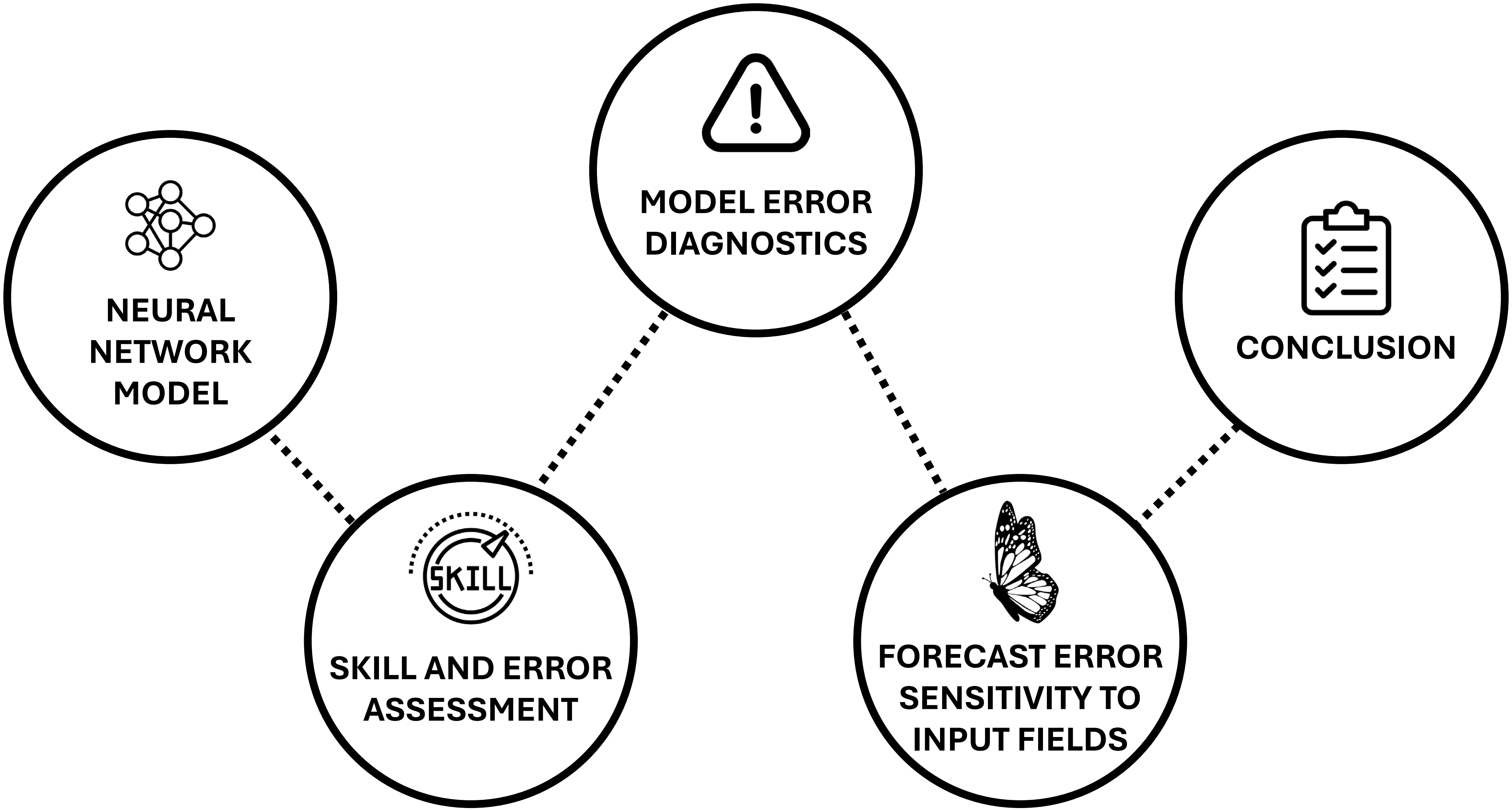
Global deep-learning weather prediction model evaluation and error diagnostics

Uroš Perkan[1]

Žiga Zaplotnik[2]

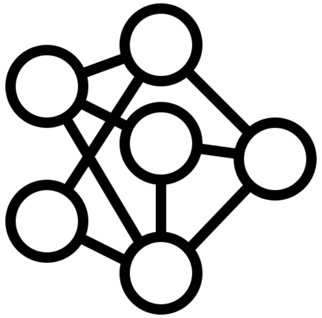
[1]University of Ljubljana
[2]ECWMF

UNIVERSITY OF LJUBLJANA | FMF Faculty of Mathematics and Physics



NEURAL NETWORK MODEL
SKILL
SKILL AND ERROR ASSESSMENT
MODEL ERROR DIAGNOSTICS
FORECAST ERROR SENSITIVITY TO INPUT FIELDS
CONCLUSION

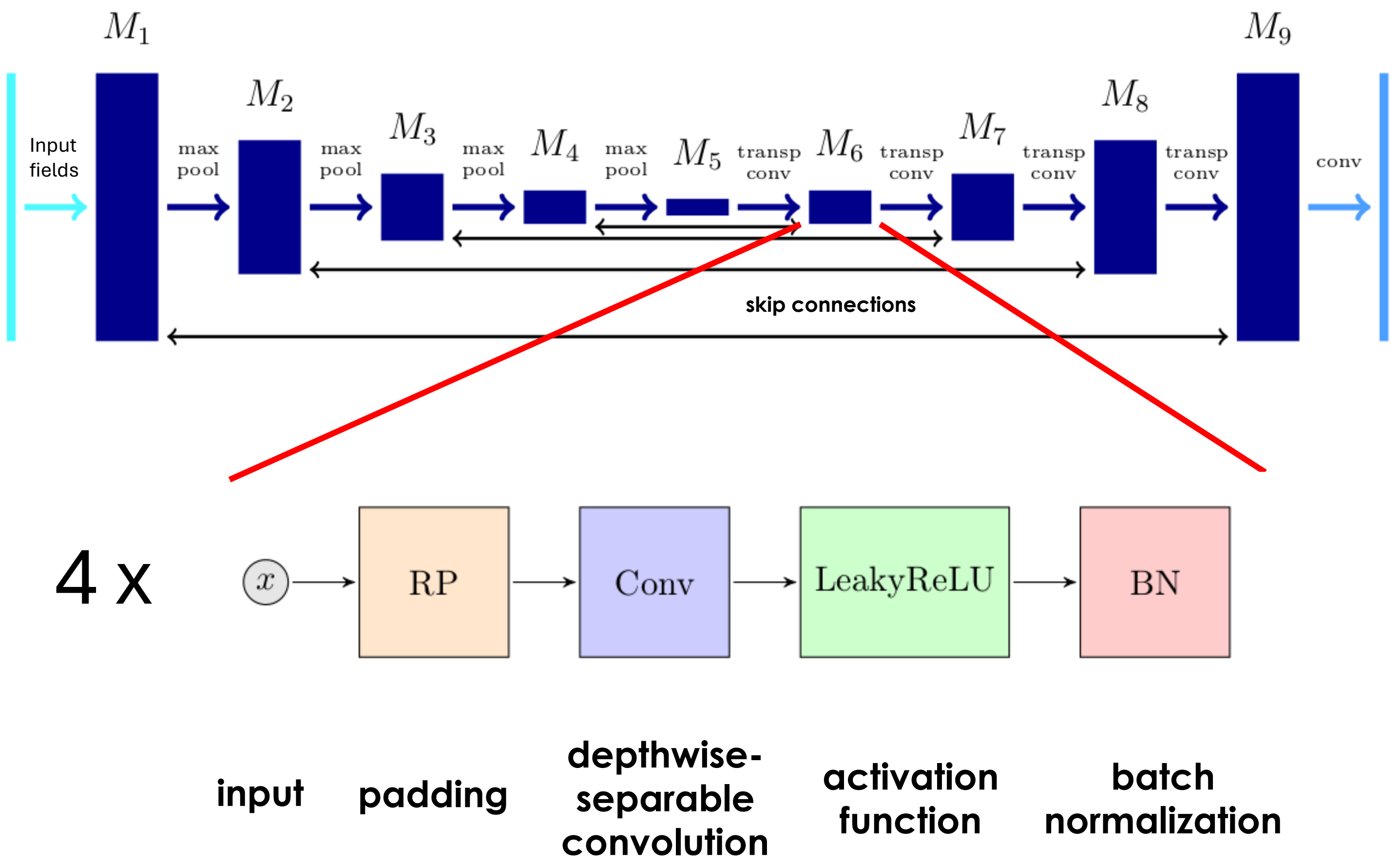
NEURAL NETWORK MODEL



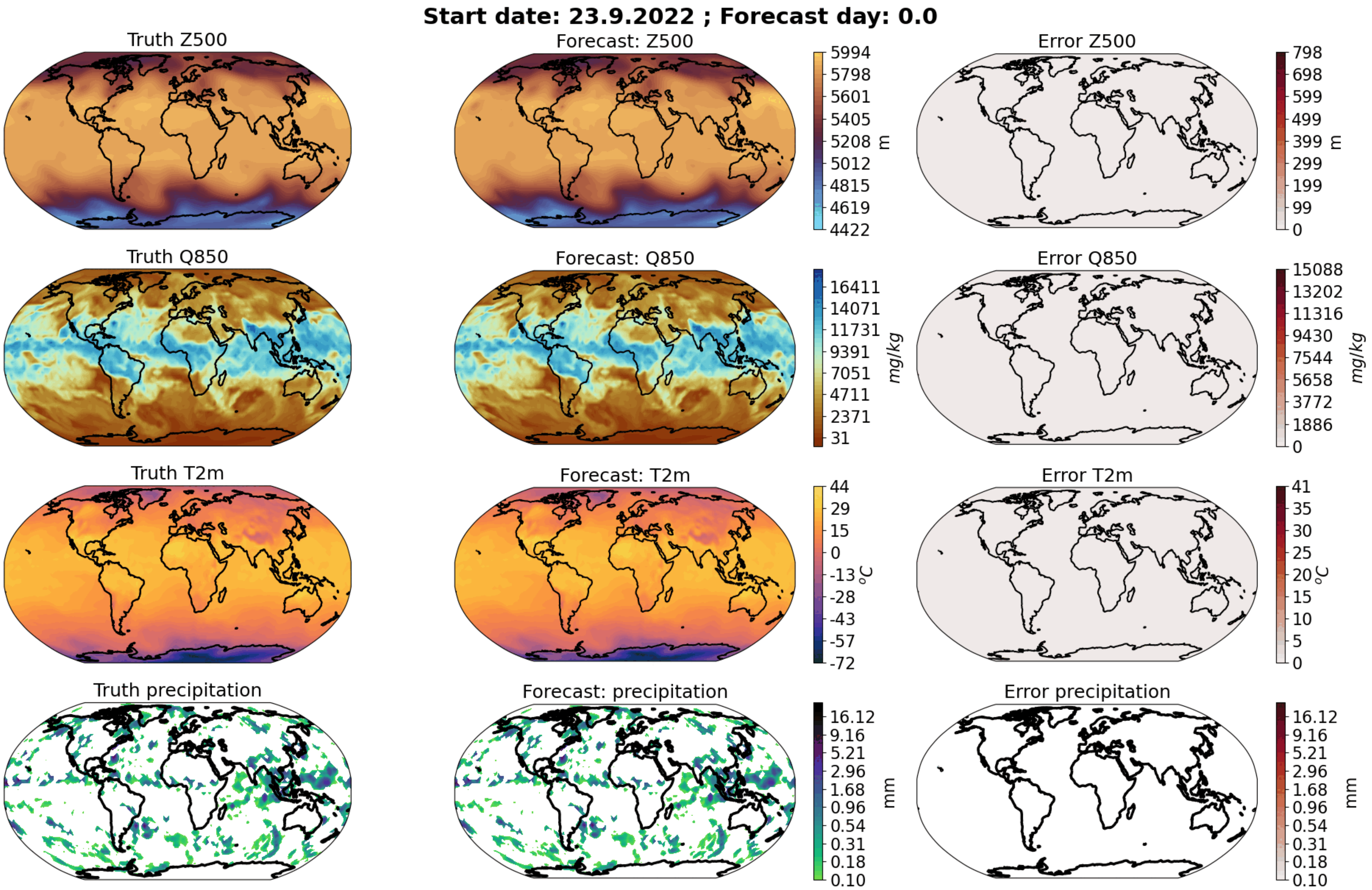
NEURAL NETWORK MODEL

ConvCastNet: convolutional weather forecasting neural network

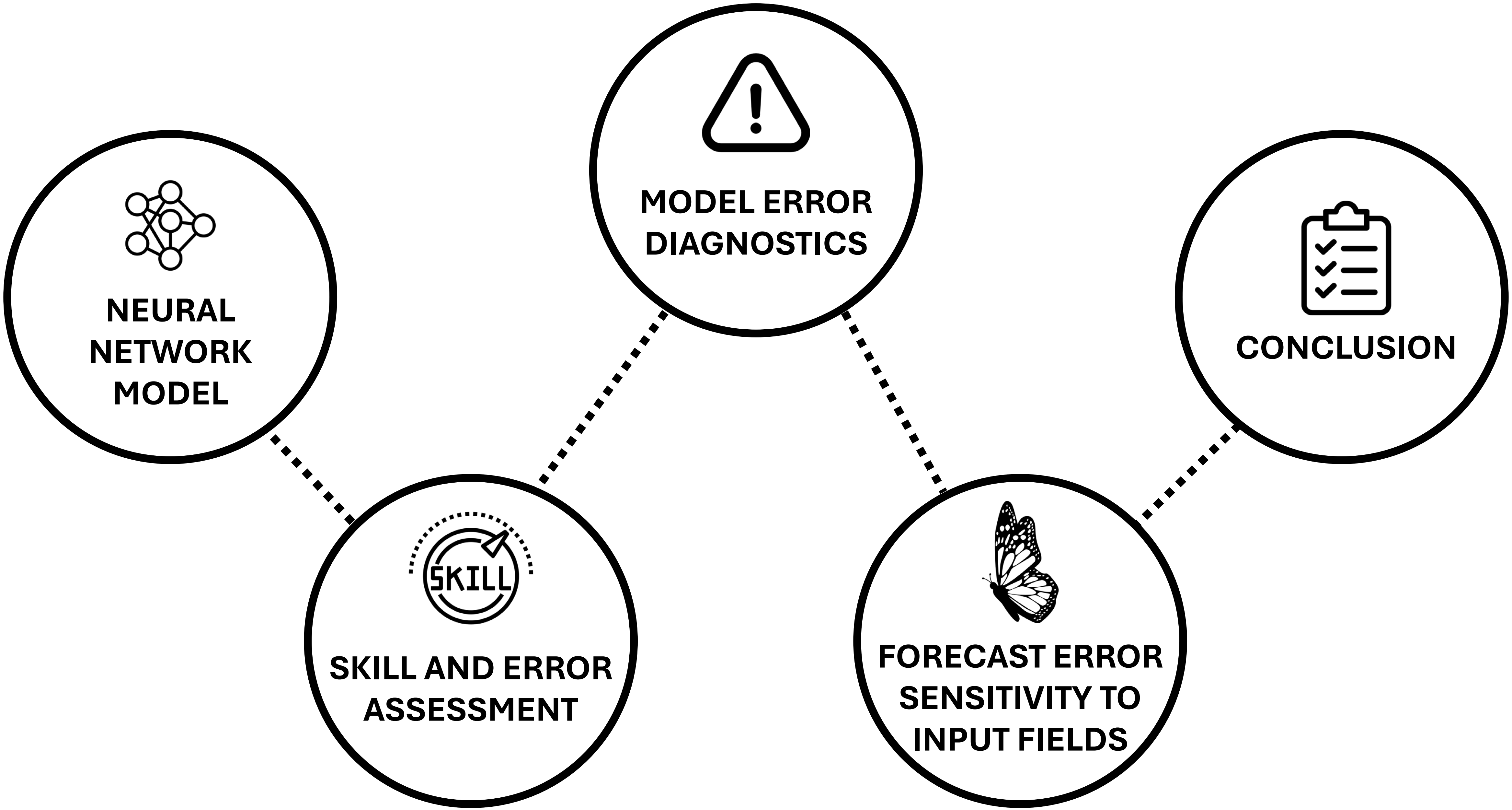
MODEL ARCHITECTURE



- Training data ERA5 reanalysis:
  - Train: 1970 – 2014
  - Validation: 2015 – 2019
  - Test: 2020 - 2022
- Model resolution: 3º
- Variables
  - Z, u, v, ω, T, q at 13 pressure levels
  - T2m, u10, v10, land and ice surface temperature, SST, mslp, tp
  - Sea ice, snow depth, soil moisture, top-of-atmosphere solar radiation, …
- Training: 4 autoregressive steps
- Time step: 12 hr



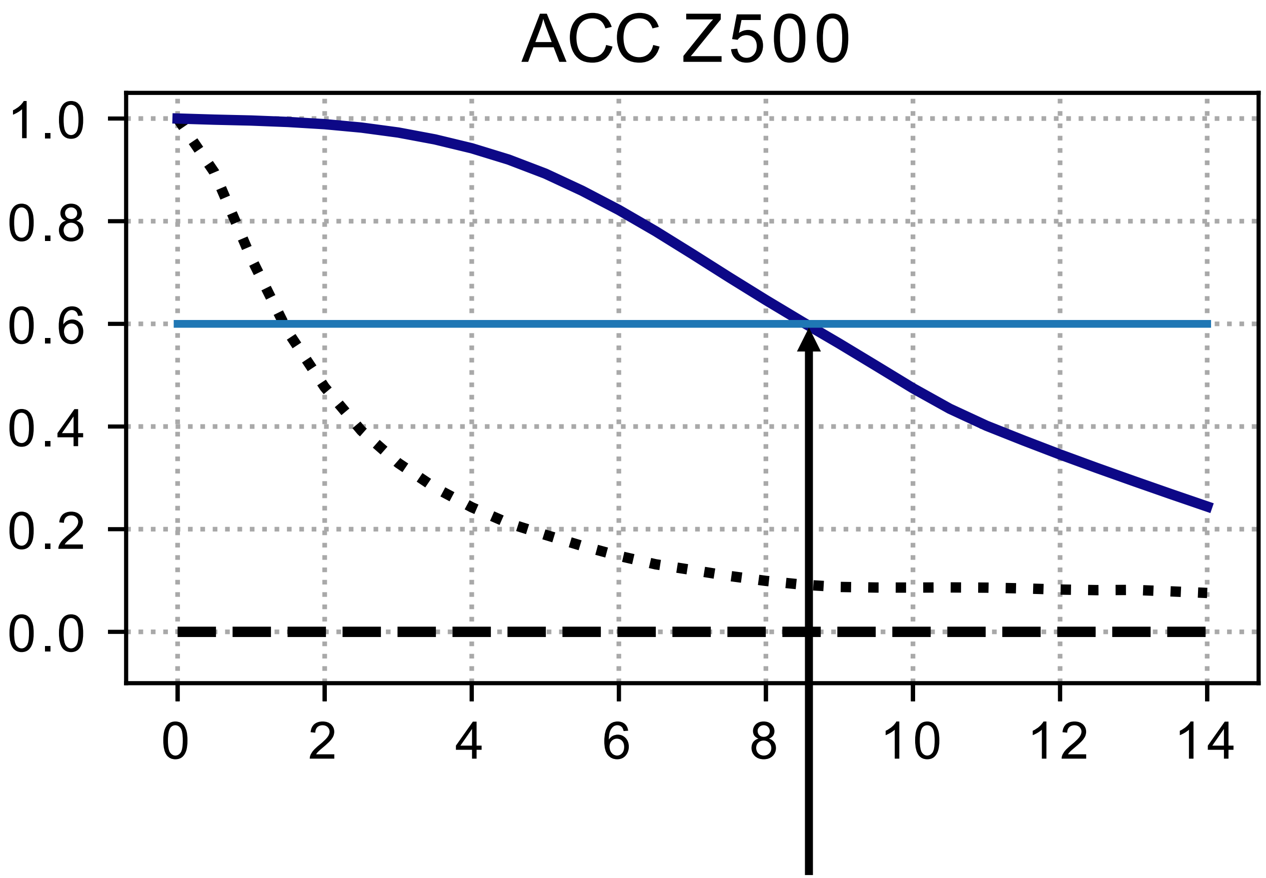
Start date: 23.9.2022 ; Forecast day: 0.0
Truth Z500
Forecast: Z500
Error Z500
5994
5798
5601
5405
5208
5012
4815
4619
4422
m
798
698
599
499
399
299
199
99
0
m
Truth Q850
Forecast: Q850
Error Q850
16411
14071
11731
9391
7051
4711
2371
31
mg/kg
15088
13202
11316
9430
7544
5658
3772
1886
0
mg/kg
Truth T2m
Forecast: T2m
Error T2m
44
29
15
0
-13
-28
-43
-57
-72
°C
41
35
30
25
20
15
10
5
0
°C
Truth precipitation
Forecast: precipitation
Error precipitation
16.12
9.16
5.21
2.96
1.68
0.96
0.54
0.31
0.18
0.10
mm
16.12
9.16
5.21
2.96
1.68
0.96
0.54
0.31
0.18
0.10
mm


NEURAL NETWORK MODEL
SKILL
SKILL AND ERROR ASSESSMENT
MODEL ERROR DIAGNOSTICS
FORECAST ERROR SENSITIVITY TO INPUT FIELDS
CONCLUSION

SKILL AND ERROR ASSESSMENT

SKILL AND ERROR ASSESSMENT

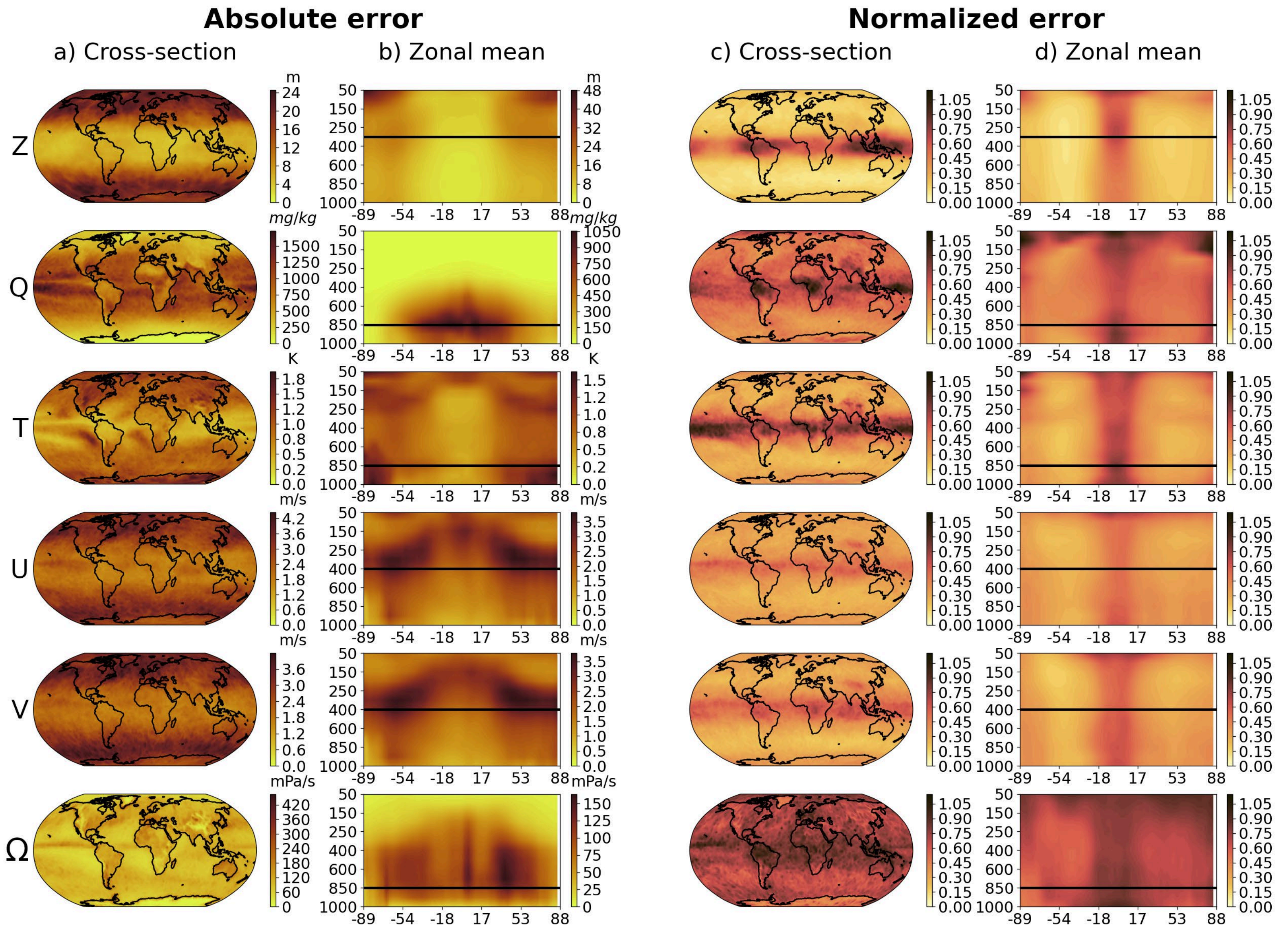
GLOBAL SKILL



ACC > 0.6 for 8.5 days

Large-scale forecast skill close to state-of-the-art ML models

SPATIAL DISTRIBUTION OF ERRORS

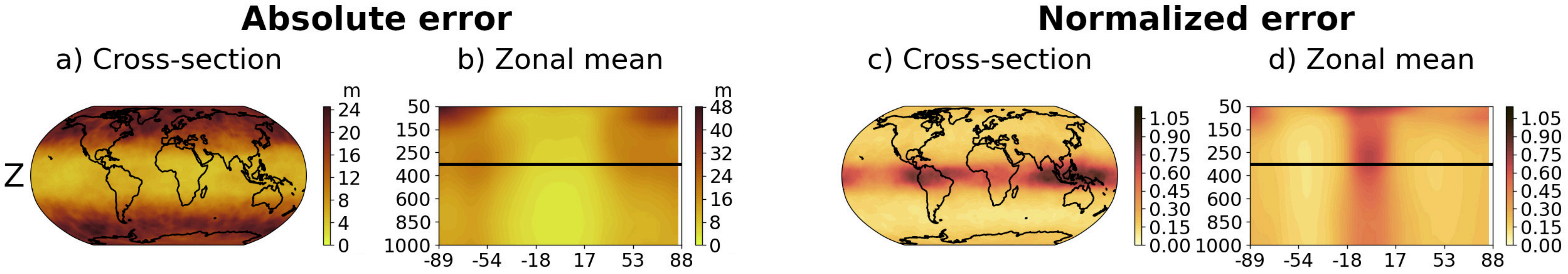


Lead time: 2 days

Absolute error:
absolute difference between forecast and ERA5 „truth“

Normailzed error:
absolute error divided by natural variability

GEOPOTENTIAL



Large absolute errors:

- Mid-to-high latitudes
- Polar stratosphere

Correspond to large natural variability

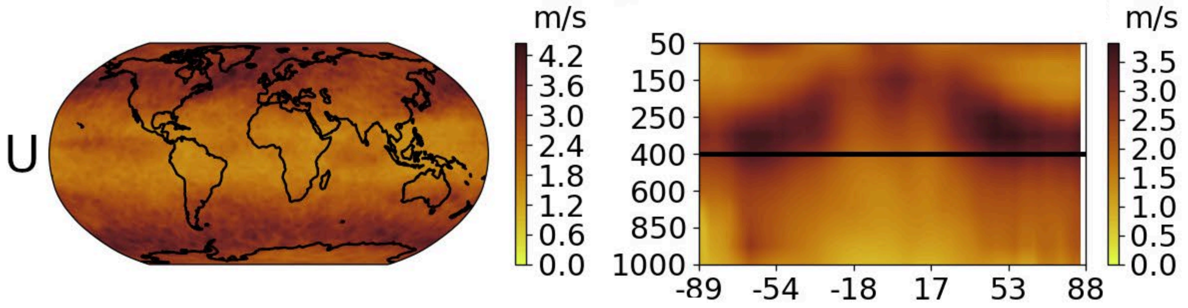
Large normalized errors:

- Tropics
- Polar and tropical stratosphere

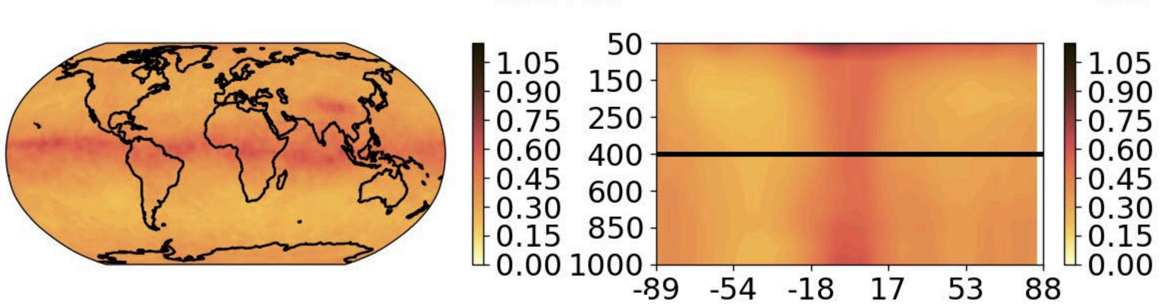
Better describes error relative to expected natural weather variability

ZONAL WIND

Absolute error



Normalized error



Large absolute errors:

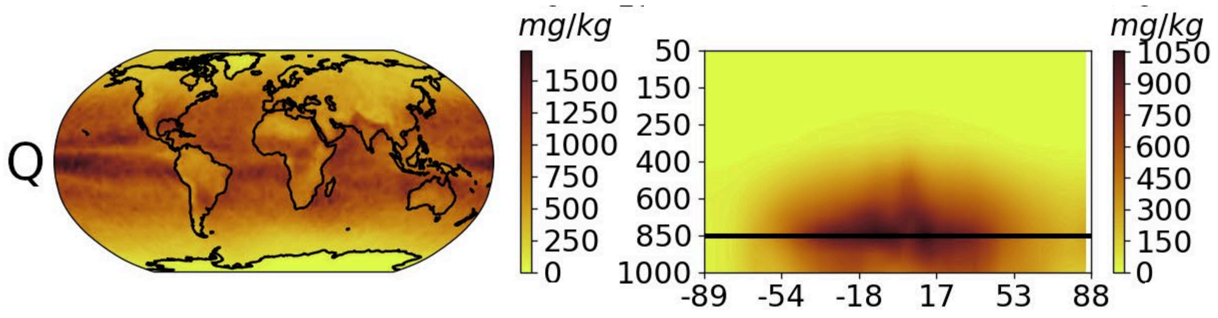
- Region of strong mid-latitude westerly jet
- Stratosphere

Large normalized errors:

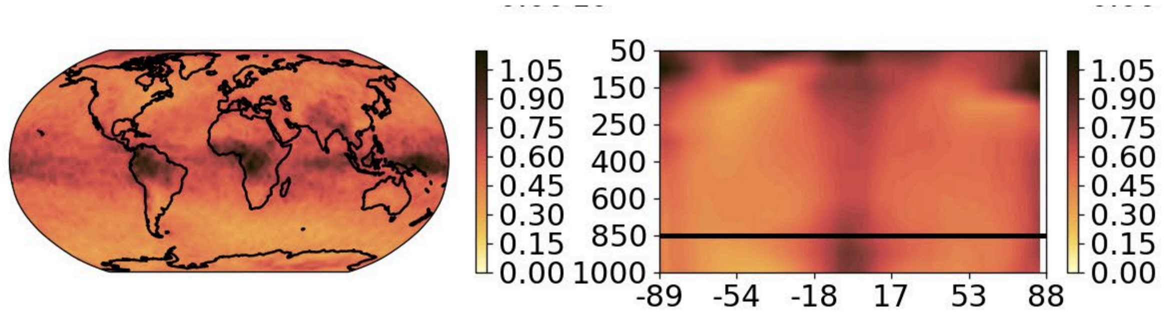
- Tropical atmosphere
- Tibetan Plateau

SPECIFIC HUMIDITY

Absolute error



Normalized error

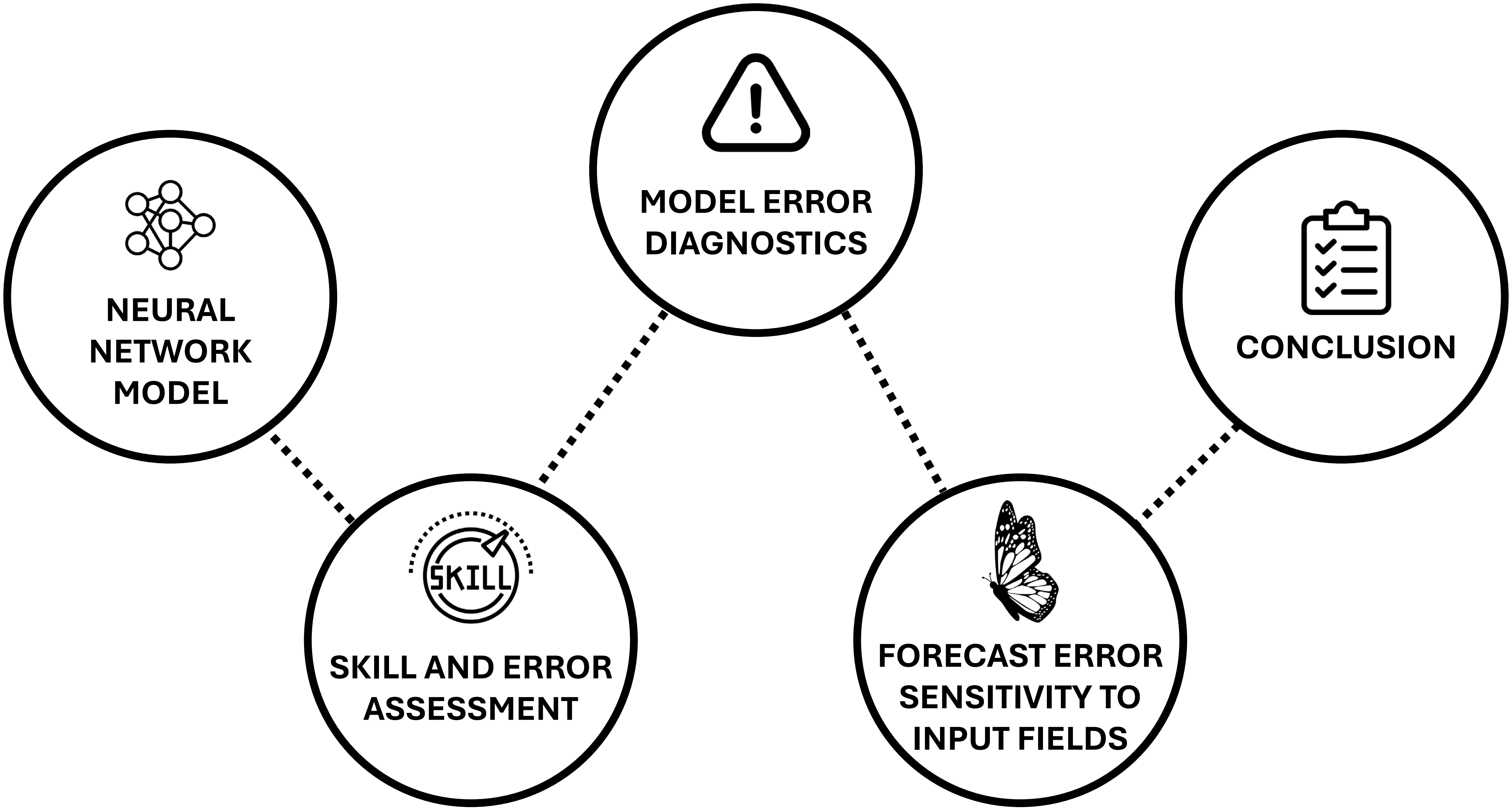


Large absolute errors:

- Above oceans
- Lower tropospere

Large normalised errors:

- Congo and Amazon basin
- Maritime continent
- Tropical oceans
- Stratosphere


NEURAL NETWORK MODEL
SKILL
SKILL AND ERROR ASSESSMENT
MODEL ERROR DIAGNOSTICS
FORECAST ERROR SENSITIVITY TO INPUT FIELDS
CONCLUSION

MODEL ERROR DIAGNOSTICS

MODEL ERROR DIAGNOSTICS

MODEL ERROR DIAGNOSTICS

- How does the forecast skill improve if we continuously replace forecasted fields with ERA5 „truth“ in:
  1. Tropics?
  2. Stratosphere?
- We perform the analysis by replacing forecasted fields in the tropics/stratosphere with ERA5 „truth“:

autoregressive input field

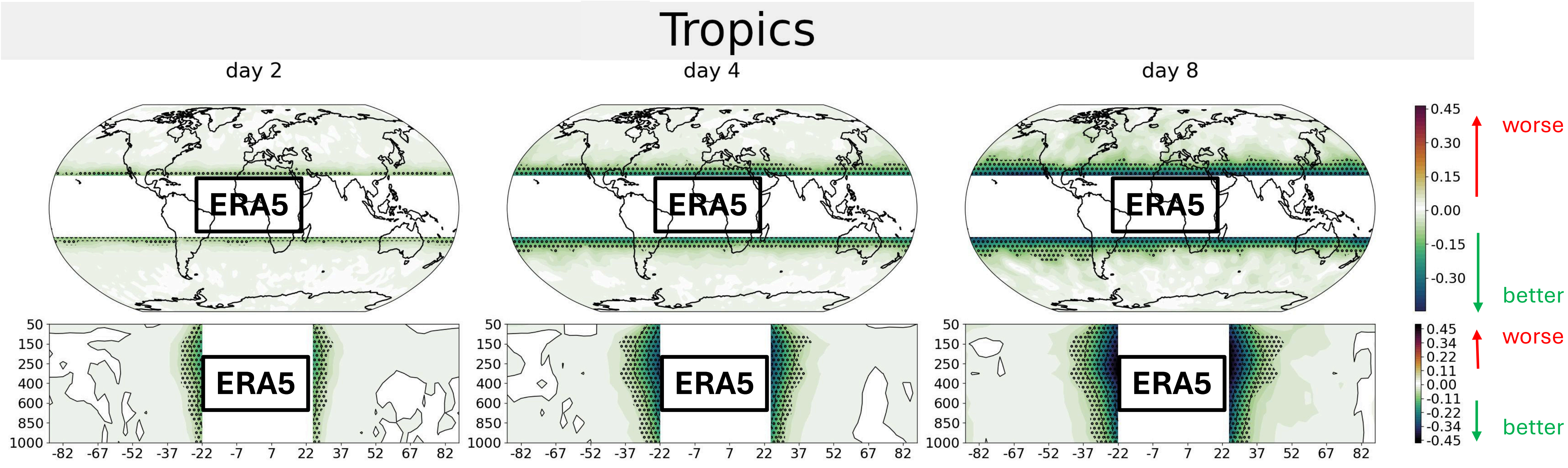
ERA5 „truth“ term

Model forecast term

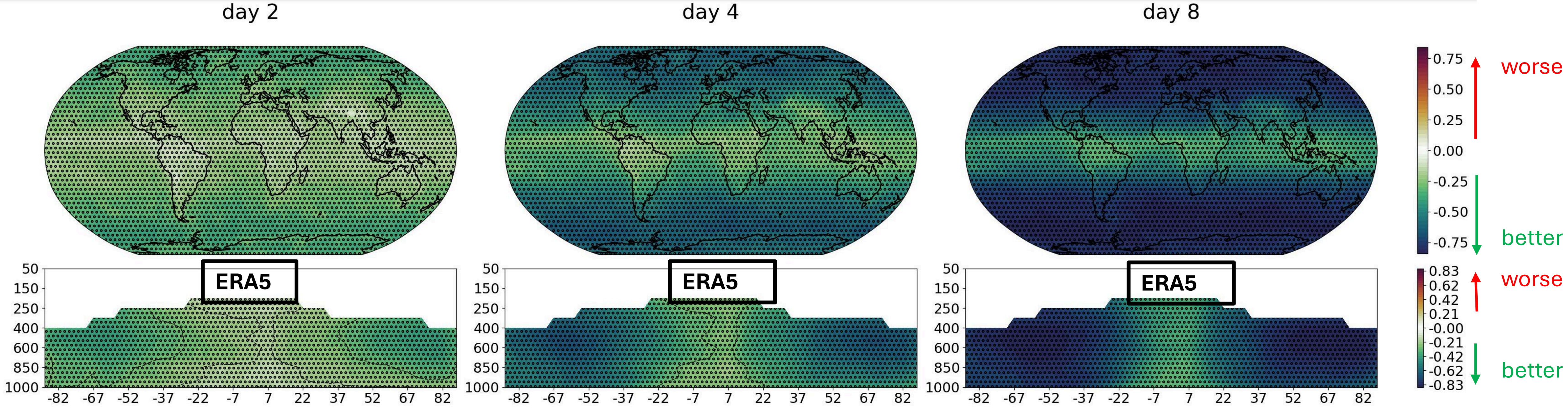
$$x(\lambda, \varphi, p, t + \Delta t) = \text{weight}_\tau(\lambda, \varphi, p) \cdot \text{ERA5}(\lambda, \varphi, p, t + \Delta t) + (1 - \text{weight}_\tau(\lambda, \varphi, p)) \cdot \mathcal{M}_{t \to t+\Delta t}(s(t))(\lambda, \varphi, p)$$

- We measure skill gain with the following metric: $1 - \dfrac{error(M(x))}{error(M(x_{forced}))}$

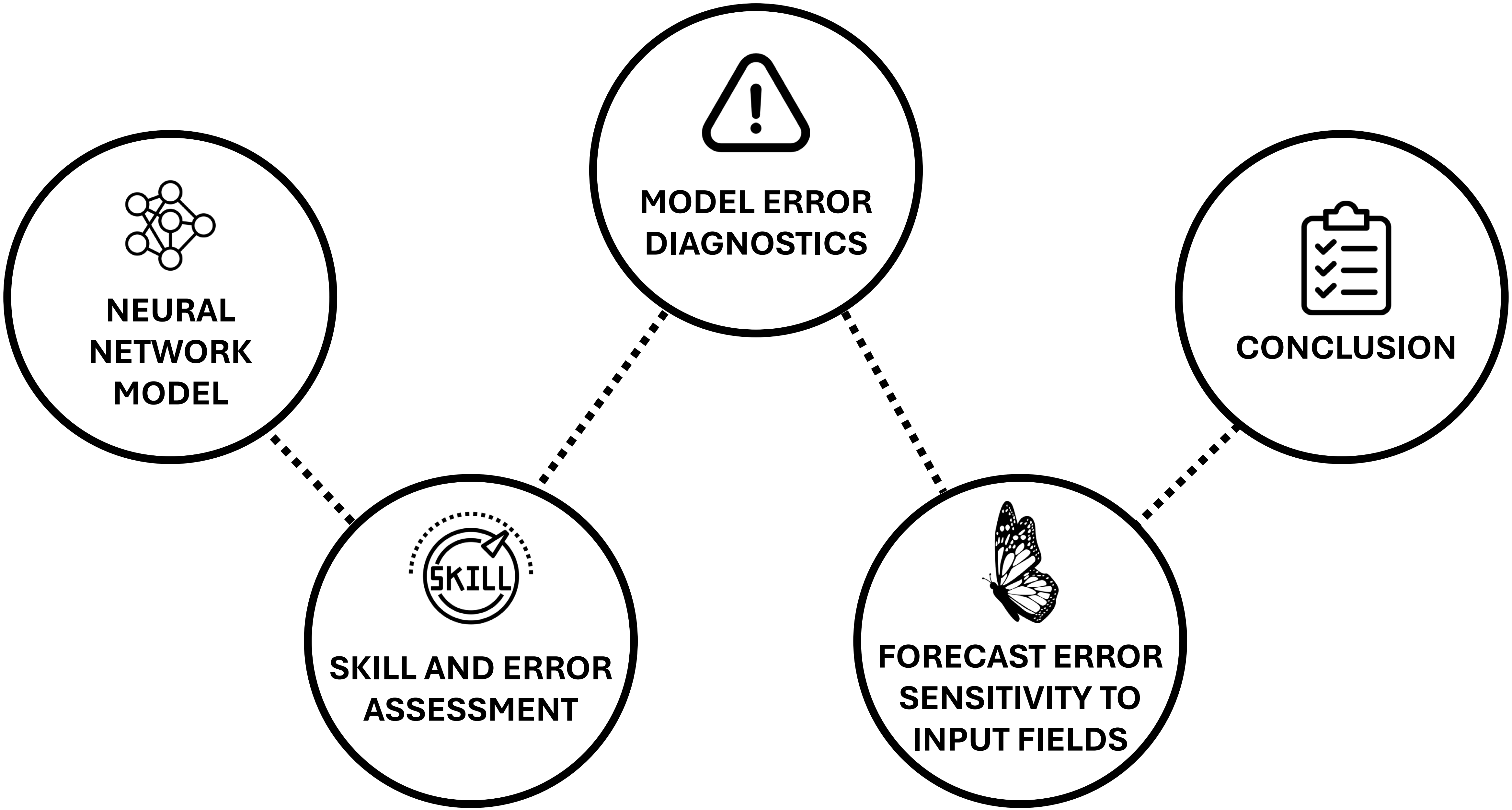
FORECAST SKILL IMPROVEMENT



Stratosphere



- This method easily diagnoses problematic regions in ML models
  - Very few issues regarding the boundary between ERA5 „truth“ and the model forecast
- We identify the stratosphere to be the key region needing better representation to improve model skill
- We could investigate other regions e.g. mountains or oceans as well


NEURAL NETWORK MODEL
SKILL
SKILL AND ERROR ASSESSMENT
MODEL ERROR DIAGNOSTICS
FORECAST ERROR SENSITIVITY TO INPUT FIELDS
CONCLUSION

FORECAST ERROR SENSITIVITY TO INPUT FIELDS

FORECAST ERROR SENSITIVITY TO INPUT FIELDS

FORECAST ERROR SENSITIVITY TO INPUT FIELDS

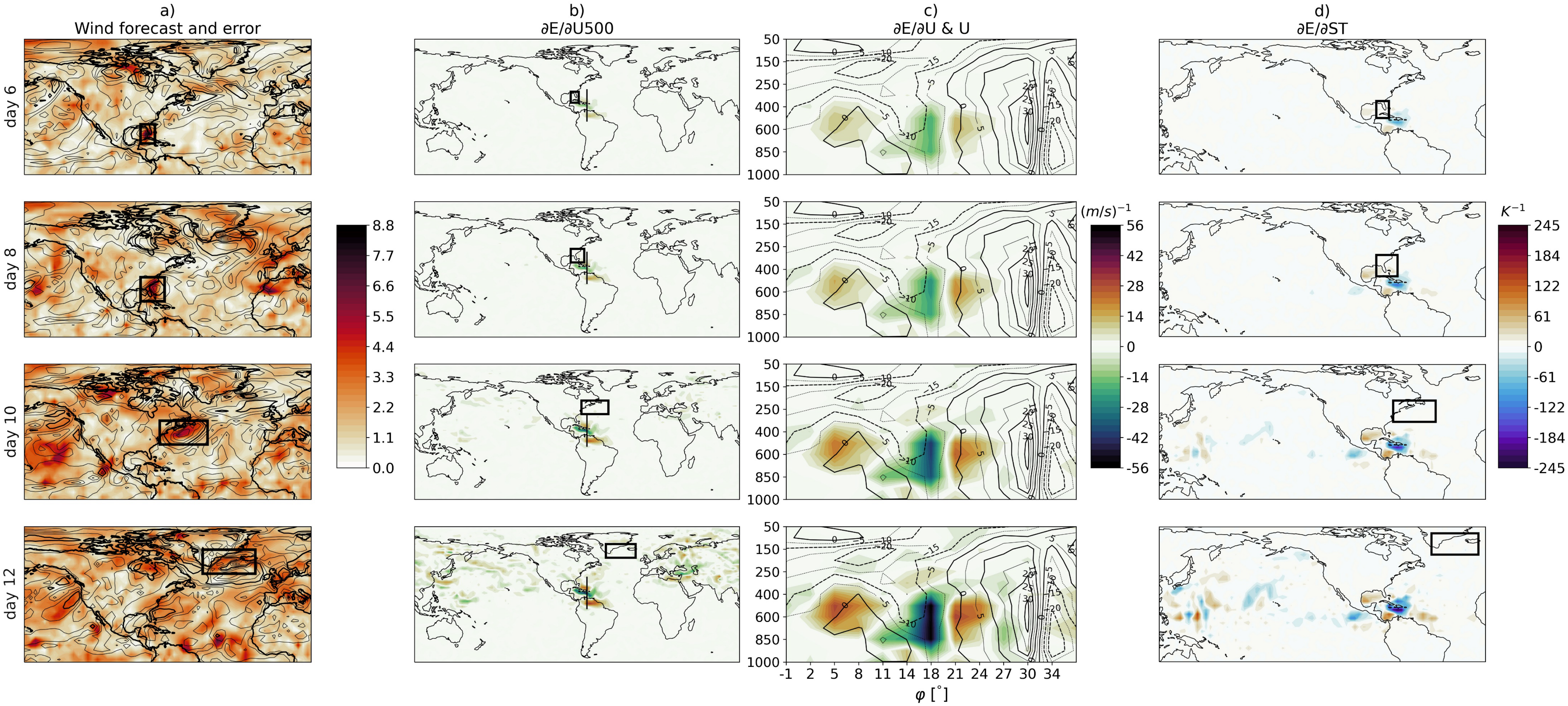
- Calculate the sensitivity of the forecast errors to the input fields:

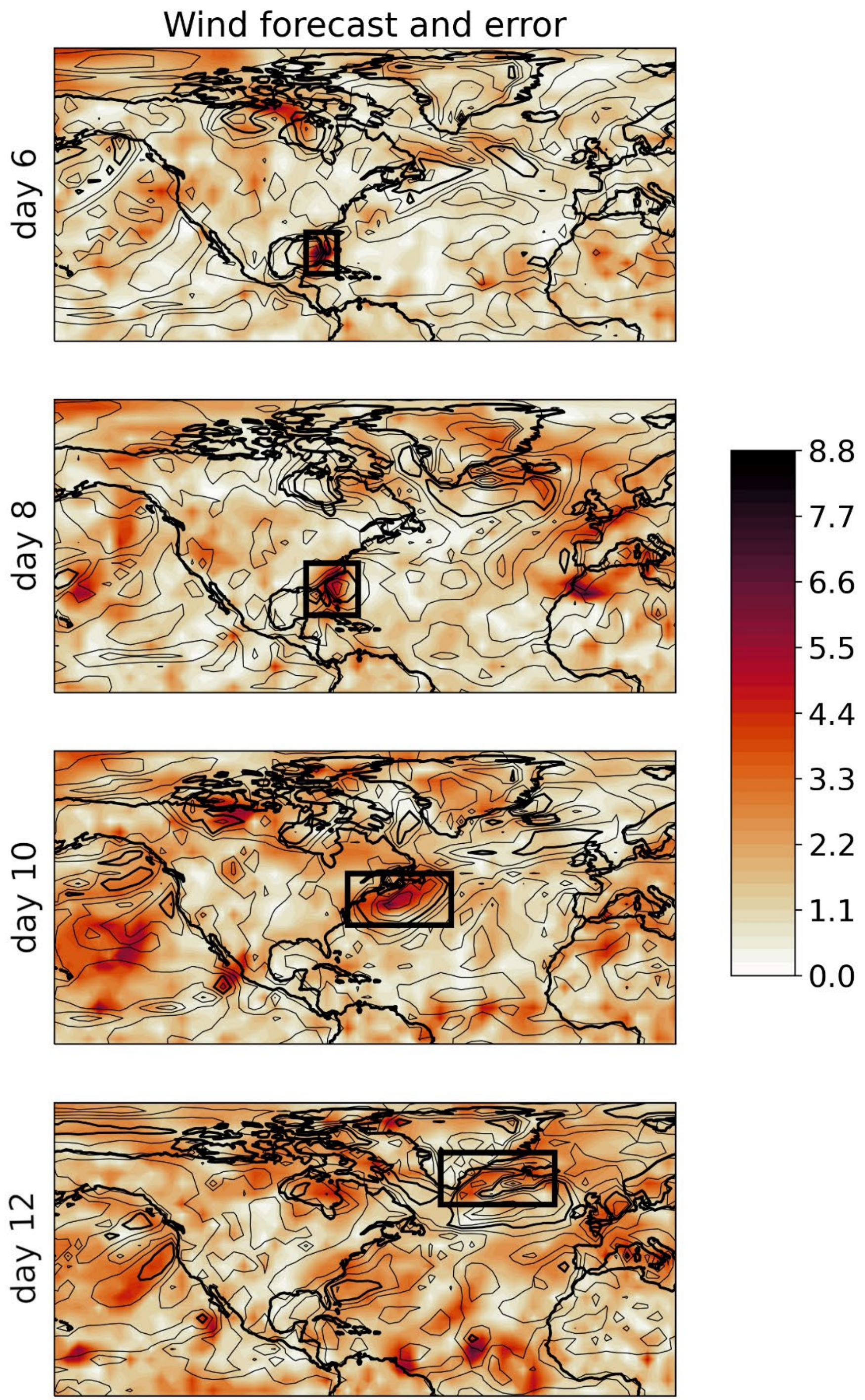
  => derivative of forecast error with respect to input fields: $\frac{\partial E}{\partial (IC)}$

  - ML models are auto-differentiable → error backpropagation

- Can it be used to improve weather forecasts?

HURRICANE IAN: forecast initialisation – September 23, 2022

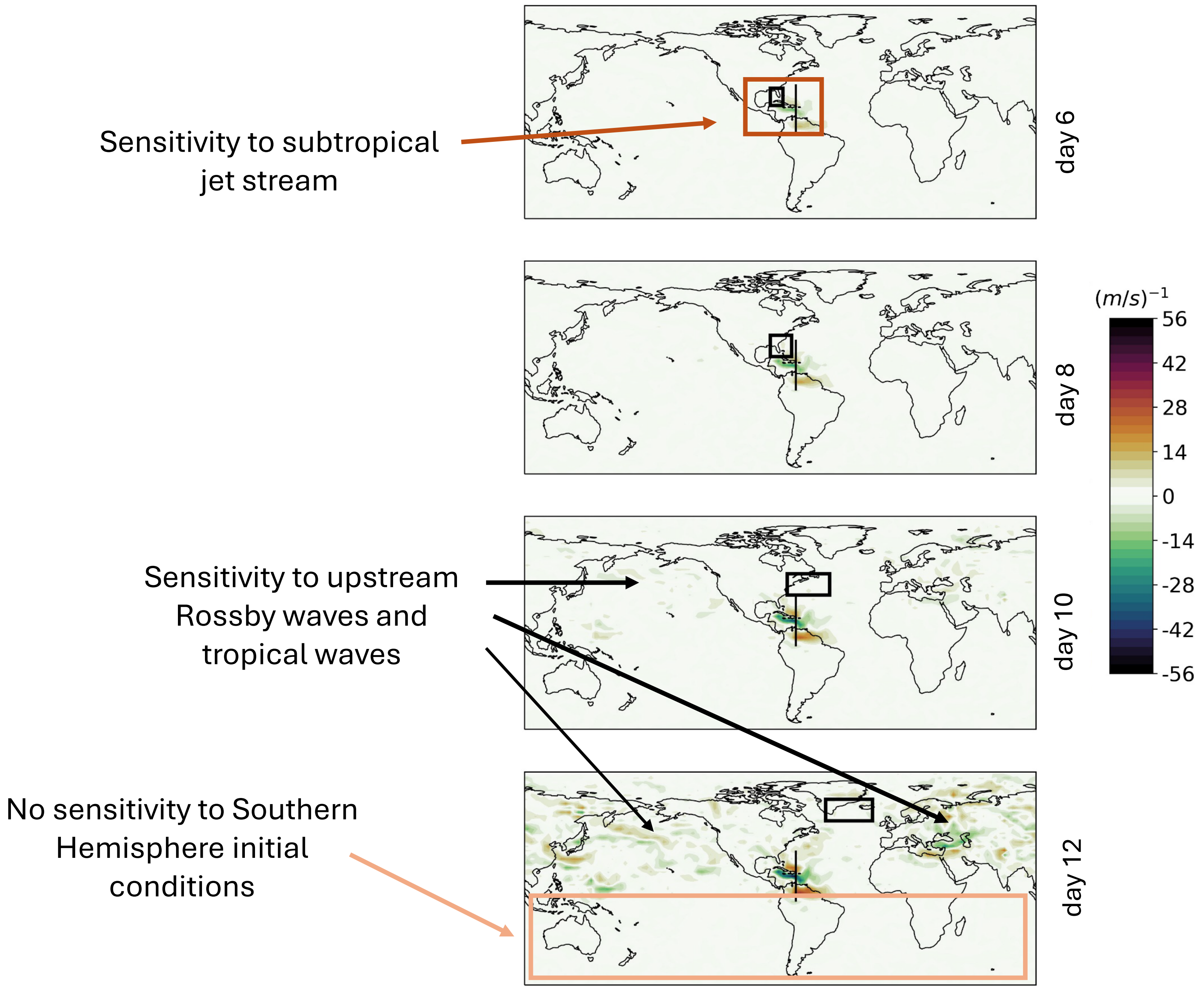




- Contours: 10 m horizontal wind speed forecast
- Colour: 10 m normalised wind forecast error
- Black box: domain for error calculation

Aim of case study: estimate the sensitivity of Hurricane Ian‘s forecast error to the initial conditions.

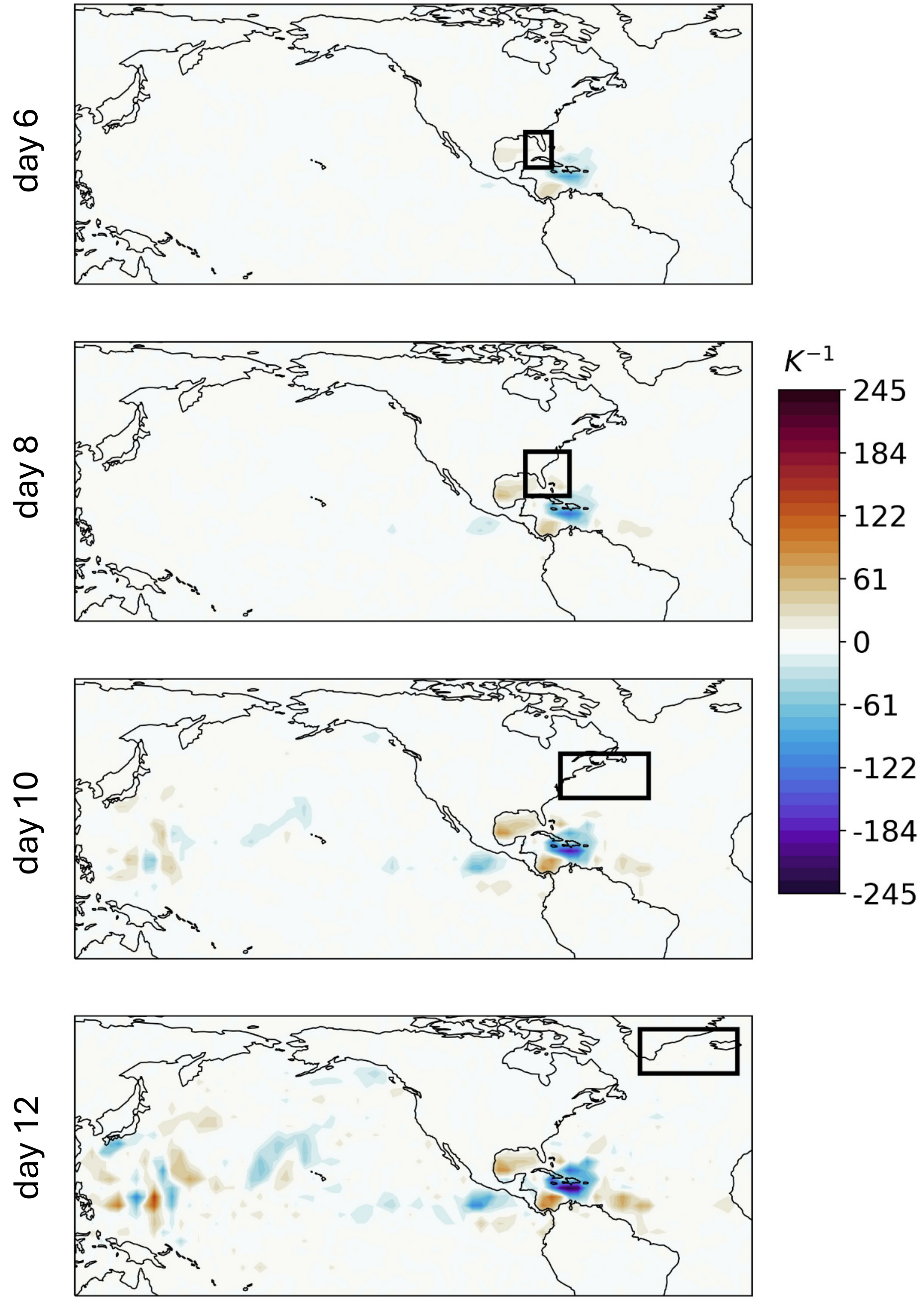
SENSITIVITY TO ZONAL WINDS



Plot: $\frac{\partial E}{\partial U500}(t=0)$

SENSITIVITY TO SURFACE TEMPERATURES

- In order to improve the forecast of hurricane Ian‘s tropical-to-extratropical transition we should improve initial condition at sensitive regions – especially at the Caribian and the Bay of Mexico.

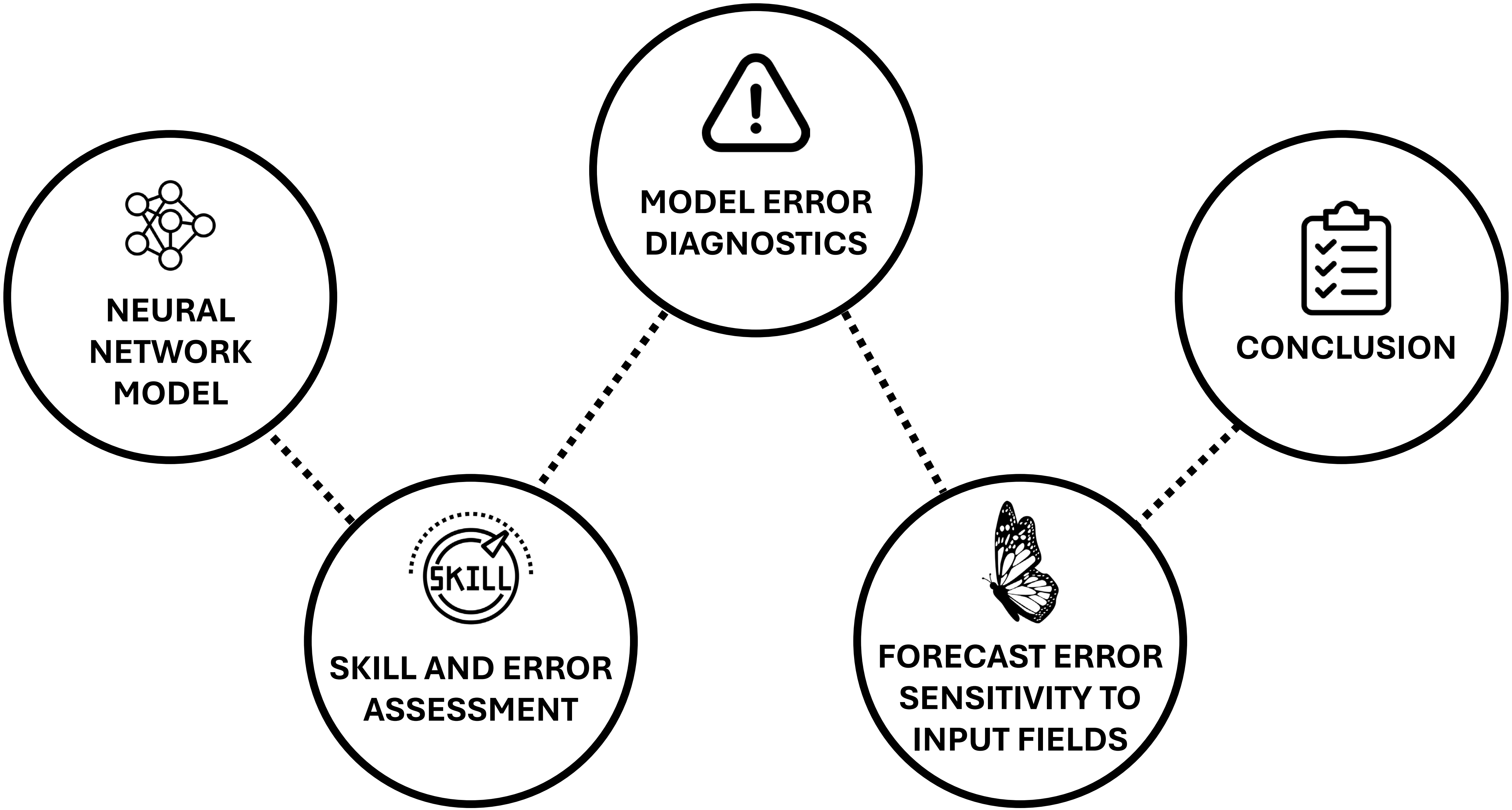


Plot: $\dfrac{\partial E}{\partial ST}(t = 0)$

POSSIBLE USE IN WEATHER PREDICTION

- Determine regions where more measurements are needed to decrease forecast uncertainty at later times
- Error calculation: use ensemble model spread as a proxy for model error

- Pros ☺:
  - Sensitivity is calculated using a fully nonlinear model (instead of an adjoint model – which is valid for a limited amount of time)
- Cons ☹:
  - We assume the perfect model


NEURAL NETWORK MODEL
SKILL
SKILL AND ERROR ASSESSMENT
MODEL ERROR DIAGNOSTICS
FORECAST ERROR SENSITIVITY TO INPUT FIELDS
CONCLUSION

CONCLUSION

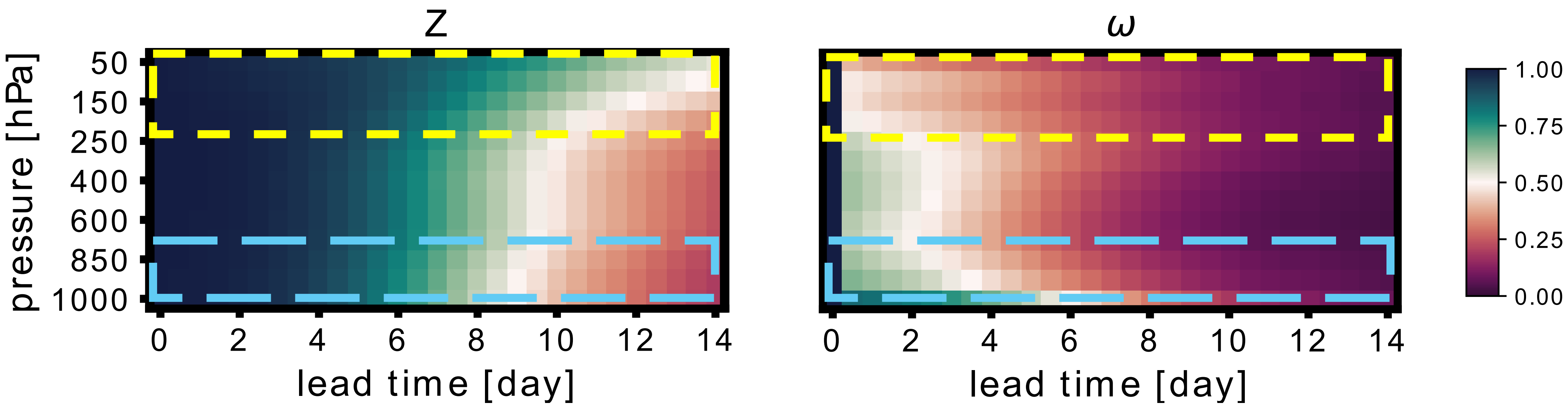
CONCLUSION

CONCLUSION

1. Regional overwriting of model forecast with the „truth“ is a simple yet efficient way for ML model error diagnostics
2. Error backpropagation could be a useful tool for:
    - Physical consistency evaluation
    - Improving initial conditions

The full analysis will be published soon.

ACC at different lead times and pressure levels

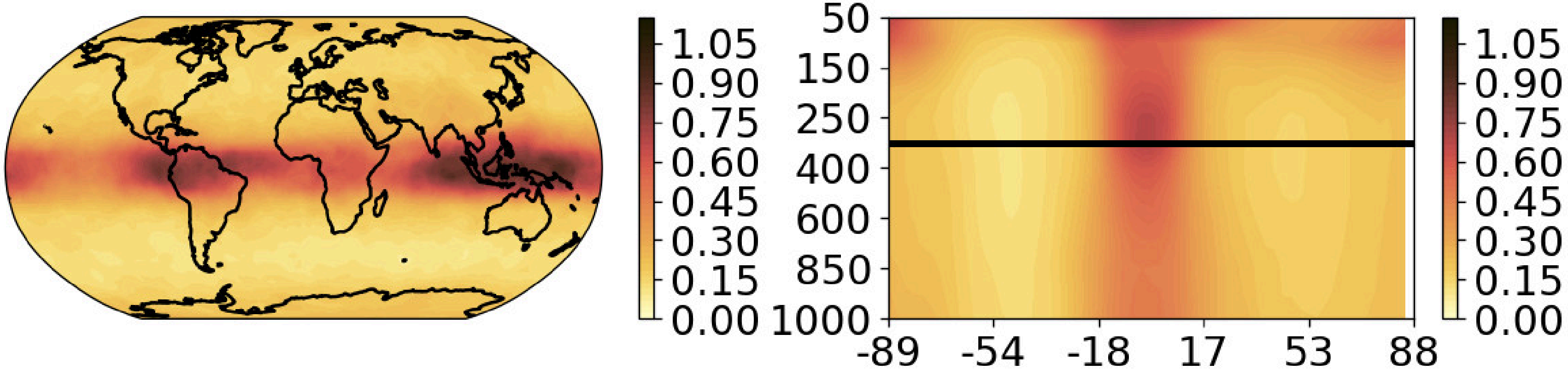


3 distinct predictability regimes:

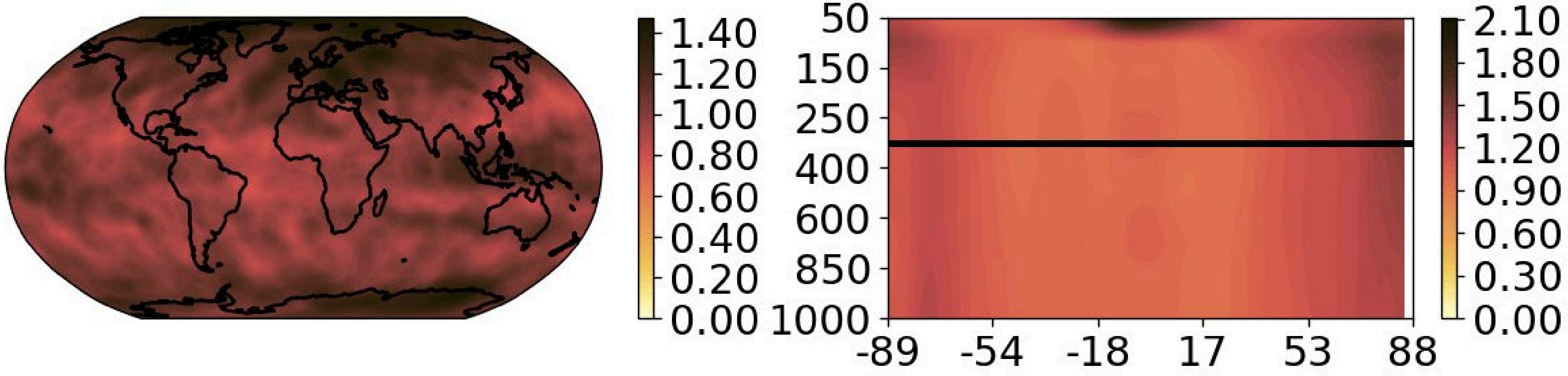
- Stratosphere
- Free troposphere
- „Planetary boundary layer“

LEAD TIME DIFFERENCE

Z: lead day 2

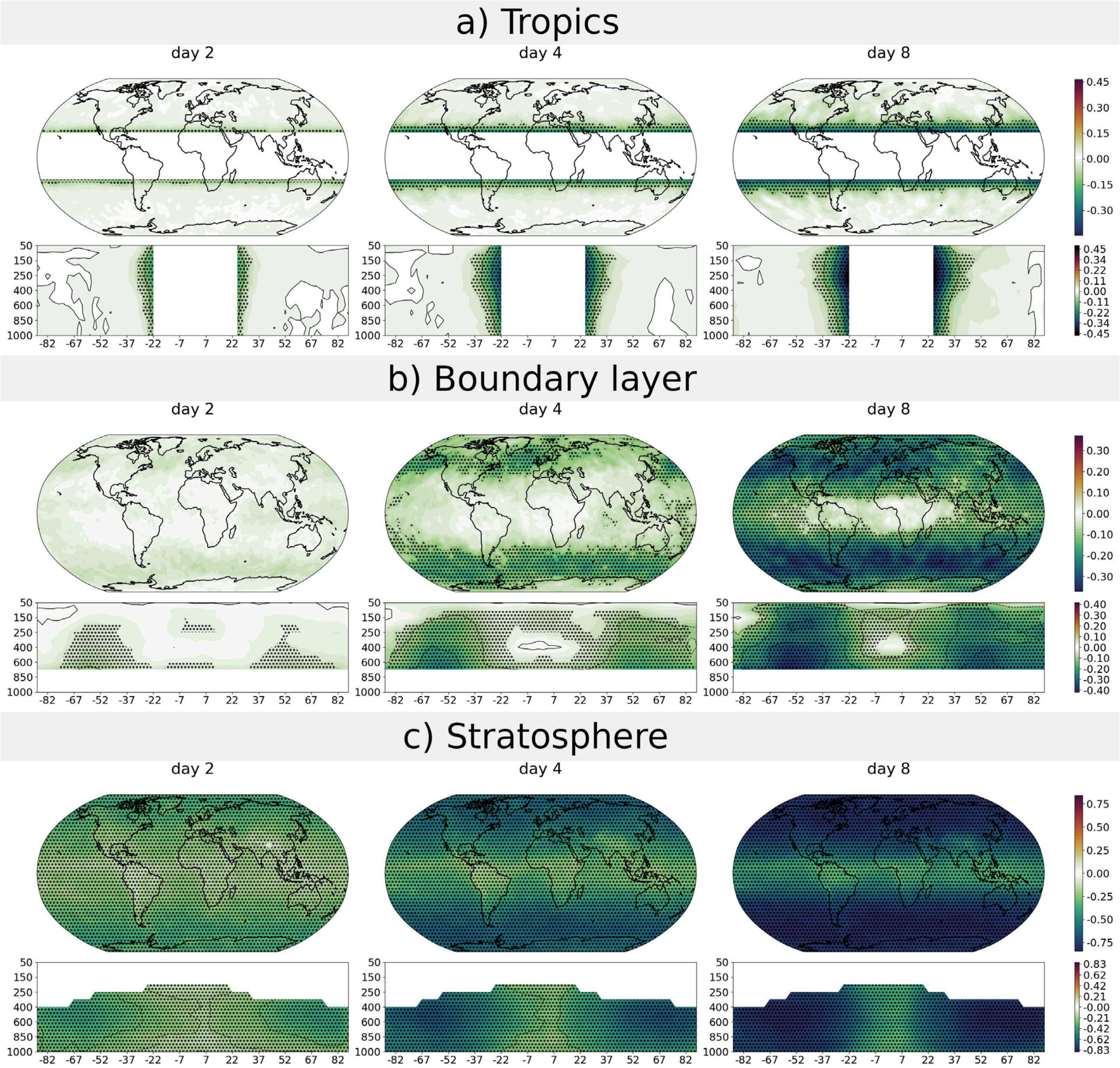


Z: lead day 10



Lead day 10 normalised error difference:

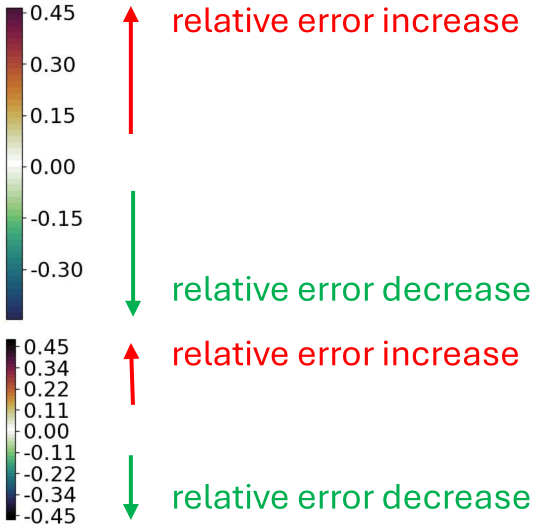
- Mid-to-high latitude errors become prevailing
- Stratospheric errors become more prominent



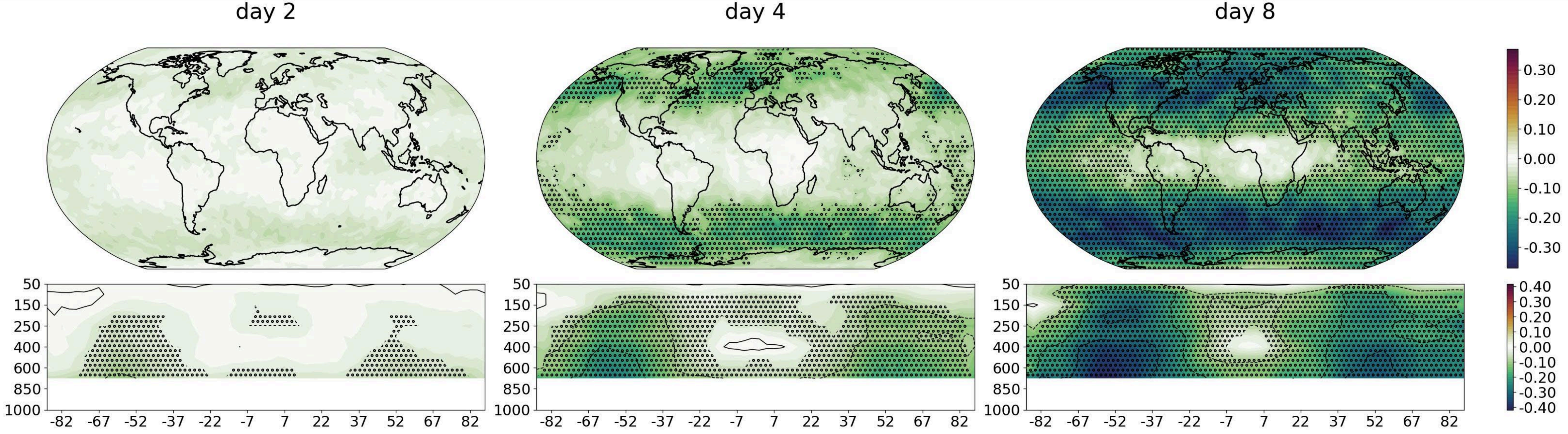
Plot:

- Verticaly and zonally averaged relative temperature error comparison

$$1 - \frac{error(M(x))}{error(M(x_{forced}))}$$



b) Boundary layer



- Statistically significant improvements spread with time:
  - Starting from mid-latitude 850 hPa ERA5 boundary
  - Surrounding tropical mid-troposphere
  - Reaching stratosphere after lead day 4
- Big improvements in mid-latitudes

MODEL FORCING METHODOLOGY

autoregressive input field

ERA5 „truth“ term

Model forecast term

$$x(\lambda, \varphi, p, t + \Delta t) = \text{weight}_\tau(\lambda, \varphi, p) \cdot \text{ERA5}(\lambda, \varphi, p, t + \Delta t) + (1 - \text{weight}_\tau(\lambda, \varphi, p)) \cdot \mathcal{M}_{t \to t+\Delta t}(s(t))(\lambda, \varphi, p)$$

Vertical weights equation: $\text{weight}_p = \frac{1}{2} \tanh \frac{p - p_0}{\Delta p} + \frac{1}{2}$

Latitude weihts equation: $\text{weight}_\varphi = \frac{1}{2} \tanh \frac{\varphi - \varphi_{\min}}{\Delta \varphi} - \frac{1}{2} \tanh \frac{\varphi - \varphi_{\max}}{\Delta \varphi} + 1$

FORECAST ERROR

Sum over predetermined domain

Standardized ERA5 fields

Standardized input fields

$$\text{error} = \sum_{\lambda,\varphi,p\in\mathcal{D}} \mid \mathcal{S}(\mathbf{ERA5}(\lambda,\varphi,p,t+\Delta t)) - \mathcal{M}_{t\to t+\Delta t}(\mathbf{s}(t))(\lambda,\varphi,p) \mid$$

$$\frac{\partial \text{error}}{\partial \mathbf{s}(t)} = \sum_{\lambda,\varphi,p\in\mathcal{D}} \operatorname{sgn}\left[\mathcal{S}(\mathbf{ERA5}(\lambda,\varphi,p,t+\Delta t)) - \mathcal{M}_{t\to t+\Delta t}(\mathbf{s}(t))(\lambda,\varphi,p)\right] \frac{\partial \mathcal{M}_{t\to t+\Delta t}(\mathbf{s}(t))(\lambda,\varphi,p)}{\partial \mathbf{s}(t)}$$

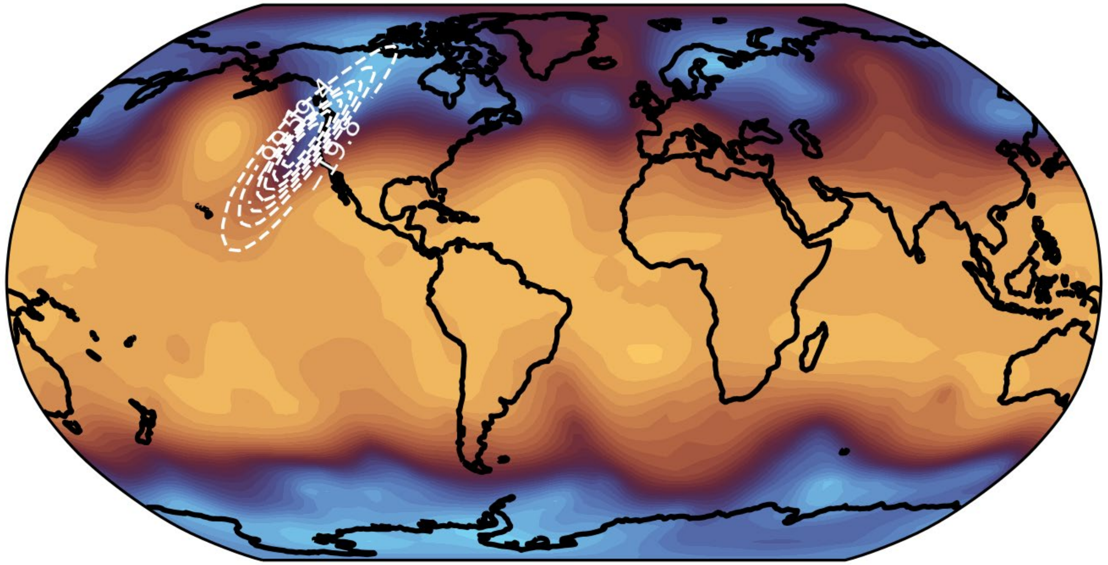
Error derivative with respect to standardized input fields

Sign function

Diferentiation of the model with respect to standardized input fields

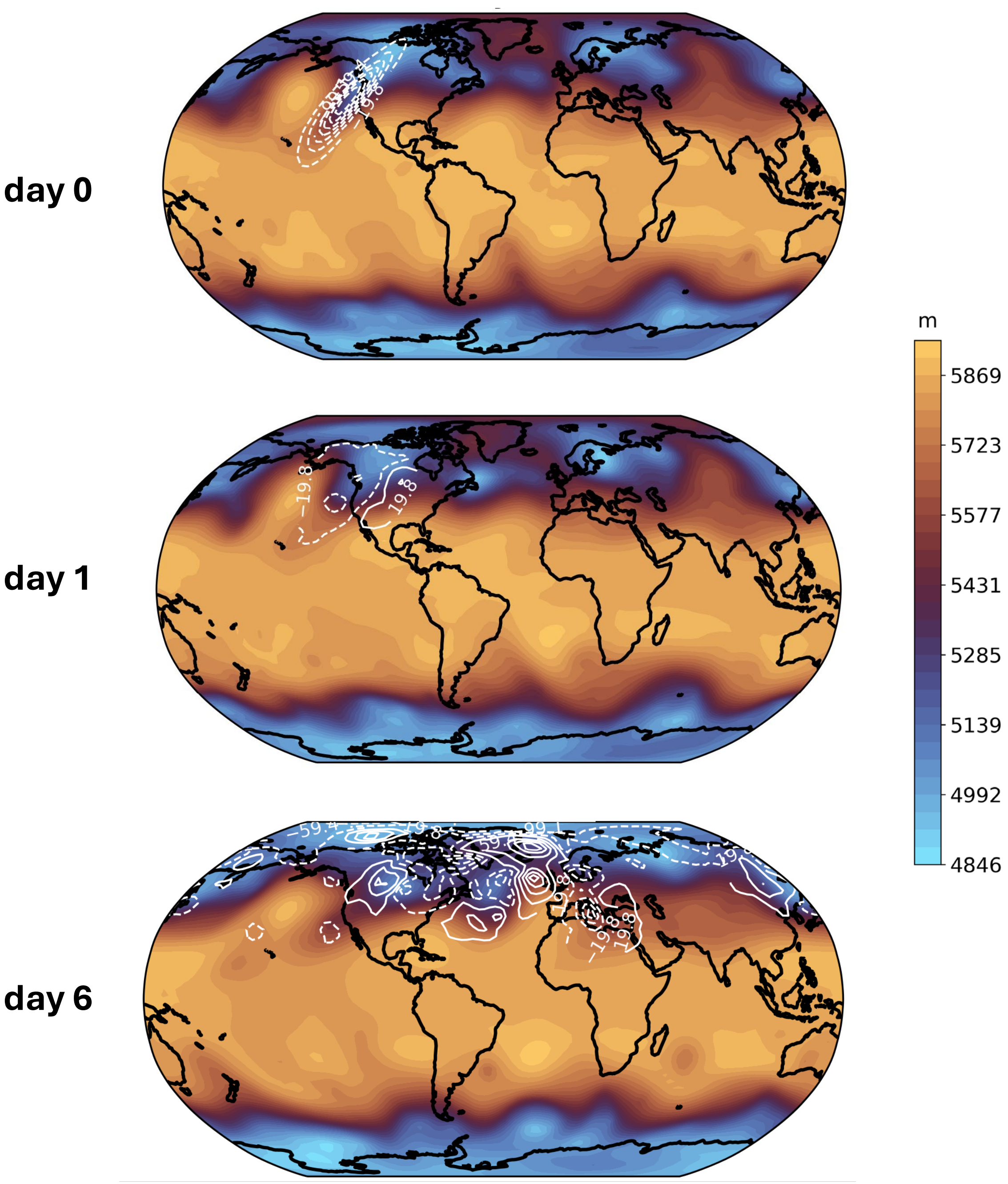
ADJUSTMENT EXPERIMENTS

Z500 initial condition perturbation



Goal:

- Compare model response with dynamics expectations
- ML models not so sensitive to instability => easy to initialize forecast



Propagation features:

- Propagates downstream
- Faster than Rossby wave phase speed
- Confined to Northern Hemisphere for first 6 forecast days