

Identification of yield-limiting factors on maize production from observational data

Identificación de los factores que limitan el rendimiento de la producción de maíz a partir de datos observacionales

Ninibeth Gibelli Sarmiento Herrera ^{1,4}, Andrés Aguilar-Ariza ^{2,5}, Jesús Hernán Camacho-Tamayo ^{3,6}

¹Cenicafé. ²Alianza Bioversity-CIAT. ³Universidad Nacional de Colombia, Facultad de Ingeniería, Departamento de Ingeniería Civil y Agrícola.

⁴✉ ninibeth.sarmiento@cafedecolombia.com; ⁵✉ a.aguilar@cgiar.org; ⁶✉ jhcamachot@unal.edu.co



<https://doi.org/10.15446/acag.v72n3.106012>

2023 | 72-3 p 241-251 | ISSN 0120-2812 | e-ISSN 2323-0118 | Rec.: 2022-11-25 Acep.: 2024-05-14

Abstract

Following the approach of Site-Specific Agriculture, this study identified the yield-limiting factors of climate, soil, and management on maize production. The information was obtained from farmers' observations on cropping events, between 2013 and 2016 in Tolima, one of the regions with the highest maize production in Colombia. Using Random Forest, factorial analysis and cluster techniques, the climate, soil and management factors related to crop yield variation were determined. Based on Random Forest regression, climate and soil factors explained 23% and 32% of yield variation, respectively. Relative humidity, average temperature, and precipitation were the most important climate factors associated with crop yield variation, while the slope and mottling were the most important soil factors. The factorial analysis in combination with cluster techniques allowed to establish groups with similar climate and soil conditions. Among those groups, the agricultural practices that favour yields, such as mechanization, fertilization, and management of grain moisture, were differentiated. The results showed an approach to characterizing productive systems by leveraging observational data.

Keywords: agricultural practices; cereals; climate; data mining; factorial analysis; soil factors.

Resumen

Siguiendo el enfoque de la Agricultura Específica de Sitio, este estudio identificó los factores limitantes de rendimiento del clima, el suelo y el manejo en la producción de maíz. La información se obtuvo de las observaciones de los agricultores sobre los eventos de cultivo, entre 2013 y 2016 en Tolima, una de las regiones con mayor producción de maíz en Colombia. Utilizando técnicas de Random Forest, análisis factorial y clúster, se determinaron los factores de clima, suelo y manejo, relacionados con la variación del rendimiento del cultivo. Mediante el análisis de regresión Random Forest, los factores de clima y suelo explicaron el 23% y el 32% de la variación del rendimiento, respectivamente. La humedad relativa, la temperatura media y la precipitación fueron los factores climáticos más importantes asociados a la variación del rendimiento, mientras que la pendiente y el moteado fueron los factores edáficos más importantes. El análisis factorial en combinación con técnicas de clúster, permitió establecer grupos con condiciones climáticas y edáficas similares. Entre esos grupos se diferenciaron las prácticas agrícolas que favorecen los rendimientos, como la mecanización, la fertilización y el manejo de la humedad del grano. Los resultados muestran un enfoque para caracterizar los sistemas productivos a partir de datos observacionales.

Palabras claves: análisis factorial; cereales; clima; factores del suelo; minería de datos; prácticas agrícolas.

Introduction

In Colombia, maize (*Zea mays* L.) production system is widespread in all thermal floors due to its adaptation to diverse agroclimatic conditions (FENALCE, 2010b). Of the total area sown, 70% corresponds to traditional maize and 30% to technified maize. The average yield between 2013 and 2017 for traditional system was between 1.4 and 1.5 t ha⁻¹ and for technified system between 4.5 and 5.4 t ha⁻¹ (MADR, 2017). The climatic, topographic and soil conditions variability have had repercussions on crop yield (Cortés *et al.*, 2013). According to the Federación Nacional de Cultivadores de Leguminosas y Cereales – FENALCE, between the years 2015 and 2016 production decreased by 3.24%, mainly due to the occurrence of El Niño phenomenon (FENALCE, 2017).

Farmers capacities must be strengthened to face the effects of climate variability on maize crop. To support the decision-making at a farm level, it is necessary to have information that allows knowing the relationship between environmental conditions, edaphic site, and crop yield. In response to the lack of historical and georeferenced data on corn production, FENALCE in collaboration with the International Centre for Tropical Agriculture - CIAT in 2013, initiated the compilation of observational data on productive events across various regions of the country. The aim was to understand the environmental and agronomic factors influencing corn cultivation. To achieve this, they conducted surveys among farmers to gather insights into their experiences and the diverse management practices employed throughout different maize production cycles. Following the methodology of Site-Specific Agriculture (AES), the productive systems were characterized in terms of management practices and environmental conditions, integrating information collected by small-scale farmers and climate information (Cock and Luna, 1996; Isaacs *et al.*, 2004; 2007). Under this methodology, information is collected from many productive events under diverse conditions, to develop data-based models that allows to generate site specific management recommendations (Jiménez *et al.*, 2009).

For the observational information analysis (Sagarin and Pauchard, 2010) the methodologies that allows to know the association between factors and system response are used (Jiménez *et al.*, 2008), from categorical and continuous variables. Multiple regression models, mixed effects models (Long *et al.*, 2017) and data mining methodologies have been used, as models based on decision trees (Delerce *et al.*, 2016; Jiménez *et al.*, 2016; Kihara *et al.*, 2015) and neural networks (Jiménez *et al.*, 2011; Miao and Niu, 2016).

The AES methodology has been applied in different crops, such as sugarcane (Isaacs *et al.*, 2007), coffee (Cock *et al.*, 2011), blackberry (Jiménez *et al.*, 2009),

lulo, (Jiménez *et al.*, 2011) and plantain (Jiménez *et al.*, 2016), to evaluate the relationships between environmental and management factors with crop response. These studies suggest that one of the most effective forms of analysis, is to establish groups of events with similar environmental conditions, prior to determining the effects of management practices on crop yield variation (Cock *et al.*, 2011; Jiménez *et al.*, 2016). Climate and soil are classified as uncontrollable factors, while management practices are factors controlled by the farmer (Jiménez *et al.*, 2016).

The aim of this study was to pinpoint the climate, soil, and management variables influencing maize yield fluctuations, based on information obtained from farmers productive experiences in Tolima, a region renowned for its significant maize production in Colombia.

Materials and methods

Research areas. This study focuses on the maize producing region of the Tolima department (Figure 1), which concentrates 15% of the total maize production in Colombia (MADR, 2017). Taking as reference the meteorological station Nataima located in the municipality of Espinal-Tolima (IDEAM), the maximum and minimum daily temperatures have average values of 33°C and 22°C, respectively. There are two rainy seasons between March-May and September-November, which determine the sowing seasons. Solar radiation has values between 16 MJ m⁻²day⁻¹ and 21 MJ m⁻²day⁻¹.

In most of maize production area, soils are of flat topography with slopes of less than 12%, moderately deep, well drained, with moderate fertility. The soils of the hillside areas where corn is grown have slopes between 12 and 50%, with a high presence of organic material. In certain areas, the effective depth is limited due to the presence of sandy layers or surface salts (IGAC, 2004).

Sources of information and data preparation. The information was obtained from 417 productive maize events, collected by FENALCE between 2013 and 2016 in 10 municipalities, through surveys applied to farmers (www.siria.fenalce.org). Yield information, sowing and harvest dates, management practices, monitoring and control of pests and diseases, and field geographical location were recorded.

Soil and land were characterized using the RASTA (Rapid Soil and Terrain Assessment) methodology (Álvarez *et al.*, 2004). We evaluated 15 variables such as slope, color, texture, structure, pH, stoniness, effective depth, among others. (Table 1). Cleaning processes and quality control of data were applied and events with incomplete information were eliminated.

Table 1. Variable description.

Name	Description	Unit
CLIMATE (Source: IDEAM)		
TX_Avg_Et1/TX_Avg_Et2	Average Maximum Temperature in vegetative/reproductive stage*	°C
TM_Avg_Et1/TM_Avg_Et2	Average Minimum temperature in vegetative/reproductive stage*	
TA_Avg_Et1/TA_Avg_Et2	Average temperature in vegetative/reproductive stage*	
Diurnal_Range_Et1/Diurnal_Range_Et2	Average diurnal range in vegetative/reproductive stage*	
RH_Avg_Et1/RH_Avg_Et2	Average relative humidity in vegetative/reproductive stage*	
SR_Accum_Et1/SR_Accum_Et2	Accumulated solar radiation in vegetative/reproductive stage*	
P_Accum_Et1/P_Accum_Et2	Accumulated rainfall in vegetative/reproductive stage*	mm
CDW_P_Freq_Et1/CDW_P_Freq_Et2	Frequency of consecutive days without rainfall in vegetative/reproductive stage*	number
TEMP_MAX_FREQ_Et1/TEMP_MAX_FREQ_Et2	Frequency of days with maximum temperature above 35 °C in vegetative/reproductive stage*	
AGRICULTURAL PRACTICES (Source: SIRIA)		
ROW_WIDTH	Spacing between plant rows*	m
PLANT_SPACING	Spacing between plant within a row*	
GROWING_PERIOD	Crop growing period from emergence to harvest*	days
TOTAL_N	Total amount of Nitrogen used*	kg ha ⁻¹
TOTAL_K	Total amount of Calcium used*	
TOTAL_P	Total amount of Phosphorus used*	
GRAIN_MOISTURE	Grain moisture at harvest	%
PLANTING_METHOD	Planting method: Manual or Mechanical**	
ENDOSPERM_COLOR	Endosperm color: white or yellow**	
CROP_NAME	Crop Common Name**	
PREVIOUS_CROP	Plant harvested the Preceding Semester**	
PLANT_POLULATION_20	Plant density at 20 days after sowing*	
HARVESTING_METHOD	Harvest method: Manual or Mechanical**	
CULTIVAR	Crop cultivar name**	
N_FERTILIZATION	Number of fertilizers	
PEST_CONTROL	Number of pest controls	
DISEASES_CONTROL	Number of disease controls	
WEED_CONTROL	Number of weed controls	
SOIL (Source: RASTA)		
SLOPE	Slope*	%
EFFECTIVE_SOIL_DEPTH	Maximum rooting depth*	cm
THICKNESS	Thickness of the first soil horizon*	
PEAT_SOIL	Is the soil black and soft? **	
DRY_REGION	Is the soil in an arid region? **	
LAYERS	Number of soil horizons in the profile	
pH	Soil pH	
CARBONATES	Has the soil carbonates? **	
HARDENED_LAYER	Has the soil profile impermeable layers? **	
MOTTLING	Has the soil yellow, red, green, or gray spots? **	
STRUCTURE	Soil structure**	
MULCH	Is there mulch on the soil? **	
VEGETAL_COVER	Is there vegetation cover on the soil? **	
TEXTURE	Soil texture**	
SOIL_RESISTANCE	Soil penetration resistance in the first soil horizon**	
DEPENDENT VARIABLE (Source: SIRIA)		
YIELD	Maize yield	ton ha ⁻¹

**Categorical variables

* Continuous variables

The information of five weather stations of the Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) was used, with records of precipitation, maximum and minimum temperature, solar radiation, and relative humidity (Figure 1). Following the methodology described by Delerce *et al.* (2016), quality control and filling of missing data was performed, using the R RMAWGEN packages (Cordano and Eccel, 2012) and SIRAD (Bojanowsky, 2015).

The crop cycle was divided into two stages from the dates of emergence, flowering, and harvest (Table 2). The first stage (Et1) corresponds to the vegetative phase and second stage (Et2) to the reproductive phase until harvest. The climate variable characteristics of each stage were estimated (Table 1).

Method for variable selection. Variable selection was conducted using the random tree-based method Random Forest (Breiman, 2001), that allows analyzing non-linear relationships between independent variables (categorical and continuous) and the response variable. We used the random Forest module from the caret package of the R software (Kuhn *et al.* 2014).

It was trained one hundred models to mitigate the method instability. For each model, 500 trees were generated (ntree=500) with a random sample of three quarters of the number of independent variables (mtry=3p/4) on each node. Training and validation data sets were randomly selected for cross validation, using 70% and 30% of the data.

The score of the independent variables was expressed in terms of «relative importance» defined as a measure of each regressor's contribution to model fit (Grömping, 2015). To avoid erroneous assignments of the variable importance, highly correlated variables were eliminated, with absolute values greater than 0.8 of the Pearson correlation for continuous variables and Cramer's V for categorical variables.

The model performance was expressed as coefficient of determination R^2 average of the cross validation in the one hundred models (Delerce *et al.* 2016). The importance of the variables was normalized and scaled by the average R^2 of the model.

Following the methods described by Jimenez *et al.* (2016), the climate and soil conditions were considered as non-controllable factors. Based on

the selected variables groups of homologous events (with climate and soil similar conditions) were established by factorial analysis and the combination of a hierarchical cluster into main components using the Ward's criterion and the k-means cluster, following the methodology described by Jiménez *et al.* (2016). The functions of the R FactoMineR package were used (Sebastien *et al.*, 2008) to apply the dimensionality reduction method (FMAD), and Hierarchical Clustering on Principal Components (HCPC) (Husson *et al.*, 2010).

Using the Kruskal-Wallis multiple comparison test at 5% significance, using the agricolae package of R (Mendiburu, 2017) the significant differences between groups was evaluated, in terms of crop yield. The variables with greater relative importance were determined by applying Random Forest, for each group of homologous events.

Variable selection was based on the relative importance of variables and their relationship with the crop yield variation, supported by the expert opinions of agronomists with knowledge of crop development in the study area.

Results and discussion

Variable reduction. We identified the variables with Pearson correlation coefficient or association greater than 0.8. The average temperature (TA_Avg) in both stages was correlated with the maximum and minimum temperature (TX_Avg y TM_Avg), the thermal range (Diurnal_Range) and frequency of days with temperatures above 35°C (TEMP_MAX_FREQ). The variables TA_Avg_Et1 and TA_Avg_Et2, were preserved since the average temperature is related to the accumulation of thermal units for the growth of the crop and presented a higher correlation with the yield.

Soil variables did not show high correlations. Management variables CROP_NAME and ENDOSPERM_COLOR showed a value of Cramer's V of 0.81. Considering the number of categories of the variable CROP_NAME (25) and that some of these corresponded to a single productive event, it was defined to use the ENDOSPERM_COLOR variable.

After the elimination process, 10 of 18 climate variables were used for model application, 19 of the 20 management variables and 15 soil variables.

Table 2. Maize Phenological stages (Ritchie and Hanway 1982).

Vegetative Growth Stages				Reproductive Development Stages			
V0 - V2	V5	V8 - V10	V12 - V14	VT - R1	R2 - R3	R3 - R4	R5 - R6
Germination and emergence	Plant population established. Growth point 20-25 mm below surface.	Cob development	Active growth-leaves and cob development. Brace root development.	Pollination. 5-10 Days	Kernel development	Grain filling. Nutrients are transported to cob.	Physiological maturity. End of mass gain. Ready for harvest 14% Moisture

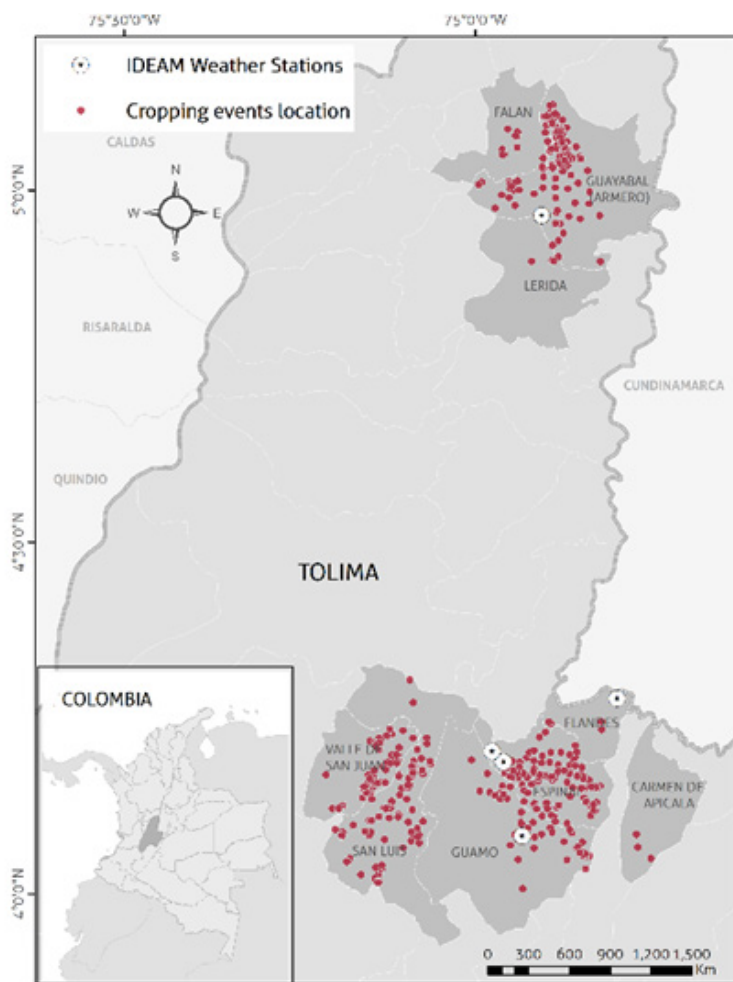


Figure 1. Location map of the study area and weather stations.

Analysis for climate and soil variables. The model obtained for the climate variables by the Random Forest method, explained 23% of the maize yield variation. The accumulated precipitation and the average temperature in the vegetative stage, the frequency of days without rain in the reproductive stage, accumulated solar radiation and relative humidity in the vegetative stage, were the most important variables (Figure 2a). For the soil variables, the model explained 32% of the yield variation, being the presence of mottling, the thickness, the presence of litter, the number of layers or horizons and the slope, the variables with greater importance (Figure 2b). The R-squared obtained by each model is an indication of the fraction of the dependent variable's (yield) variance the model could explain with the climate or soil factors by separate (Shmueli, 2010).

Several authors report better performance of Random Forest in relation to multiple regression models (Jeong *et al.* 2016; Sandri and Zuccolotto 2006), due to its high performance against non-linear interactions, as is the case of productive

systems, in which there are multiple interactions between biological, physical, physiological and crop management factors. Unlike methods based on linear regression, Random Forest does not require categorical variables to be transformed to fit a distribution.

Relationships between crop yield and climate and soil factors. To delve into the intricate connections among key variables and crop yield, were utilized partial dependence graphs derived from obtained models for continuous variables (Figures 3 and 6). These graphs illustrate the fluctuation of crop yield as depicted by gray lines, with the black line denoting the average variation. Additionally, for categorical variables, the boxplots showcase the distribution of crop yield and its correlation with each variable.

Partial dependence graphs of precipitation (Figure 3a), and solar radiation (Figure 3d) accumulated in the vegetative stage, did not show a clear response on crop yield variations. Other studies reported that maize crop requires between 550 to 650 mm, well

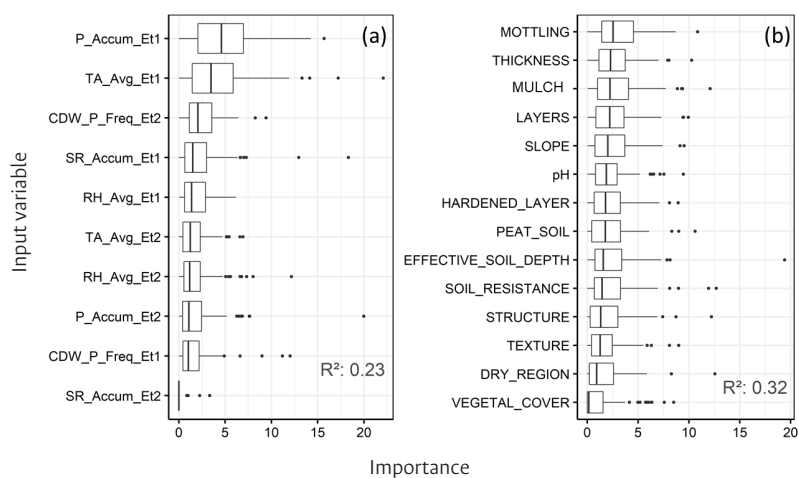


Figure 2. Boxplot of variables importance of Random Forest models (a) Climate; (b) Soil variables.

distributed, during its vegetative cycle (FENALCE 2010a) and low rainfall in this stage limit the growth and crop yield (Westcott *et al.* 2005), additionally, solar radiation in interaction with the absorption of nutrients and temperature, influences the growth of plants (Morales-Ruiz *et al.* 2016). The results obtained for these two variables can be related to the irrigation application that favors the crop development, however, there is no information associated with the productive events.

The impact of temperature during the vegetative stage on yield becomes notably adverse when temperatures surpass 28.5°C, as depicted in Figure 3b. Maize thrives within a temperature spectrum of 10 to 29°C, with the optimal average range falling between 24 and 26°C. However, temperatures exceeding 28°C hinder the crop's water absorption capacity (Cortés *et al.*, 2013).

Frequencies number of continuous days without rainfall higher than 0.15, which corresponds to 15 ± 2 days, decrease the yield (Figure 3c). Most of the crop water consumption occurs during flowering where a deficit of two days can reduce yields by 22%, and 50% when it is from 6 to 8 days. Runge (1968) reported maize yield responses to interactions between maximum temperature and precipitation between 25 and 15 days after anthesis, finding that when precipitation is low (equal to or less than 44 mm for 8 days), the yield can be reduced by 1.2 to 3.2%.

In terms of relative humidity, values higher than 76% generated lower yields (Figure 3e). According to Oke (2016) elevated average relative humidity levels correlate with a heightened incidence of diseases such as Tar Spot Complex (TSC), leading to a subsequent decrease in maize yield. In Colombia, FENALCE (2010a) reported that the occurrence of temperatures between 17 and 22°C, and relative humidity greater than 75% encourage the development of TSC and gray spot complex.

The presence of mottling in soil was the soil variable with higher relative importance in the crop yield (Figure 3f). In soils with mottling presence, lower yields were found. Mottling is yellow, red, blue, green, or gray spots mixed on the horizon in little or large amounts and indicate poor drainage and lack of oxygen for the roots (Álvarez *et al.*, 2004). According to the charts of partial dependence, there is not a higher variation on yield in relation to the first horizon thickness greater (Figures 3g and 3i). For the events evaluated thickness varies between 10 to 60 cm. FENALCE (2010a) states that deep soils, with high content of organic matter and good moisture retention capacity are ideal for maize crop.

Fields with presence of mulch in the soil, presented lower yields (Figure 3h) with significant differences. This result is not consistent with FENALCE (2010a) recommendations for soil conservation and other studies. Kamar *et al.* (2018) found higher yields in soils covered by mulch or litter, since it allows to conserve soil moisture. Soils with slopes greater than 20% have lower yields (Fig. 3j). High slopes favor erosion and affect water infiltration, which can reduce yield (Marques da Silva and Silva 2008).

Supported by the expertise of FENALCE agronomists, those variables for which clear and representative relationships of productive conditions were not found were eliminated. The average temperature and relative humidity in Et1 and the frequency of days without rain in Et2 were selected as climatic limiting factors. The presence of mottling and the slope were selected as soil factors.

Climate and soil homologous events. Under similar climate and soil conditions, differences in crop yield can be explained from the variation in management practices (Cock *et al.* 2011). According to Jiménez *et al.* (2016), if the variance in climate and soil factors (not controllable) through groupings is reduced, the variations in yield will be mostly related to the management practices conducted in the crop.

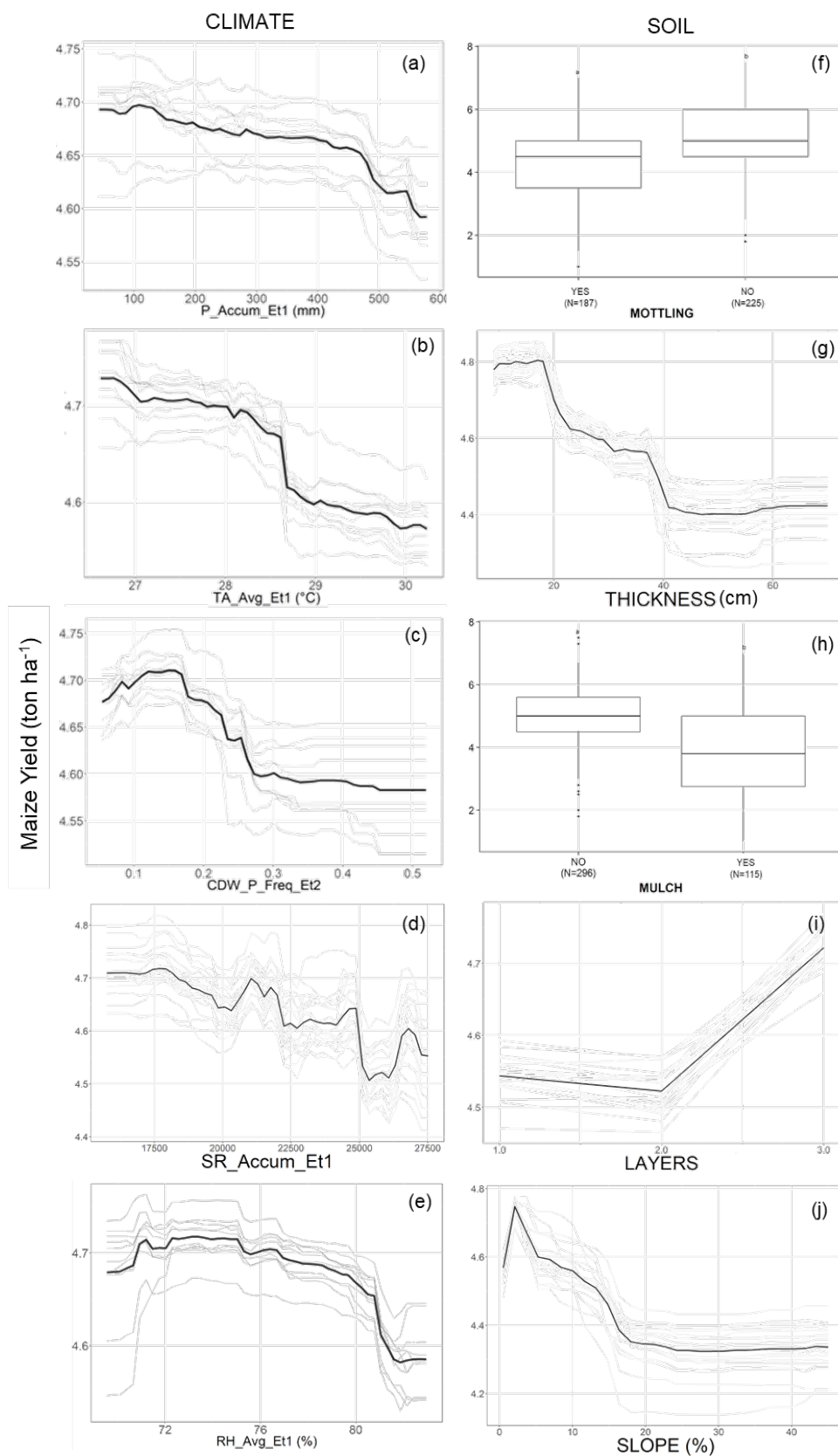


Figure 3. Partial dependence plots and Boxplot of the most relevant climate and soil variables.

Applying factorial analysis of selected climate and soil variables, the first five components explained 60.74% of the variance. Under Ward's criteria, the method suggested four groups of homologous events. According to the Kruskal Wallis test with a significance of 5%, significant differences were found in the mean values of crop yield between the groups found.

Groups 1 and 2 present higher crop yields, with statistically equal average values of 5.09 and 4.93 t ha⁻¹, followed by group 3 with a value of 4.5 t ha⁻¹ and group 4 with 3.7 t ha⁻¹ (Figure 4).

Relationships between management factors and crop yield. To understand how management practices by farmers can significantly affect crop yield, Random Forest was executed using the management information associated with the productive events.

For groups 1 and 2 that have higher yields, based on R², management variables just explained, 8% and 36% of yield variation, respectively. For groups 3 and 4, management variables explained between 67% and 19% respectively (Figure 5). Due to the model low performance, no variables were selected for group 1.

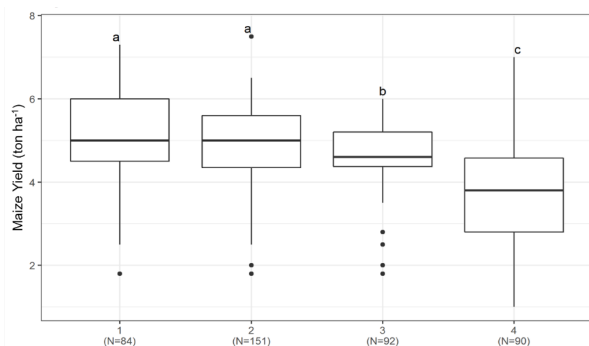


Figure 4. Yield distribution across the homologous events of climate and soil. Lowercase letters show the results of the Kruskal-Wallis test, with statistically similar clusters grouped by the same letter.

The harvest method and the distance between plants were the most important variables for group 2, the total nitrogen and the planting method for group 3, for group 4 and for the previous crop and the number of fertilizations.

Group 2 events, in which mechanical harvesting took place, present higher yields than manual harvest (Figure 6a). Mechanical harvesting has greater efficiency in the work execution, decreases the risks and allows obtaining a uniform and clean product. Additionally, mechanical harvesting can be performed with relatively high humidity (20 -25%) which reduces the number of fallen plants (FENALCE, 2010a). The distance between plants higher than 20 cm represent lower yields (Figure 6b), given the direct relationship between planting density and yield.

For group 3, total doses of nitrogen applied during the cycle, equal to 100 kg ha⁻¹, and are suitable for crop (Figure 6c), with smaller quantities, lower yields are obtained. Based on events data, was not found differences of yield between 100 kg ha⁻¹ and doses higher than 300 kg ha⁻¹. Events in which mechanical harvesting was conducted (Fig. 6d), have higher yields than manual harvesting. Under mechanical harvest, the efficiency is greater, and the uniformity in the distance and depth of sowing is guaranteed if the correct equipment calibration is conducted. (FENALCE, 2010a).

For group 4, events in which crops were rotated between rice and maize, have higher yields than those in which the previous crop was maize (Figure 6e). Crop rotation has several advantages, among which are the increase in the content of organic matter, the improvement of soil structure and microbial activity and the interruption of pests and diseases cycles (Benitez et al., 2017).

The number of fertilizer applications indicates better yields under the application of fractional fertilization in 3 applications (Figure 6f). Split

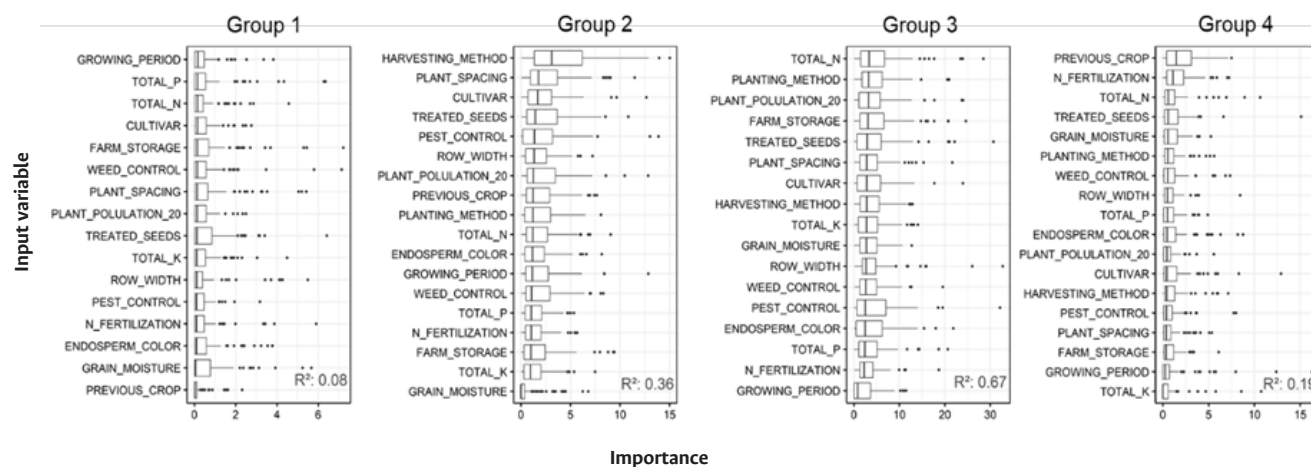


Figure 5. Boxplots of agricultural practices importance of Random Forest models by group.

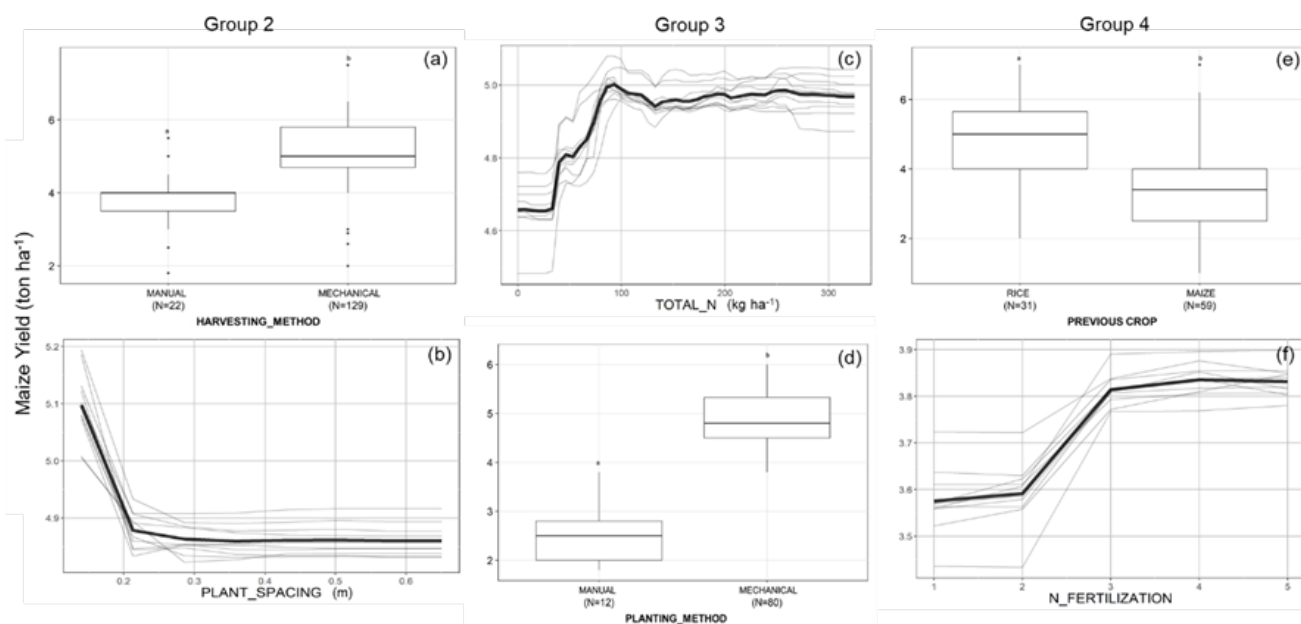


Figure 6. Partial dependence plots and Boxplot of the most relevant agricultural practices by group. Lowercase letters show the results of the Kruskal-Wallis test, with statistically similar clusters grouped by the same letter.

fertilizer application in the stages in which the crop requires more nutrients, ensures a good foliar expansion, use of sunlight and, therefore, higher yields (FENALCE, 2010a).

Conclusions

The Random Forest, factorial analysis and cluster techniques allowed to identify the climate, soil, and management variable of major importance in explaining the variation of maize yield in Tolima, from the observational information associated with productive events. The climate and soil factors by separate explained 23% and 32% of the yield variability, respectively.

The analysis by groups of homologous events, with similar conditions of climate and soil, allowed to differentiate management practices such as fertilization, mechanization and crops rotation and their relation to yield.

The results obtained provide an approach to characterizing productive systems by leveraging observational data and employing data mining techniques, all under the guidance of experts. This approach aids in identifying the most recommended management practices tailored to specific sites.

The lack of detailed information on some management practices and soil data limited the analysis of the relationships between variables and yield variation. Therefore, it is especially important to include other crop aspects, continue with collecting information and define, based on expert knowledge, the variables that characterize the productive systems.

One of the possible applications of this methodology is the identification of farmers who better manage crops and who can share their experiences with other farmers to improve their productivity.

Acknowledgements

To the Federacion Nacional de Cultivadores de Cereales y Leguminosas, FENALCE, for providing the information for the development of this research and for participating in the results evaluation. To Daniel Jiménez PhD., senior scientist at CGIAR, for his advice in the development of this research.

References

- Álvarez, Diana Milena, Marcela Estrada, and James H. Cock. (2004). "RASTA Rapid Soil and Terrain Assessment: Guía Práctica Para La Caracterización Del Suelo y Del Terreno." Cali, Colombia.
- Benitez, M.-S., Osborne, S. L., & Lehman, R. M. (2017). Previous crop and rotation history effects on maize seedling health and associated rhizosphere microbiome. *Scientific Reports*, 7(1), 15709. <https://doi.org/10.1038/s41598-017-15955-9>
- Bojanowsky, J. S. (2015). Sirad: Functions for calculating daily solar radiation and evapotranspiration. <https://cran.r-project.org/web/packages/sirad/sirad.pdf>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 1–33. <https://doi.org/10.1017/CBO9781107415324.004>
- Cock, J., & Luna, C. A. (1996). Analysis of large commercial databases for decision making. In *Sugar 2000 Symposium* (pp. 24–25). CSIRO. https://www.researchgate.net/publication/312586486_Analysis_of_large_commercial_databases_for_decision_making

- Cock, J. H., Álvarez, D. M., Estrada, M. (2004). RASTA rapid soil and terrain assessment: Guía práctica para la caracterización del suelo y del terreno. Cali, Colombia.
- Cock, J., Oberthür, T., Isaacs, C., Läderach, P. R., Palma, A., Carbonell, J., Victoria, J., et al. (2011). Crop management based on field observations: Case studies in sugarcane and coffee. *Agricultural Systems*, 104(9), 755–769. <https://doi.org/10.1016/j.agsy.2011.07.001>
- Cordano, E., & Eccel, E. (2012). RMAWGEN: Multi-site autoregressive weather generator (R Package). <https://cran.r-project.org/web/packages/RMAWGEN/RMAWGEN.pdf>
- Cortés B., C. A., Bernal, J., Díaz A., E., & Méndez, J. (2013). Uso del modelo Aquacrop para estimar rendimientos para el cultivo de maíz en los departamentos de Córdoba, Meta, Tolima y Valle de Cauca. FAO.
- Delerce, S., Dorado, H., Grillon, A., Rebolledo, M. C., Prager, S. D., Patiño, V. H., Garcés Varón, G., & Jiménez, D. (2016). Assessing weather-yield relationships in rice at local scale using data mining approaches. *PloS ONE*, 11(8), e0161620. <https://doi.org/10.1371/journal.pone.0161620>
- FENALCE. (2010a). Aspectos técnicos de la producción de maíz en Colombia. Federación Nacional de Cerealistas.
- FENALCE. (2010b). El cultivo de maíz, historia e importancia. El Cerealista.
- FENALCE. (2017). Informe de gestión 2017-A. http://fenalce.org/siembras/archivos_lt/lt_246IG-FNC-2017-A.pdf
- Grömping, U. (2015). Variable importance in regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(2), 137–152. <https://doi.org/10.1002/wics.1346>
- Husson, F., Josse, J., & Pagès, J. (2010). Principal component methods - hierarchical clustering - partitional clustering: Why would we need to choose for visualizing data? Technical Report of the Applied Mathematics Department (Agrocampus). http://www.sthda.com/english/upload/hpc_husson_josse.pdf
- Instituto Geográfico Agustín Codazzi (IGAC). (2004). Estudio general de suelos y zonificación de tierras. Departamento de Tolima.
- Isaacs, C., Carrillo, C., Carbonell, J. A., Anderson, A., & Ortiz, U. (2004). Desarrollo de un sistema interactivo de información en web con el enfoque de agricultura específica por sitio (Serie Técnica 34). Cenicaña. https://www.cenicana.org/pdf_privado/serie_tecnica/st_34/st_34.pdf
- Isaacs, C. H., Carbonell, J. A., Amaya, A., Torres, J. S., Victoria, J. I., Quintero, R., Palma, A. E., & Cock, J. H. (2007). Site specific agriculture and productivity in the Colombian sugar industry. In *Proceedings of the 26th Congress International Society of Sugar Cane Technologists (ISSCT)* (pp. 339–350). Proc. Int. Soc. Sugar Cane Technol., 26.
- Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K. M., Gerber, J. S., Reddy, V. R., & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *PloS ONE*, 11(6), e0156571. <https://doi.org/10.1371/journal.pone.0156571>
- Jiménez, D., Cock, J., Jarvis, A., García, J., Satizábal, H. F., Van Damme, P., Pérez-Urbe, A., & Barreto-Sanz, M. A. (2011). Interpretation of commercial production information: A case study of lulo (*Solanum quitoense*), an under-researched Andean fruit. *Agricultural Systems*, 104(3), 258–270. <https://doi.org/10.1016/j.agsy.2010.10.004>
- Jiménez, D., Cock, J., Satizábal, H. F., Barreto S, M. A., Pérez-Urbe, A., Jarvis, A., & Van Damme, P. (2009). Analysis of Andean blackberry (*Rubus glaucus*) production models obtained by means of artificial neural networks exploiting information collected by small-scale growers in Colombia and publicly available meteorological data. *Computers and Electronics in Agriculture*, 69(2), 198–208. <https://doi.org/10.1016/j.compag.2009.08.008>
- Jiménez, D., Dorado, H., Cock, J., Prager, S. D., Delerce, S., Grillon, A., Andrade Bejarano, M., Benavides, H., & Jarvis, A. (2016). From observation to information: Data-driven understanding of on farm yield variation. *PloS ONE*, 11(3), e0150015. <https://doi.org/10.1371/journal.pone.0150015>
- Jiménez, D., Pérez-Urbe, A., Satizábal, H. F., Barreto, M., Van Damme, P., & Tomassini, M. (2008). A survey of artificial neural network-based modeling in agroecology. In B. Prasad (Ed.), *Soft computing applications in industry* (Vol. 226, pp. 1–17). Springer. https://doi.org/10.1007/978-3-540-77465-5_13
- Kamar, S. S. A., Khan, M. H., & Uddin, M. S. (2018). Effect of irrigation and mulch on maize yield (*Zea mays*) in southern areas of Bangladesh. *Journal of Agricultural Crop Research*, 6(June), 28–37. <http://www.sciencewebpublishing.net/jacr/archive/2018/June/pdf/Kamar%20et%20al.pdf>
- Kihara, J., Tamene, L. D., Massawe, P., & Bekunda, M. (2015). Agronomic survey to assess crop yield, controlling factors and management implications: A case-study of Babati in northern Tanzania. *Nutrient Cycling in Agroecosystems*, 102(1), 5–16. <https://doi.org/10.1007/s10705-014-9648-3>
- Kosaki, T., Wasano, K., & Juo, A. S. R. (1989). Multivariate statistical analysis of yield-determining factors. *Soil Science and Plant Nutrition*, 35(4), 597–607.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., et al. (2023). Caret: Classification and regression training. <https://cran.r-project.org/package=caret>
- Long, N. V., Assefa, Y., Schwalbert, R., & Ciampitti, I. A. (2017). Maize yield and planting date relationship: A synthesis-analysis for US high-yielding contest-winner and field research data. *Frontiers in Plant Science*, 8, 2106. <https://doi.org/10.3389/fpls.2017.02106>
- Ministerio de Agricultura y Desarrollo Rural (MADR). (2017). Evaluaciones agropecuarias municipales. Bogotá, Colombia. <http://www.agronet.gov.co>
- Marques da Silva, J. R., & Silva, L. L. (2008). Evaluation of the relationship between maize yield spatial and temporal variability and different topographic attributes. *Biosystems Engineering*, 101(2), 183–190. <https://doi.org/10.1016/j.biosystemseng.2008.07.003>
- Mendiburu, F. de. (2013). Statistical procedures for agricultural research (R package version 1.3-68). <https://cran.r-project.org/web/packages/agricolae/agricolae.pdf>
- Miao, J., & Niu, L. (2016). A survey on feature selection. *Procedia Computer Science*, 91, 919–926. <https://doi.org/10.1016/J.PROCS.2016.07.111>
- Morales-Ruiz, A., Loeza-Corte, J. M., Díaz-López, E., Morales-Rosales, E. J., Franco-Mora, O., Mariezcurrena-Berasáin, M. D., & Estrada-Campuzano, G. (2016). Efficiency on the use of radiation and corn yield under three densities of sowing. *International Journal of Agronomy*, 2016, 6959708. <https://doi.org/10.1155/2016/6959708>
- Oke, O. F. (2016). Effects of agro-climatic variables on yield of *Zea mays* L. in a humid tropical rainforest agroecosystem. *African Journal of Agricultural Research*, 6(1), 148–151. <https://core.ac.uk/download/pdf/234664471.pdf>

- Ritchie, S. W., & Hanway, J. J. (1982). How a corn plant develops (Special Report No. 48). Iowa State University of Science and Technology, Cooperative Extension Service. <https://publications.iowa.gov/18027/1/How%20a%20corn%20plant%20develops001.pdf>
- Runge, E. C. A. (1968). Effects of rainfall and temperature interactions during the growing season on corn yield. *Agronomy Journal*, 60(5), 503. <https://doi.org/10.2134/agronj1968.00021962006000050018x>
- Sagarin, R., & Pauchard, A. (2010). Observational approaches in ecology open new ground in a changing world. *Frontiers in Ecology and the Environment*, 8(7), 379–386. <https://doi.org/10.1890/090001>
- Sandri, M., & Zuccolotto, P. (2006). Variable selection using random forests. In *Data analysis, classification and the forward search* (pp. 263–270). Springer. https://doi.org/10.1007/3-540-35978-8_30
- Sebastien, L., Julie, J., & Francois, H. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, 25, 1–18. <https://doi.org/10.18637/jss.v025.i01>
- Shmueli, G. (2010). “To Explain or to Predict?.” *Statist. Sci.* 25 (3) 289 – 310. <https://doi.org/10.1214/10-STS330>
- Westcott, N. E., Hollinger, S. E., & Kunkel, K. E. (2005). Use of real-time multisensor data to assess the relationship of normalized corn yield with monthly rainfall and heat stress across the central United States. *Journal of Applied Meteorology*, 44(11), 1667–1676. <https://doi.org/10.1175/JAM2303.1>