Yield forecasting with machine learning and small data

Michele Meroni, Franz Waldner, Lorenzo Seguini, Herve Kerdiles, Filip Szabo, Felix Rembold

Food Security Unit - Joint Research Centre

Supported by the Antoine Royer Award

ESA-LPS 2022

Background

- **Forecasting** crop production is important for monitoring **food security in Africa**
- **Satellite remote sensing** has become instrumental to estimate both components of production (**yield and area**)
- **Yield empirical models** rely on the correlation between meteo variables, VIs and biophysical properties of the crops
- Yield **data** needed to train the models, and usually **the more the better**

In regional yield forecasting we use official yield statistics available at admin level

small data with poorly characterized quality

challenging for machine learning and especially for deep learning

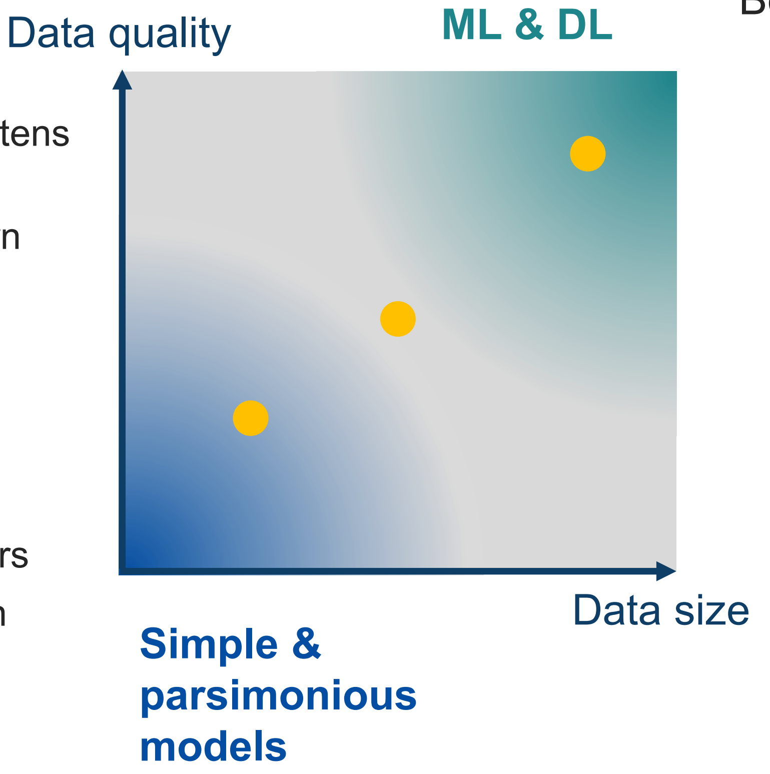
Space of data quality and availability

Our study

- Yield statistics for tens of admin units for tens of years
- Prior knowledge of where crops are grown *(general and static cropland map)*

Worst case scenario

- Yield stats at few admin units for few years
- No information on where crops are grown



Best case scenario

- Yield measured in situ *(geolocated crop cuts, yield maps from harvesters)*
- Prior knowledge of where each crop type is grown *(yearly crop type maps, including in-season for current year)*
- Large sample size in both space and time dimension

Objectives

Q: understand if and to what extent machine learning (ML) and deep learning (DL) methods can improve the accuracy of regional crop yield forecasts

Develop an operational/reusable workflow for regional yield forecasting

Ensure smooth technology transfer to interested African partners by:

- developing scripts using free and open software
- making use of public satellite and climate data

European Commission

Demonstrate the workflow in Algeria, where cereal production faces high inter-annual variability

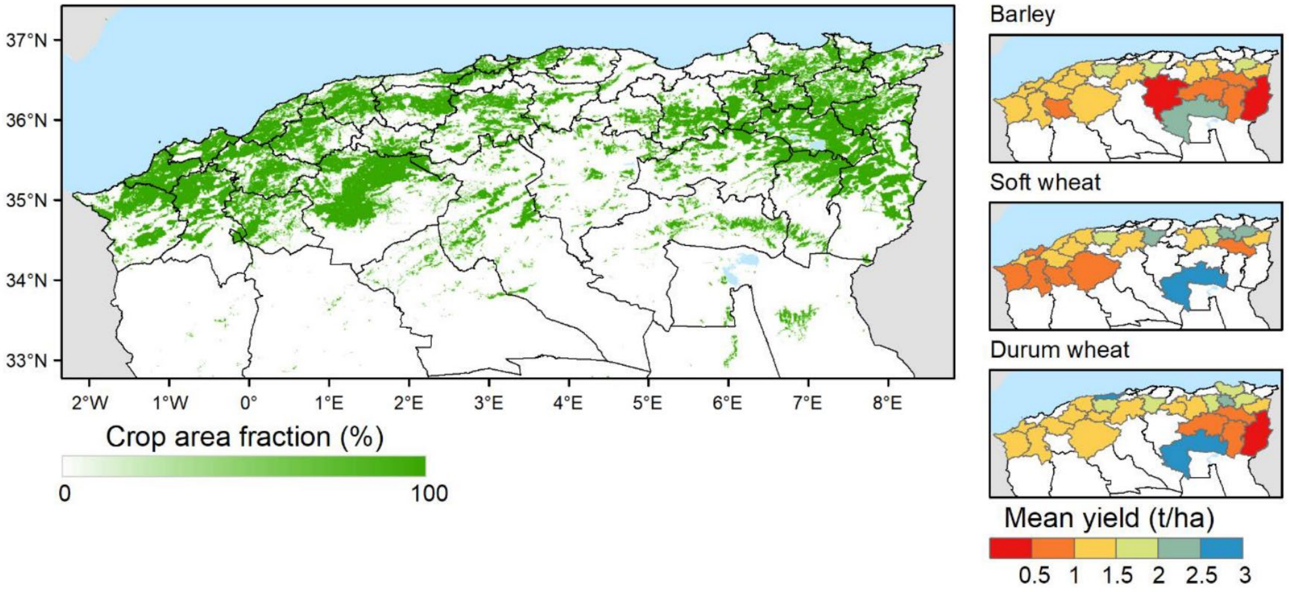
Soft wheat, durum wheat, barley

t/ha

Yields range from 0.5 t/ha to 2.5 t/ha and are highly influenced by climate variability

20+ Provinces per crop representing 90% of the national mean crop production

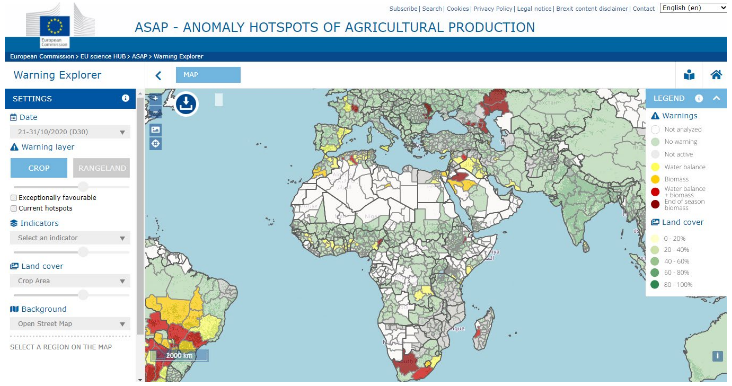
16 years of yield stats (2002-2018)



European Commission

We forecast yields every month during the growing season

- Predictors downloaded from the JRC Early Warning System ASAP as tabular data - aggregated in time (10-day) and in space (GAUL1 Admin Unit) using crop area fraction image from GlobeLand 30m

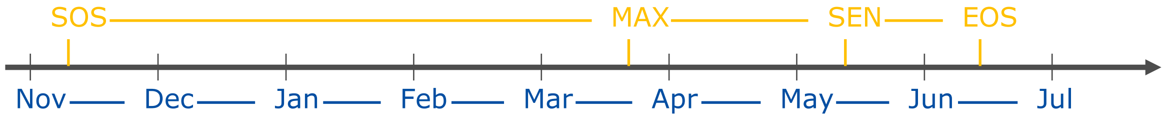


https://mars.jrc.ec.europa.eu/asap

- Predictors aggregated at monthly time step

Category	Variable	Aggregation	Source
Satellite obs of vegetation	**NDVI**	Avg, max	MODIS 1 km
Meteorology	**Precipitation**	Sum	CHIRPS 5 km
	Temperature	Avg, min, max	ECMWF 25 km
	Global radiation	Sum	

- Yield are predicted monthly using all observations obtained so far in the season (incomplete information)



ML workflow

1. Empirical definition of predictor sets to be tested
2. Automatic predictor selection
3. Identification of model parameters and evaluation of model accuracy through nested cross-validation

Machine learning workflow

→ This workflow is repeated for a large number of model configurations

Predictor sets: guided selection of predictors

We defined six sets of predictors (all variables, only RS, only Met, and reduced sets).

No predictor contains information about soil, irrigation, management practices, etc.

	Variables used						
	Remote sensing		**Metereology**				
Set name	**NDVI (avg)**	**NDVI (max)**	**Rad (sum)**	**Rain (sum)**	**T (avg)**	**T (min)**	**T (max)**
RS&Met	●	●	●	●	●	●	●
RS	●	●					
Met			●	●	●	●	●
RS&Met-	●			●	●		
RS-	●						
Met-			●	●	●		

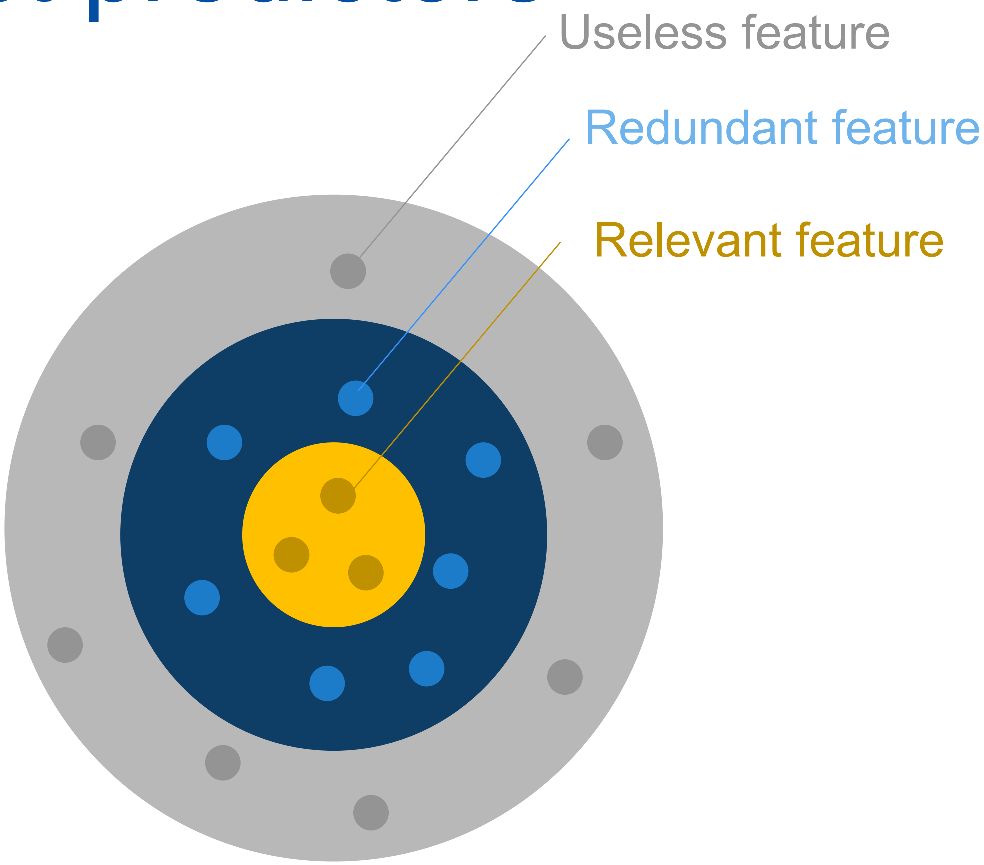
One way to convey all this missing information is to use the IDs of administrative units as predictors (one-hot encoding, thus one additional feature per unit).

Assumption: unobserved effect = *f(admin unit)*

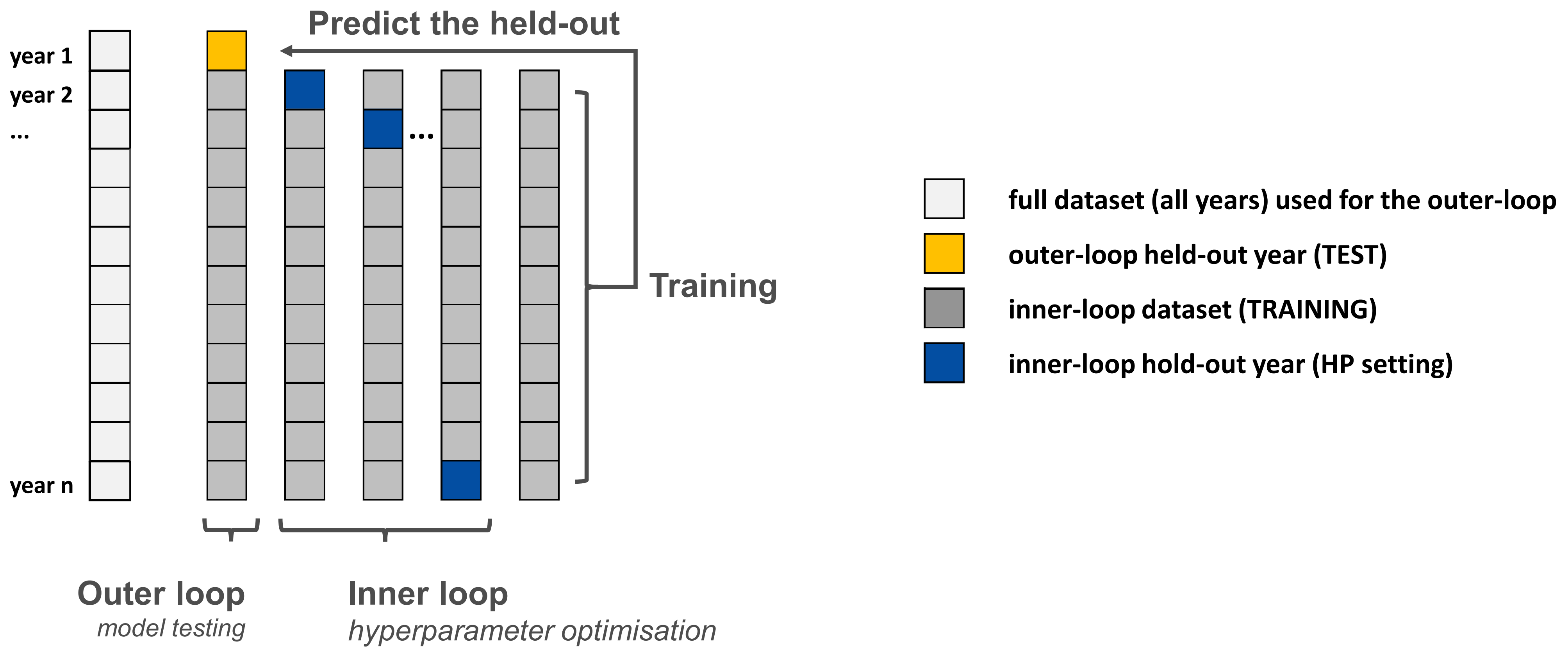
Automatic predictor selection: the best K predictors are not the K best predictors

We want to select K features that **collectively** have the strongest predictive value.

Use of “***Maximum Relevance - Minimum Redundancy***”: select the predictors that have maximum relevance with respect to the target variable and minimum redundancy with respect to the other selected predictors.

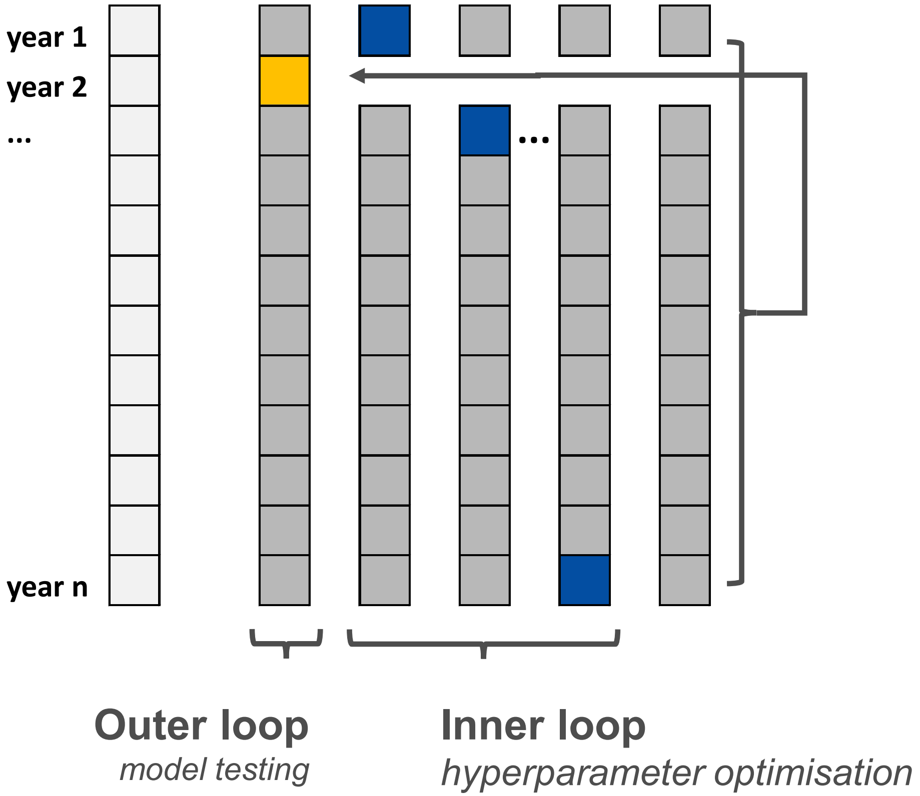


Training, validation and testing with small data avoiding info leakage: nested cross-validation



European Commission

Training, validation and testing with small data avoiding info leakage: nested cross-validation



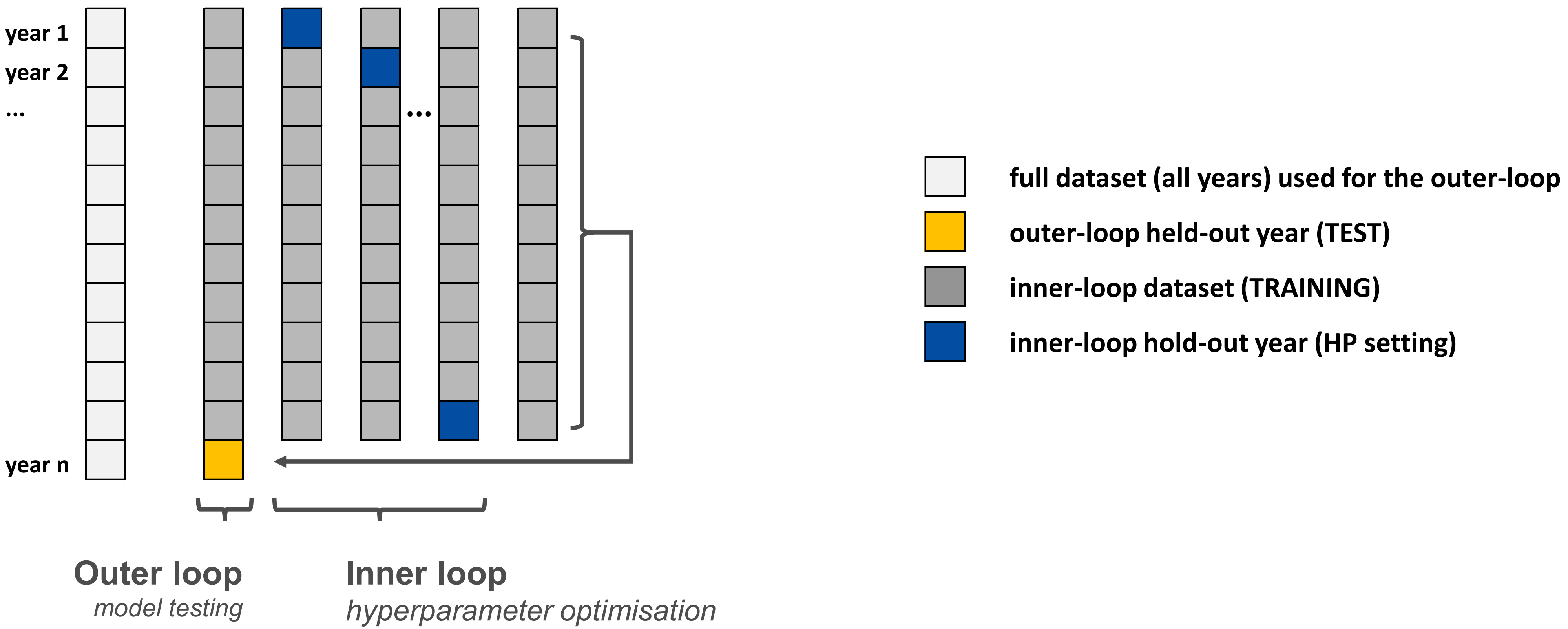
full dataset (all years) used for the outer-loop

outer-loop hold-out year (TEST)

inner-loop dataset (TRAINING)

inner-loop hold-out year (HP setting)

Training, validation and testing with small data avoiding info leakage: nested cross-validation



We tested six machine-learning models

LASSO
Linear regressor that performs variable selection and regularization.

Support Vector Regression (linear)
Regressor that finds the optimal regression hyperplane so that most training samples lie within a certain margin around it.

Support Vector Regression (rbf)
SVR mapping the input data to a high dimensional feature space using non-linear kernel functions, here, radial basis functions.

Random Forest
Ensemble regressor that averages the output of multiple regression trees.

Gradient boosting regression
Ensemble of shallow trees in sequence where each new tree minimises the residuals of the previous tree.

MultiLayer Perceptron
Artificial neural network that uses a nonlinear weighted combination of the features to predict the target variable.

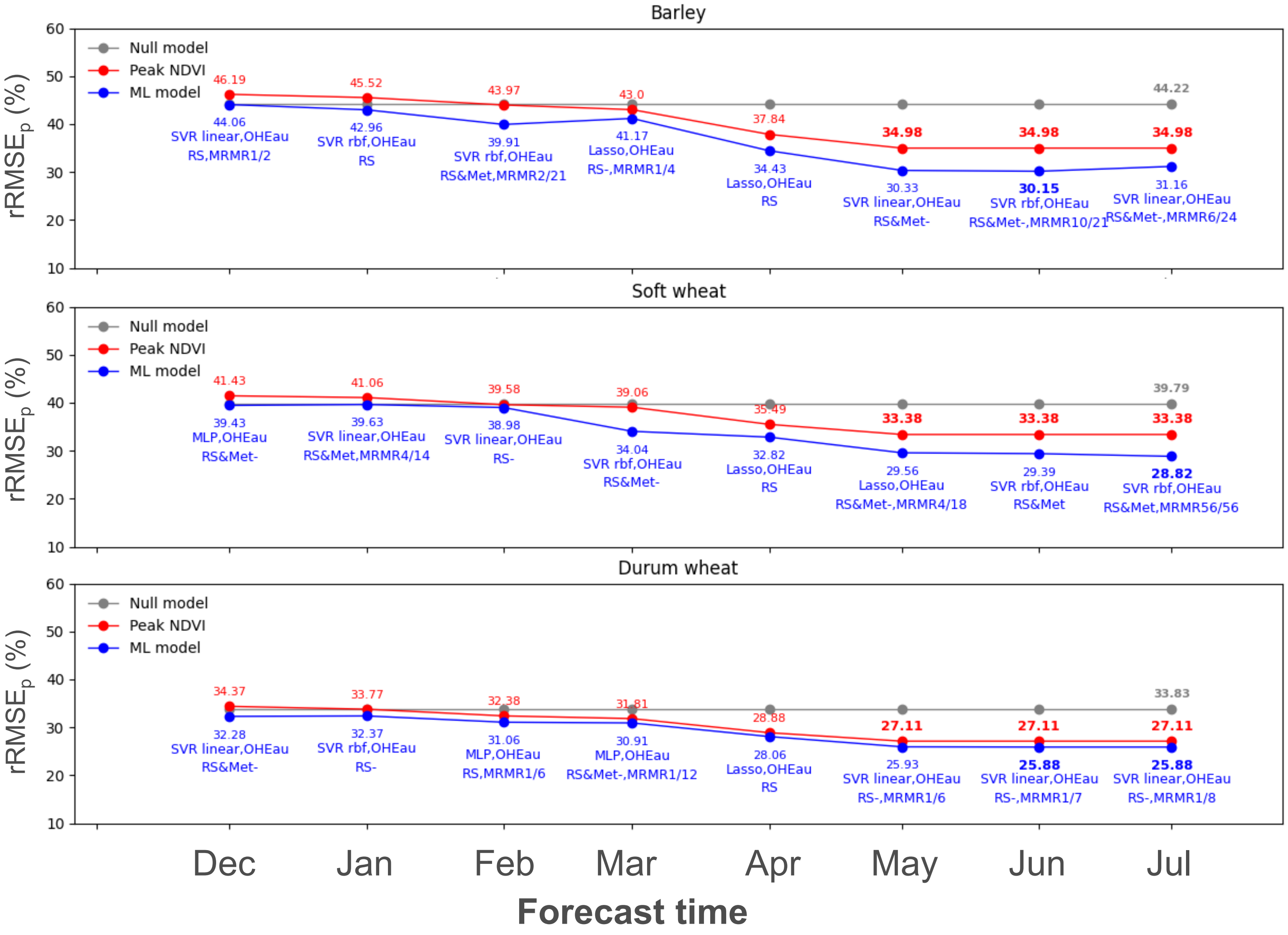
And compared with 2 benchmarks

$\bar{Y}$

Null model
Average observed yield per administrative unit.

Peak NDVI
Linear regression by administrative unit between the maximum NDVI value (peak) and yield

Machine learning is better than the benchmarks but no one method is consistently better



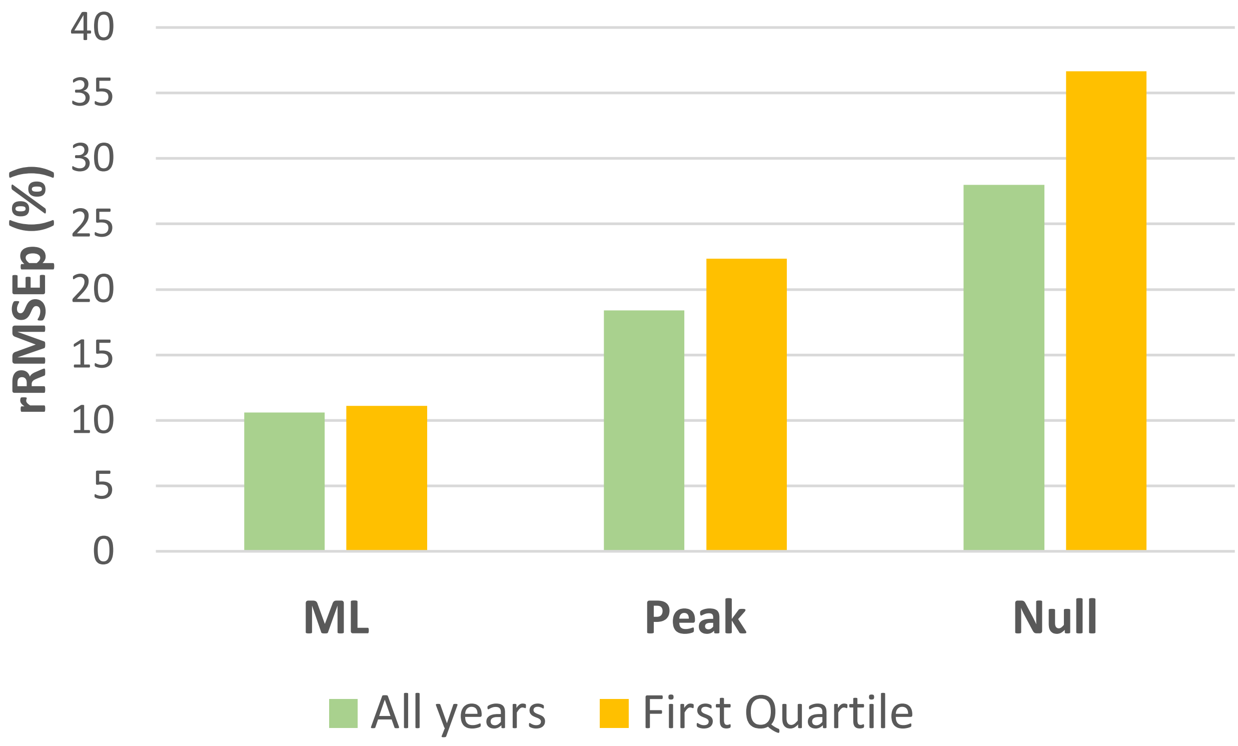
- The best machine-learning models were always more accurate than the peak NDVI model regardless of the forecast month.
- Support vector regression is the most frequently selected algorithm, followed by Lasso and MultiLayer Perceptron.
- Accuracy flattens out in May.
- Admin unit OHE & predictor selection helped increase accuracy most of the cases

Machine-learning estimates are more reliable in low-yielding years

When focusing on low-yield years (first quartile of yield distribution), the forecast accuracy of ML models remained nearly constant, unlike benchmark models.

Soft wheat – national level error

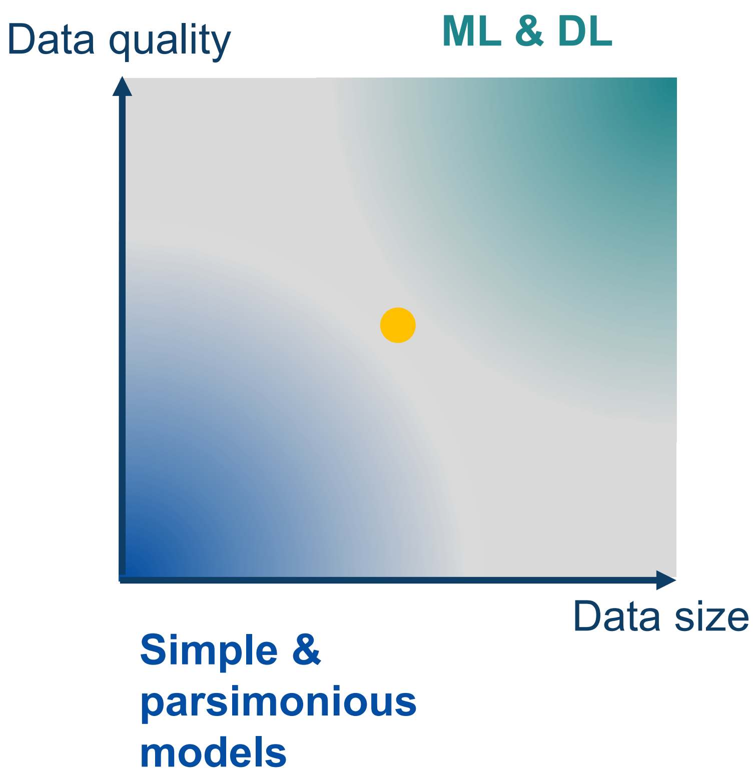
All years vs. First Quartile



ML twice as accurate as the peak in FQ

More analysis on ML results in *Meroni et al., 2021, Yield forecasting with machine learning and small data: What gains for grains?, AFM, https://doi.org/10.1016/j.agrformet.2021.108555*

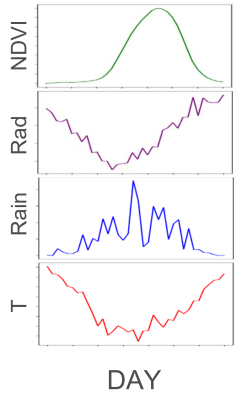
Enough data for ML, what about DL?



We tested two types of DL models

1D-CNN

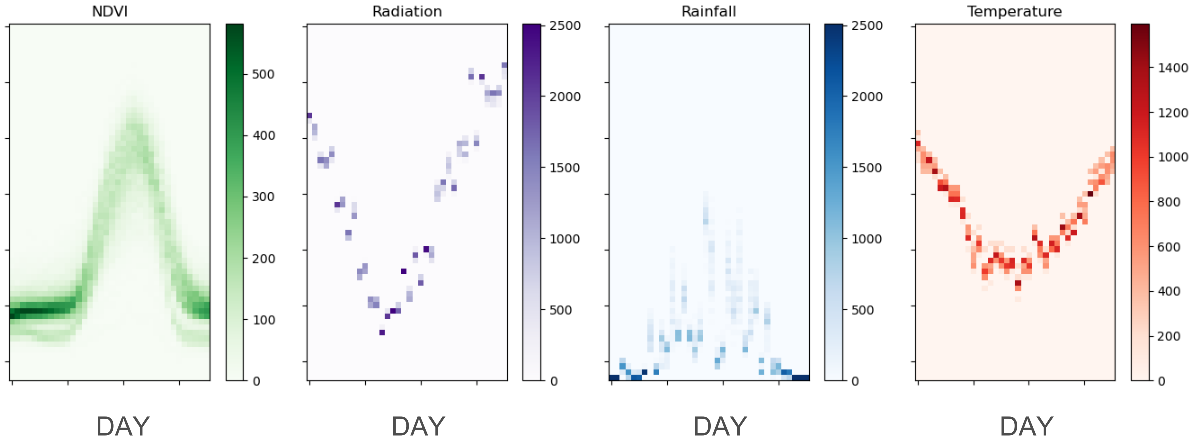
Kernels slide along 1 dimension
(time input time series)
The time series are admin level
average of input variables



2D-CNN

Kernels slide along 2 dimensions of the "image"

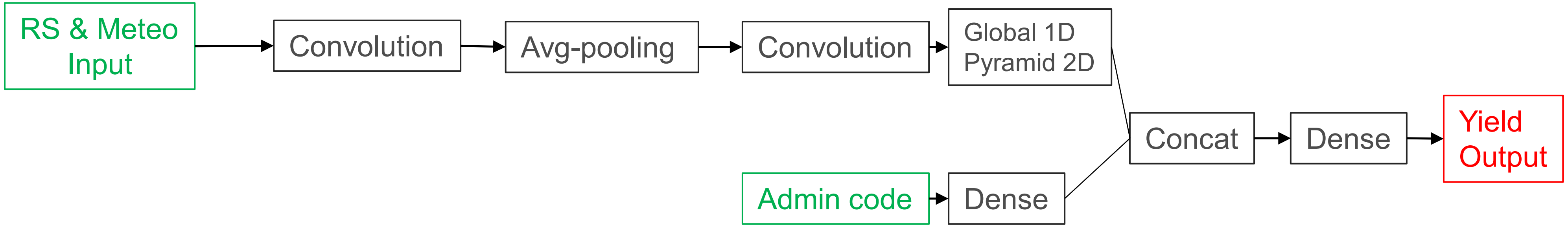
The band of the "image" are admin level
histograms (y) over (x) time of input variables
(carrying information about distributions of input
variables)



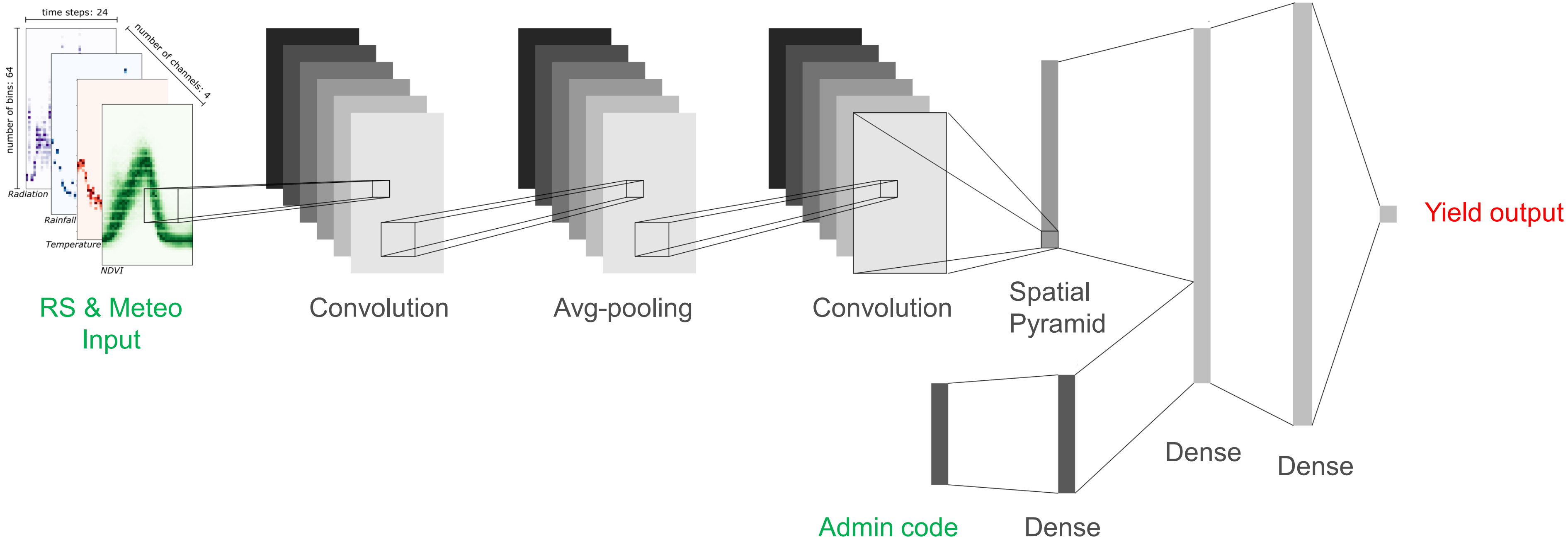
End to end approach: no feature selection as in ML workflow

Simple architectures given small data size

1-D and 2-D CNN flow:



Graphical example, 2-D CNN



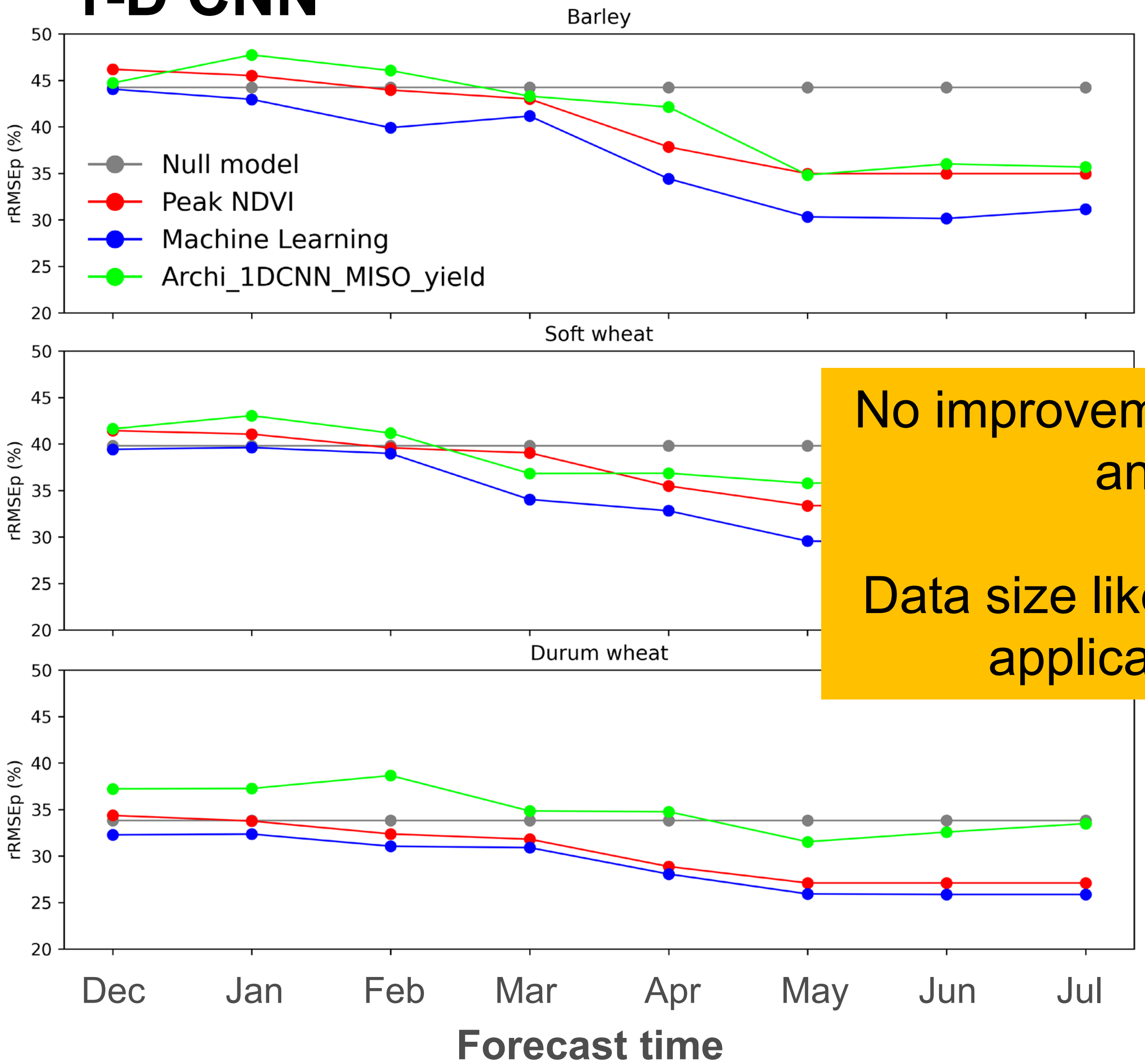
Hyper-parameters:

- Number of convolutional units
- Convolution kernel size
- Pooling size (1D), Pyramid bins (2D)
- Number of dense layers (at the end)
- Dropout rate
- Learning rate
- Number of ending dense layers
- Number of epochs
- Batch size

European Commission

CNN results vs. ML

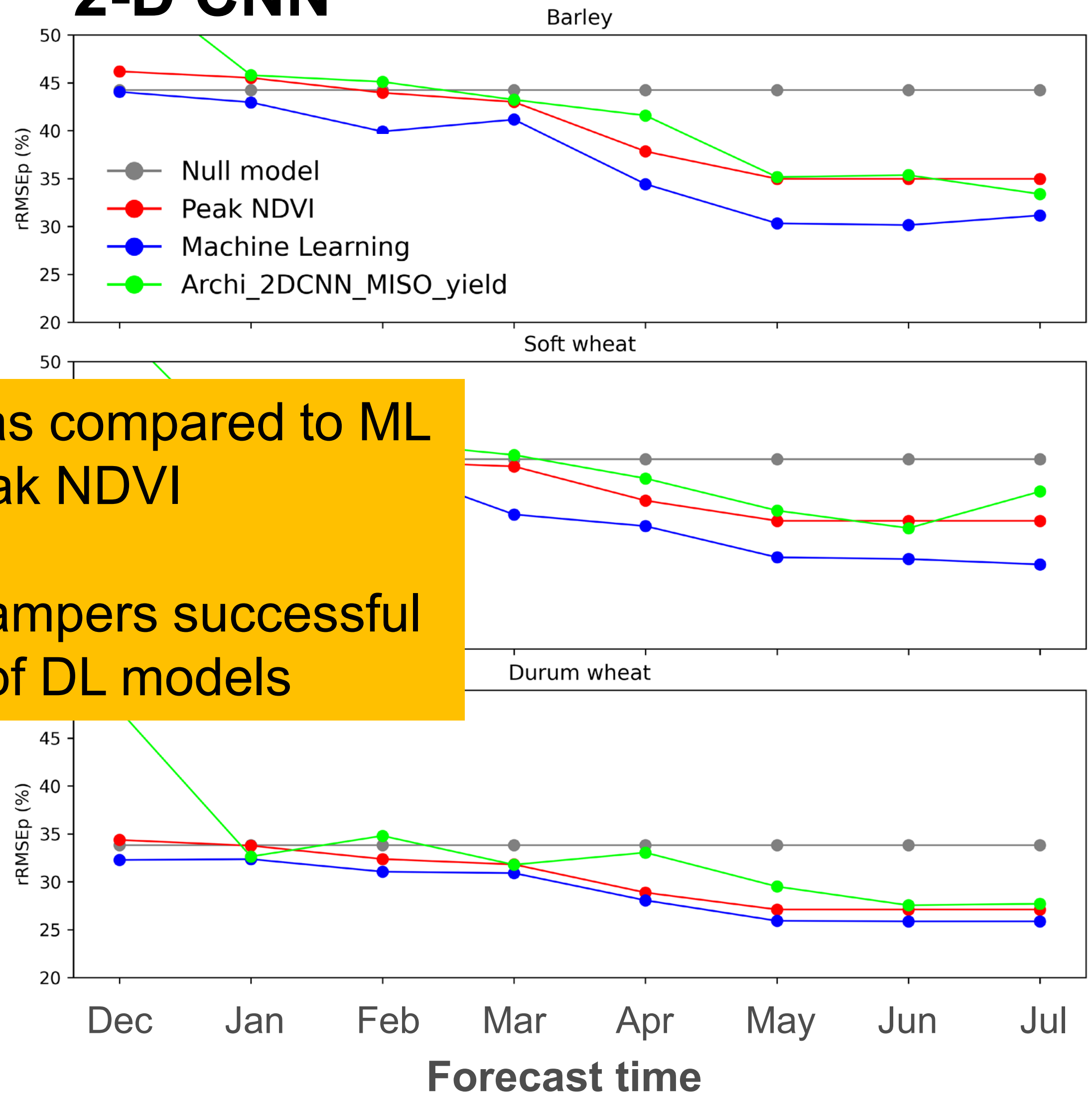
1-D CNN



No improvement as compared to ML and peak NDVI

Data size likely hampers successful application of DL models

2-D CNN



Conclusions

We presented a **generic and reusable machine learning workflow** to forecast crop yields with small, public, climate and satellite time series.

Our workflow is fully automated and identifies the best model configuration for prediction during the growing season.

We deployed our workflow in Algeria:

- the **best machine learning model always outperformed simple benchmarks** but no single model nor predictor set combination consistently delivered the best forecasts
- **data smallness prevented CNNs to improve accuracy**

→ **Model parameterization and rigorous testing are paramount** but time and resource consuming (1 month on a computer cluster).

The ML workflow is used operationally to produce near real-time forecasts in Algeria (2021 & 2022) and currently being tested in South Africa.

Partnerships with other African institutions welcome!

European Commission

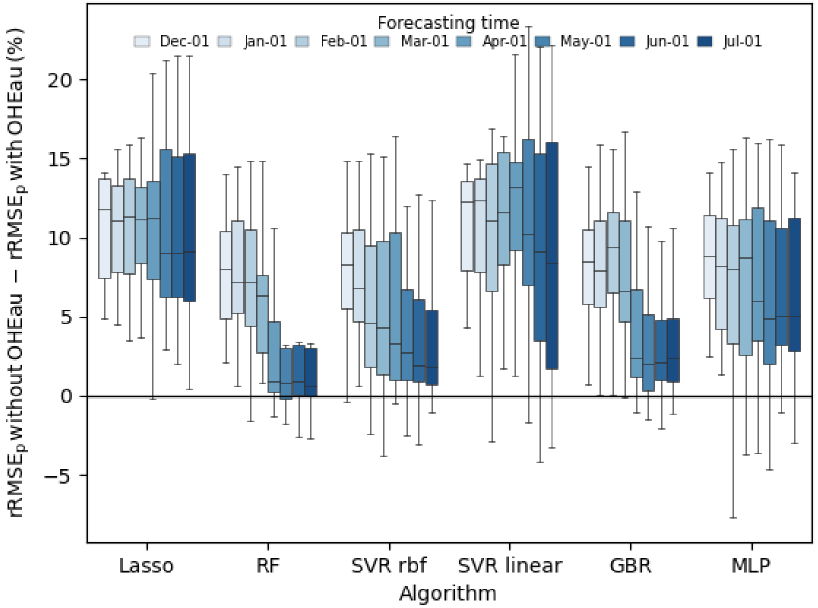
Thank you

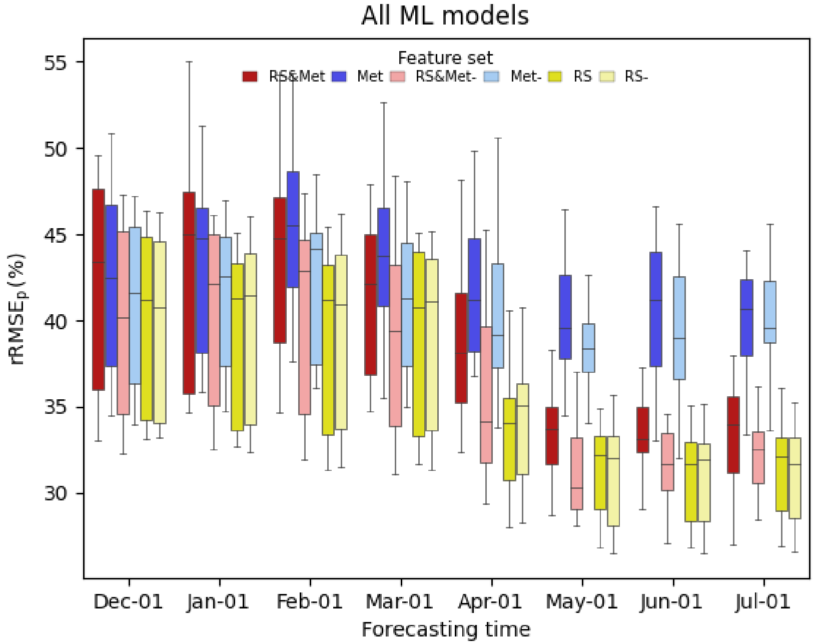
European Commission

Which predictors matter most?

One-hot encoded predictors (administrative units) boost the accuracy of all model configurations. They reduced the RMSE on average by 13 to 1%.

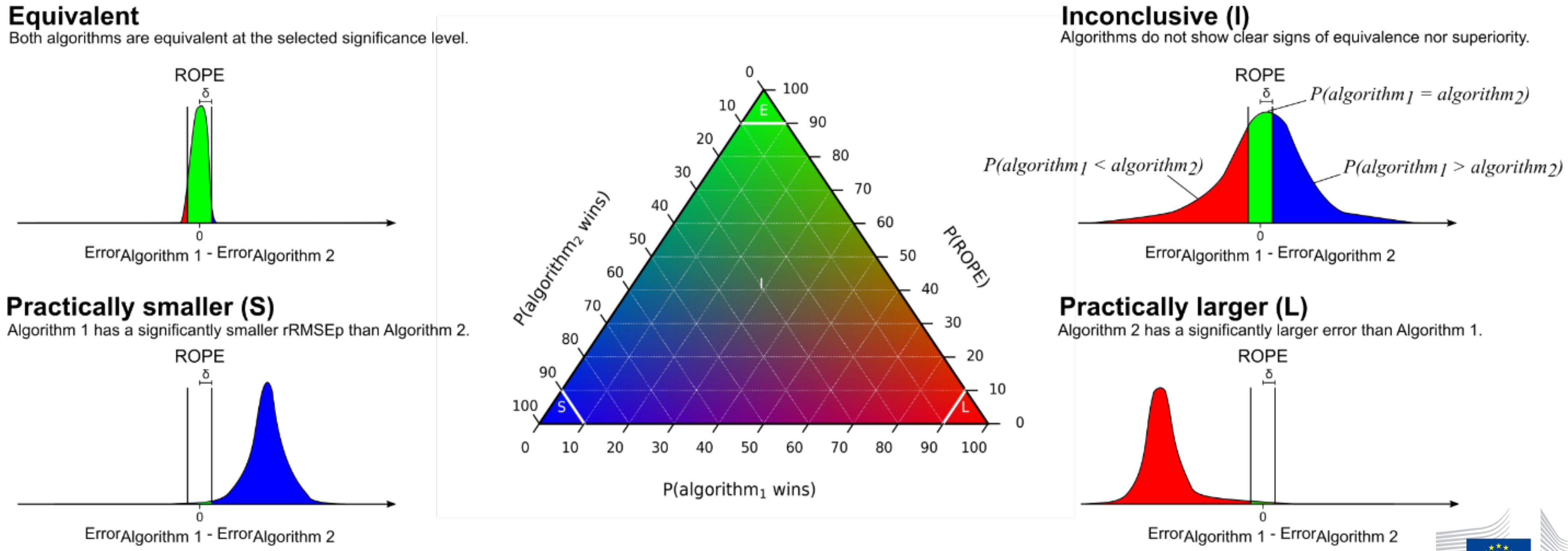
Climate predictors were mainly relevant early in the season. From April onwards, they did not allow a sensible reduction of the error unlike remote sensing features.



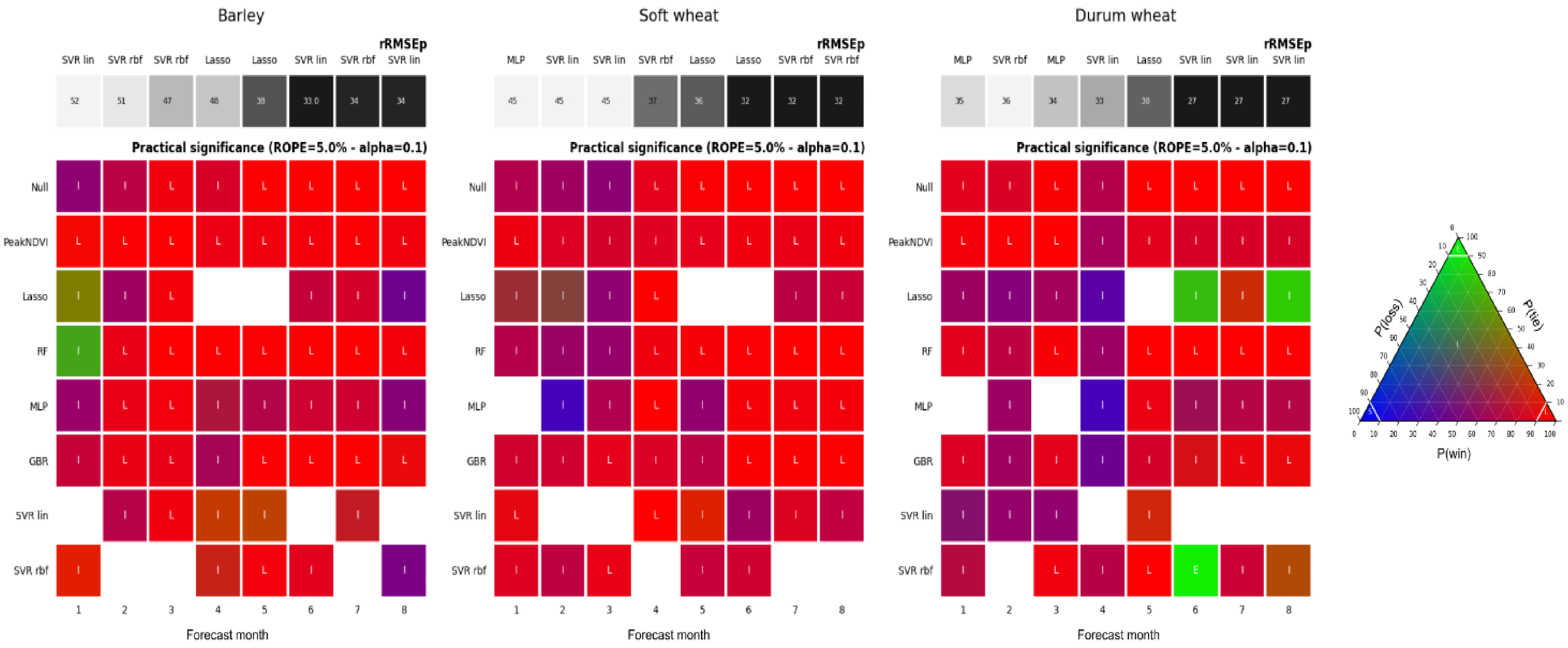


Are gains significant?

We evaluate the significance of the difference between methods using Bayesian testing. The goal of Bayesian testing is to compute the probability than one method is superior to another by a given margin.



Most differences are significant or the test is inconclusive



European Commission