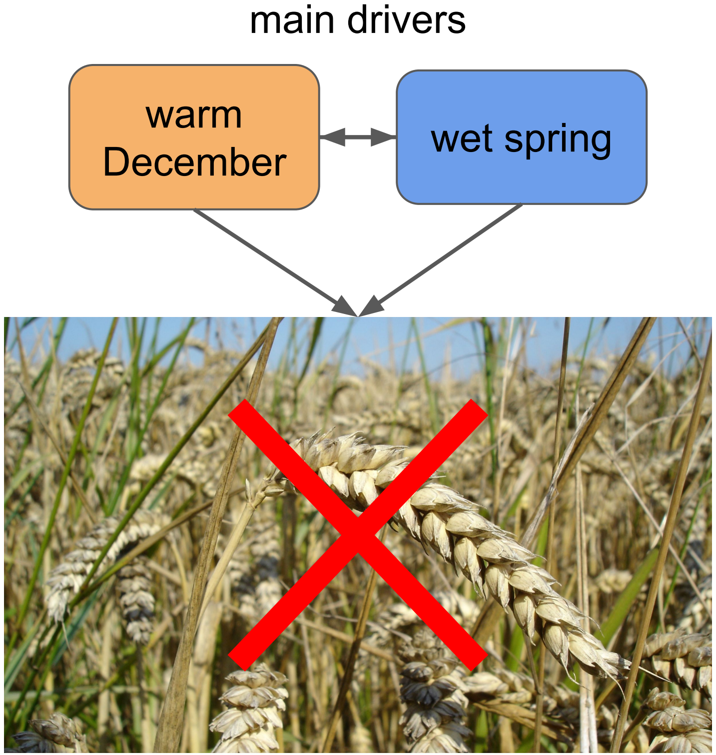
Identifying compound meteorological drivers of extreme wheat yield loss using LASSO regression

Christoph Sauter, Cristina Deidda, Leila Rahimi, Pauline Rivoire, Elisabeth Tschumi, Johannes Vogel, Karin van der Wiel, Jakob Zscheischler

https://commons.wikimedia.org/wiki/File:Wheat_close-up.JPG

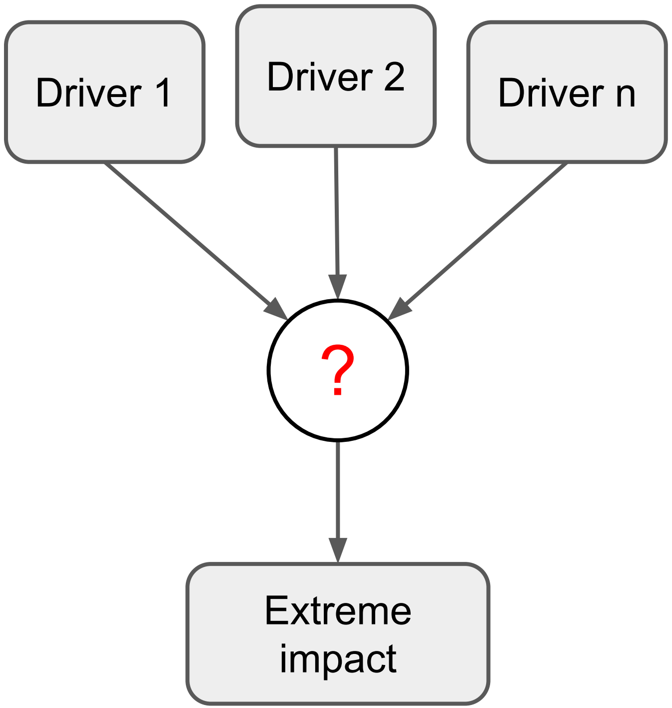
Motivation: 2016 crop failure in France

- In 2016 France experienced an unprecedented crop failure
- The drivers of this event were identified as a combination of an unusually warm December and a wet spring (Ben-Ari et al., 2018)
- Both of these drivers in combination lead to the extreme impact



Motivation

- An extreme impact can be caused by a combination of several meteorological drivers (i.e., a *compound event*, Zscheischler et al., 2018); These drivers are often unknown
- How to identify multiple (unknown) drivers of extreme impacts?



Aim of this study

Identify meteorological conditions that lead to crop failure through statistical modelling.

- Use model simulations to have large amount of consistent data available

- Use LASSO regression to identify explanatory variables of crop failure

Data

- We use EC-Earth to provide 1600 years of daily meteorological input data for the crop model APSIM (simulates wheat)

- Crop yield depends on meteorological conditions during growing season

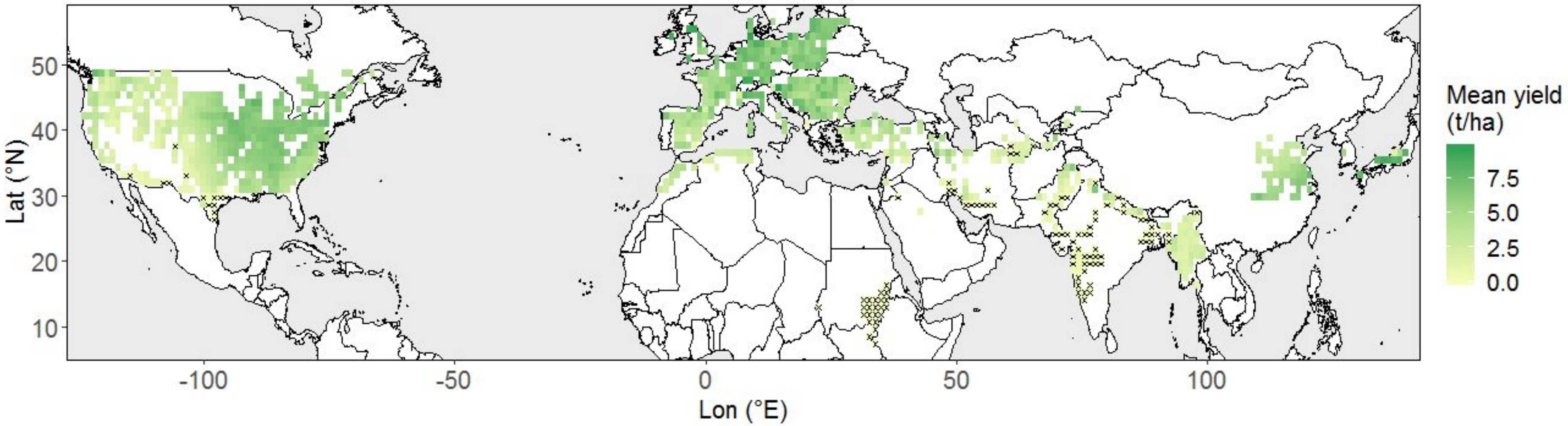
EC-Earth: Meteorological boundary conditions → APSIM: Crop model → Yearly crop yield

Input variables:

- Dewpoint temperature
- Maximum temperature
- Minimum temperature
- Shortwave radiation
- Precipitation
- Wind speed

https://commons.wikimedia.org/wiki/File:Wheat_fields_at_N%C3%B6bbel%C3%B6v,_Lund_-_panoramio_(1).jpg

Mean yield

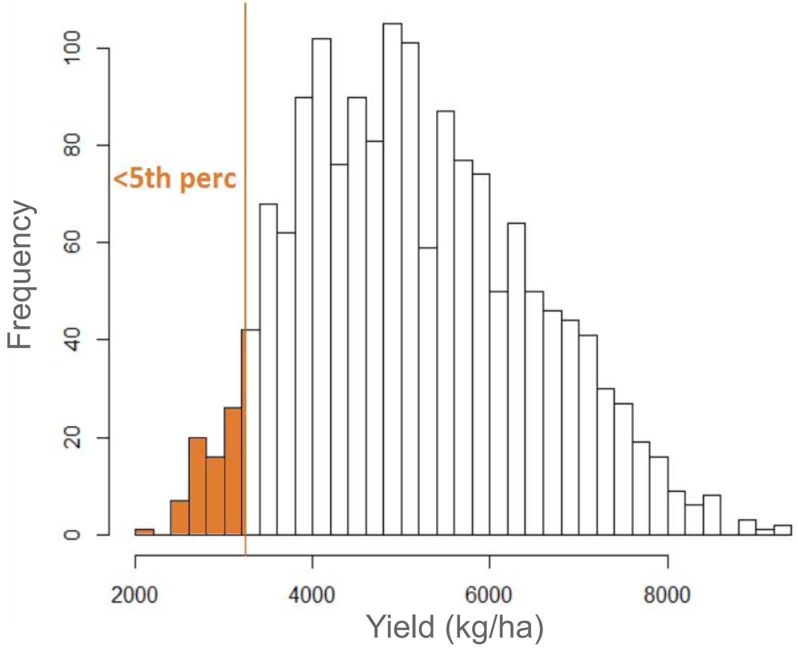


We exclude grid points with mean yield below the 10th percentile (crosses).

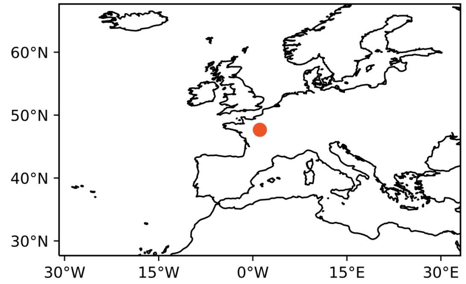
Defining extreme impacts

- We define the lowest 5 percent of crop yields as extreme yield loss years ("bad years")

- All remaining years are considered "normal years"



Exploratory analysis (single grid point)

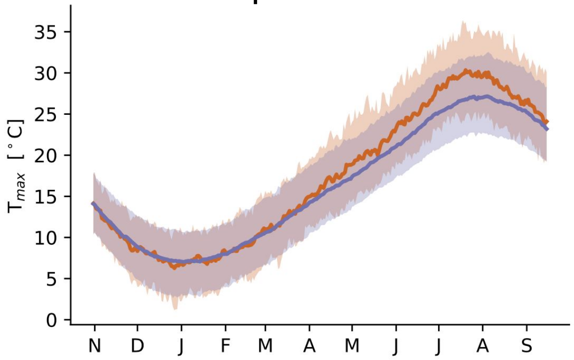


- Daily evolution of meteorological variables over the course of the year for a grid point in France

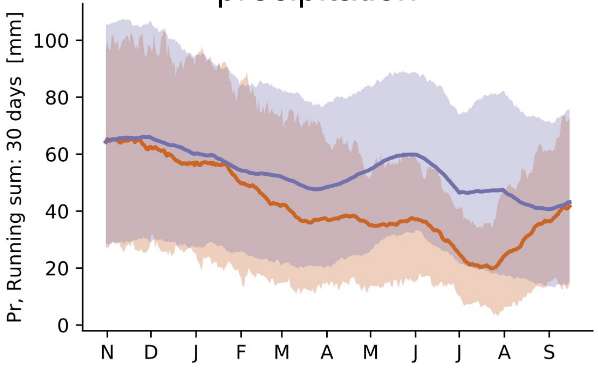
- Exploratory composite analysis of meteorological variables: some variables (not all) seem to be linked to bad yield at certain times during the growing season

Bad years
Normal years

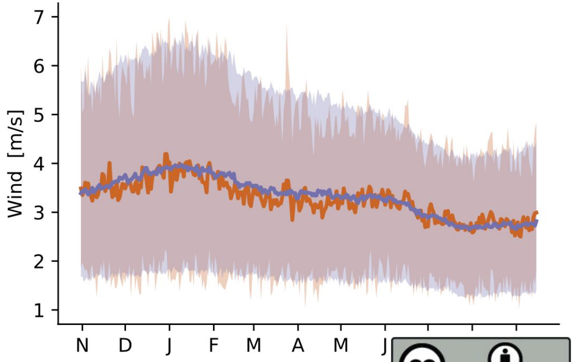
maximum temperature



30-day sum of precipitation



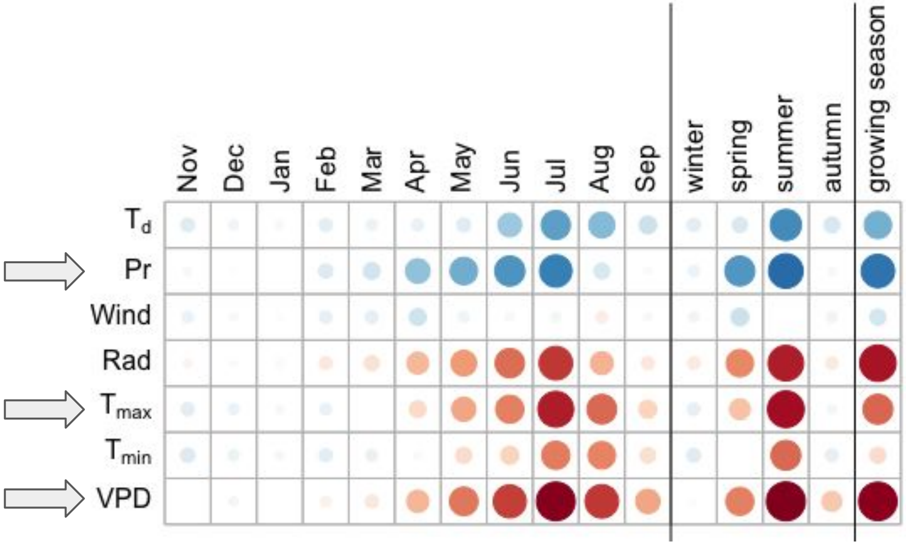
wind speed



Selecting relevant variables (and months)

- Correlation between crop yields and meteorological variables

- We use this analysis to narrow down the meteorological variables to be investigated further: precipitation (Pr), maximum temperature (Tmax), vapor pressure deficit (VPD, calculated based on met. variables)



Meteorological drivers used in the analysis

In addition to the Tmax, VPD and Pr, 7 extreme indices were considered (Vogel et al. 2019)

Variable name	Description	Type
T_{max}	Maximum Temperature	Monthly mean
VPD	Vapour-pressure deficit	Monthly mean
Pr	Precipitation	Monthly mean
dtr	Mean diurnal temperature range in the growing season	Climate extreme indicator
frs	Number of frost days in the growing season	Climate extreme indicator
TXx	Maximum temperature in the growing season	Climate extreme indicator
TNn	Minimum temperature in the growing season	Climate extreme indicator
Rx5day	Maximum five day precipitation sum in the growing season	Climate extreme indicator
TX90p	Number of warm days in the growing season with daily maximum temperature above the 90^{th} percentile [a]	Climate extreme indicator
TN10p	Number of cold days in the growing season with daily minimum temperature below the 10^{th} percentile [a]	Climate extreme indicator

[a] Note: Percentiles are grid-point based, i.e. they are representative for the local climate

A logistic regression model

- Goal: statistical approach that identifies predictors that have a strong impact on a target variable

- Yield considered as the binary response, 0=bad year, 1=normal year

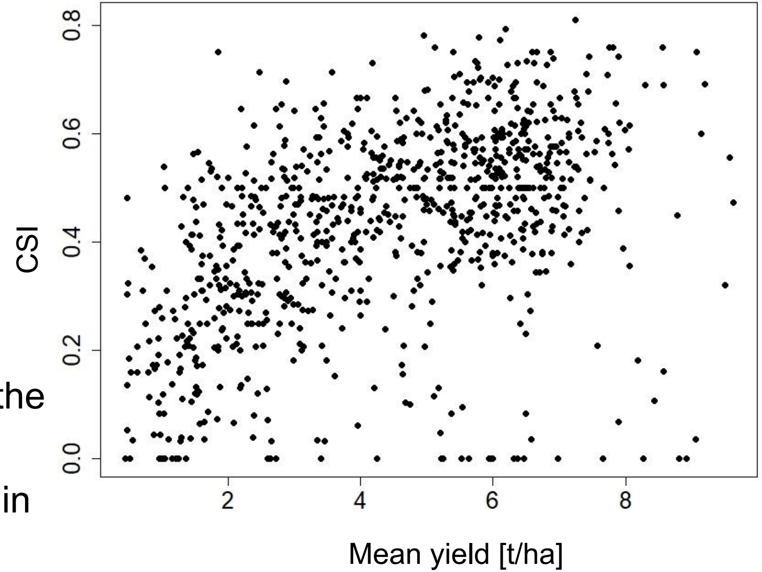
- LASSO logistic regression (Tibshirani 1996)
    - Optimizes the fit through regularization (prevents overfitting)
    - Account for correlation between variables
    - Automatic variable selection: Removes unimportant variables and keeps only relevant variables that contribute towards prediction of the impact

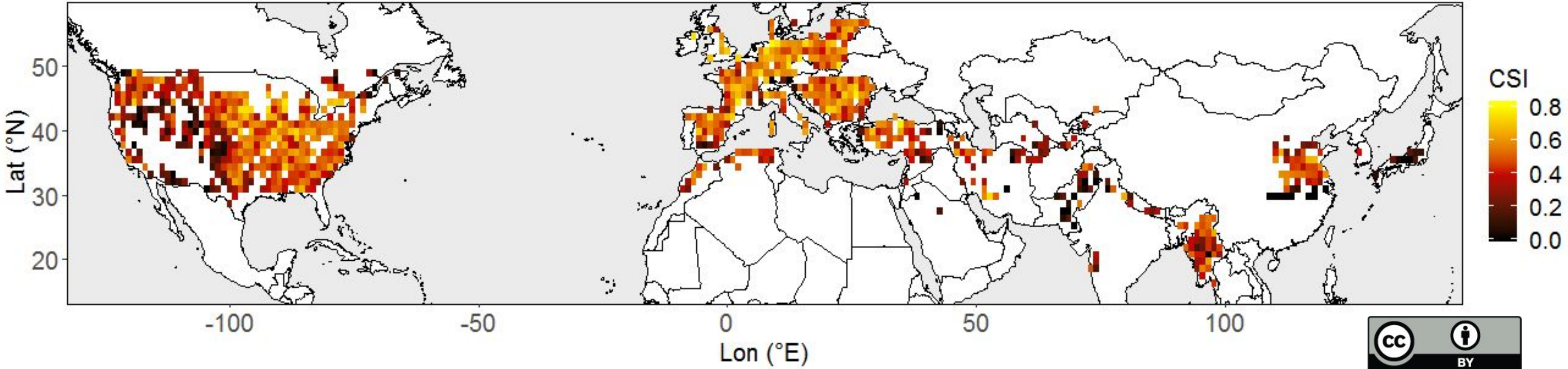
LASSO model performance

Evaluation of the model using Critical Success Index (CSI)

$$\boxed{\mathrm{CSI} = \frac{\mathrm{Hits}}{\mathrm{Hits} + \mathrm{Misses} + \mathrm{False\ Alarms}}}$$

- Mean CSI is 0.43
- High performance in Central Europe and in the Corn Belt in the USA
- Low performance in the vicinity of the Rocky Mountains and in many Asian grid points
- Strong correlation between CSI and mean annual yield

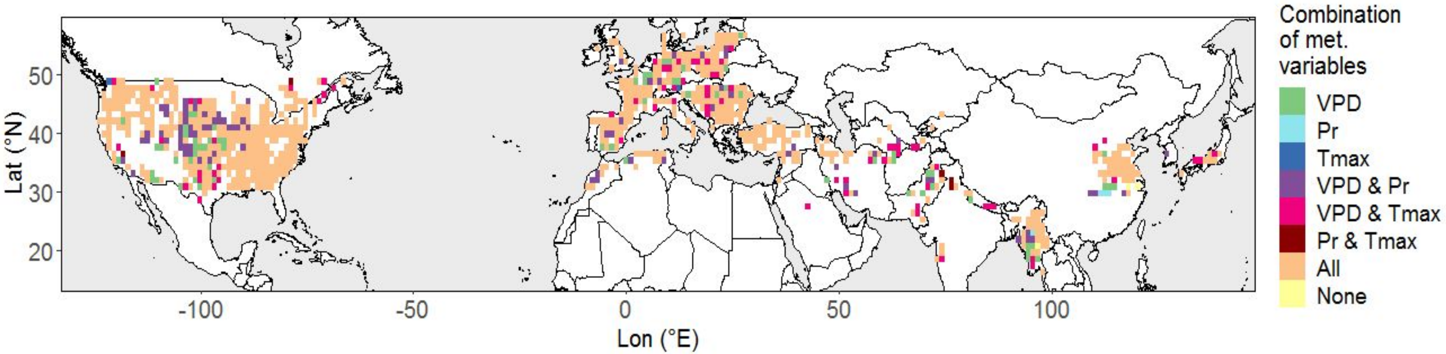




Analysing compounding effects (1/2)

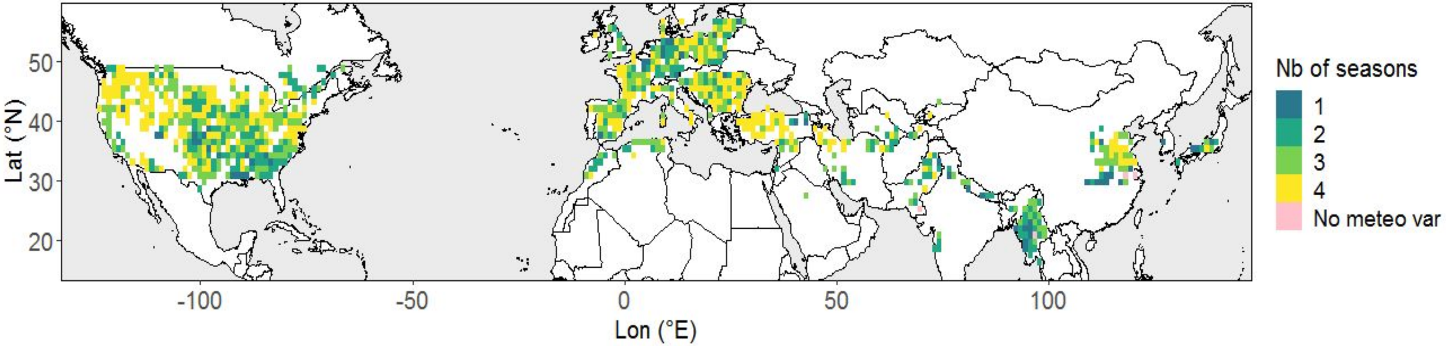
Which **meteorological variables** are included?

- VPD is included in 97.3% of grid points
- 73.0% of grid points include all variables



How many **seasons** are included?

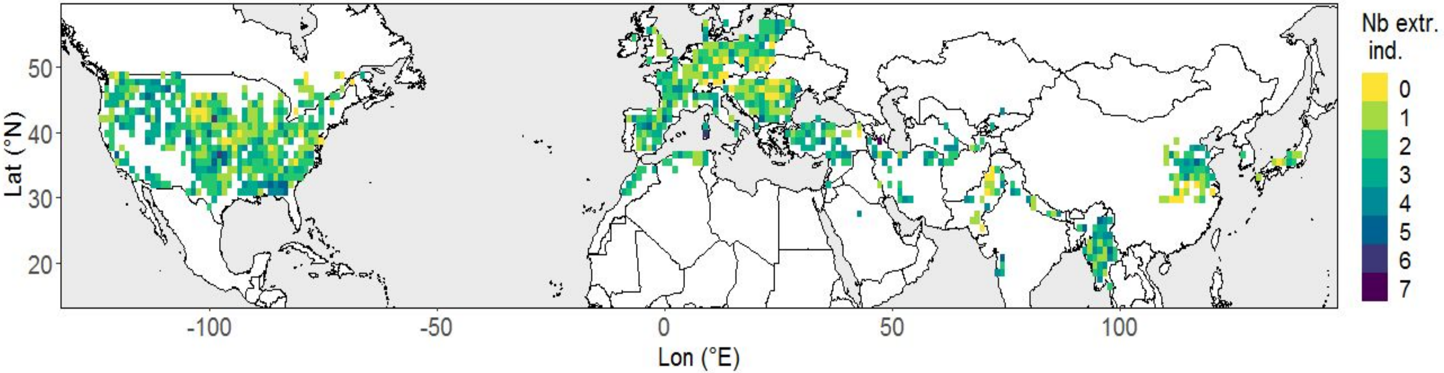
- Increasing number from SE to NW in the USA
- Number of included seasons is low in southern Asia



Analysing compounding effects (2/2)

How many **extreme indicators** are included?

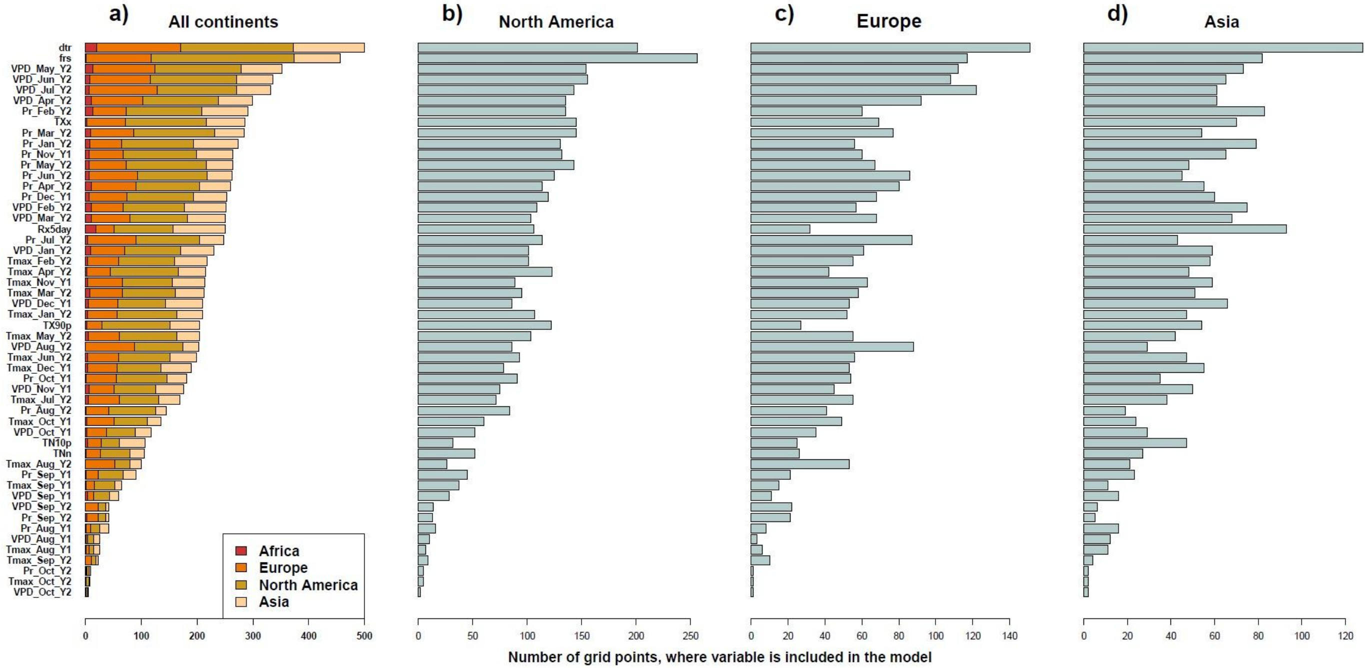
- median number of extreme indices kept = 2
- very few areas without any extreme indices kept



Variable importance

- In Europe and North America, VPD and Pr in spring to early summer are the prevailing monthly predictors

- Diurnal temperature range (dtr) and number of frost day (frs) are the most frequent extreme indicators, maximum 5-day precipitation (Rx5day) is especially common in Asia



The extension "Y1" signifies that the respective variable month belongs to the first year of the growing season, while "Y2" signifies it belongs to the latter.

Discussion

- Strong relationship between CSI and mean yield: LASSO model performs better for location with high mean crop yield
- In most areas, nearly all meteorological predictors and at least two seasons are relevant for explaining bad years → compounding effect between multiple climate drivers and seasons
- VPD was kept as explanatory variable in almost all grid points
- Diurnal temperature range, number of frost days, and the hottest day of the year are the most relevant extreme indicators to explain crop failure
- Consideration of predictors at the exact timing of phenological stages instead of at a monthly scale might further improve model accuracy

Conclusions

- We have presented an automatic approach to identify the most relevant meteorological predictors of extreme impacts using simulated wheat yields as an example

- LASSO regression can successfully
  - identify the drivers of extreme bad yield
  - identify during which time of the year the meteorological variables play a critical role for causing crop failure
  - integrate a large number of predictors

- The findings can be compared to real-world observations and the approach can be applied to other climate-related impact

References

Ben-Ari, T., Boé, J., Ciais, P., Lecerf, R., Van der Velde, M., and Makowski, D.: Causes and implications of the unforeseen 2016 extreme yield loss in the breadbasket of France, Nature communications, 9, 1–10, 2018.

Tibshirani, R.: Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society, 58, 267–288, 1996.

Vogel, E., Donat, M. G., Alexander, L. V., Meinshausen, M., Ray, D. K., Karoly, D., Meinshausen, N., and Frieler, K.: The effects of climate extremes on global agricultural yields, Environmental Research Letters, 14, 054 010, 2019.

Zscheischler, J., Westra, S., van den Hurk, B., Seneviratne, S, I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., and Zhang, X.: Future climate risk from compound events, Nature Climate Change, 8, 469–477, 2018.