans 1. Crop Science, 17(3), pp.399-404.

Lawson, I.Y., et al., 1995. Effect of organic matter on the growth, nodulation, and nitrogen fixation of soybean grown under acid and saline conditions. Soil science and plant nutrition, 41(4), pp.721-728. https://doi.org/10.1080/00380768.1995.10417022

Leffel, R.C., et al., 1992. Nitrogen Metabolism of Normal and High-Seed-Protein Soybean. Crop Science 32, 747–750. https://doi.org/10.2135/cropsci1992.0011183X003200030034x

Lemaire, G., Ciampitti, I., 2020. Crop Mass and N Status as Prerequisite Covariables for Unraveling Nitrogen Use Efficiency across Genotype-by-Environment-by-Management Scenarios: A Review. Plants 9, 1309. https://doi.org/10.3390/plants9101309

Lemaire, G., Gastal, F., 1997. N Uptake and Distribution in Plant Canopies, in: Diagnosis of the Nitrogen Status in Crops. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–43. https://doi.org/10.1007/978-3-642-60684-7_1

Lemaire, G., Gilles, B., Fern, J.A., Gastal, F., Makowski, D., Ciampitti, I.A., 2021. Revisiting the critical nitrogen dilution curve for tall fescue : A quantitative synthesis 131. https://doi.org/10.1016/j.eja.2021.126380

Lemaire, G., Salette, J., SIGOGNE, M., TERRASSON, J.-P., 1984. Relation entre dynamique de croissance et dynamique de prélèvement d'azote pour un peuplement de graminées fourragères. I. — Etude de l'effet du milieu. Agronomie 4, 423–430. https://doi.org/10.1051/agro:19840503

Lindemann, W.C., Ham, G.E., 1979. Soybean Plant Growth, Nodulation, and Nitrogen Fixation as Affected by Root Temperature. Soil Science Society of America Journal 43, 1134–1137. https://doi.org/10.2136/sssaj1979.03615995004300060014x

Linn, D.M., Doran, J.W., 1984. Effect of water-filled pore space on carbon dioxide and nitrous oxide production in tilled and nontilled soils. Soil Science Society of America Journal, 48(6), pp.1267-1272. https://doi.org/10.2136/sssaj1984.03615995004800060013x

Mahal, N.K., Sawyer, J.E., Iqbal, J., Sassman, A.M., Mathur, R., Castellano, M.J., 2022. Role of sulfur mineralization and fertilizer source in corn and soybean production systems. Soil Sci. Soc. Am. J. 86, 1058–1071. https://doi.org/10.1002/saj2.20417

Mastrodomenico, et al., 2013. The response and recovery of nitrogen fixation activity in soybean to water deficit at different reproductive developmental stages. Environmental and Experimental Botany, 85, pp.16-21.

Matcham, E.G., Vann, R.A., Lindsey, L.E., Gaska, J.M., Lilley, D.T., Ross, W.J., Wright, D.L., Knott, C., Lee, C.D., Moseley, D., Singh, M., Naeve, S., Irby, J.T., Wiebold, W., Kandel, H., Lofton, J., Inman, M., Conley, S.P., 2021. Foliar fertilizers rarely increase yield in United States soybean. https://doi.org/10.1002/agj2.20889

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, Second Edition. CRC Press.

McCulloch, R., et al., 2024. BART: Bayesian Additive Regression Trees.

Monnens, D., et al., 2023. Rising vapor-pressure deficit increases nitrogen fixation in a legume crop. New Phytologist 239, 54–65. https://doi.org/10.1111/nph.18929

Moro Rosso, L.H., et al., 2023. Temporal variation of soil N supply defines N fixation in soybeans. European Journal of Agronomy 144, 126745. https://doi.org/10.1016/j.eja.2023.126745

Mourtzinis, S., et al., 2018. Soybean response to nitrogen application across the United States: A synthesis-analysis. Field Crops Research 215, 74–82. https://doi.org/10.1016/j.fcr.2017.09.035

Murphy, P.N.C., Stevens, R.J., 2010. Lime and Gypsum as Source Measures to Decrease Phosphorus Loss from Soils to Water. Water Air Soil Pollut 212, 101–111. https://doi.org/10.1007/s11270-010-0325-0

NASS, U., 2017. National Agricultural Statistics Service. Quick Stats. USDA National Agricultural Statistics Service. https://hdl.handle.net/10113/AA5765

Ney, B., Dore, T., Sagan, M., 1997. Grain Legumes, in: Diagnosis of the Nitrogen Status in Crops. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 107–117. https://doi.org/10.1007/978-3-642-60684-7_6

NRCS, U., 2025. Soil Survey Geographic Database (SSURGO) Access [WWW Document]. URL https://www.arcgis.com/sharing/rest/content/items/2af9435fb0f447258c38e8d9609a34cd/info/metadata/metadata.xml?format=default&output=html (accessed 3.17.25).

Oenema, O., et al., 2003. Approaches and uncertainties in nutrient budgets: implications for nutrient management and environmental policies. European Journal of Agronomy, Element Balances as Sustainability Tools 20, 3–16. https://doi.org/10.1016/S1161-0301(03)00067-4

Ortez, O.A., et al., 2018. Exploring Nitrogen Limitation for Historical and Modern Soybean Genotypes. Agronomy Journal 110, 2080–2090. https://doi.org/10.2134/agronj2018.04.0271

Osborne, S.L., Riedell, W.E., 2006. Starter Nitrogen Fertilizer Impact on Soybean Yield and Quality in the Northern Great Plains. https://doi.org/10.2134/agronj2006.0089

Patterson, T.G. and LaRue, T.A., 1983. Nitrogen fixation by soybeans: seasonal and cultivar effects, and comparison of estimates 1. Crop Science, 23(3), pp.488-492.

Peoples, M.B. and Herridge, D.F., 1990. Nitrogen fixation by legumes in tropical and subtropical agriculture. Advances in agronomy, 44, pp.155-223.

Peoples, M.B., et al., 2021. Quantifying country-to-global scale nitrogen fixation for grain legumes: I. Reliance on nitrogen fixation of soybean, groundnut and pulses. Plant Soil 469, 1–14. https://doi.org/10.1007/s11104-021-05167-6

Plummer, M., 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, in: Proceedings of the 3rd International Workshop on Distributed Statistical Computing. Vienna, Austria, pp. 1–10.

Purcell, L., King, A., 1996. Drought and Nitrogen Source Effects on Nitrogen Nutrition, Seed Growth, and Yield in Soybean ,. https://doi.org/10.1080/01904169609365173

Purcell, L.C., 2009. Physiological Responses of N_2 Fixation to Drought and Selecting Genotypes for Improved N2 Fixation, in: Nitrogen Fixation in Crop Production. John Wiley & Sons, Ltd, pp. 211–238. https://doi.org/10.2134/agronmonogr52.c8

Purcell, L.C., et al., 1997. Biomass accumulation and allocation in soybean associated with genotypic differences in tolerance of nitrogen fixation to water deficits. Plant and Soil 196, 101–113. https://doi.org/10.1023/A:1004289609466

Purcell, L.C., et al., 2004. Soybean N2 Fixation Estimates, Ureide Concentration, and Yield Responses to Drought. Crop Science 44, 484–492. https://doi.org/10.2135/cropsci2004.4840

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (accessed 3 February 2023).

R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Raper, C.D., Parsons, L.R., Patterson, D.T., Kramer, P.J., 1977. Relationship between Growth and Nitrogen Accumulation for Vegetative Cotton and Soybean Plants. Bot. Gaz. 138, 129–137. https://doi.org/10.1086/336907

Ray, D.K., et al., 2013. Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLoS One 8, e66428. https://doi.org/10.1371/journal.pone.0066428

Rotundo, L., Westgate, M.E., 2009. Meta-analysis of environmental effects on soybean seed composition 110, 147–156. https://doi.org/10.1016/j.fcr.2008.07.012

Sadras, V.O., et al., 2016. Interactions between water and nitrogen in Australian cropping systems: physiological, agronomic, economic, breeding and modelling perspectives. Crop Pasture Sci. 67, 1019–1053. https://doi.org/10.1071/CP16027

Salvagiotti, F., Cassman, K.G., Specht, J.E., Walters, D.T., Weiss, A., Dobermann, A., 2008. Nitrogen uptake, fixation and response to fertilizer N in soybeans: A review. F. Crop. Res. 108, 1–13. https://doi.org/10.1016/j.fcr.2008.03.001

Salvagiotti, F., Castellarín, J.M., Miralles, D.J., Pedrol, H.M., 2009. Sulfur fertilization improves nitrogen use efficiency in wheat by increasing nitrogen uptake. F. Crop. Res. 113, 170–177. https://doi.org/10.1016/j.fcr.2009.05.003

Salvagiotti, F., et al., 2009. Sulfur fertilization improves nitrogen use efficiency in wheat by increasing nitrogen uptake. Field Crops Research 113, 170–177. https://doi.org/10.1016/j.fcr.2009.05.003

Santachiara, G., Borrás, L., Salvagiotti, F., Gerde, J.A., Rotundo, J.L., 2017. Relative importance of biological nitrogen fixation and mineral uptake in high yielding soybean cultivars. Plant Soil 418, 191–203. https://doi.org/10.1007/s11104-017-3279-9

Santachiara, G., et al., 2019. Nutritional and environmental effects on biological nitrogen fixation in soybean: A meta-analysis. Field Crops Research 240, 106–115. https://doi.org/10.1016/j.fcr.2019.05.006

Santachiara, G., Salvagiotti, F., Gerde, J.A., Rotundo, J.L., 2018. Does biological nitrogen fixation modify soybean nitrogen dilution curves? F. Crop. Res. 223, 171–178. https://doi.org/10.1016/j.fcr.2018.04.001

Santos, C., Salmerón, M., Purcell, L.C., 2019. Soybean Phenology Prediction Tool for the US Midsouth. Agric. Environ. Lett. 4, 190036. https://doi.org/10.2134/ael2019.09.0036

Schipanski, M.E., et al, 2010. Understanding the variability in soybean nitrogen fixation across agroecosystems. Plant Soil 329, 379–397. https://doi.org/10.1007/s11104-009-0165-0

Serraj, R., et al., 1999. Symbiotic N2 fixation response to drought. Journal of Experimental Botany 50, 143–155. https://doi.org/10.1093/jxb/50.331.143

Sinclair, T.R. and De Wit, C.T., 1976. Analysis of the carbon and nitrogen limitations to soybean yield 1. Agronomy Journal, 68(2), pp.319-324.

Sinclair, T.R., de Wit, C.T., 1975. Photosynthate and Nitrogen Requirements for Seed Production by Various Crops. Science (80-.). 189, 565–567. https://doi.org/10.1126/science.189.4202.565

Sinclair, T.R., et al., 1987. Relative Sensitivity of Nitrogen and Biomass Accumulation to Drought in Field-Grown Soybean1. Agronomy Journal 79, 986–991. https://doi.org/10.2134/agronj1987.00021962007900060007x

Sinclair, T.R., et al., 2017. Limited-transpiration response to high vapor pressure deficit in crop species. Plant Science 260, 109–118. https://doi.org/10.1016/j.plantsci.2017.04.007

Smith, P.M.C., Atkins, C.A., 2002. Purine Biosynthesis. Big in Cell Division, Even Bigger in Nitrogen Assimilation. Plant Physiology 128, 793–802. https://doi.org/10.1104/pp.010912

Soil Conservation Service, (SCS), 1972. National Engineering Handbook. Hydrology Section 4, Chapters 4/10.

Sparks, A.H., 2018. nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. Journal of Open Source Software 3, 1035. https://doi.org/10.21105/joss.01035

Tadesse, M.G., Vannucci, M., 2021. Handbook of Bayesian Variable Selection. CRC Press.

Tamagno, S., Balboa, G.R., Assefa, Y., Kovács, P., Casteel, S.N., Salvagiotti, F., García, F.O.,

Stewart, W.M., Ciampitti, I.A., 2017. Nutrient partitioning and stoichiometry in soybean: A synthesis-analysis. F. Crop. Res. 200, 18–27. https://doi.org/10.1016/j.fcr.2016.09.019

Tamagno, S., et al., 2018. Interplay between nitrogen fertilizer and biological nitrogen fixation in soybean: implications on seed yield and biomass allocation. Sci Rep 8, 17502. https://doi.org/10.1038/s41598-018-35672-1

Team, S.D., 2022. Stan: A C++ Library for Probability and Sampling, Version 2.16.0. https//mc-stan.org (acessed 3 February 2023).

Therneau, T., et al., 2013. Rpart: Recursive Partitioning. R Package Version 4.1-3. http://CRAN.R-project.org/package=rpart (accessed 10 October 2024)

Tian, C.F., et al., 2012. Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. Proc. Natl. Acad. Sci. U.S.A. 109, 8629–8634. https://doi.org/10.1073/pnas.1120436109

Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. Proc. Natl. Acad. Sci. 108, 20260–20264. https://doi.org/10.1073/pnas.1116437108

Tremblay, N., Bouroubi, Y.M., Bélec, C., Mullen, R.W., Kitchen, N.R., Thomason, W.E., Ebelhar, S., Mengel, D.B., Raun, W.R., Francis, D.D., Vories, E.D., Ortiz-Monasterio, I., 2012. Corn Response to Nitrogen is Influenced by Soil Texture and Weather. Agron. J. 104, 1658–1671. https://doi.org/10.2134/agronj2012.0184

Unkovich, M., et al., 2008. Measuring Plant-associated Nitrogen Fixation in Agricultural Systems. ACIAR Monograph No.136.

Unkovich, M., Herridge, D., Peoples, M., Cadisch, G., Boddey, R., Giller, K., et al. (2008). *Measuring Plant-Associated Nitrogen Fixation in Agricultural Systems.* Canberra: Australian Centre for International Agricultural Research, 258.

Unkovich, M.J., et al., 1994. Potential precision of the δ15N natural abundance method in field estimates of nitrogen fixation by crop and pasture legumes in south-west Australia. Aust. J. Agric. Res. 45, 119–132. https://doi.org/10.1071/ar9940119

USDA NASS, 2021a. Acreage, ISSN: 1949-1522 Released June 30, 2021, by the National Agricultural Statistics Service (NASS), Agricultural Statistics Board, United States Department of Agriculture (USDA). 1–48. https://www.nass.usda.gov/Publications/Todays_Reports/reports/acrg0621.pdf (acessed 3 February 2023).

USDA NASS, 2021b. 2021 Crop Production Annual Summary. https://www.nass.usda.gov/Newsroom/2022/01-12-2022.php (acessed 3 February 2023).

USDA, 2022. Oilseeds: World Markets and Trade. Foreign Agric. Serv. 1–35. https://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf (acessed 3 February 2023).

USDA, 2023. Soybean 2023 World Production. https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=2222000 (accessed 6.11.23).

USDA, 2024. US farmers expect to plant less corn and more soybean acres. https://www.nass.usda.gov/Newsroom/2024/03-28-2024.php (accessed 4.30.24).

van Kessel, C., Hartley, C., 2000. Agricultural management of grain legumes: has it led to an increase in nitrogen fixation? Field Crops Research 65, 165–181. https://doi.org/10.1016/S0378-4290(99)00085-4

Wakefield, J., 2013. Bayesian Inference, in: Wakefield, J. (Ed.), Bayesian and Frequentist Regression Methods. Springer, New York, NY, pp. 85–151. https://doi.org/10.1007/978-1-4419-0925-1_3

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org (accessed 3 February 2023).

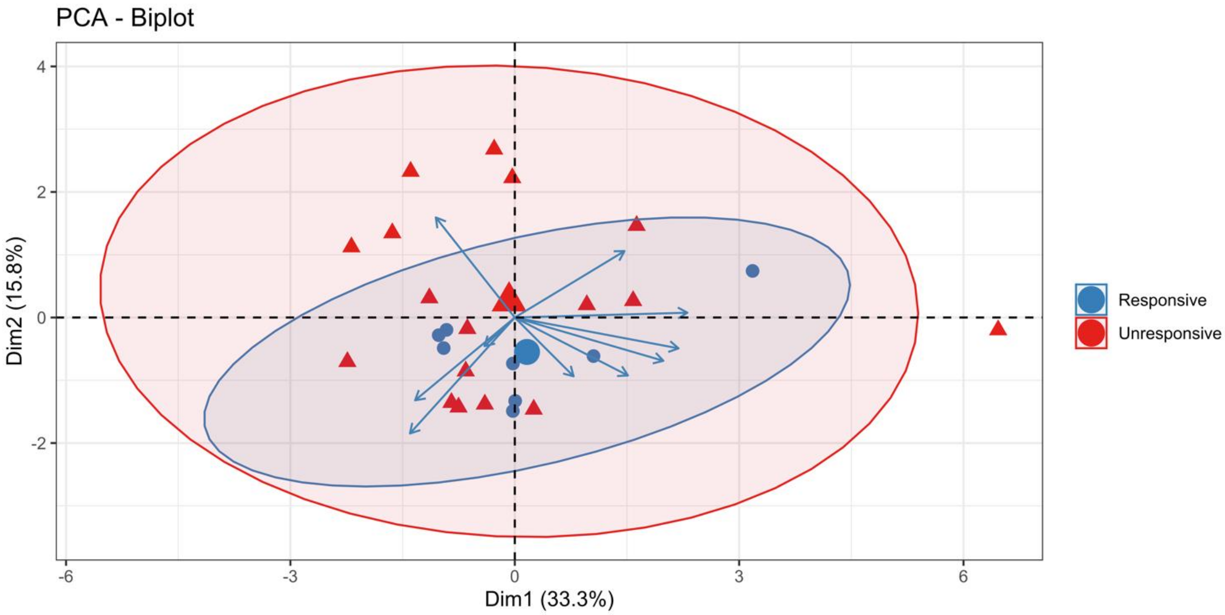
Wickham, H., Averick, M., Bryan, J., Chang, W., Mcgowan, L.D.A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin, T., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., 2019. Welcome to the Tidyverse Tidyverse package 4, 1–6. https://doi.org/10.21105/joss.01686

Wood, S.N., 2017. Generalized Additive Models: An Introduction with R, Second Edition, 2nd ed. Chapman and Hall/CRC, Boca Raton. https://doi.org/10.1201/9781315370279

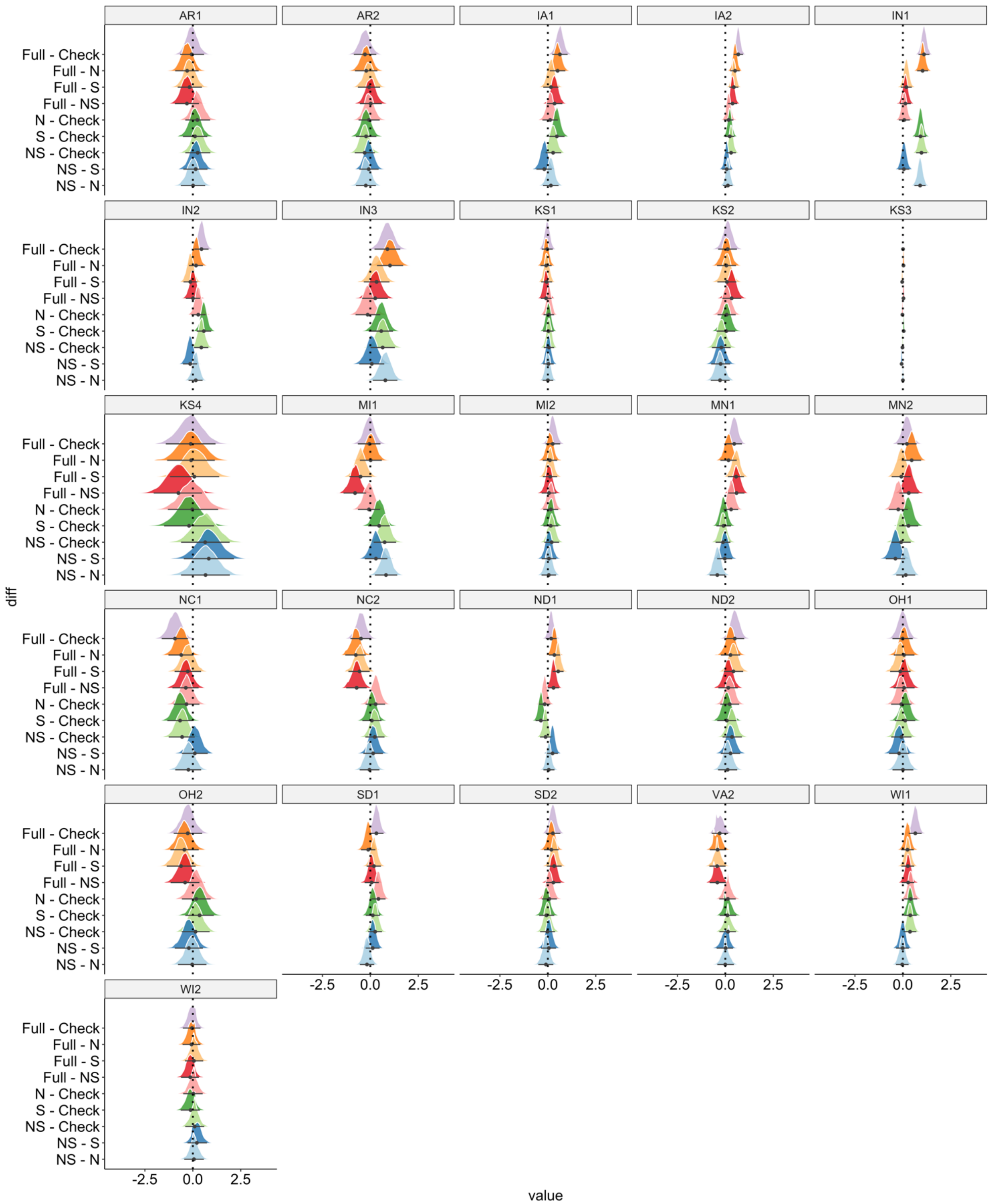
Zapata, F., Danso, S.K.A., Hardarson, G., Fried, M., 1987. Time Course of Nitrogen Fixation in Field-Grown Soybean Using Nitrogen-15 Methodology 1. Agron. J. 79, 172–176. https://doi.org/10.2134/agronj1987.00021962007900010035x

Zheng, H., et al., 2009. Classification and regression tree (CART) for analysis of soybean yield variability among fields in Northeast China: The importance of phosphorus application rates under drought conditions. Agriculture, Ecosystems & Environment 132, 98–105. https://doi.org/10.1016/j.agee.2009.03.004

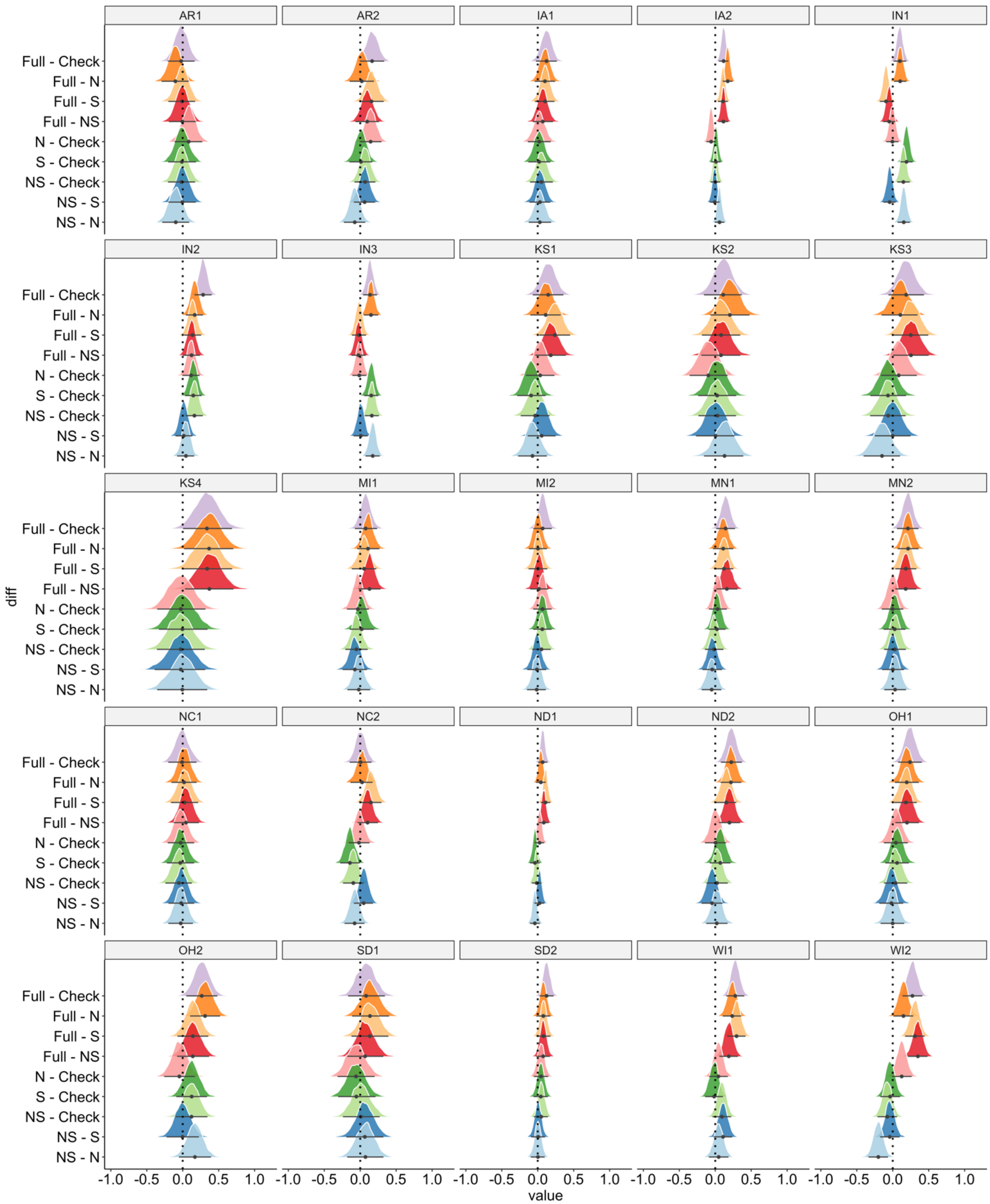
Appendix A - Chapter 2 Supplementary Materials



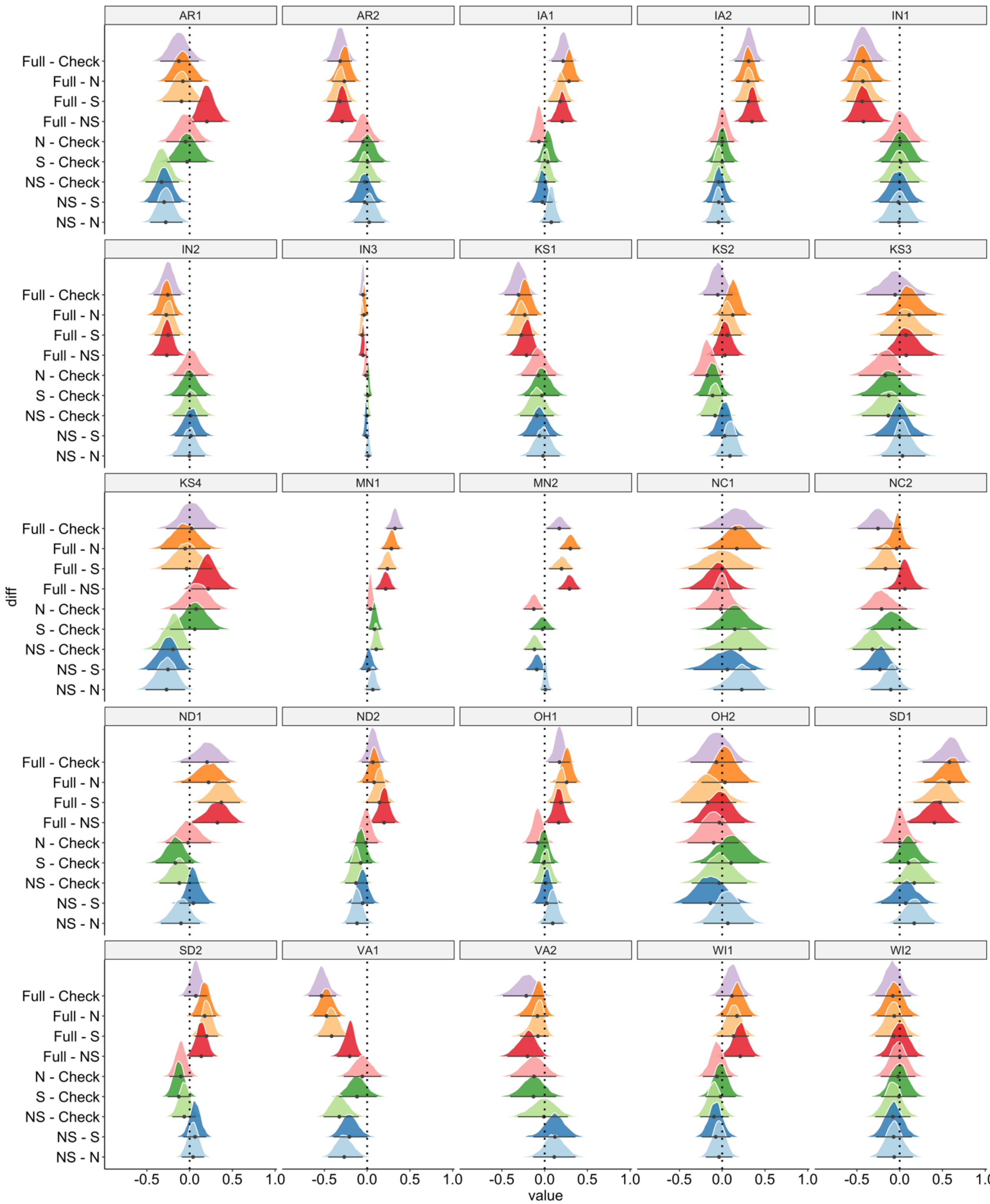
Appendix Figure A.1. Principal component analysis of the two types of response to fertilization. Different colors and ellipses represent the two groups of response.



Appendix Figure A.2. Treatment pairwise comparisons of posterior distributions for each of the sites in the yield analysis.



Appendix Figure A.3. Treatment pairwise comparisons of posterior distributions for each of the sites in the NNI analysis.



Appendix Figure A.4. Treatment pairwise comparisons of posterior distributions for each of the sites in the Ndfa analysis.

Appendix Table A.1 Site-year descriptors, including geographic region, soil type, and management information.

State	Site	City	Soil Order	Soil Subgroup	Soybean Variety	Seeding rate	Row spacing	Tillage system	Irrigated or Dryland	Planting	Harvesting
						seeds/ha	cm			mm/dd/yy	
AR	1	Newport	Dundee	Typic Endoaqualfs	P49T62E	345800	76.2	Conventional	Irrigated	5/27/21	10/20/21
	2	Colt	Dexter	Typic Glossaqualfs	P49T62E	345800	76.2	Conventional	Irrigated	6/16/21	10/26/21
IA	1	Ames	Clarion	Typic Hapludolls	P28T14E	345000	76.2	Field cultivator in the spring	Rainfed	5/6/21	10/13/21
	2	Sutherland	Galva	Typic Hapludolls	P23A15	345000	76.2	Soil finisher in the spring	Rainfed	5/10/21	10/6/21
IN	1	West Lafayette	Drummer	Typic Endoaquolls	P34T21SE	-	-	-	-	5/25/21	10/11/21
	2	Wannatah	Palms	Terric Haplosaprists	P30T99	-	-	-	-	5/20/21	9/30/21
	3	LaCrosse	Maumee	Typic Endoaquolls	P26T23E	-	-	-	-	5/5/21	9/28/21
KS	1	Topeka	Eudora	Fluventic Hapludolls	AG40X70	350000	76.2	Vertical tilled corn stalks	Irrigated	5/12/21	10/4/21
	2	Kiro	Rossville	Cumulic Hapludolls	AG40X70	350000	76.2	No-Till	Rainfed	5/12/21	10/8/21
	3	Scandia	Crete	Pachic Udertic Argiustolls	P39A45X	350000	76.2	No-Till	Irrigated	5/12/21	10/12/21
	4	Ashland Bottoms	Belvue	Typic Udifluvents	P39A45X	320000	76.2	No-Till	Rainfed	5/30/21	10/8/21
MI	1	Campus	Riddles	Typic Hapludalfs	GH2011E3	320000	38.1	-	-	5/25/21	10/1/21
	2	SVREC	Tappan	Typic Epiaquolls				-	-	5/25/21	10/1/21
MN	1	Saint Paul	Waukegan	Typic Hapludolls	XO2181E	345000	76.2	Conventional	Irrigated	4/30/21	9/22/21
	2	Minnesota Lake	Marna	Vertic Endoaquolls	XO2181E	345000	76.2	Conventional	Rainfed	5/5/21	10/10/21
NC	1	North Carolina	Goldsboro	Aquic Paleudults	AG56X8	321100	91	Conventional	Rainfed	6/17/21	11/2/21
	2	Roper	Portsmouth	Typic Umbraquults	AG56X8	321100	76.2	No-Till	Irrigated	6/18/21	10/21/21
ND	1	Fargo	Fargo	Typic Epiaquerts	0819R2X	457000	30.5	No-Till	Rainfed	5/10/21	9/29/21
	2	Lisbon	Barnes	Calcic Hapludolls	0819R2X	457000	30.5	Conventional	Rainfed	5/7/21	9/28/21
OH	1	South Charleston	Kokomo	Typic Argiaquolls	P34T21SE	370500	38.1	Minimum Till	Rainfed	5/21/21	11/9/21
	2	South Charleston	Crosby	Aeric Epiaqualfs	P34T21SE	370500	38.1	Minimum Till	Rainfed	5/21/21	11/9/21
SD	1	Aurora	Brandt	Calcic Hapludolls	P18T26R2	370000	76.2	Conventional	Rainfed	5/21/21	9/30/21
	2	Brookings	Vienna	Calcic Hapludolls	P18T26R2	370000	76.2	Conventional	Rainfed	5/21/21	10/12/21
VA	1	Suffolk	Kenansville	Arenic Hapludults	P49T62E	346000	38.1	No-Till	Rainfed	5/6/21	
	2	Painter	Bojac	Typic Hapludults	-	-	-	No-Till	Rainfed		
WI	1	Arlington	Plano	Typic Argiudolls	Renk RS219NX	345680	76.2	Chisel plow soil finisher	Rainfed	4/29/21	9/28/21
	2	Stratford	Loyal	Oxyaquic Glossudalfs	AG14X8	345680	76.2	Chisel plow soil finisher	Rainfed	5/7/21	10/19/21

Appendix Table A.2. Initial soil characterization by site, at 0-15 cm, 0-20 cm, and 20-61 cm depths.

State	Site	City	pH	CEC	SOM	Sand	Silt	Clay	P	K	Ca	Mg	Na	NO_3	NH^4	SO_4^{-2}	NO_3	NH^4	SO_4^{-2}
					------------- % -------------				---------- mg kg-1 ----------										
			0 - 15 cm											*0 - 20 cm*			*20 - 61 cm*		
AR	1	Newport	6	8.5	1.5	56	35	9	112.1	87.4	724.3	128.9	17.3	-	-	-	9.8	2	4.7
	2	Colt	7.6	10.2	1.9	28	59	13	22.2	98.2	1712.1	265.9	46.5	-	-	-	7.8	2.6	19
IA	1	Ames	6.4	28.7	4.7	42	34	24	42.8	268.2	3444.8	491.3	9.3	5.3	5.6	16.6	21.5	9.2	3.5
	2	Sutherland	5.9	32.2	4.9	16	52	32	19	212.4	3157.9	598.6	20.8	13.2	4.2	3.5	8.1	2.7	6.1
IN	1	West Lafayette	6.42	20.76	3.12	-	-	-	63.4	206.2	697	2447.4	-	-	-	9.4	-	-	-
	2	Wannatah	5.78	17.14	2.66	-	-	-	45	138.6	1777.4	427.8	-	-	-	18.2	-	-	-
	3	LaCrosse	6.16	12.24	2.16	-	-	-	29.8	186.4	1399.4	311.4	-	-	-	11	-	-	-
KS	1	Topeka	6.77	11.7	2	32	56	12	16.7	228	1629.6	168	5.3	6.1	4.9	3.6	5	4.5	3.2
	2	Kiro	5.49	23.4	3	16	62	22	27.3	419.7	2077.1	186.2	5.6	8.9	6.8	5.3	9.9	6.4	4.2
	3	Scandia	7.5	9.8	1.2	46	42	12	46	186.4	1658.8	124.3	6.7	3.1	3.1	2	3.2	2.7	2.1
	4	Ashland Bottoms	6	9.5	1.4	56	36	8	36.1	152.7	847.9	110.4	5.1	3.1	2.2	2.4	2.1	1.87	1.8
MI	1	Campus	6.1	9.9	2.1	66	22	12	88.2	83.3	699.9	145.2	3.4	9.5	3.6	2.5	-	-	-
	2	SVREC	7.9	35.2	2.1	34	34	32	11	169.8	4013.5	468.5	8.4	13.9	7.6	5.6	-	-	-
MN	1	Saint Paul	5.1	23.4	3.5	26	54	20	98.5	179.8	1581.4	231.8	9.3	12.5	1.6	3.1	11.5	1.6	6
	2	Minnesota Lake	6	32.1	5.2	34	36	30	32.3	197	3315	643.5	10.1	7.4	3.8	2.3	5.8	2.8	4
NC	1	Rocky Mount	6.8	-	0.9	78	14	8	48	70	277	38	4	3.9	3.2	13.8	-	-	-
	2	Roper	6.1	-	12	79	17	4	62	205	1274	88	21	11.3	4	5.5	-	-	-
ND	1	Fargo	7.6	37.7	4.8	34	14	52	77.6	610.1	4369.5	1519.4	53.5	2.2	3	82.5	3.3	2.6	30
	2	Lisbon	7.7	28.4	3.9	36	34	30	41.7	258.4	4154.4	810	26.5	1.5	2.5	32.5	1.3	2.8	6.1
OH	1	South Charleston	7.1	19.8	3	18	58	24	67	149.2	2161	559.8	71.9	6.6	5.9	3.4	3.8	4.2	2.3
	2	South Charleston	6.5	19	2.8	22	54	24	38	174.6	1923.5	511.5	13.9	17.5	11.2	4	12.5	6.8	3
SD	1	Aurora	6.4	28.5	4.4	26	48	26	12.5	176.2	2881.1	706.6	14	5.8	3.1	7.9	6.4	4.1	6.3
	2	Brookings	7.3	25.9	4.8	34	42	24	22.3	176.2	3221.6	868.8	27.9	13.2	5.2	7.6	7	2.6	7.2
VA	1	Suffolk	5.9	11.2	1.7	62	22	16	95.2	160.5	538.9	83.9	4.9	-	-	5.6	6.2	10.2	6
	2	Painter	6.1	7.7	1.9	80	12	8	87.8	84.5	478.3	69.2	2.7	-	-	16.3	3.7	6.1	52.9
WI	1	Arlington	6.2	20.7	3.8	20	60	20	42.2	242.8	1742.6	496.6	5.6	7.8	4.1	2.9	9.1	4.3	2.7
	2	Stratford	6.7	17.9	3.9	28	60	12	78.3	143.6	1671.2	466.5	38.5	16.3	3.9	8.7	8.3	5.2	8.1

Appendix Table A.3. Response and uncertainty magnitudes by site and treatment pairwise comparison.

Site	Comparison	Response			Uncertainty		
		Seed Yield (Mg ha^{-1})	NNI	Ndfa (%)	Seed Yield (Mg ha^{-1})	NNI	Ndfa (%)
AR1	NS - N	0.0121	-0.0969	-27.9680	1.2656	0.3745	38.0986
AR1	NS - S	0.1481	-0.0054	-29.8276	1.2598	0.3809	37.9232
AR1	NS - Check	0.2533	-0.0164	-32.8617	1.3051	0.3761	39.0759
AR1	S - Check	0.1132	-0.0118	-2.9986	1.2692	0.3883	46.5059
AR1	N - Check	0.2392	0.0831	-4.7878	1.2939	0.3848	45.5691
AR1	Full - NS	-0.3088	-0.0054	20.1279	1.2544	0.3763	35.3912
AR1	Full - S	-0.1726	-0.0087	-9.5734	1.2639	0.3710	43.8973
AR1	Full - N	-0.2976	-0.1022	-7.9798	1.2270	0.3778	43.3586
AR1	Full - Check	-0.0544	-0.0213	-12.5848	1.2504	0.3732	45.2990
AR2	NS - N	-0.2366	-0.0763	2.3693	1.0376	0.2975	35.4764
AR2	NS - S	-0.0834	0.0610	-2.9162	1.0052	0.2939	35.4874
AR2	NS - Check	-0.2974	0.0675	-2.4763	1.0468	0.2900	35.3136
AR2	S - Check	-0.2237	0.0096	0.5053	1.0654	0.3167	36.9787
AR2	N - Check	-0.0734	0.1443	-4.8545	1.1021	0.3037	35.7563
AR2	Full - NS	0.0164	0.0969	-29.2775	1.1398	0.3280	26.9444
AR2	Full - S	-0.0650	0.1563	-31.9055	1.1591	0.3372	28.3258
AR2	Full - N	-0.2147	0.0214	-26.6834	1.1715	0.3370	27.7159
AR2	Full - Check	-0.2816	0.1650	-31.5882	1.1712	0.3411	28.7973
IA1	NS - N	0.1583	0.0289	7.7503	0.8112	0.3101	20.9659
IA1	NS - S	-0.1847	0.0311	-2.6660	0.7985	0.2912	23.7886
IA1	NS - Check	0.2735	0.0513	1.1261	0.8115	0.3174	22.9860
IA1	S - Check	0.4684	0.0200	3.6664	0.8029	0.3041	23.4278
IA1	N - Check	0.1199	0.0219	-6.7376	0.8051	0.3181	20.6380
IA1	Full - NS	0.3478	0.0699	20.5415	0.8022	0.2801	25.8502
IA1	Full - S	0.1614	0.1024	18.0632	0.7958	0.2716	26.0362
IA1	Full - N	0.5076	0.0999	28.3987	0.8095	0.2837	24.3047
IA1	Full - Check	0.6287	0.1217	21.6453	0.8010	0.2902	25.6553
IA2	NS - N	0.1299	0.0591	-4.4234	0.4923	0.1349	27.7906
IA2	NS - S	0.0676	-0.0072	-4.0540	0.4838	0.1358	27.5224
IA2	NS - Check	0.2977	0.0030	-4.1844	0.4977	0.1340	27.1874
IA2	S - Check	0.2304	0.0099	-0.0268	0.4816	0.1352	27.9100
IA2	N - Check	0.1682	-0.0571	0.2792	0.4955	0.1323	27.6534
IA2	Full - NS	0.3733	0.1139	34.8986	0.4774	0.1340	27.7990
IA2	Full - S	0.4406	0.1066	30.7666	0.4740	0.1366	28.3002
IA2	Full - N	0.5049	0.1733	30.3394	0.4874	0.1361	27.3049
IA2	Full - Check	0.6728	0.1166	30.8380	0.4936	0.1335	29.2397
IN1	NS - N	0.8989	0.1529	-0.9095	0.5814	0.1720	45.7391
IN1	NS - S	0.0518	-0.0432	-1.5728	0.5696	0.1726	45.2371
IN1	NS - Check	0.9699	0.1479	-0.2118	0.5668	0.1751	45.9986
IN1	S - Check	0.9172	0.1915	1.4362	0.5712	0.1710	45.5660
IN1	N - Check	0.0718	-0.0049	0.6777	0.5753	0.1709	46.8707
IN1	Full - NS	0.1308	-0.0495	-42.3226	0.5638	0.1720	40.6975
IN1	Full - S	0.1842	-0.0937	-43.9072	0.5723	0.1753	41.3105
IN1	Full - N	1.0290	0.1029	-43.1025	0.5725	0.1736	40.5364
IN1	Full - Check	1.1025	0.0989	-42.3355	0.5684	0.1778	40.6152
IN2	NS - N	0.1533	0.0448	-0.3462	0.7320	0.2374	37.0800
IN2	NS - S	-0.1369	0.0142	1.5329	0.7362	0.2341	37.3863
IN2	NS - Check	0.4420	0.1624	1.1421	0.7428	0.2377	40.7568
IN2	S - Check	0.5737	0.1468	-0.3164	0.7547	0.2328	40.4443
IN2	N - Check	0.2827	0.1170	1.6772	0.7396	0.2383	40.5710

IN2	Full - NS	0.0085	0.1222	-26.7129	0.7496	0.2367	29.5808
IN2	Full - S	-0.1295	0.1368	-25.2507	0.7216	0.2376	29.3142
IN2	Full - N	0.1648	0.1664	-27.3077	0.7276	0.2394	29.5575
IN2	Full - Check	0.4506	0.2837	-25.6025	0.7466	0.2386	31.3448
IN3	NS - N	0.7818	0.1737	1.2095	1.2855	0.1980	7.4245
IN3	NS - S	0.0764	0.0078	-1.2202	1.3002	0.1980	7.1801
IN3	NS - Check	0.6429	0.1613	-0.1285	1.2674	0.1958	7.5026
IN3	S - Check	0.5680	0.1539	1.0889	1.2678	0.1993	7.2608
IN3	N - Check	-0.1393	-0.0130	-1.3587	1.2863	0.1946	7.6112
IN3	Full - NS	0.2524	-0.0242	-4.8240	1.3365	0.1993	8.0734
IN3	Full - S	0.3251	-0.0164	-6.0547	1.3406	0.1962	7.7452
IN3	Full - N	1.0242	0.1500	-3.6183	1.3633	0.2010	8.0080
IN3	Full - Check	0.8886	0.1364	-4.9338	1.3453	0.1951	8.1528
KS1	NS - N	0.0006	-0.0732	-1.8666	0.5531	0.3990	38.6466
KS1	NS - S	0.0309	0.0549	-6.1174	0.5420	0.3866	38.8240
KS1	NS - Check	0.0577	-0.0365	-9.1105	0.5460	0.4000	39.4204
KS1	S - Check	0.0292	-0.0921	-3.2293	0.5426	0.3951	40.7022
KS1	N - Check	0.0552	0.0341	-7.2682	0.5537	0.3941	39.9769
KS1	Full - NS	-0.0918	0.1810	-21.2103	0.5663	0.4253	31.0227
KS1	Full - S	-0.0623	0.2376	-27.3907	0.5705	0.4304	32.6795
KS1	Full - N	-0.0877	0.1107	-23.2656	0.5816	0.4213	31.2398
KS1	Full - Check	-0.0330	0.1453	-30.5561	0.5779	0.4249	32.1988
KS2	NS - N	-0.2789	0.1293	9.1184	0.9817	0.5463	30.0958
KS2	NS - S	-0.2482	0.0042	2.8761	0.9932	0.5361	32.3155
KS2	NS - Check	-0.2090	0.0315	-8.3146	0.9765	0.5231	32.7264
KS2	S - Check	0.0446	0.0270	-11.2402	0.9767	0.5286	33.1887
KS2	N - Check	0.0673	-0.0958	-17.5145	0.9459	0.5281	31.9563
KS2	Full - NS	0.3271	0.0797	3.1785	0.9882	0.5397	33.0772
KS2	Full - S	0.0784	0.0808	5.9406	0.9969	0.5330	32.3638
KS2	Full - N	0.0515	0.2030	12.4714	0.9917	0.5448	31.6867
KS2	Full - Check	0.1195	0.1109	-5.1308	0.9787	0.5258	33.7374
KS3	NS - N	0.0017	-0.1493	3.3133	0.0977	0.5074	50.6153
KS3	NS - S	-0.0638	-0.0014	-0.5073	0.0992	0.4962	56.0639
KS3	NS - Check	-0.0209	-0.0644	-13.2766	0.0992	0.4925	62.6768
KS3	S - Check	0.0426	-0.0663	-12.8365	0.0983	0.4959	64.0311
KS3	N - Check	-0.0232	0.0823	-16.7960	0.0963	0.4983	61.8295
KS3	Full - NS	0.0295	0.2537	7.6848	0.0988	0.4937	60.7991
KS3	Full - S	-0.0347	0.2508	7.3706	0.0983	0.4887	61.0957
KS3	Full - N	0.0309	0.1057	11.0480	0.0963	0.4987	59.1215
KS3	Full - Check	0.0083	0.1866	-5.3583	0.0983	0.4938	68.7302
KS4	NS - N	0.6624	-0.0052	-27.0355	2.5419	0.6969	46.2977
KS4	NS - S	0.8319	-0.0268	-25.3625	2.6141	0.7039	44.7753
KS4	NS - Check	0.6426	-0.0310	-19.3826	2.5303	0.6652	42.5321
KS4	S - Check	-0.1991	-0.0041	5.9688	2.6038	0.6776	56.9367
KS4	N - Check	-0.0292	-0.0261	7.5853	2.6215	0.6710	58.6255
KS4	Full - NS	-0.7515	0.3704	21.7561	2.5229	0.6753	44.5313
KS4	Full - S	0.0773	0.3397	-3.5131	2.5644	0.6921	59.1868
KS4	Full - N	-0.0911	0.3667	-5.2827	2.4877	0.6852	57.9741
KS4	Full - Check	-0.1115	0.3364	2.2120	2.5929	0.6814	58.0736
MI1	NS - N	0.8153	-0.0196	NA	1.1622	0.3159	NA
MI1	NS - S	0.2906	-0.0749	NA	1.1771	0.3403	NA
MI1	NS - Check	0.7463	-0.0526	NA	1.1760	0.3108	NA
MI1	S - Check	0.4597	0.0200	NA	1.1304	0.3387	NA
MI1	N - Check	-0.0716	-0.0335	NA	1.1597	0.3068	NA
MI1	Full - NS	-0.7978	0.1286	NA	1.1800	0.3067	NA
MI1	Full - S	-0.5119	0.0542	NA	1.1554	0.3408	NA

MI1	Full - N	0.0141	0.1085	NA	1.1445	0.3100	NA
MI1	Full - Check	-0.0516	0.0755	NA	1.1753	0.3089	NA
MI2	NS - N	0.0503	-0.0142	NA	0.7569	0.2616	NA
MI2	NS - S	0.0550	-0.0119	NA	0.7565	0.2620	NA
MI2	NS - Check	0.1986	0.0525	NA	0.7709	0.2598	NA
MI2	S - Check	0.1469	0.0635	NA	0.7730	0.2554	NA
MI2	N - Check	0.1521	0.0667	NA	0.7760	0.2565	NA
MI2	Full - NS	0.0591	0.0168	NA	0.7635	0.2552	NA
MI2	Full - S	0.1188	0.0045	NA	0.7477	0.2606	NA
MI2	Full - N	0.1078	0.0033	NA	0.7587	0.2515	NA
MI2	Full - Check	0.2557	0.0683	NA	0.7563	0.2570	NA
MN1	NS - N	-0.4202	-0.0496	6.6175	0.8109	0.2735	16.4604
MN1	NS - S	-0.0295	-0.0420	1.9797	0.7895	0.2690	17.9504
MN1	NS - Check	-0.1236	-0.0177	10.6231	0.7746	0.2576	14.6236
MN1	S - Check	-0.0964	0.0241	8.6031	0.7618	0.2359	13.3647
MN1	N - Check	0.2982	0.0322	3.9581	0.7661	0.2475	10.2130
MN1	Full - NS	0.5804	0.1613	21.6423	0.8473	0.2934	21.3902
MN1	Full - S	0.5593	0.1206	23.7161	0.8468	0.2689	21.1653
MN1	Full - N	0.1630	0.1109	28.3590	0.8527	0.2836	20.3054
MN1	Full - Check	0.4603	0.1445	32.5002	0.8198	0.2627	18.8627
MN2	NS - N	0.1504	0.0318	0.7620	0.9870	0.3016	11.1399
MN2	NS - S	-0.3892	0.0013	-9.2602	1.0078	0.3035	21.4157
MN2	NS - Check	-0.0927	0.0292	-11.9245	1.0054	0.3021	22.3596
MN2	S - Check	0.2924	0.0278	-2.4789	0.9953	0.3028	27.2854
MN2	N - Check	-0.2389	-0.0019	-12.7325	0.9913	0.2993	22.0309
MN2	Full - NS	0.3066	0.1801	29.0087	1.0013	0.2942	23.6221
MN2	Full - S	-0.0805	0.1812	19.7447	0.9897	0.2925	27.2528
MN2	Full - N	0.4587	0.2126	30.0348	1.0139	0.2915	23.1187
MN2	Full - Check	0.2107	0.2107	17.1077	1.0012	0.3001	27.9272
NC1	NS - N	-0.2311	-0.0269	22.8777	1.3515	0.3417	60.5727
NC1	NS - S	0.1033	-0.0213	5.9780	1.3466	0.3573	73.2289
NC1	NS - Check	-0.5600	-0.0528	21.1613	1.3513	0.3693	72.3994
NC1	S - Check	-0.6660	-0.0343	14.7218	1.3142	0.3778	65.2480
NC1	N - Check	-0.3311	-0.0280	-1.4444	1.3425	0.3673	52.2892
NC1	Full - NS	-0.3666	0.0454	-5.3855	1.3657	0.3500	60.7942
NC1	Full - S	-0.2647	0.0253	-0.4382	1.3792	0.3790	75.3922
NC1	Full - N	-0.6032	0.0190	17.1583	1.3062	0.3492	62.0159
NC1	Full - Check	-0.9316	-0.0085	15.2538	1.3666	0.3930	73.6699
NC2	NS - N	-0.0454	-0.0780	-10.4701	1.0282	0.2771	42.8904
NC2	NS - S	0.1466	0.0478	-23.1065	1.0174	0.2644	42.7360
NC2	NS - Check	0.2308	-0.0961	-31.9131	1.0002	0.2875	50.1133
NC2	S - Check	0.0788	-0.1439	-8.5210	0.9981	0.2721	54.2544
NC2	N - Check	0.2784	-0.0175	-21.1817	1.0151	0.2892	53.6931
NC2	Full - NS	-0.7146	0.1019	6.1751	1.0704	0.2830	38.2738
NC2	Full - S	-0.5639	0.1487	-16.5886	1.1142	0.2948	43.4971
NC2	Full - N	-0.7582	0.0249	-3.5612	1.0646	0.3081	40.9999
NC2	Full - Check	-0.4886	0.0055	-25.3567	1.0733	0.3037	49.4222
ND1	NS - N	0.0472	-0.0404	-10.1123	0.5773	0.1371	45.6150
ND1	NS - S	0.2443	0.0246	4.3408	0.5733	0.1364	34.7425
ND1	NS - Check	-0.1193	-0.0145	-12.0515	0.5890	0.1362	46.6652
ND1	S - Check	-0.3628	-0.0387	-16.6695	0.5775	0.1362	44.1988
ND1	N - Check	-0.1673	0.0257	-1.8645	0.5766	0.1306	52.7621
ND1	Full - NS	0.2936	0.0836	32.5030	0.5875	0.1373	52.5487
ND1	Full - S	0.5379	0.1080	37.0483	0.5708	0.1348	49.3684
ND1	Full - N	0.3408	0.0433	22.0888	0.5816	0.1346	57.5153
ND1	Full - Check	0.1722	0.0688	20.3466	0.5703	0.1327	56.2375

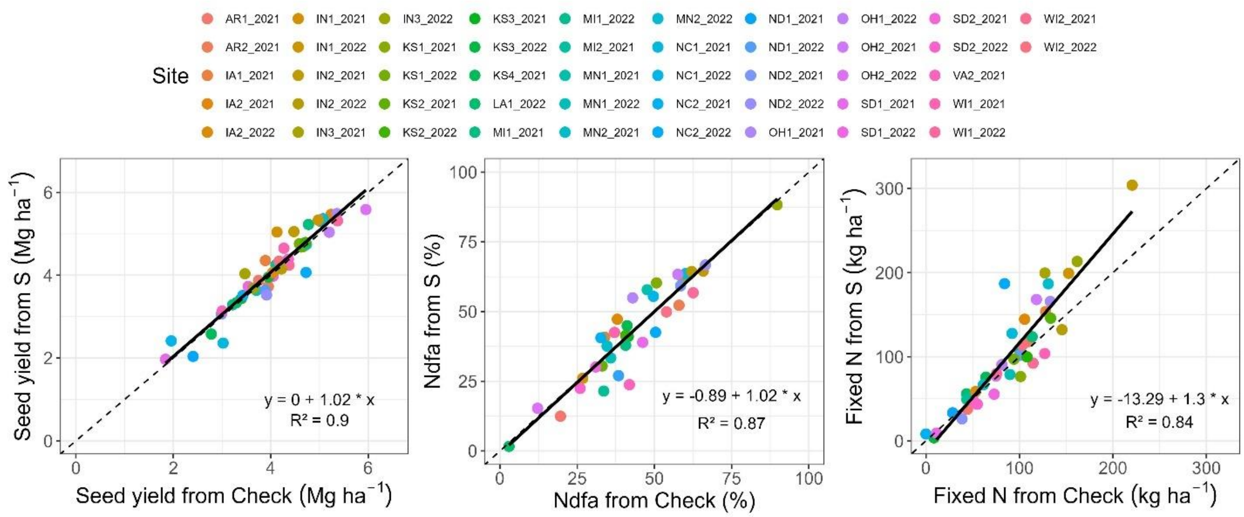
ND2	NS - N	0.1197	0.0207	-11.8237	0.9951	0.2965	26.4440
ND2	NS - S	0.2687	-0.0452	-5.3494	1.0153	0.2913	26.3645
ND2	NS - Check	0.3448	0.0258	-13.0422	1.0087	0.2974	26.1890
ND2	S - Check	0.0763	0.0704	-7.5527	1.0011	0.2885	26.1144
ND2	N - Check	0.2277	0.0040	-1.3394	0.9861	0.2936	26.7697
ND2	Full - NS	0.1511	0.1976	19.7742	1.0102	0.2916	25.9161
ND2	Full - S	0.4203	0.1535	14.4107	1.0031	0.2871	26.5079
ND2	Full - N	0.2682	0.2181	8.0855	1.0078	0.2879	25.9798
ND2	Full - Check	0.4939	0.2231	6.6437	0.9871	0.2914	25.7542
OH1	NS - N	0.0029	-0.0013	9.3990	1.0873	0.3291	25.5892
OH1	NS - S	-0.1968	-0.0189	2.5244	1.0579	0.3315	25.6966
OH1	NS - Check	-0.0749	0.0393	1.2176	1.0576	0.3220	25.8558
OH1	S - Check	0.1222	0.0599	-1.2404	1.0641	0.3274	25.4775
OH1	N - Check	-0.0806	0.0435	-8.1370	1.0594	0.3337	25.6118
OH1	Full - NS	0.0630	0.1982	16.2565	1.0934	0.3283	24.7476
OH1	Full - S	-0.1363	0.1823	18.8921	1.0656	0.3177	24.5437
OH1	Full - N	0.0621	0.1944	25.7368	1.0735	0.3331	24.6656
OH1	Full - Check	-0.0137	0.2393	17.3855	1.0594	0.3288	24.9141
OH2	NS - N	-0.0331	0.1697	6.6208	1.4856	0.4457	58.5976
OH2	NS - S	-0.2068	-0.0034	-13.6983	1.4669	0.4488	67.9766
OH2	NS - Check	0.1394	0.1223	-3.3701	1.4589	0.4396	64.5776
OH2	S - Check	0.3504	0.1245	10.3996	1.4861	0.4280	68.6895
OH2	N - Check	0.1774	-0.0469	-9.8810	1.4996	0.4278	60.1692
OH2	Full - NS	-0.4053	0.1399	-3.1988	1.4984	0.4445	59.7279
OH2	Full - S	-0.6175	0.1412	-17.0610	1.5030	0.4283	64.9143
OH2	Full - N	-0.4430	0.3107	3.1976	1.4791	0.4159	55.0074
OH2	Full - Check	-0.2653	0.2645	-6.7436	1.4832	0.4322	60.6563
SD1	NS - N	-0.1751	0.0728	16.9800	0.7480	0.5040	47.7749
SD1	NS - S	0.1215	0.0642	6.9217	0.7293	0.5147	54.8906
SD1	NS - Check	0.2504	0.0088	16.9778	0.7231	0.5112	48.6195
SD1	S - Check	0.1254	-0.0548	10.2745	0.7361	0.5199	46.6314
SD1	N - Check	0.4211	-0.0589	0.0678	0.7346	0.5143	39.1546
SD1	Full - NS	0.0625	0.0686	40.5497	0.7410	0.5009	52.6822
SD1	Full - S	0.1893	0.1333	47.4517	0.7457	0.5125	52.8860
SD1	Full - N	-0.1099	0.1356	58.0696	0.7506	0.5135	47.7643
SD1	Full - Check	0.3146	0.0814	58.1919	0.7424	0.5111	49.8287
SD2	NS - N	-0.0991	0.0030	4.0148	0.7823	0.1935	26.2533
SD2	NS - S	0.0503	0.0052	6.2668	0.7741	0.1941	26.8643
SD2	NS - Check	-0.0252	0.0483	-6.2871	0.7781	0.1992	27.4825
SD2	S - Check	-0.0747	0.0430	-12.7011	0.7815	0.1986	27.3181
SD2	N - Check	0.0763	0.0455	-10.2429	0.7879	0.2010	26.9361
SD2	Full - NS	0.2880	0.0753	13.5788	0.7567	0.1905	27.3883
SD2	Full - S	0.3316	0.0802	19.8493	0.7723	0.1940	27.3751
SD2	Full - N	0.1834	0.0778	17.5462	0.7677	0.1934	27.2188
SD2	Full - Check	0.2624	0.1227	7.2508	0.7656	0.1959	27.1300
VA2	NS - N	-0.0061	NA	11.2001	0.8649	NA	49.5758
VA2	NS - S	0.0115	NA	11.9758	0.8560	NA	49.3171
VA2	NS - Check	0.1336	NA	-0.9597	0.8668	NA	58.7985
VA2	S - Check	0.1066	NA	-13.0491	0.8645	NA	53.5231
VA2	N - Check	0.1286	NA	-12.4376	0.8594	NA	53.3852
VA2	Full - NS	-0.4179	NA	-20.0420	0.8505	NA	42.0361
VA2	Full - S	-0.4288	NA	-7.8611	0.8485	NA	33.3329
VA2	Full - N	-0.4052	NA	-8.4653	0.8586	NA	33.2242
VA2	Full - Check	-0.2965	NA	-21.5121	0.8571	NA	48.5643
WI1	NS - N	-0.0323	0.0477	-3.6653	0.6601	0.2729	31.3294
WI1	NS - S	-0.0132	0.1089	-7.5732	0.6687	0.2580	33.7754

WI1	NS - Check	0.3716	0.0919	-9.5450	0.6688	0.2601	33.5873
WI1	S - Check	0.3848	-0.0153	-1.8790	0.6528	0.2480	35.2025
WI1	N - Check	0.4069	0.0435	-6.0331	0.6775	0.2612	34.2446
WI1	Full - NS	0.2759	0.1884	21.1458	0.6555	0.2574	35.2238
WI1	Full - S	0.2593	0.2964	13.4100	0.6657	0.2539	37.0710
WI1	Full - N	0.2408	0.2365	17.3103	0.6607	0.2596	36.0957
WI1	Full - Check	0.6487	0.2791	11.4885	0.6749	0.2499	36.6861
WI2	NS - N	0.0589	-0.2004	-6.4187	1.0218	0.2712	41.9201
WI2	NS - S	0.2105	-0.0421	-7.0749	1.0185	0.2710	42.0992
WI2	NS - Check	0.0984	-0.0769	-7.9911	0.9929	0.2694	41.3975
WI2	S - Check	-0.1183	-0.0365	-0.8603	0.9988	0.2747	41.8031
WI2	N - Check	0.0293	0.1244	-1.6717	1.0174	0.2753	40.6979
WI2	Full - NS	-0.1344	0.3496	0.1180	0.9294	0.2734	41.6847
WI2	Full - S	0.0571	0.3091	-7.1063	0.9501	0.2707	42.9765
WI2	Full - N	-0.0709	0.1487	-6.3320	0.9306	0.2754	41.8124
WI2	Full - Check	-0.0444	0.2734	-8.0854	0.9234	0.2738	41.4176

Appendix B - Chapter 3 Supplementary Materials

Appendix Table B.1 Initial soil characterization by site and season, at 0-15 cm, 0-20 cm, and 20-60 cm.

		0 – 15 cm											0 – 20 cm			20 – 60 cm		
Site	Season	pH	CEC	SOM	Sand	Silt	Clay	P	K	Ca	Mg	Na	NO_3	NH^4	SO_4^{-2}	NO_3	NH^4	SO_4^{-2}
				--------------- % ---------------				--------------------- mg kg^{-1} ---------------------					-------- mg kg^{-1} --------			-------- mg kg^{-1} --------		
IA1	2021	6.4	28.7	4.7	42.0	34.0	24.0	42.8	268.2	3444.8	491.3	9.3	5.3	5.6	-	21.5	9.2	-
IA2	2021	5.9	32.2	4.9	16.0	52.0	32.0	19.0	212.4	3157.9	598.6	20.8	13.2	4.2	-	8.1	2.7	-
IN1	2021	6.4	20.8	3.1	16.5	48.7	27.0	63.4	206.2	2447.4	697.0	-	-	-	9.4	-	-	-
IN2	2021	5.8	17.1	2.7	43.4	34.9	18.0	45.0	138.6	1777.4	427.8	-	-	-	18.2	-	-	-
IN3	2021	6.2	12.2	2.2	71.3	17.7	10.5	29.8	186.4	1399.4	311.4	-	-	-	11	-	-	-
KS1	2021	6.8	11.7	2.0	32.0	56.0	12.0	16.7	228.0	1629.6	168.0	5.3	6.1	4.9	3.6	5	4.5	3.2
KS2	2021	5.5	23.4	3.0	16.0	62.0	22.0	27.3	419.7	2077.1	186.2	5.7	8.9	6.8	5.3	9.9	6.4	4.2
KS3	2021	7.5	9.8	1.2	46.0	42.0	12.0	46.0	186.4	1658.8	124.3	6.7	3.1	3.1	2	3.2	2.7	2.1
KS4	2021	6.0	9.5	1.4	56.0	36.0	8.0	36.1	152.7	847.9	110.4	5.1	3.1	2.2	2.4	2.1	1.87	1.8
MI1	2021	6.1	9.9	2.1	66.0	22.0	12.0	88.2	83.3	699.9	145.2	3.4	9.5	3.6	2.5	-	-	-
MN1	2021	5.1	23.4	3.5	26.0	54.0	20.0	98.5	179.8	1581.4	231.8	9.3	12.5	1.6	-	11.5	1.6	-
MN2	2021	6.0	32.1	5.2	34.0	36.0	30.0	32.3	197.0	3315.0	643.5	10.1	7.4	3.8	-	5.8	2.8	-
ND2	2021	7.7	28.4	3.9	36.0	34.0	30.0	41.7	258.4	4154.4	810.0	26.5	1.5	2.5	-	1.3	2.8	-
OH1	2021	7.1	19.8	3.0	18.0	58.0	24.0	67.0	149.2	2161.0	559.8	71.9	6.6	5.9	3.4	3.8	4.2	2.3
OH2	2021	6.5	19.0	2.8	22.0	54.0	24.0	38.0	174.6	1923.5	511.5	13.9	17.5	11.2	4	12.5	6.8	3
SD1	2021	6.4	28.5	4.4	26.0	48.0	26.0	12.5	176.2	2881.1	706.6	14.0	5.8	3.1	-	6.4	4.1	-
SD2	2021	7.3	25.9	4.8	34.0	42.0	24.0	22.3	176.2	3221.6	868.8	27.9	13.2	5.2	-	7	2.6	-
WI1	2021	6.2	20.7	3.8	20.0	60.0	20.0	42.2	242.8	1742.6	496.6	5.6	7.8	4.1	-	9.1	4.3	-
WI2	2021	6.7	17.9	3.9	28.0	60.0	12.0	78.3	143.6	1671.2	466.5	38.5	16.3	3.9	-	8.3	5.2	-
OH1	2022	5.5	24.1	4.4	12.0	56.0	32.0	66.0	153.0	2226.0	383.0	6.0	8.9	6.8	4.7	4.9	3.4	6.2
OH2	2022	6.8	11.5	2.9	18.0	61.0	21.0	176.0	104.0	1794.0	269.0	6.0	9.1	4.6	7	4.5	2.9	5.6
WI2	2022	7.3	13.5	4.1	18.0	64.0	18.0	49.0	177.0	1761.0	506.0	8.0	14.3	4.1	24.7	7.7	4	21.1
WI1	2022	6.5	12.4	3.3	10.0	67.0	23.0	22.0	176.0	1624.0	456.0	17.0	13.4	14.3	5.2	14.4	8.7	2.4
SD1	2022	6.5	18.2	3.9	18.0	60.0	22.0	27.0	215.0	2585.0	565.0	14.0	4	4.1	6	9	1.2	9
SD2	2022	7.1	24.5	4.3	19.0	52.0	29.0	11.0	155.0	3084.0	1030.0	34.0	9.5	5.9	5	14	2.6	26
ND1	2022	7.9	32.5	4.0	12.0	43.0	45.0	41.0	348.0	4190.0	1269.0	28.0	6.3	6.3	5	3.3	4.2	5.3
ND2	2022	6.9	21.1	4.1	28.0	40.0	32.0	33.0	294.0	2859.0	723.0	9.0	2.8	5.3	7.9	4.6	3.3	4.9
MN1	2022	5.9	18.1	3.1	20.0	56.0	24.0	87.0	135.0	1672.0	303.0	9.0	10.6	7.2	5.5	11.3	4.9	2.5
MN2	2022	5.5	27.1	4.7	31.0	38.0	31.0	17.0	125.0	2643.0	461.0	8.0	9.8	10.3	7.6	9.6	8.1	6.7
KS1	2022	6.6	8.7	1.8	39.8	46.6	13.6	35.0	173.2	1427.4	131.6	6.6	5.98	6.18	8.48	4.26	2.84	4.4
KS2	2022	6.8	7.9	1.6	45.6	42.4	12.0	39.2	138.2	1317.0	120.2	5.6	8.18	5.36	9.14	5.1	3.16	5.16
KS3	2022	6.2	25.6	3.4	10.0	59.0	31.0	53.0	276.0	2906.0	438.0	19.0	5	6.7	8	4.8	6.1	6.5
IA2	2022	6.0	27.1	4.5	12.5	56.0	31.5	27.0	109.0	2981.5	547.5	6.0	14.65	7.4	3.3	-	-	-
IN1	2022	6.4	22.0	3.5	16.5	48.7	27.0	36.4	117.6	2717.6	722.0	-	-	-	6.6	-	-	-
IN3	2022	6.8	10.6	2.0	71.3	17.7	10.5	37.8	112.8	1411.4	335.2	-	-	-	10.4	-	-	-



Appendix Figure B.1 Relationship between Check and S treatments for seed yield (Mg ha^{-1}), Ndfa (%), and fixed-N (kg ha^{-1}) across sites (dashed line represents the 1:1 line).

Appendix C - Chapter 4 Supplementary Materials

Appendix Table C.1. Study characterization by site × year, sowing date, soybean maturity group (MG), and initial soil characteristics measured (pH, soil organic matter – OM, phosphorus – P, potassium – K, cation exchange capacity – CEC, soil texture, nitrate – NO3-, and ammonium – NH4+). Soil nitrate and ammonium were collected at 0.6 m depth, while other soil variables at 0.15 m.

Site × year	Sowing date	MG	pH	OM (%)	P (ppm)	K (ppm)	CEC	Sand (%)	Silt (%)	Clay (%)	NO_3^- (ppm)	NH_4^+ (ppm)
IA1 (2021)	05/06/2021	2.8	6.4	4.7	42.8	268.2	28.7	42	34	24	15.9	7.9
IA2 (2021)	05/10/2021	2.3	5.9	4.9	19.0	212.4	32.2	16	52	32	9.7	3.2
IN1 (2021)	05/25/2021	3.4	6.42	3.12	63.4	206.2	20.8	17	49	27	-	-
IN3 (2021)	05/05/2021	2.6	6.16	2.16	29.8	186.4	12.2	71	18	11	-	-
IN2 (2021)	05/20/2021	3	5.78	2.66	45.0	138.6	17.1	43	35	18	-	-
KS1 (2021)	05/12/2021	4	6.77	2	16.7	228.0	11.7	32	56	12	5.3	4.6
KS2 (2021)	05/12/2021	4	5.49	3	27.3	419.7	23.4	16	62	22	9.5	6.5
KS3 (2021)	05/12/2021	3.9	7.5	1.2	46.0	186.4	9.8	46	42	12	3.1	2.8
KS4 (2021)	05/30/2021	3.9	6	1.4	36.1	152.7	9.5	56	36	8	2.4	2.0
NC1 (2021)	06/17/2021	5.6	6.8	0.9	48.0	70.0	-	78	14	8	-	-
ND1 (2021)	05/10/2021	0.8	7.6	4.8	77.6	610.1	37.7	34	14	52	2.9	2.7
ND2 (2021)	05/07/2021	0.8	7.7	3.9	41.7	258.4	28.4	36	34	30	1.4	2.7
OH1 (2021)	05/21/2021	3.4	7.1	3	67.0	149.2	19.8	18	58	24	4.7	4.7
OH2 (2021)	05/21/2021	3.4	6.5	2.8	38.0	174.6	19.0	22	54	24	14.0	8.2
SD1 (2021)	05/21/2021	1.8	6.4	4.4	12.5	176.2	28.5	26	48	26	6.1	3.7
SD2 (2021)	05/21/2021	1.8	7.3	4.8	22.3	176.2	25.9	34	42	24	9.0	3.4
AR1 (2021)	05/27/2021	4.9	6	1.5	112.1	87.4	8.5	56	35	9	-	-
AR2 (2021)	06/16/2021	4.9	7.6	1.9	22.2	98.2	10.2	28	59	13	-	-
MN1 (2021)	04/30/2021	2.1	5.1	3.5	98.5	179.8	23.4	26	54	20	11.7	1.6
MN2 (2021)	05/05/2021	2.1	6	5.2	32.3	197.0	32.1	34	36	30	6.3	3.1
WI1 (2021)	04/29/2021	2.1	6.2	3.8	42.2	242.8	20.7	20	60	20	8.6	4.2
WI2 (2021)	05/07/2021	1.4	6.7	3.9	78.3	143.6	17.9	28	60	12	10.9	4.7
IA2 (2022)	05/13/2022	2.2	6	4.5	27.0	109.0	27.1	13	56	32	-	-
IN1 (2022)	05/12/2022	3.4	6.44	3.54	36.4	117.6	22.0	17	49	27	-	-

IN3 (2022)	05/12/2022	3	6.8	2.04	37.8	112.8	10.6	71	18	11	-	-
KS1 (2022)	05/09/2022	3.9	6.62	1.84	35.0	173.2	8.7	40	47	14	4.8	3.9
KS2 (2022)	06/13/2022	3.9	6.8	1.64	39.2	138.2	7.9	46	42	12	6.1	3.9
KS3 (2022)	06/20/2022	3.9	6.2	3.4	53.0	276.0	25.6	10	59	31	4.8	6.2
ND1 (2022)	05/25/2022	0.7	7.9	4	41.0	348.0	32.5	12	43	45	4.3	4.9
ND2 (2022)	05/24/2022	0.9	6.9	4.1	33.0	294.0	21.1	28	40	32	4.0	3.9
OH1 (2022)	05/12/2022	3.4	5.5	4.4	66.0	153.0	24.1	12	56	32	6.2	4.5
OH2 (2022)	05/12/2022	3.4	6.8	2.9	176.0	104.0	11.5	18	61	21	6.0	3.4
SD1 (2022)	05/21/2022	1.3	6.5	3.9	27.0	215.0	18.2	18	60	22	7.3	2.1
SD2 (2022)	05/27/2022	1.3	7.1	4.3	11.0	155.0	24.5	19	52	29	12.4	3.7
MN1 (2022)	05/05/2022	1.8	5.9	3.1	87.0	135.0	18.1	20	56	24	11.0	5.6
MN2 (2022)	05/31/2022	1.8	5.5	4.7	17.0	125.0	27.1	31	38	31	9.6	8.7
WI1 (2022)	05/16/2022	2.1	6.5	3.3	22.0	176.0	12.4	10	67	23	13.9	10.5
WI2 (2022)	05/06/2022	2.1	7.3	4.1	49.0	177.0	13.5	18	64	18	9.8	4.0

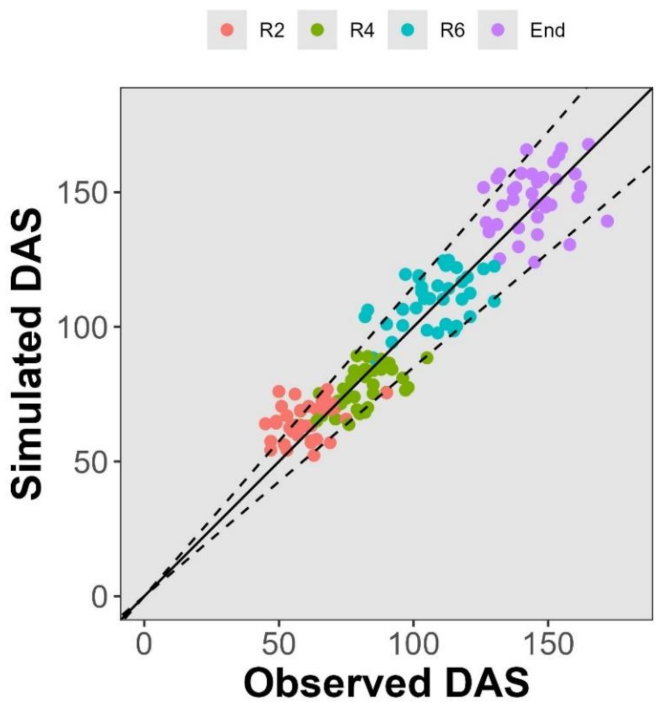
Appendix Table C.2 Growing season weather characterization across the 38 experiments (site-year combinations).

ID	Mean maximum air temperature (°C)	Mean minimum air temperature (°C)	Mean average air temperature (°C)	Seasonal water supply (mm)	Mean vapor pressure deficit (kPa)
AR1 (2021)	30.61	18.48	24.54	439.42	0.66
AR2 (2021)	31.13	19.69	25.41	337.31	0.75
IA1 (2021)	26.42	13.08	19.75	386.33	0.89
IA2 (2021)	25.49	11.87	18.68	365.00	0.89
IA2 (2022)	26.39	12.27	19.33	269.75	0.95
IN1 (2021)	28.16	15.89	22.03	506.48	0.72
IN1 (2022)	27.90	15.06	21.48	409.19	0.76
IN2 (2021)	26.23	14.22	20.22	530.86	0.69
IN3 (2021)	25.01	12.93	18.97	618.24	0.64
IN3 (2022)	25.73	13.04	19.39	388.11	0.66
KS1 (2021)	28.74	16.07	22.40	663.70	0.77
KS1 (2022)	29.76	16.32	23.04	562.86	1.00

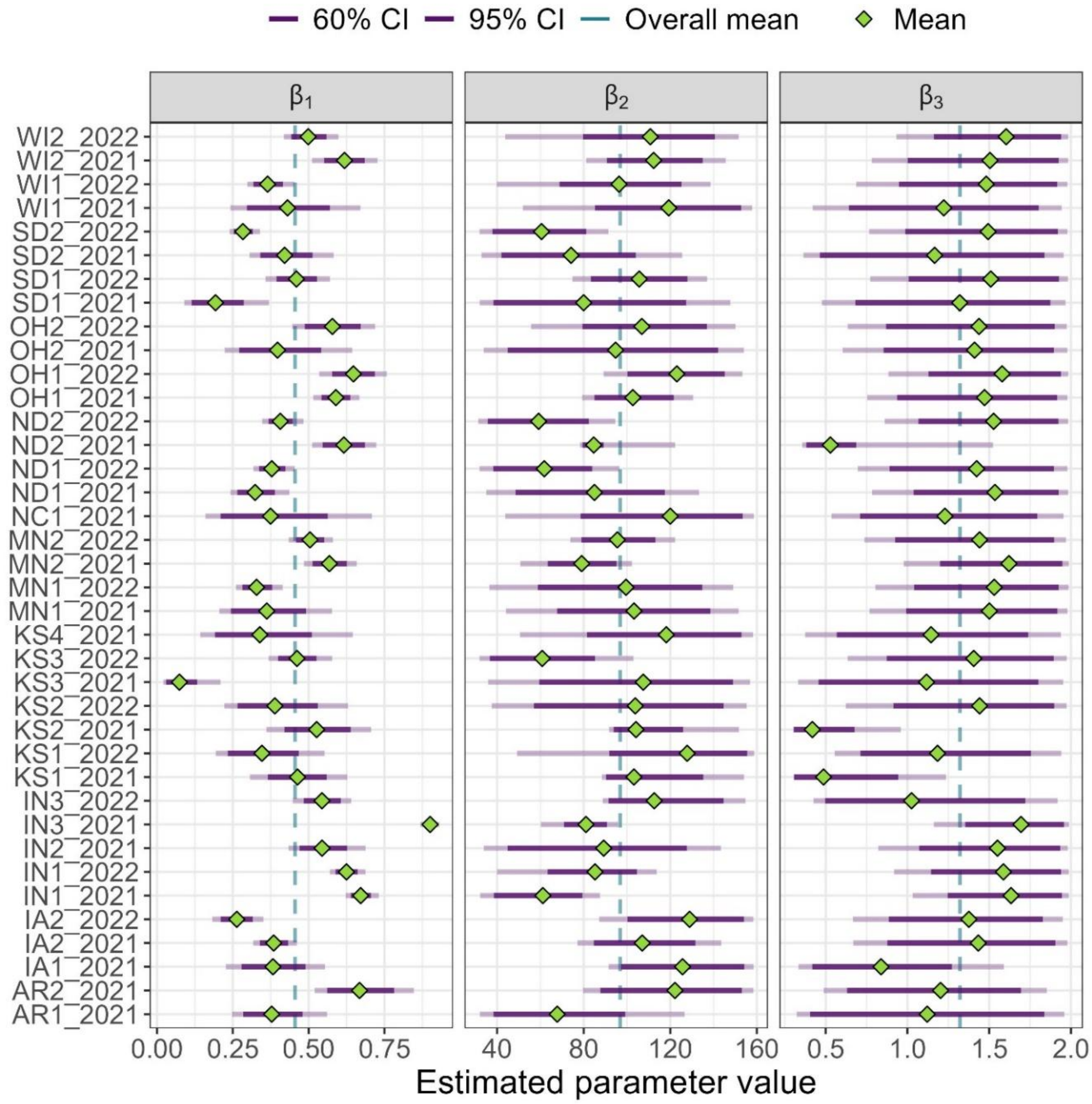
KS2 (2021)	28.75	16.12	22.44	663.70	0.77
KS2 (2022)	30.53	16.31	23.42	434.34	1.09
KS3 (2021)	28.07	14.80	21.43	386.59	0.97
KS3 (2022)	30.63	16.63	23.63	462.53	1.27
KS4 (2021)	29.41	16.87	23.14	504.95	1.03
MN1 (2021)	24.67	12.61	18.64	244.86	0.84
MN1 (2022)	24.51	13.34	18.92	399.29	0.76
MN2 (2021)	24.33	11.72	18.02	394.72	0.82
MN2 (2022)	25.11	12.75	18.93	414.27	0.77
NC1 (2021)	28.92	17.07	22.99	669.80	0.64
ND1 (2021)	24.26	9.78	17.02	403.35	0.91
ND1 (2022)	24.07	10.86	17.46	382.78	0.70
ND2 (2021)	24.35	10.16	17.26	295.15	0.98
ND2 (2022)	24.21	11.27	17.74	334.01	0.77
OH1 (2021)	26.52	14.64	20.58	632.71	0.64
OH1 (2022)	26.39	14.34	20.37	544.83	0.59
OH2 (2021)	26.52	14.64	20.58	632.71	0.64
OH2 (2022)	26.39	14.34	20.37	544.83	0.59
SD1 (2021)	25.97	12.00	18.98	316.99	0.92
SD1 (2022)	25.10	11.70	18.40	403.61	0.80
SD2 (2021)	26.05	11.91	18.98	344.17	0.92
SD2 (2022)	26.09	12.29	19.19	325.37	0.84
WI1 (2021)	23.68	10.89	17.29	451.36	0.57
WI1 (2022)	22.99	11.13	17.06	608.33	0.56
WI2 (2021)	22.90	9.42	16.16	749.81	0.36
WI2 (2022)	23.45	11.40	17.42	653.29	0.54

Appendix Table C.3. Parameters influencing seasonal N2 fixation (Ndfa, %) across 38 site × year combinations (ID). The table details β1 (peak of Ndfa), β2 (timing of the peak), and β3 (growth and decay rate) for each ID throughout the season. It includes median estimates (Q500) and corresponding 95% (Q025 and Q975), and 60% (Q200 and Q800) credible intervals.

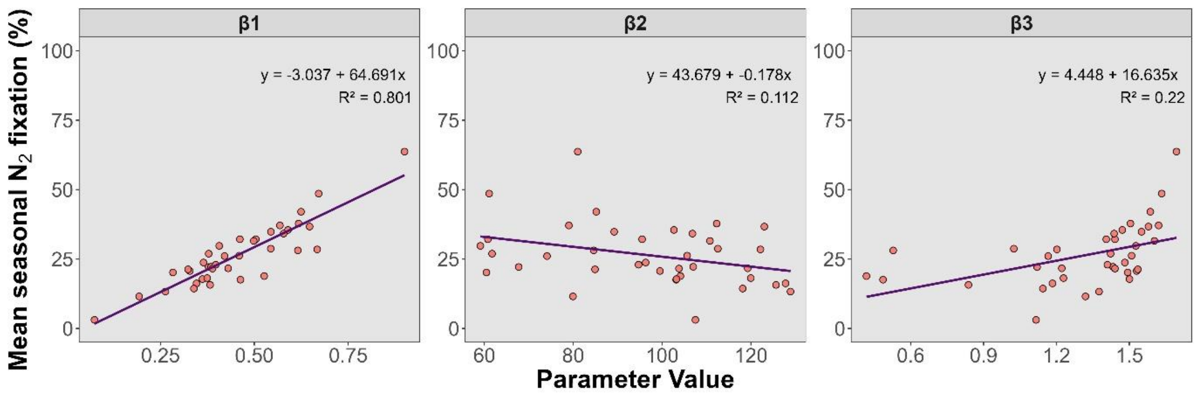
ID	Q025			Q100			Q500			Q900			Q975		
	$β_1$	$β_2$	$β_3$	$β_1$	$β_2$	$β_3$	$β_1$	$β_2$	$β_3$	$β_1$	$β_2$	$β_3$	$β_1$	$β_2$	$β_3$
AR1 (2021)	0.248	32	0.323	0.285	38.305	0.404	0.378	67.739	1.12	0.48	99.228	1.837	0.561	126.532	1.959
AR2 (2021)	0.52	79.67	0.487	0.561	87.674	0.63	0.668	122.153	1.202	0.782	153.074	1.694	0.847	158.385	1.852
IA1 (2021)	0.228	91.522	0.333	0.279	97.264	0.417	0.382	125.652	0.838	0.49	154.179	1.272	0.554	158.454	1.59
IA2 (2021)	0.317	77.083	0.668	0.339	84.661	0.877	0.385	107.029	1.433	0.432	131.613	1.904	0.461	143.549	1.977
IA2 (2022)	0.182	87.16	0.666	0.21	100.166	0.886	0.263	128.908	1.375	0.316	154.043	1.829	0.351	158.389	1.95
IN1 (2021)	0.622	32.142	1.032	0.639	38.451	1.246	0.671	61.098	1.633	0.705	79.373	1.944	0.731	87.521	1.986
IN1 (2022)	0.57	39.895	0.918	0.589	63.311	1.145	0.625	85.191	1.587	0.662	104.553	1.94	0.687	113.694	1.986
IN2 (2021)	0.434	33.883	0.821	0.469	44.866	1.072	0.544	89.225	1.55	0.626	127.622	1.934	0.688	143.377	1.98
IN3 (2021)	0.869	60.162	1.161	0.881	70.954	1.353	0.9	81.02	1.694	0.92	90.751	1.957	0.93	95.768	1.988
IN3 (2022)	0.448	88.744	0.425	0.484	91.45	0.496	0.544	112.564	1.025	0.606	144.528	1.72	0.64	154.742	1.918
KS1 (2021)	0.307	88.37	0.301	0.365	90.358	0.306	0.463	103.178	0.486	0.559	135.245	0.943	0.627	154.069	1.236
KS1 (2022)	0.194	49.158	0.554	0.233	91.803	0.712	0.346	127.793	1.184	0.467	155.497	1.755	0.552	158.759	1.941
KS2 (2021)	0.361	91.685	0.301	0.42	93.807	0.305	0.526	104.133	0.418	0.639	125.901	0.676	0.705	151.738	0.959
KS2 (2022)	0.222	37.419	0.623	0.265	56.98	0.914	0.389	103.803	1.441	0.53	144.657	1.897	0.629	155.306	1.973
KS3 (2021)	0.021	35.833	0.331	0.03	59.368	0.456	0.074	107.505	1.116	0.133	148.942	1.8	0.21	156.85	1.952
KS3 (2022)	0.368	31.803	0.634	0.399	36.583	0.873	0.461	60.808	1.405	0.526	85.269	1.894	0.578	102.981	1.975
KS4 (2021)	0.143	50.518	0.373	0.191	81.483	0.567	0.339	118.139	1.144	0.51	152.877	1.738	0.645	158.164	1.938
MN1 (2021)	0.206	44.081	0.767	0.244	67.714	0.992	0.362	103.236	1.501	0.491	138.507	1.917	0.576	151.5	1.979
MN1 (2022)	0.261	36.308	0.804	0.281	58.86	1.04	0.328	99.519	1.53	0.38	134.794	1.925	0.414	149.087	1.985
MN2 (2021)	0.485	50.684	0.976	0.513	63.357	1.198	0.568	79.024	1.621	0.626	95.207	1.947	0.658	102.273	1.986
MN2 (2022)	0.434	73.722	0.737	0.459	78.97	0.924	0.504	95.514	1.441	0.551	113.111	1.897	0.58	122.153	1.97
NC1 (2021)	0.16	43.732	0.535	0.21	78.486	0.709	0.374	119.977	1.229	0.563	153.461	1.794	0.709	158.592	1.954
ND1 (2021)	0.242	34.974	0.783	0.265	48.426	1.038	0.323	84.91	1.535	0.388	117.424	1.924	0.436	133.27	1.983
ND1 (2022)	0.318	31.825	0.695	0.337	38.216	0.891	0.378	61.744	1.423	0.424	83.91	1.895	0.454	96.421	1.976
ND2 (2021)	0.512	78.326	0.355	0.546	79.448	0.381	0.616	84.616	0.528	0.686	89.181	0.686	0.723	122.234	1.524
ND2 (2022)	0.347	31.274	0.861	0.367	35.642	1.067	0.406	59.124	1.527	0.448	82.36	1.923	0.483	94.576	1.982
OH1 (2021)	0.515	79.214	0.754	0.542	84.972	0.937	0.59	102.67	1.471	0.637	121.658	1.915	0.667	130.51	1.979
OH1 (2022)	0.535	89.18	0.883	0.577	100.153	1.129	0.647	123.018	1.579	0.718	145.154	1.938	0.757	153.293	1.984
OH2 (2021)	0.223	33.752	0.602	0.271	44.908	0.854	0.397	94.703	1.41	0.541	142.121	1.895	0.644	153.998	1.977
OH2 (2022)	0.446	55.643	0.635	0.487	79.41	0.869	0.578	106.843	1.438	0.671	136.978	1.9	0.719	150.11	1.975
SD1 (2021)	0.09	31.895	0.476	0.114	38.297	0.681	0.193	79.959	1.319	0.286	127.298	1.874	0.369	147.737	1.966
SD1 (2022)	0.358	74.739	0.772	0.394	83.285	1.008	0.46	105.65	1.509	0.527	127.934	1.926	0.57	137.069	1.981
SD2 (2021)	0.305	32.78	0.363	0.341	41.985	0.464	0.421	74.111	1.166	0.513	104.05	1.838	0.582	125.443	1.955
SD2 (2022)	0.239	31.845	0.766	0.253	37.733	0.986	0.283	60.471	1.493	0.316	81.095	1.919	0.339	91.362	1.979
WI1 (2021)	0.242	51.924	0.42	0.296	85.089	0.641	0.43	119.28	1.222	0.57	152.787	1.802	0.671	157.801	1.943
WI1 (2022)	0.298	39.846	0.686	0.318	68.813	0.947	0.365	96.314	1.481	0.415	125.123	1.916	0.453	138.574	1.978
WI2 (2021)	0.511	81.093	0.782	0.551	90.715	1.002	0.618	112.293	1.505	0.685	134.966	1.924	0.727	145.563	1.982
WI2 (2022)	0.419	43.662	0.932	0.443	79.686	1.161	0.499	110.77	1.603	0.559	140.581	1.94	0.598	151.586	1.984



Appendix Figure C.1 Observed and simulated phenology stages (full-bloom: R2; full-pod: R4; full-seed: R6; and crop harvest) across different locations. The crop phenology simulation was performed using the APSIM Next-Generation and the DSSAT v4.8.2. The accuracy of the simulated crop phenology stages (from R2 to Harvest) was assessed against observed data, showing a root mean square error (RMSE) of 11 days and a relative root mean square error (RRMSE) of 12%. Boundary dashed lines represent ± 15% deviation of the 1:1 solid line.



Appendix Figure C.2 Parameters influencing seasonal N_2 fixation (Ndfa, %) in non-linear models across 38 site-year combinations, which includes β_1 (peak of Ndfa), β_2 (timing of the peak), and β_3 (growth and decay rate) throughout the season. The green diamond indicates the median estimate, while the purple shading represents the posterior distribution credible intervals at 95% (lightest), 60% (darkest) levels.



Appendix Figure C.3. Relationship between mean seasonal Ndfa (nitrogen derived from the atmosphere, %) and model parameters including β_1 (peak of Ndfa), β_2 (timing of the peak), and β_3 (growth and decay rate).

Appendix Equation C.1.

$$y_{ij} | \beta_1, \beta_2, \beta_3, \phi_j \sim beta(z_{ij} \cdot \phi_j, (1 - z_{ij}) \cdot \phi_j) \quad \text{(C1a)}$$

$$z_{ij} = \beta_{1j} \cdot e^{\left(-\left(\frac{DAE_{ij} - \beta_{2j}}{DAE_{ij} \cdot \beta_{3j}}\right)^2\right)} \quad (C1b)$$

$$\beta_{1j} \sim beta(1, 1) \quad (C1c)$$

$$\beta_{2j} \sim uniform(30, 160) \quad (C1d)$$

$$\beta_{3j} \sim uniform(0.3, 2) \quad (C1e)$$

$$\phi_j \sim uniform(1, 60) \quad (C1f)$$

Here the parametrization of $beta(z_{ij} \cdot \phi_j, (1 - z_{ij}) \cdot \phi_j)$ is the standard parameterization (i.e., $beta(\alpha, \beta)$) where $E(y_{ij} \mid \beta_1, \beta_2, \beta_3, \phi_j) = z_{ij}$ and $Var(y_{ij} \mid \beta_1, \beta_2, \beta_3, \phi_j) = \frac{z_{ij}(1 - z_{ij})}{1 + \phi_j}$.

The data model (likelihood) **(C1a)** was assumed to be a beta distribution, where y_{ij} represents the Ndfa observation at the time i in location j. The term z_{ij} is the conditional expected value of the beta distribution, while ϕ_j is a dispersion parameters that scales the conditional variance of the beta distribution. The expected value z_{ij} **(C1b)**, was modeled by a non-linear function implementing DAE as the predictor variable. The parameter β_1 refers to the maximum value or peak of Ndfa, β_2 describes the time (units of DAE) of the peak and β_3 define the rate of growth and decay of the function.

To accommodate all possible values for the peak of N_2 fixation, a non-informative beta distribution with a uniform probability density between 0 and 1 was assumed **(C1c)**. For the timing of the peak, prior information from Ciampitti et al. (2021), Córdova et al. (2019), and Zapata et al. (1987) was used to define a uniform distribution with specified minimum and maximum hyperparameters **(C1d)**. Similarly, a uniform distribution was chosen for the growth and decay rates (Ciampitti et al., 2021) to accommodate a wide range of possible shapes for seasonal N_2 fixation **(C1e)**. Lastly, a uniform distribution was also assumed for the precision parameter, enabling the model to account for different degrees of variability in the observations **(S1f)**.

Appendix Equation C.2.

$$y_i|\beta_0, \beta_1, \sigma^2 \sim normal(\mu_i, \sigma^2) \qquad (C2a)$$

$$\mu_i = \beta_0 + \beta_1 \cdot x_i \qquad (C2b)$$

$$\beta_0, \beta_1 \sim normal(0, 10^6) \qquad (C2c)$$

$$\sigma^2 \sim uniform(0, 10^6) \qquad (C2d)$$

The data model considered a normal distribution for the i^{th} observation of both seed yield and protein and can be written as in **C2a**. For the process model, μ_i was defined as a linear effect where β_1 is the linear parameter (slope), and x_i the calculated mean seasonal N_2 fixation of each location

(C2b). For the parameter models, non-informative and normal distributed priors were defined for the intercept (β_0) and slope (β_1) **(C2c)**, and a uniform was assigned to the variance (σ^2) **(C2d)**.

Appendix Equation C.3.

$$y_i = \sum_{j=1}^{m} f\left(x; T_j, M_j\right) + \epsilon, \qquad \epsilon \sim N(0, \sigma^2) \qquad (C3a)$$

$$\mu_{ij} \sim N(0, \sigma_\mu^2) \qquad (C3b)$$

$$\sigma^2 \sim \frac{\nu\lambda}{\chi_\nu^2} \qquad (C3c)$$

where y_i represents the response variable, which is the median estimate of the posterior distributions of the β_1, β_2, and β_3 parameters for location *i* **(C3a)**.

The function $f\left(x; T_j, M_j\right)$ is a conditional tree determined by T_j (how covariates x are split based on the node decision rules) and M_j (predicted values at the tree leaves). For the tree structure (T_j), the chance of a node being non-terminal decreases as the tree becomes deeper. This is controlled by the formula α(1+d)−β, and α was set to 0.95 and β set to 2, as suggested by Chipman et al., (2010).

For each covariate x, the function $f\left(x; T_j, M_j\right)$ assigns μ_{ij} from M_j. The final predicted value is the sum of all terminal node predictions (μ_{ij}) from all trees in the model **(C3b)**. For μ_{ij}, a normal distribution with mean 0 and variance σ_μ^2 was assumed, where $\sigma_\mu^2 = \frac{0.5}{k\sqrt{m}}$, shrinking μ_{ij} toward zero to limit the influence of individual trees. The error term ϵ follows a normal distribution with mean 0 and variance σ^2 following an inverse χ^2 prior distribution with default hyperparameters ν (degrees of freedom) and λ (scale) **(C3c)**.