Seasonal forecasting of impacts of droughts on agriculture in Mexico using a principal component regression approach

Roberto A. Real-Rangel (rrealr@iingen.unam.mx)[1,2], Adrián Pedrozo-Acuña[1,3], J. Agustín Breña-Naranjo[1,3]

[1] National Autonomous University of Mexico, Institute of Engineering, Coyoacan, CDMX, Mexico
[2] National Autonomous University of Mexico, Programme of Master and Doctorate in Engineering, Coyoacan, CDMX, Mexico
[3] Mexican Institute of Water Technology, Jiutepec, Mor., Mexico

Seasonal forecasting of impacts of droughts on agriculture in Mexico using a principal component regression approach

Drought monitoring and forecasting allows impact mitigation measures to be adopted in early stages of an event to reduce the vulnerability of a wide range of environmental, economical and social sectors. In Mexico, there are drought monitoring systems at national and regional levels that follow up on these events, such as the Drought Monitor in Mexico, and the North American Drought Monitor, but seasonal drought forecasting is still a pending task.

The present study aims to fill this gap by applying a methodology that uses data derived from a globally available atmospheric reanalysis product and a principal component regression-based model aimed at predicting the impacts of droughts on rainfed crops associated with soil moisture deficits, estimated by means of the standardized soil moisture index (SSI).

Using the state of Guanajuato (Central-Northern Mexico) as a case study, the model generated yielded RMSE values of 0.74, using regional and global hydrological, climatic and atmospheric variables as predictors with four-month lead-time.

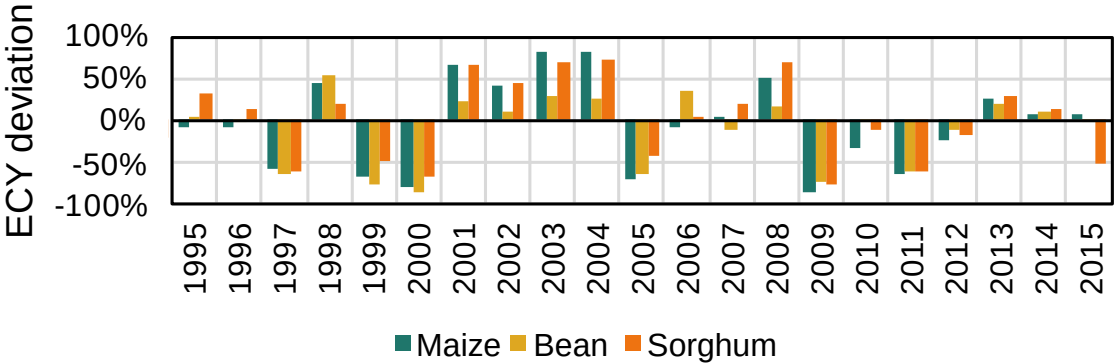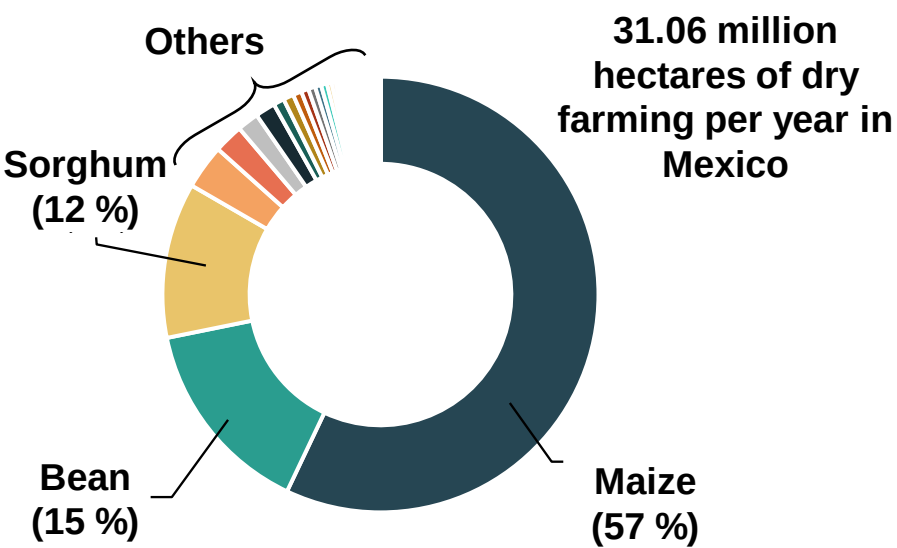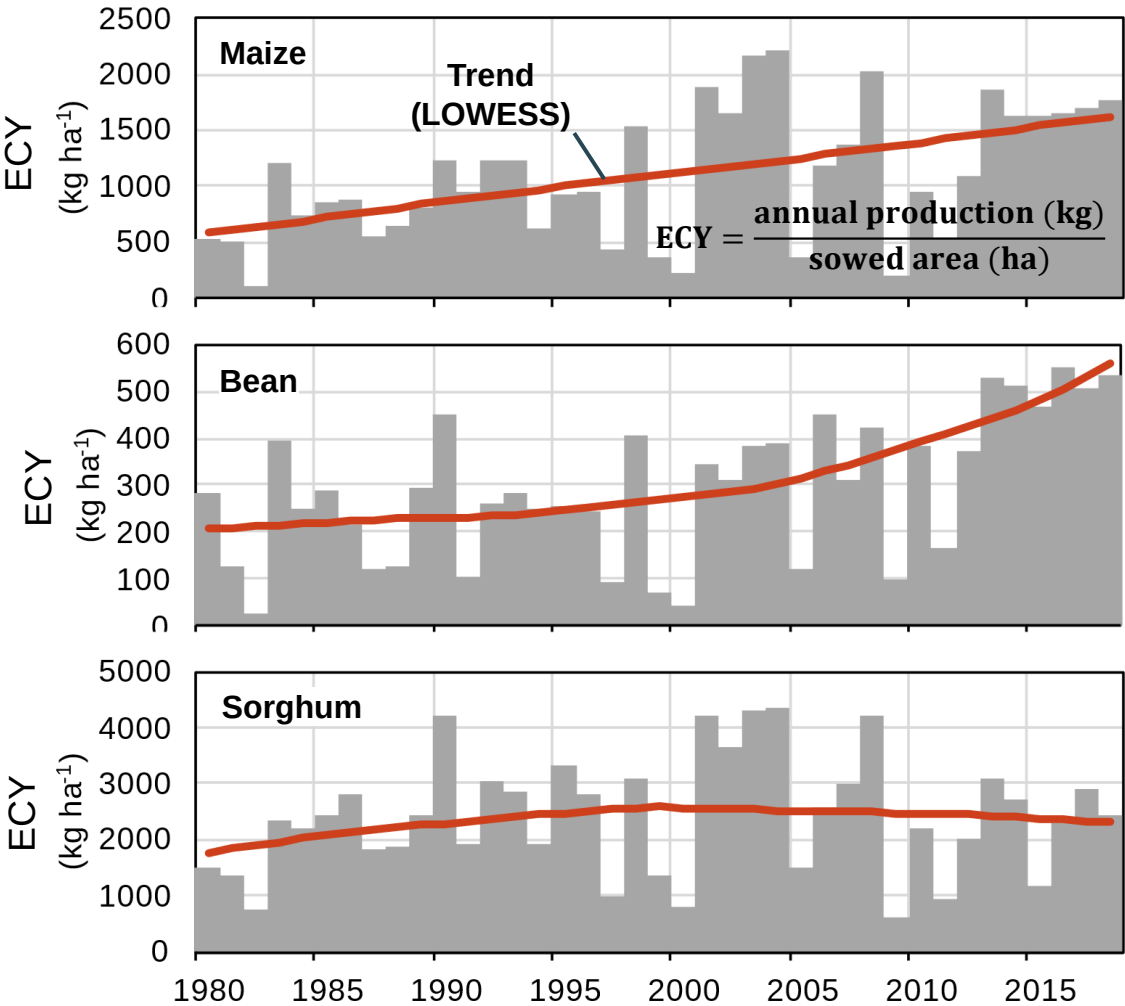Hydrological, climatic and atmospheric variables from MERRA-2

MERRA-2's vs. ground-based precipitation
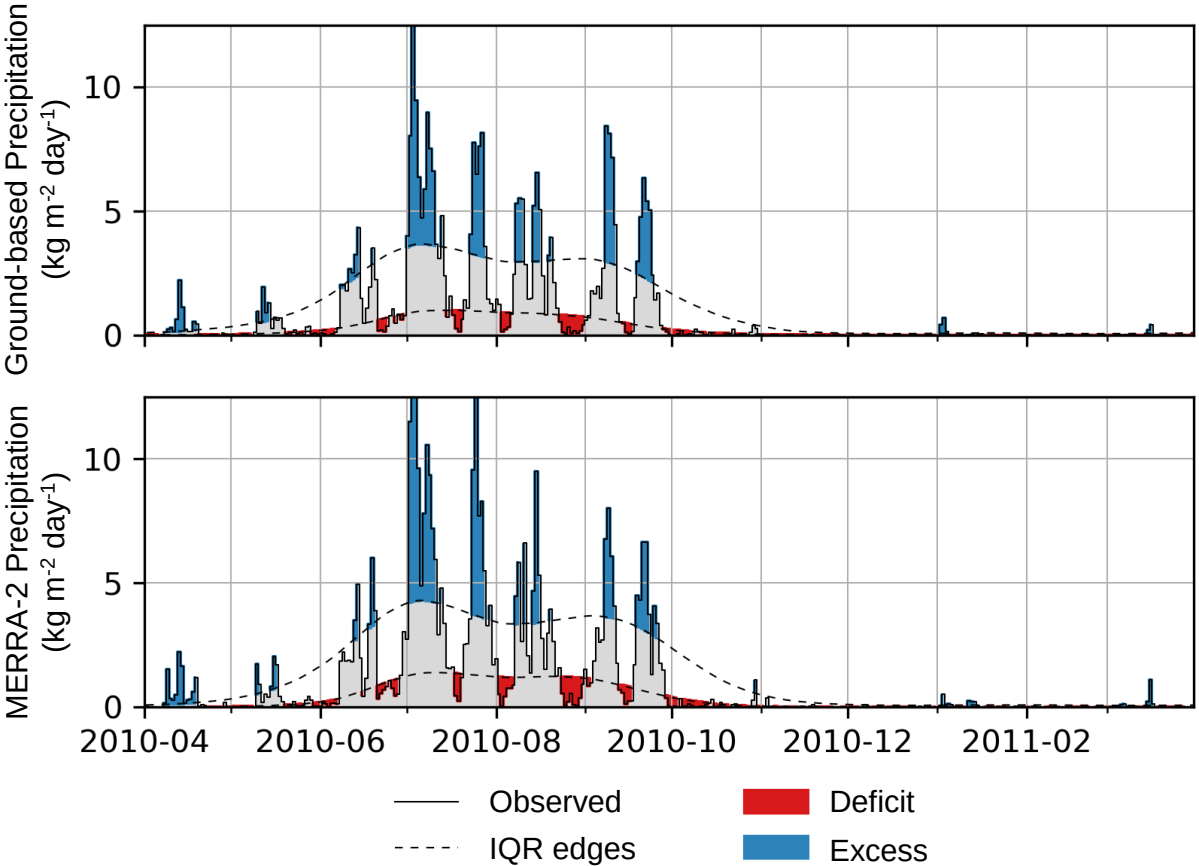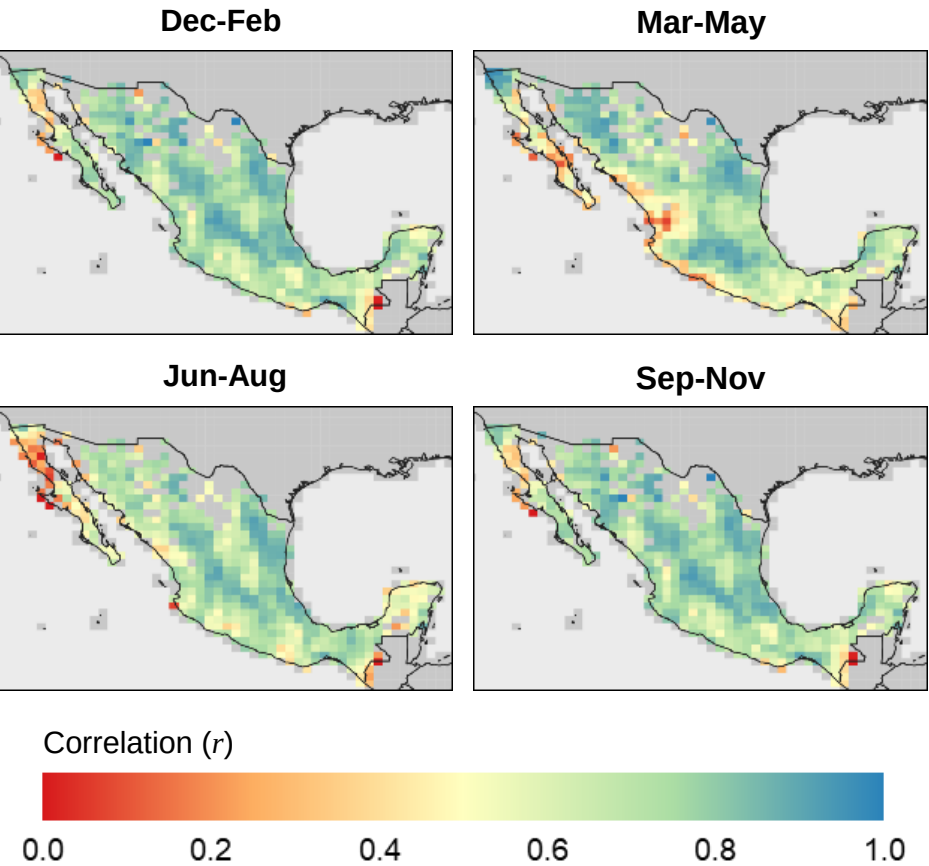
Gelaro, R., McCarty, W., Suárez *et al.* (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

IQR: Interquartile range

Annual agricultural production from SIAP of the Ministry of Agriculture and Rural Development of Mexico (SADER)

$$ECY = \frac{\text{annual production (kg)}}{\text{sowed area (ha)}}$$

31.06 million hectares of dry farming per year in Mexico

Maize (57 %)

Bean (15 %)

Sorghum (12 %)

Others

Maize — Bean — Sorghum

ECY: Effective crop yield

SSI settings with strongest correlations

Crop	Temporal scale	Month of calculation	r_S	95 % r_S-confidence limits	
				Lower	Upper
Maíze	7 months	Dec	0.79	0.63	0.89
Bean	3 months	Oct	0.56	0.25	0.79
Sorghum	7 months	Jan	0.65	0.39	0.83

Probability of detection
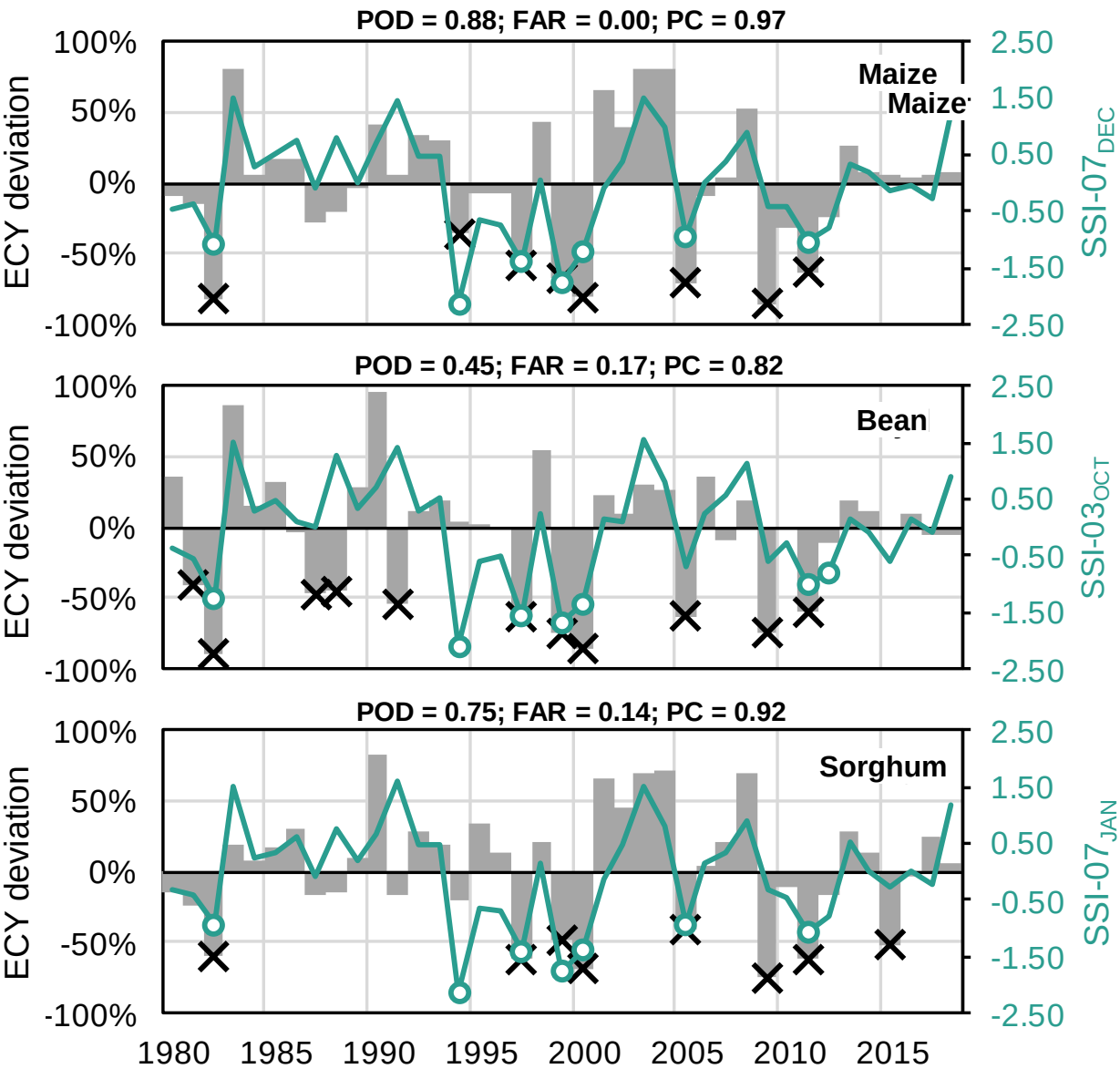
$$POD = \frac{a}{a+c}$$

False alarm ratio

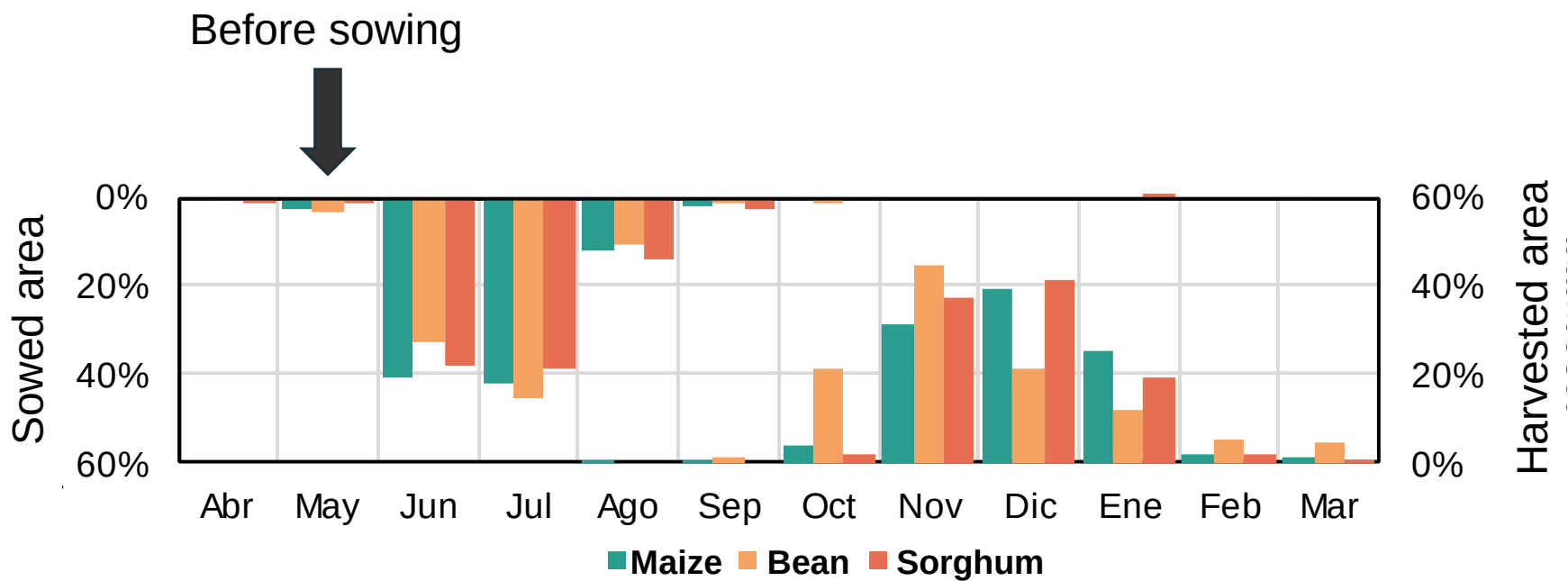$$FAR = \frac{b}{a+b}$$

Proportion correct

$$PC = \frac{a+d}{a+b+c+d}$$

	A reduction in ECY was observed	No reduction in ECY was observed
SSI indicated a drought	Hit (*a*)	False alarm (*b*)
SSI did not indicate a drought	Miss (*c*)	Correct negative (*d*)

ECY: Effective crop yield

POD = 0.88; FAR = 0.00; PC = 0.97 (Maize, SSI-07$_{DEC}$)

POD = 0.45; FAR = 0.17; PC = 0.82 (Bean, SSI-03$_{OCT}$)

POD = 0.75; FAR = 0.14; PC = 0.92 (Sorghum, SSI-07$_{JAN}$)

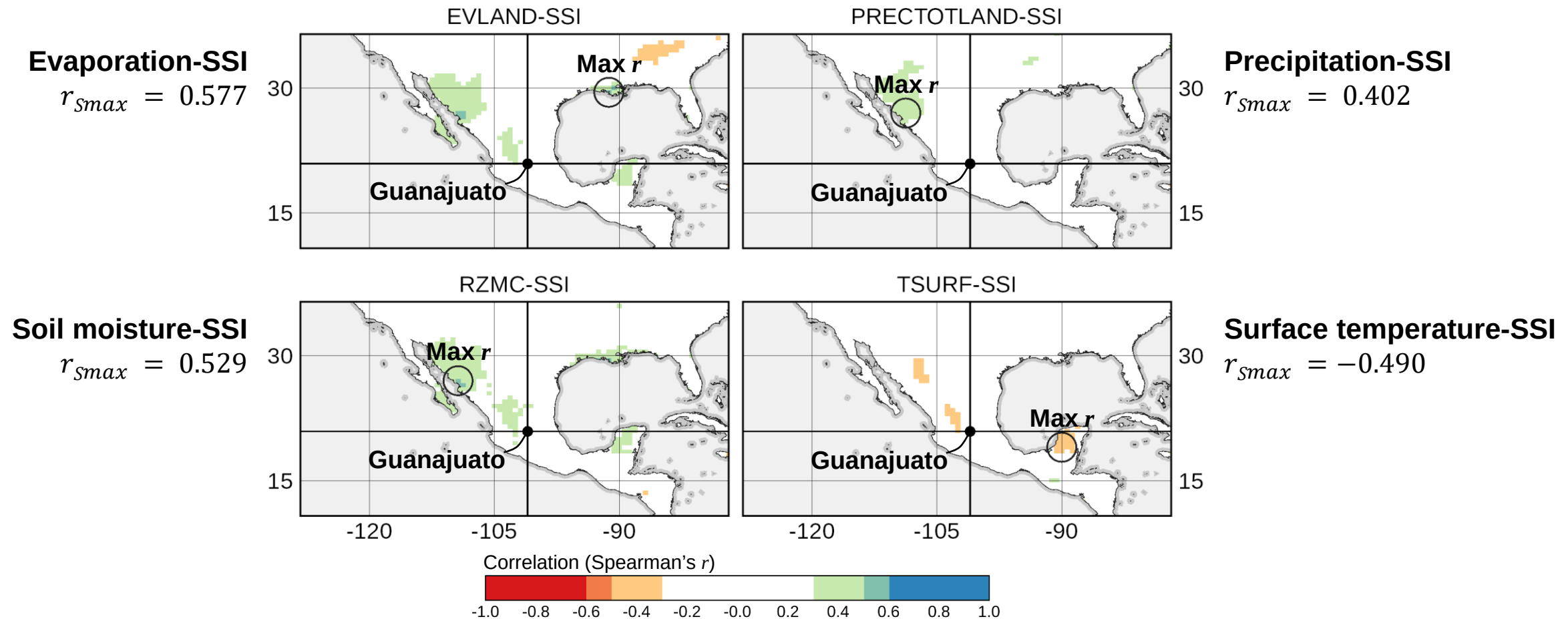**Seasonality of sowing and harvesting of maize, beans and sorghum**
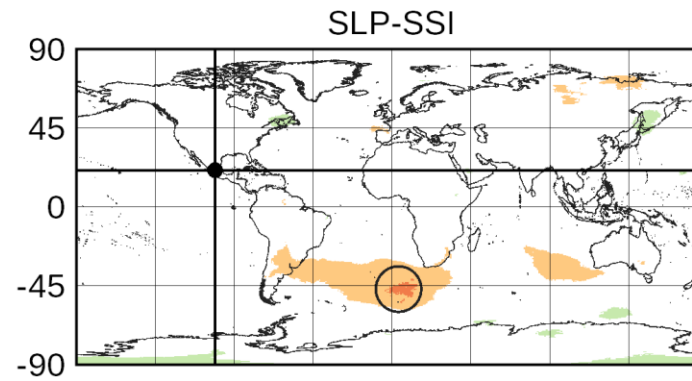

Average seasonality of maize, bean and sorghum crops in the spring-summer cycle in the state of Guanajuato (Mexico), in the period 2004-2018.

Variables aggregated in 3 months (February-April)

Evaporation-SSI
$r_{Smax} = 0.577$

Precipitation-SSI
$r_{Smax} = 0.402$

Soil moisture-SSI
$r_{Smax} = 0.529$

Surface temperature-SSI
$r_{Smax} = -0.490$

Atmospheric pressure-SSI
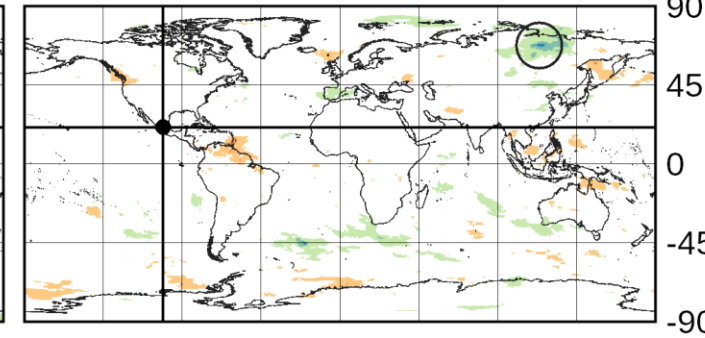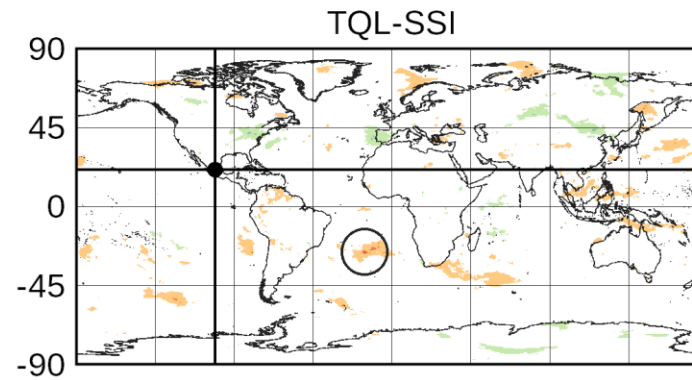
$r_{Smax} = -0.541$

Total precipitable ice water-SSI

$r_{Smax} = 0.641$

Total precipitable liquid water-SSI

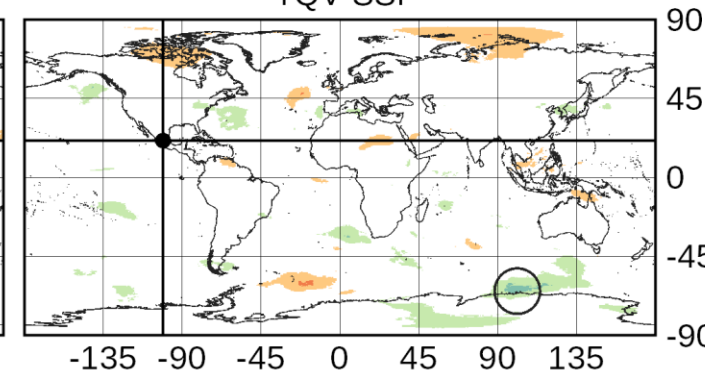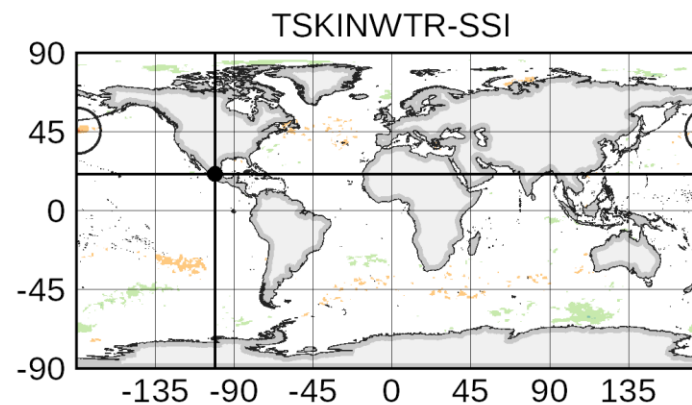$r_{Smax} = -0.627$

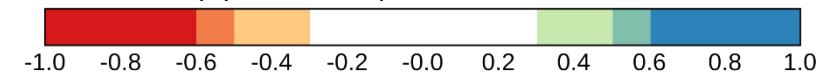**Total precipitable wáter vapor-SSI**

$r_{Smáx} = 0.583$

Sea surface temperature-SSI

$r_{Smáx} = -0.577$

Correlation (Spearman's r)

Explanatory variables for the variation of SSI-07$_{DEC}$ (impacts on maize) in the state of Guanajuato (Mexico) and their location.

Variable	r_S	Region	
		Box (x_{min}, y_{min}, x_{max}, y_{max})	Reference
Relations at regional level			
EVLAND	0.577	29.25, –91.563, 29.75, –90.938	Gulf of Mexico Coastal Plain (USA)
PRECTOTLAND	0.402	26.75, –109.063, 27.25, –108.438	Coast south of Sonora, Mexico
RZMC	0.529	26.75, –109.688, 27.25, –109.063	South of Sonora (Mexico)
TSURF	–0.490	18.75, –90.313, 19.25, –89.688	Yucatan Peninsula (Mexico)
Relations at global level			
SLP	–0.541	–47.25, 3.438, –46.75, 4.063	South Atlantic Ocean
TQI	0.641	67.25, 113.438, 67.75, 114.063	Central Siberian Plateau (Russia)
TQL	–0.627	–26.25, –15.938, –25.75, –15.313	South Atlantic Ocean
TQV	0.583	–65.25, 100.938, –64.75, 101.563	Davis Sea (Antarctic Ocean)
TSKINWTR	–0.577	45.25, –179.688, 45.75, –179.063	North Pacific Ocean

How many predictors?
Which predictors?
How many principal components?

Parameters and hiperparameters

Cross-validation

Best parameters

Data set

Training subset

Testing subset

Model training

Final evaluation

Example of iteration within the cross-validation process

EVLAND | PRECTOTLAND | RZMC | TSURF | SLP | TQI | TQL | TQV | TSKINWTR

EVLAND | SLP | TSKINWTR ← 3 predictors

1st CP | 2nd CP | 3th CP

1st CP | 2nd CP ← 2 principal components

$$\widehat{y} = b_0 + b_1 x_1 + b_2 x_2$$

How many predictors?
Which predictors?
How many principal components?

Parameters and hiperparameters

$$R^2(y, \widehat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

Cross-validation

Best parameters

Data set

Training subset

Testing subset

Model training

Final evaluation

Example of iteration within the cross-validation process

| EVLAND | PRECTOTLAND | RZMC | TSURF | SLP | TQI | TQL | TQV | TSKINWTR |

| EVLAND | SLP | TSKINWTR | ⬅ 3 predictors

| 1st CP | 2nd CP | 3th CP |

| 1st CP | 2nd CP | ⬅ 2 principal components

$$\widehat{y} = b_0 + b_1 x_1 + b_2 x_2$$

How many predictors?
Which predictors?
How many principal components?

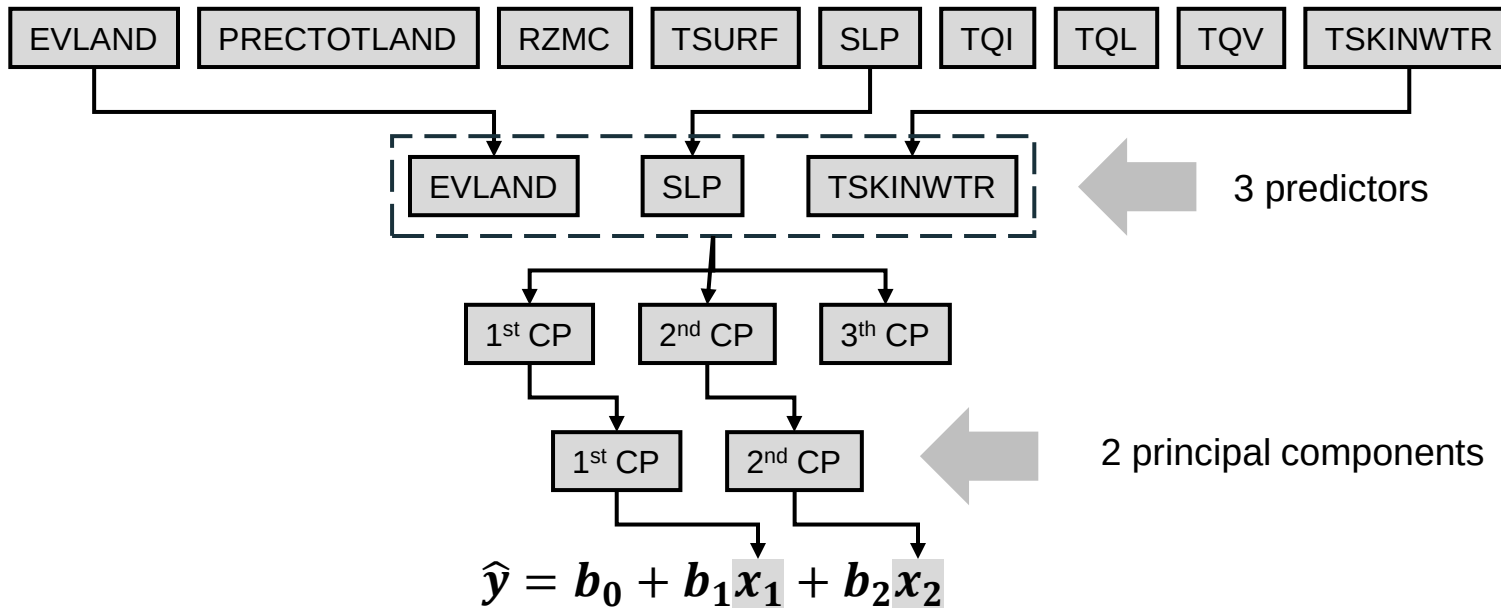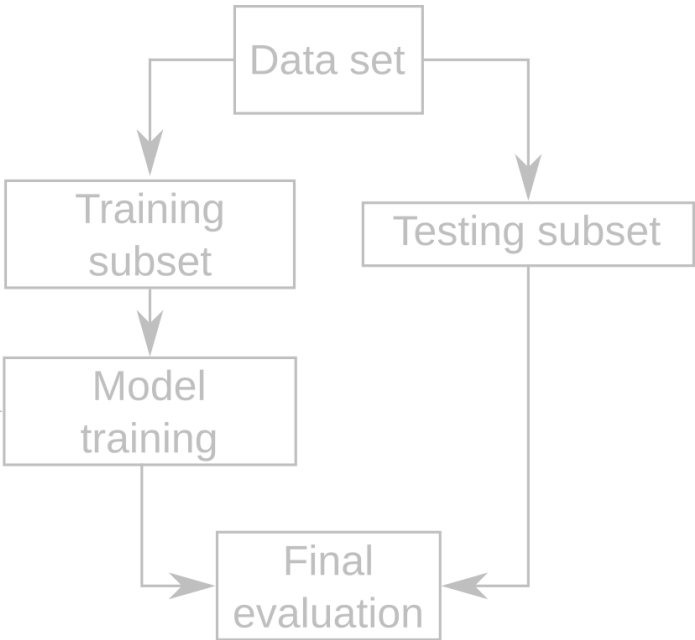Parameters and hiperparameters

Cross-validation

Best parameters

Data set

Training subset

Testing subset

Model training

Final evaluation

Original set

Training | Test

Training folds | Testing fold

1st it. $\rightarrow E_1$

2nd it. $\rightarrow E_2$

3th it. $\rightarrow E_3$

4th it. $\rightarrow E_4$

5th it. $\rightarrow E_5$

$E = \dfrac{\sum\limits_{i=1}^{5} E_i}{5}$

Slope $(\beta_1) = \dfrac{dy}{dx}$

Intercept (β_0)

Y

X

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_m x_m$$

$$\hat{y} = -0.037 + 0.513(\text{PC1}) - 0.022(\text{PC2}) + 0.006(\text{PC3})$$
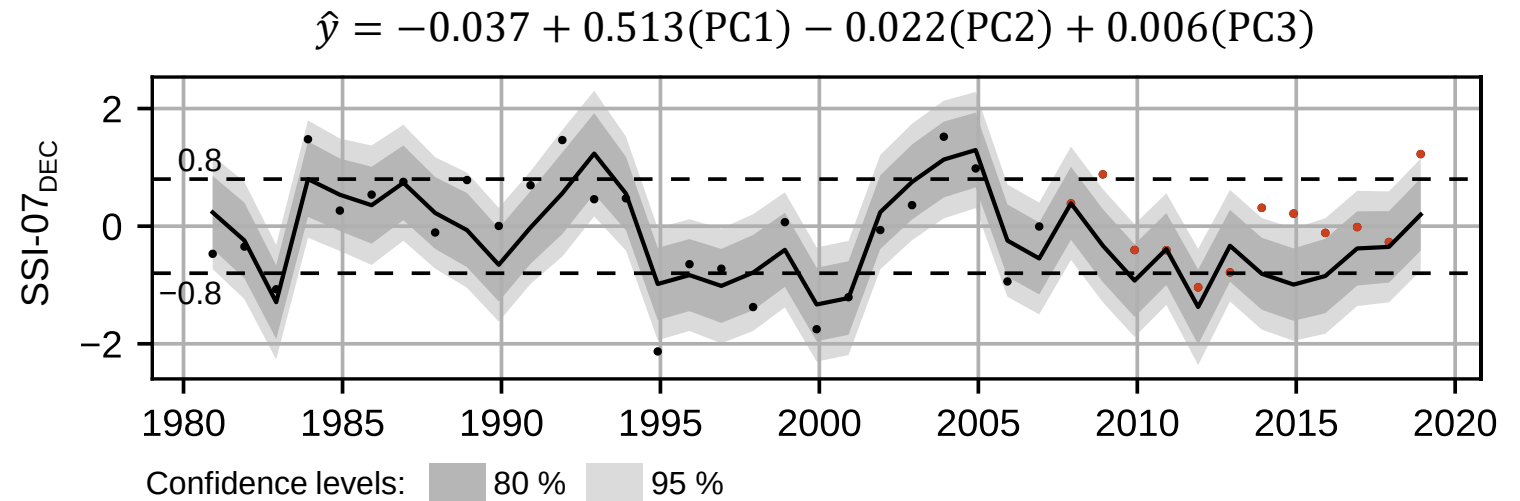
$y = \textbf{SSI-07}_{\textbf{DEC}}$ (impacts on maize)

Predictors subset

RZMC, SLP, TQL, TQV and TSKINWTR

$\text{RMSE}_{\text{training}} = 0.54$
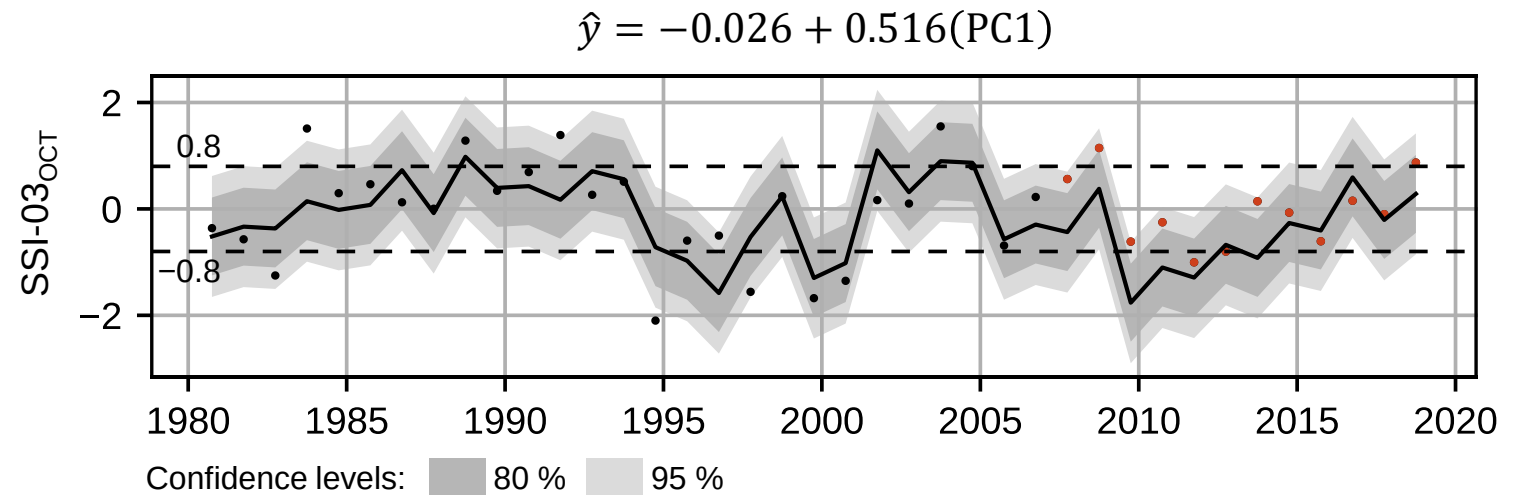$\text{RMSE}_{\text{testing}} = 0.74$

$$\hat{y} = -0.026 + 0.516(\text{PC1})$$

$y = \textbf{SSI-03}_{\textbf{OCT}}$ (impacts on bean)

Predictors subset

TQI, TQL, TSKINWTR

$\text{RMSE}_{\text{training}} = 0.65$
$\text{RMSE}_{\text{testing}} = 0.68$

Relatively high errors.

Synthesis

- Locally weighted scatterplot smoothing (LOWESS) models were applied in conjunction with a multiplicative decomposition model (as suggested by Lu *et al.*, 2017) to remove the trend from the time series of effective crop yield (ECY).

- A method (proposed by Real-Rangel *et al.*, 2020) was applied to determine the most appropriate target variable to detect agricultural impacts.

- A correlation analysis was carried out, including nine candidate predictors at the global and regional scales.

- For the prediction of soil moisture, a linear regression model was applied together with a principal component analysis (Mortensen *et al.*, 2018).

Ongoing work

- The ECY trend model has been changed from the LOWESS model to the Empirical Mode Decomposition (EMD) model to avoid multiple local maxima and minima in some cases (not presented here), contrary to the definition of a trend (Wu *et al.*, 2007).

- ECY itself has been defined as the target variable (instead of SSI).

- The multi-model inference approach (MMI; Burnham & Anderson, 2004) is being compared with the "best model" approach presented here.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, *33*(2), 261–304.

Lu, J., Carbone, G. J., & Gao, P. (2017). Detrending crop yield data for spatial visualization of drought impacts in the United States, 1895–2014. Agricultural and Forest Meteorology, 237–238, 196–208.

Mortensen, E., Wu, S., Notaro, M., Vavrus, S., Montgomery, R., De Piérola, J., Sánchez, C., & Block, P. (2018). Regression-based season-ahead drought prediction for southern Peru conditioned on large-scale climate variables. *Hydrology and Earth System Sciences*, *22*(1), 287–303.

Real-Rangel, R. A., Pedrozo-Acuña, A., Breña-Naranjo, J. A., & Alcocer-Yamanaka, V. H. (2020). A drought monitoring framework for data-scarce regions. *Journal of Hydroinformatics*, *22*(1), 170–185.

Wu, Z., Huang, N. E., Long, S. R., & Peng, C.-K. (2007). On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences*, *104*(38), 14889–14894.

Thanks!

See you at the chat
Session HS2.4 | Display D246 | Abstract ID number <u>EGU2020-21650</u>
Chat time: Thursday, 7 May 2020, 16:15-18:00 (CEST)

Contact
E-mail: <u>rrealr@iingen.unam.mx</u>
Twitter: @rrealrangel
LinkedIn: roberto-real-rangel