living planet symposium | BONN 23–27 May 2022

TAKING THE PULSE OF OUR PLANET FROM SPACE

Improving predictions of crop yield loss in years of severe droughts by integrating Earth observation and climate data in a machine learning framework. A case study for the Pannonian basin

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

23 May 2022

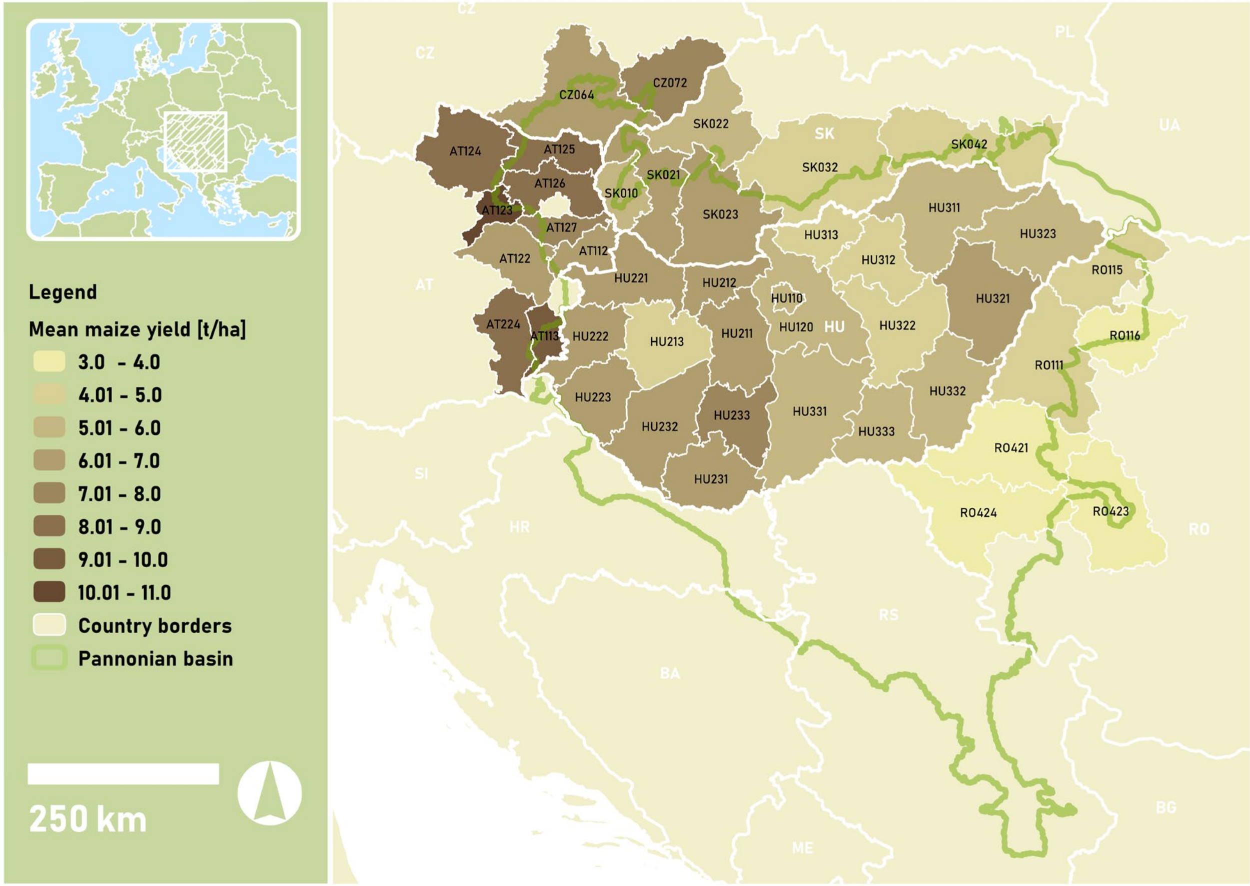
Motivation

Pannonian basin

- Sheltered with relatively low levels of precipitation (< 600 mm/year)
- High dependency of population on agriculture: 10-20% of population; >70% of area

Several drought episodes in the last decades caused significant crop yield losses

Further exacerbated by climate change



Funded by: esa

Consortium: eodc, TU Wien GEO, CzechGlobe, SPACE-SI

Introduction

TECHNISCHE UNIVERSITÄT WIEN

esa

Many crop yield models have been applied

- process based and machine learning

Application over large areas possible using EO data, reanalysis and interpolated datasets

Extreme weather events complicate accurate crop yield forecasts

The ARYA crop yield forecasting algorithm: Application to the main wheat exporting countries

B. Franch[a,b,*], E. Vermote[c], S. Skakun[b,c], A. Santamaria-Artigas[b,c], N. Kalecinski[b,c], J.-C. Roger[b,c], I. Becker-Reshef[b], B. Barker[b], C. Justice[b], J.A. Sobrino[a]

Towards regional grain yield forecasting with 1 km-resolution EO biophysical products: Strengths and limitations at pan-European level

Raúl López-Lozano[a,*], Gregory Duveiller[a,b], Lorenzo Seguini[a], Michele Meroni[a], Sara García-Condado[a], Josh Hooker[a], Olivier Leo[a], Bettina Baruth[a]

Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices

Anikó Kern[a], Zoltán Barcza[b,c,d,*], Hrvoje Marjanović[e], Tamás Árendás[f], Nándor Fodor[f], Péter Bónis[f], Péter Bognár[a], János Lichtenberger[a,g]

In-season performance of European Union wheat forecasts during extreme impacts

glar, S. Garcia Condado, S. Karetsos, R. Lecerf, & M. van den Berg

Seasonal weather forecasts for crop yield modelling in Europe

Pierre Cantelaube & Jean

Statistical modelling of drought-related yield losses using soil moisture-vegetation remote sensing and multiscalar indices in the south-eastern Europe

Vera Potopová[a,*], Miroslav Trnka[b,c], Pavel Hamouz[a], Josef Soukup[a], Tudor Castraveț[d]

Yield estimation and forecasting for winter wheat in Hungary using time series of MODIS data

Péter Bognár[a], Anikó Kern[a], Szilárd Pásztor[a], János Lichtenberger[b], Dávid Koronczay[a] and Csaba Ferencz[a]

Objectives

Apply a random forest based crop yield forecasting system and thoroughly assess its forecasting skill in normal years and severe drought years

Assessment of the contribution of the different explanatory variables to the model skill at different times during the growing season

Data

TECHNISCHE UNIVERSITÄT WIEN

41 NUTS3 regions; 2002-2016

- Wheat and maize (harvested approx. in July and September)

Different datasets representing:

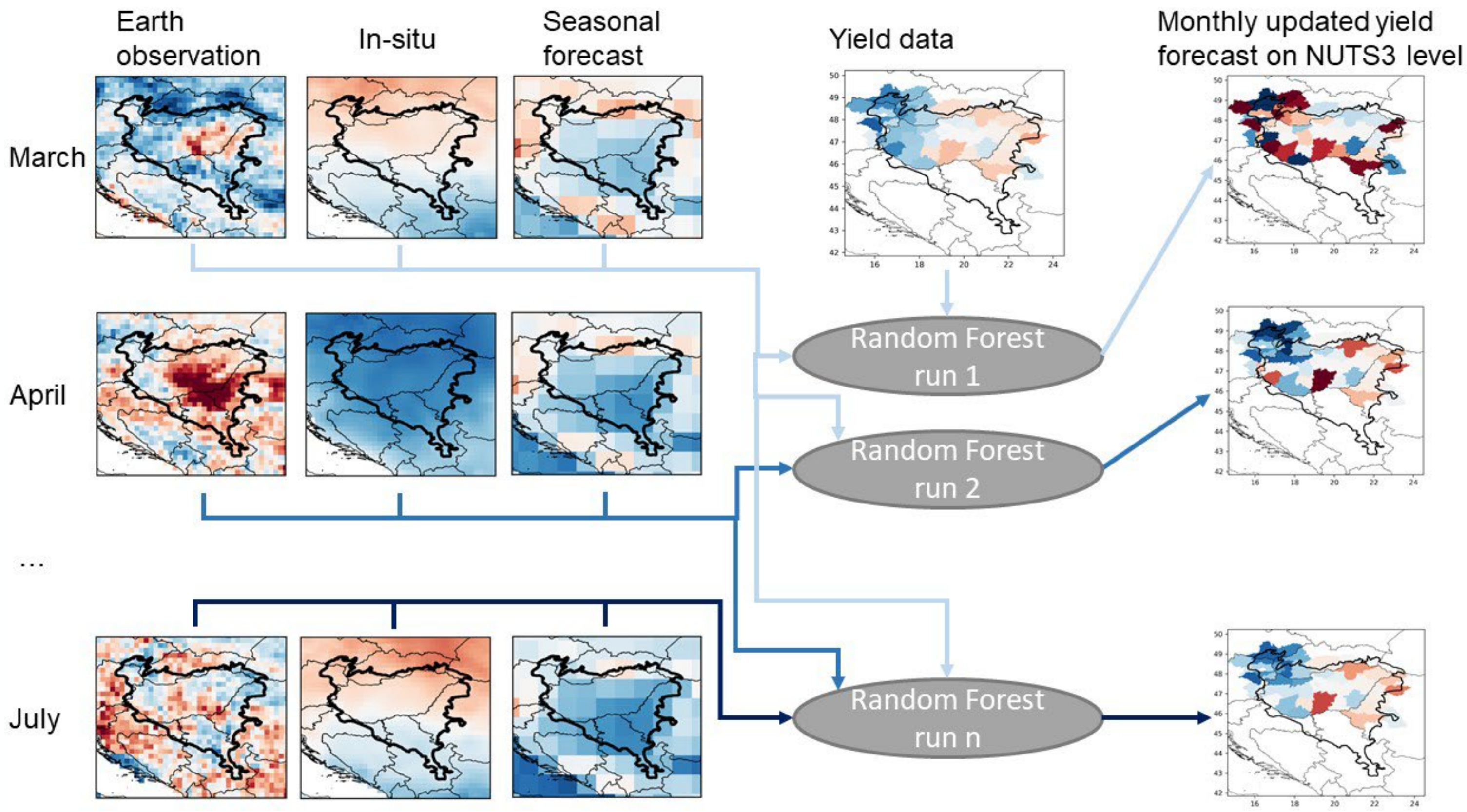
- Temperature and water availability as key drivers of wheat and maize growth
- State of the plants: VOD, NDVI, LAI
- Drought indices SPEI and ESI for specific drought information
- Seasonal forecast show conditions for remaining growing season

Dataset	Source	Spatial Resolution	Temporal Resolution
Yield data	DriDanube	NUTS3 level	yearly
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
Reanalysis			
Diurnal Temperatures	ERA5-Land	0.1°	daily
Growing Degree Days	ERA5-Land	0.1°	monthly
SPEI (1 month)	Based on ERA5	0.25°	monthly
SPEI (3 months)	Based on ERA5	0.25°	monthly
ESI (1 month)	Based on MODIS	0.05°	weekly
ESI (3 months)	Based on MODIS	0.05°	weekly
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

Methods

Crop yield **forecasts** with lead times **up to 4 months** before harvest

- **Random forest** to combine input datasets
  - **Feature importance** to assess impact of predictors
- **Monthly** updated with latest data



Methods

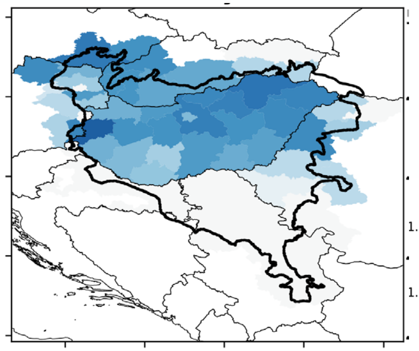
Cross-validation leaving 3 years out

Full timeseries 2002-2016 with test data for 41 regions

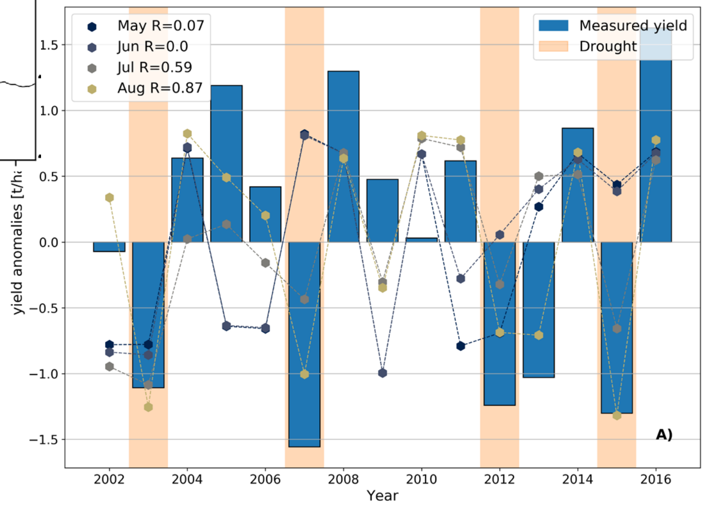
Validation for different skills of model:

- Regional performance (A)
  - Validation per NUTS3 region
- Pannonian basin mean yield forecast (B)
  - Mean forecasted NUTS3 regions vs. mean observed yield
- Yearly performance
  - Validation forecast of all NUTS3 region within individual year
- Overall performance (C)
  - All forecasted NUTS3 yields of all years vs. Observed yields
- Drought year performance (C)
  - As overall but only for drought years

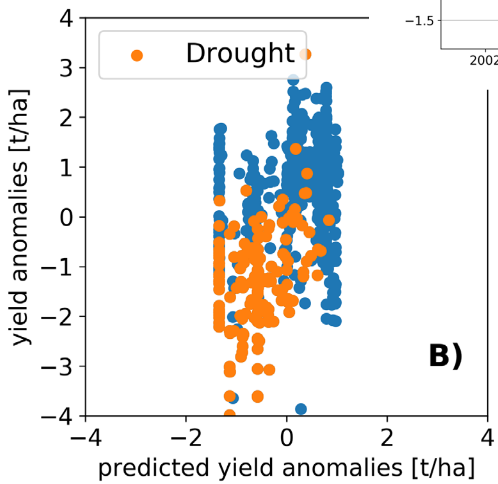
A



B



C



Discussion

TECHNISCHE UNIVERSITÄT WIEN

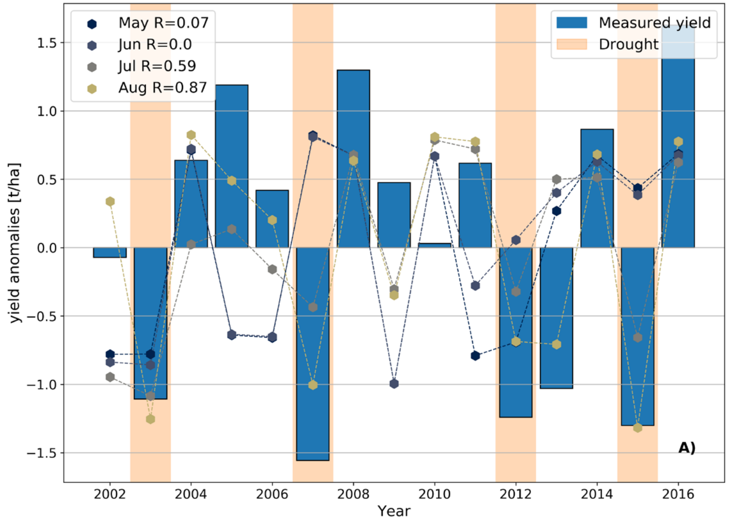
esa

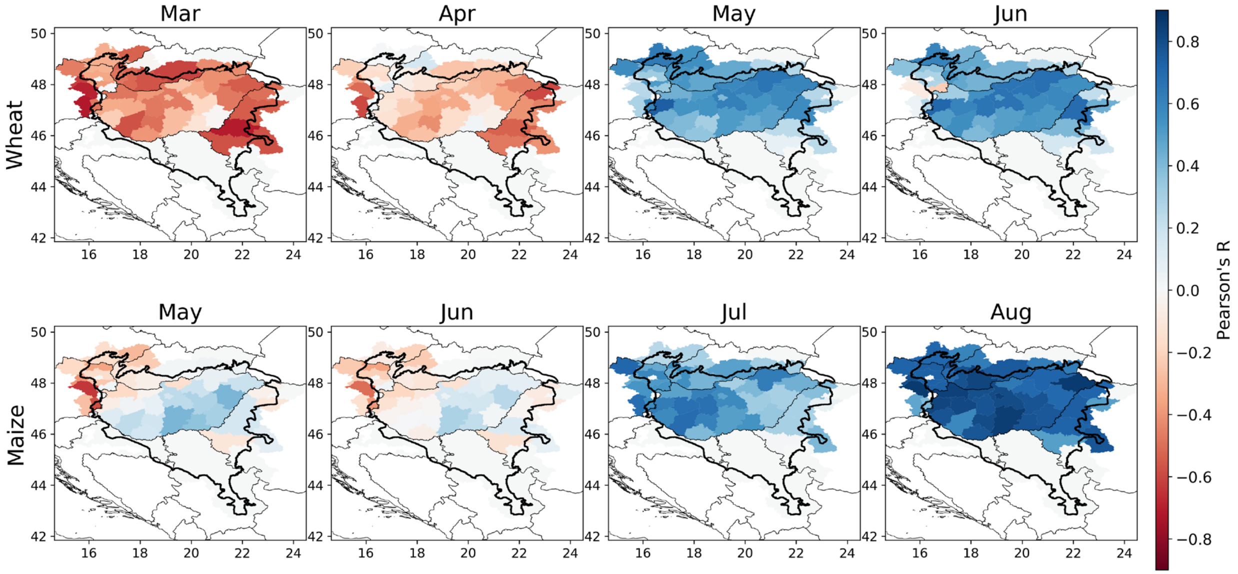
What can be forecasted by the model?

All regions have high correlations to observed yields 2 months before harvest (Regional performance and Pannonian basin mean)

Highest correlations to predict Pannonian basin mean maize yield the month before harvest (R=0.87)

Model early detects negative anomalies in drought years





Discussion

What cannot be forecasted by the model?

Model does not do well to distinguish the crop yields between different NUTS3 regions

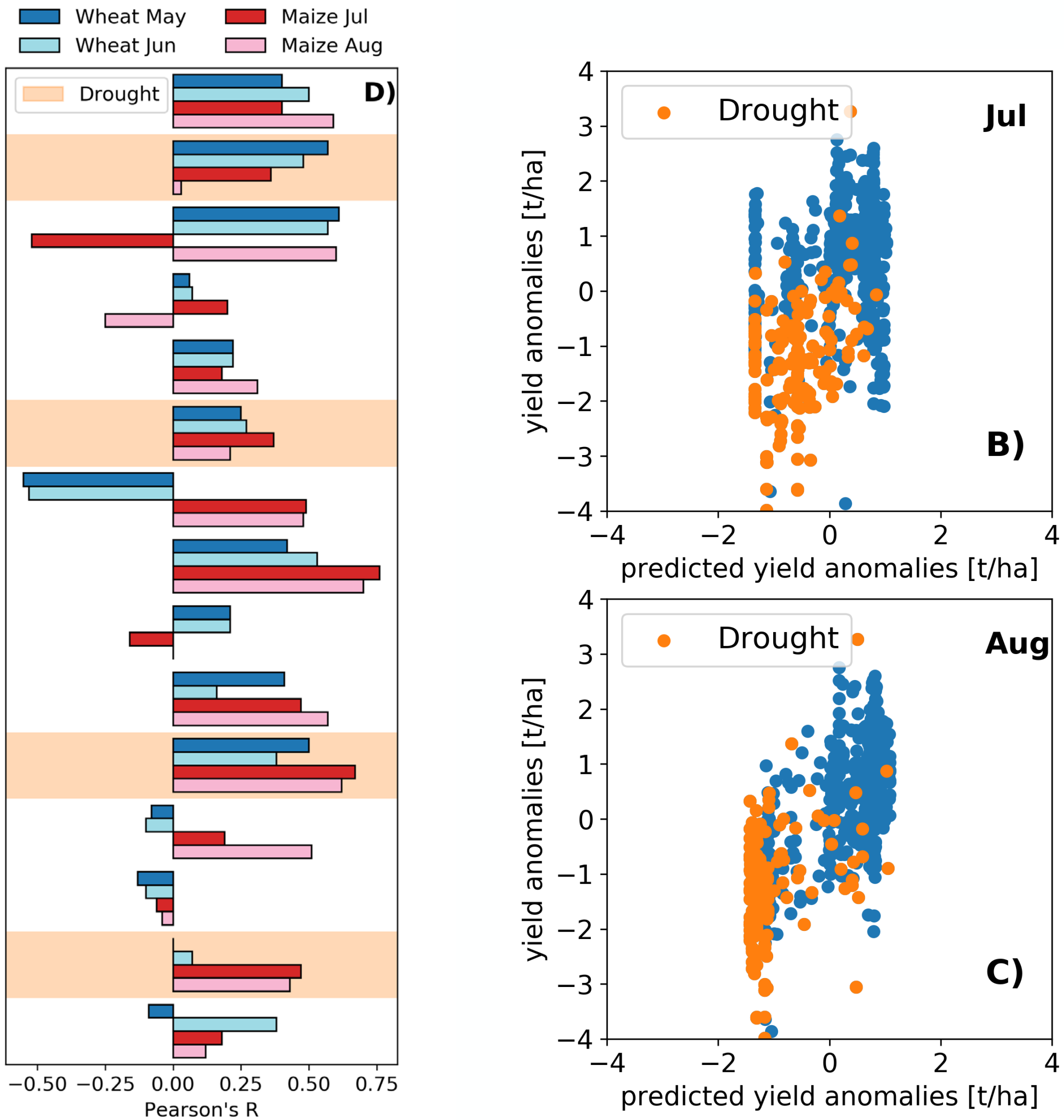
->Low yearly performance

High spatial autocorrelations of yields and predictors

Coarse resolution of predictors

-> Impacts overall performance and drought year performance

	Overall (R)	Drought (R)
Maize Jul	0.51	0.36
Maize Aug	0.67	0.33
Wheat May	0.41	0.47
Wheat Jun	0.47	0.44

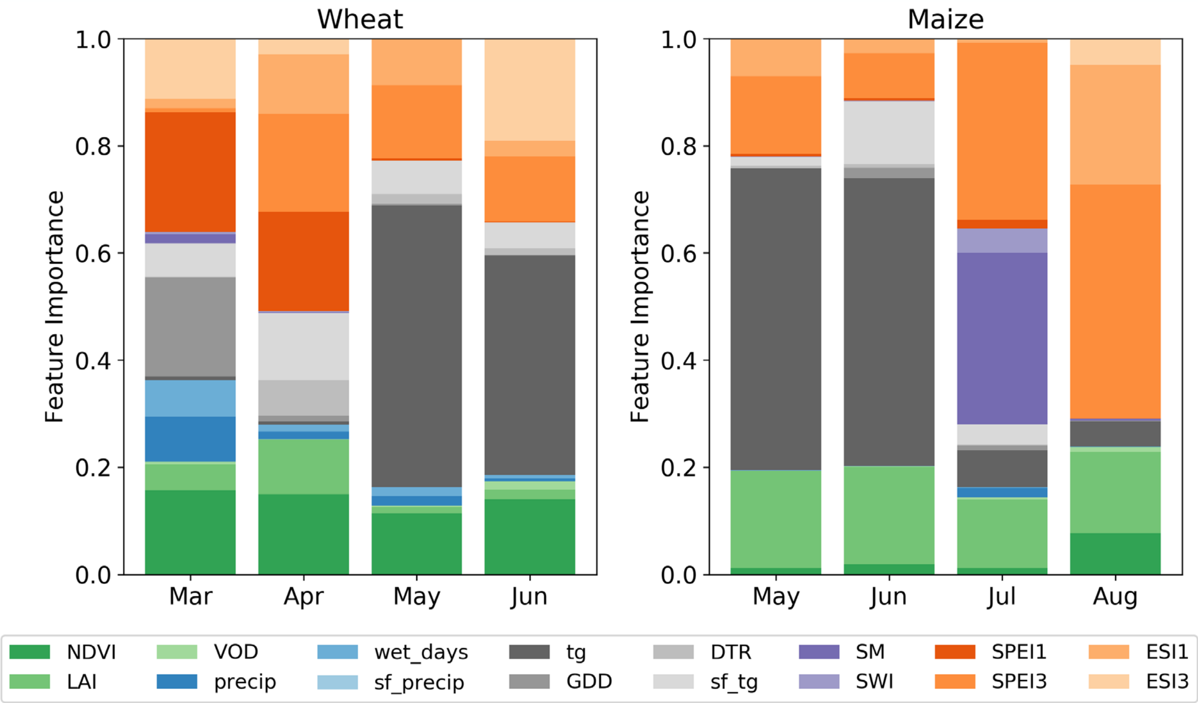


Discussion

Large changes of the feature importance from first to last two months

Wheat model mainly dependent on temperature

Maize on water availability (SPEI3 and soil moisture)



Conclusions

- Crop yield forecasts show reasonable performance from around two months before harvest for interannual variabilities
- High spatial autocorrelations within individual years leads to poorer performance to compare regions
- Crop yield losses in years of severe droughts underestimated - negative anomalies are correctly early detected
- Wheat is mainly dependent on the temperature - maize on water availability

Next steps:

- Improving spatial and temporal resolutions to improve regional model performance; toward field-scale prediction
- Using novel EO datasets to better capture key variables like temperature and water availability
    - Sentinel-1 soil moisture, LSTM temperature

Thank you for your attention!

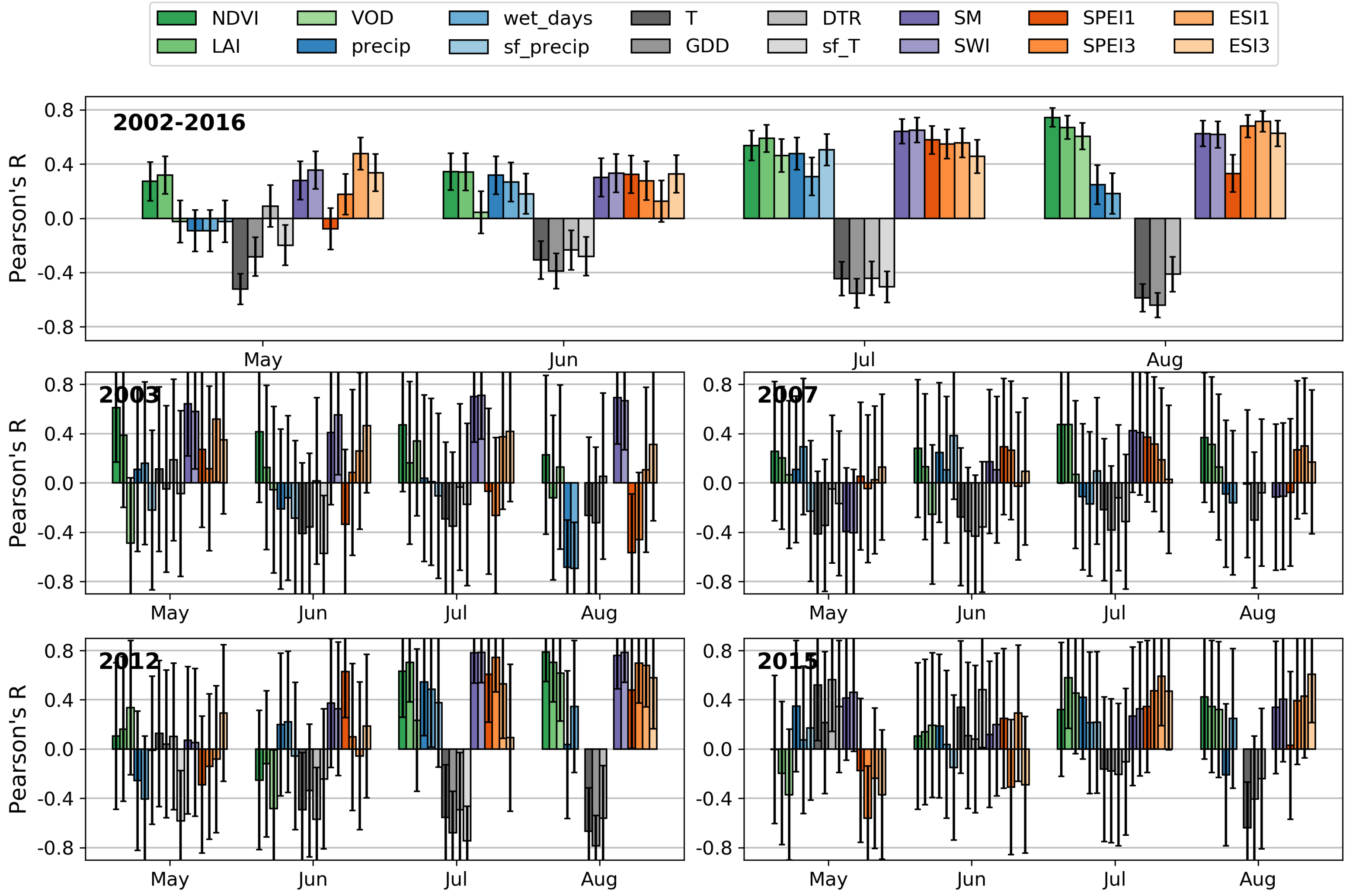
Results

Correlation analysis of input parameters and final wheat yields:

- Clear pattern for all years
- Increasing closer to harvest
- Temperature-related negativ, rest positive
- EO and temperature highest correlations
- Barely significant in individual years

Maize yields:

- Higher than for wheat
- EO and SPEI highest correlations
- Barely significant in individual years



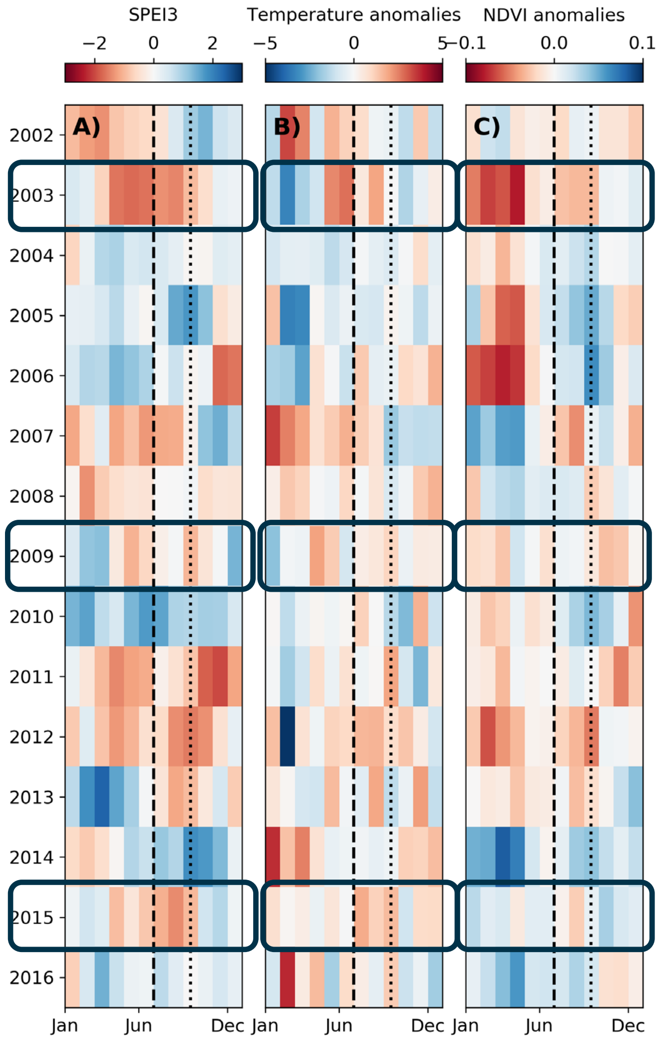
Discussion

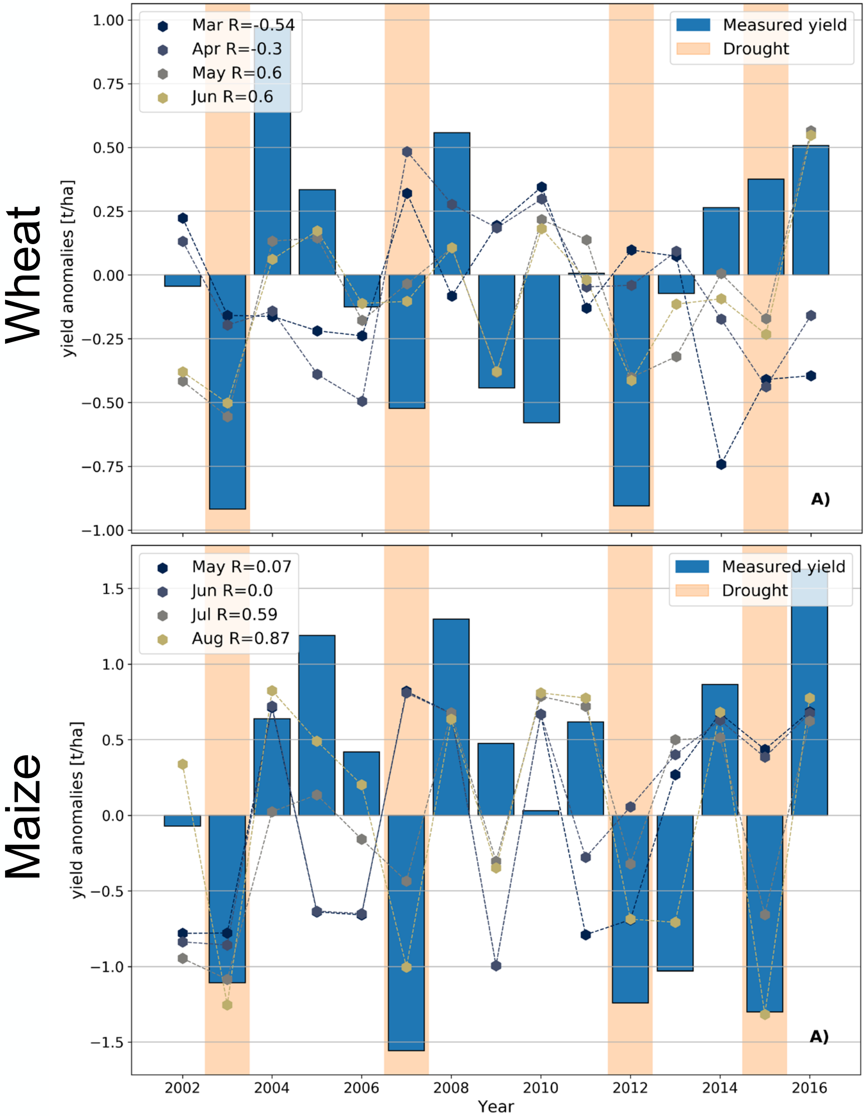
Model can mostly be explained by variability of the three main input parameters:

2003: long-lasting severe drought starting from around March

2009: High fluctations over the months, lead to unreliable (long-term) forecasts

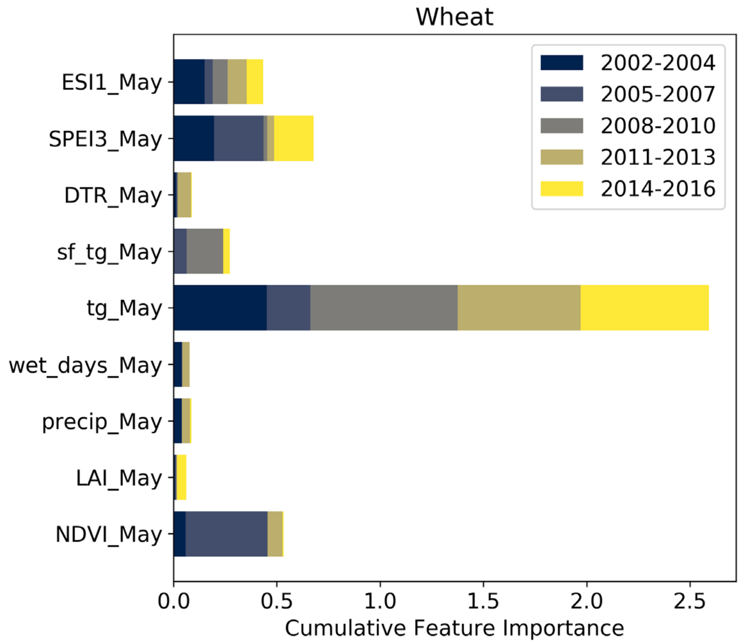
2015: Wheat not yet impacted much by drought, but maize



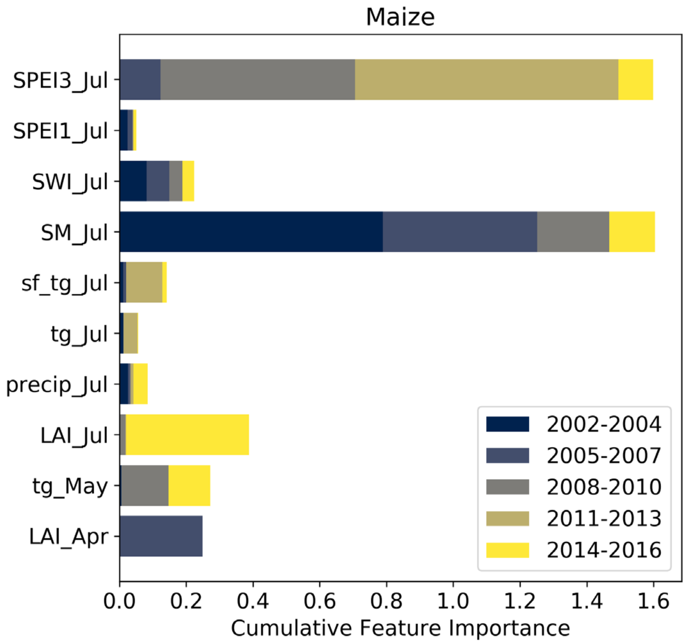


Crop yield of maize and wheat are highly dependent on the conditions in the last two months before the harvest

Appendix



Wheat
ESI1_May
SPEI3_May
DTR_May
sf_tg_May
tg_May
wet_days_May
precip_May
LAI_May
NDVI_May
2002-2004
2005-2007
2008-2010
2011-2013
2014-2016
0.0
0.5
1.0
1.5
2.0
2.5
Cumulative Feature Importance



Maize
SPEI3_Jul
SPEI1_Jul
SWI_Jul
SM_Jul
sf_tg_Jul
tg_Jul
precip_Jul
LAI_Jul
tg_May
LAI_Apr
2002-2004
2005-2007
2008-2010
2011-2013
2014-2016
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Cumulative Feature Importance