Data pipeline to assess the spatio-temporal yield variation within a field in response to precipitation

Emmanuela van Versendaal[1], Carlos M. Hernandez[1], Peter Kyveryga[2], Trevor Hefley[3], P.V. Vara Prasad[1], and Ignacio A. Ciampitti[1].

Department of Agronomy[1], Kansas State University, 1712 Claflin Rd., Manhattan, KS, 66506, USA.

Science Agronomy[2], John Deere; 9505 Northpark Drive, Urbandale, IA, 50131, USA.

Department of Statistics[3], Kansas State University, 205 Dickens Hall, 1116 Mid-Campus Drive North, Manhattan, KS, 66506, USA.

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.
Rainfed agricultural systems are dependent on weather, particularly precipitation. Understanding the spatio-temporal variability of on-farm yield productivity is crucial for agronomic decision-making. This study develops a data pipeline to assess crop yield variability within a field in response to precipitation, utilizing historical yield monitor data farm and daily precipitation records from the CHIRPS dataset. A farm located in Illinois, containing nine years of maize and seven years of soybean yield monitor data, is used as case study. Daily precipitation data was used to calculate synthetic variables representing cumulative precipitation over different periods. Generalized Additive Models (GAMs) were employed to identify spatially varying yield responses to cumulative precipitation. This was followed by a Classification and Regression Tree (CART) analysis to categorize these responses varying response within the field. The study revealed that: (i) the optimal precipitation period for explaining yield variation differs across fields and crops; (ii) certain field areas showed more yield variation related to precipitation; and (iii) similarly high-yield variation areas can react differently to the same precipitation levels. These findings underscore the potential of integrating spatial data science with field-level data to refine management practices in precision agriculture.

Keywords.
Spatio-temporal variation, On-farm yield variation, yield response to precipitation; yield monitor data.

1. Introduction

Rainfed agriculture is the dominant farming method globally. Crop yields in rainfed systems are often 50% below their potential due to varying water availability, influenced by environmental, soil, and climatic factors. Inter-annual climate variability significantly contributes to yield variability, with 31-43% of soybean and maize yield fluctuations linked to changes in temperature and precipitation during the growing season (Lobell et al., 2009; Wani et al., 2009).

Despite extensive research on the impacts of climate variability on regional crop yields, there is a noticeable lack of detailed studies on intra-field yield responses to precipitation. This knowledge gap is significant for optimizing variable rate management practices within fields. The adoption of Geographic Information Systems (GIS) and yield monitor data has fostered interest in site-specific management, allowing for the adjustment of inputs like fertilizers and water to match the varying conditions within a field. Such practices not only cater to the unique characteristics of each field section but also enhance resource efficiency and productivity. However, more detailed research is needed to better understand how specific field conditions, such as soil moisture and topography, affect yield responses to precipitation and to tailor management practices accordingly. This is crucial for advancing precision agriculture and maximizing the potential of rainfed crops (Basso et al., 2010, 2009; Bullock et al., 2002; Fu et al., 2014; Krause et al., 2020; Moore and Tyndale-Biscoe, 1999; Mourtzinis et al., 2017; Plant, 2001; Teasdale and Cavigelli, 2017; Xu et al., 2022).

Following this rationale, this study aims to develop a data pipeline to quantify the spatio-temporal yield variation within a field in response to precipitation, using many years of yield-precipitation records.

2. Materials and Methods

2.1. Data

One farm in the United States served as case study to evaluate the data pipeline. The Farm, located in Illinois, covers 20 hectares and provided nine years of maize (Case study 1) and seven years of soybean (case study 2) yield monitor data. The yield monitor data were cleaned using an outlier detection method followed by the local Moran procedure (Córdoba, 2014). Then, the data were interpolated using inverse distance weighting interpolation (Gräler et al., 2016) to create a 6x6 m grid, enabling the overlay of yield monitor data across years.

Daily precipitation data of the farm was gathered form CHIRPS dataset (Funk et al., 2015, 2014), and used to calculate synthetic variables representing cumulative precipitation over different periods. Since planting date was not available, these periods start from the beginning of the historical planting date (April 20th for soybean) (Baum et al., 2019; Deines et al., 2023; Mourtzinis et al., 2017), and range from 10 to 200 days later, in 10-day increments.

2.2. Models

The analysis of on-farm yield variation in response to cumulative precipitation was conducted using Generalized Additive Models (GAMs) for each crop (Wood, 2017). In this framework, the first level of the model (data layer) describes the crop yield (Y_i) for the i^{th} observation. The crop yield (Y_i) was assumed to follow a normal distribution with expected value μ_i and variance σ^2:

$$Y_i \sim Normal(\mu_i, \sigma^2) \tag{1}$$

In the second level of the model, μ_i was defined as:

$$\mu_i = \beta_0 + f(Lon_i, Lat_i, CP_i), \tag{2}$$

where β_0 is the intercept; and $f(Lon_i, Lat_i, CP_i)$ is a tensor product smooth function of latitude (X_i), longitude (Y_i), and cumulative precipitation (CP_i) with basis functions thin plate spline for the spatial coordinates (Wood, 2003) and cubic regression spline for cumulative precipitation variable.

The GAMs were applied using the mgcv package in R, with model selection guided by the Akaike Information Criterion (AIC) to identify the best cumulative precipitation period for explaining yield variation. Model performance was assessed using metrics such as Relative Root Mean Square Error (RRMSE), Kling-Gupta Efficiency (KGE), and percentages of Lack of Precision (PLP) and Accuracy (PLA) from the Metrica package (Correndo et al., 2022).

Spatial variations in yield responses were derived from the GAM, determining the maximum yield at each coordinate through the first derivative of the yield with respect to cumulative precipitation. This maximum is found where the derivative equals zero, or nearest to zero when a true maximum isn't achieved. In cases of a constant yield response, the lowest cumulative precipitation extreme is used. The specific cumulative precipitation that optimizes yield was also pinpointed. Growth rates from the onset to the peak yield and decline rates post-peak were computed. If peak yield is unattainable, a zero rate is recorded, and the data includes a marker indicating that peak yield was not reached due to precipitation levels at that location.

A Classification and Regression Tree (CART) approach was utilized to differentiate responses to cumulative precipitation across the field (Breiman, 1998). This technique, categorizes observations into increasingly similar groups based on their characteristics. The cumulative precipitation that maximizes yield served as response variable, while other spatial parameters derived from the GAM analysis—such as Maximum Yield, Growth Rate, Decay Rate, and whether the maximum yield was achieved—acted as predictors. Additionally, CART configurations were adjusted to only perform splits that accounted for at least 10% of the variation in the data, ensuring significant segmentation.

3. Results

3.1. Model evaluation

Based on AIC, the optimum cumulative precipitation period for explaining yield variation changed across case studies. For Maize (case study 1), the optimum period spanned 100 days starting from May 1st. In the soybean case study (case study 2), the optimum period was 90 days beginning on April 20th. In general, the models using the optimal cumulative precipitation periods demonstrated a strong performance (KGE = 0.66-0.91, RRMSE = 11-17%).

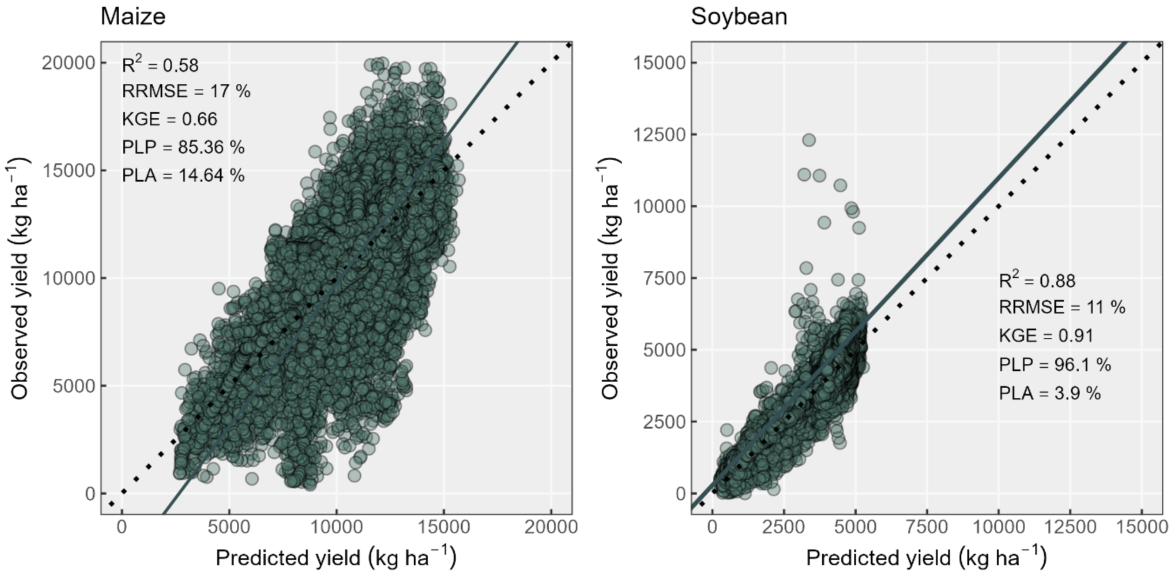

Figure 1. Model evaluation across maize (left panel) and soybean (right panel) case studies. Points represent yield in each cell within the field in a specific year.

3.2. Spatially varying response parameters

Figure 2 shows the ranges and spatial distribution of response parameters across maize and soybean case studies. In both case studies, a small fraction of the total field area did not achieve maximum yield within the specified cumulative precipitation range—specifically, 0.7% in maize (case study 1) and 0.1% in soybean (case study 2). The maize case study exhibited the highest variability in maximum achievable yield, with a coefficient of variation (CV) of 10%, compared to the soybean case study, which had a CV of 6%.

The spatial distribution of cumulative precipitation that maximizes yield does not consistently align with the observed patterns of maximum yield throughout the field. This divergence is due to the fact that identical maximum yield levels can be reached under varying precipitation conditions.

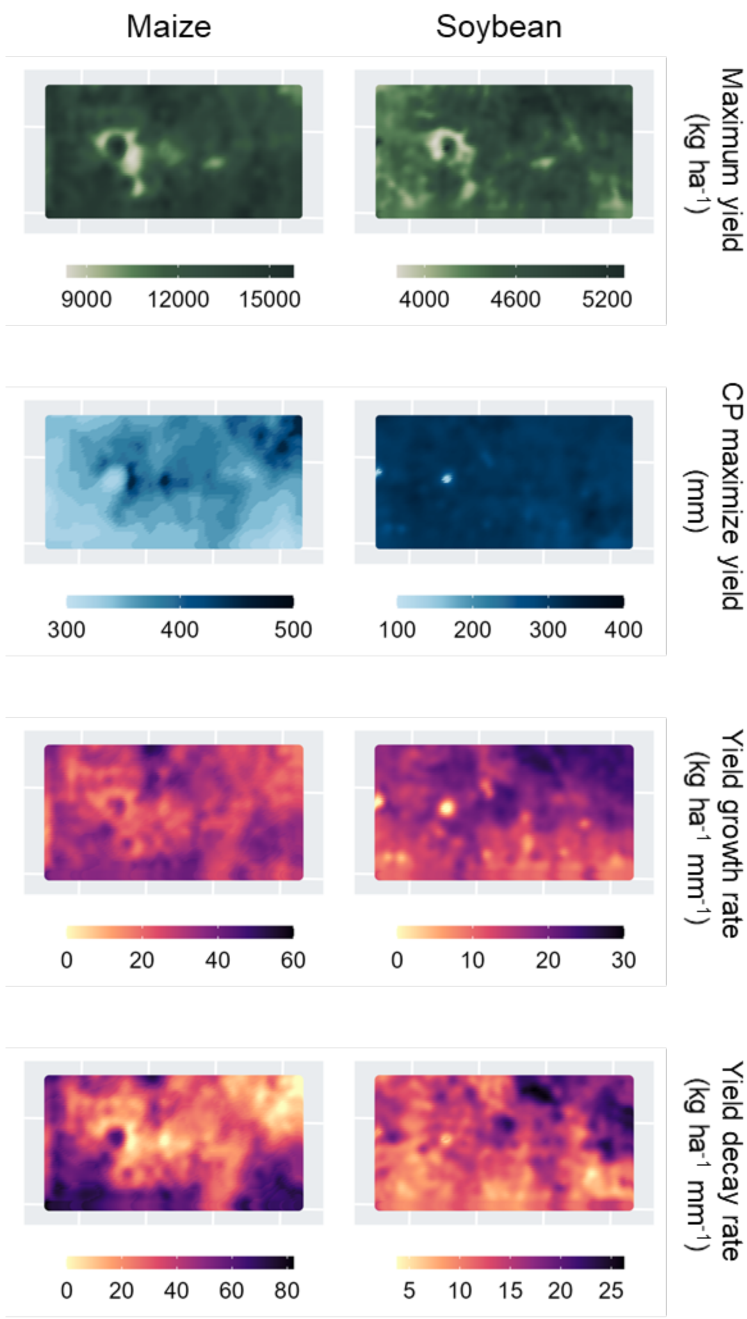

Figure 2. Maps of the spatially varying response parameters across Maize (left panels) and Soybean (right panels) case studies. Rows represent each spatially varying parameter. In all parameters, darker the color, higher the parameter's value. CP: cumulative precipitation.

3.3. Classify type of yield response to cumulative precipitation

The CART analysis revealed groups with varying type of responses to cumulative precipitation across both case studies. Maize case study obtained four groups and groups were categorized based on the decay rate, while soybean case study obtained three groups and groups were categorized based on the growth rate. In maize case study, group 4 requires the most cumulative precipitation to maximize yield, compared to groups 1, 2, and 3. Group 1 has the highest decay rate, leading to substantial yield reductions when precipitation exceeds the optimal range. The decay rates for groups 2 and 3 are 30–37 kg ha^{-1} mm^{-1} and 17–25 kg ha^{-1} mm^{-1}, respectively, while group 4 has the lowest at 4–10 kg ha^{-1} mm^{-1}. In soybean case study, Group 4 required the most cumulative precipitation to achieve maximum yield, whereas groups 1, 2, and 3 required less. Group 1, needing the least precipitation, achieved the highest yield and growth rate. In contrast, Group 4, with the highest precipitation requirement, had lower yields and growth rates (Figure 3).

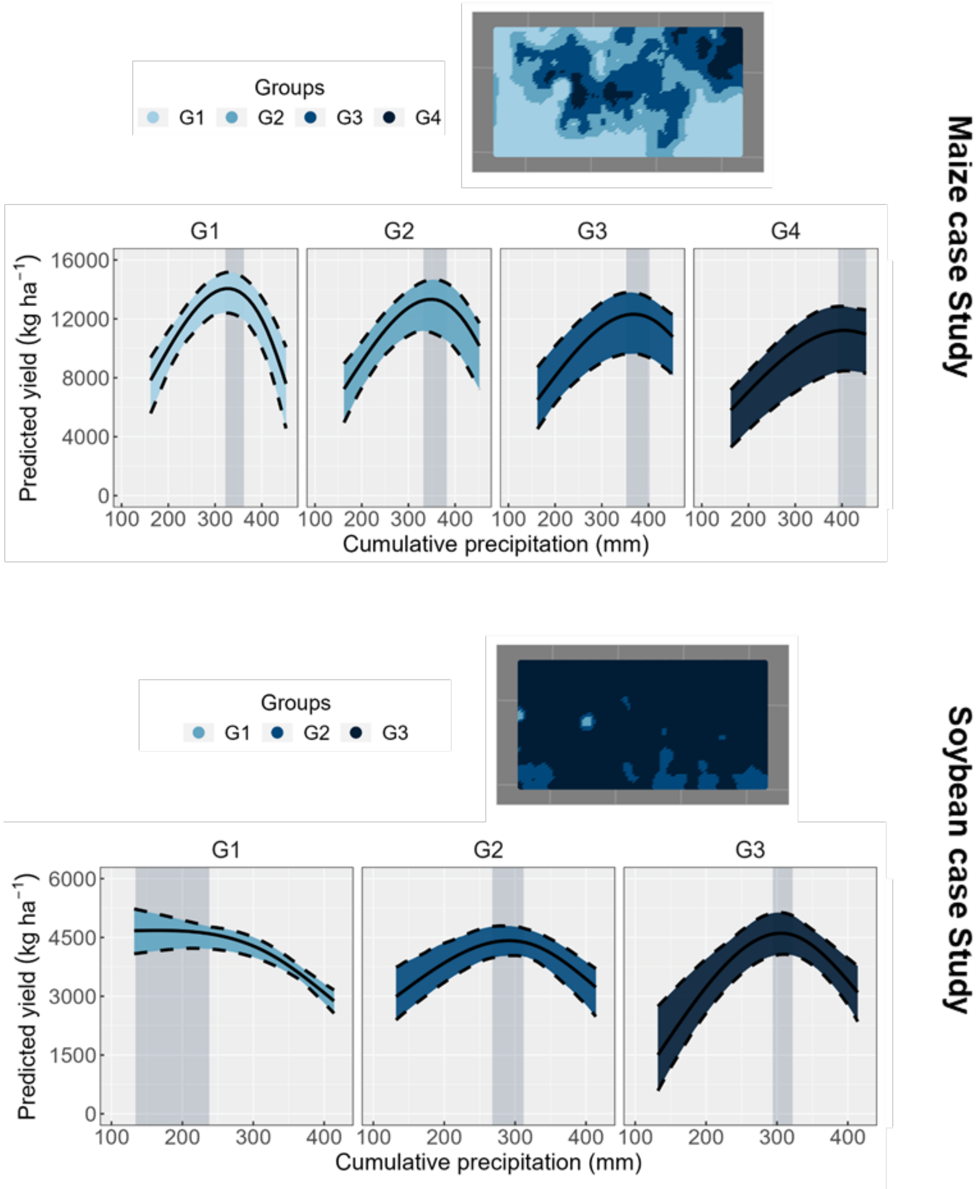

Figure 3. Maps illustrating spatial distribution of groups with varying yield response to cumulative precipitation within the field in each case study. Below each map, curves of response of yield to cumulative precipitation in each group. Top panels represent maize case study. Bottom panels represent soybean case study. Each color represents a group. Solid black line corresponds to the mean yield. Dashed black lines correspond to the 95% confidence intervals.

4. Conclusion

This research highlights the importance for farmers of recognizing spatial and temporal fluctuations in yield due to varying cumulative precipitation. Key findings from this study reveal: (i) the optimal precipitation period for explaining yield variation differs across fields and crops; (ii) certain field areas showed more yield variation related to precipitation; and (iii) similarly high-yield variation areas can react differently to the same precipitation levels. Further research should leverage high-resolution remote sensing to expand upon these findings and apply the established data framework to conduct field sampling and variable rate on-farm experiments.

References

Basso, B., Cammarano, D., Chen, D., Cafiero, G., Amato, M., Bitella, G., Rossi, R., Basso, F., 2009. Landscape Position and Precipitation Effects on Spatial Variability of Wheat Yield and Grain Protein in Southern Italy. J Agronomy Crop Science 195, 301–312. https://doi.org/10.1111/j.1439-037X.2008.00351.x

Basso, B., Cammarano, D., Troccoli, A., Chen, D., Ritchie, J.T., 2010. Long-term wheat response to nitrogen in a rainfed Mediterranean environment: Field data and simulation analysis. European Journal of Agronomy 33, 132–138. https://doi.org/10.1016/j.eja.2010.04.004

Baum, M.E., Archontoulis, S.V., Licht, M.A., 2019. Planting Date, Hybrid Maturity, and Weather Effects on Maize Yield and Crop Stage. Agronomy Journal 111, 303–313. https://doi.org/10.2134/agronj2018.04.0297

Breiman, L. (Ed.), 1998. Classification and regression trees, 1. CRC Press repr. ed. Chapman & Hall/CRC, Boca Raton, Fla.

Bullock, D.S., Lowenberg-DeBoer, J., Swinton, S.M., 2002. Adding value to spatially managed inputs by understanding site-specific yield response. Agricultural Economics 27, 233–245. https://doi.org/10.1111/j.1574-0862.2002.tb00119.x

Córdoba, M.A., 2014. Herramientas estadísticas para el monitoreo y uso de la variabilidad espacial del rendimiento y propiedades de suelo intralote (Doctoral dissertation). Universidad Nacional de Córdoba.

Correndo, A.A., Rosso, L.H.M., Hernandez, C.H., Bastos, L.M., Nieto, L., Holzworth, D., Ciampitti, I.A., 2022. metrica: an R package to evaluate predictionperformance of regression and classification point-forecastmodels. JOSS 7, 4655. https://doi.org/10.21105/joss.04655

Deines, J.M., Swatantran, A., Ye, D., Myers, B., Archontoulis, S., Lobell, D.B., 2023. Field-scale dynamics of planting dates in the US Corn Belt from 2000 to 2020. Remote Sensing of Environment 291, 113551. https://doi.org/10.1016/j.rse.2023.113551

Fu, Q., Wang, Q., Shen, X., Fan, J., 2014. Optimizing water and nitrogen inputs for winter wheat cropping system on the Loess Plateau, China. J. Arid Land 6, 230–242. https://doi.org/10.1007/s40333-013-0225-4

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci Data 2, 150066. https://doi.org/10.1038/sdata.2015.66

Funk, C.C., Peterson, P.J., Landsfeld, M.F., Pedreros, D.H., Verdin, J.P., Rowland, J.D., Romero, B.E., Husak, G.J., Michaelsen, J.C., Verdin, A.P., 2014. A quasi-global precipitation time series for drought monitoring: U.S. (Data Series No. 832), Geological Survey Data Series 832.

Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-Temporal Interpolation using gstat. The R Journal 8, 204. https://doi.org/10.32614/RJ-2016-014

Krause, M.R., Crossman, S., DuMond, T., Lott, R., Swede, J., Arliss, S., Robbins, R., Ochs, D., Gore, M.A., 2020. Random forest regression for optimizing variable planting rates for corn and soybean using topographical and soil data. Agronomy Journal 112, 5045–5066. https://doi.org/10.1002/agj2.20442

Lobell, D.B., Cassman, K.G., Field, C.B., 2009. Crop Yield Gaps: Their Importance, Magnitudes, and Causes. Annu. Rev. Environ. Resour. 34, 179–204. https://doi.org/10.1146/annurev.environ.041008.093740

Moore, G.A., Tyndale-Biscoe, J.P., 1999. Estimation of the Importance of Spatially Variable Nitrogen Application and Soil Moisture Holding Capacity to Wheat Production. Precision Agriculture 1, 27–38. https://doi.org/10.1023/A:1009973802295

Mourtzinis, S., Gaspar, A.P., Naeve, S.L., Conley, S.P., 2017. Planting Date, Maturity, and Temperature Effects on Soybean Seed Yield and Composition. Agronomy Journal 109, 2040–2049. https://doi.org/10.2134/agronj2017.05.0247

Plant, R.E., 2001. Site-specific management: the application of information technology to crop production. Computers and Electronics in Agriculture 30, 9–29. https://doi.org/10.1016/S0168-1699(00)00152-6

Teasdale, J.R., Cavigelli, M.A., 2017. Meteorological fluctuations define long-term crop yield patterns in conventional and organic production systems. Sci Rep 7, 688. https://doi.org/10.1038/s41598-017-00775-8

Wani, S.P., Rockström, J., Oweis, T.Y., 2009. Rainfed agriculture: unlocking the potential, Comprehensive assessment of water management in agriculture series. CABI, Wallingford.

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

Wood, S.N., 2003. Thin Plate Regression Splines. Journal of the Royal Statistical Society Series B: Statistical Methodology 65, 95–114. https://doi.org/10.1111/1467-9868.00374

Xu, R., Li, Y., Guan, K., Zhao, L., Peng, B., Miao, C., Fu, B., 2022. Divergent responses of maize yield to precipitation in the United States. Environ. Res. Lett. 17, 014016. https://doi.org/10.1088/1748-9326/ac3cee