Paris
2022

Is it possible to predict the occurrence of extreme agricultural yield losses and their impact of commodity prices?

david.makowski@inrae.fr

Outline

- Why many people are trying to predict crop yields?
- How to assess risk of crop yield loss?
- Why machine learning can be useful (in principle)?

Outline

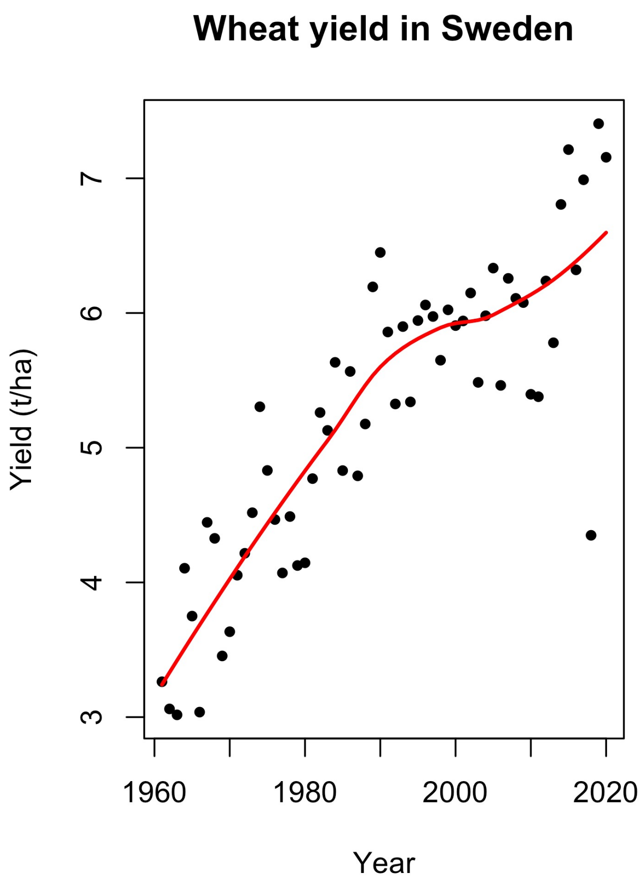
- Why many people are trying to predict crop yields?
- How to assess risk of crop yield loss?
- Why machine learning can be useful (in principle)?

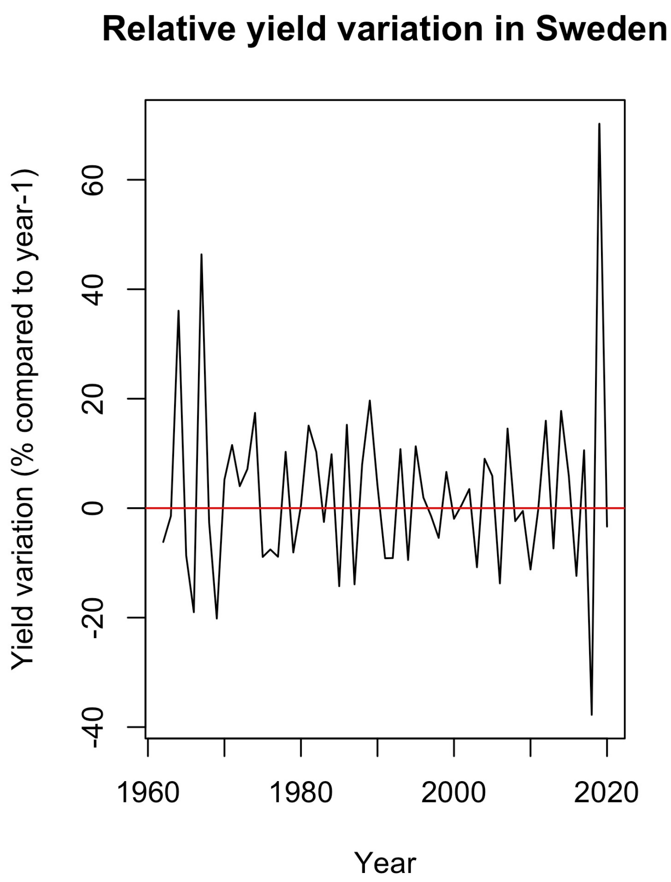
What is a « crop yield »?

Amount of crop product per unit area of land:

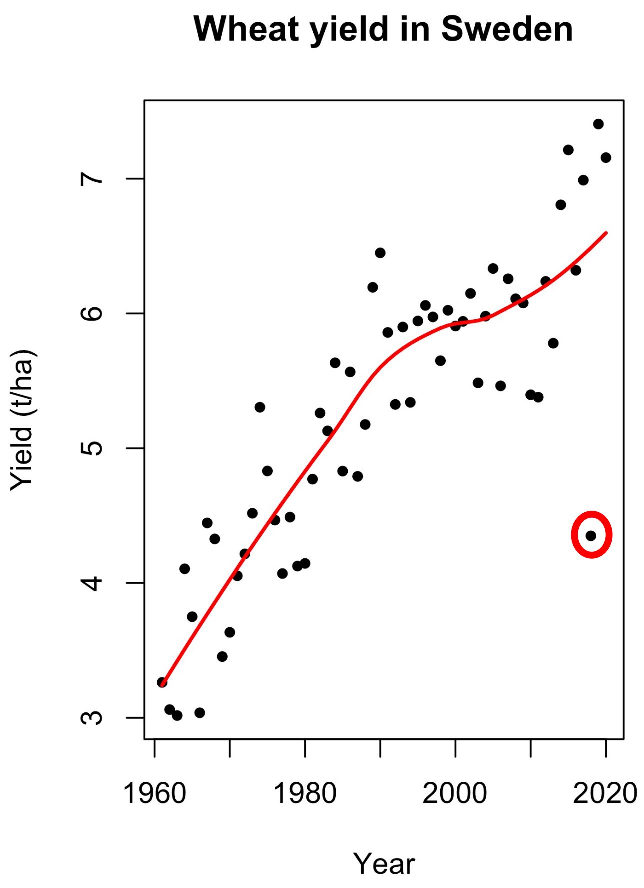
- Tons of wheat grain per ha
- Tons of corn grain per ha
- Tons of biomass per ha
- Tons of sugar per ha
- Kcal per ha
- Gj per ha

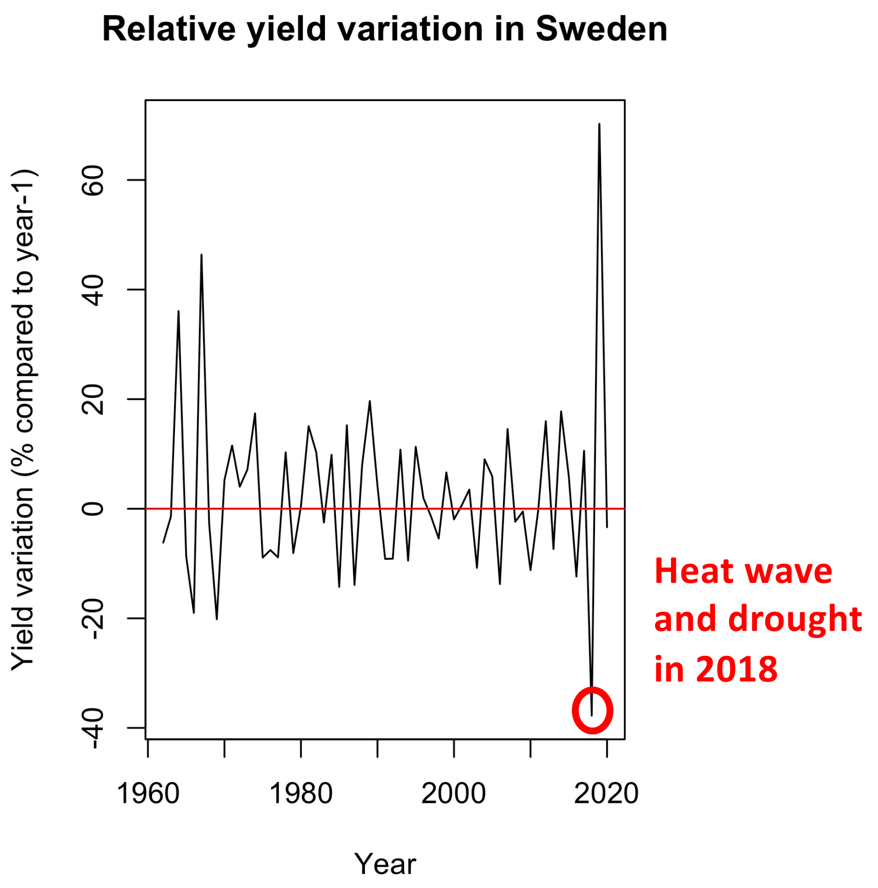
Crop yield variability





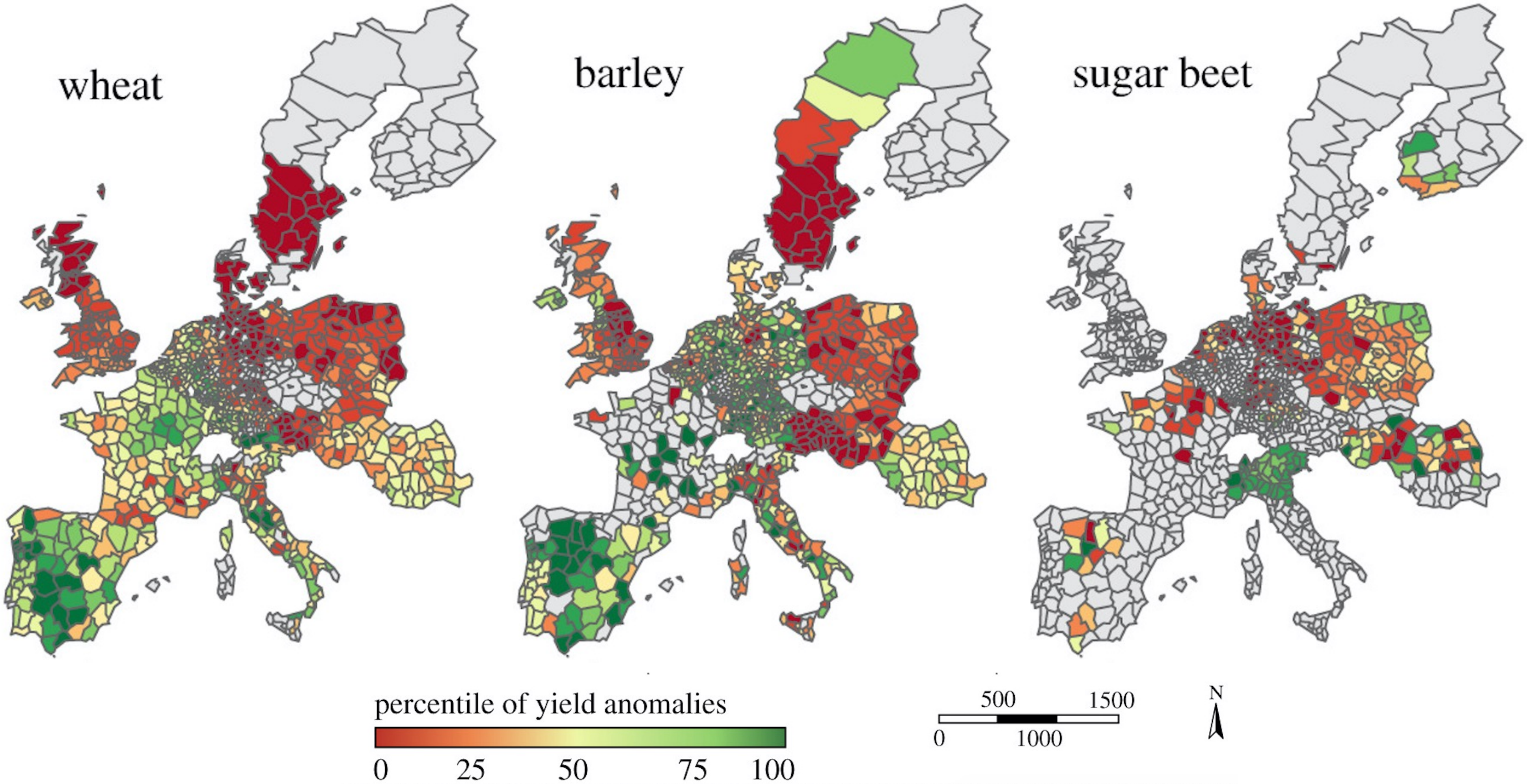
Crop yield variability



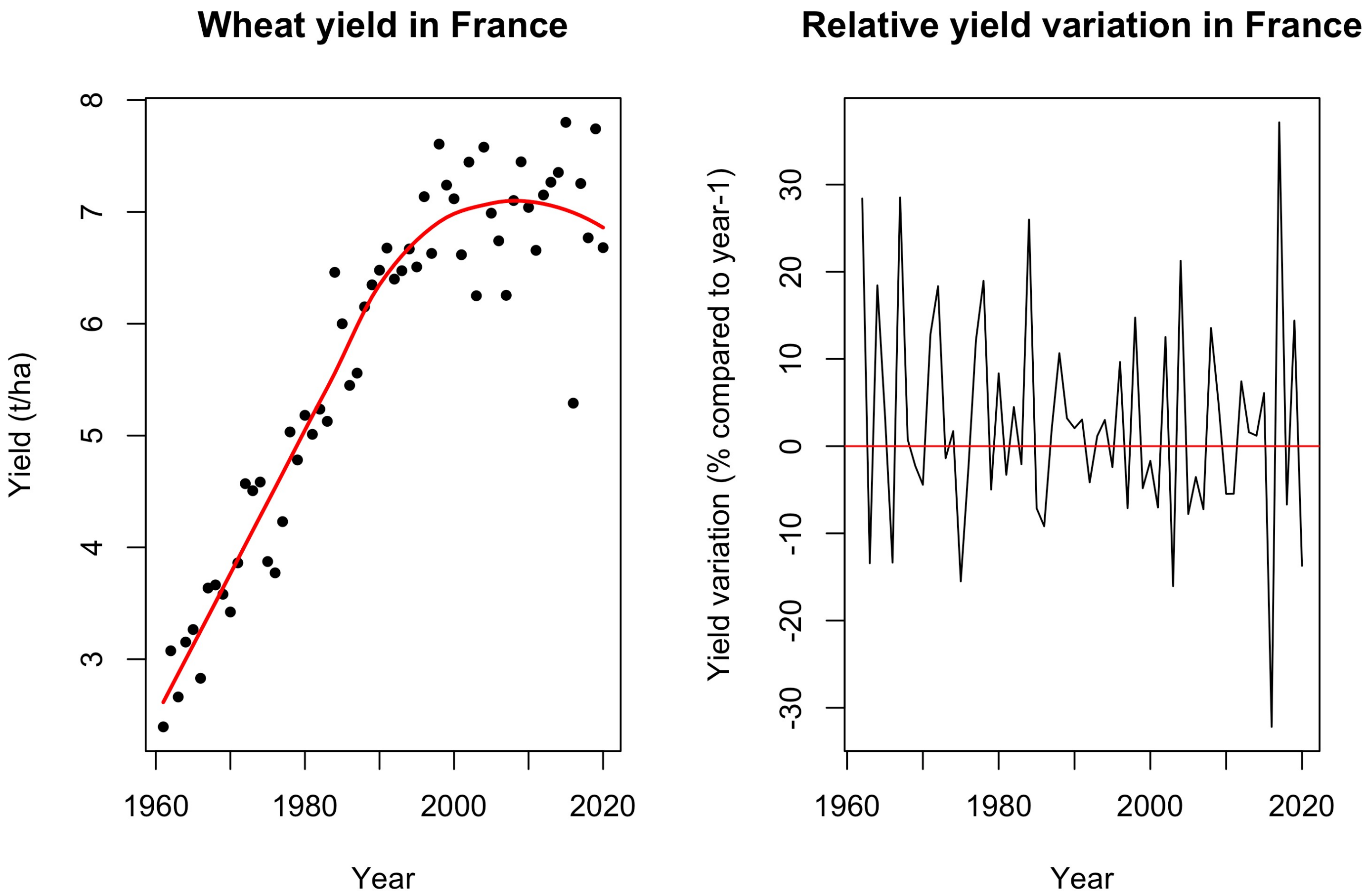


Yield losses and gains in 2018 in Europe

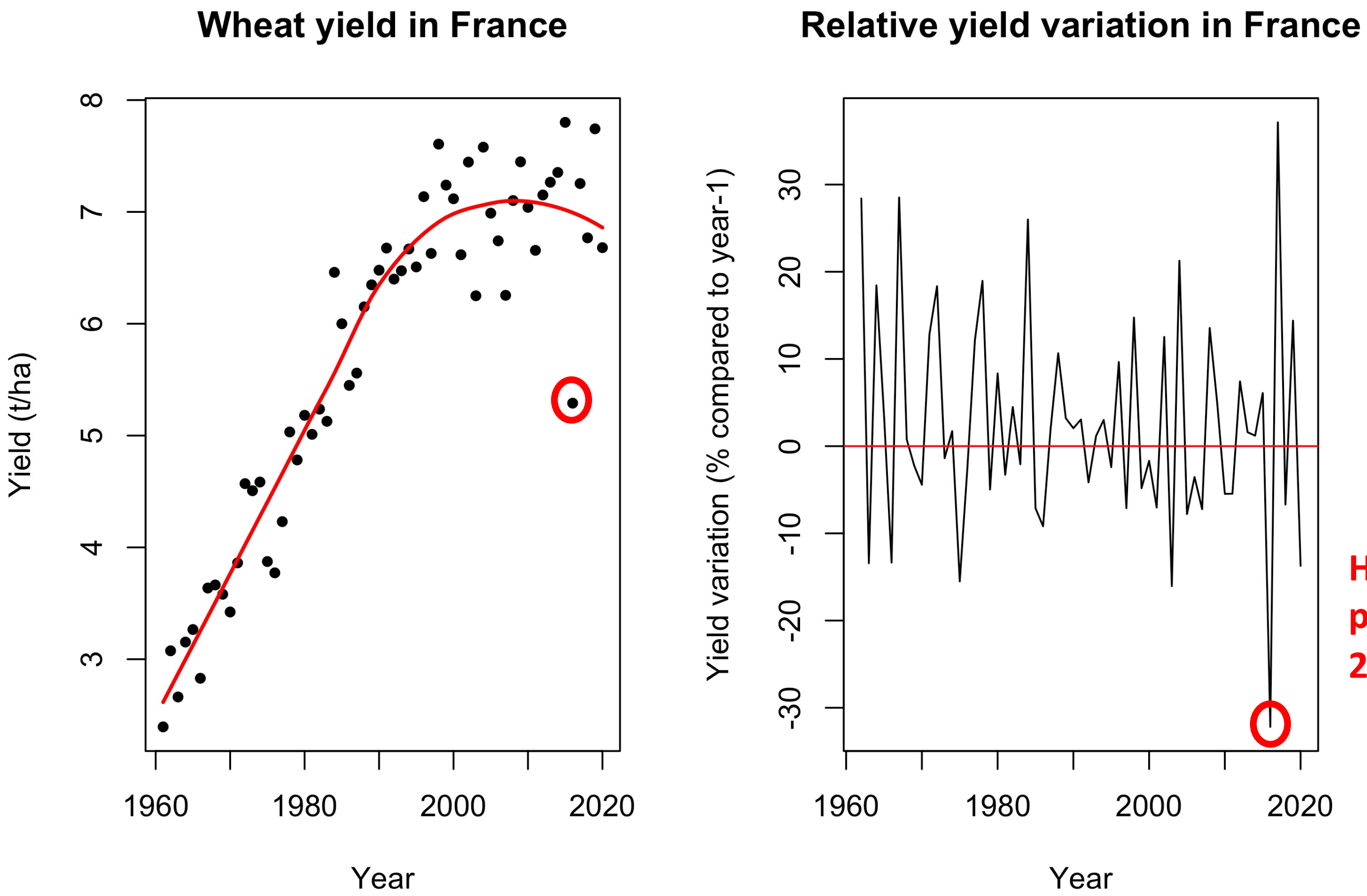
https://doi.org/10.1098/rstb.2019.0510



Crop yield variability

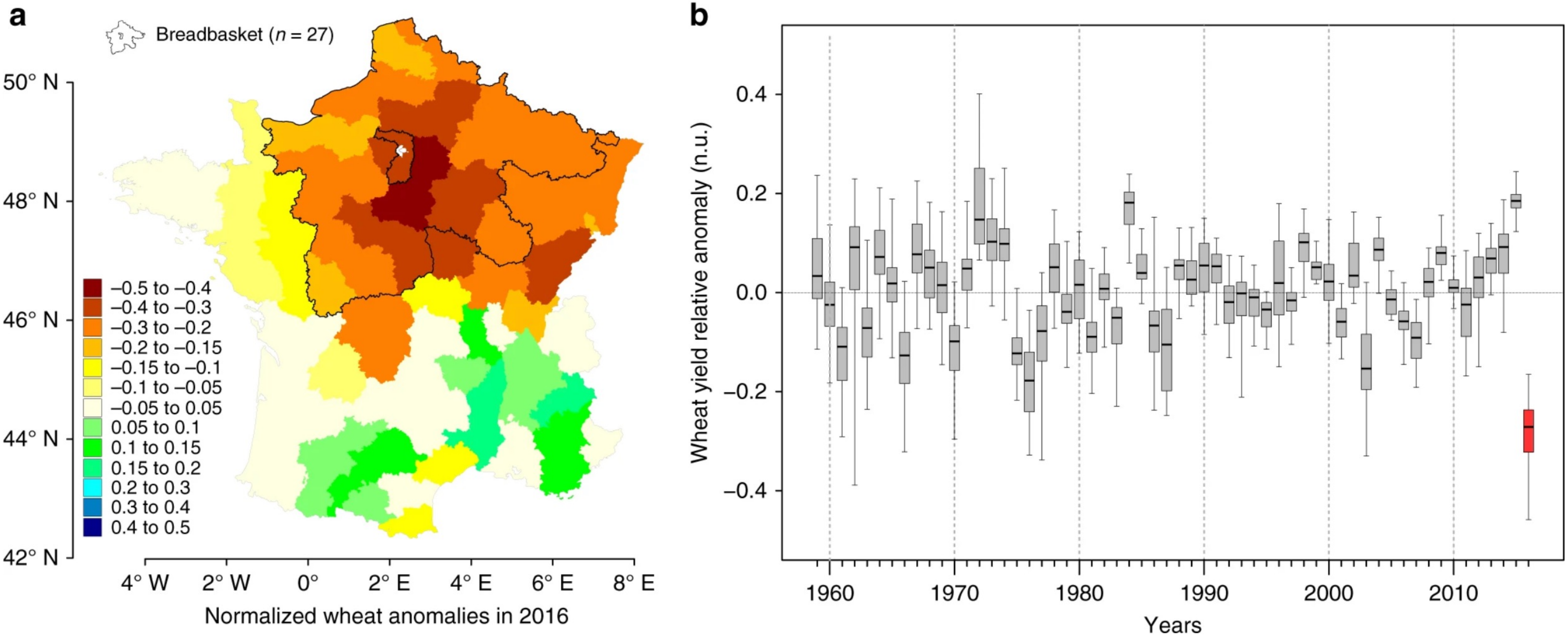


Crop yield variability



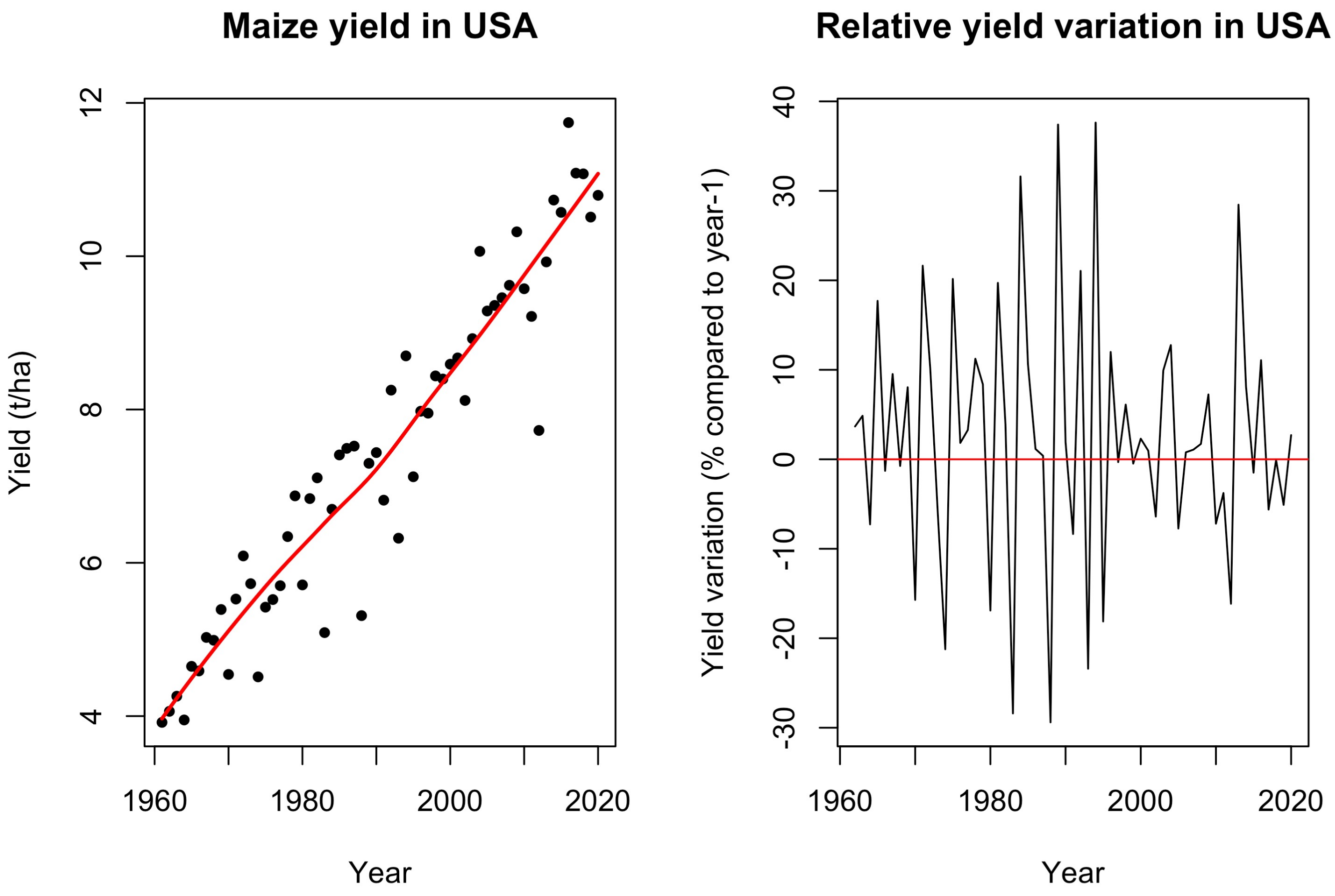
Heavy rainfall and plant diseases in 2016

From: Causes and implications of the unforeseen 2016 extreme yield loss in the breadbasket of France

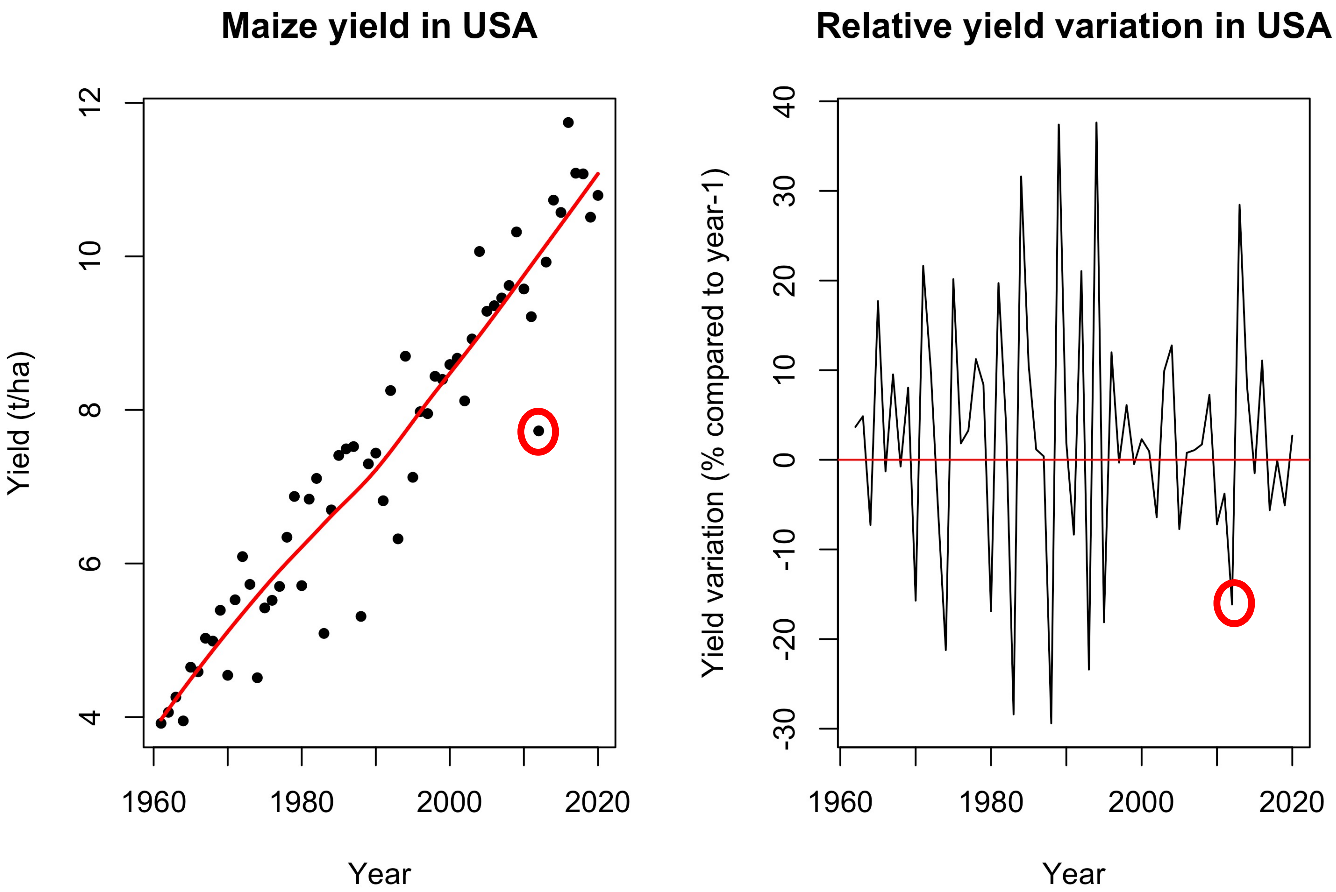


doi.org/10.1038/s41467-018-04087-x

Crop yield variability



Crop yield variability



Drought in 2012

Prices are impacted by yield shocks

Who predict crop yields?

- Private companies
- Public organizations

Issued: 26 April 2021

JRC MARS Bulletin Vol. 29 No 4

European Commission

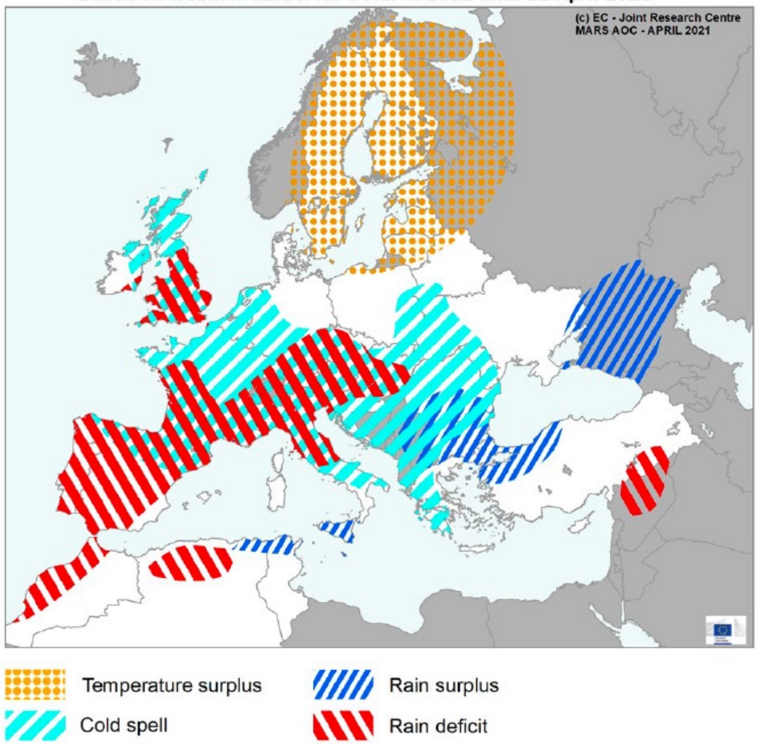
JRC MARS Bulletin
Crop monitoring in Europe
April 2021

Limited impacts of cold spells on annual crops

Crop	Yield t/ha				
	Avg 5yrs	March Bulletin	MARS 2021 forecasts	%21/5yrs	% Diff March
Total cereals	5.33	5.53	**5.52**	+ 3.6	-
Total wheat	5.47	5.67	**5.64**	+ 3.0	− 0.5

AREAS OF CONCERN - EXTREME WEATHER EVENTS

Based on weather data from 1 March 2021 until 21 April 2021



Outline

- Why many people are trying to predict crop yields?
- How to assess risk of crop yield loss?
- Why machine learning can be useful (in principle)?

Two main types of risk assessement

- Short-term predictions

  A few weeks or a few months before harvest

- Long-term analysis

  Yield projection several decades ahead

Two main types of risk assessement

- Short-term predictions

  A few weeks or a few months before harvest

- Long-term analysis

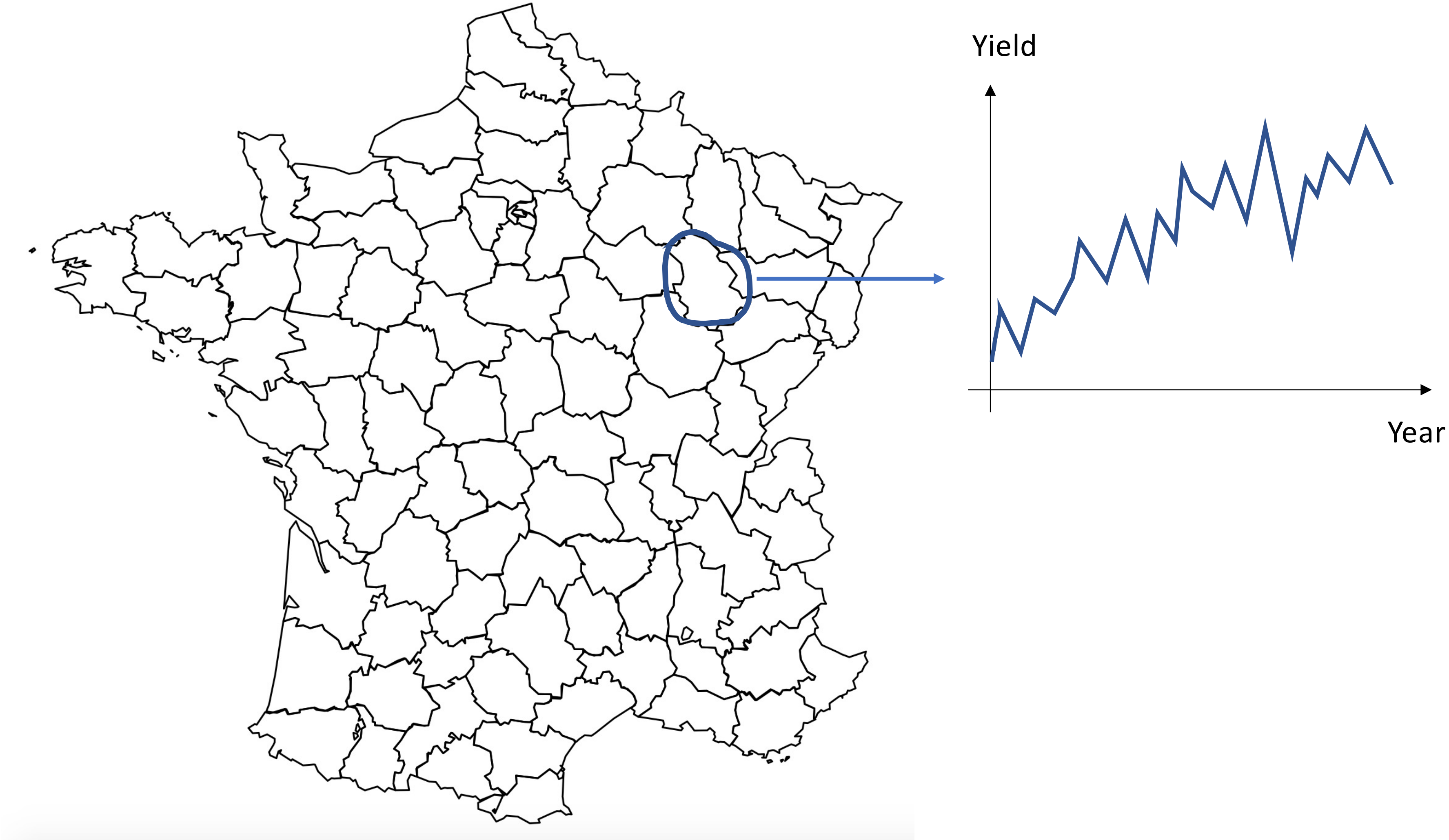
  Yield projection several decades ahead

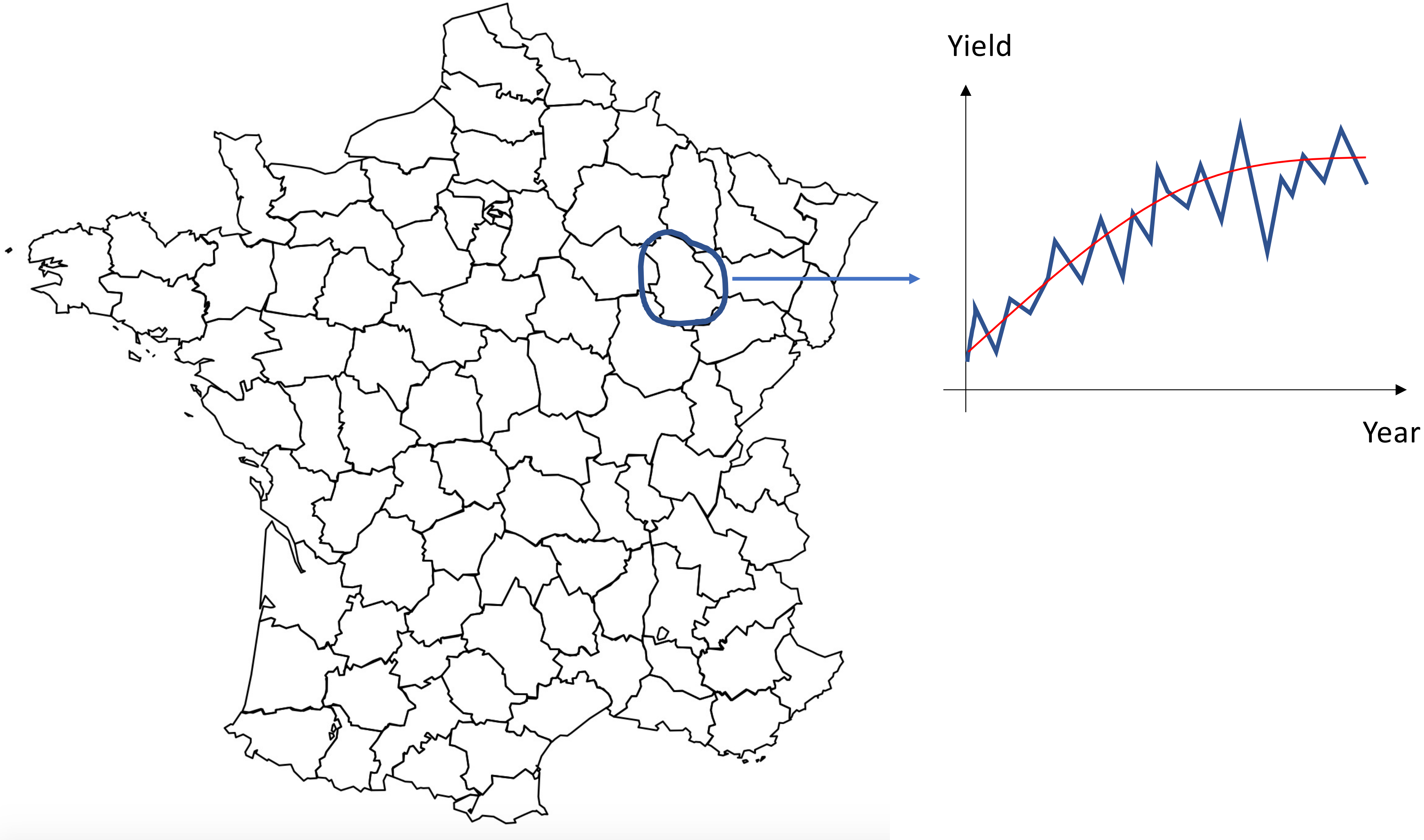
Sources of information

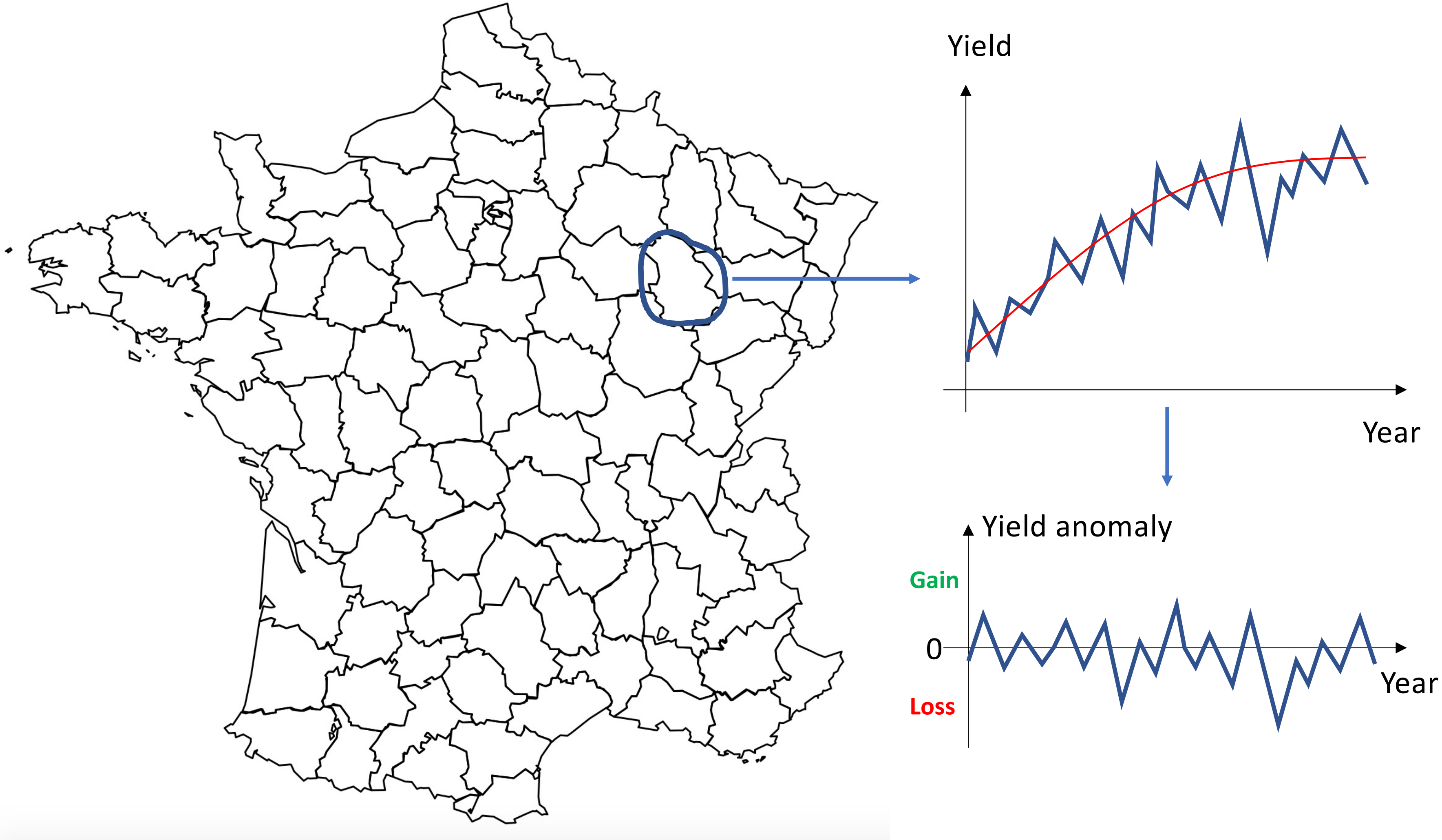
- Official yield statistics
- Expert knowledge
- Climate and remote sensing
- Field surveys and experiments
- Process-based dynamic crop models

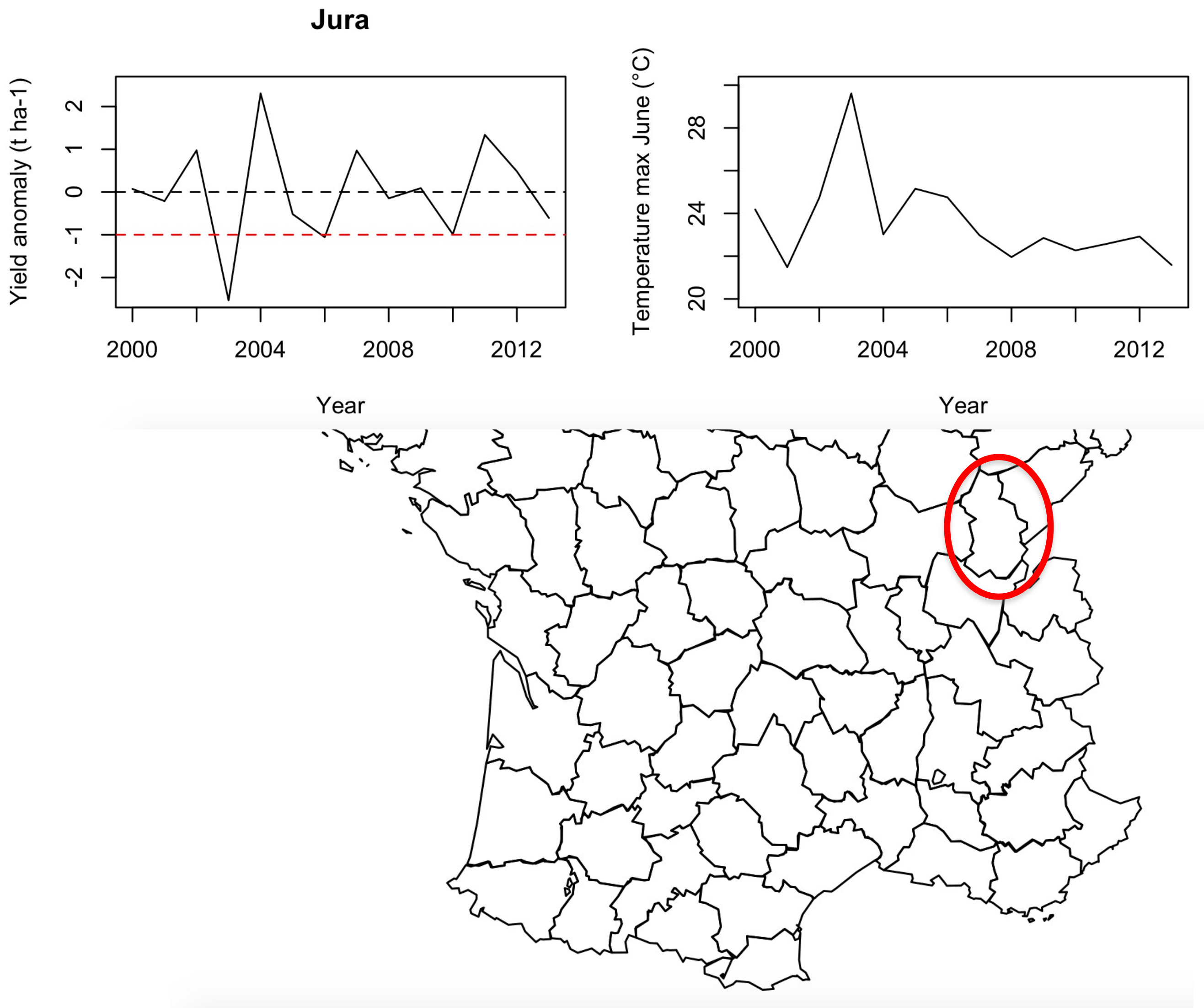
Sources of information

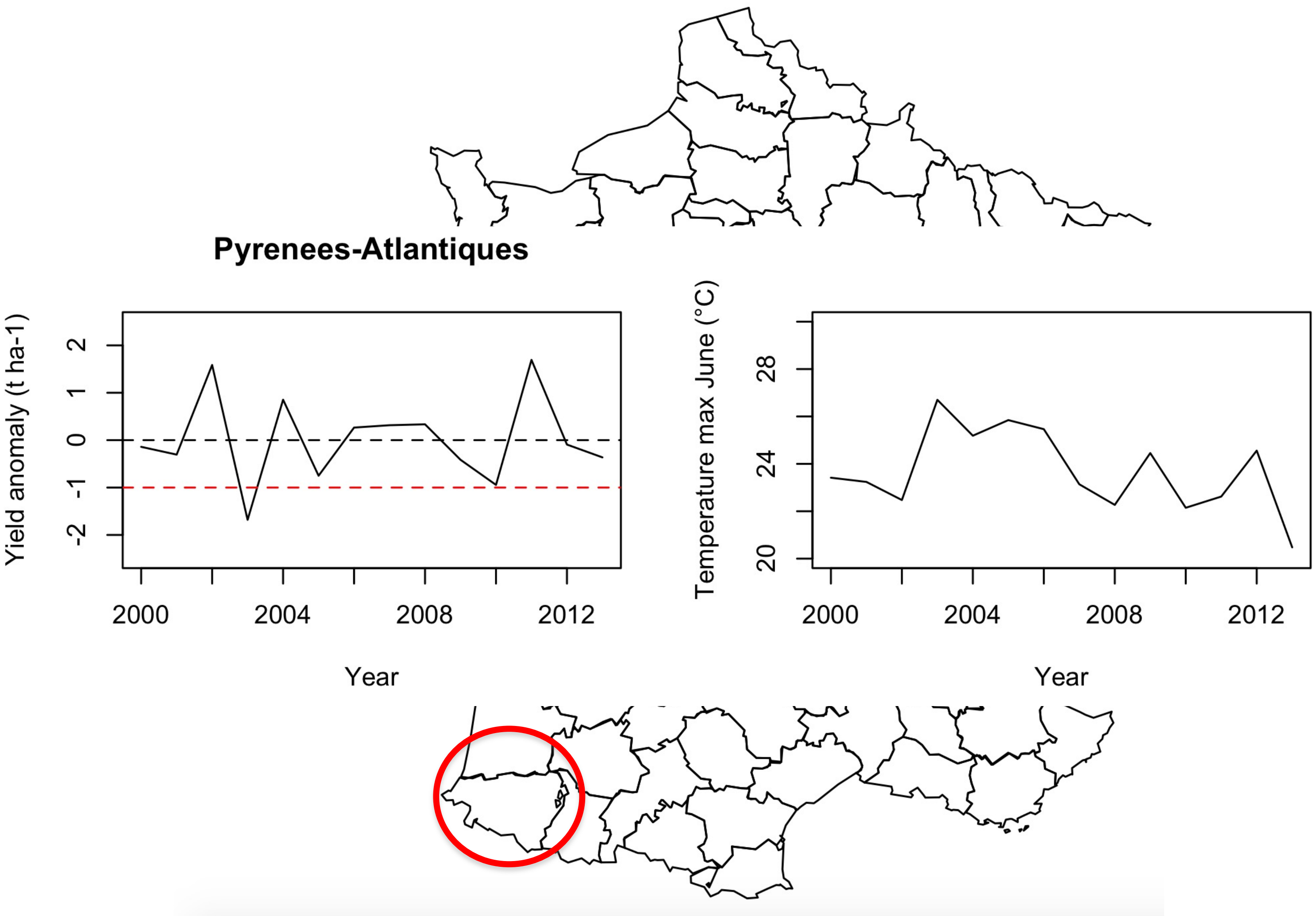
- Official yield statistics
- Expert knowledge
- Climate and remote sensing
- Field surveys and experiments
- Process-based dynamic crop models


Yield
Year


Yield
Year


Yield
Year
Yield anomaly
Gain
0
Year
Loss


Jura
Yield anomaly (t ha-1)
2
1
0
-1
-2
2000
2004
2008
2012
Year
Temperature max June (°C)
28
24
20
2000
2004
2008
2012
Year


Pyrenees-Atlantiques
Yield anomaly (t ha-1)
2
1
0
-1
-2
2000
2004
2008
2012
Year
Temperature max June (°C)
28
24
20
2000
2004
2008
2012
Year

Location	Year	Weather variable X1	Weather variable X2	…	Yield	Yield trend	Yield anomaly Y	Severe yield loss Z
1	1989	20.5	12.0		5.2	5	+0.2	No
1	1990	28.1	-1.2		3.1	5.5	-2.4	Yes
…	…	…	…		…	…	…	…
2	1989	22.7	21.4		5.4	5.8	-0.4	No
2	1990	24.8	9.7		6.9	6.1	+0.8	No
…	…	…	…		…	…	…	…

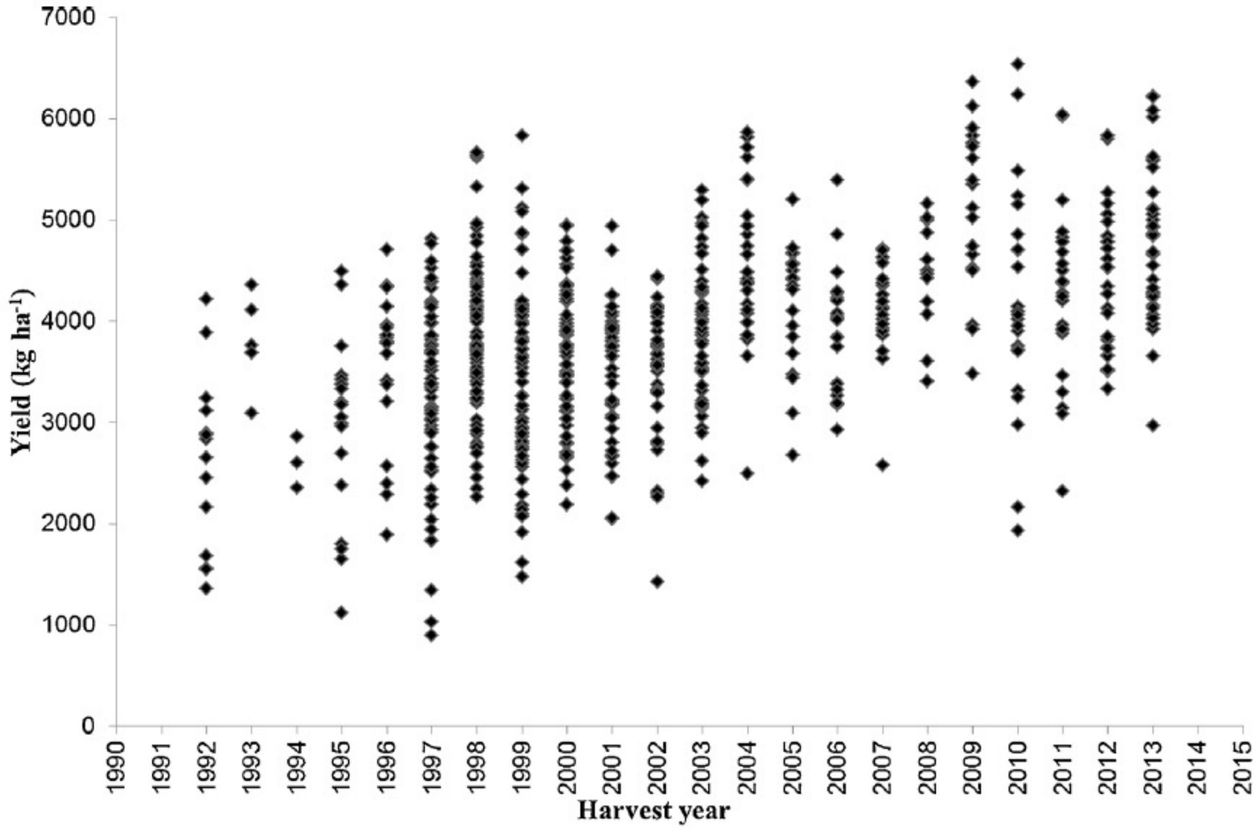
Location	Year	Weather variable X1	Weather variable X2	…	Yield	Yield trend	Yield anomaly Y	Severe yield loss Z
1	1989	20.5	12.0		5.2	5	+0.2	No
1	1990	28.1	-1.2		3.1	5.5	-2.4	Yes
…	…	…	…		…	…	…	…
2	1989	22.7	21.4		5.4	5.8	-0.4	No
2	1990	24.8	9.7		6.9	6.1	+0.8	No
…	…	…	…		…	…	…	…

$$Y = X\theta + \varepsilon$$

Yield anomaly

Weather & soil variables

Prediction of oilseed rape yields in Denmark



Sharif et al. (2017) https://doi.org/10.1016/j.eja.2016.09.015

Prediction of oilseed rape yields in Denmark

$$\log\left(Yield_j\right) = b_0 + b_1 \times YEAR_j + \sum_{i=1}^{n} b_{2i} \times TEMP_{ij} + \sum_{i=1}^{n} b_{3i}$$

$$\times RAD_{ij} + \sum_{i=1}^{n} b_{4i} \times PREC_{ij} + \sum_{i=1}^{n} b_{5i} \times TEMP_{ij}{}^2 + \sum_{i=1}^{n} b_{6i}$$

$$\times RAD_{ij}{}^2 + \sum_{i=1}^{n} b_{7i} \times PREC_{ij}{}^2 + b_8 \times SOIL_j + b_9 \times Pre_{CROP\,j}$$

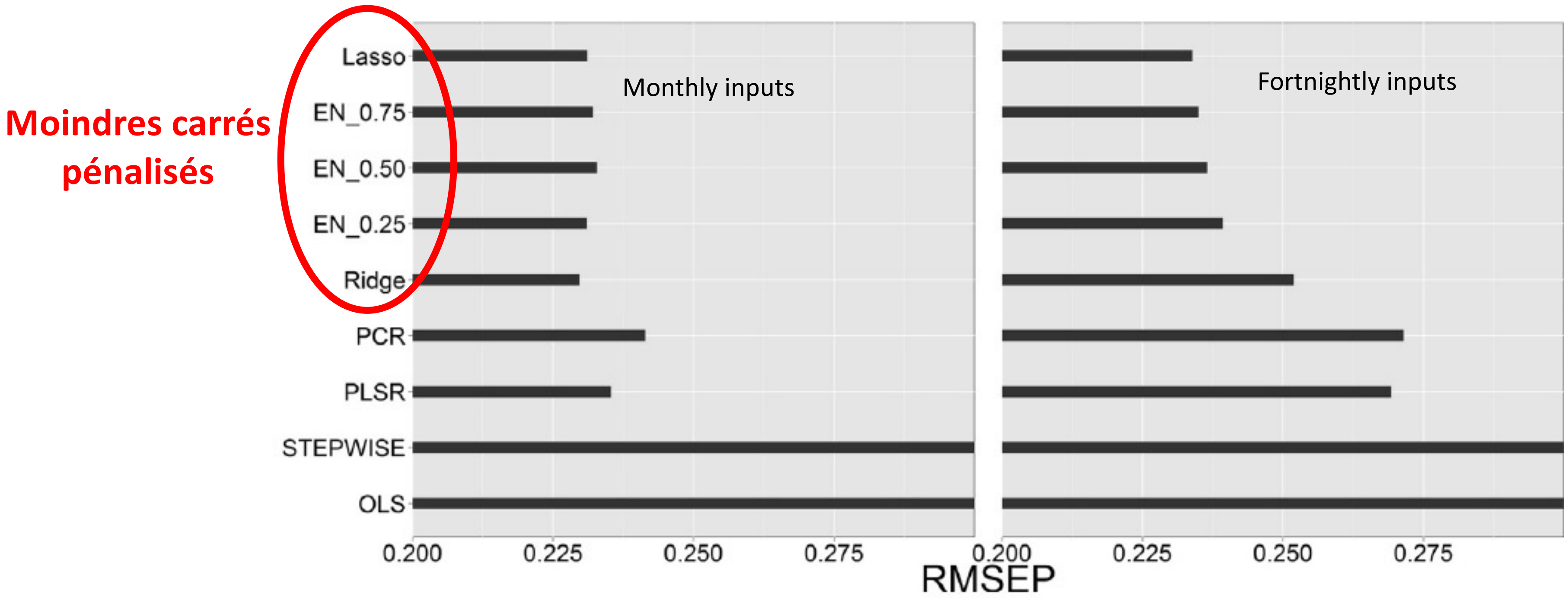
$$+ b_{10} \times Sowing_j + b_{11} \times Sowing_j{}^2 + \varepsilon_j$$

Sharif et al. (2017) https://doi.org/10.1016/j.eja.2016.09.015

Penalized regression (LASSO)

$$\sum_{i=1}^{p}\left(y_i - b_0 - \sum_{j=1}^{q} b_j \,.x_{ij}\right)^2 + \lambda \sum_{j=1}^{q} \left|b_j\right|$$

Prediction of oilseed rape yields in Denmark



Sharif et al. (2017) https://doi.org/10.1016/j.eja.2016.09.015

Locati on	Year	Weather variable X1	Weather variable X2	…	Yield	Yield trend	Yield anomaly Y	Severe yield loss Z
1	1989	20.5	12.0		5.2	5	+0.2	No
1	1990	28.1	-1.2		3.1	5.5	-2.4	Yes
…	…	…	…		…	…	…	…
2	1989	22.7	21.4		5.4	5.8	-0.4	No
2	1990	24.8	9.7		6.9	6.1	+0.8	No
…	…	…	…		…	…	…	…

Locati on	Year	Weather variable X1	Weather variable X2	…	Yield	Yield trend	Yield anomaly Y	Severe yield loss Z
1	1989	20.5	12.0		5.2	5	+0.2	No
1	1990	28.1	-1.2		3.1	5.5	-2.4	Yes
…	…	…	…		…	…	…	…
2	1989	22.7	21.4		5.4	5.8	-0.4	No
2	1990	24.8	9.7		6.9	6.1	+0.8	No
…	…	…	…		…	…	…	…

Binary classification rule based on one single index:

If $X_1 > T$, Z=« yes »
otherwise, Z=« no »

Location	Year	Weather variable X1	Weather variable X2	…	Yield	Yield trend	Yield anomaly Y	Severe yield loss Z
1	1989	20.5	12.0		5.2	5	+0.2	No
1	1990	28.1	-1.2		3.1	5.5	-2.4	Yes
…	…	…	…		…	…	…	…
2	1989	22.7	21.4		5.4	5.8	-0.4	No
2	1990	24.8	9.7		6.9	6.1	+0.8	No
…	…	…	…		…	…	…	…

Classification based on combinations of several variables

Binary variable

$$Z \sim Bern(\pi)$$

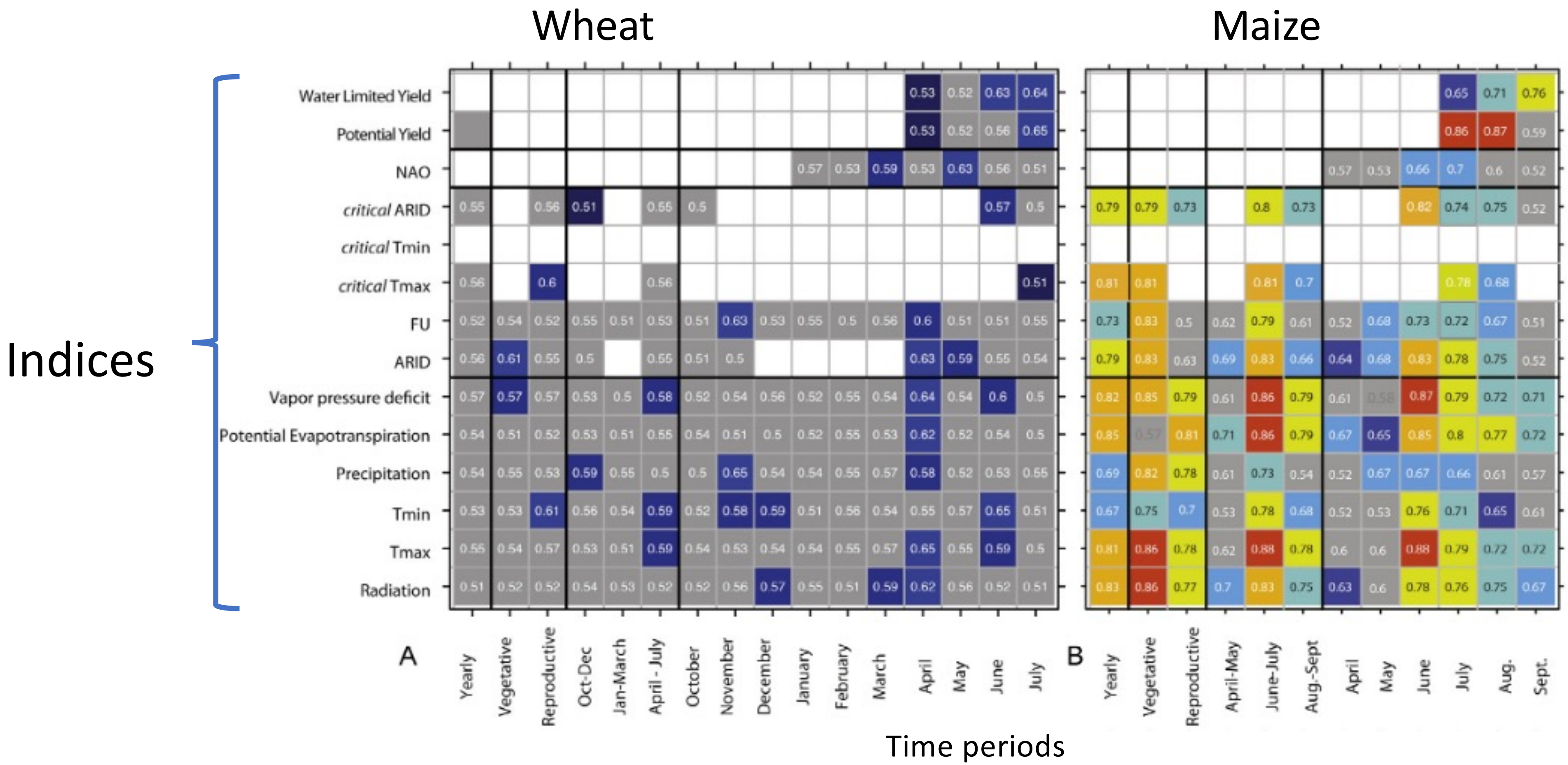
$$logit(\pi) = X\theta$$

Prob. of severe yield loss

Weather & soil variables

Accuracy of classification rules (AUC)

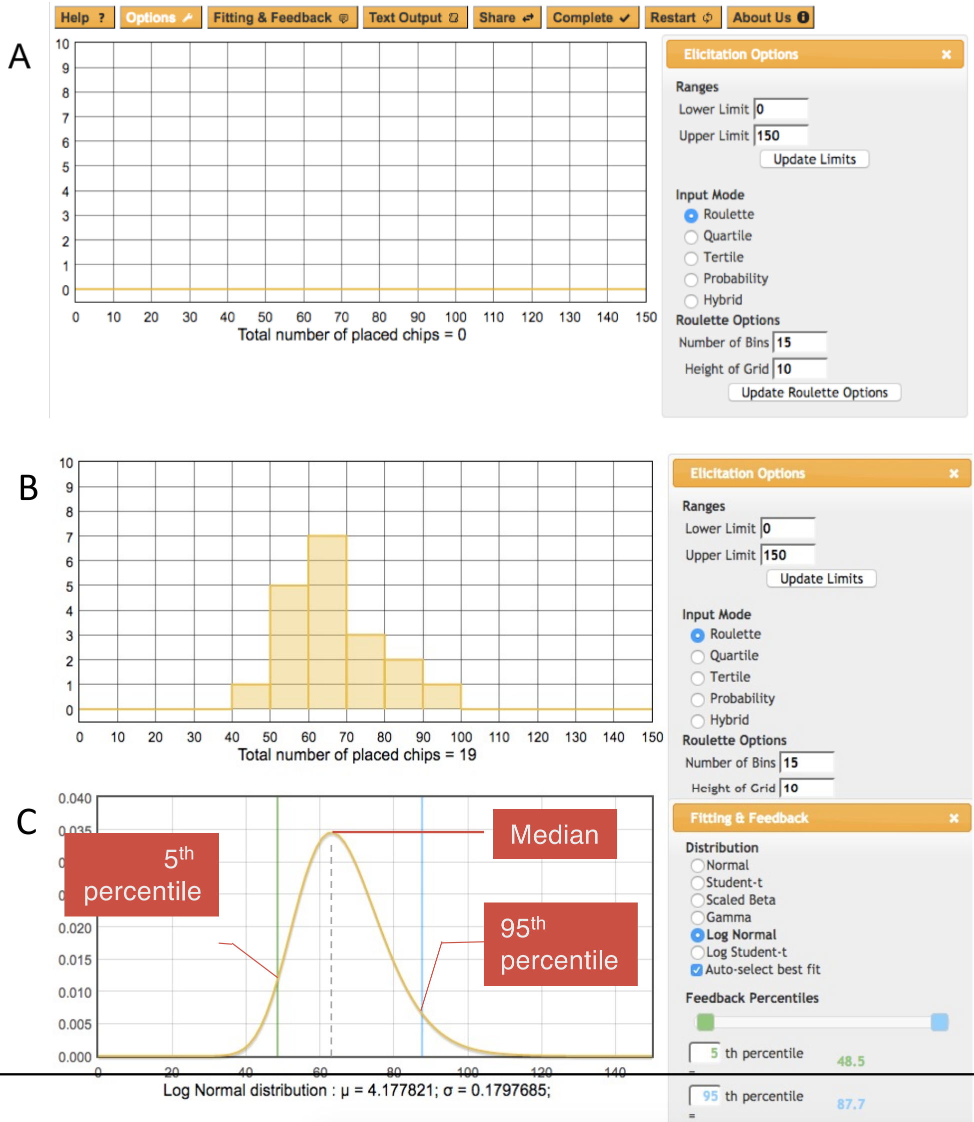
https://doi.org/10.1016/j.agrformet.2016.01.009



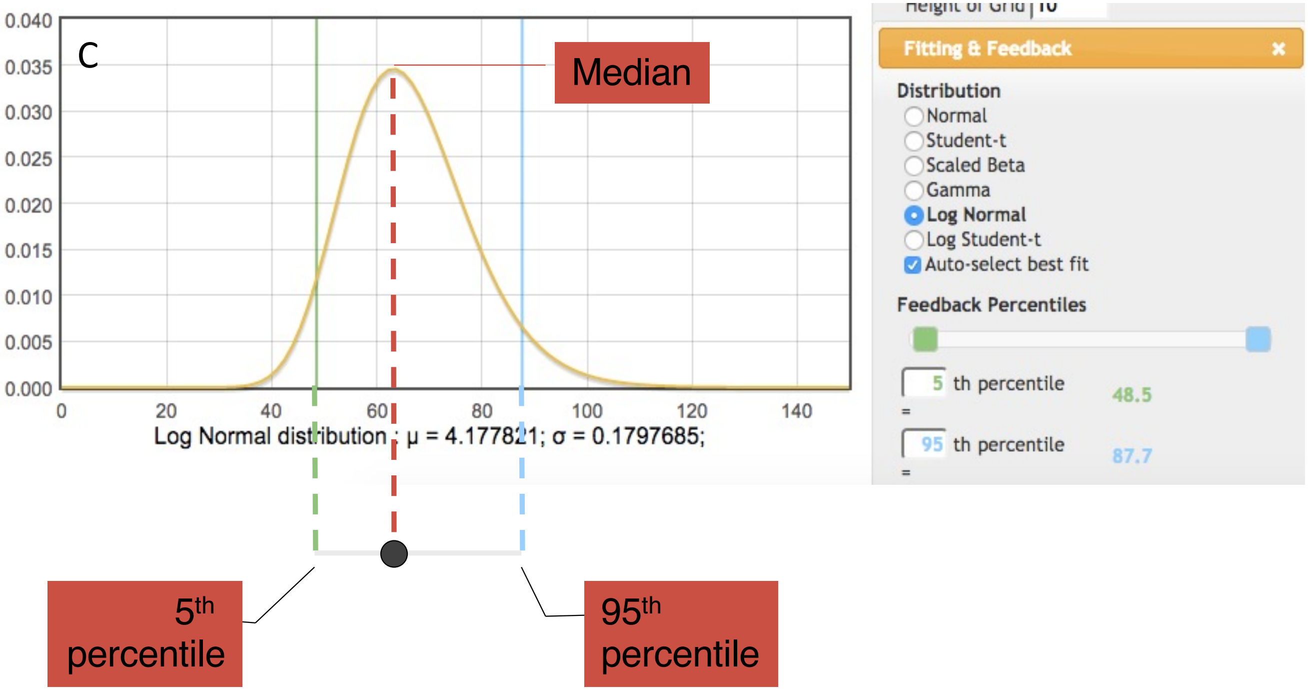
Sources of information

- Expert knowledge
- Climate and remote sensing
- Field surveys and experiments
- Process-based dynamic crop models

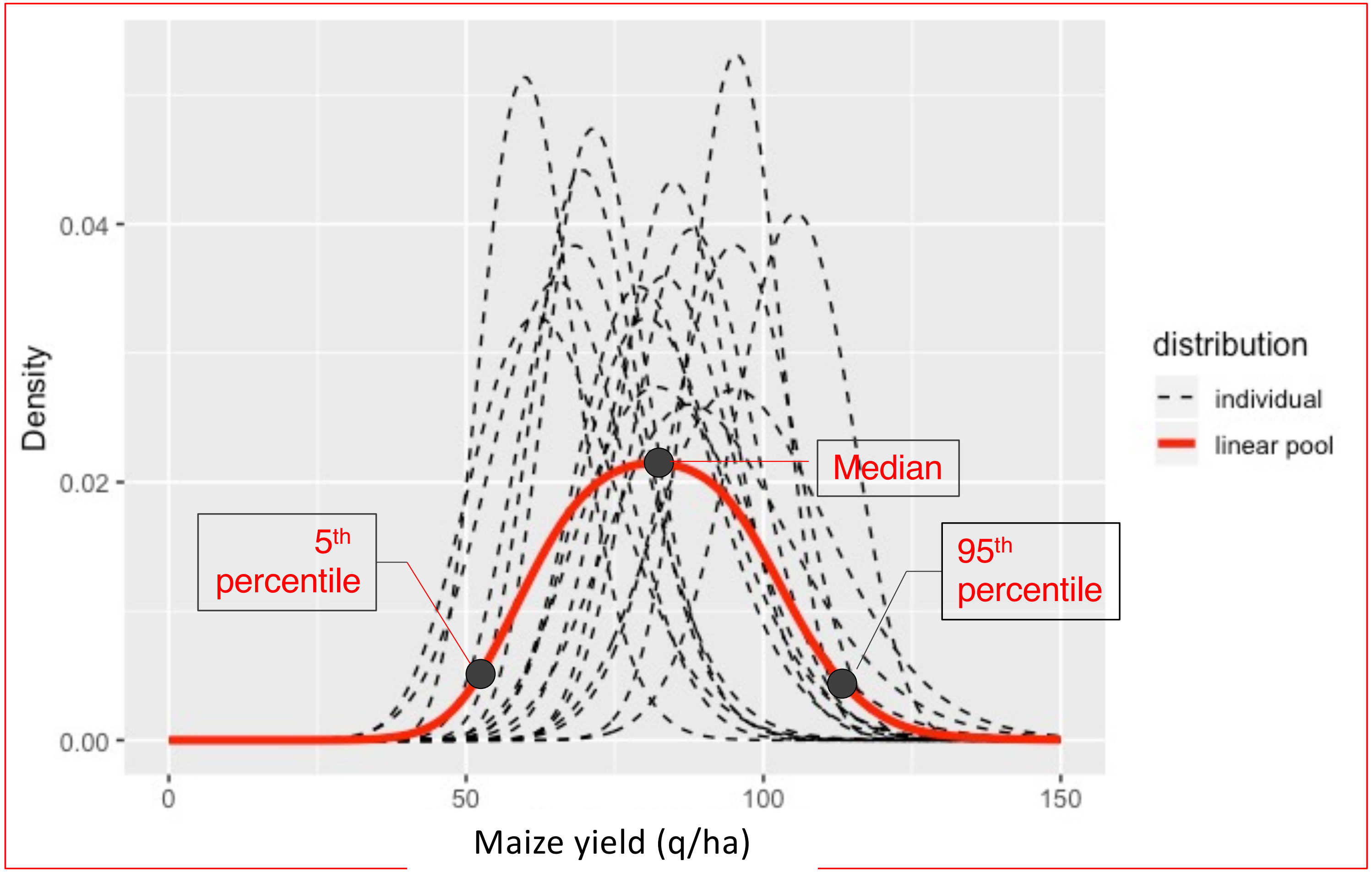
Match Tool ® : A tool for probabilistic expert elicitation



Individual elicitation

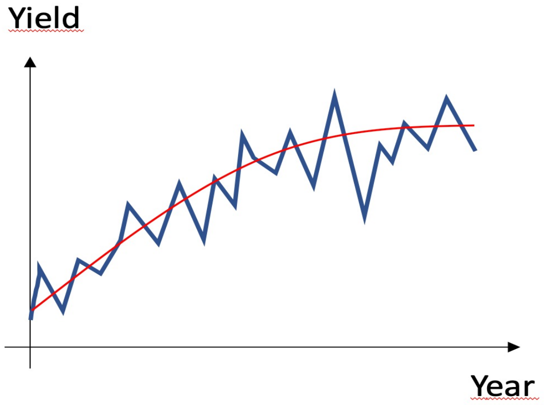


Elicitation of group of experts: https://licite.fr/licite/



Sources of information

- Expert knowledge
- Climate and remote sensing
- Field surveys and experiments
- Process-based dynamic crop models



SATELLITE

Vegetation indices

$$Y = X\theta + \varepsilon$$

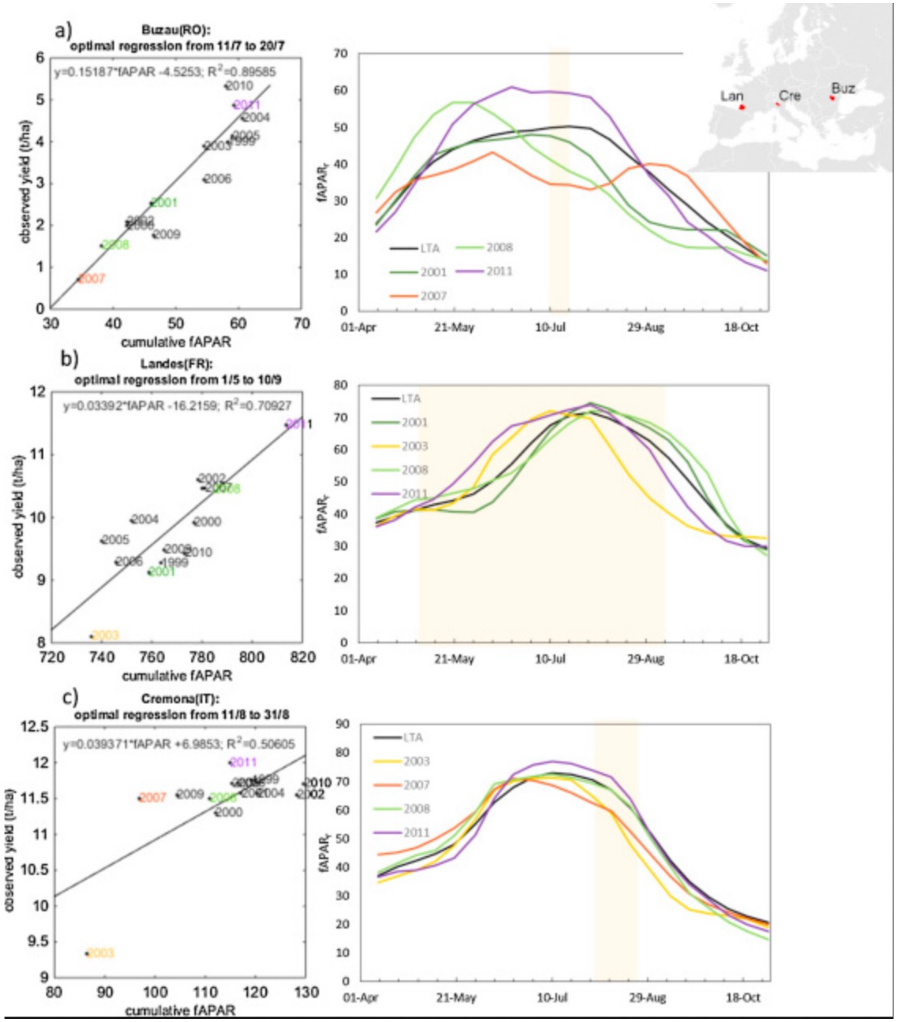
Yield forecast in new sites and/or years
after model fitting

Maize yield

Site 1

Site 2

Site 3



Index fAPAR

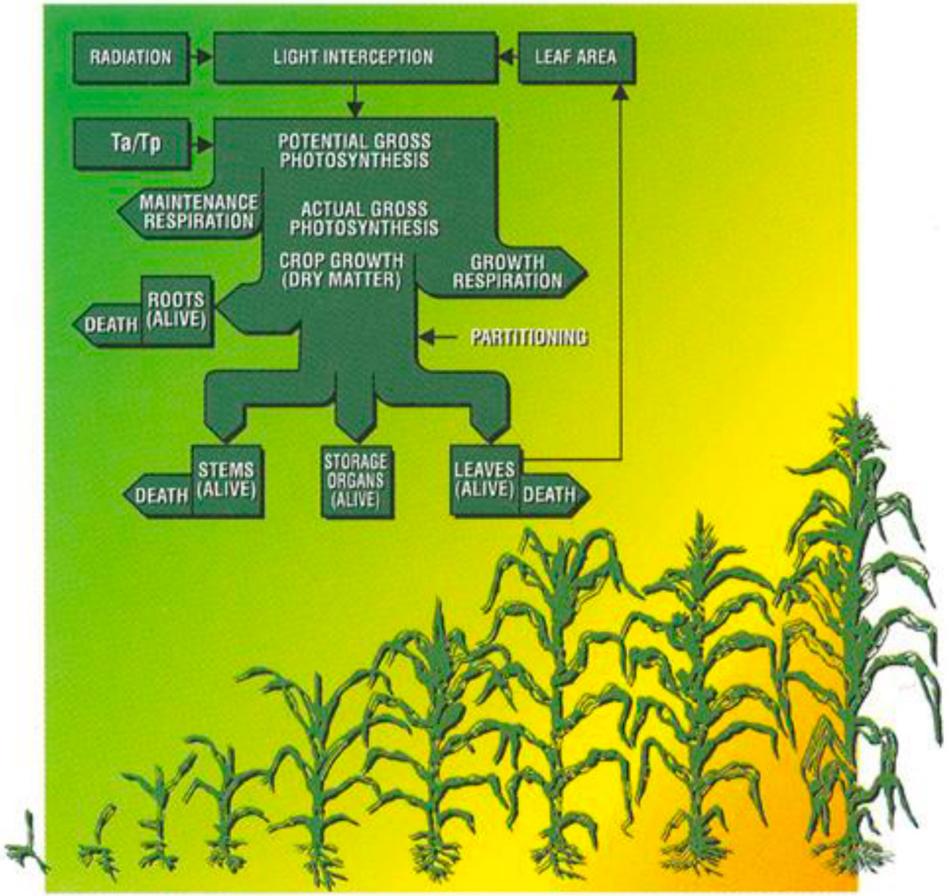
Lopez-Lozano et al. 2015
https://doi.org/10.1016/j.agrformet.2015.02.021

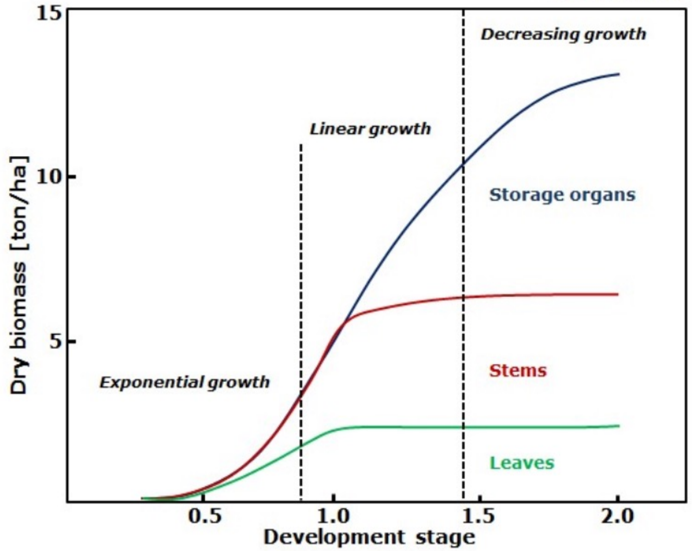
Sources of information

- Expert knowledge
- Climate and remote sensing
- Field surveys and experiments
- Process-based dynamic crop models

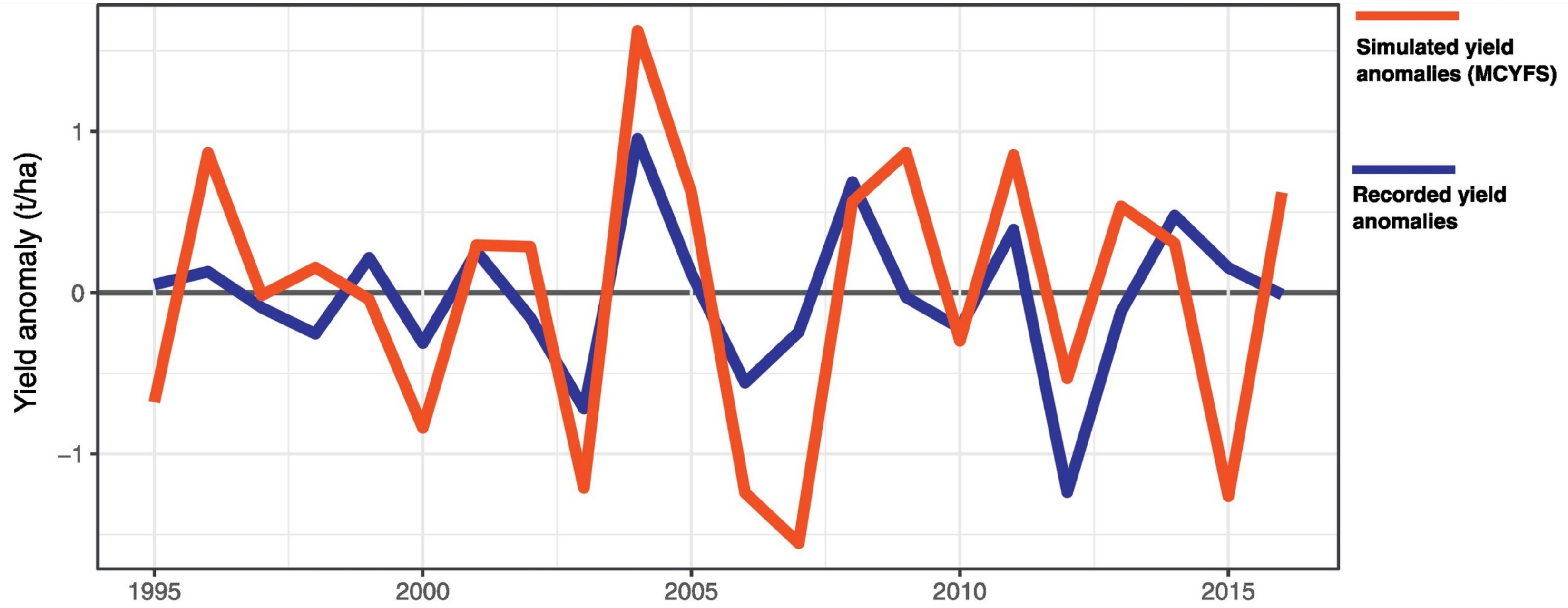
WOFOST Control Centre 2.1 and WOFOST 7.1.7

WOFOST: process-based model used for crop yield forecast in Europe





Wheat yield in Czech Republic



https://doi.org/10.1016/j.agsy.2018.05.002

Ceglar et al. 2019

Outline

- Why many people are trying to predict crop yields?
- How to assess risk of crop yield loss?
- Why machine learning can be useful (in principle)?

Advantages of machine learning

- Flexible and data-driven -> low bias
- Take nonlinear responses and interactions into account
- Can combine different types of information
  - ✓ Climate inputs,
  - ✓ Soil characteristics
  - ✓ Cropping practices
  - ✓ Remote sensing data
  - ✓ Process-based model simulations
- Can be trained from official yield statistics
- Accuracy can be easily evaluated

Example: Two-month ahead yield forecasts in Europe

Field Crops Research 276 (2022) 108377

Contents lists available at ScienceDirect

Field Crops Research

journal homepage: www.elsevier.com/locate/fcr

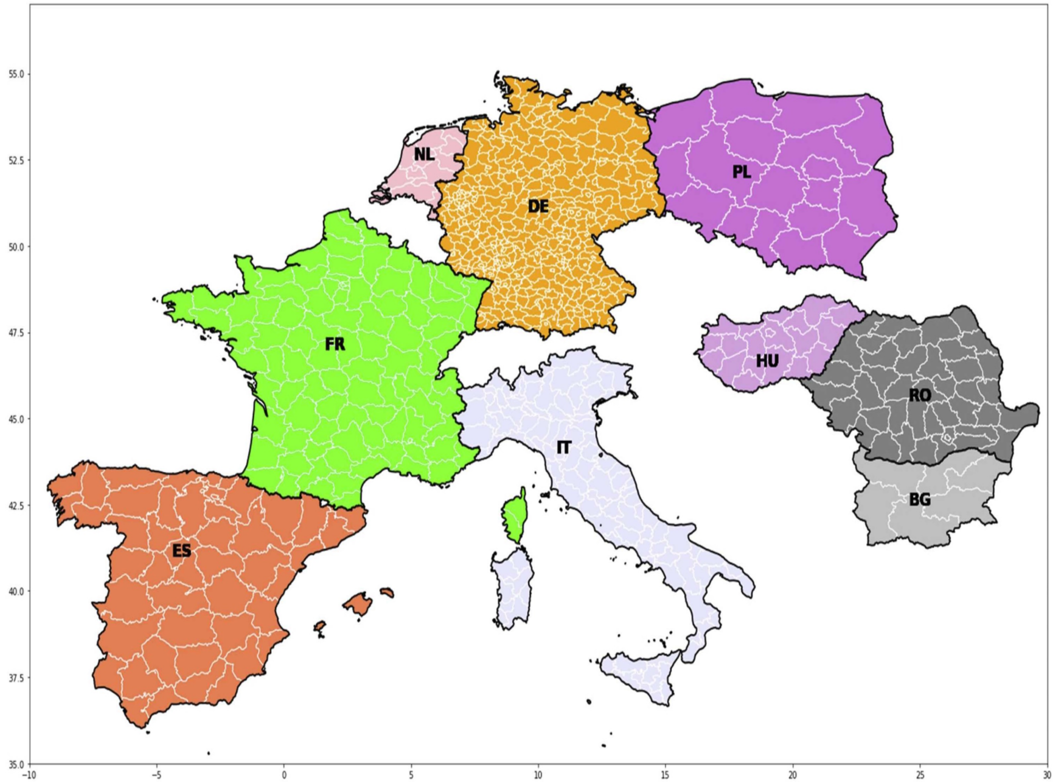
Machine learning for regional crop yield forecasting in Europe

Dilli Paudel [a,*,1], Hendrik Boogaard [b], Allard de Wit [b], Marijn van der Velde [d], Martin Claverie [d], Luigi Nisini [d], Sander Janssen [b], Sjoukje Osinga [a], Ioannis N. Athanasiadis [c]

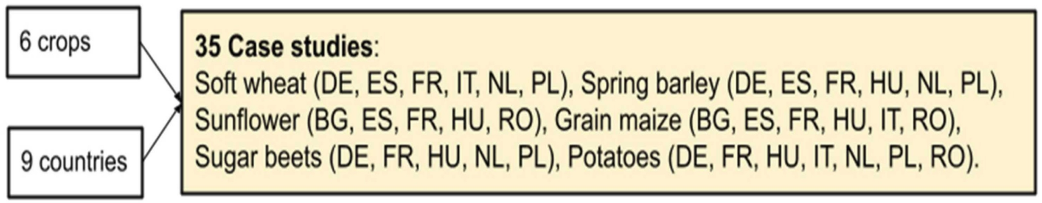
Algorithms tested

- Ridge
- SVM
- K-nearest neighbors regression
- Gradient boosting decision trees

Paudel et al. 2022
/doi.org/10.1016/j.fcr.2021.108377



(a)



(b)

- Four categories of features
  - ✓Process-based model simulation outputs,
  - ✓Weather observations,
  - ✓Remote sensing indicators,
  - ✓Soil water holding capacity.

- Output
  - ✓Yield

Paudel et al. 2022
/doi.org/10.1016/j.fcr.2021.108377

Features

A.4.1 WOFOST Indicators

- POT_YB (potential dry weight biomass) (kg ha^{-1})
- POT_YS (potential dry weight storage organs) (kg ha^{-1})
- WLIM_YB (water-limited dry weight biomass) (kg ha^{-1})
- WLIM_YS (water-limited dry weight storage organs) (kg ha^{-1})
- PLAI (potential leaf area divided by surface area) (m^2 m^{-2})
- WLAI (water-limited leaf area divided by surface area) (m^2 m^{-2})
- DVS (development stage (0-200))
- RSM (percentage of soil water holding capacity)
- TWC (sum of water limited transpiration (cm))
- TWR (sum of potential transpiration (cm))

A.4.2 Meteo Indicators

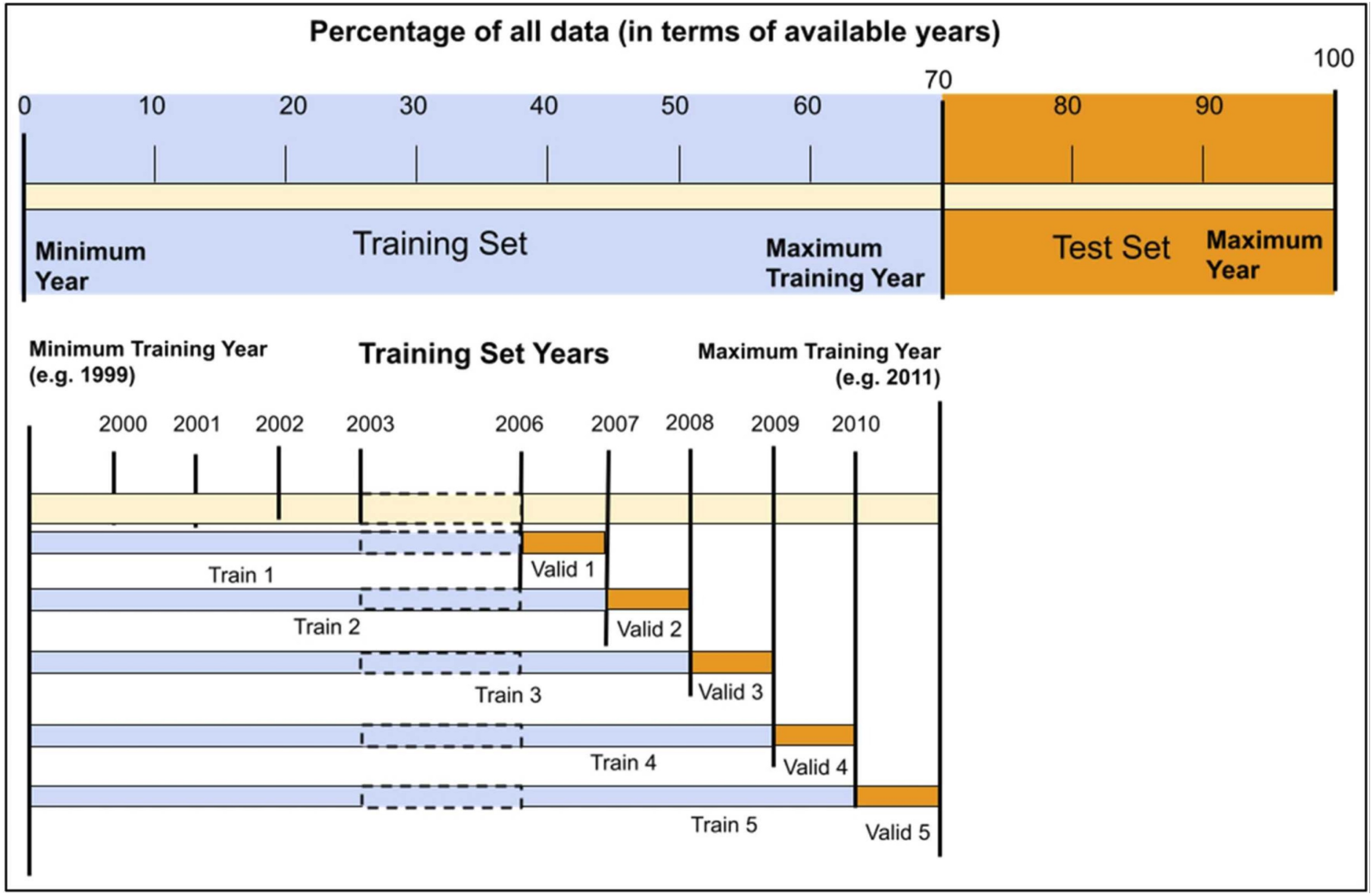
- TMAX (maximum daily air temperature) (℃)
- TMIN (minimum daily air temperature) (℃)
- TAVG (average daily air temperature (℃))
- VPRES (average daily vapour pressure (hPa))
- WSPD (average daily wind speed at 10 m (m s^{-1}))
- PREC (sum of daily precipitation (mm))
- ET0 (sum of daily evapotranspiration of short vegetation (Penman-Monteith, Allen et al., 1998) (mm))
- RAD (sum of daily global incoming shortwave radiation (KJ m^{-2} d^{-1}))
- RELH (average daily relative humidity (%))
- CWB (climate water balance, calculated as PREC - ET0)

A.4.3 Remote Sensing Indicators

FAPAR (Fraction of Absorbed Photosynthetically Active Radiation (Smoothed))

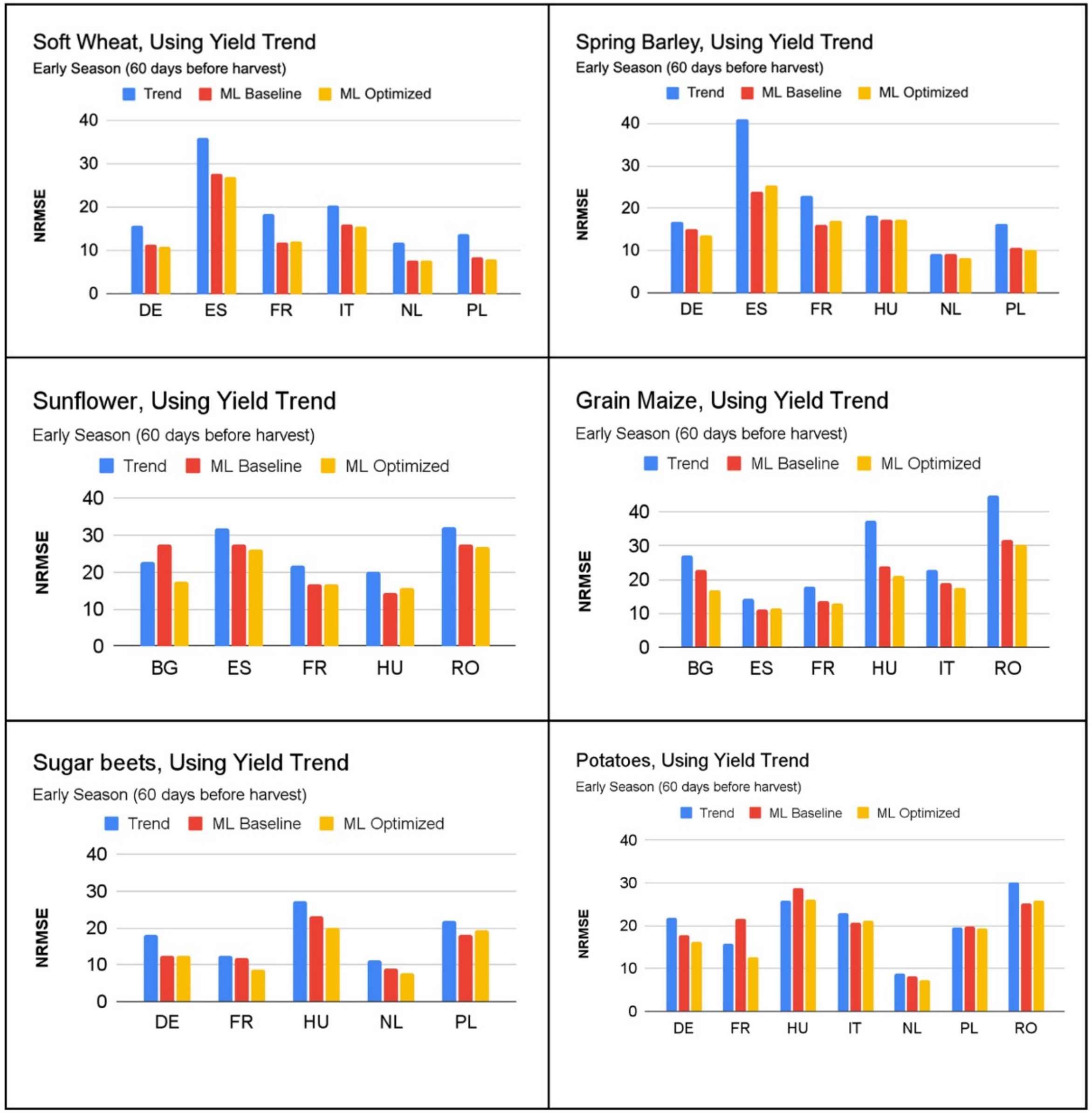
A.4.4 Soil Moisture Indicators

- SM_WP (wilting point)
- SM_FC (field capacity)
- SM_SAT (saturation)
- DEPTH (rooting depth)
- SM_WHC (water holding capacity)

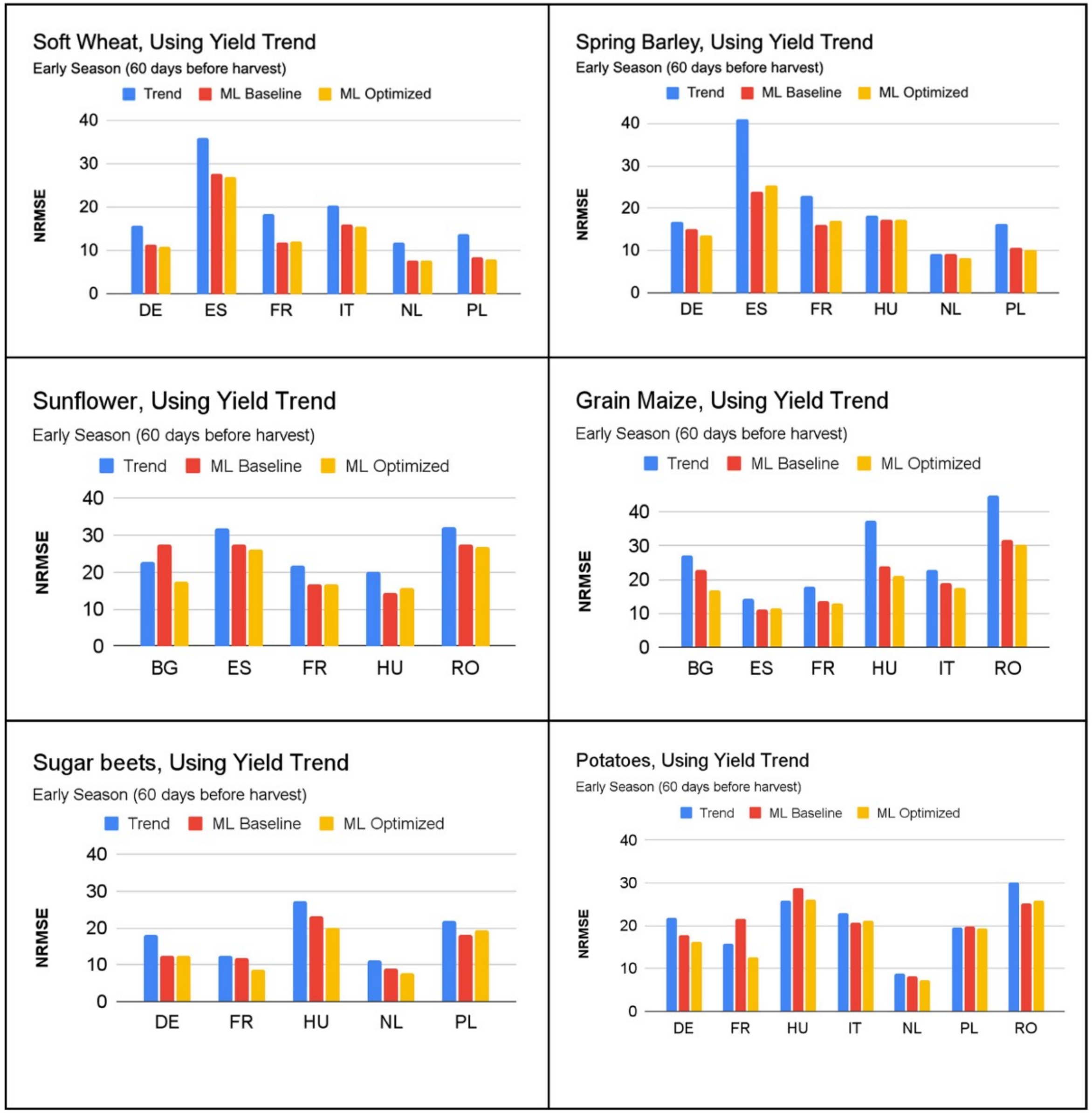


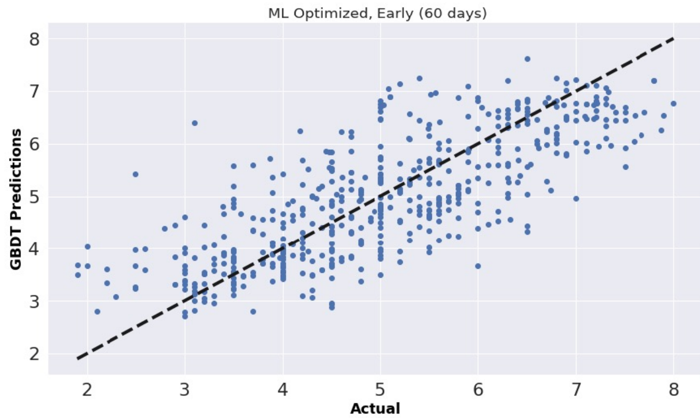
Features selection & Hyperparameter optimisation Test

Paudel et al. 2022
/doi.org/10.1016/j.fcr.2021.108377

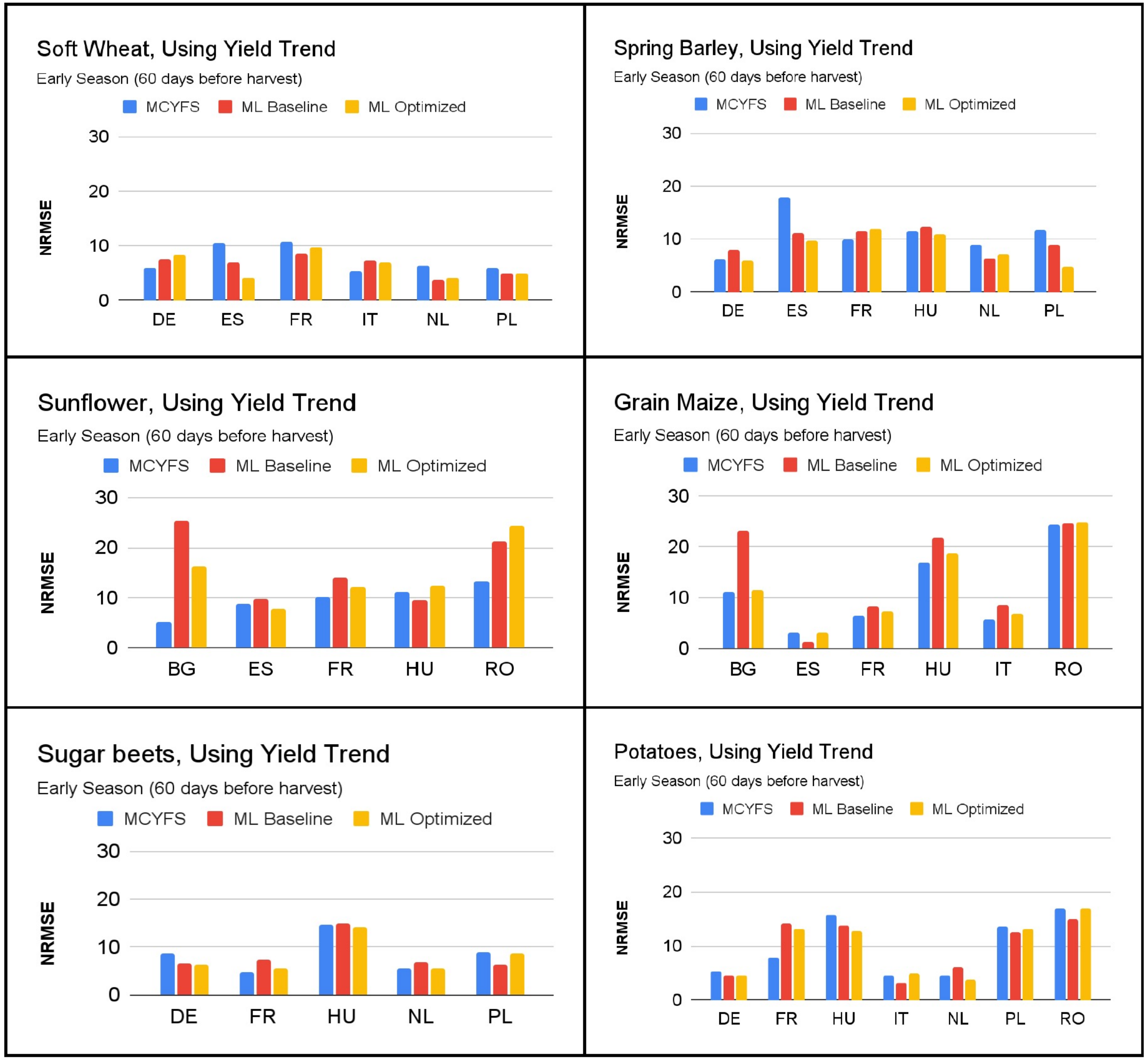


Paudel et al. 2022
/doi.org/10.1016/j.fcr.2021.108377



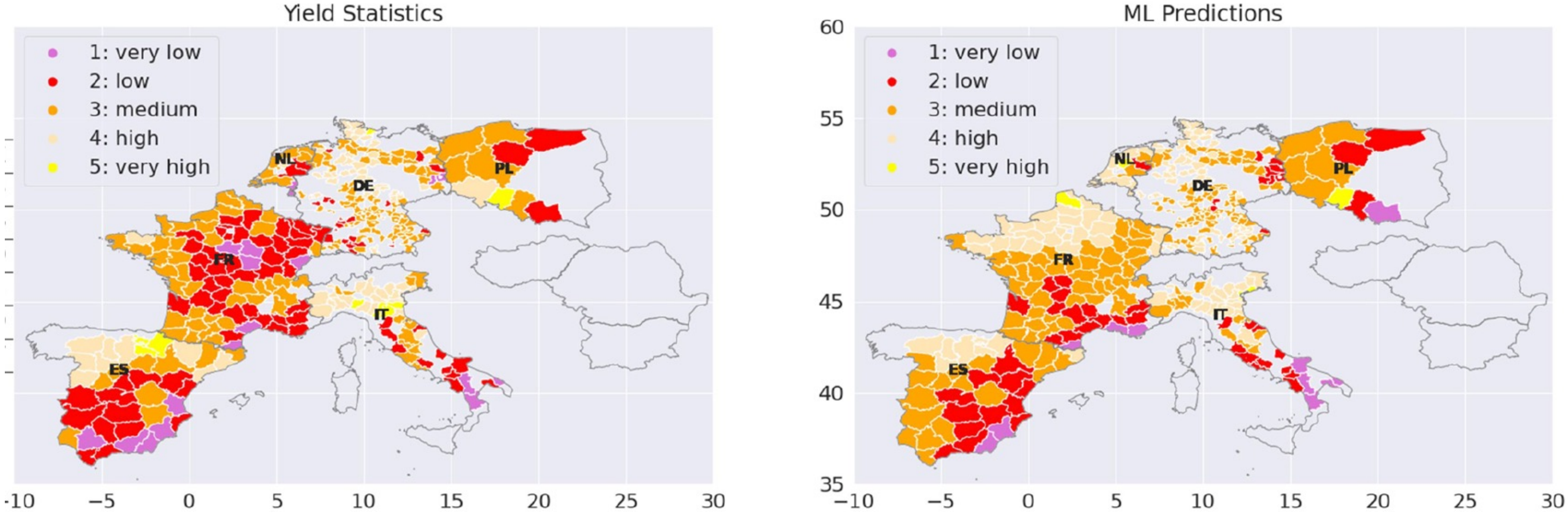


Paudel et al. 2022
/doi.org/10.1016/j.fcr.2021.108377



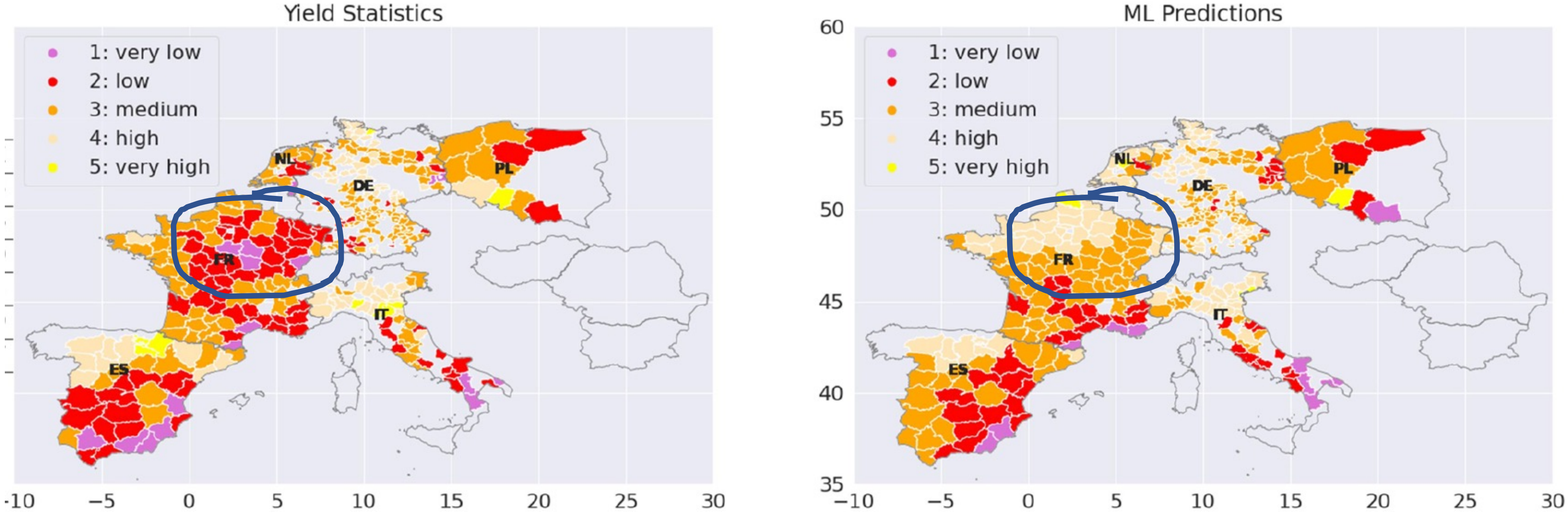
Paudel et al. 2022
/doi.org/10.1016/j.fcr.2021.108377

Predicted wheat yield classes *vs*. Observed yield classes in 2016



Paudel et al. 2022
/doi.org/10.1016/j.fcr.2021.108377

Predicted wheat yield classes *vs*. Observed yield classes in 2016



Paudel et al. 2022
/doi.org/10.1016/j.fcr.2021.108377

Conclusions

- Various types of information are available to predict yields
- Traditional methods based on regression and process-based models often fail to capture extreme yield loss occurrences
- Machine learning is attractive but does not always perform better
- Extreme yield loss events are still very difficult to predict

- Why? Short yield time series!

Perspective

- Try to predict agricultural prices directly, not yields

Acknowledgment

- Bernard Schauberger (PIK)
- Tamara Ben-Ari (INRAE)
- Damien Beillouin (CIRAD)
- JRC Ispra