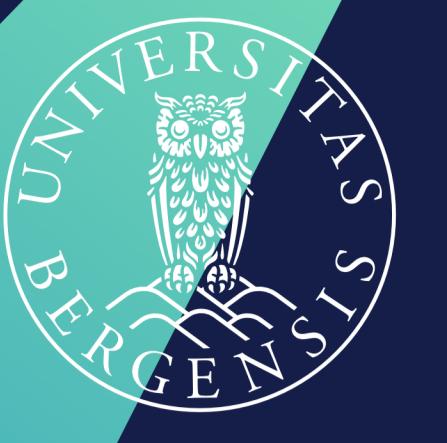


Evaluating the physical realism of neural networks with object-based Explainable AI

Application to the prediction of heavy rainfall in Western Norway

Robin Guillaume-Castel, Camille Li, Stefan Sobolowski
Geofysisk Institutt, University of Bergen & Bjerknes Centre for Climate Research

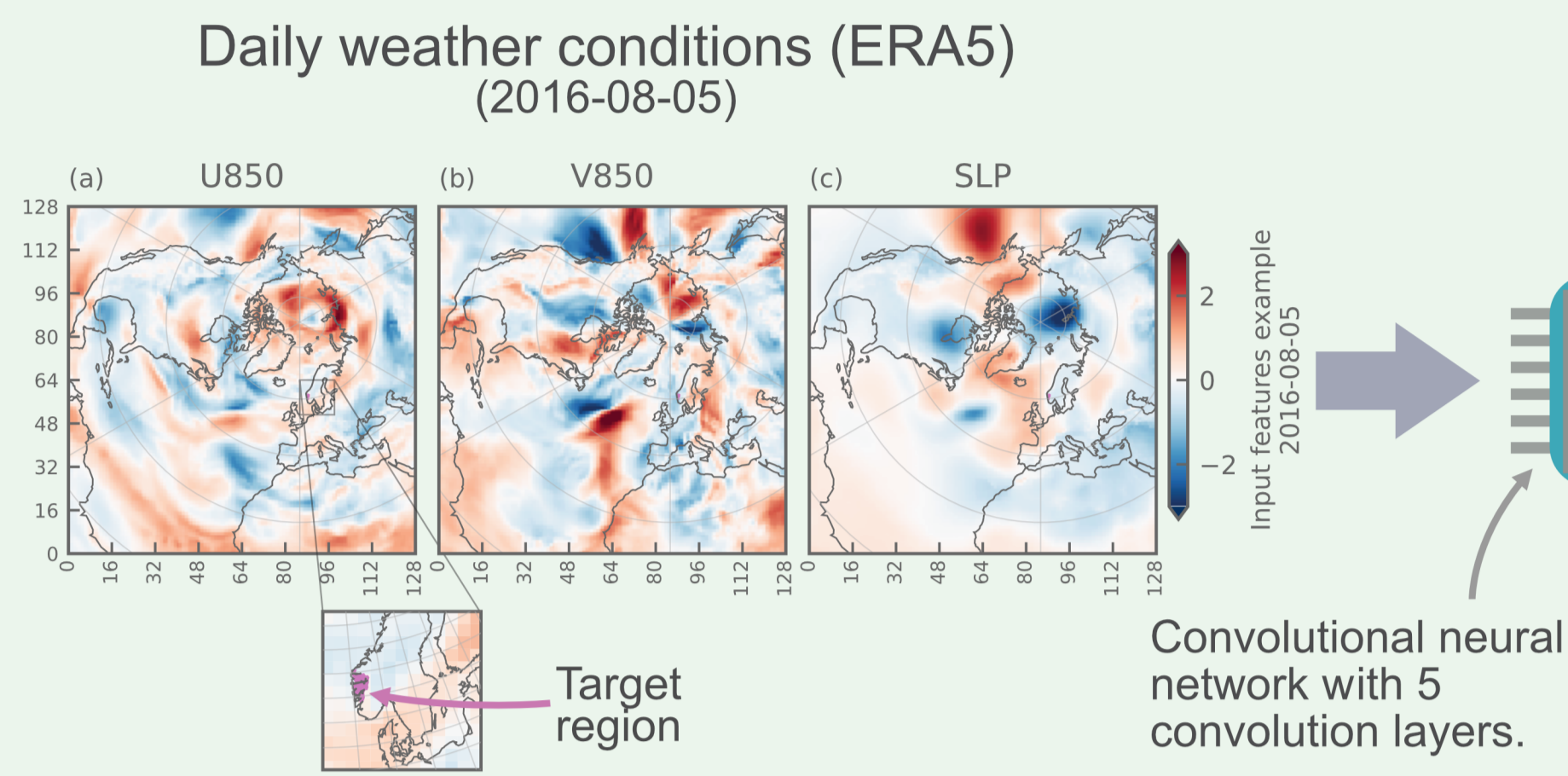
robin.guillaume-castel@uib.no



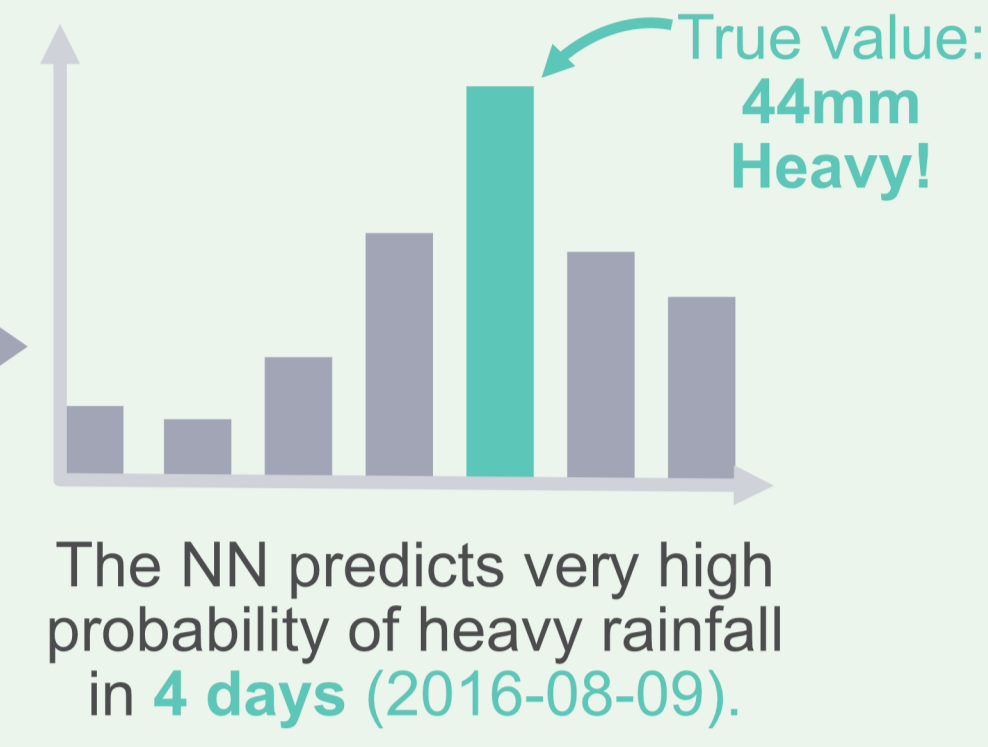
LEAD AI

Introduction

1 We train a neural network (NN) to predict heavy rainfall events



Probability of West Norway heavy rainfall event in the next days:



2 We want to assess the physical realism of the neural network's decisions with explainable AI (XAI)

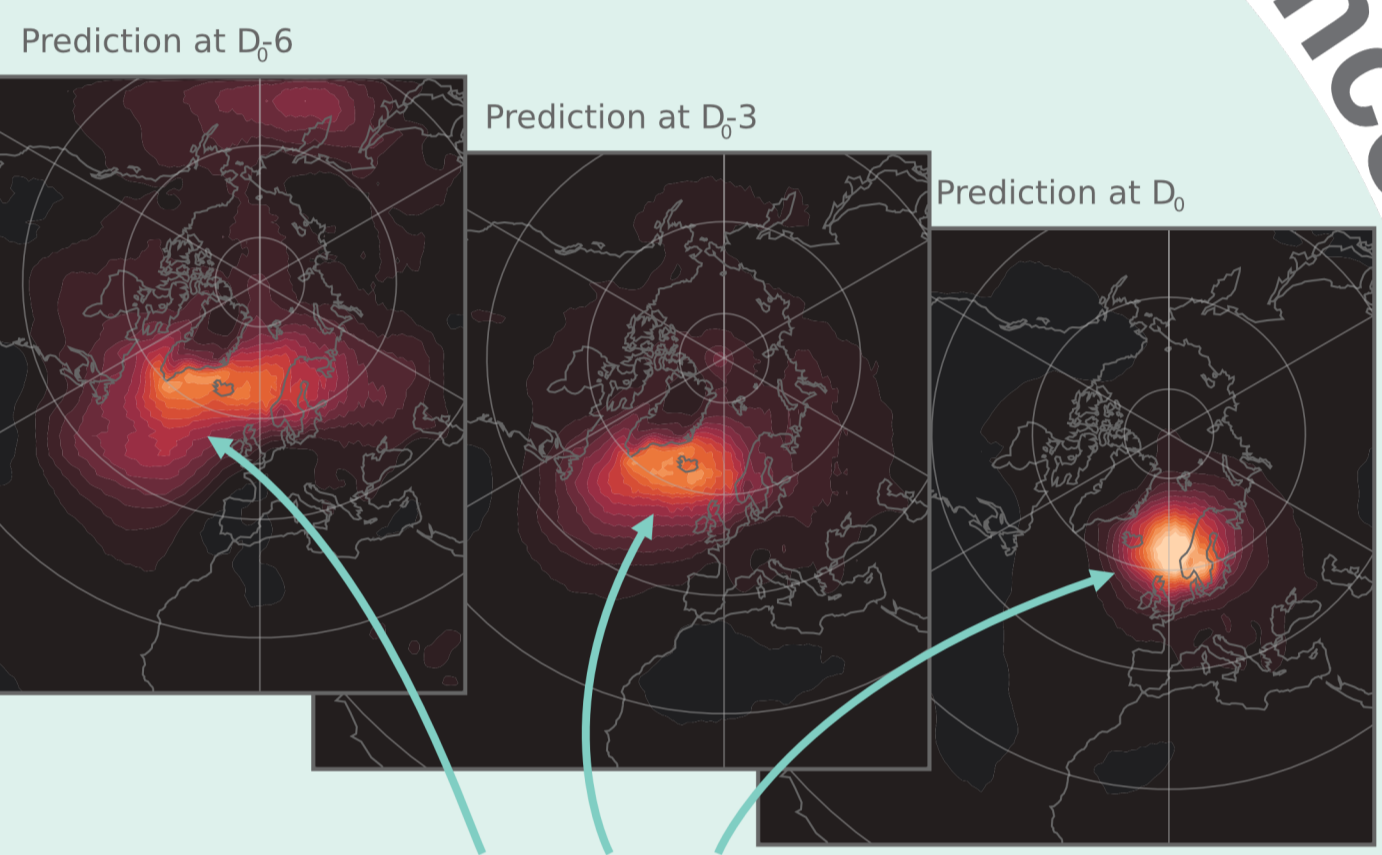
Compute **relevance** maps with Layer-wise Relevance Propagation:



Object-based XAI relevance provides a more quantitative and more robust assessment of physical realism than aggregated relevance

Aggregated XAI relevance

Average relevance over positive predictions allow to study global decision patterns



This highlights that the NN focuses on the correct geographical region

BUT

The assessment is largely qualitative.

We can say that the neural network focuses on:

The right region ✓

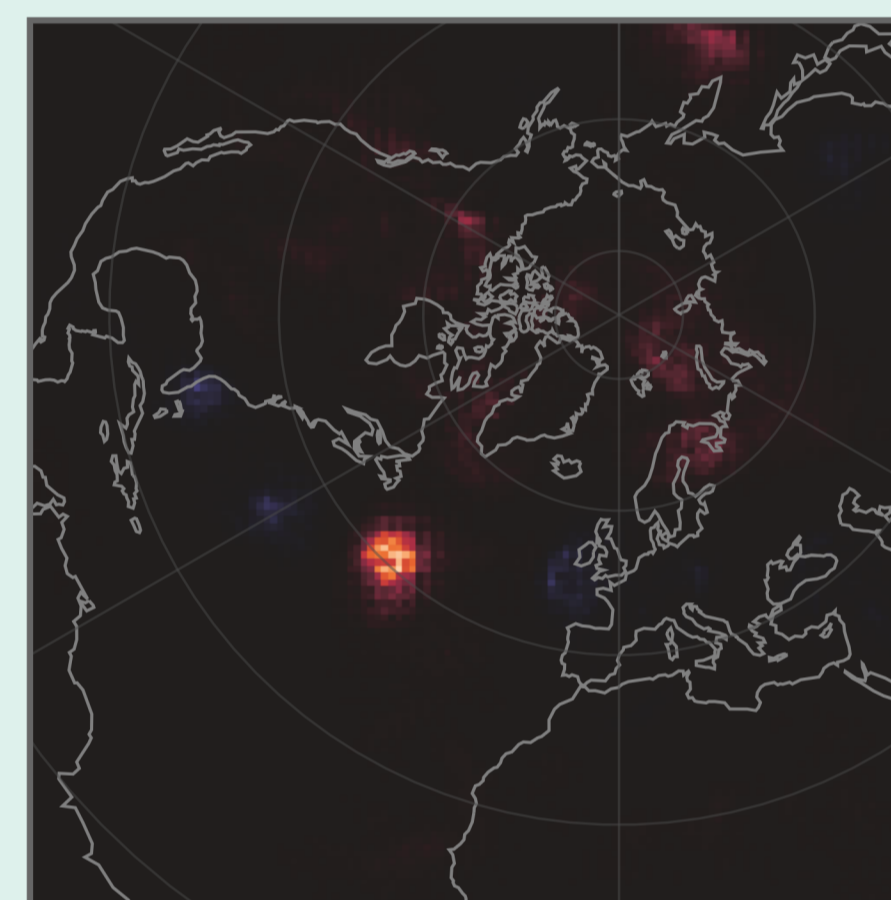
The right physical processes ?

Object-based XAI relevance

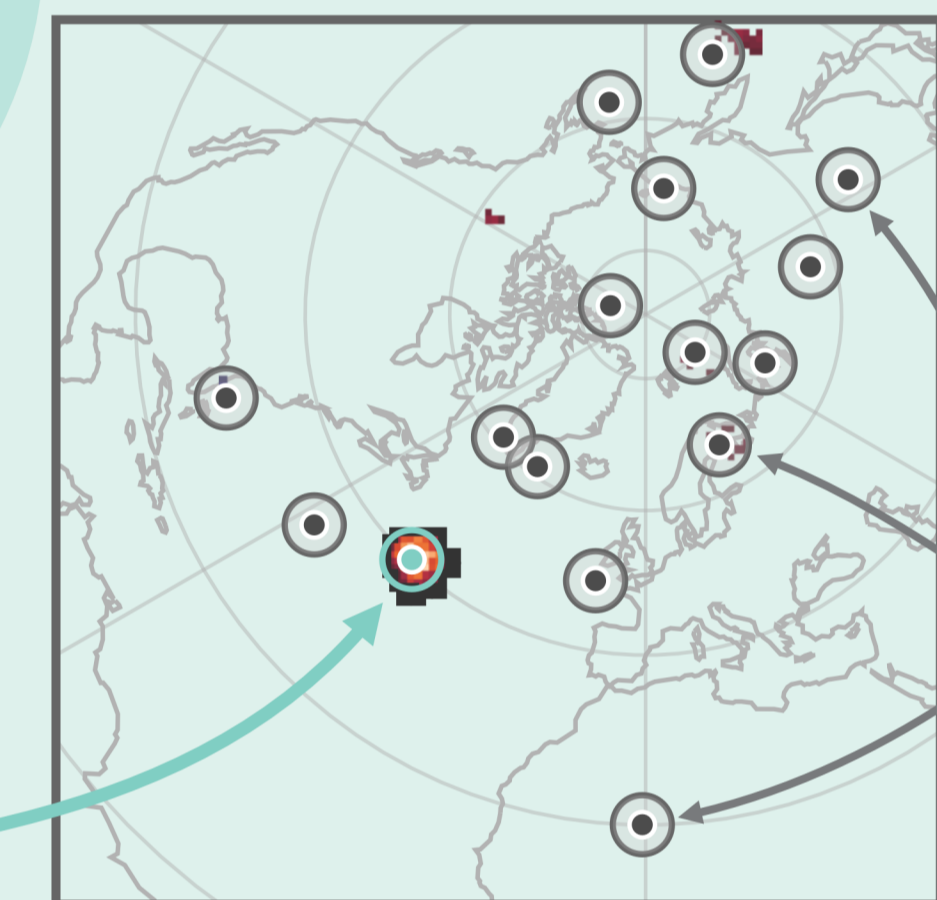
Methods

Cyclone-associated processes are the main drivers of heavy rainfall in Western Norway:

Individual predictions can be associated with cyclones



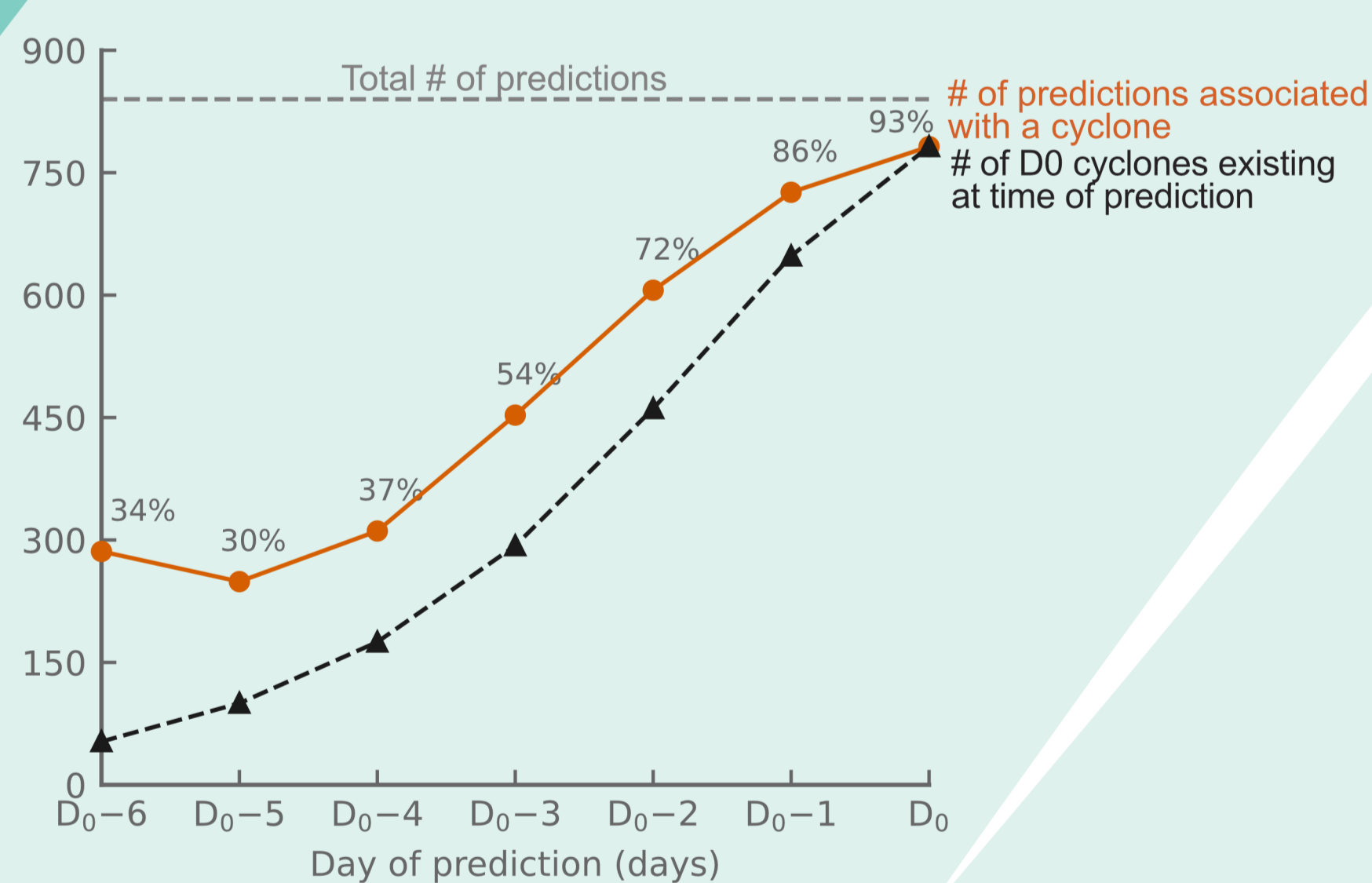
This **cyclone** is associated to the neural network's prediction



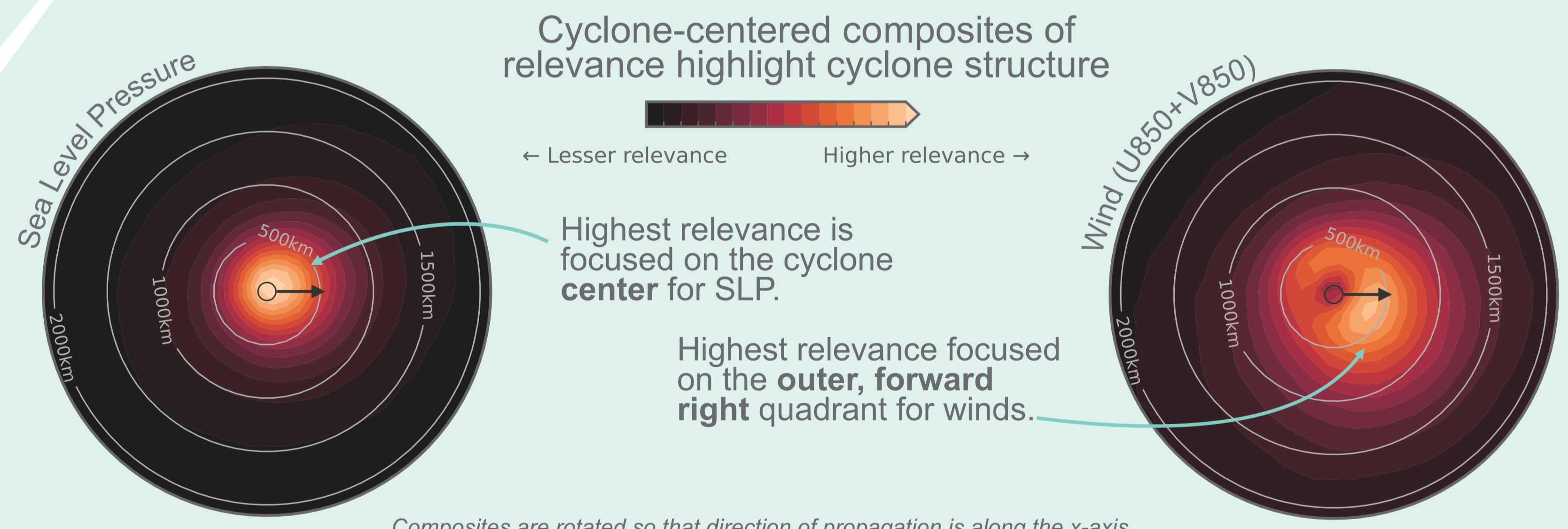
Results

Do cyclones contribute to the neural network's predictions?

1 High relevance patches coincide with cyclones



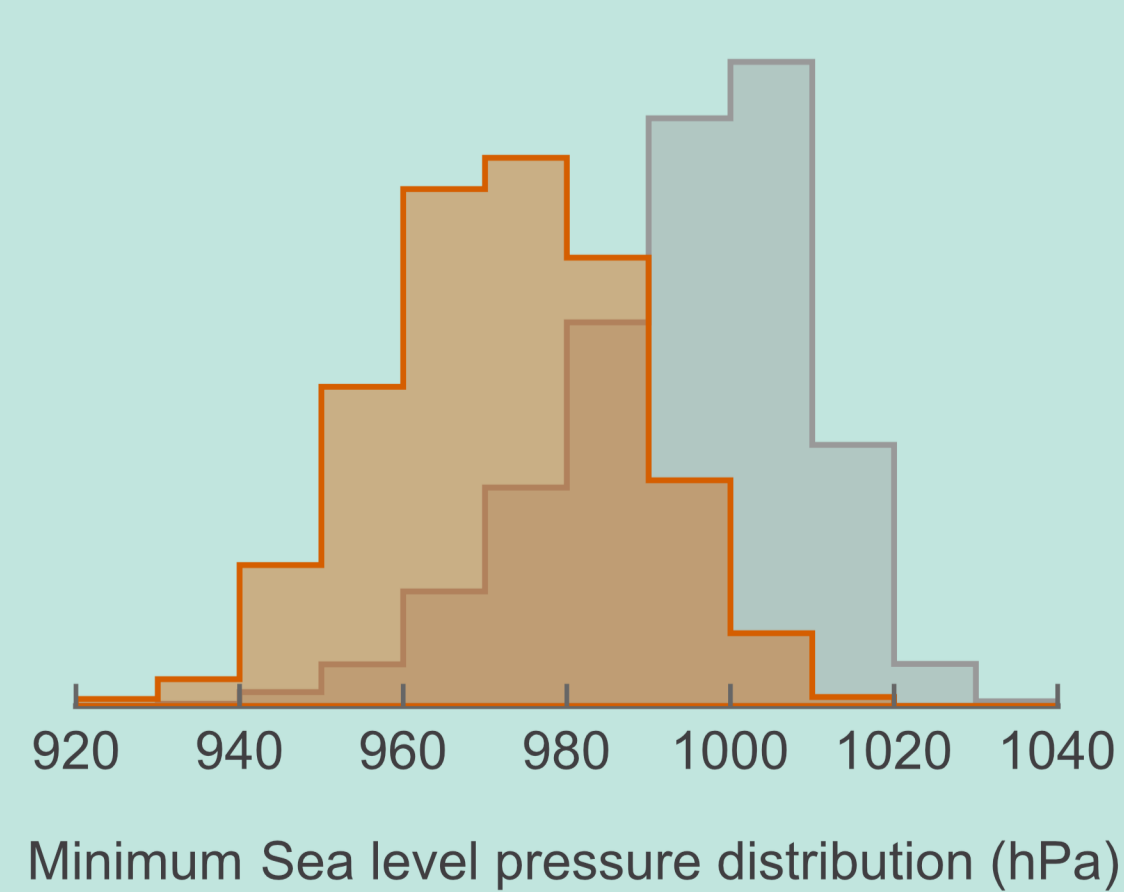
2 Relevance patches are not just collocated with cyclones but are a signature of cyclone dynamics



Which cyclones contribute to the predictions?

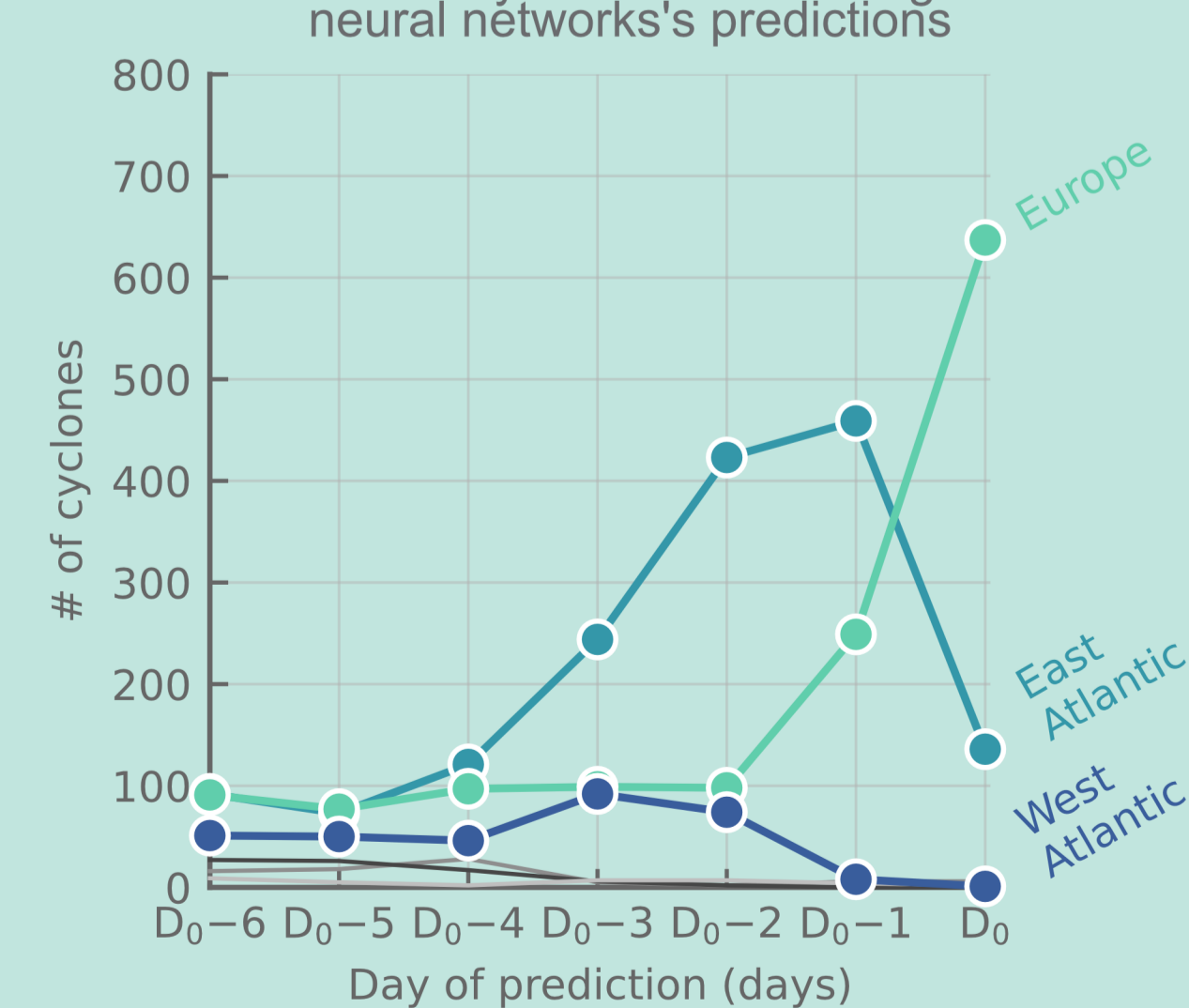
3 The NN favors stronger cyclones that can lead to heavy precipitation

■ Cyclones contributing to the NN's predictions
□ All low pressure systems



4 The NN follows cyclones upstream in the North Atlantic storm track

Locations of cyclones contributing to the neural networks's predictions



Conclusion

- Object-based XAI relevance provides a stronger and more quantitative framework for physical assessment of neural networks than aggregated XAI relevance.
- Our neural network's predictions of heavy rainfall occurrence in Western Norway are based on **North Atlantic cyclones**, aligning with physical understanding.
- We can build on this work to create **physical benchmarks** for more advanced AI weather prediction models, like *Pangu*, *GraphCast* or *AIFS*.



Co-funded by the European Union, GA#101126560; Bergen research and training program for future AI leaders across the disciplines, LEAD AI.