Random forest-based crop yield forecasting in the Pannonian Basin and its skill in years of severe drought

E. Bueechi, M. Fischer, L. Crocetti, M. Trnka, A. Grlj, W. Dorigo

Consortium:

Background

In the last decades, droughts have heavily affected agricultural production in the Pannonian Basin caused by a generally increasing frequency of heatwaves and dry conditions [1]. As most fields in this region are only rain-fed, agriculture there is particularly vulnerable to droughts [2]. The already challenging conditions for crop production are expected to worsen due to climate change [3]. The Pannonian Basin is even considered as the region with the most negative impacts of climate change on crop production in Europe [4]. A potential tool to support the adaption to these challenging circumstances is crop yield forecasting. This has proven being a vital tool to minimise socio-economic impacts of crop losses [5].

Data

Crop yield data

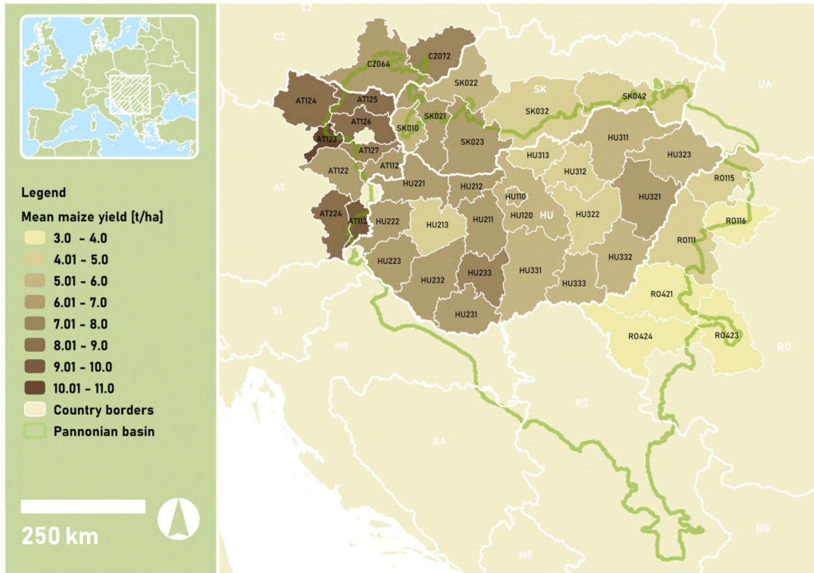

Fig. 1: Overview of the study area

Predictors

Dataset	Source	Spatial Resolution	Temporal Resolution
Earth Observation			
Soil Moisture	ESA CCI	0.25°	daily
Soil Water Index	ESA CCI	0.25°	daily
VOD Ku-Band	VODCA	0.25°	daily
NDVI	CGLS	0.01°	10-daily
LAI	CGLS	0.01°	10-daily
ESI (1 and 3 months)	MODIS	0.05°	weekly
Reanalysis			
Temperature	ERA5-Land	0.1°	daily
Growing Degree Days	ERA5-Land	0.1°	monthly
SPEI (1 and 3 months)	ERA5	0.25°	monthly
Seasonal forecasts			
Precipitation	ECMWF	1°	monthly
Temperature	ECMWF	1°	monthly
In situ data			
Temperature	E-OBS	0.25°	daily
Precipitation	E-OBS	0.25°	daily
Fraction of wet days	E-OBS	0.25°	monthly

Tab. 1: List of used predictors

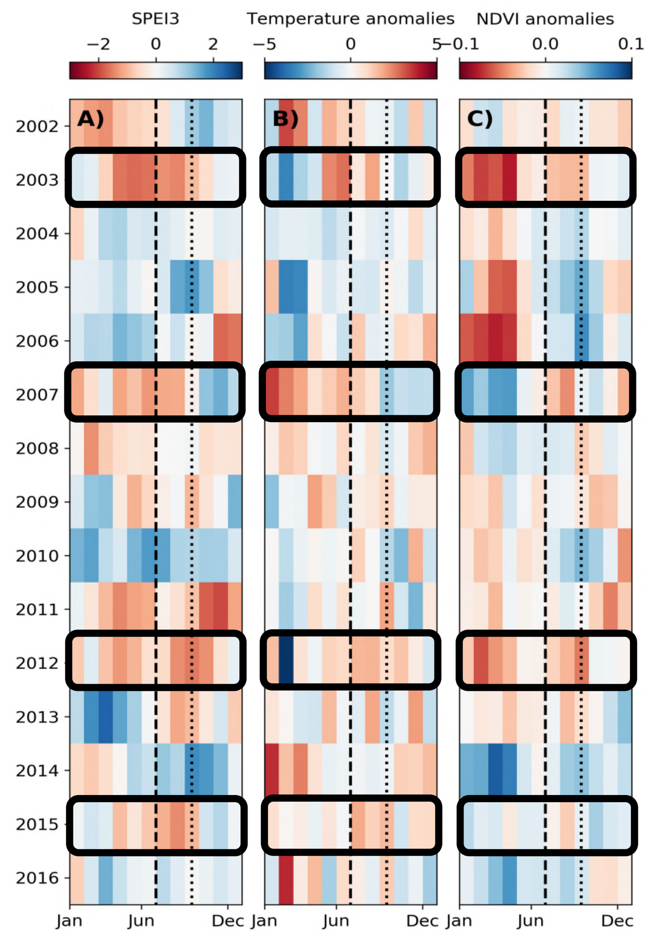

Fig. 2: Anomalies of three predictors – severe drought years highlighted

Methods

Yearly yield anomalies of maize and winter wheat are forecasted for various districts in the Pannonian Basin (Fig. 1) from 2002-2016. Monthly forecasts are made for each growing season, starting around three months before harvest.

The forecasts are cross-validated by using each consecutive 3-years time period once as testing set. After an initial model run using all predictors, predictors are reduced using recursive feature elimination and manually removing predictors with large cross-correlation to other predictors.

Feature importance of the predictors are calculated to get an impression how the impact of the predictors change over the months.

Results

Crop yield forecast

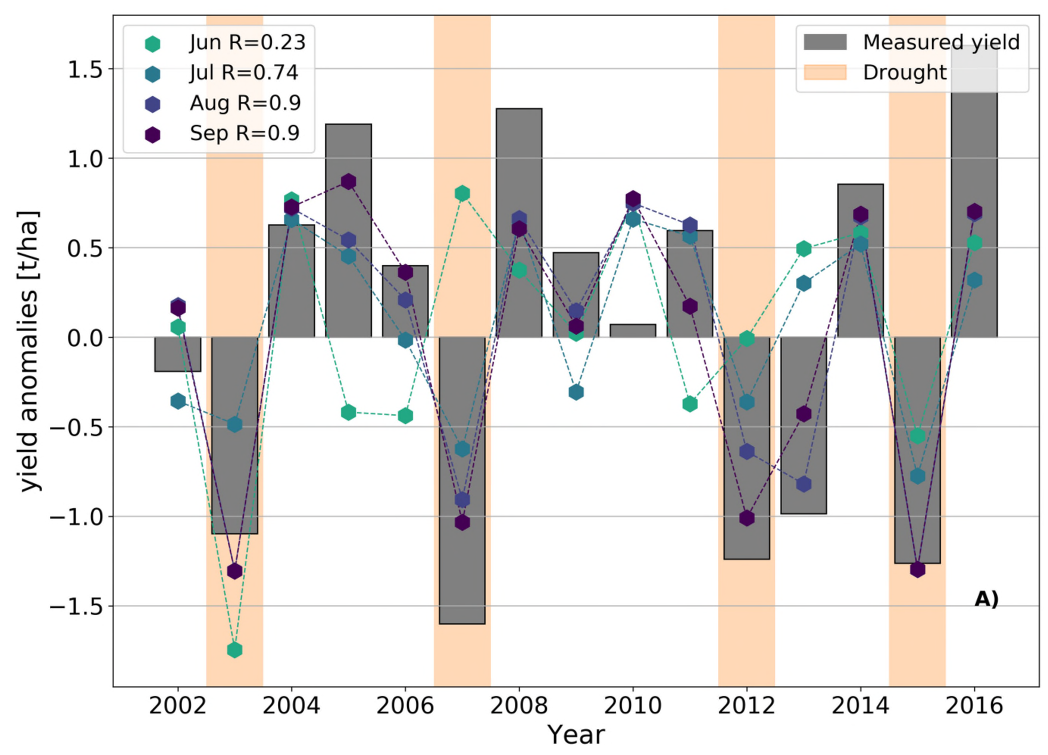

Fig. 3: Forecasted and observed mean maize yield anomalies over the Pannonian Basin

The model underestimates extremes of high and low crop yields (Fig. 3)
Forecasts in drought years early detect crop yield losses

Validation

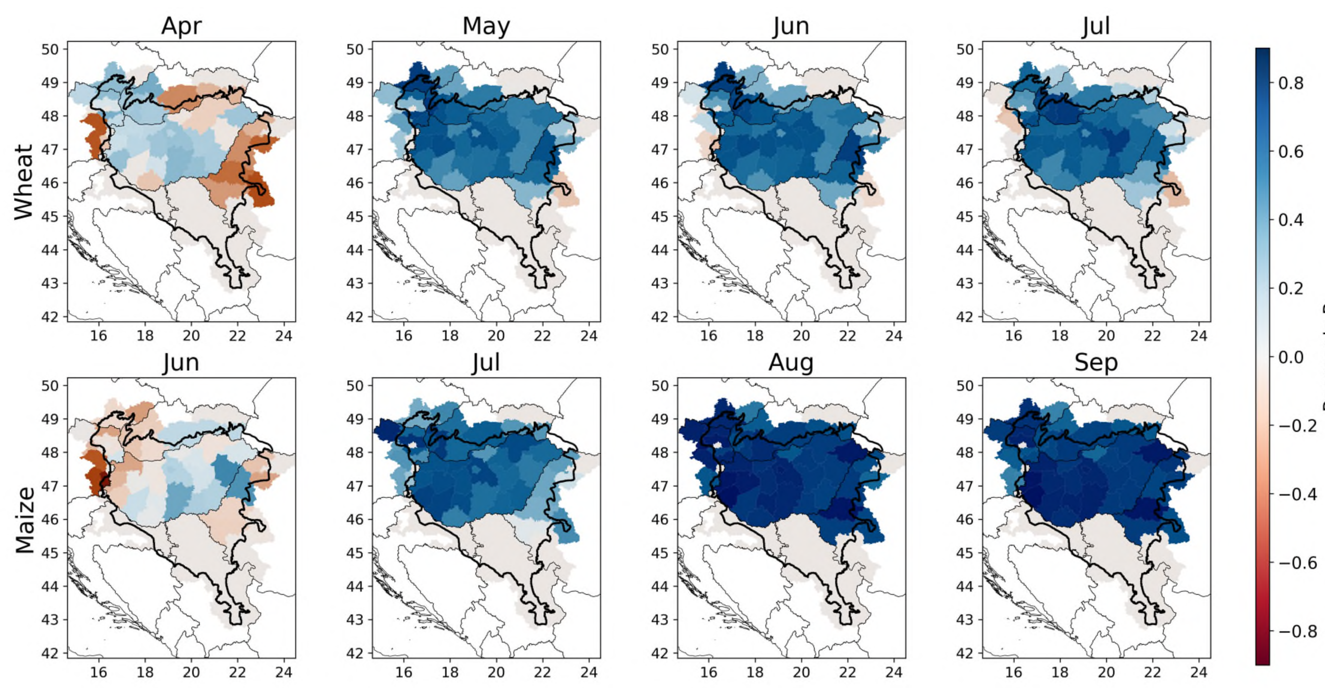

Fig. 4: Correlations of forecasted and observed crop yield anomalies

Crop yield of maize and wheat are highly dependent on the conditions in the last two months before harvest. This leads to highest performances of crop yield forecasts in these months.

Feature importance

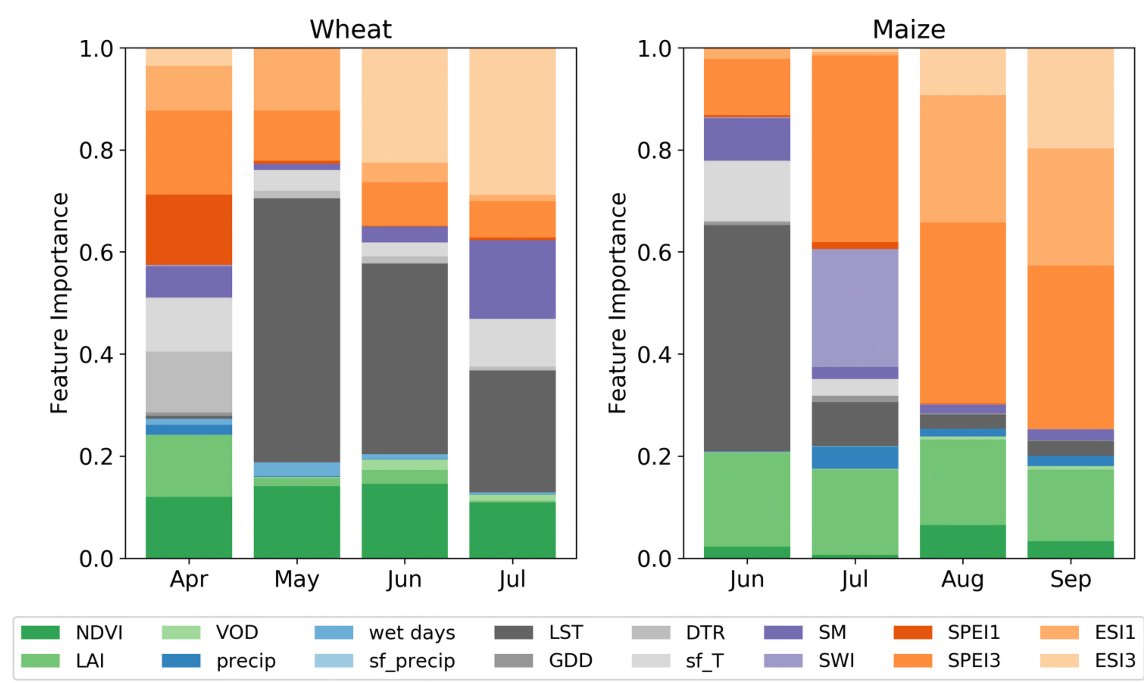

Fig. 5: Cumulative feature importances of the predictors

Key driver of wheat forecast model is temperature - moisture availability (SPEI/ESI) for maize. Impacts of predictors are largely dependent on the forecast month

Predictor analysis

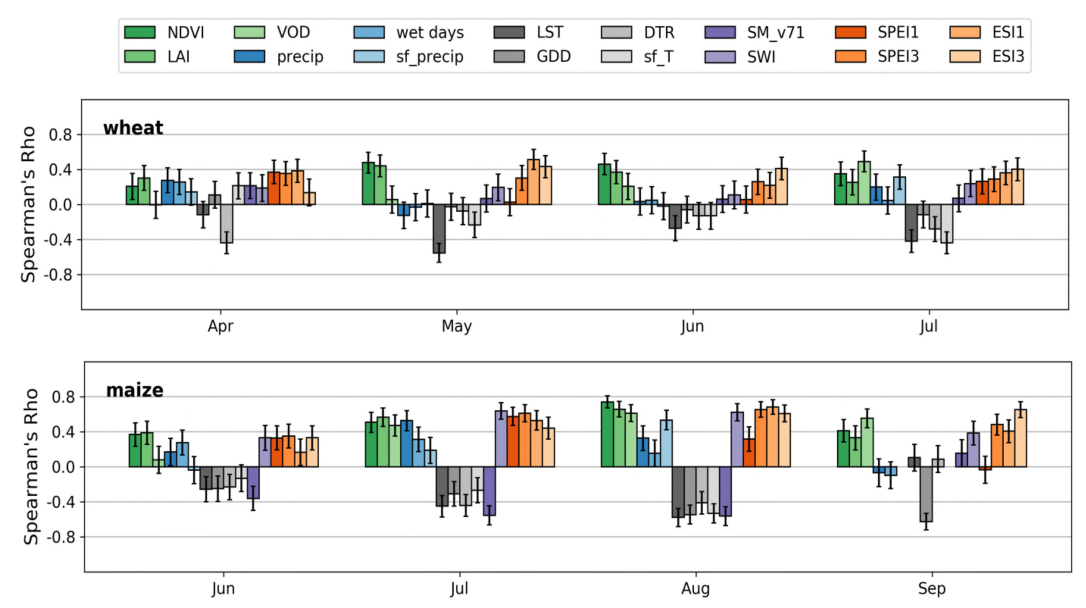

Fig. 6: Correlations of explanatory variables and yield anomalies

Correlations of explanatory variables and crop yields are increasing towards harvest: all, except precipitation, relatively high in July and August for maize. After harvest correlations are decreasing.

Conclusion

Key findings

- Wheat and maize crops can be forecasted around **two months before harvest** with a good performance (Fig. 3 & 4)
- Crop yield losses in years of **severe drought are underestimated by the forecasts** but the model correctly detects negative anomalies (Fig. 3)
- **Good performance** to predict **interannual variabilities** of the yields for the districts (Fig. 4)
- **Bad performance** to distinguish crop yields **between regions** within individual years
- **Wheat yields** largely dependent on **temperature**; **maize yields** on **water availability** (Fig. 5)

Next steps

The results affirm earlier findings of the large dependency on water availability for maize and temperature for wheat in the last weeks before harvest. Future work should focus on how these conditions can be better represented for the modelling, by either considering other input datasets or increasing the temporal and spatial resolutions.
Yield forecasts in severe drought years will require further improvement in the Pannonian Basin. A more thorough analysis of the seasonal forecast and potentially other machine learning techniques can help to do so.

Acknowledgement
The results presented here have been developed in the framework of the project „DryPan: Novel EO data for improved agricultural drought impact forecasting in the Pannonian basin“ funded by the European Space Agency (ESA).

References

[1] Crocetti, L. et al. (2020). Earth Observation for agricultural drought monitoring in the Pannonian Basin (southeastern Europe): current state and future directions. *Regional Environmental Change*, *20*(123). https://doi.org/10.1007/s10113-020-01710-w; **[2]** Antofie, T. et al. (2015). Estimating the water needed to end the drought or reduce the drought severity in the Carpathian region. *Hydrology and Earth System Sciences*, *19*(1), 177–193. https://doi.org/10.5194/hess-19-177-2015; **[3]** Kis, A. et al. (2020). Multi-scenario and multi-model ensemble of regional climate change projections for the plain areas of the pannonian basin. Idojaras, 124(2), 157–190. https://doi.org/10.28974/idojaras.2020.2.2; **[4]** Olesen, J. E. et al. (2011). Impacts and adaptation of European crop production systems to climate change. In European Journal of Agronomy (Vol. 34, Issue 2, pp. 96–112). Elsevier. https://doi.org/10.1016/j.eja.2010.11.003; **[5]** Ceglar, A. et al. (2018). Land-surface initialisation improves seasonal climate prediction skill for maize yield forecast. Scientific Reports, 8(1), 1322

Contact
MSc Emanuel Büechi
CLIMERS - GEO - TU WIEN
emanuel.bueechi@geo.tuwien.ac.at