AGROSAVIA
Corporación colombiana de investigación agropecuaria

EGU General Assembly 2020

Universidad de Ibagué

Site-specific management zones delineation and Yield prediction for rice based cropping system using on-farm data sets in Tolima (Colombia)

Sofiane Ouazaa, Oscar Barrero, Yeison Mauricio Quevedo Amaya, Nesrine Chaali, Omar Montenegro Ramos

Sofiane Ouazaa
Ph.D. Irrigation Science
AGROSAVIA
souazaa@agrosavia.co

May 7th, 2020

MINAGRICULTURA GOBIERNO DE COLOMBIA

Colombia is facing mounting pressures on soil and water use:

i) Periodic variations in the availability of water that range from floods to drought periods, negatively affect soil, agricultural and livestock production systems;

ii) Water supply in drought period is limited for all uses;

iii) Competition for water will intensify among agricultural, urban, and environmental users;

iv) Short and long-term climate trends is exacerbating the problems associated with water scarcity;

Source: SSWM.

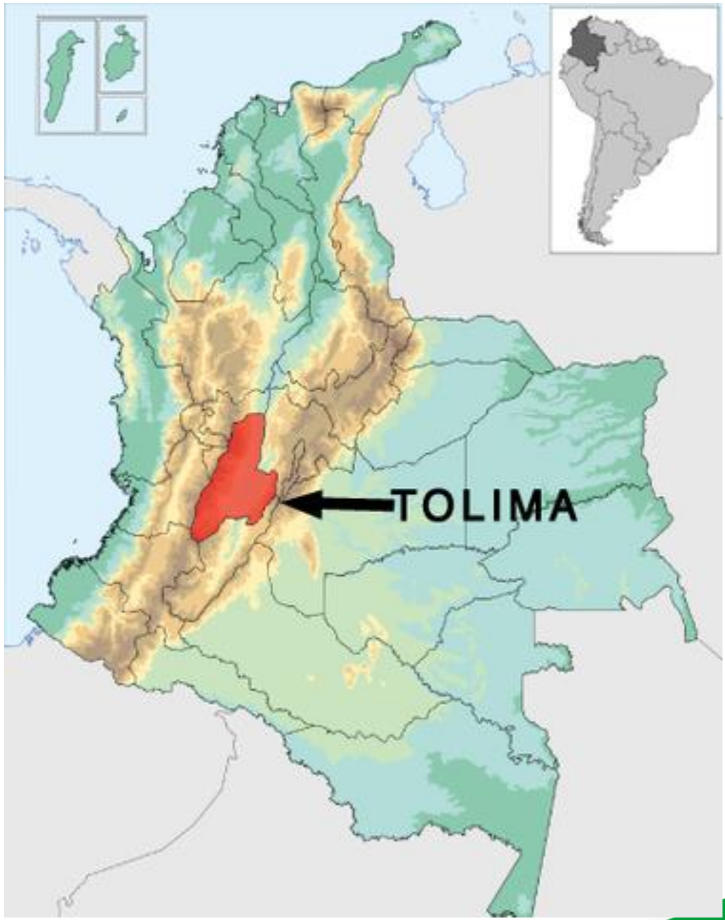
Objective

The aim of this study was

1) to predict rice yield using on farm data set and machine learning,
2) to compare delimited management zones (MZ) for rice-based cropping system with physiological parameters and within field variation yield.

Tolima Region

- ✓ <u>Located in the </u>Andina region.
- ✓ Tolima : Major rice producing department
- ✓ Most number of Agricultural Water District
- ✓ Very inefficient use of water resources
- ✓ No irrigation scheduling decision
- ✓ Soil degradation



Source: Colombia-SA.

EGU General Assembly 2020

Universidad de Ibagué

AGROSAVIA
Corporación colombiana de investigación agropecuaria

Problem Tree Analysis

High Within Field Yield Variability

Environmental issues of agricultural production

Soil degradation and irrigation water losses

Inefficient use of soil and irrigation water affecting farm profitability

Lack of knowledge and information about spatial variability of soil properties

Lack of knowledge and information about soil-water-plant-atmosphere continuum

Inefficient use of irrigation water

Colombian Farmers generally lack adequate means and incentives to know crops' water use, actual irrigation applications, crops' yield response to different water management practices.

Methodology

- Identification of spatial variability of soil properties and crop yield response within field

- A 72 sampling points spatially distributed were defined in a 5 hectares plot at the research center Nataima, Agrosavia.

- For each sampling point, physical and chemical properties, biomass and relative chlorophyll content were determined at different vegetative stages.

- A multispectral camera mounted to an Unmanned Aerial Vehicle (UAV) was used to acquire multispectral images over the rice canopy in order to estimate vegetation indices.

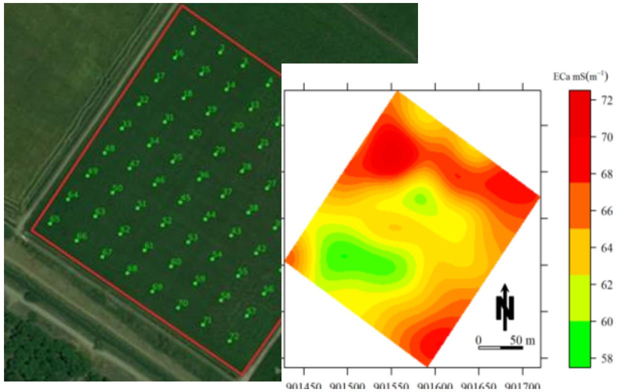
Methodology

- Five nonlinear models and two multilinear algorithms were employed to estimate rice yield. The fuzzy cluster analysis algorithm was used to classify soil data into two to six MZ.

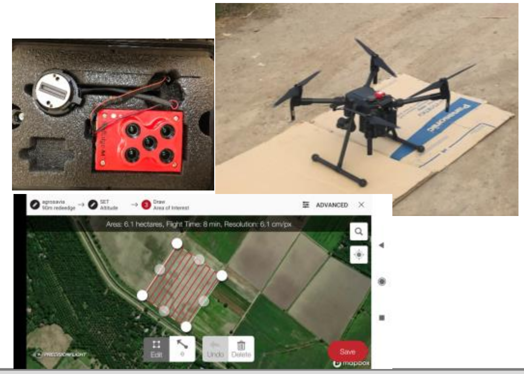
- The appropriate number of MZ was determined according to the results of a fuzziness performance index and normalized classification entropy.

Methodology

Universidad de Ibagué
AGROSAVIA
Corporación colombiana de investigación agropecuaria

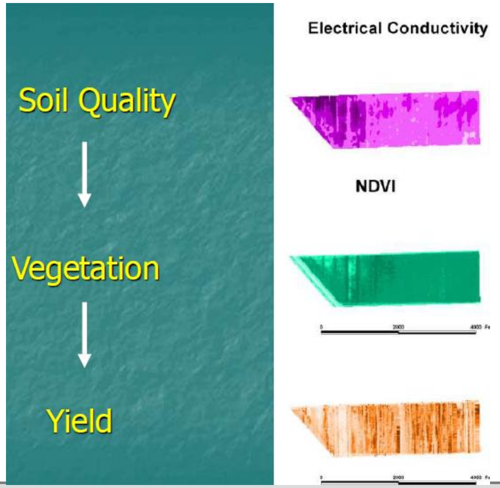


- Soil sampling and spatial distribution

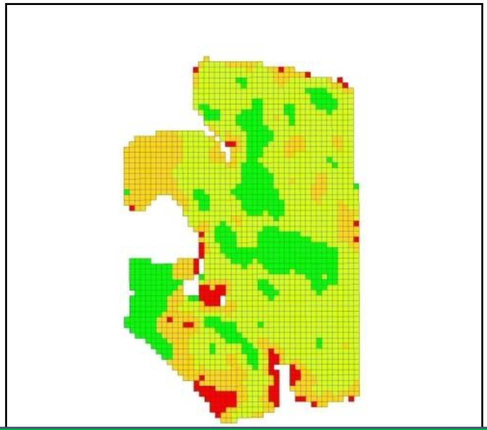


- UAV Remote sensing application
- Multispectral camera
- Crop health imagery

- Yield monitor
- GIS: computerized map to provide information



1) Prediction of rice yield
2) Definition of management zones



Prescription Maps as a Decision Support System

I- Rice Yield Prediction Model

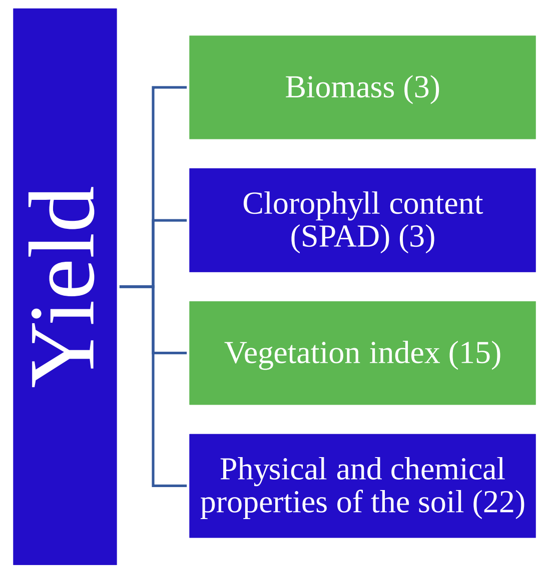
Rice Yield Prediction Model

Study Area

5 ha plot cropped with rice at the experimental farm of Agrosavia (CI Nataima)

Soil and Plant Sampling

Soil and plant sampling were carried out following a grid of 25 x 25 m^2, for 72 sampling points. The points were georeferenced and soil samplings were taken within two depths (0-10 cm, and 10-20 cm).



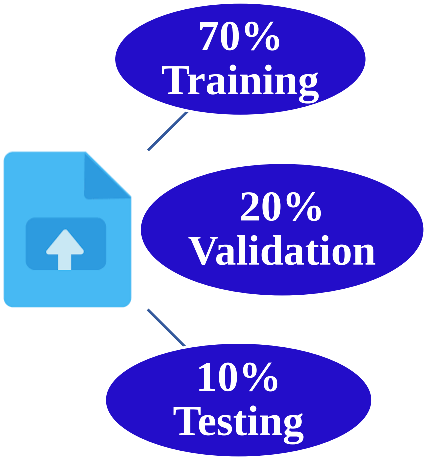
EGU General Assembly 2020

Universidad de Ibagué

AGROSAVIA

Corporación colombiana de investigación agropecuaria

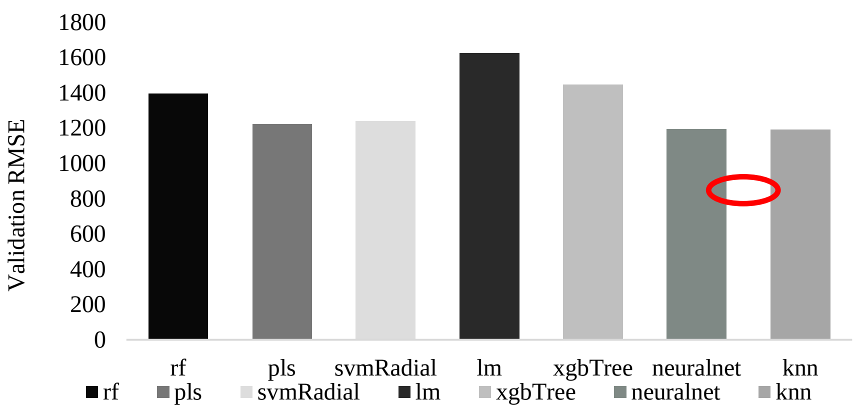
Database

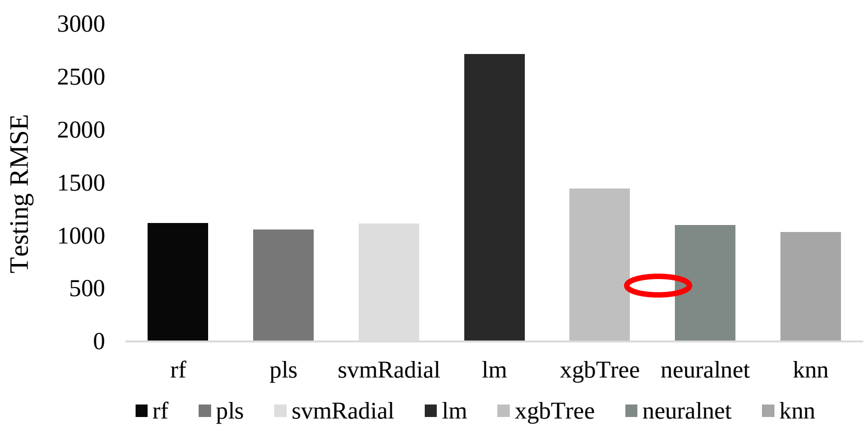


Machine Learning Regression Algorithms

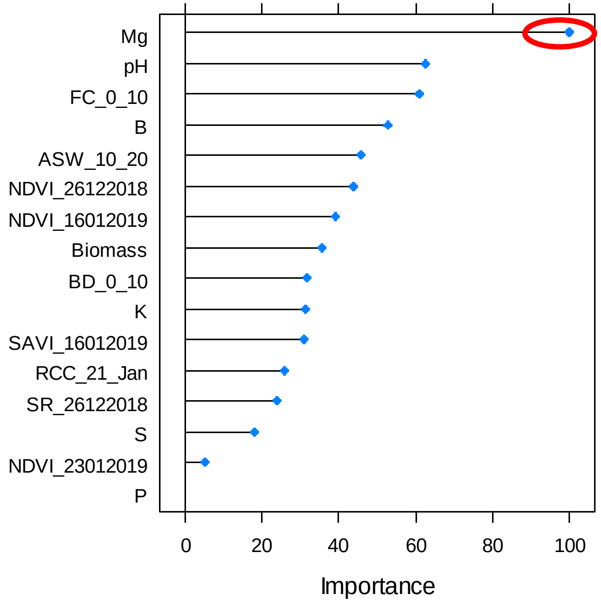
Random Forest	PLS	svmRadial	xgbTree
lm	Neuralnet	KNN	

Results

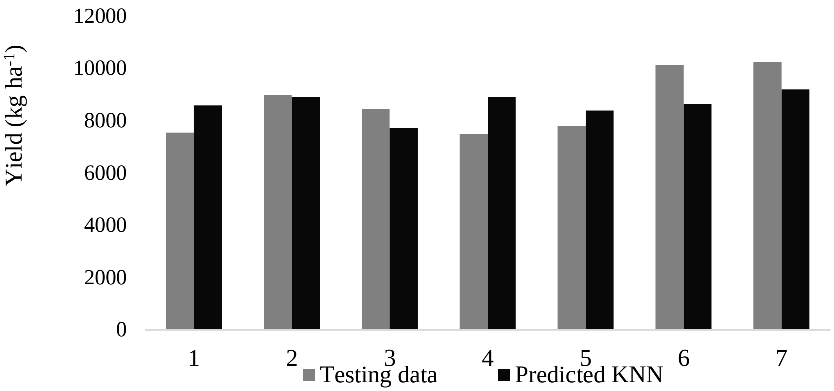




Root mean square error (RMSE) results for a) validation and b) testing



Importance of covariates used in KNN model.



Comparison between the real and predicted yield at testing points.

The worst case was the MLR with an RMSE of 2712.26 kg-ha^{-1} in the training dataset, while KNN regression was the best with 1029.69 kg-ha^{-1}.

KNN regression algorithm had an average absolute error of 10.74%

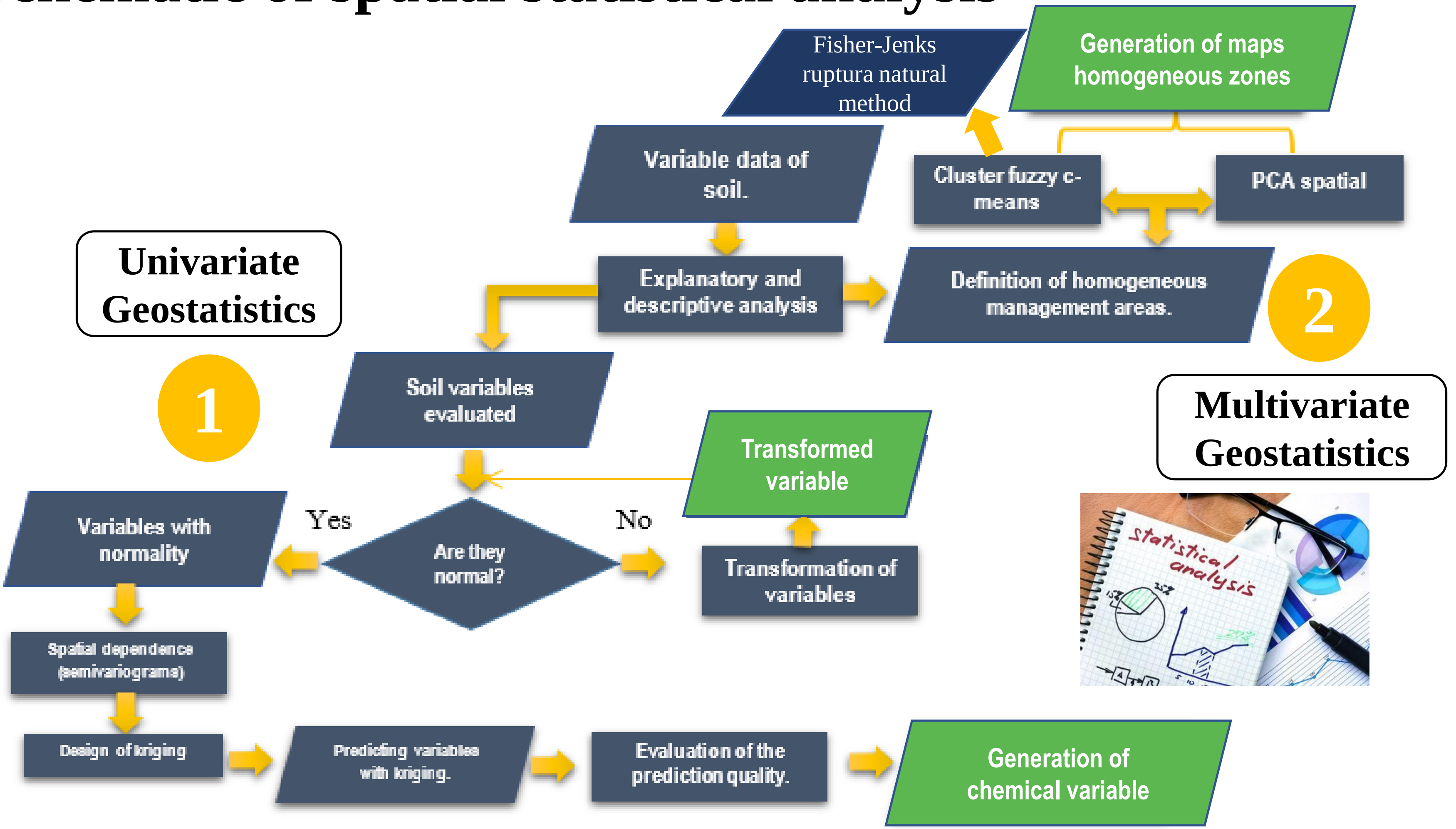
II- Definition of Management zones

EGU General Assembly 2020

Universidad de Ibagué

AGROSAVIA

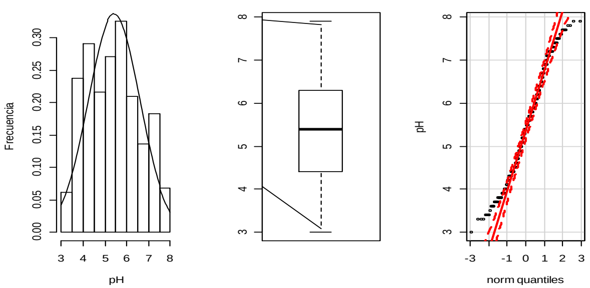
Schematic of spatial statistical analysis



Comparison of soil properties and yield between several zones throught mixed linear models.

Statistic analysis

1 Univariate Geostatistics

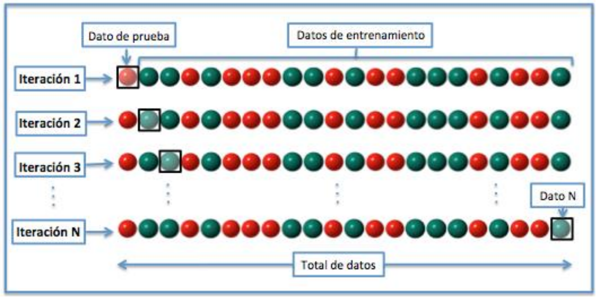


Dispersión

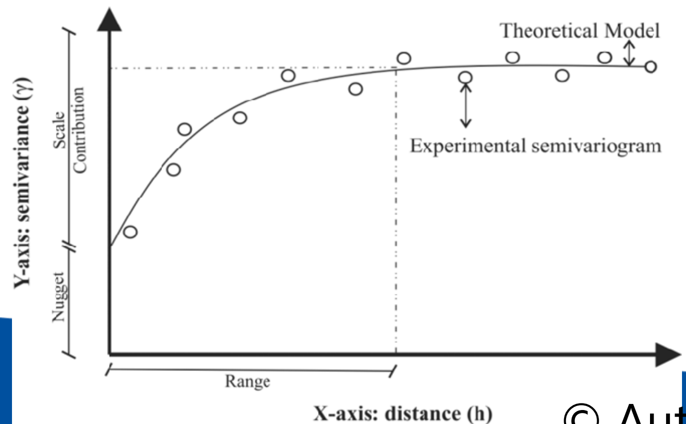
Tendencia central

Forma

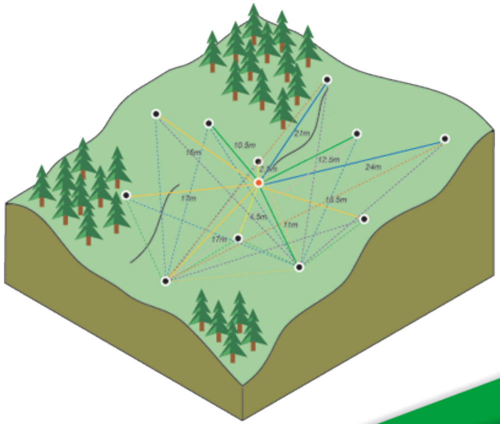
3 Estimation methods and selection of the best model (Cross validation)



2 Calculation of the experimental semivariogram



4 Calculation of interpolation by universal kriging (UK)

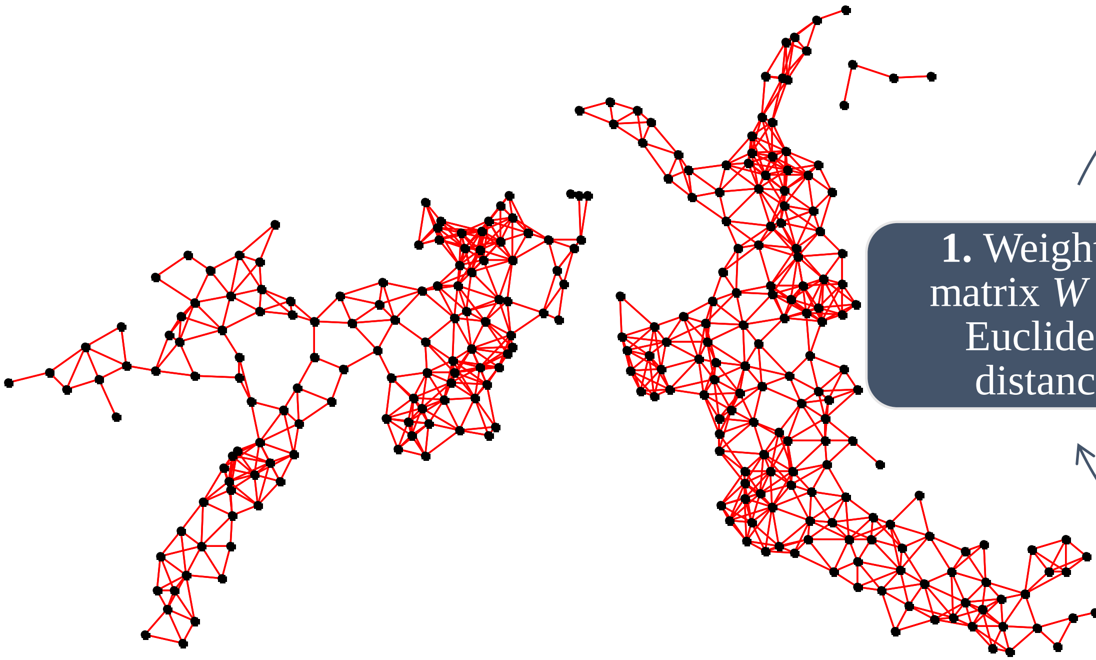


EGU General Assembly 2020

Universidad de Ibagué

AGROSAVIA
Corporación colombiana de investigación agropecuaria

Analysis of spatial principal components (MULTISPATI).



1. Weighting matrix *W nxn*. Euclidean distance.

2. The sPCs were calculated and the associated eigenvalues equivalent to the spatially structured variance were obtained.

3. Spatial autocorrelation in sPCs was analyzed with the Moran index.

4. UK on semivariograms of sPC1 and sPC2 interpolations.

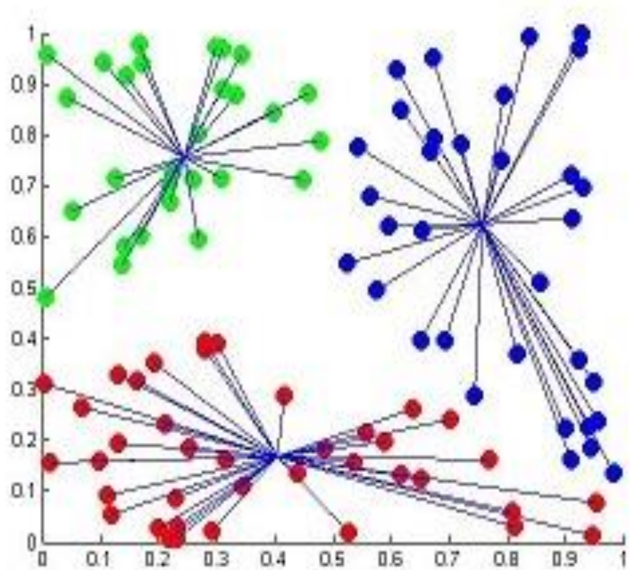
(Schabenberger & Pierce, 2002; Bivand, 2008; Dray, Said and Debias, 2008; Córdoba et al., 2012).

EGU General Assembly 2020

Universidad de Ibagué

AGROSAVIA
Corporación colombiana de investigación agropecuaria

Fuzzy c-means cluster algorithm



(Vendrusculo & Kaleita, 2011; Meyer et al., 2018; Behera et al., 2018).

1. Euclidean distance was used.

2. *Fuzziness* exponent of 1.3 Córdoba *et al.*, (2016), maximum number of iterations of 500

3. Between two and eight homogeneous zones. Internal validation measures.

4. Representation of homogeneous zones on the map.

For one sPC Fisher-Jenks natural rupture method was used.

Some of the packages used in R

R (R Core Team, 2019) version 3.5.2

sp	maptools	geoR
gstat	ggplot2	adespatial
spdep	lattice	e1071
nlme	classInt	

raster (Hijmans, 2017), sp (Pebesma y Bivand, 2005), maptools (Bivand y Lewin, 2017), geoR (Ribeiro y Diggle, 2016), gstat (Pebesma, 2004), ade4 (Dray y Dufour, 2007), spdep (Bivand y Piras, 2005), adegraphics (Siberchicot et al., 2017).

Results

Table 1. Chemical and hydrophysical properties plot IV-4 C.I. Nataima for the first 20 cm deep.

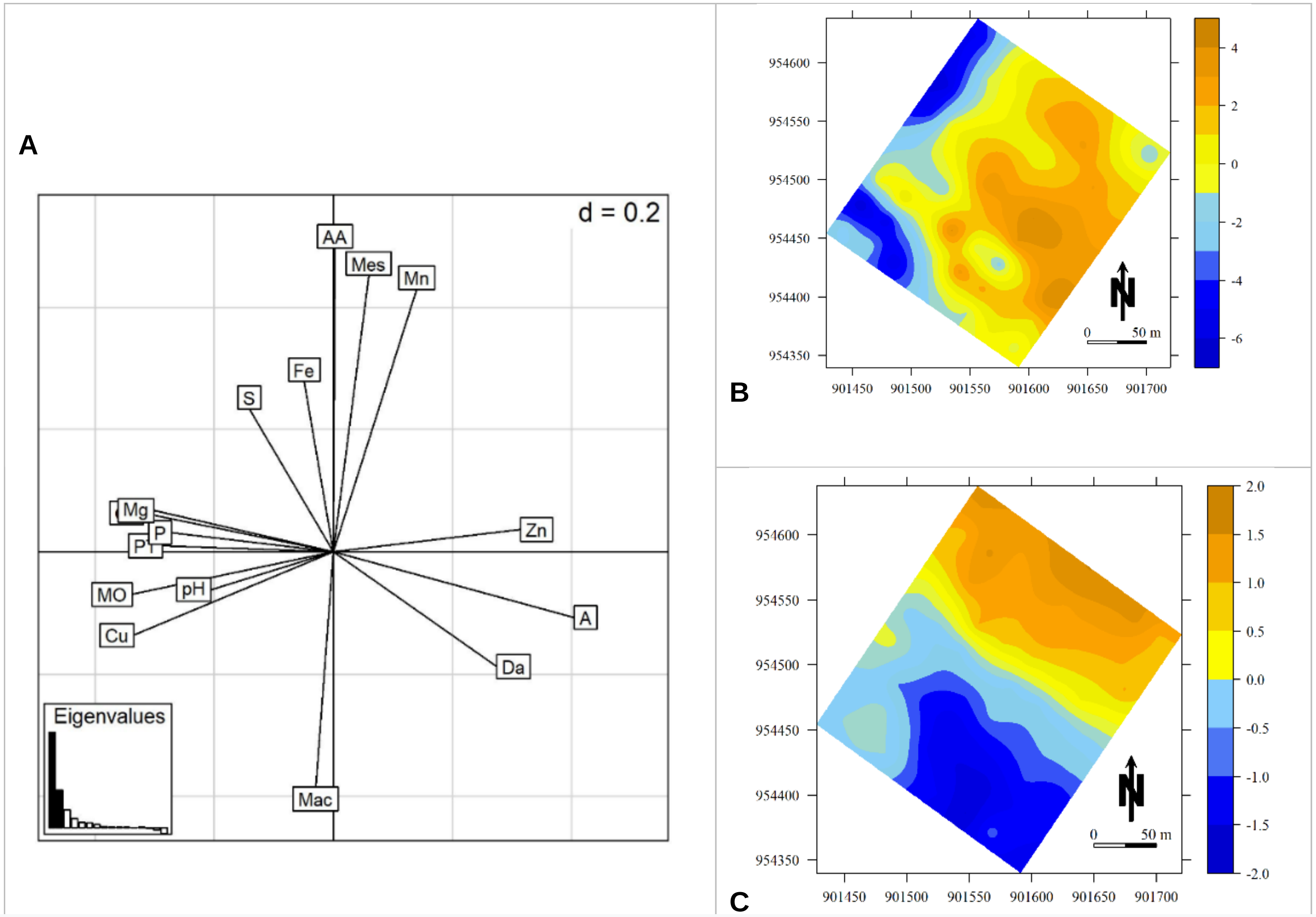
Variable	Unity	Average	Mín	Max	Skewness	Kurtosis	C.V. %
pH		6.07	5.57	6.50	-0.01	2.91	3.08
M.O.	%	1.28	0.92	1.97	0.83	2.93	19.82
P	mg kg^{-1}	24.72	11.39	52.69	0.80	2.94	38.26
S	mg kg^{-1}	11.65	6.16	21.86	0.69	3.33	28.75
Ca	Cmolc/kg^{1}	5.25	2.91	8.96	0.64	2.66	26.28
Mg	Cmolc/kg^{1}	1.55	0.90	2.45	0.29	2.40	22.05
K	Cmolc/kg^{1}	0.15	0.09	0.22	0.27	2.64	18.96
B	mg kg^{-1}	0.45	0.33	0.65	0.70	2.96	16.70
Fe	mg kg^{-1}	75.65	51.44	135.93	1.17	4.63	22.50
Cu	mg kg^{-1}	3.85	1.87	7.64	0.67	2.54	37.90
Mn	mg kg^{-1}	5.39	2.03	10.02	0.47	2.65	28.75
Zn	mg kg^{-1}	2.66	1.48	3.84	-0.13	1.98	22.71
Sand	%	49.00	5.06	68.60	-0.85	3.05	31.70
Clay	%	13.63	7.12	23.30	0.41	2.04	31.61
AW	%	7.32	4.60	10.21	0.09	1.96	20.48
BD	g cm^{-3}	1.56	1.23	1.84	-0.36	2.67	8.43
Macropores	%	3.67	1.56	15.13	3.66	24.02	48.57
Mesopores	%	6.51	3.21	10.71	0.22	2.63	23.30
Micropores	%	31.63	23.12	45.64	0.74	3.05	15.82
Total porosity	%	41.81	28.98	57.05	0.46	2.93	13.46

EGU General Assembly 2020

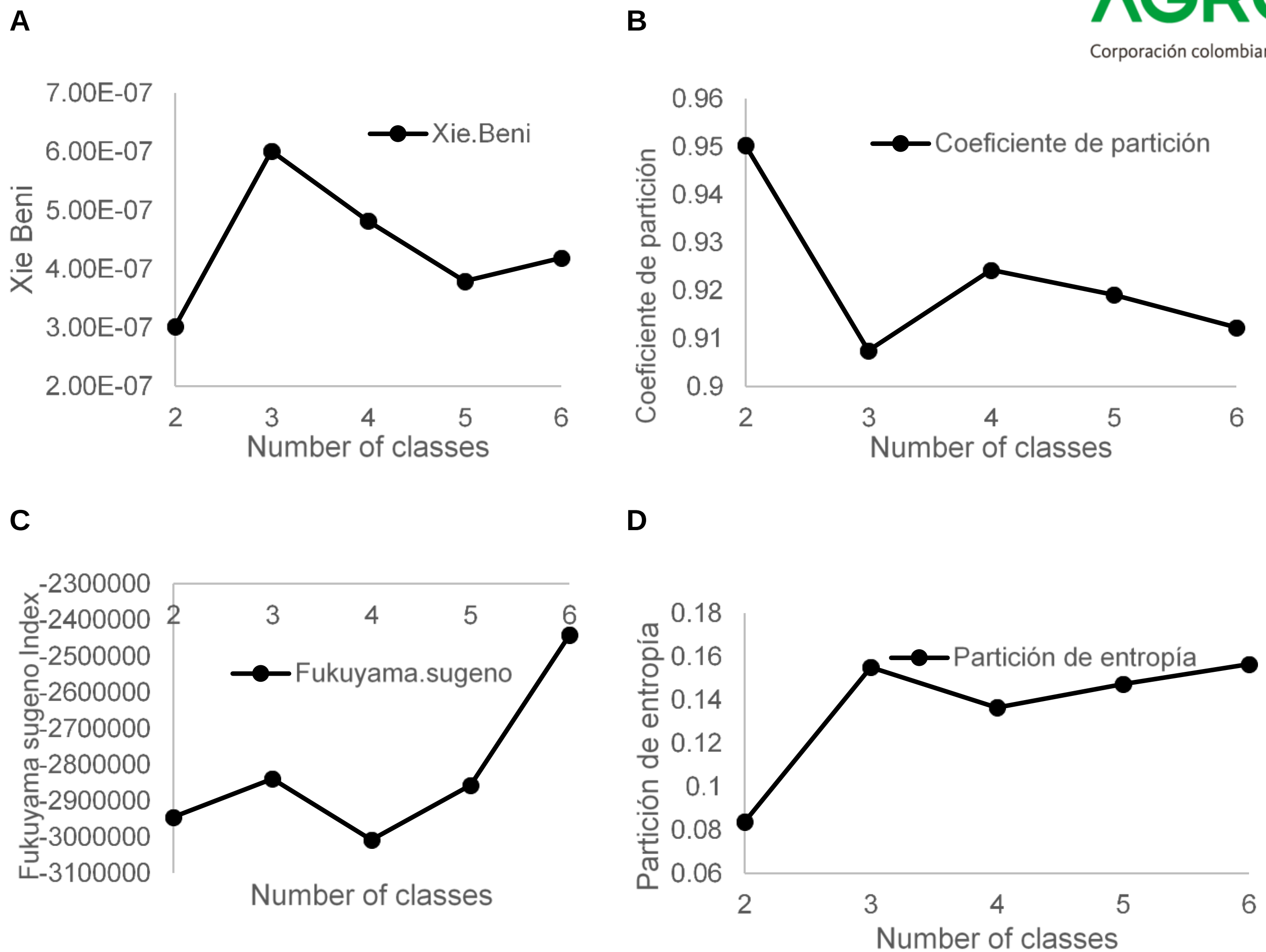
Universidad de Ibagué

AGROSAVIA
Corporación colombiana de investigación agropecuaria

Spatial principal components

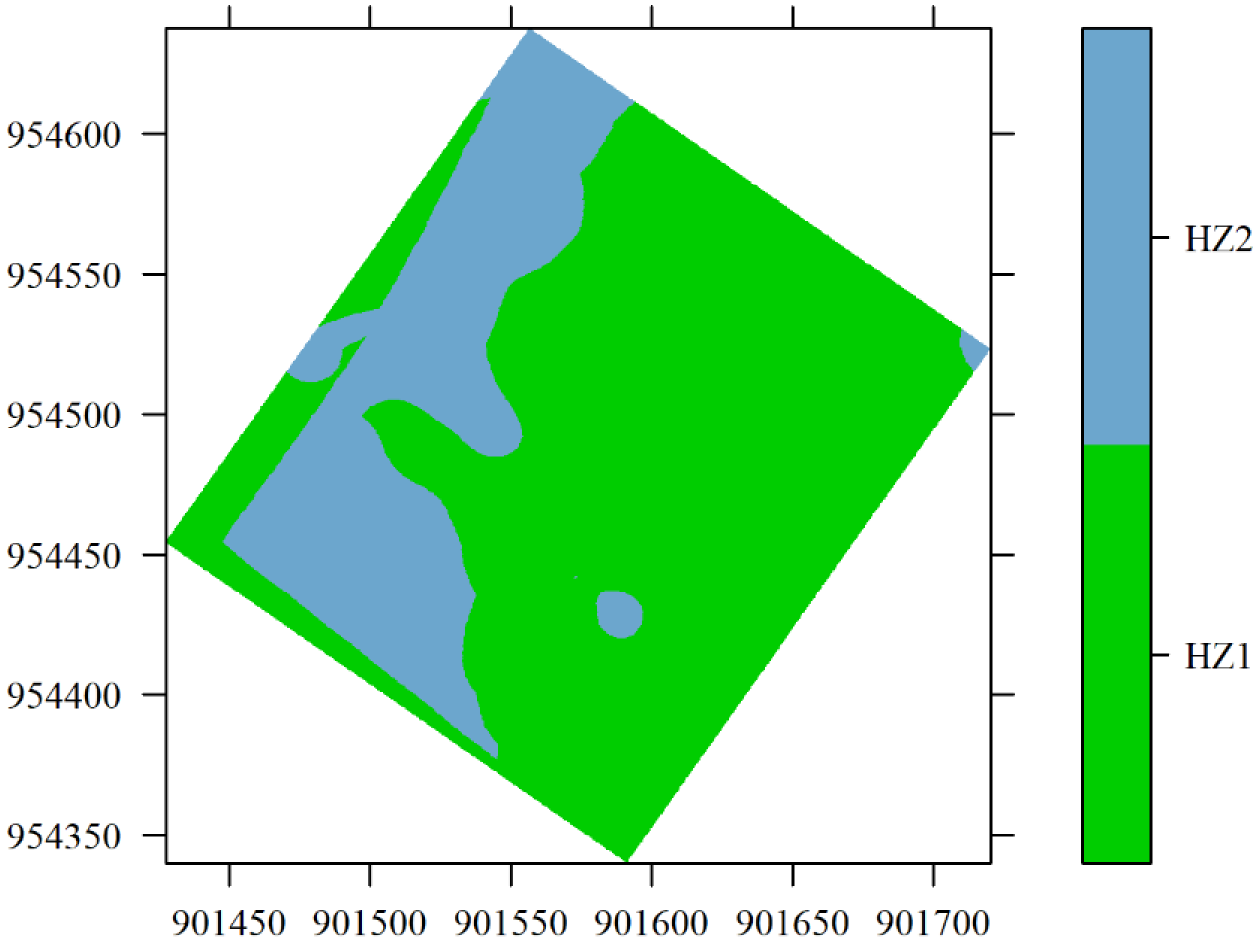


The principal spatial component analysis (sPC) showed significant spatial autocorrelation (p-value <0.001) in the first two sPC, with IM values of 0.721 and 0.629. The first component explained 62.25% of the spatial variability and the second 20.64%, for a total of 82.89% together for the two sPC.



Indices used to establish the optimal cluster number of the fuzzy k-means algorithm. A. Xie and Beni Index. B. Partition coefficient. C. Fukuyama-Sugeno D. Entropy partition.

Homogeneous zones



Hydrophysical and chemical properties of the homogeneous zones defined by the fuzzy k-means algorithm in the first two main components of the study area.

	Zone 1		**Zone 2**			
Soil property	**Mean±SE**	**CV**	**Mean±SE**	**CV**	**T o W**	**p-value**
Area (ha)	3.24 (64.8%)		1.76 (35.2%)			
Sand (%)	**57.43±1.21**	**14.40**	**33.15±2.7**	**40.68**	390.00*	**<0.001**
AW (%)	7.36±0.22	20.08	7.23±0.31	21.61	0.36	0.7191
Bd (g cm-3)	**1.61±0.02**	**6.77**	**1.48±0.03**	**8.96**	**4.40**	**<0.001**
Macrop (%)	3.56±0.3	57.70	3.89±0.22	28.87	1064.0*	0.0731
Mesop (%)	6.69±0.24	24.16	6.16±0.25	20.52	1.44	0.1537
PT (%)	**39.76±0.63**	**10.92**	**45.66±1.16**	**12.75**	**-4.86**	**<0.001**
S (mg kg^{-1})	11.1±0.52	31.91	12.68±0.55	21.49	-1.96	0.0556
Mn (mg kg^{-1})	5.64±0.26	31.66	4.93±0.32	32.19	1.67	0.1003
pH	**6.03±0.02**	**2.74**	**6.15±0.04**	**3.31**	**-2.71**	**0.0084**
OM (%)	**1.17±0.03**	**15.46**	**1.49±0.05**	**16.54**	**-6.19**	**<0.001**
P (mg kg-1)	**21.31±1.09**	**35.18**	**31.15±1.91**	**30.63**	**-4.82**	**<0.001**
Ca (Cmolc/kg^{-1})	**4.67±0.15**	**21.39**	**6.33±0.27**	**21.42**	**-5.91**	**<0.001**
Mg (Cmolc/kg^{-1})	**1.41±0.04**	**19.80**	**1.82±0.06**	**16.18**	**-5.72**	**<0.001**
Fe (mg kg^{-1})	75.01±2.3	21.01	76.85±3.89	25.32	-0.43	0.6658
Zn (mg kg^{-1})	**2.91±0.06**	**15.25**	**2.18±0.11**	**26.38**	**6.04**	**<0.001**
Cu (mg kg^{-1})	**3.22±0.17**	**35.71**	**5.03±0.25**	**24.40**	**-6.25**	**<0.001**
B (mg kg^{-1})	**0.43±0.01**	**15.89**	**0.49±0.01**	**14.69**	**-3.61**	**0.0006**

AW= Available water. Bd= Bulk density. TP= Total porosity. OM=Organic Matter. CV=Coeficiente of variation. * Wilcoxon test

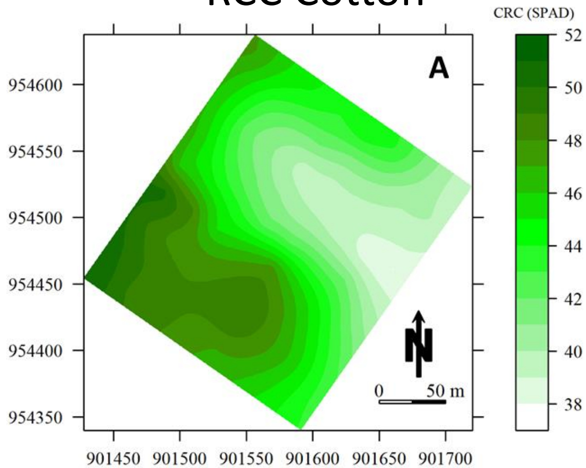
EGU General Assembly 2020

Universidad de Ibagué

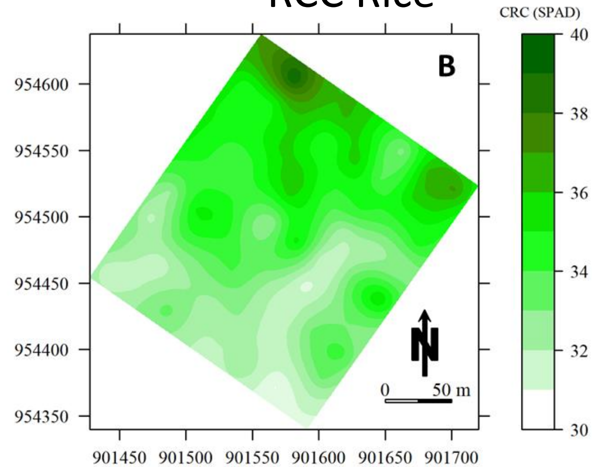
AGROSAVIA
Corporación colombiana de investigación agropecuaria

The relative chlorophyll content (RCC)of cotton and maize crops showed a similar spatial distribution pattern to delimited MZ

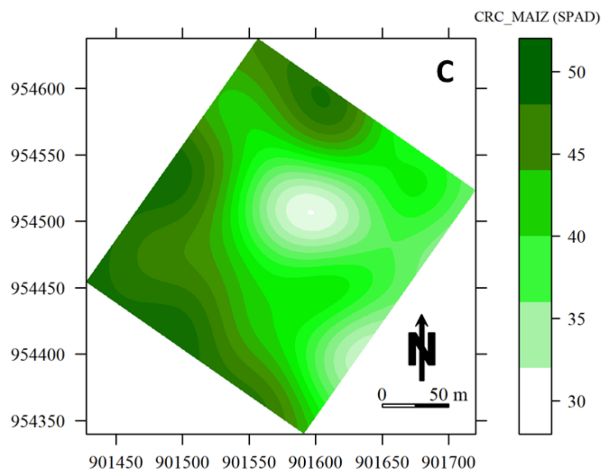
RCC Cotton



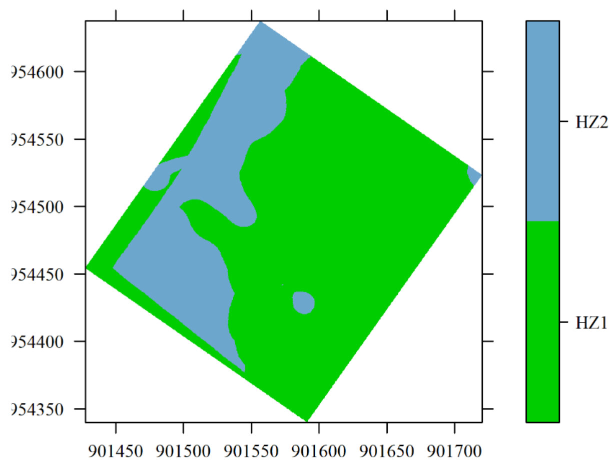
RCC Rice



RCC Maiz



Delimited HZ



Conclusion

Model yield Prediction

- The best performance rice yield prediction model was obtained by K-Nearest Neighbors (KNN) regression algorithm with an average absolute error of 10.74%.
- The Multiple Linear regression (MLR) showed the worst performance.
- These findings show the importance of machine learning could have for supporting decisions in agriculture processes management.

Definition of MZ

- The cluster analyses revealed that two zones was the optimal number of classes based on different criteria.
- Delineated zones were evaluated and revealed significant differences ($p \leq 0.05$) in some soil properties.
- The relative chlorophyll content of cotton and maize crops showed a similar spatial distribution pattern to delimited MZ.
- The results demonstrate the ability of the proposed procedure to delineate a farmer's field into zones based on spatially varying soil and crop properties that should be considered for irrigation and fertilization management.

THANK YOU