Assessing field experts yield maize estimations with satellite information derived from Sentinel-2

Diego Carcedo[1], Diego Pons[1,2], Cesar Alonso[3], Silvina Fiant[3] Marcelo Scavuzzo[1], Maria Victoria Marinelli[1]

[1] Instituto Gülich – Comision Nacional de Actividades Espaciales - CONAE
Ruta C45 - Km 8 - CP 5187 Falda del Cañete - Córdoba
dcarcedo@conae.gov.ar, scavuzzo@conae.gov.ar

[2] Instituto Nacional de Tecnologia Agropecuaria
Ruta Nac. 9 - Km 636 Estación Experimental Agropecuaria Manfredi, EEA, X5988 Manfredi, Córdoba, Argentina
pons.diego@inta.gob.ar

[3] Bolsa de Cereales de la Provincia de Cordoba
Av. Gral. Francisco Ortíz de Ocampo 317, FND, X5000 Córdoba, Argentina
calonso@bccba.org.ar, sfiant@bccba.org.ar

Abstract

Accurate and timely crop yield forecasts are critical for making informed agricultural policies and investments, as well as increasing market efficiency and stability. Earth observation data from space can contribute to agricultural monitoring, including crop yield assessment and forecasting. In this study, we present a multiple regression site-specific crop yield model based on the Normalized Difference Vegetation Index (NDVI) extracted from Sentinel – 2 data at 10 meters resolution calibrated with yield corn estimations reported by local experts at a field level.

1 Introduction

Argentina is a country in which the agro-industrial export sector, particularly the agricultural sector, participates to a large extent in the gross domestic product.[1] The main crops produced in the Pampa region are corn, wheat and soybeans. Particularly the corn, which participates with a large percentage in the national annual production [8], is the crop that has the highest initial production cost and is the most susceptible to reduce its production in the face of extreme climatic events.[5]

The anticipated knowledge of the yield of the corn crop would allow the producer to implement management measures both to improve the probable low yields or to consolidate estimation of high yields, as well as to design more convenient future commercialization strategies. Also, this information could be used for insurance companies to provide producers with new policies tied to the estimated returns. Finally, it would allow both the state and reference institutions in the sector to make and report projections on the total expected corn yield for a given year. It is discounted that this type of information would assist to stabilize prices with greater certainty in the market of purchase and sale of grain.[10]

The use of remote monitoring tools, in particular the processing of time series of images of Sentinel-2 sensor, from which to derive information on the phenological condition and physiology of crops is a frequent practice.[14] Studies of yield estimation with vegetation indexes derived from satellite information have shown the existing relationship between vegetation indexes evaluated in the growing season and yield.[13][7]

The objective of the present work is to study the relationship of field estimations made by local field experts in collaboration with a local market institution *"Bolsa de Cereales de Cordoba" (BCCBA)* and a time series of vegetation index images of Sentinel-2 sensor.

2 Methodology

2.1 Study area

The area under study includes the departments of Río Primero and Marcos Juárez of the Province of Córdoba. The territory of Marcos Juárez belongs to the Pampeana phytogeographic region, it has a moderate to gently rolling plain, with slope gradients ranging between 3 and 0.5. The annual average temperature is 17ºC and the thermal amplitude of 14ºC, with a frost-free period that reaches 257 days. The average annual precipitation is distributed in a range of 850 mm to the west and exceeds 900 mm to the east. The highest percentage of rainfall is observed during the summer season. However, the water balance only shows excesses during part of spring, autumn and early winter. The water deficit presents a variation of 100 mm to the East and 160 mm to the West.[4] This makes the area particularly suitable for agricultural activities.
The territory of Río Primero belongs to the Chaqueña phytogeographic region. It presents soils with good physical and chemical conditions for agricultural use, but which are fragile once they have been destitute of the vegetation cover under which they were developed. The regional pluviometry has a distribution with a range of 750 mm to the west and 800 mm to the east, with a seasonal distribution of monsoon type. The water deficit presents a variation of 180 mm to the East and 240 mm to the West.[4]

2.2 Data

The BCCBA has trained over the last 10 years local experts to provide every 15 days a yield forecast from a field level. These experts constitute a net that is regularly fed with 200 locality specific reports. We studied 16 localities of these data set, 8 corresponding to Río Primero department (RP) and 8 localities of Marcos Juárez department (MJ). We used the last report received before harvest (Column RTO of the Table 1). As Chipanshi (2015) states the critical period of the corn crop is strongly related to the final yield performance. [2]The Office of Environmental Risk, of the National Agroindustry Secretary studying the water deficit and the crop water consumption (Kc) established the critical period for maize 20 days before and 10 days after flowering. [3]
In the Cloud Computing service of Google Earth Engine (GEE), a serie of 14 cloud-free images of Sentinel-2 was identified for the 11/1/2016 - 20/2/2017 maize critical period, and 13 images were selected for the same period of the 2017/2018 campaign. The NDVI was calculated for each image of the serie. The integral, mean, variance and maximum were extracted at pixel level for each serie.
An agricultural land use cover layer built up by BCCBA, where the crops presented in each paddock for every season are detailed at parcel level, was used to determine the corn coverage. The spatial mean per maize paddock was computed for each variable derived from NDVI for a 5 km locality buffer. (Table 1).
The productivity index (PI) is a basis for judgment developed by *Instituto Nacional de Tecnologia Agropecuaria(INTA)* that takes acount numerous properties of the soil like: Drainage, effective depth, surface texture, subsurface texture, salinity, alkalinity, content of organic matter, capacity of cation exchange, slope, stoniness and rockiness, water erosion and current and potential wind. Is an index that relates a surface area to its potential crop productivity [12]

2.3 Analysis

A correlation model analysis was carried out between the 4 variables (i.e integral, mean, variance, maximum) and the performance reports of the collaborators of the BCCBA of corn for 2017 and 2018 for each locality under study.
Finally, a multiple regression model was constructed in order to determine the effect of year and the different localities.

DEPARTMENT	LOCALITIES	YEAR	NDVI				PI	YIELD
			INT	VAR	MEAN	MAX		
Marcos Juárez	Arias	2017	19.87	0.061	0.503	0.808	55.5	109
		2018	11.536	0.022	0.561	0.745	55.516	88
	Corral de Bustos	2017	18.362	0.06	0.487	0.818	84.259	90
		2018	10.877	0.025	0.577	0.766	84.46	85
	Inriville	2017	14.769	0.059	0.401	0.794	74.651	113
		2018	9.825	0.029	0.527	0.762	79.832	75
	Los Surgentes	2017	15.617	0.059	0.417	0.79	66.094	115
		2018	10.218	0.033	0.54	0.773	74.589	120
	Monte Buey	2017	16.495	0.059	0.442	0.793	86.484	114
		2018	10.001	0.032	0.53	0.764	86.019	85
	Camilo Aldao	2018	9.454	0.034	0.502	0.759	75.007	120
		2017	16.817	0.064	0.446	0.814	83.279	115
	Leones	2018	11.852	0.029	0.575	0.778	85.258	74
		2017	18.905	0.067	0.47	0.808	85.033	98
	Marcos Juarez	2018	10.934	0.027	0.546	0.758	85.539	70
		2017	17.468	0.059	0.458	0.792	85.513	90
Río Primero	Capilla de los Remedios	2017	7.921	0.067	0.414	0.795	63.523	93
		2018	3.21	0.046	0.353	0.677	63.311	51
	Montecristo	2017	7.805	0.058	0.395	0.732	71.484	96
		2018	3.327	0.047	0.337	0.689	72.28	63
	Río Primero	2017	7.583	0.046	0.364	0.697	60.151	107
		2018	3.503	0.037	0.351	0.633	63.696	65
	Chalacea	2018	4.666	0.041	0.316	0.598	58.439	58
		2017	13.425	0.048	0.368	0.704	58.282	80
	El Crispin	2018	3.76	0.055	0.389	0.711	62.717	40
		2017	7.899	0.057	0.393	0.736	62.794	92
	Monte del Rosario	2018	3.649	0.055	0.375	0.698	62.659	69
		2017	7.46	0.05	0.377	0.703	62.076	115
	Piquillin	2018	3.341	0.051	0.358	0.706	72.518	83
		2017	7.393	0.057	0.373	0.754	70.848	75
	Villa Santa Rosa de Rio Primero	2018	3.661	0.039	0.376	0.642	43.05	57
		2017	8.486	0.05	0.416	0.712	42.759	87

Table 1: Grain yield, NDVI and Productivity index (PI) in 2017 and 2018 for corn paddocks in the Province of Cordoba, Argentina.

The general equation of the model is:

$$Y = \beta_0 + \beta_1 NDVI_\alpha + \beta_2 T + \beta_3 L + \epsilon \tag{1}$$

Where $NDVI_\alpha$ represents the four different variables modelled. T is the year effect, L is the locality effect and the ϵ is the error term.

3 Results

Linear models showed correlations between the variables (Figure 1) and indicated a significant relationship between the yield estimations and the NDVI variables (Table 2). The differences observed between the performance of the different variables was low. The standardized mean squared error is low, which indicates that the distribution of the data to the line of greatest adjustment is small (Table 2).

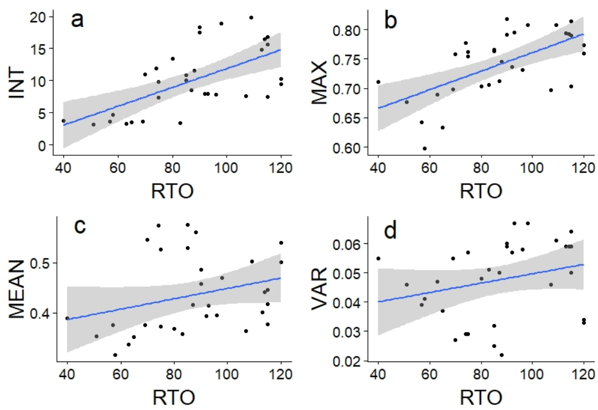

Figure 1: Linear models between NDVI derived variables and yield performance reported by experts (RTO). a) INT refers to integral NDVI variable. The data appears to be consolidated in the best fit regression line. b) MAX refers to the maximum NDVI variable. The data is condensed in the best fit line. c) MEAN refers to the mean NDVI variable. The data appears to be slowly correlated as is extensively distributed and outliers can be identified. d) VAR refers to the variance NDVI variable. The data is widely distributed, presenting outliers.

Variable	**Model**	**Statistics**		
		R2	**RMSE**	**NRMSE**
INTEGRAL	m1=lm(INT)∼YIELD	0.38**	3.65	0.18
MEDIA	m2=lm(MEAN)∼YIELD	0.53*	0.07	0.22
VARIANZE	m3=lm(VAR)∼ YIELD	0.30*	0.009	0.25
MAXIMO	m4=lm(MAX)∼YIELD	0.40**	0.04	0.19

$^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Linear models adjusted between NDVI variables and yield.

The coefficient of determination R^2 was calculated (Table 2) for each model analyzed, as well as the mean squared error (RMSE), which indicates how concentrated the data is in the line of best fit. Since the mean square error depends on the scale of the data, the standardized mean square error was also calculated. (NRMSE = RMSE / (Max (dependent variable) - Min (Dependent variable)) (Table 2).

Studying the residuals of the linear models a patron was discovered, revealing that the models proposed were not suitable for represent the behaviour of the data. As its presented in the Equation (1) a multiple regression model was constructed, taking in account the effect of the years and the localities (Table 3). The results obtained, evidence a relation between the field experts yield maize estimations and the satellite information derived from Sentinel-2, particularly with the NDVI integral and the maximum variable. The low value of NRMSE (Table 3) reports the goodness of fit.

The effect of the year is due to the climatological differenced observed between the seasons. The 2016/2017 was a humid year for the northern part of the study area and a dried season for the southern part of the study area, meanwhile 2017/2018 campaign was a slightly above mean precipitation season for all the area. The effect of the locality was linearly modelled with the productivity index data (R^2 : $0.9632, P =< 0.01$). The high correlation observed explained the site-specific differences among the different localities.

Variable	Model	Statistics		
		R2	RMSE	NRMSE
INTEGRAL	m5=lm(MAX) ∼RTO + AÑO + LOCALIDAD	**0.95***	0.74	0.04
MEDIA	m6=lm(INT) ∼RTO + AÑO + LOCALIDAD	0.64	0.03	0.11
VARIANZE	m7=lm(VAR) ∼RTO + AÑO + LOCALIDAD	0.49	0.006	0.14
MAXIMO	m8=lm(MEAN) ∼RTO + AÑO + LOCALIDAD	**0.84***	0.01	0.06

$^{*}p < 0.01$

Table 3: The multiple regression models adjusted between vegetation index derived variables and reported yield.

4 Conclusions

Even when the relation between vegetation indexes derived from satellite information and field crop yield measures is already widely described in the literature [15] [11], as well as remote sensed calibration of crop models [9], the relation of satellite derived indexes and estimated yield field data has not yet been sufficiently explored. In this study a relationship was established between the variables derived from NDVI and the maize field estimates made by BCCBA experts for the 2016/2017 and 2017/2018 campaigns in the Marcos Juárez and Rio Primero departments.
The 2016/2017 season presented setbacks due to rainfall above the historical average in the areas of the department of Marcos Juarez. The anomaly of some extreme data of NDVI (e.g. Marcos Juarez, Camilo Aldao) may be due to this excess of water in the area. The campaign 2017/2018 in general was a good campaign in which no floods or droughts were reported in the main agricultural areas.
The multiple regression proved to be a suitable model to capture the variability of maize yield experts reports and its terrain dispersion.

5 Future works

The performance of different vegetation indexes (i.e LAI, EVI, TNDVI), the behaviour of the model using co-variables (i.e drainage, salinity, rainfall), as well as, calibration of locally performance estimation models fed with vegetation indices (e.g ARYA) and processed based models. (e.g SAFY) will be explored.[6]

References

[1] Guillermo D Andrea. La Contribución del Agro Mitos y Realidades. *Ciclo Alumni Agronegocios*, 2019.

[2] Aston Chipanshi, Yinsuo Zhang, Louis Kouadio, Nathaniel Newlands, Andrew Davidson, Harvey Hill, Richard Warren, Budong Qian, Bahram Daneshfar, Frederic Bedard, and Gordon Reichert. Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape. *Agricultural and Forest Meteorology*, 206(June):137–150, 2015.

[3] J L Dardanelli, D J Collino, M E Otegui, and V O Sadras. Bases funcionales para el manejo del agua en los sistemas de producción de los cultivos de grano. *Producción de granos Bases funcionales para su manejo*, Editorial(JANUARY 2003):383, 2003.

[4] Carlos Daza Ghida and Carina Sánchez. Zona agroeconómica homogénea. Córdoba. page 268, 2009.

[5] Bolsa de Cereales de Cordoba. Departamento de Economia. Informe de Mercados Agricolas 31, 2019.

[6] B. Franch, E. F. Vermote, S. Skakun, J. C. Roger, I. Becker-Reshef, E. Murphy, and C. Justice. Remote sensing based yield monitoring: Application to winter wheat in United States and Ukraine. *International Journal of Applied Earth Observation and Geoinformation*, 76(July 2018):112–127, 2019.

[7] Ferran Gascon. Sentinel-2 for agricultural monitoring. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2018-July:8166–8168, 2018.

[8] INDEC. Informe de Avance del Nivel de Actividad - Abril 2019, 2009.

[9] Xiuliang Jin, Lalit Kumar, Zhenhai Li, Haikuan Feng, Xingang Xu, Guijun Yang, and Jihua Wang. A review of data assimilation of remote sensing and crop models, 2018.

[10] David M. Johnson. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment*, 141:116–128, 2014.

[11] David M. Johnson. A comprehensive assessment of the correlations between field crop yields and commonly used MODIS products. *International Journal of Applied Earth Observation and Geoinformation*, 52:65–81, 2016.

[12] Gorgas Juan Antonio. *INTA – Secretaria de Ambiente de Córdoba*, (2011).

[13] Marie Julie Lambert, Pierre C.Sibiry Traoré, Xavier Blaes, Philippe Baret, and Pierre Defourny. Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. *Remote Sensing of Environment*, 2018.

[14] Chunhua Liao, Jinfei Wang, Taifeng Dong, Jiali Shang, Jiangui Liu, and Yang Song. Using spatio-temporal fusion of Landsat-8 and MODIS data to derive phenology, biomass and yield estimates for corn and soybean. *Science of the Total Environment*, 650:1707–1721, 2019.

[15] Sergii Skakun, Eric Vermote, Jean-Claude Roger, and Belen Franch. Combined Use of Landsat-8 and Sentinel-2A Images for Winter Crop Mapping and Winter Wheat Yield Assessment at Regional Scale. *AIMS Geosciences*, 3(2):163–186, 2017.