Using Explainable AI and Transfer Learning to understand and predict the maintenance of Atlantic blocking with limited observational data

Huan Zhang (NYU), Justin Finkel (MIT), Dorian S. Abbot (UChicago), Edwin P. Gerber (NYU), Jonathan Weare (NYU)

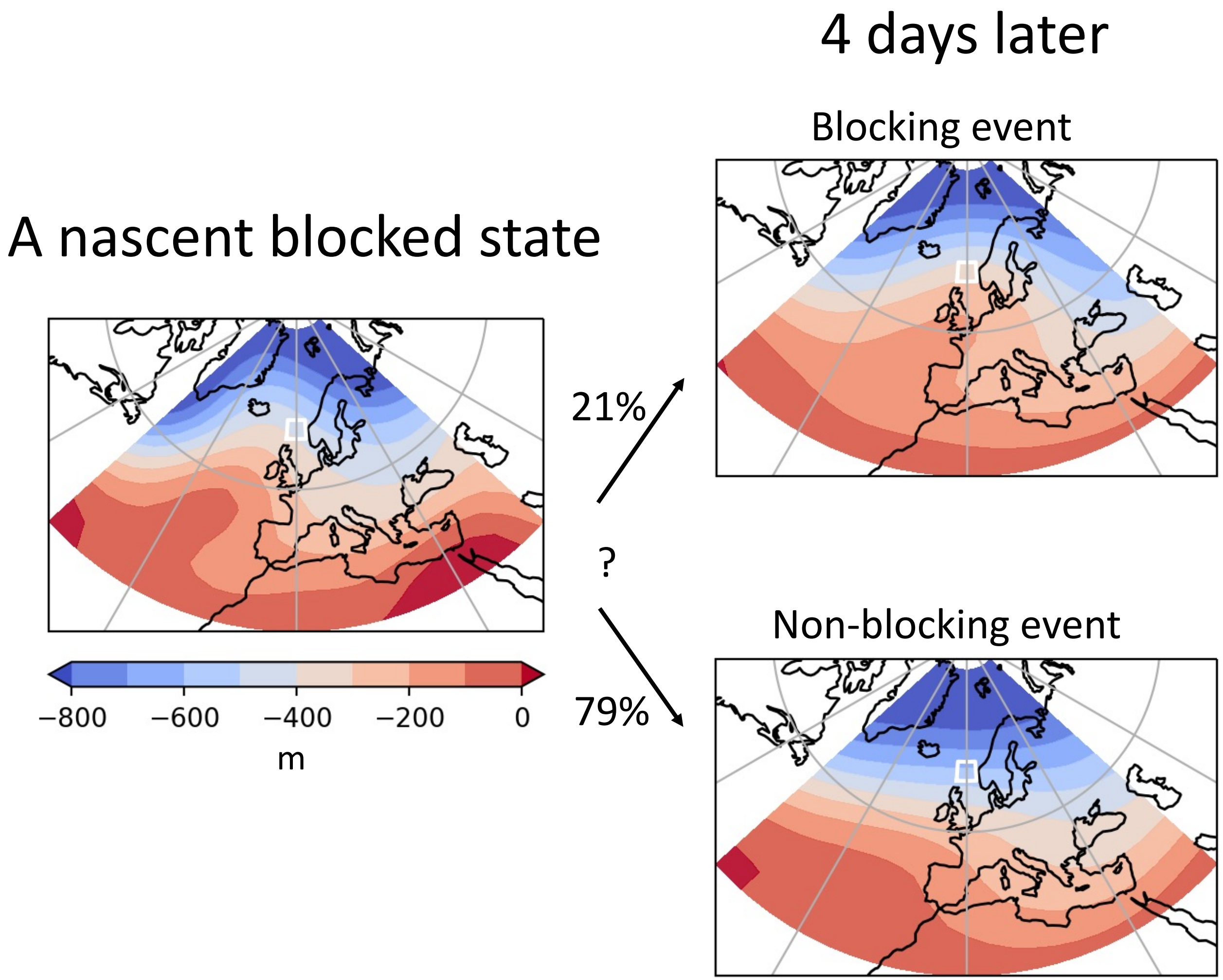
Blocking and Extreme Weather in a Changing Climate, US CLIVAR Workshop

This work is supported by

Predict: Does a blocked state remain blocked?

Given the onset of a blocked state, what is the likelihood that the flow will remain blocked for an extended period (≥ 5 days) ?

Conditional probabilistic forecasting



Two questions

- Can a data-driven approach predict persistent blocking events?
  If so, how?

- Can we do this within the limits of the observational record?

Predict: Does a blocked state remain blocked?

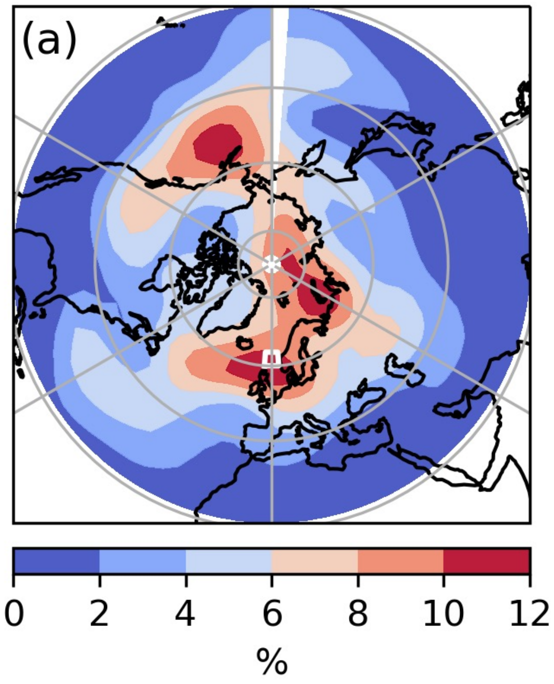
- Classification problem in machine learning
- But a plausible neural network training requires lots of data!
- Idealized model: Marshall-Molteni (3-layer QG equations)

ERA5 comes later!

Marshall, J., and F. Molteni, 1993: Toward a Dynamical Understanding of Planetary-Scale Flow Regimes. *J. Atmos. Sci.*, **50**, 1792–1818

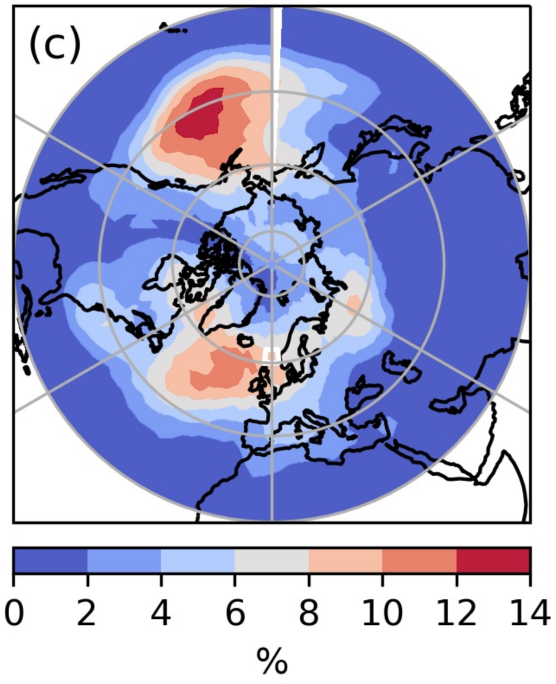
Model statistics comparison

Marshall-Molteni model



Blocking events days percent

ERA5 data



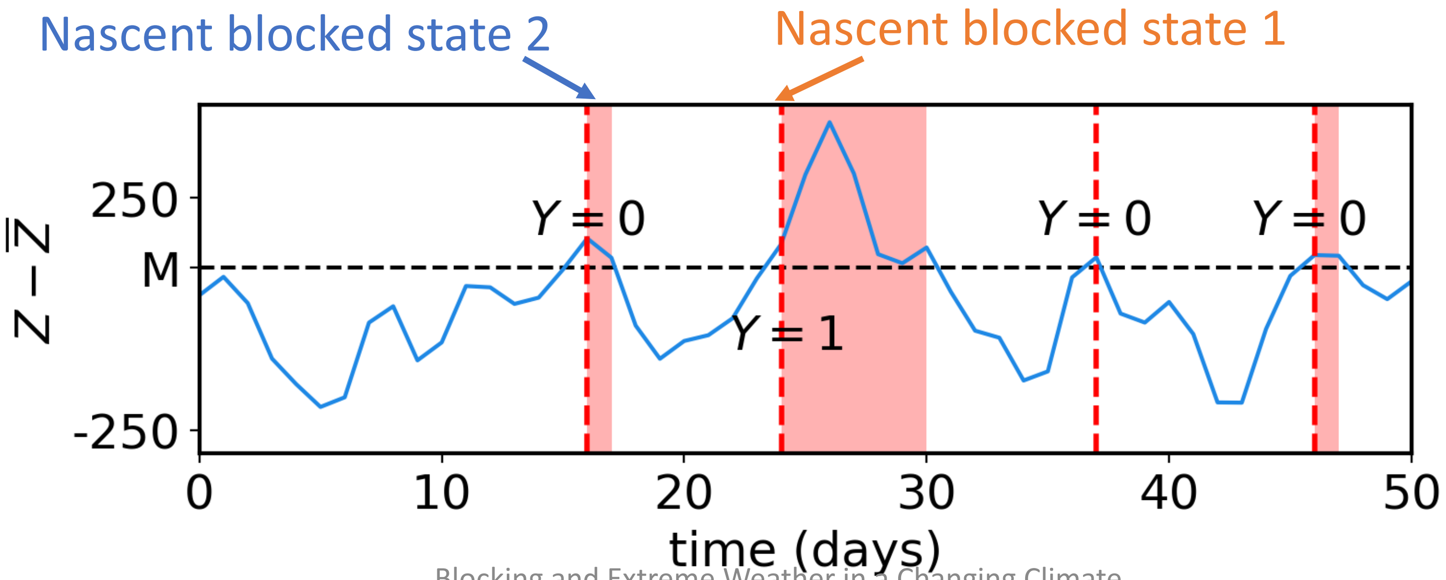
Blocking events days percent

Neural Network training

- We train a convolutional neural network (CNN) $p_\theta(x)$ to approximate

$$p(x) = \mathbb{P}(X \text{ starts a blocking event} | X = x)$$

- X_n are nascent blocked states (18 x 90 x 3 grid map)
- Y_n are binary variables: 1 if X_n starts a blocking event, 0 if not.



Performance Metrics

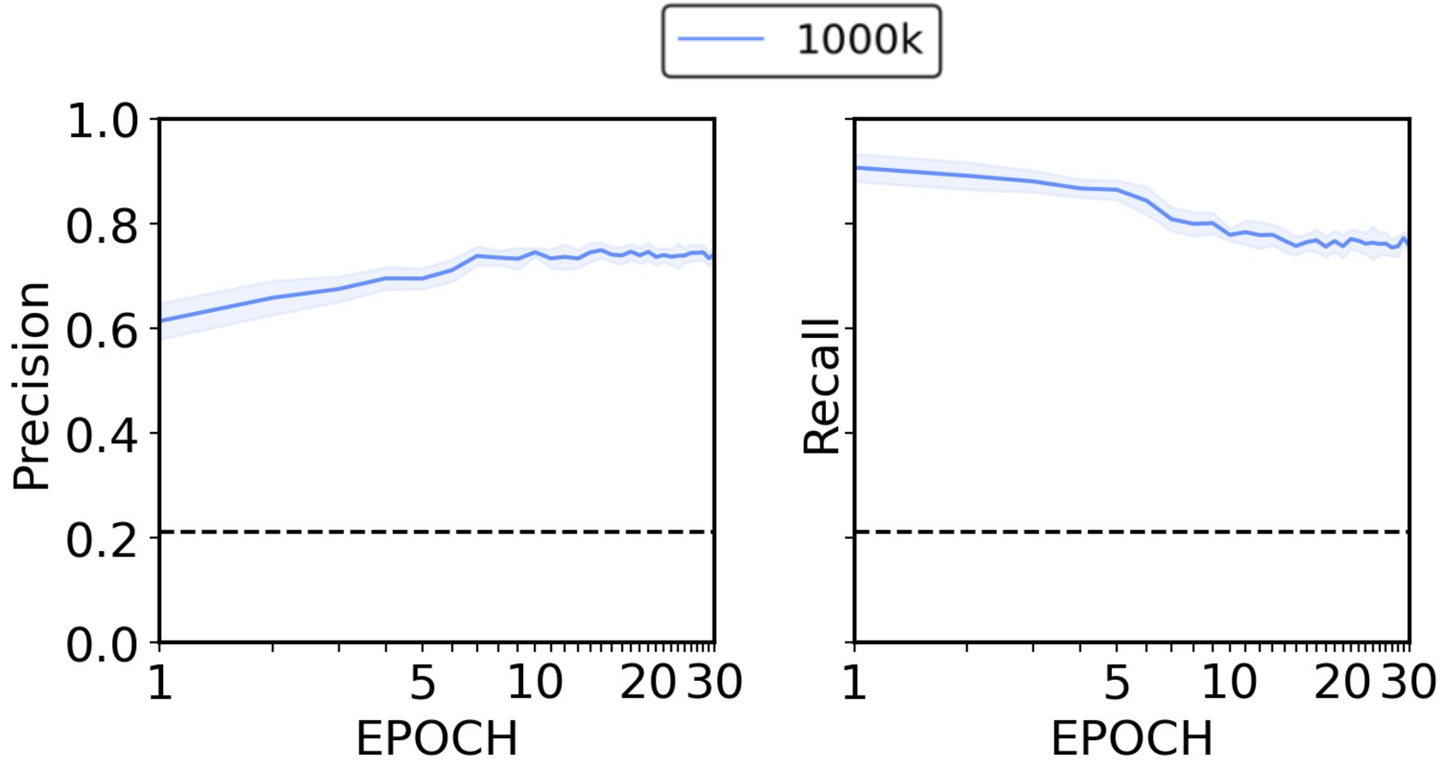
$$\text{precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

- Precision means “When you forecast an event, what is that probability you are correct”.

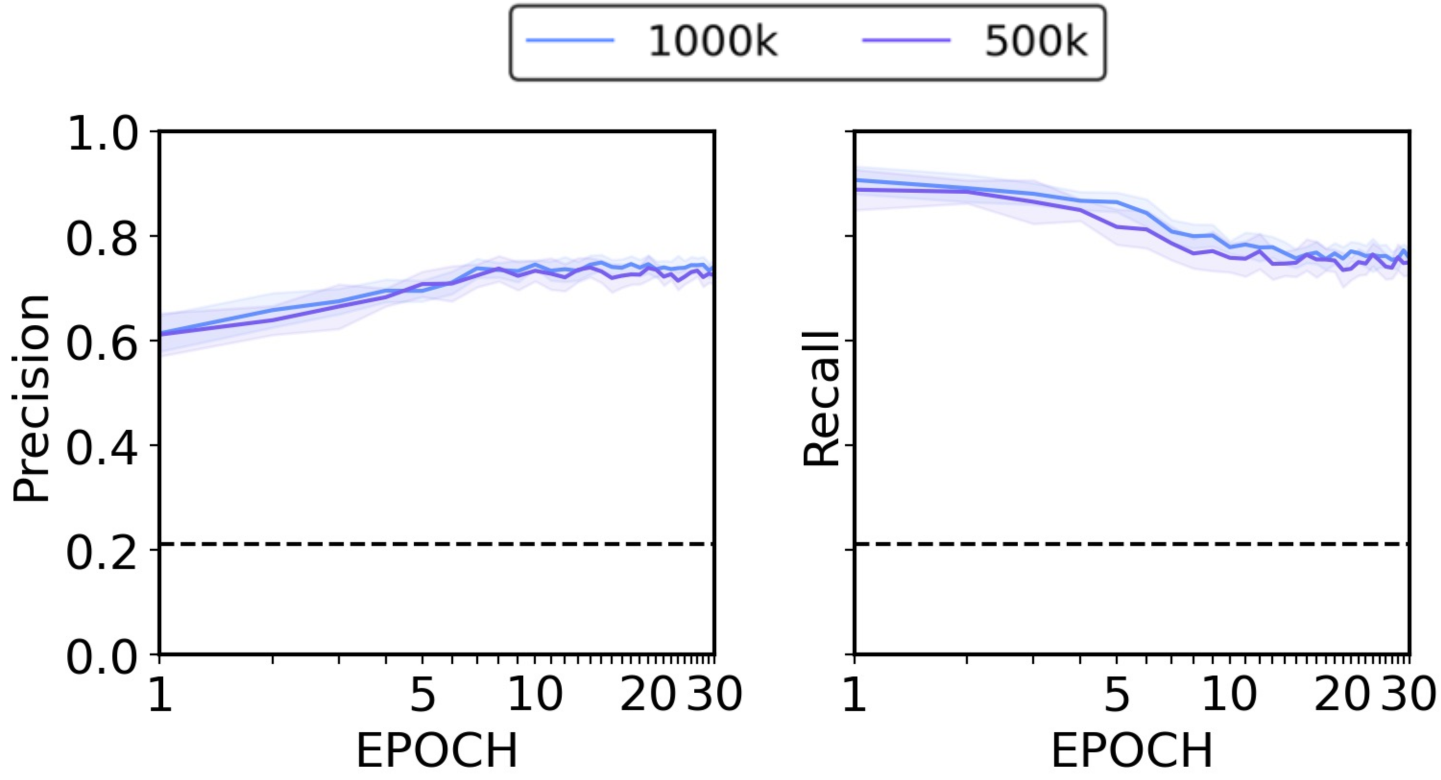
$$\text{recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

- Recall means “What fraction of the total number of events do you forecast”.

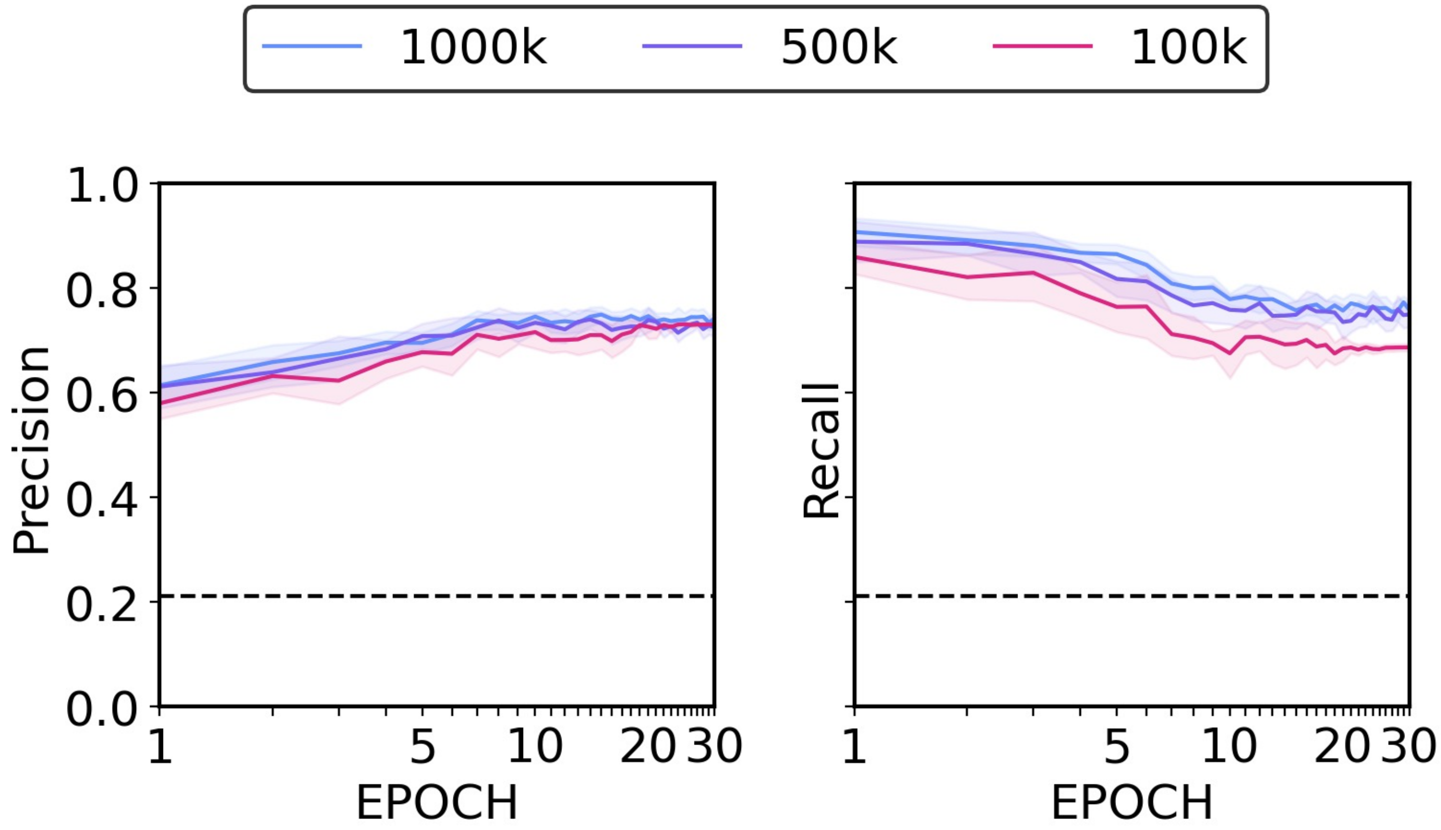
CNN training result



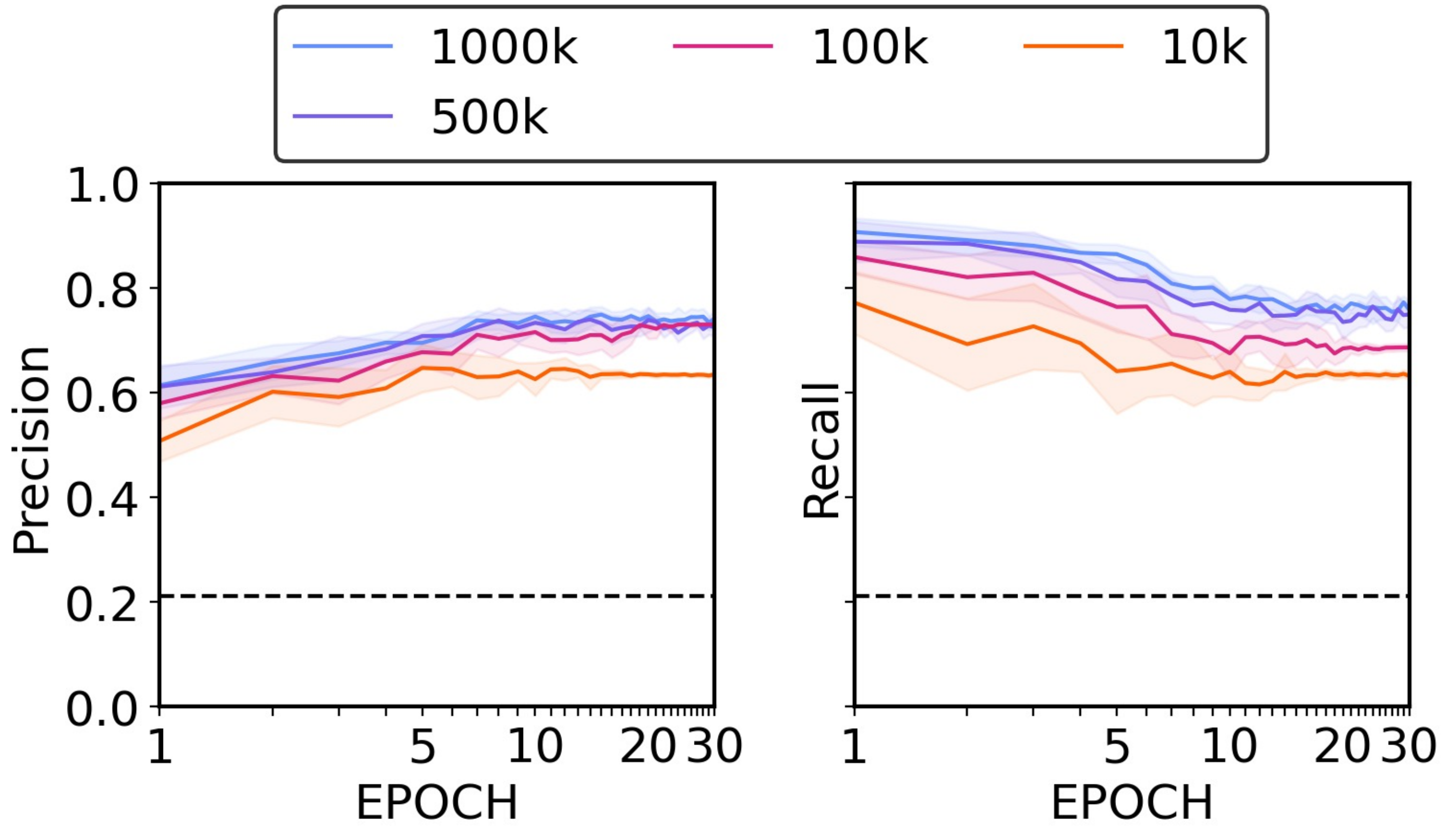
CNN training result



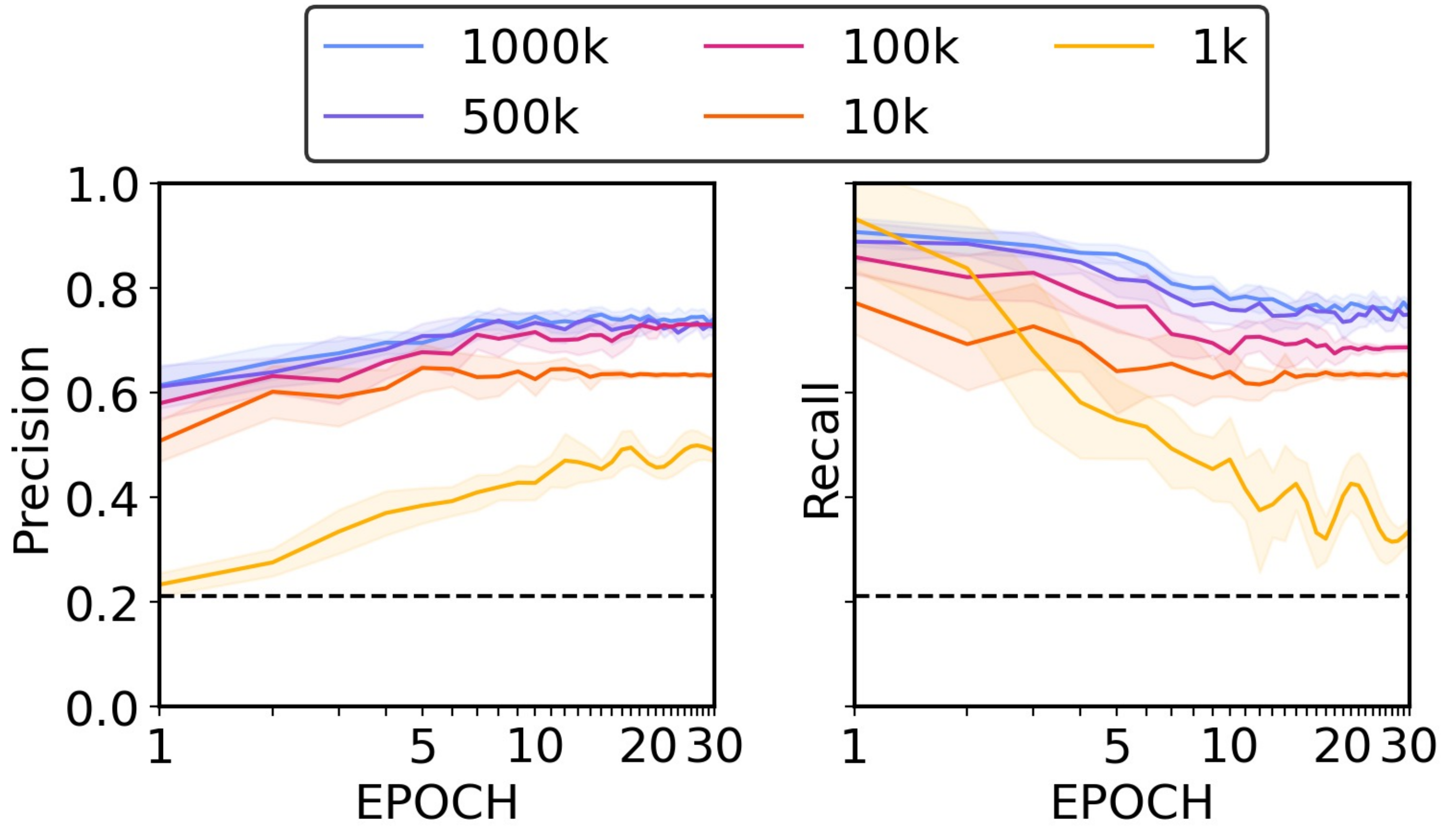
CNN training result



CNN training result



CNN training result



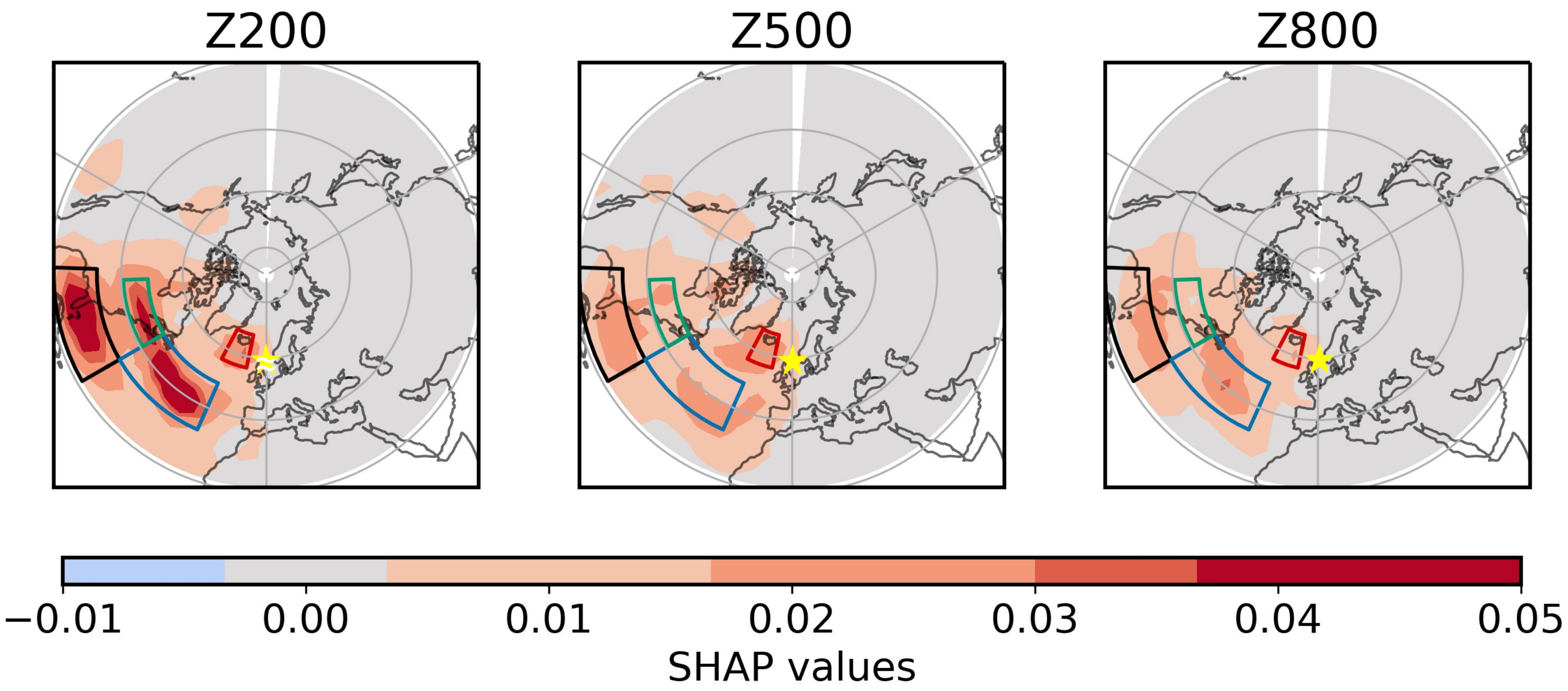
SHapley Additive exPlanation (SHAP) values

- Intuitively, given a function $p(\boldsymbol{x})$, SHAP assigns an importance value ϕ_i to each feature x_i of the argument $\boldsymbol{x} \in \mathbb{R}^d$:

$$p(\boldsymbol{x}) = \mathbb{E}[p(\boldsymbol{x})] + \sum_{i=1}^{d} \phi_i(\boldsymbol{x}).$$

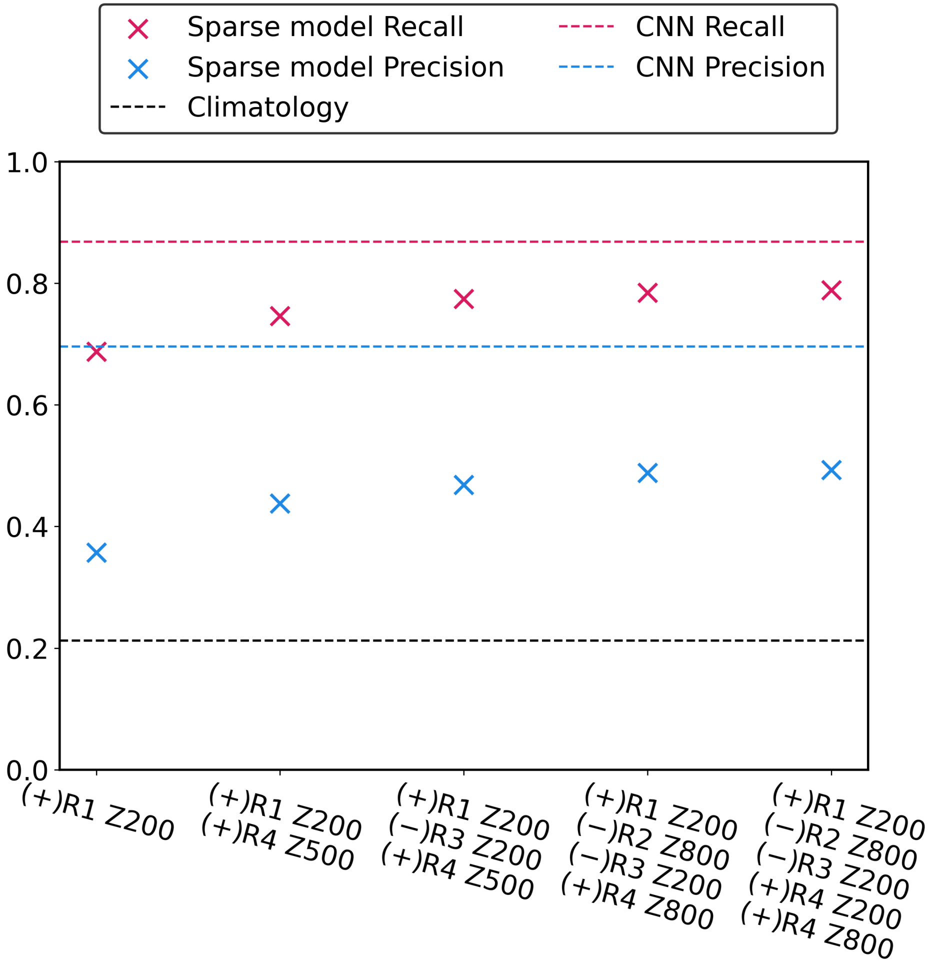
- SHAP values quantify how much is gained by incorporating information from each component x_i

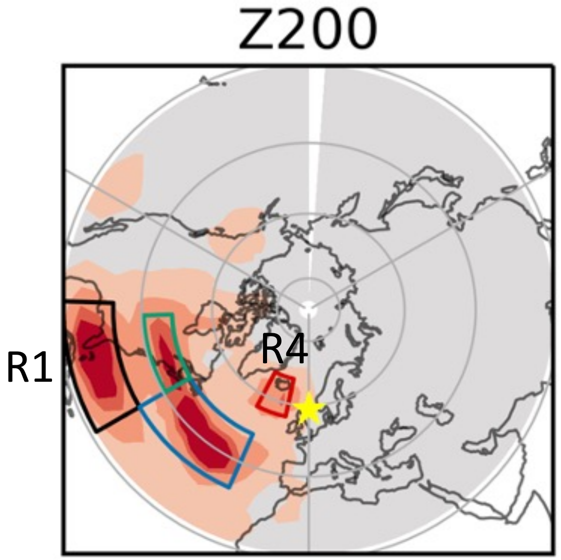
SHAP values can select important features



Learned features + Logistic regression can recover predictive skills

$$p(\boldsymbol{x}) = \frac{1}{1 + e^{-(\boldsymbol{k}\cdot\boldsymbol{x}+b)}}$$

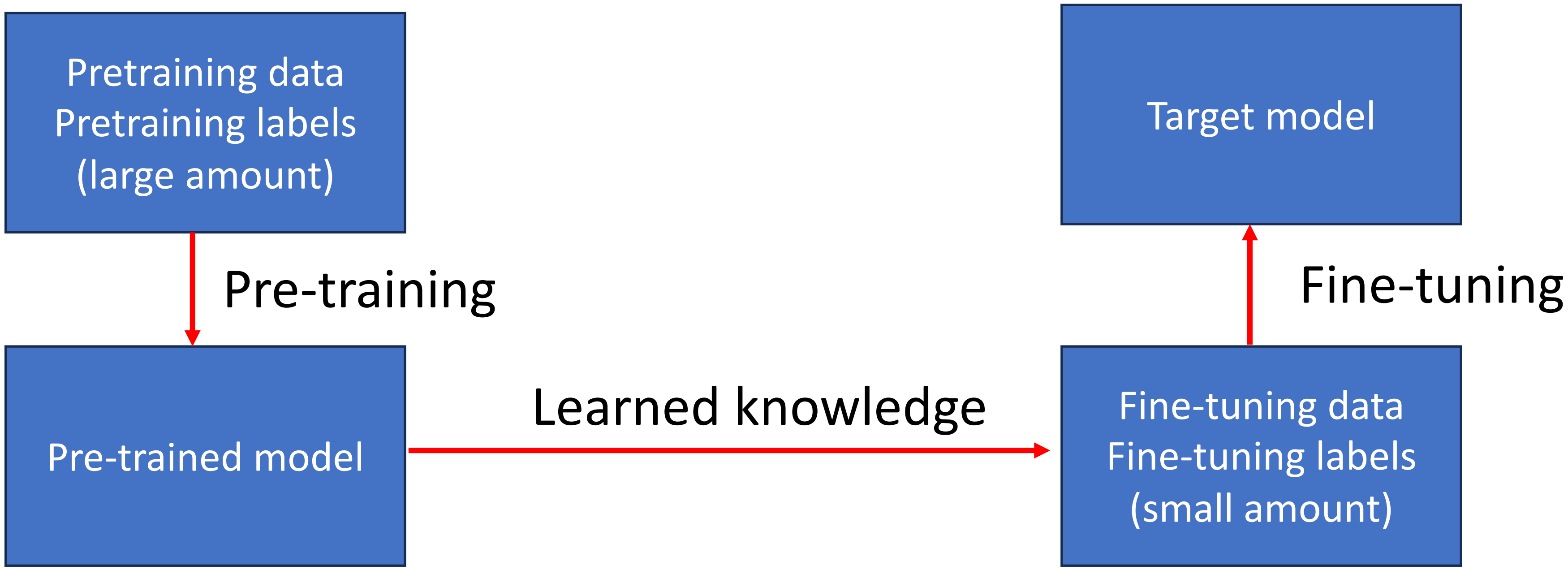




Question 2
Can we make the prediction with events from the observational record (ERA5)?

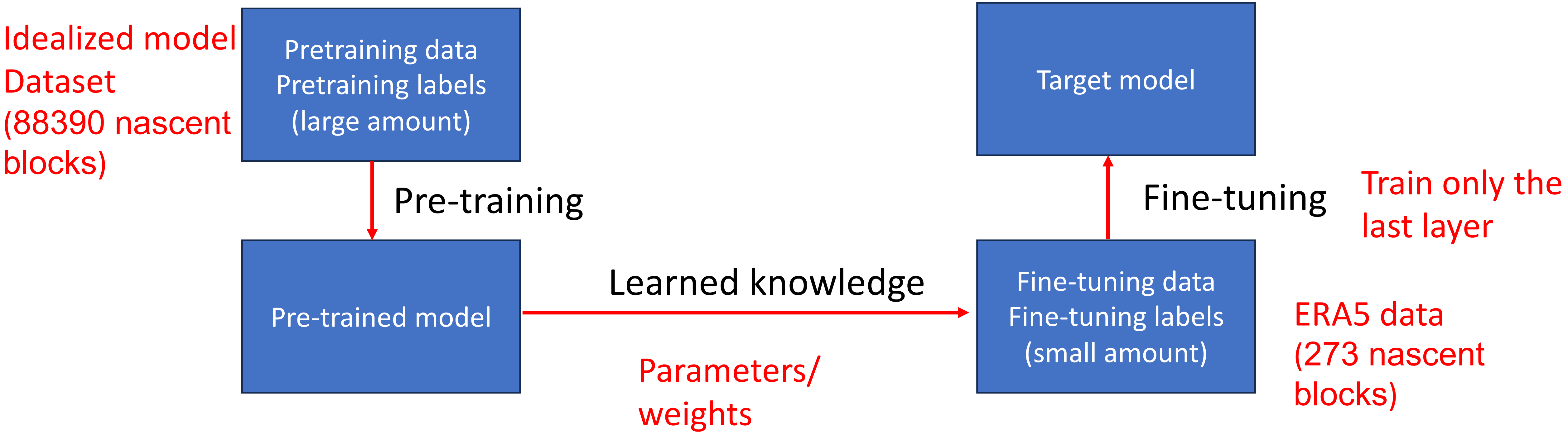
Transfer learning

- The model takes knowledge gained from solving one problem and applies it to a different but related problem.



Transfer learning

- The model takes knowledge gained from solving one problem and applies it to a different but related problem.



Transfer learning outperforms direct training

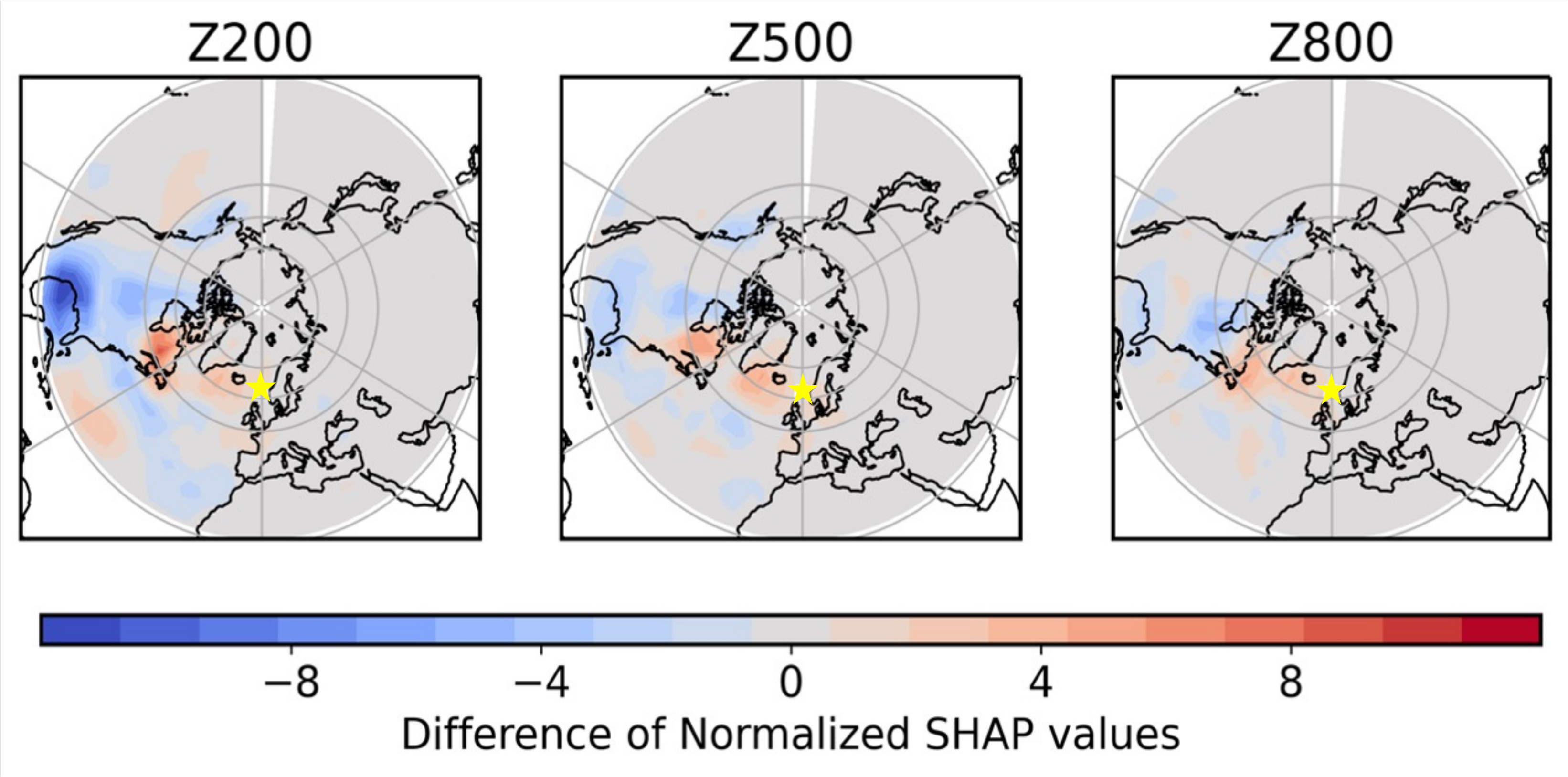
5-day blocking events forecast

	Climatology	Direct Training (DT)	Transfer Learning (TL)
Best Precision	0.31	0.45	0.45
Best Recall	0.31	0.61	0.82

7-day blocking events forecast

	Climatology	Direct Training (DT)	Transfer Learning (TL)
Best Precision	0.07	0.21	0.22
Best Recall	0.08	0.48	0.76

What has transfer learning learned by fine-tuning on the ERA5 dataset?



Two questions

- Can a data-driven approach predict persistent blocking events? If so, how?

  Yes! Explainable AI (SHAP values) identify important regions upstream of the blocking that indicate it will persist. The upstream wave train is consistent with our synoptic understanding of blocking, but the CNN gains greater precision from more subtle features.

- Can we do this within the limits of the observational record?

  Yes (to some degree)! Transfer learning allows us to combine features learned from an atmospheric model with limited events in ERA5, making more robust and accurate predictions for extreme events than possible with direct training alone.

Thanks!

Marshall-Molteni model

Marshall, J., and F. Molteni, 1993: Toward a Dynamical Understanding of Planetary-Scale Flow Regimes. *J. Atmos. Sci.*, **50**, 1792–1818

- 3-layer quasi-geostrophic potential vorticity equation

$$\partial_t q_j + J(\psi_j, q_j) = -D_j + S_j, j = 1, 2, 3$$

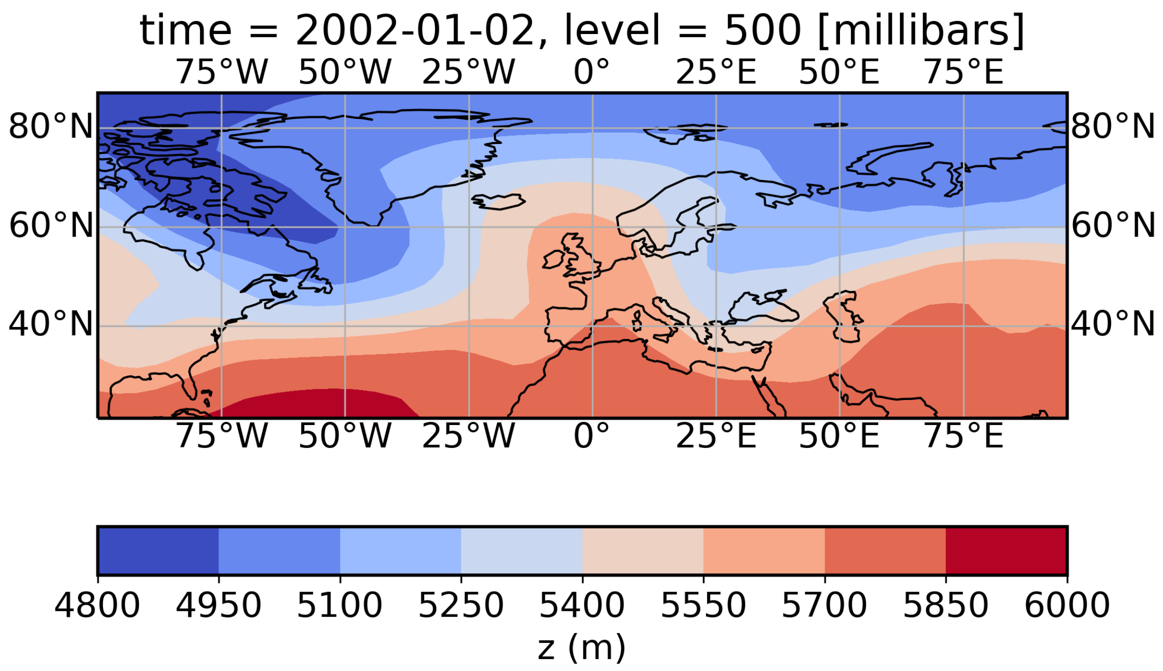
- $J(A, B) = A_x B_y - A_y B_x$ is the Jacobian operator.
- $-D_j$ is the operator defining the radiative damping, surface friction, and hyper diffusion: all relax model to state of rest
- S_j is the forcing, computed from the observed data to give the model a realistic mean state:

$$S_j = \overline{J(\psi_j, q_j)} + \overline{D}_j$$

Blocking events

Despite its importance, blocking events are not predicted well!

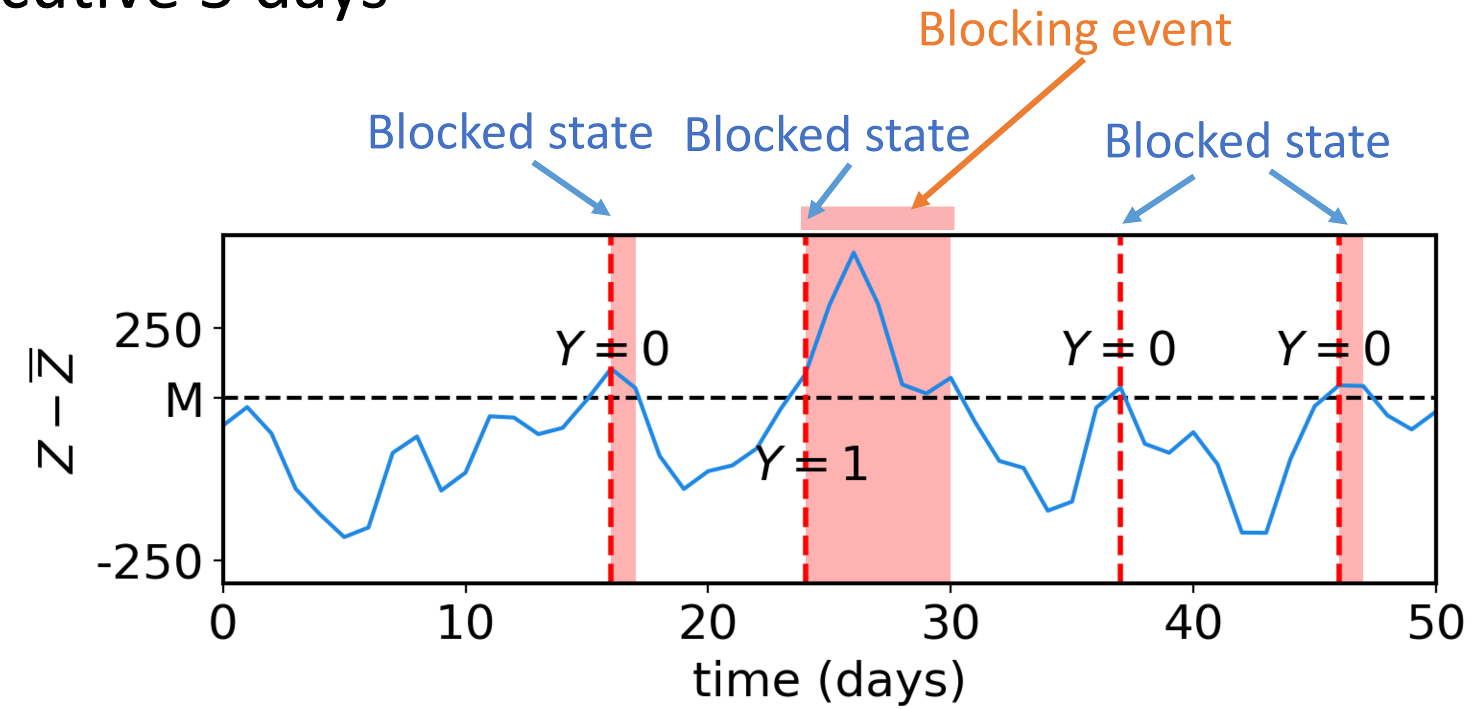
- High-amplitude, quasi-stationary anticyclonic high-pressure anomalies
- Leads to regional extreme weather



Aftermath of wildfires near Moscow, 4 August 2010. Credit: ITAR-TASS News Agency / Alamy Stock Photo.

Criteria for the blocking events

- Key ingredients: high pressure + persistence
- Dole and Gordon criteria: Z500 anomalies > M(e.g. 150m) for at least consecutive 5 days



Composite maps for Marshall Molteni