

Before answering the comments done by the referees, the authors want to mention a modification done in the training of the ML model.

Validation/Training split

During a recent analysis using this same model, we have realised that the validation years were not fixed in each random seed run, which resulted in each run having different train/validation split. As the common way to proceed is to fix the validation years (and was the original purpose in the framework) the code has been modified to fix the validation years to those that had been already specified in the manuscript (see table B1):

Validation years: 1950, 1952, 1955, 1956, 1957, 1987, 1992, 1994, 1999, 2001, 2006, 2008, 2010

The results of the training and explanation using SHAP values have not changed much, as was expected. However, we keep the validation years fixed across different random seed initialization runs, as is the common procedure.

Computation of uncertainty bars in mean SHAP plots

An additional fix has been applied to the computation of error bars in the mean shap results shown in figures: 5, 18, 20, 21. A mistake was found in the code that made the error bars incorrect.

— Referee Comments —

We answer here to the comments made by Referee 2 in the minor revision.

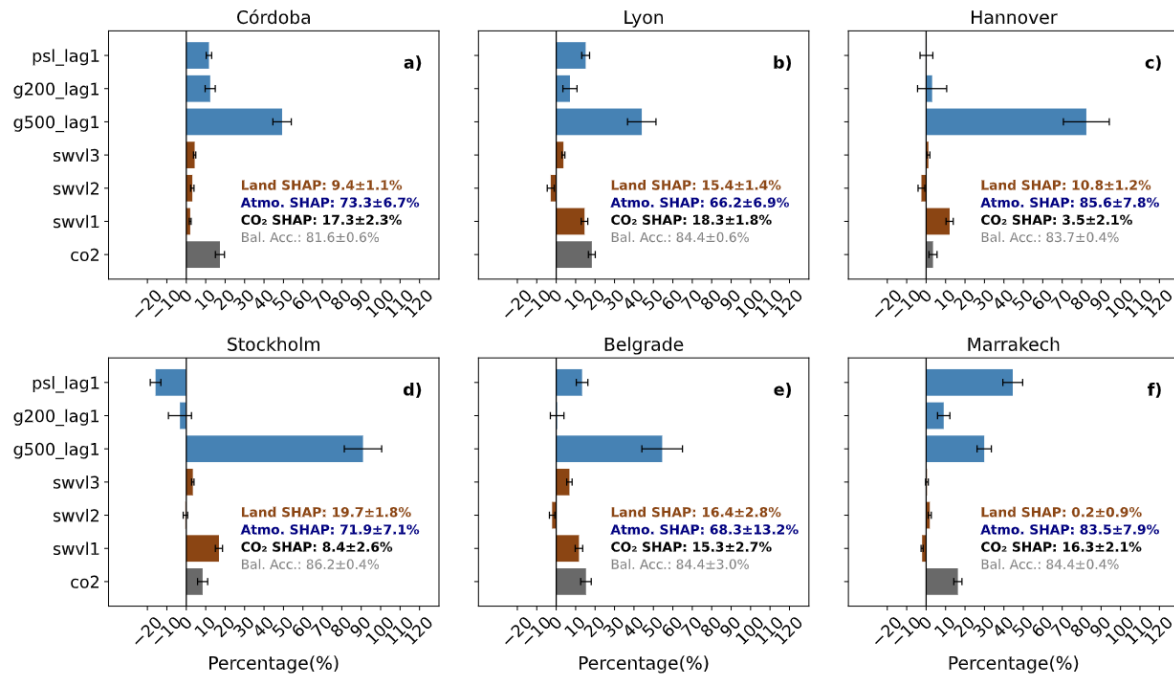
The authors want to thank the referee for his/her time and help reviewing the submitted manuscript.

CO2 SHAP heterogeneity (follow-up)

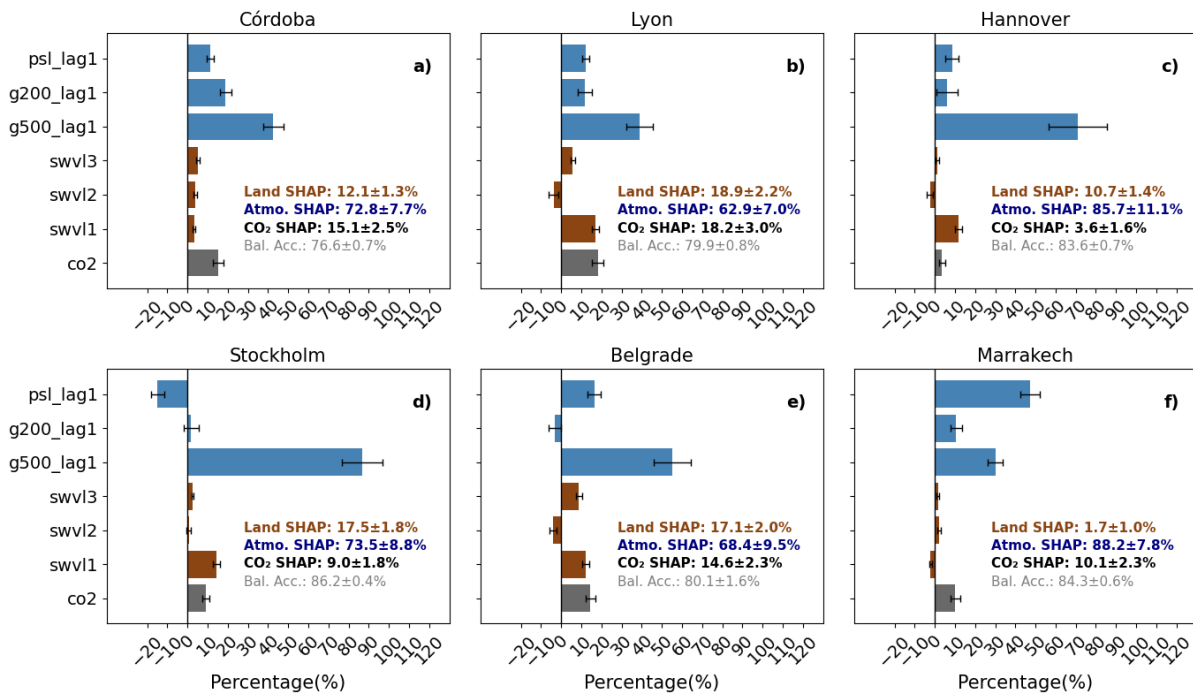
Thanks for the follow-up on this matter. The authors agree that the interpretation on the CO2 feature needs to be reviewed. In the set-up of the manuscript, this feature alone does not allow to identify the role of CO2 alone, as including a time index can yield similar results due to the fact that both increase monotonically in the recent years.

To test this, we have carried out the experiment suggested by the referee and used a normalized time index instead of CO2. Below we show the original Mean SHAP results and the ones obtained using the time index:

Original:



Replacing CO₂ with a normalized time index:



Comparing both figures we can see that results are quite similar except for Córdoba and Marrakech, where we see a higher importance when using CO₂ instead of the time_index.

The authors still think that using CO₂ instead of a plain time index is more suitable as it is closer to the actual physical processes that cause the long-term warming.

We suggest pointing this out in the final version of the manuscript. We suggest including this text in section 2.1 L96, where the CO2 index is defined:

Old L96: By explicitly including CO2, our framework controls for the long-term warming trend as a potential additional driver of heat extremes, allowing for a cleaner attribution of the physical drivers

New L96: *In our set-up, the CO2 feature acts as a proxy for the multi-decadal trend signal of heatwave probability. This trend represents an aggregate signal combining forced anthropogenic warming, long-term shifts in local temperature drivers, and low-frequency internal climate variability. Using a normalized time index yields similar attribution results (not shown), since recent years are associated with higher CO2 concentrations. However, we use CO2 concentrations because this is closer to the actual physical processes causing the long-term warming. Additionally, using CO2 concentrations to account for the trend remains valid under potential future conditions when CO2 concentrations stabilise or even decline.*

Additionally, we re-write the sentences in the manuscript that identify the importance of the CO2 feature with an anthropogenic forcing:

- L17: They also confirm the enhanced role of the land component in regions of Northern Europe and reveal a contribution of the anthropogenic factor through CO2 concentrations, even for specific events

New: They also confirm the enhanced role of the land component in regions of Northern Europe and reveal a contribution of the CO2 concentrations, even for specific events.

- L407: CO2 SHAP contributions are substantial in locations where we detect a trend in the yearly extreme count, suggesting that the model captures the long-term forcing signal.

New: CO2 SHAP contributions are substantial in locations where we detect a trend in the yearly extreme count. This trend represents an aggregate signal combining forced anthropogenic warming, long-term shifts in local temperature drivers, and low-frequency internal climate variability

Physical-understanding claim (follow-up)

The authors agree that tempering the statements clarified the purpose of the study. The claim highlighted by the referee was removed in the last submission of the revised manuscript.

Additional points noticed when skimming manuscript again

§2.2 (soil moisture): The sentence "As for soil moisture, both SPI and SPEI grid-points

are averaged spatially over a 100km x 100km region" reads as though spatial averaging had been mentioned before, but I don't believe that is the case

Thank you for pointing this out. The following clarification has been added in L92 of the revised manuscript before the text highlighted by the referee appears:

All land-surface variables are taken as the spatial average over a 100kmx100km region to avoid focusing on a single grid-point.

L122: "(see 3)", the section reference appears to be broken, should probably read "see Sect. 3"?

This has been corrected

We answer here to the comments made by Referee 3 in the minor revision.

The authors want to thank the referee for his/her time and help reviewing the submitted manuscript.

Technical corrections:

- Line 79: section 2 -> Sect. 2

This has been corrected

- Line 103: section 2.4 -> Sect. 2.4

This has been corrected

- Line 132: sec 3 -> Sect 3.2

This has been corrected

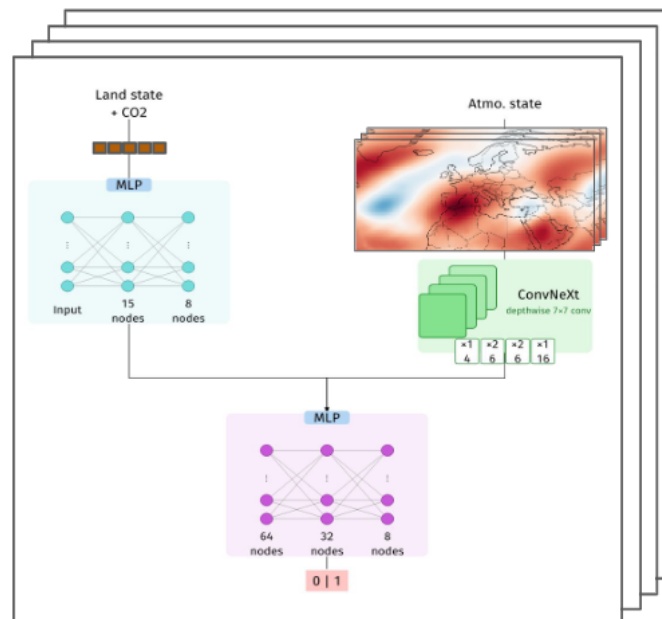
- Line 188: change to “we trust the class imbalance by training the model with a weighted loss function.”

Thanks for the suggestion. We re-writte the sentence in the revised manuscript like this:

“...we treat the class imbalance by training the model with a weighted loss function.”

- Figure 3: Reader would profit from an improvement. Include details of the ConvNeXt. Currently the written description is only of limited value due to its complexity.

The figure has been updated to make it more descriptive of the model.



The description of the two MLPs has been improved. Additionally, the 1D nature of the Land State and CO2 features has been clarified graphically compared to the spatial maps of the atmospheric features.

Regarding the ConvNext model, the depth-wise convolution has been specified as well as the layer dimensions and block depths. The authors refer to the original ConvNext paper in the manuscript for more details on the model like the specific ConvNext block

- Line 304: higher-latitude? Is Lyon considered high latitude?

Thanks for pointing this out. With “higher-latitude”, the authors want to express that Lyon is a higher-latitude compared to Córdoba, resulting in larger role of the land modulation.

- Line 462: uper -> upper

This has been corrected

- Line 467: analysied -> analysed

This has been corrected

- Line 479: “We also analyse two case studies to further study the explainability technique” This to me is a misleading formulation since the case studies are meant to provide further insight on the networks predictive behaviour rather than on the explanation techniques used to analyse the predictions.

Thanks for pointing this out. The authors agree that the sentence can be rephrased in a clearer manner. The phrase has been re-written following the referee’s suggestion:

“We also analyse two case studies to provide further insights on the networks predictive behaviour.”