Predicting in-season crop yield within fields: A Sentinel-2 time series based monitoring approach

Eatidal Amin[1], Luca Pipia [2], Santiago Belda [3], Gregor Perich [4], Lukas Valentin Graf [4], Helge Aasen [5], Shari Van Wittenberghe [1], José Moreno [1], Jochem Verrelst [1]

[1]University of Valencia, Spain; [2]Institut Cartogràfic i Geològic de Catalunya, Spain; [3]University of Alicante, Spain; [4]Institute of Agricultural Sciences, ETH Zürich, Switzerland; [5]Division Agroecology and Environment, Agroscope, Switzerland

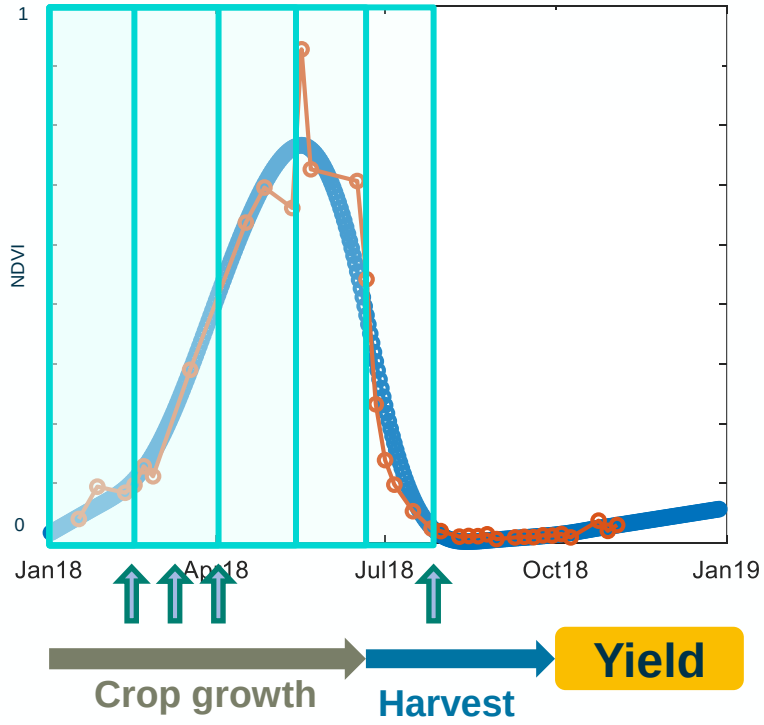
ESA UNCLASSIFIED - For ESA Official Use Only

Objective: In-season crop yield forecasting

Timely assessment of crop productivity to optimize management strategies and assess potential yield loss

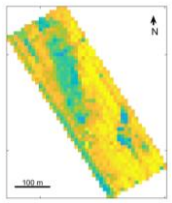
- Remote sensing **time series analysis**
- Regular **tracking of crop development**
- Predict **crop yield within the growing season** at pixel scale

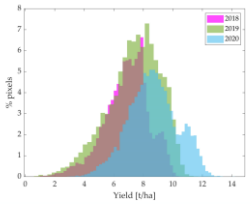
Crop yield: harvest product (e.g. grain) per unit cropped area [t/ha]



Workflow

Field data





Remote sensing time series

Machine Learning Regression Algorithms (MLRA)

Gaussian Process Regression (GPR)
Kernel Ridge Regression (KRR)
Random Forest (RF)

MLRA models validation

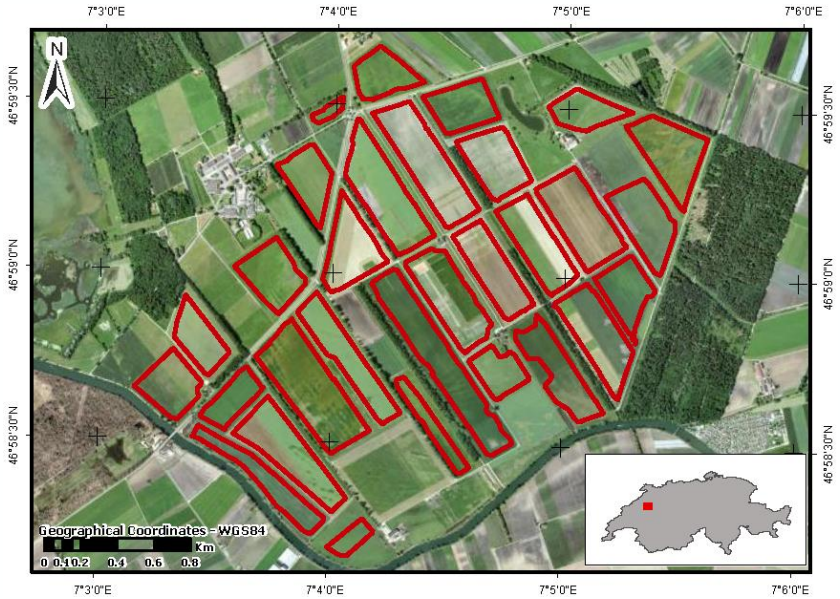
Yield forecasting accuracy assessment

In-season yield mapping

Amin, E., Pipia, L., Belda, S., Perich, G., Graf, L. V., Aasen, H., Van Wittenberghe, S., Moreno, J. & Verrelst, J. (2024). In-season forecasting of within-field grain yield from Sentinel-2 time series data. International Journal of Applied Earth Observation and Geoinformation, 126, 103636.

Study area, field data and satellite imagery

Field data Northwestern Switzerland



- Cereal crops: **wheat, barley and triticale**
- Combine harvester yield data (10m)
- **2018 – 2020**

Sentinel-2, 10m Vegetation indices (VIs)

NDVI

$$NDVI = \frac{NIR - red}{NIR + red}$$

Kernel NDVI (kNDVI)

$$kNDVI = \frac{k(NIR,NIR) - k(NIR,red)}{k(NIR,NIR) + k(NIR,red)}$$

Non-linear enhanced version

Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., ... & Running, S. W. (2021). A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances*, *7*(9), eabc7447.

Temperature data

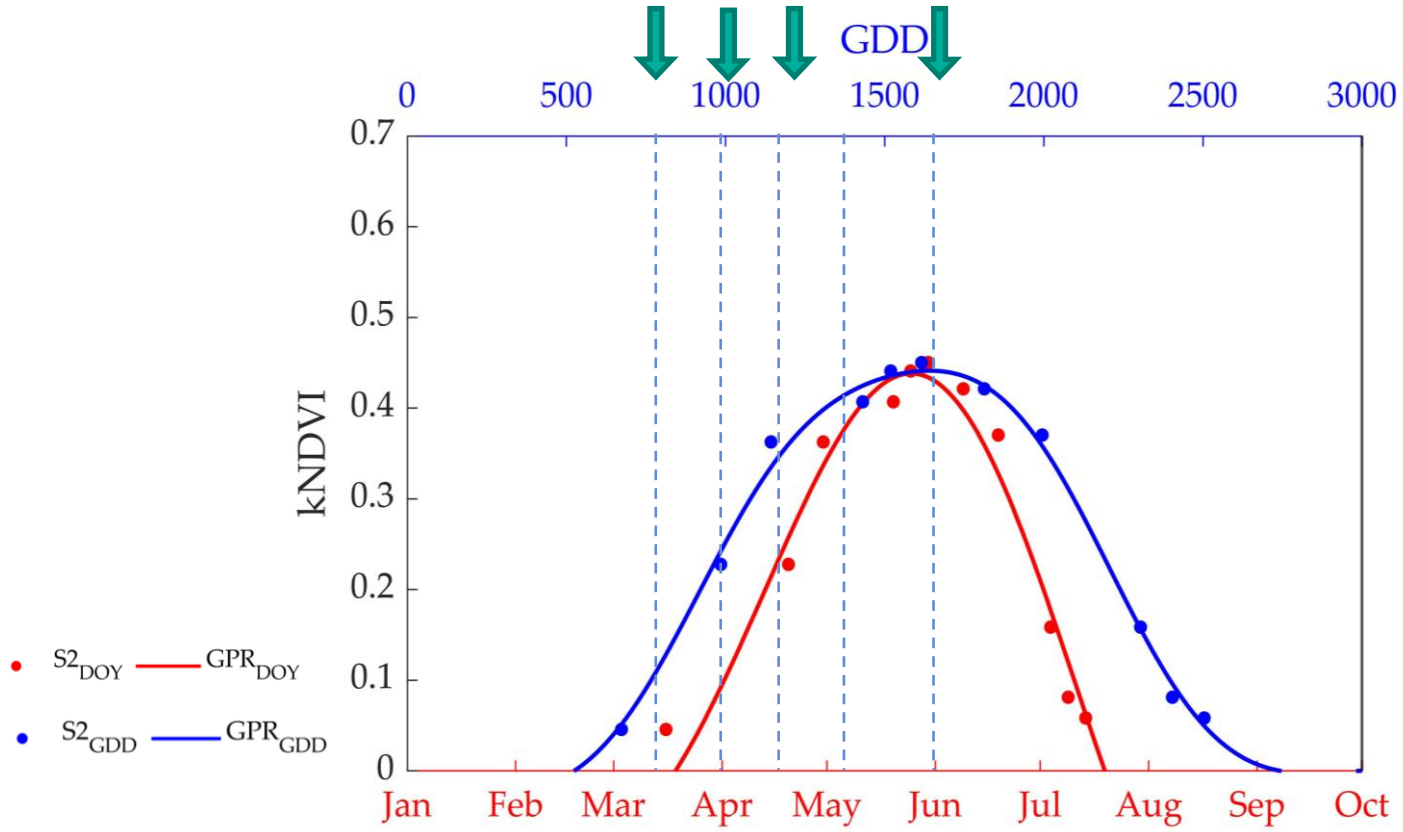
- 2m above ground air temperatures 1 × 1 km grid tile (MeteoSwiss)
- Growing Degree Days (GDD)

$$GDD = [(T_{max} + T_{min})/2] - T_{base}$$

NDVI, kNDVI time series and yield forecasting

GPR Gap-filling. Reconstruction of continuous time series for any time and year.

- **D**ay-**o**f-**y**ear **(DOY)** Regular time series (5-stp)
- **G**rowing **D**egree **D**ays **(GDD)** Cumulative temperate based time series (5-stp)



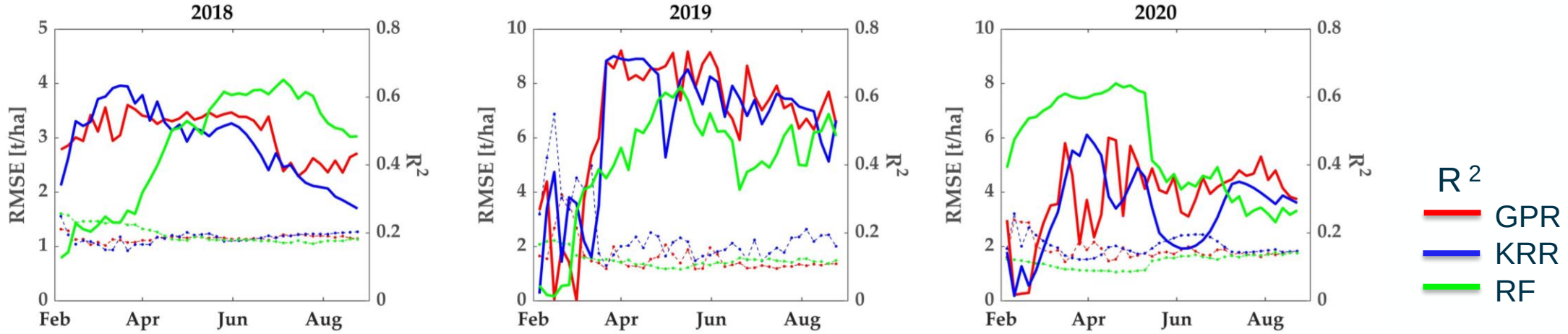
Yield field data
- Training (60%)
- Testing (40%)

MLRA
GPR, KRR, RF
900 training samples
400 testing samples

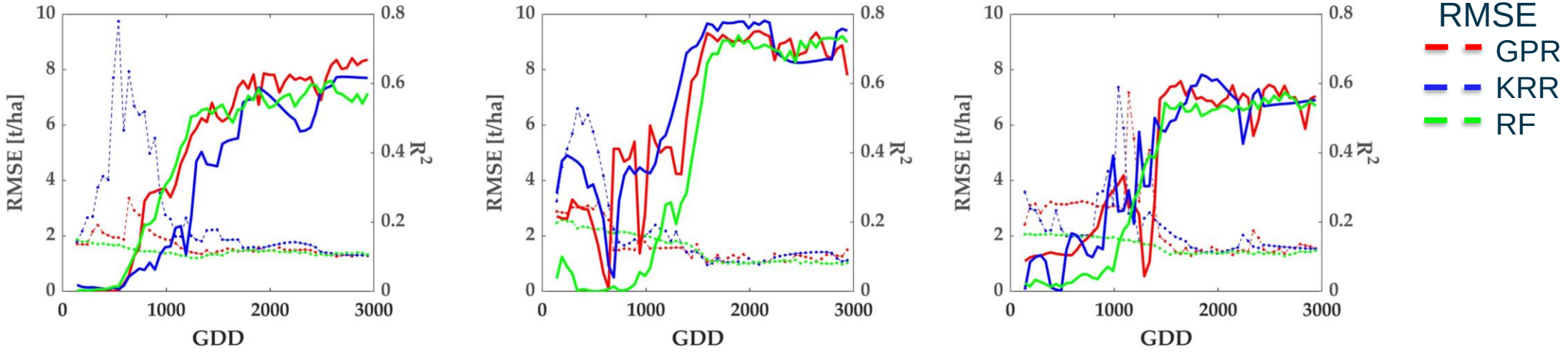
Results: Yield forecasting DOY vs GDD

- Performance in accuracy and error: GDD >> DOY

DOY

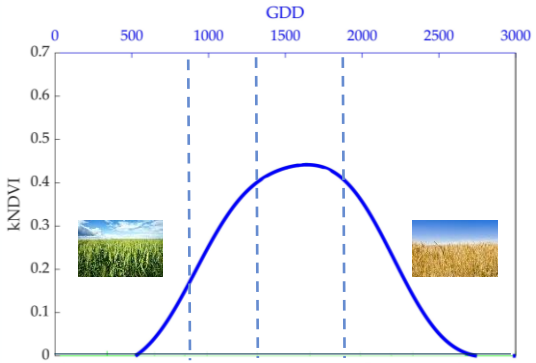


GDD



Results: Yield forecasting (same year)

- **Same-year training and testing (2019)**
- Performance in accuracy and error: kNDVI > NDVI



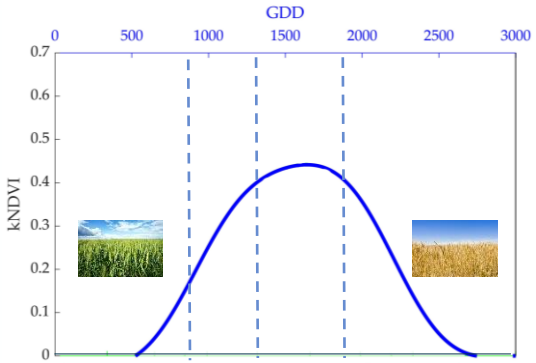
2019	900 GDD		1300 GDD		1800 GDD	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
NDVI	0.13	1.99	0.45	2.31	0.76	1.03
kNDVI	0.34	1.87	0.55	2.15	0.77	1.05

2019	900 GDD		1300 GDD		1800 GDD	
kNDVI	R^2	RMSE	R^2	RMSE	R^2	RMSE
GPR	0.43	1.50	0.34	1.58	0.73	1.20
KRR	0.34	1.87	0.55	2.15	0.77	1.05
RF	0.03	2.04	0.23	1.76	0.70	1.06

++ Accuracy -- Error

Results: Yield forecasting (unseen years)

- **Training 2019. Testing 2018 and 2020**
- General decrease in accuracy and higher errors
- Similar performance of all MLRA models tested



2018	900 GDD		1300 GDD		1800 GDD	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
GPR	0.37	1.54	0.46	1.32	0.64	1.07
KRR	0.20	1.69	0.45	1.52	0.65	1.10
RF	0.00	1.89	0.22	1.48	0.69	1.00

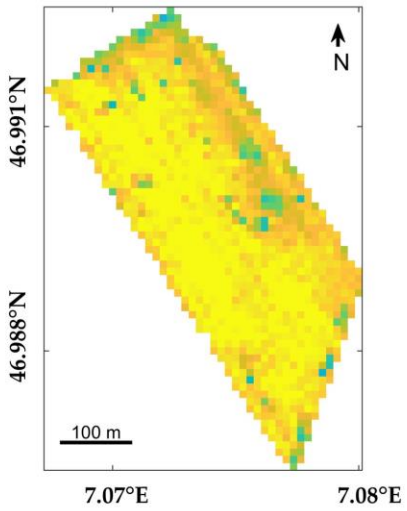
2020	900 GDD		1300 GDD		1800 GDD	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
GPR	0.10	2.20	0.14	2.17	0.52	2.61
KRR	0.10	2.48	0.34	2.90	0.51	2.58
RF	0.08	1.86	0.14	2.02	0.60	2.37

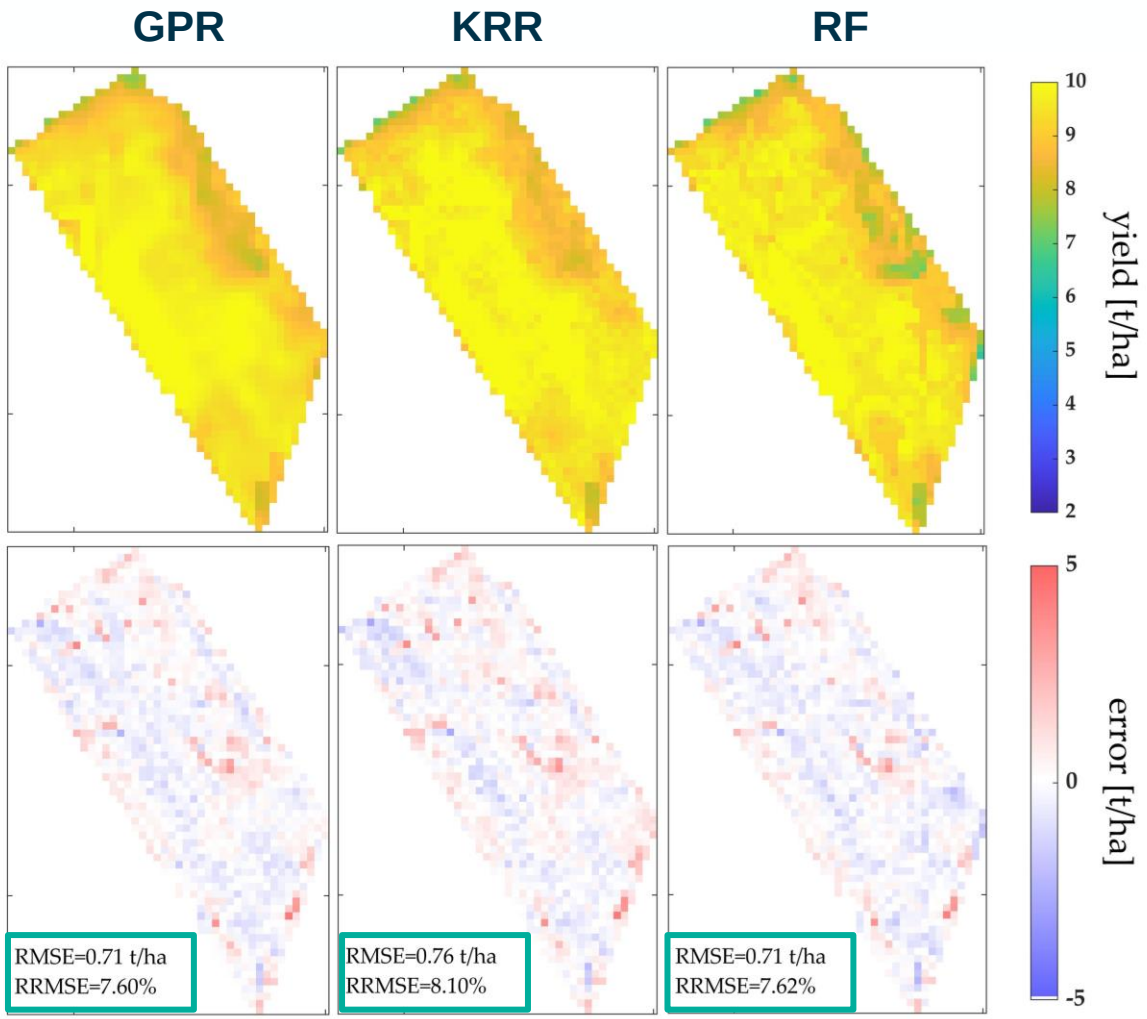
++ Accuracy -- Error

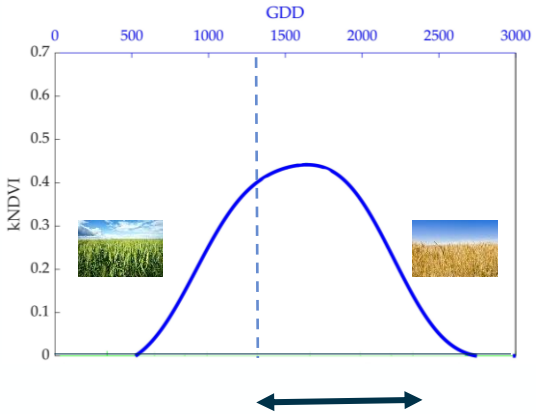
Results: Yield mapping at 1300 GDD (before harvest)

- Same-year training and testing (2019)

Reference yield map



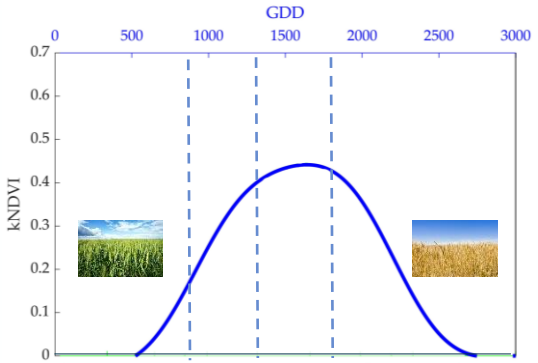




~2 months before harvest

Results: Yield seasonal mapping

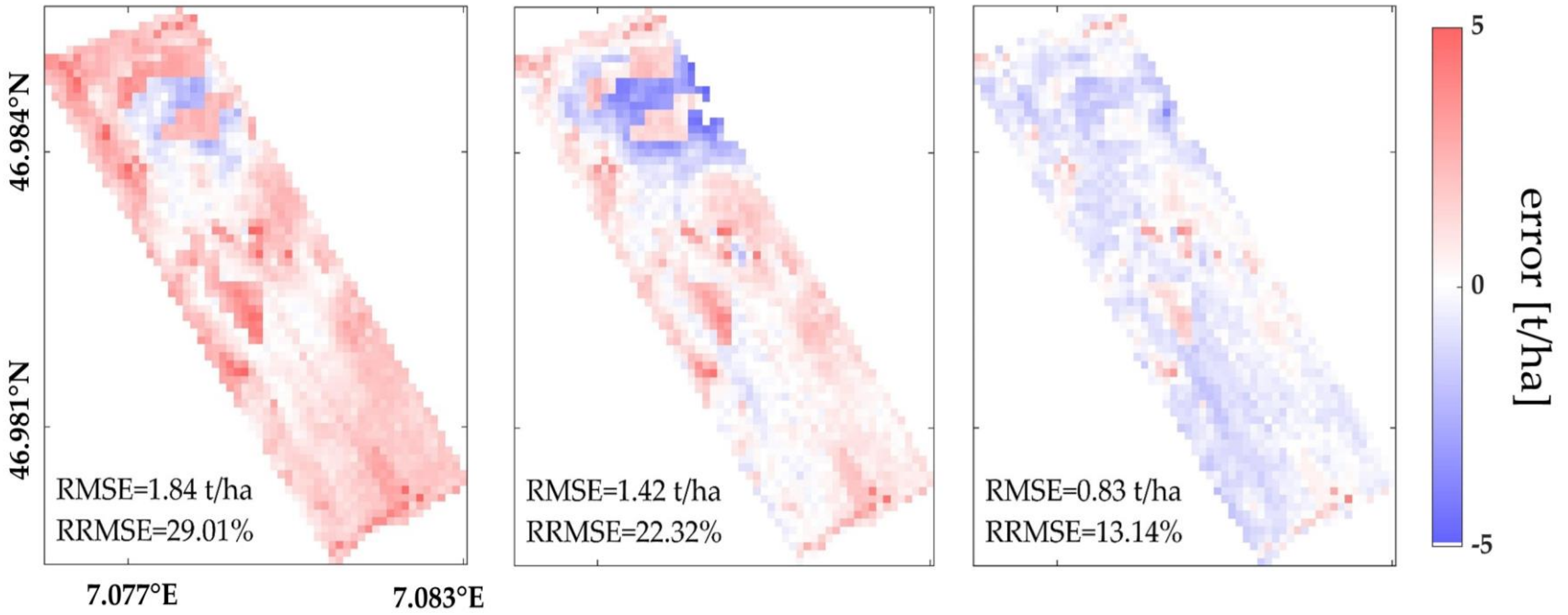
- Left-out year scenario (training 2019, testing 2018)
- Consistent improvement in accuracy across the crop growing season



900 GDD

1300 GDD

1800 GDD

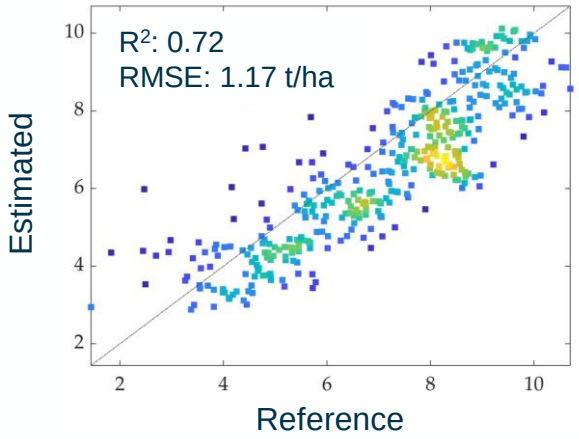


Results: Yield estimation at 2700 GDD (after harvest)

❑ **GPR**

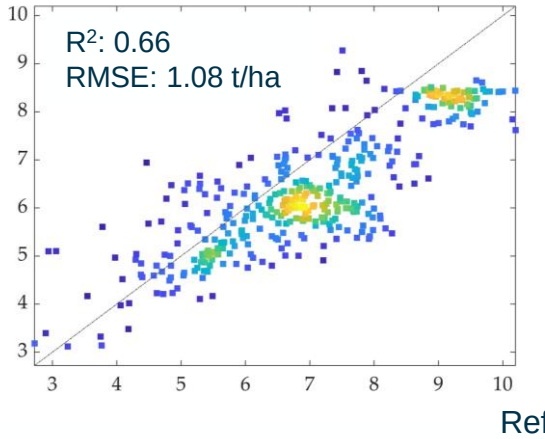
Same-year training and testing

2019

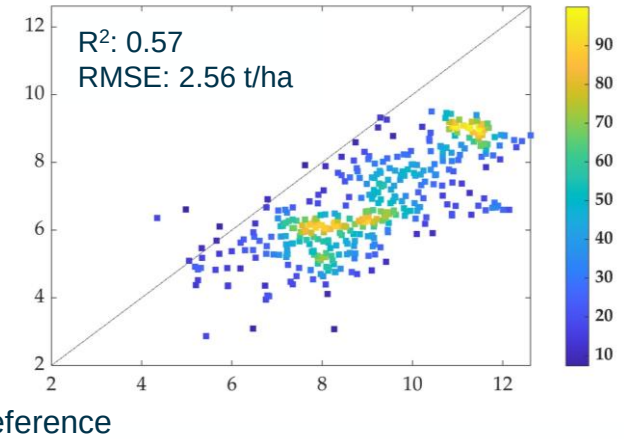


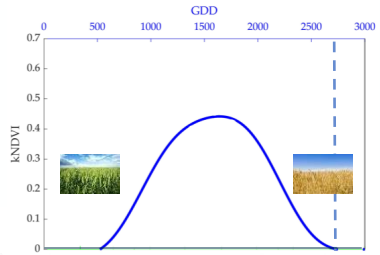
Left-out years

2018

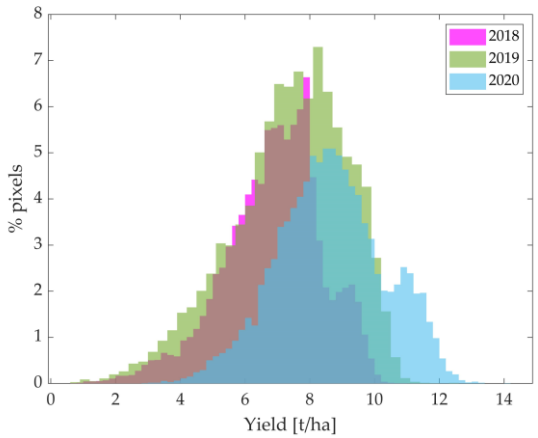


2020





Yield field dataset



Decrease in prediction accuracy of MLRA models

- ❑ Worse performance in unseen years with atypical yields
- ❑ 2020's exceptional higher yields represent a remarkable deviation from standard predictions

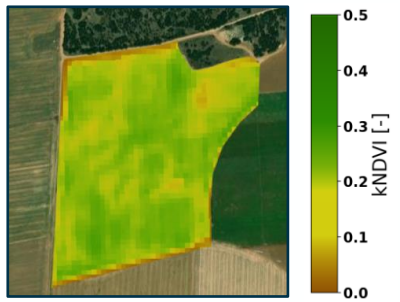
Yield forecasting at 1300 GDD (parcel aggregated yield data)

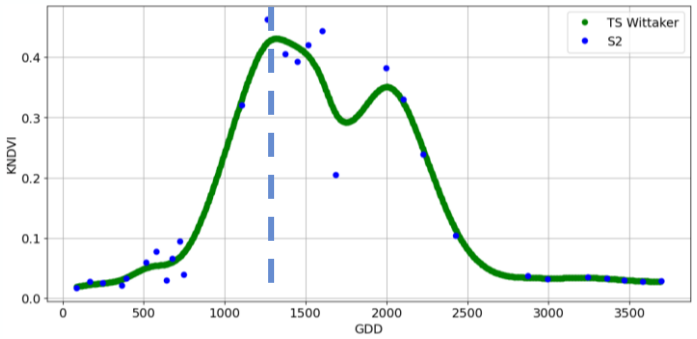
Valladolid (Spain), 2018

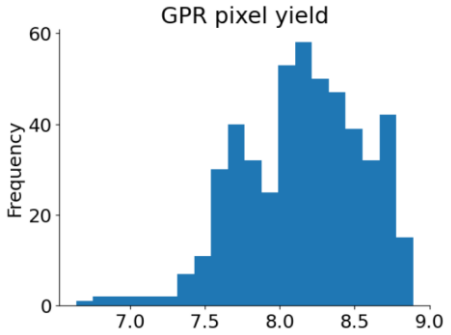
- ❑ **Parcel aggregated** yield data
- ❑ Whittaker gap-filling
- ❑ GPR yield model (active learning)

Input data	Source
kNDVI	Sentinel-2
GDD	ERA5

Wheat







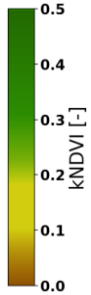
1300 GDD

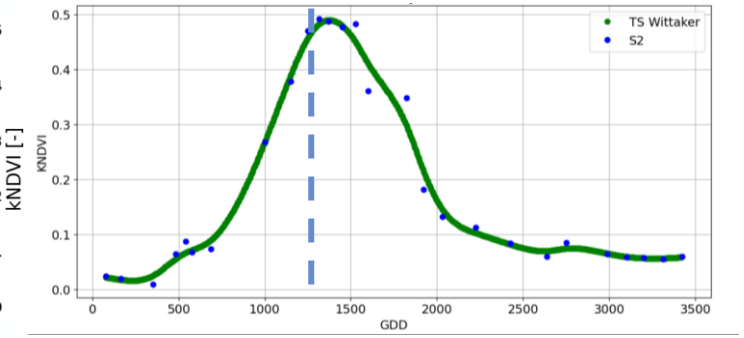
Parcel total yield

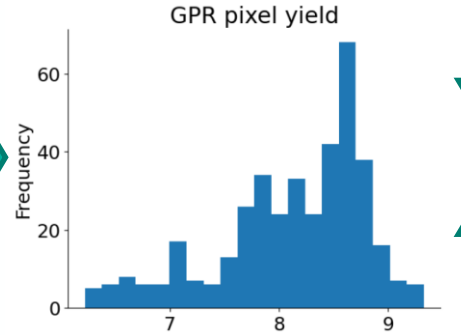
Predicted	Reference
4008.15 t/ha	4700 t/ha

~ 14%

Barley







Predicted	Reference
3185.80 t/ha	3000 t/ha

~ 6%

Conclusions

- ✓ **GDD-based VIs time series enabled** forecasting of **winter cereals** yield **before of harvest**.
- ✓ Reliable and **accurate yield forecasts were obtained** around the peak of the growing season (**flowering stage**).
- ✓ MLRA models showed a consistent **increase in forecasting accuracy along** with the crop progressive **phenology development**.

Forecasting challenge (R&D):

- ❖ Availability and accessibility of field data sets.
- ❖ Integration of management practices and crop calendar information.
- ❖ Standard accuracy/uncertainty of final products produced.

Thank you!

VNIVERSITAT DE VALÈNCIA

IPL IMAGE PROCESSING LABORATORY