Early forecasting of maize yields and prices using vegetation satellite products

TESTE Florian, Atos, Paris-Saclay University

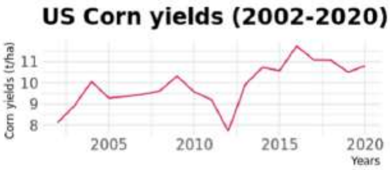
Supervisor: D.Makowski (INRAE), Ph. Ciais (LSCE)

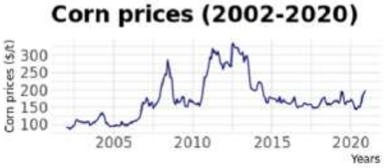
18/11/2022

Outline

Context

Maize prices highly impacted by supplies:[2, 5, 6, 8, 14]





Remote sensing & crop monitoring

How?

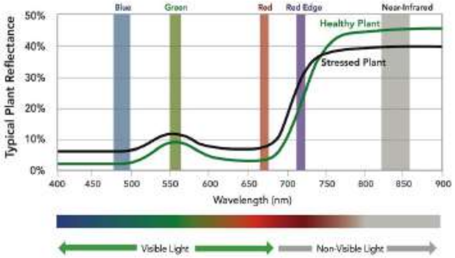
Satellite products and crop monitoring:

- Vegetation indices (NDVI, SAVI...)[11, 13]
- Biophysical parameters (LAI,fAPAR,FVC...)[7, 12, 10]

Why?

Advantages:

- Large area coverage
- low-cost
- repetitive coverage (up to 1 day)
- near real-time information



Remote sensing & crop monitoring

Examples of satellite products

(a) LAI July 2012,severe drougth, annual yield of 7,73 t/ha (FAOSTAT)

(b) LAI July 2016, annual yield of 11,2 t/ha (FAOSTAT)

Figure 1: LAI july values & annual yield

Goals & Motivations

Goals:

- Predict the impact of maize production variation on prices based on **satellite images available during the season**

Motivations

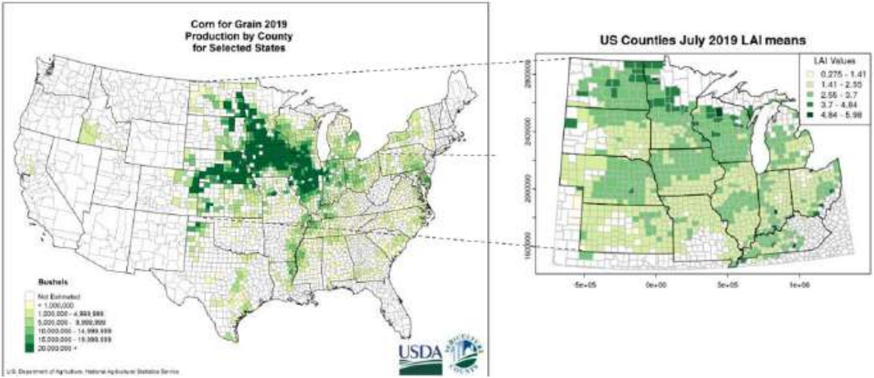
- Near-real time price predictions
- Avoiding the use of regional agricultural production, demand, progress & condition reports
- Mitigate food crisis through production shock forecasting

Data

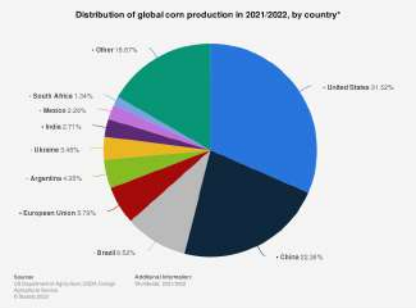
Data & remote sensing products:

- LAI from 1981 to 2018 (GLASS).
  - Spatial resolution: $0.05^\circ \approx 5.5$km (at the equator)
  - Temporal resolution: 8 days
- Maize mask (USDA)
  - Spatial resolution: 30m
- Maize prices (USD/tonne) (1961-2021) (World banks).
- US annual yield (hectograms per hectare) data from 1961 to 2021 (FAOSTAT).

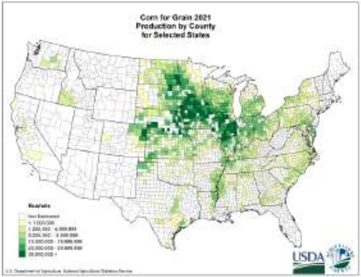
Study area


Corn for Grain 2019
Production by County
for Selected States
Bushels
Not Estimated
USDA
U.S. Department of Agriculture, National Agricultural Statistics Service
US Counties July 2019 LAI means
LAI Values
0.275 - 1.41
1.41 - 2.55
3.7 - 4.84
4.84 - 5.98
-5e+05
0e+00
5e+05
1e+06

Study area



(a) World maize production, source: USDA,statista



(b) US maize production by counties, source: USDA

Method

Response variables

The maize prices relatives changes Δp_{mt} from year to year is expressed as follow:

$$\Delta p_{mt} = \frac{p_{mt} - p_{mt-1}}{p_{mt-1}} \tag{1}$$

where p_{mt} is the maize prices for the $m'th$ month of the year t.

The maize yield relatives change $\Delta yield_t$ from year to year for a specific region is expressed as follows:

$$\Delta yield_t = \frac{yield_t - yield_{t-1}}{yield_{t-1}} \tag{2}$$

where $yield_t$ denotes the maize yield for the year t.

Method

We define a binary variable Δp^{b}_{mt} equal to one in case of price increase ($\Delta p_{mt} > 0$) and to zero otherwise.

We define a binary variable $\Delta yield^{b}_{t}$ equal to one in case of price increase ($\Delta yield_{t} > 0$) and to zero otherwise.

Method

The gridded datasets (remote sensing products, here LAI) at time t and spatial location/pixel s,with $t \in [1, T]$ and $s \in [1, S]$,is represented by the matrix X

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1S} \\ x_{21} & x_{22} & \cdots & x_{2S} \\ \vdots & \vdots & \vdots & \vdots \\ x_{T1} & x_{T2} & \cdots & x_{TS} \end{pmatrix} Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} \tag{3}$$

S= 123 000, T=38

Method

Predictors variables

The LAI monthly mean (LAI_{mt}) over all pixels in month m of the year t, the monthly pixel average can be computed as:

$$\overline{LAI}_{mt} = \frac{1}{S}\left(\sum_{s=1}^{S} LAI_{smt}\right) \tag{4}$$

Where S is the number of pixels in the time series, m represents the month and t the year.

Method

The space–time decomposition of EOF analysis can be written as follow:

$$X = A\Lambda U^T \tag{5}$$

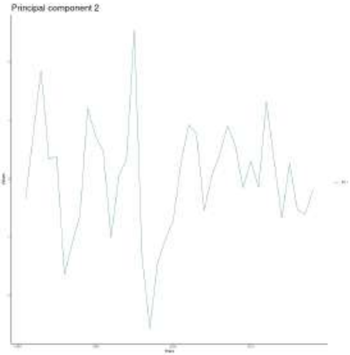
$$X = \sum_{k=1}^{M} \lambda_k a_k u_k^T \tag{6}$$

Formulated as an optimal set of orthonormal spatial functions u_k and time expansion functions, also known as expansion coefficients (EC), a_k, where M is the number of functions (modes), and $M = min(T, S)$.
(R packages wql)

Method

EOF output examples

(c) EOF 2nd mode: computed from the LAI july time series



(d) EOF 2nd Principal components: computed from the LAI july time series

Figure 2: EOF outputs

Method

Binomial Models

Price predictions:

$$\Pr(\Delta p^b_{mt} | a_{1t}, .., a_{kt}..a_{Mt}) = 1 = \frac{e^{\beta_0 + \sum_{k=1}^{M} \beta_k a_{kt}}}{1 + e^{\beta_0 + \sum_{k=1}^{M} \beta_k a_{kt}}}$$

Yield predictions:

$$\Pr(\Delta yield^b_{mt} | a_{1t}, .., a_{kt}..a_{Mt}) = 1 = \frac{e^{\beta_0 + \sum_{k=1}^{M} \beta_k a_{kt}}}{1 + e^{\beta_0 + \sum_{k=1}^{M} \beta_k a_{kt}}}$$

LASSO variables selection

Method

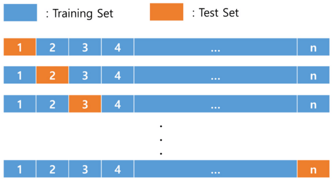


Figure 3: LOOCV method. Source: Cha et al.

LOOCV Algorithm:

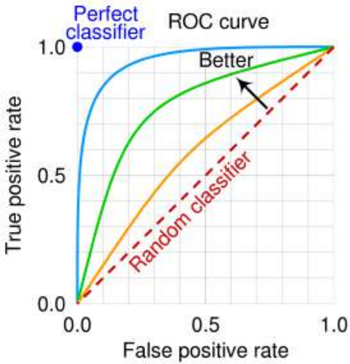
- Split the entire data set of size T into:
  Orange = year selected in the test data set
  Blue = years selected in the training data set
- Fit the model using the training data set
- Run the model using the test dataset
- Repeat previous steps *T* times, and compute AUC.

Method

AUC

The Area Under the ROC Curve **(AUC)** is a measurement of the total area beneath the entire ROC curve from (0,0) to (1,1).

- AUC ranges from **0 to 1**.
- A model with **100% correct predictions** has an **AUC of 1.0**.
- An **AUC of 0.5** is similar to a **random guess**.



Results

Maize prices variation:LAI means

The AUC of Corn prices variation prediction

LAI means

	LAI month								
Monthly prices	LAI February	LAI march	LAI April	LAI May	LAI June	LAI July	LAI August	LAI September	LAI October
March	0.68	—	—	—	—	—	—	—	—
April	0.74	0.71	—	—	—	—	—	—	—
May	0.73	0.64	0.49	—	—	—	—	—	—
June	0.70	0.60	0.50	0.57	—	—	—	—	—
July	0.62	0.54	0.58	0.57	0.57	—	—	—	—
August	0.43	0.57	0.49	0.54	0.58	0.75	—	—	—
September	0.47	0.54	0.51	0.51	0.48	0.64	0.68	—	—
October	0.45	0.54	0.50	0.49	0.47	0.64	0.72	**0.82**	—
November	0.49	0.53	0.53	0.46	0.55	0.61	0.69	0.74	0.60
December	0.45	0.52	0.54	0.48	0.48	0.64	0.73	**0.78**	0.61

Results

Products means

Key points:

The averaging method gives:

- Poor overall prediction results
- End-of-year price forecasting is possible using end-of-season products.

Results

Maize prices variation:GLM LASSO

The AUC of corn price variation prediction

LAI based models

	Input predictors								
Monthly prices	EOF February	EOF March	EOF April	EOF May	EOF June	EOF July	EOF August	EOF September	EOF October
March	0.55	—	—	—	—	—	—	—	—
April	0.45	0.66	—	—	—	—	—	—	—
May	0.46	0.54	0.82	—	—	—	—	—	—
June	0.45	0.62	0.89	0.86	—	—	—	—	—
July	0.54	0.85	0.87	0.66	0.83	—	—	—	—
August	0.39	0.99	0.89	0.71	0.90	0.52	—	—	—
September	0.93	0.96	0.89	0.84	0.89	0.59	0.62	—	—
October	0.72	0.82	0.94	0.81	0.82	0.57	0.49	0.64	—
November	0.68	0.92	0.94	0.73	0.82	0.64	0.64	0.69	0.84
December	0.61	0.93	0.99	0.73	0.66	0.60	0.55	0.59	0.84

Results

Yield variation: Satellite products means

The AUC of yield variation prediction
Satellite products means

	Products
Month	**LAI averaged**
February	0.53
March	0.55
April	0.51
May	0.58
June	0.50
July	0.45
August	0.56
September	0.66
October	0.57

Results

Yield variation: GLM LASSO

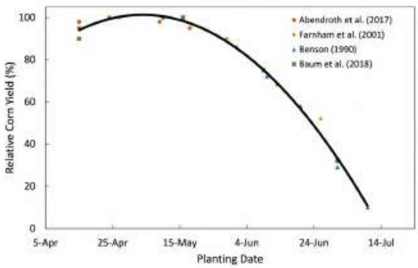
The AUC of yield variation prediction
GLM LASSO based models

	Input predictors
Month	**EOF LAI**
February	**0.93**
March	**0.76**
April	0.72
May	**0.91**
June	**0.84**
July	0.71
August	0.65
September	**0.73**
October	**0.92**

Discussion

How can such early prediction results be explained?

- Pre-season LAI probably highlight winter/early spring sowing conditions:
  - Pre-growing season weather condition impact corn yield[9]
  - A late sowing date is associated with a decrease in yield[3, 4, 1]



Conclusions

- The use of vegetation satellite products allows for early forecasting of annual increase vs. decrease of maize yields and prices
- EOF better than map average
- EOF probably capture maize sowing conditions in early spring

Perspectives

- Experiment with different dimension reduction methods (Neural networks: Variational Auto-encoder):
  - EOF is primarily a linear transformation, but auto-encoders can take complex nonlinear functions into account.
- Applications to other crops and regions:
  - Wheat Durum $\longrightarrow$ Canada is the primary producer, and production is concentrated in a relatively small and constrained area.
  - Soybean $\longrightarrow$ the USA is one the largest producer($\approx$ 30%, and it is grown in the same production area as Maize
  - Rice in South $\longrightarrow$ Eastern Asia

Thank You

References I

[1] Lori J. Abendroth et al. "Yield-Based Corn Planting Date Recommendation Windows for Iowa". In: *Crop, Forage & Turfgrass Management* 3.1 (2017). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2134/cftm2017.02.0015, cftm2017.02.0015. ISSN: 2374-3832. DOI: 10.2134/cftm2017.02.0015. URL: https://onlinelibrary.wiley.com/doi/abs/10.2134/cftm2017.02.0015 (visited on 11/16/2022).

[2] Ryan Bain and T. Randall Fortenbery. "IMPACT OF CROP CONDITION REPORTS ON NATIONAL AND LOCAL WHEAT MARKETS". In: *J. Agric. Appl. Econ.* 49.1 (Feb. 2017), pp. 97–119. ISSN: 1074-0708, 2056-7405. DOI: 10.1017/aae.2016.31. URL: https://www.cambridge.org/core/product/identifier/S1074070816000316/type/journal_article (visited on 09/04/2022).

[3] M. E. Baum, S. V. Archontoulis, and M. A. Licht. "Planting Date, Hybrid Maturity, and Weather Effects on Maize Yield and Crop Stage". In: *Agronomy Journal* 111.1 (2019). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2134/agronj2018.04.0297, pp. 303–313. ISSN: 1435-0645. DOI: 10.2134/agronj2018.04.0297. URL: https://onlinelibrary.wiley.com/doi/abs/10.2134/agronj2018.04.0297 (visited on 11/16/2022).

References II

[4] G. O. Benson. "Corn Replant Decisions: A Review". In: *Journal of Production Agriculture* 3.2 (1990). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2134/jpa1990.0180, pp. 180–184. ISSN: 2689-4114. DOI: 10.2134/jpa1990.0180. URL: https://onlinelibrary.wiley.com/doi/abs/10.2134/jpa1990.0180 (visited on 11/16/2022).

[5] An Cao and Michel A. Robe. *Market Uncertainty and Sentiment Around USDA Announcements*. Rochester, NY, Nov. 1, 2021. URL: https://papers.ssrn.com/abstract=3993272 (visited on 09/04/2022).

[6] An N. Q. Cao et al., eds. *County-level USDA Crop Progress and Condition data, machine learning, and commodity market surprises*. 2022. DOI: 10.22004/ag.econ.322281.

[7] Deborah V. Gaso, Andrés G. Berger, and Verónica S. Ciganda. "Predicting wheat grain yield and spatial variability at field scale using a simple regression or a crop model in conjunction with Landsat images". In: *Computers and Electronics in Agriculture* 159 (Apr. 1, 2019), pp. 75–83. ISSN: 0168-1699. DOI: 10.1016/j.compag.2019.02.026. URL: https://www.sciencedirect.com/science/article/pii/S0168169918302072 (visited on 09/04/2022).

References III

[8] Georg V. Lehecka. “The Value of USDA Crop Progress and Condition Information: Reactions of Corn and Soybean Futures Markets”. In: *undefined* (2014). URL: `https://www.semanticscholar.org/paper/IMPACT-OF-CROP-CONDITION-REPORTS-ON-NATIONAL-AND-Bain-Fortenbery/dbf39570ca8089065e451504331676e06c7dbaf4` (visited on 09/04/2022).

[9] Ziyi Li et al. “Assessing the impacts of pre-growing-season weather conditions on soil nitrogen dynamics and corn productivity in the U.S. Midwest”. In: *Field Crops Research* 284 (Aug. 1, 2022), p. 108563. ISSN: 0378-4290. DOI: `10.1016/j.fcr.2022.108563`. URL: `https://www.sciencedirect.com/science/article/pii/S0378429022001344` (visited on 11/08/2022).

References IV

[10] Raúl López-Lozano et al. “Towards regional grain yield forecasting with 1km-resolution EO biophysical products: Strengths and limitations at pan-European level”. In: *Agricultural and Forest Meteorology* 206 (June 15, 2015), pp. 12–32. ISSN: 0168-1923. DOI: 10.1016/j.agrformet.2015.02.021. URL: https://www.sciencedirect.com/science/article/pii/S0168192315000702 (visited on 09/05/2022).

[11] Franciele Morlin Carneiro et al. “Comparison between vegetation indices for detecting spatial and temporal variabilities in soybean crop using canopy sensors”. In: *Precision Agric* 21.5 (Oct. 1, 2020), pp. 979–1007. ISSN: 1573-1618. DOI: 10.1007/s11119-019-09704-3. URL: https://doi.org/10.1007/s11119-019-09704-3 (visited on 09/05/2022).

[12] Yi Peng et al. “Remote prediction of yield based on LAI estimation in oilseed rape under different planting methods and nitrogen fertilizer applications”. In: *Agricultural and Forest Meteorology* 271 (June 15, 2019), pp. 116–125. ISSN: 0168-1923. DOI: 10.1016/j.agrformet.2019.02.032. URL: https://www.sciencedirect.com/science/article/pii/S0168192319300905 (visited on 09/04/2022).

References V

[13] Chenghai Yang and Gerald L. Anderson. "Mapping Grain Sorghum Yield Variability Using Airborne Digital Videography". In: *Precision Agriculture* 2.1 (Sept. 1, 2000), pp. 7–23. ISSN: 1573-1618. DOI: 10.1023/A:1009928431735. URL: https://doi.org/10.1023/A:1009928431735 (visited on 09/05/2022).

[14] Rotem Zelingher, David Makowski, and Thierry Brunelle. "Assessing the Sensitivity of Global Maize Price to Regional Productions Using Statistical and Machine Learning Methods". In: *Front. Sustain. Food Syst.* 5 (June 2, 2021), p. 655206. ISSN: 2571-581X. DOI: 10.3389/fsufs.2021.655206. URL: https://www.frontiersin.org/articles/10.3389/fsufs.2021.655206/full (visited on 09/04/2022).

Method

Performances metrics

ROC curve

An ROC curve (receiver operating characteristic curve) is a graph that depicts the performance of a classification model across all classification thresholds.
This curve depicts two parameters:

- True Positive Rate
- False Positive Rate

True Positive Rate (TPR) is defined as follows:

$$TPR = \frac{TP}{TP + FN}$$

False Positive Rate (FPR) is defined as follows:

$$FPR = \frac{FP}{FP + TN}$$