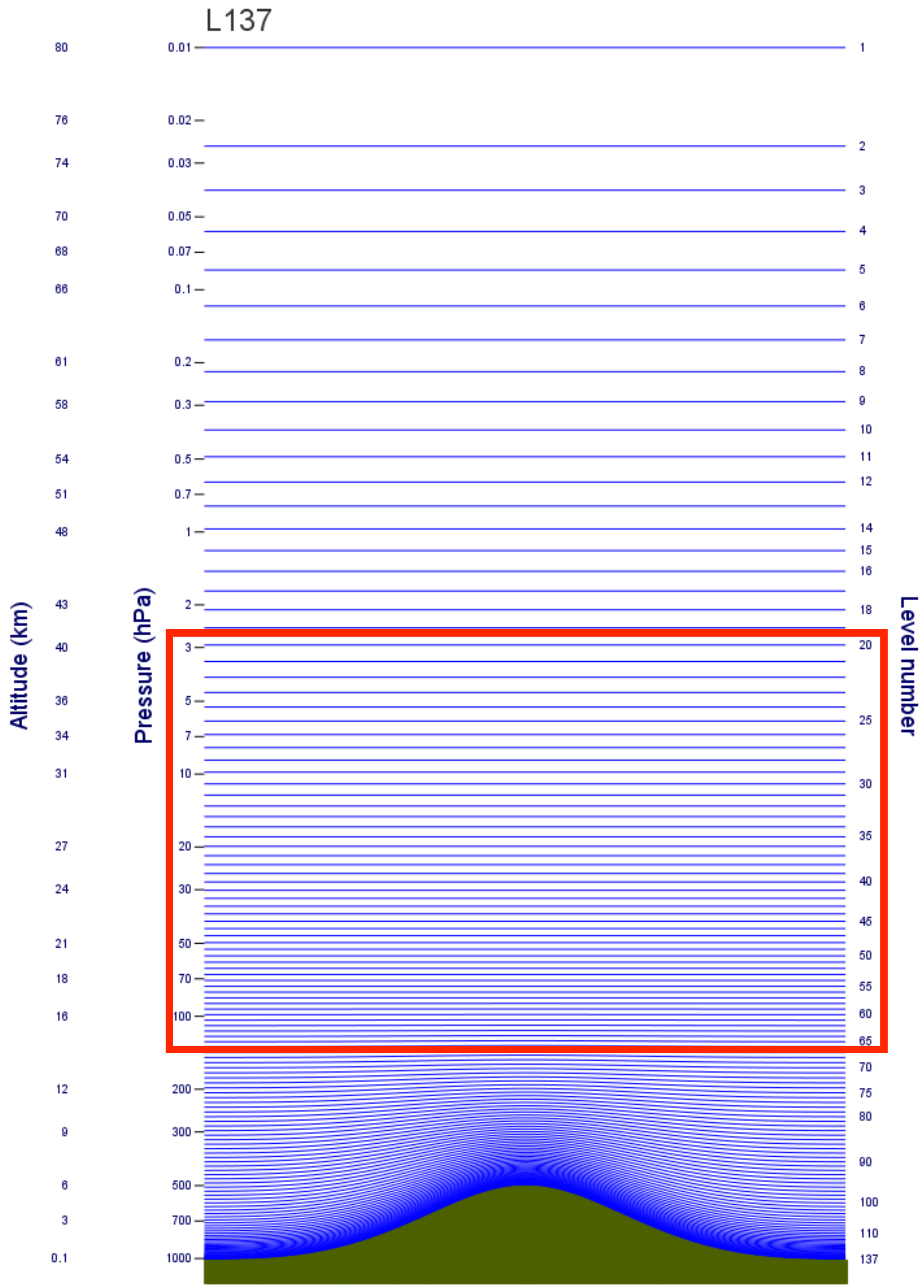
3D Bias Correction with Deep Learning in the Integrated Forecasting System

Thorsten Kurth*, David Hall (NVIDIA)
Patrick Laloyaux, Peter Dueben (ECMWF)

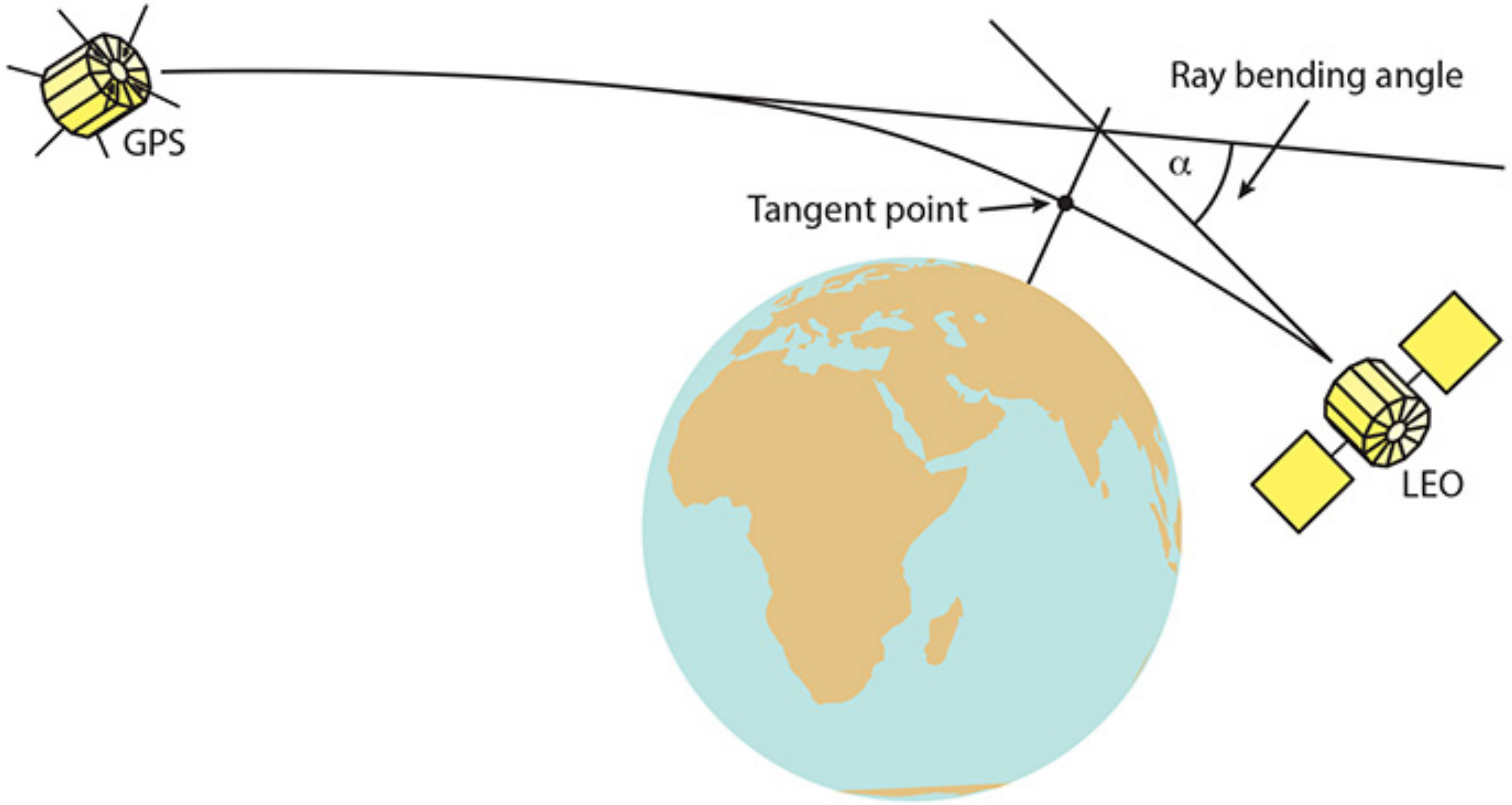
Joint IS-ENES3/ESiWACE2 Workshop, 18.03.2021

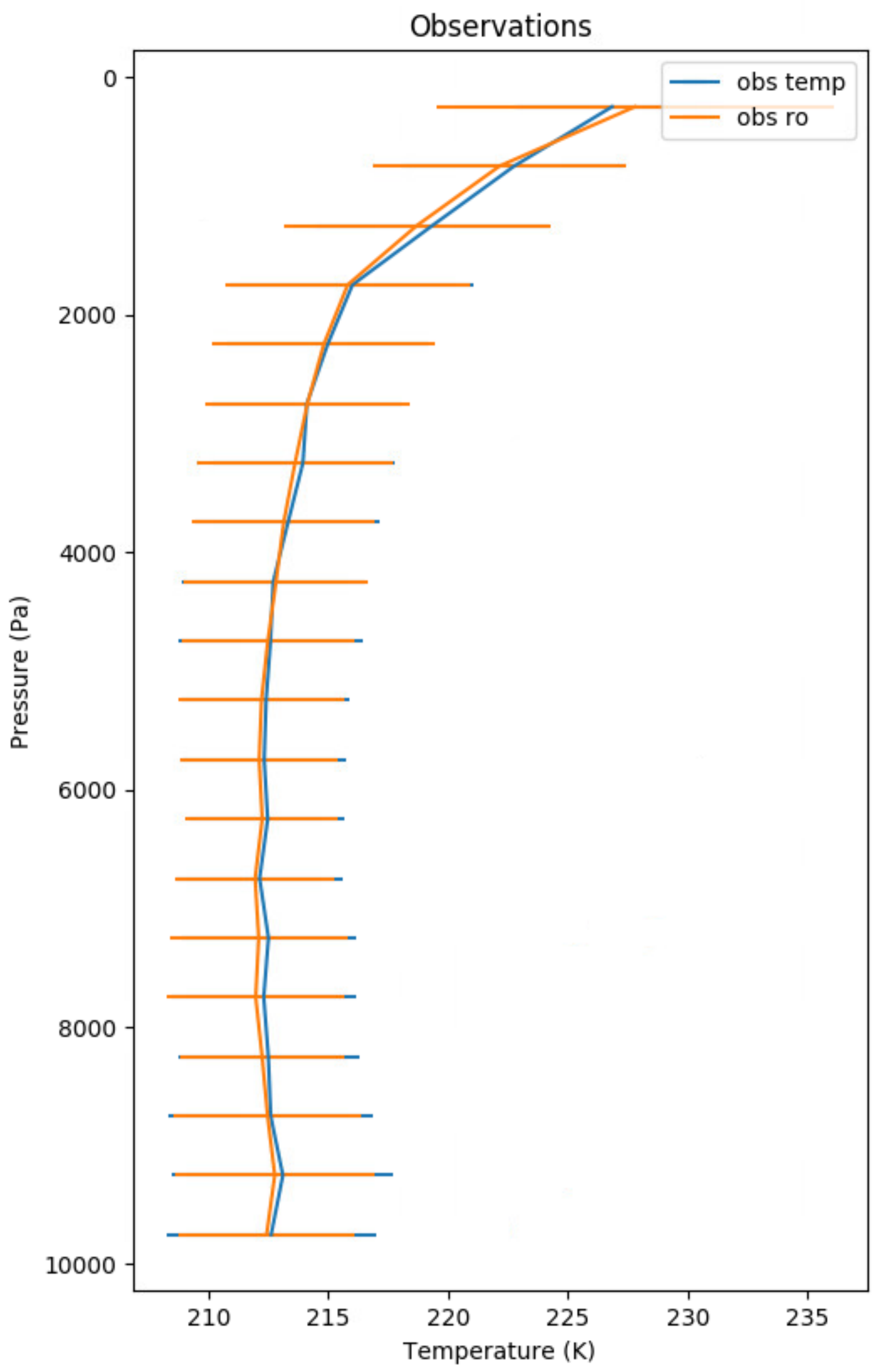
Problem Description

- ECMWF is developing their own atmospheric model: Integrated Forecasting System (IFS)
- like all models, it has systematic uncertainties
- systematic uncertainties largest for altitudes 15-40 km (levels ~20-65)
- improve data by observing this part using instruments, e.g. sounding balloons and satellite (radio occultation)



Radio Occultation

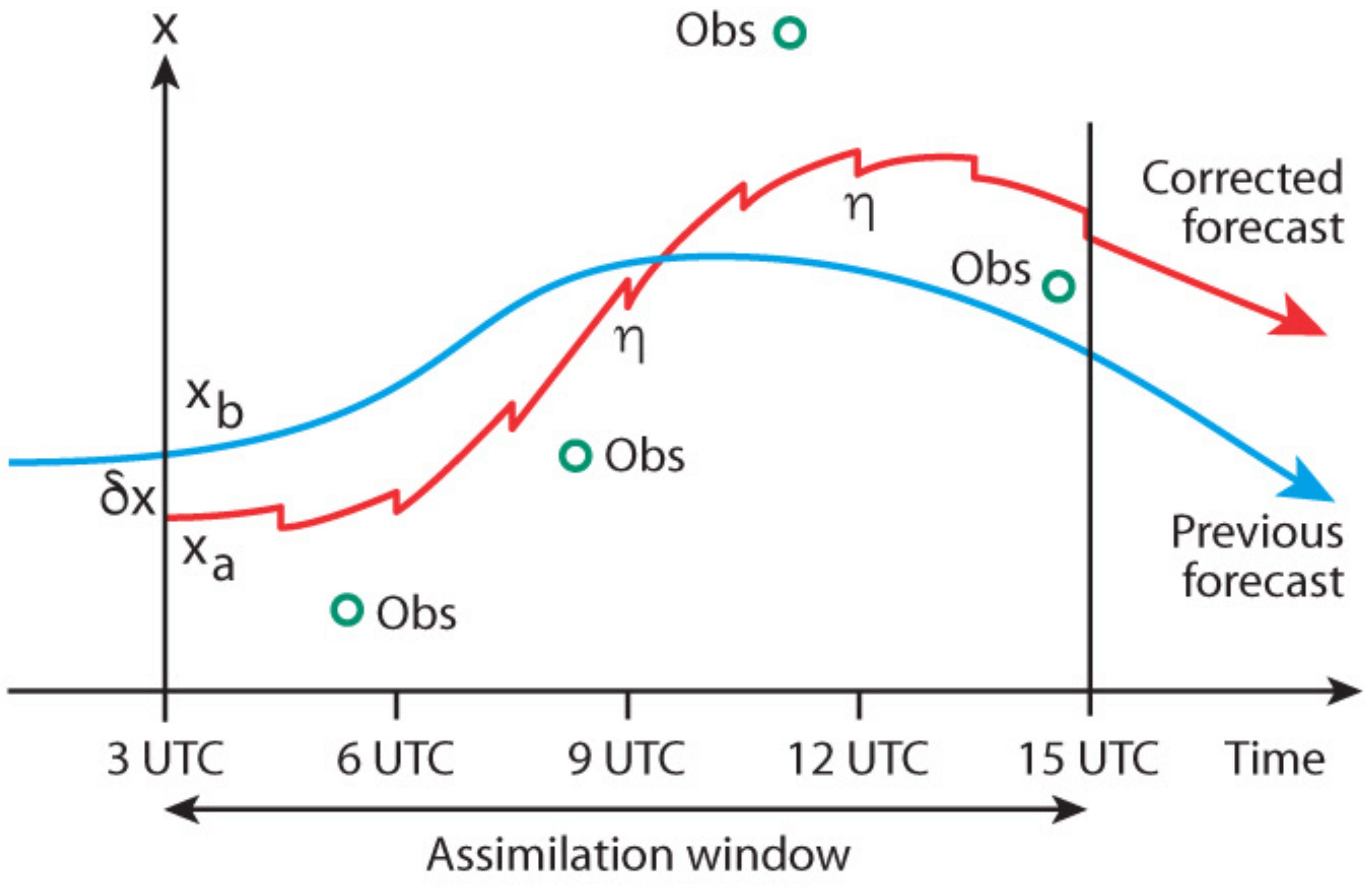




- observe bending of radio signals in atmosphere between GPS satellite and low-earth-orbiting satellite (LEO)
- as LEO moves behind earth, bending profile can be obtained
- infer temperature profile from that bending profile
- accuracy similar to conventional probes but better coverage

Weak-constrained-4D-Var

- correct model using observations
- estimation of bias term main focus of this talk

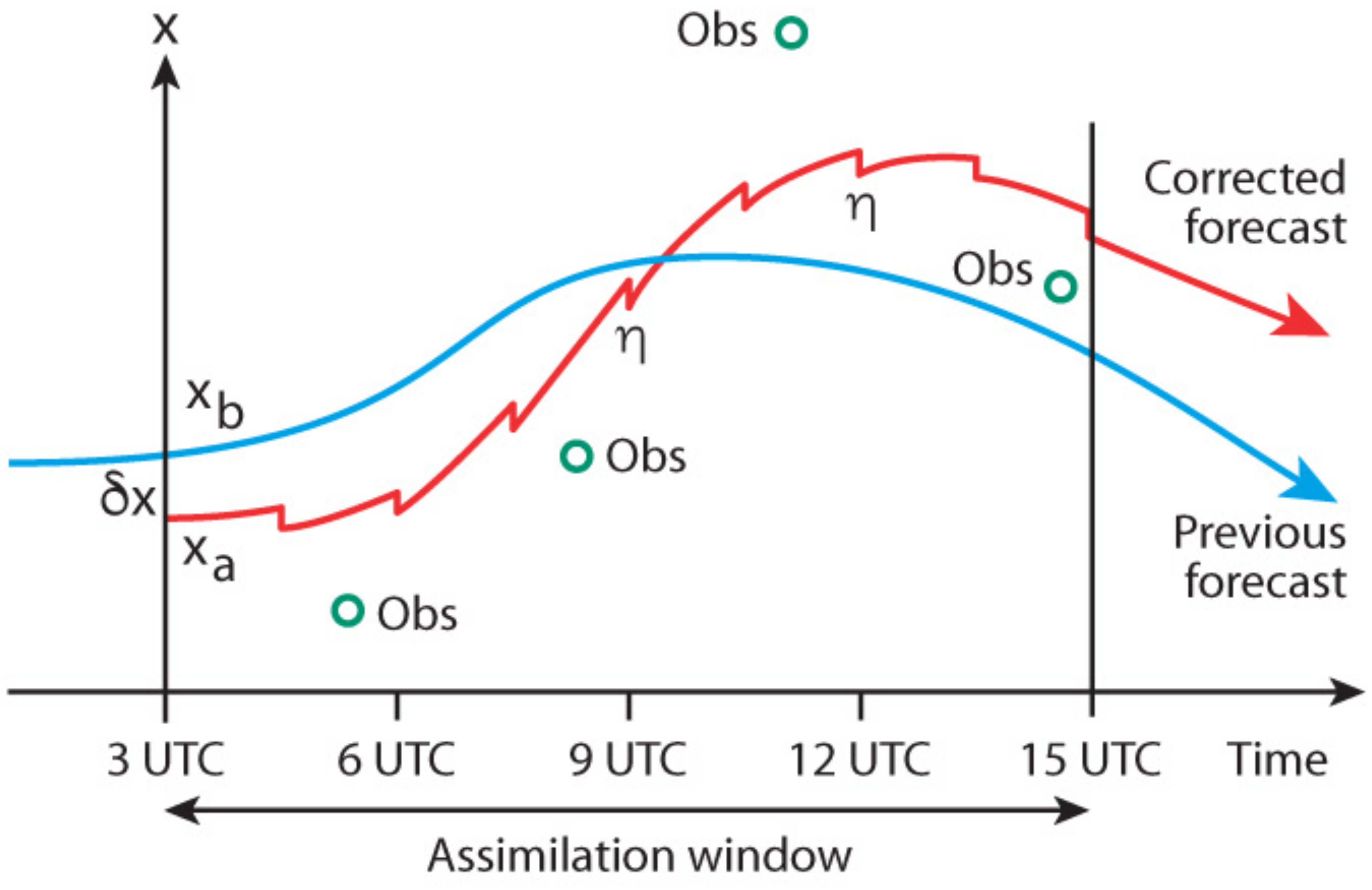


$$
\begin{aligned}
J(x_0, \beta, \eta) = & \frac{1}{2}(x_0 - x_b)^T \mathbf{B}^{-1}(x_0 - x_b) \\
& + \frac{1}{2} \sum_{k=0}^{K} \left(y_k - b(x_k, \beta) - \mathcal{H}(x_k)\right)^T \\
& \quad \mathbf{R}_k^{-1}\left(y_k - b(x_k, \beta) - \mathcal{H}(x_k)\right) \\
& + \frac{1}{2}(\beta - \beta_b)^T \mathbf{B}_\beta^{-1}(\beta - \beta_b) \\
& + \frac{1}{2}(\eta - \eta_b)^T \mathbf{Q}^{-1}(\eta - \eta_b)
\end{aligned}
$$

$$x_k = \mathcal{M}(x_{k-1}) + \eta$$

Weak-constrained-4D-Var

- correct model using observations
- estimation of bias term main focus of this talk



$$J(x_0, \beta, \eta) = \frac{1}{2}(x_0 - x_b)^T \mathbf{B}^{-1}(x_0 - x_b)$$
$$+ \frac{1}{2}\sum_{k=0}^{K} \big(y_k - b(x_k, \beta) - \mathcal{H}(x_k)\big)^T$$
$$\mathbf{R}_k^{-1}\big(y_k - b(x_k, \beta) - \mathcal{H}(x_k)\big)$$
$$+ \frac{1}{2}(\beta - \beta_b)^T \mathbf{B}_\beta^{-1}(\beta - \beta_b)$$
$$+ \frac{1}{2}(\eta - \eta_b)^T \mathbf{Q}^{-1}(\eta - \eta_b)$$

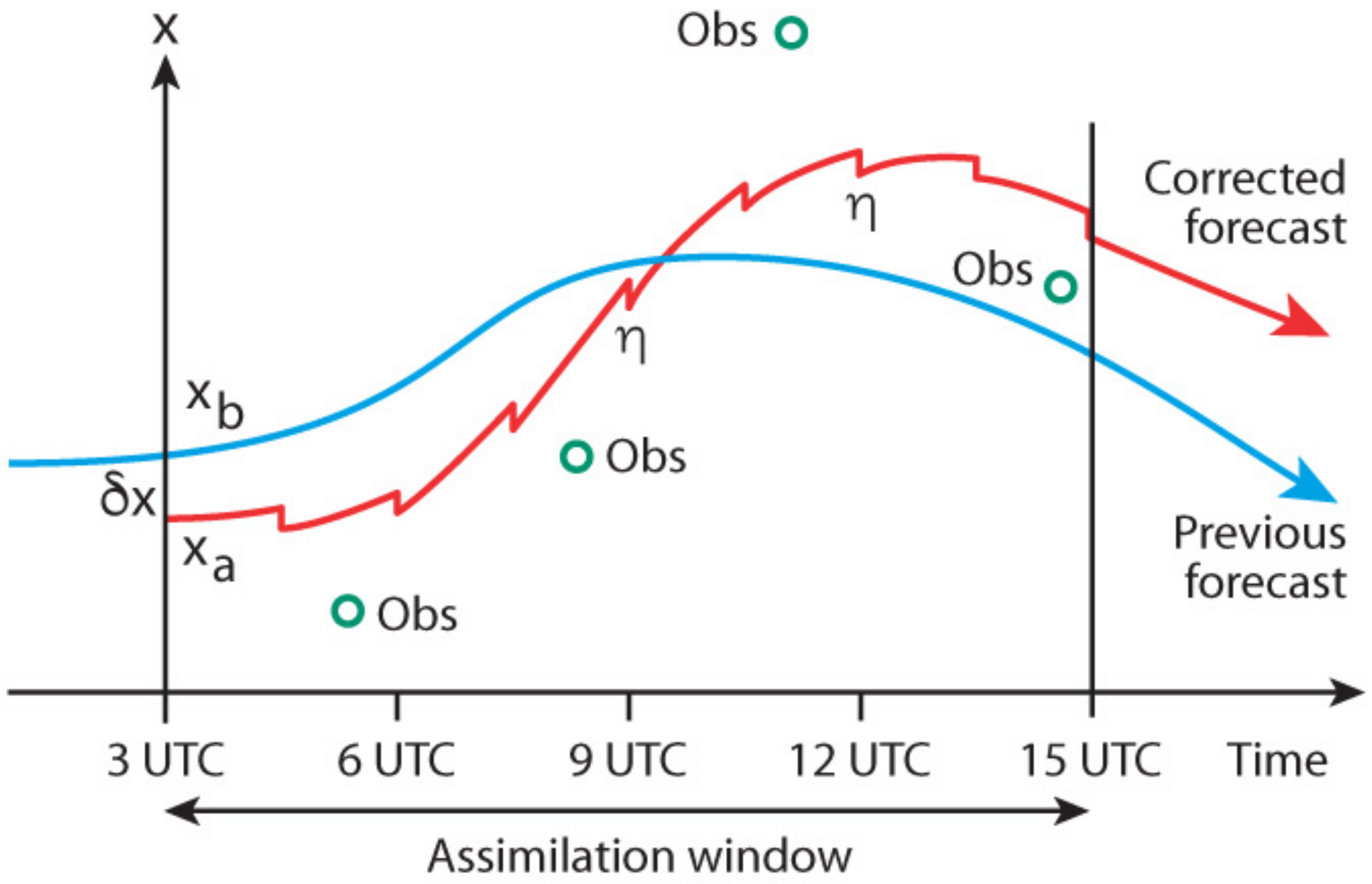
Model bias

$$x_k = \mathcal{M}(x_{k-1}) + \eta$$

Weak-constrained-4D-Var

- correct model using observations
- estimation of bias term main focus of this talk

$$J(x_0,\beta,\eta) = \frac{1}{2}(x_0 - x_b)^T \mathbf{B}^{-1}(x_0 - x_b)$$
$$+ \frac{1}{2}\sum_{k=0}^{K}\big(y_k - b(x_k,\beta) - \mathcal{H}(x_k)\big)^T$$
$$\mathbf{R}_k^{-1}\big(y_k - b(x_k,\beta) - \mathcal{H}(x_k)\big)$$
$$+ \frac{1}{2}(\beta - \beta_b)^T \mathbf{B}_\beta^{-1}(\beta - \beta_b)$$
$$+ \frac{1}{2}(\eta - \eta_b)^T \mathbf{Q}^{-1}(\eta - \eta_b)$$



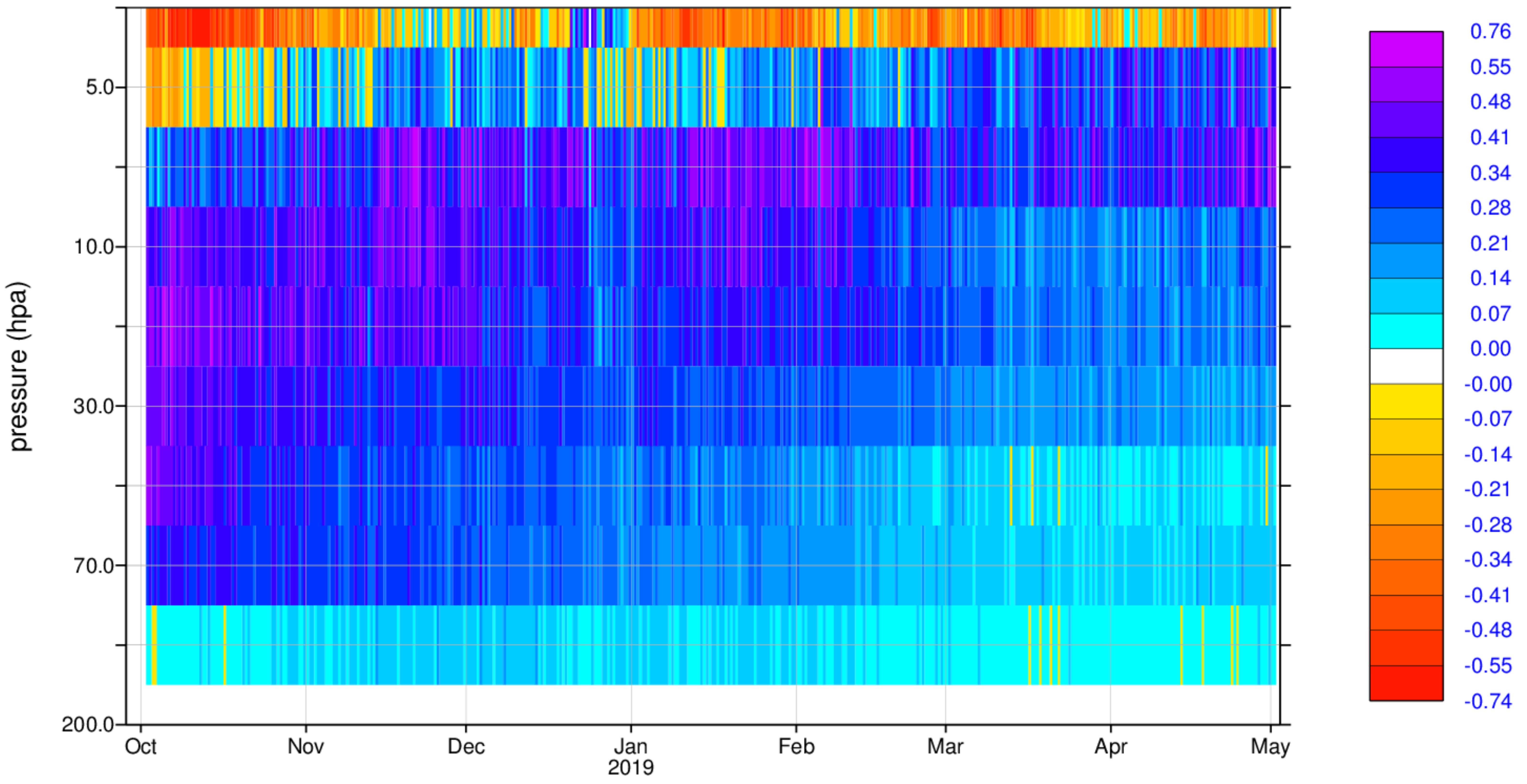
Model bias

$$x_k = \mathcal{M}(x_{k-1}) + \eta$$

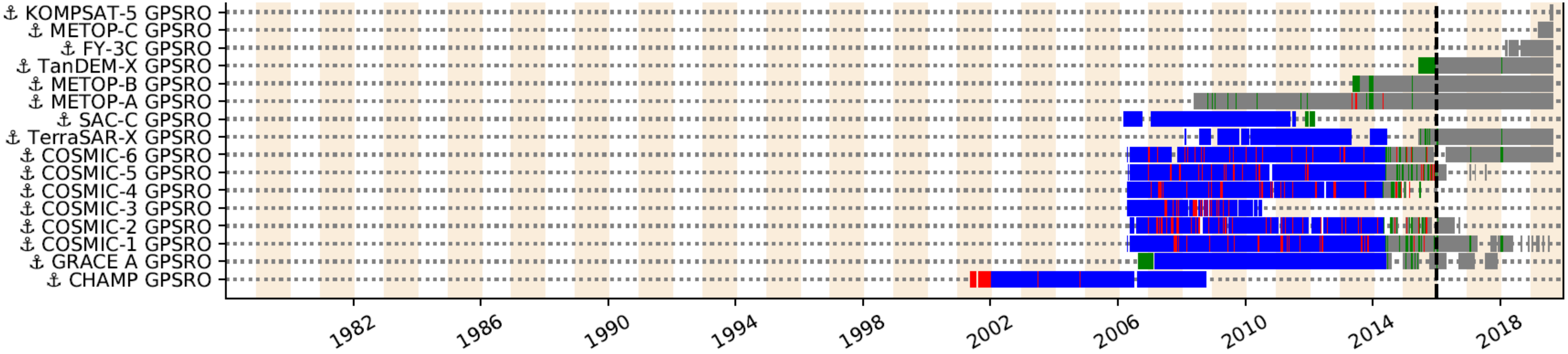
$$x_k = (1 + D_\eta) \circ \mathcal{M}(x_{k-1}) + \tilde{\eta}, \quad |\tilde{\eta}| \ll |\eta|$$

Successful Weak-Constrained-4D-Var

- online learning: bias reduced over time
- model improves so that fit bias is reduced
- goal: train a neural network which improves the model such that fitted bias will be small

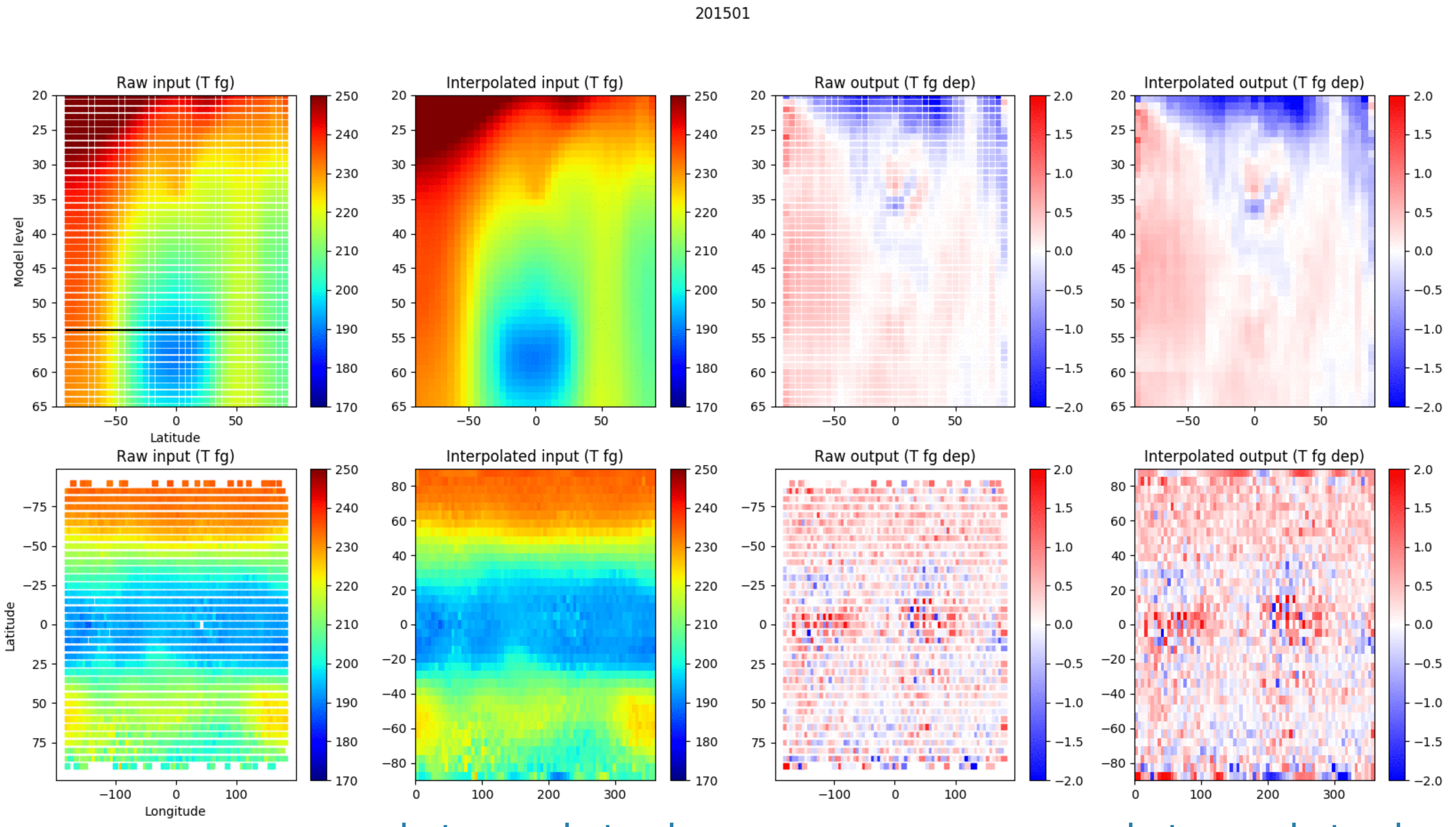


Dataset: GNNS-RO Observations



- reliable, high-quality data since ~2007
- use recent data from 2008-2021 (train on 2008-2019, test on 2020+)
- average over 2 or 10 days (hi-res, hi-noise vs lo-res, lo-noise)

Dataset: Visualization



vertical
x-section

horizontal
x-section

raw data

interpolated
input

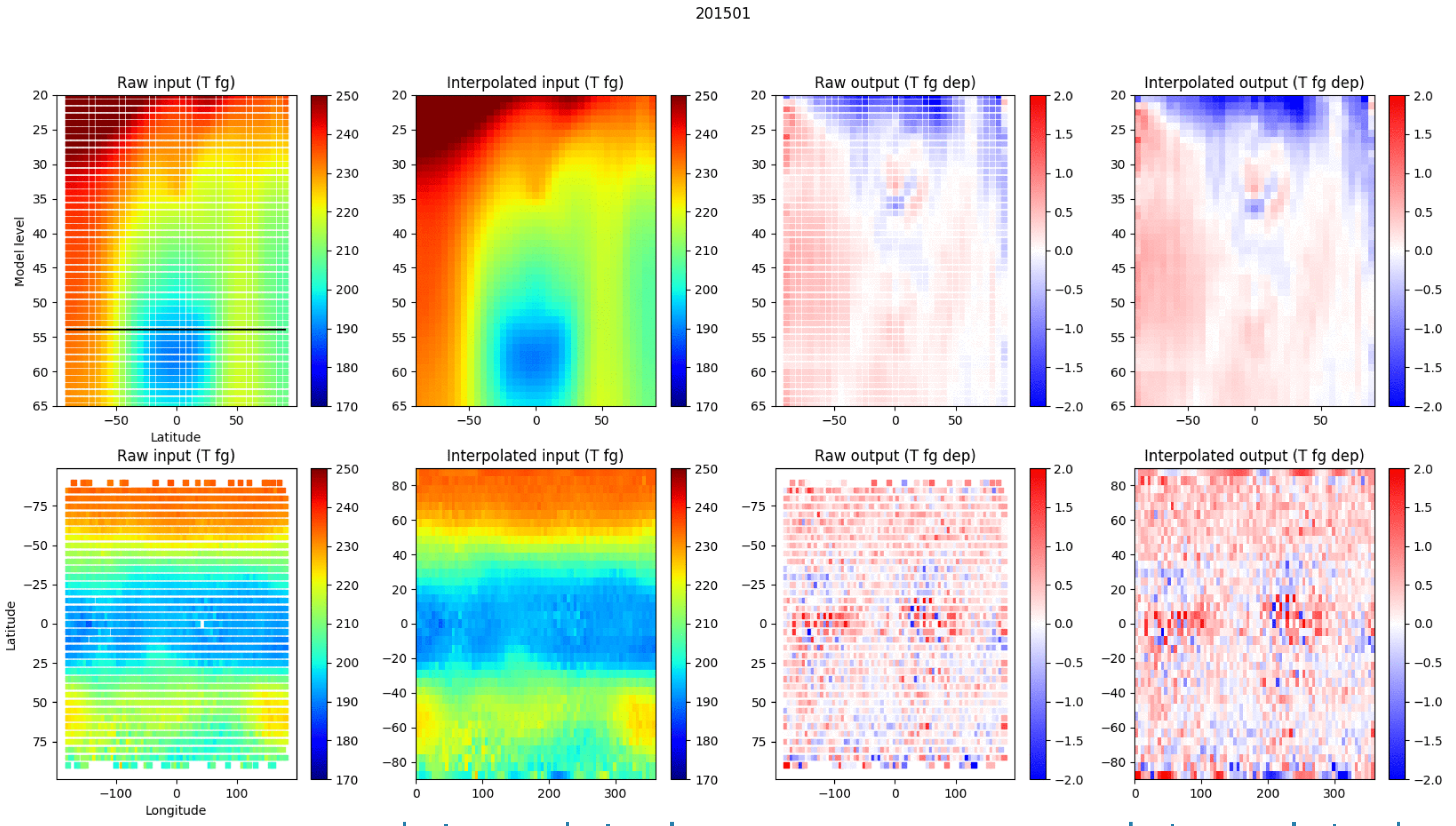
raw output

interpolated
output

Dataset: Visualization

vertical x-section

horizontal x-section



raw data

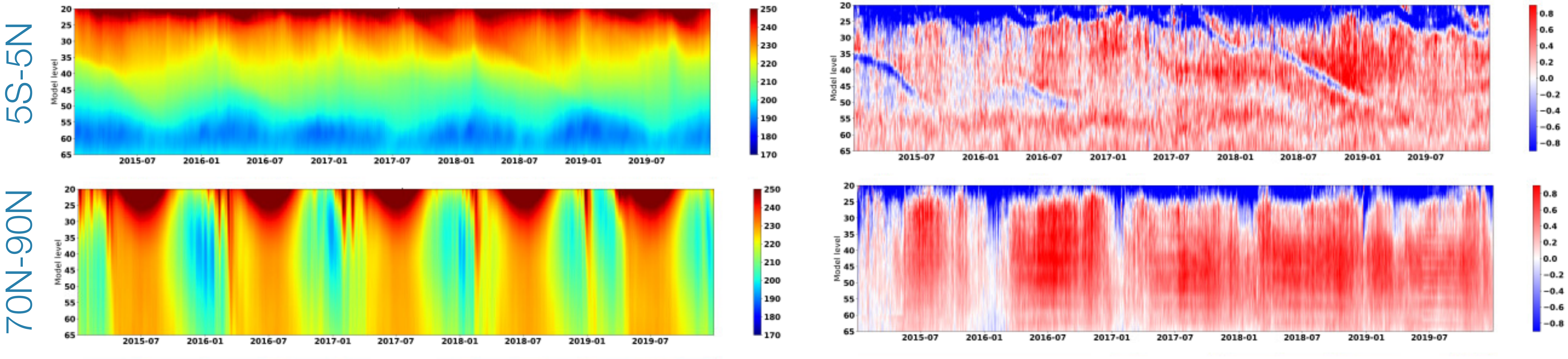
interpolated input

raw output

interpolated output

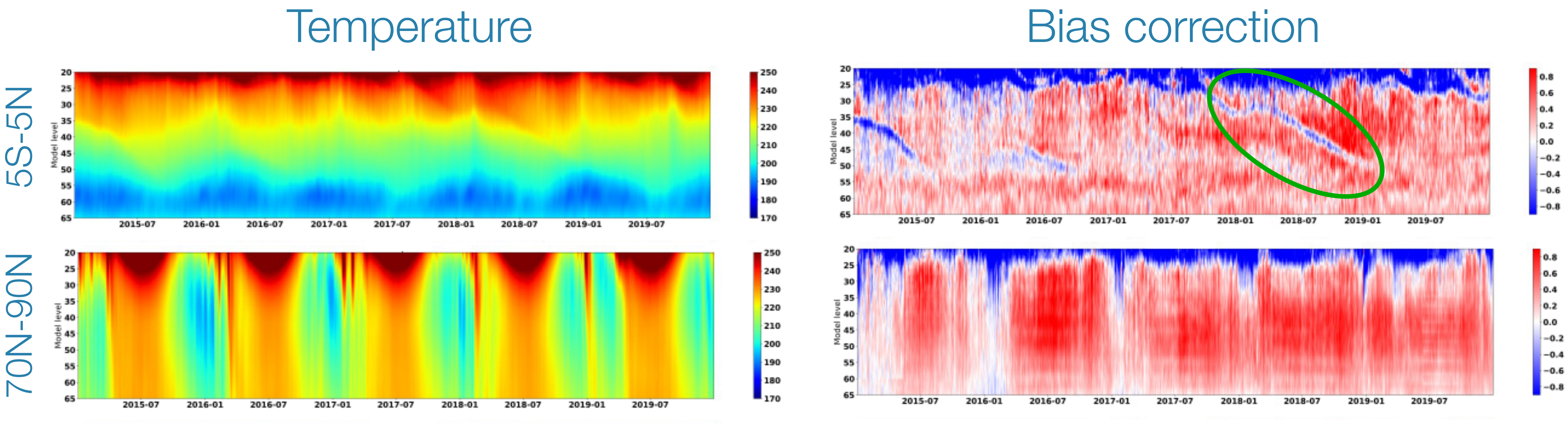
Challenges

Temperature | Bias correction



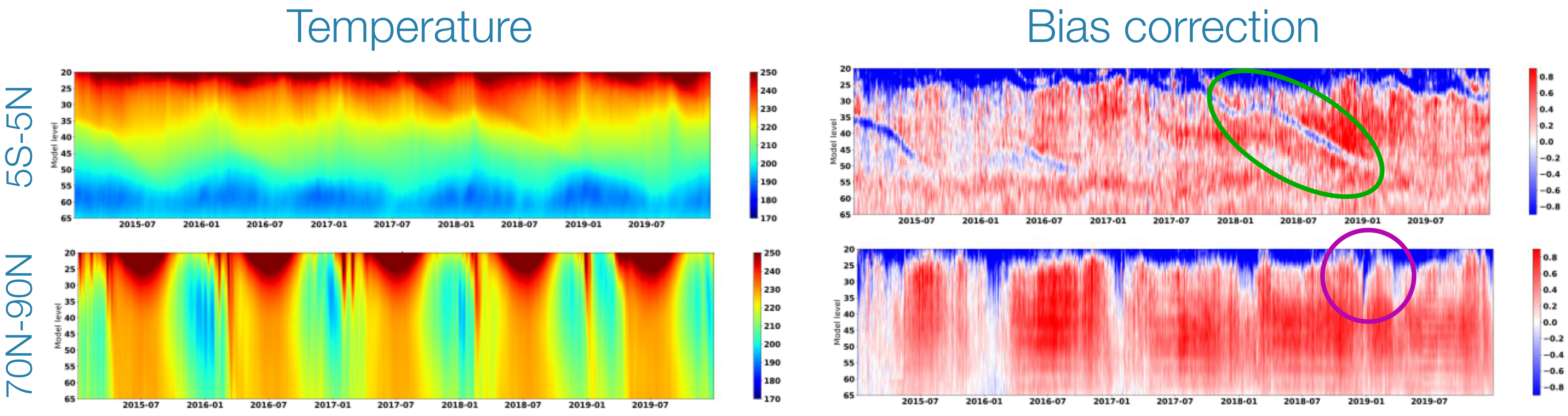
- can a neural network learn and predict features of the bias?
- positive correction below level 30, negative above level 30
- Quasi-Biennial Oscillation (QBO)
- sudden stratospheric warming events (SSW)

Challenges



- can a neural network learn and predict features of the bias?
- positive correction below level 30, negative above level 30
- Quasi-Biennial Oscillation (QBO)
- sudden stratospheric warming events (SSW)

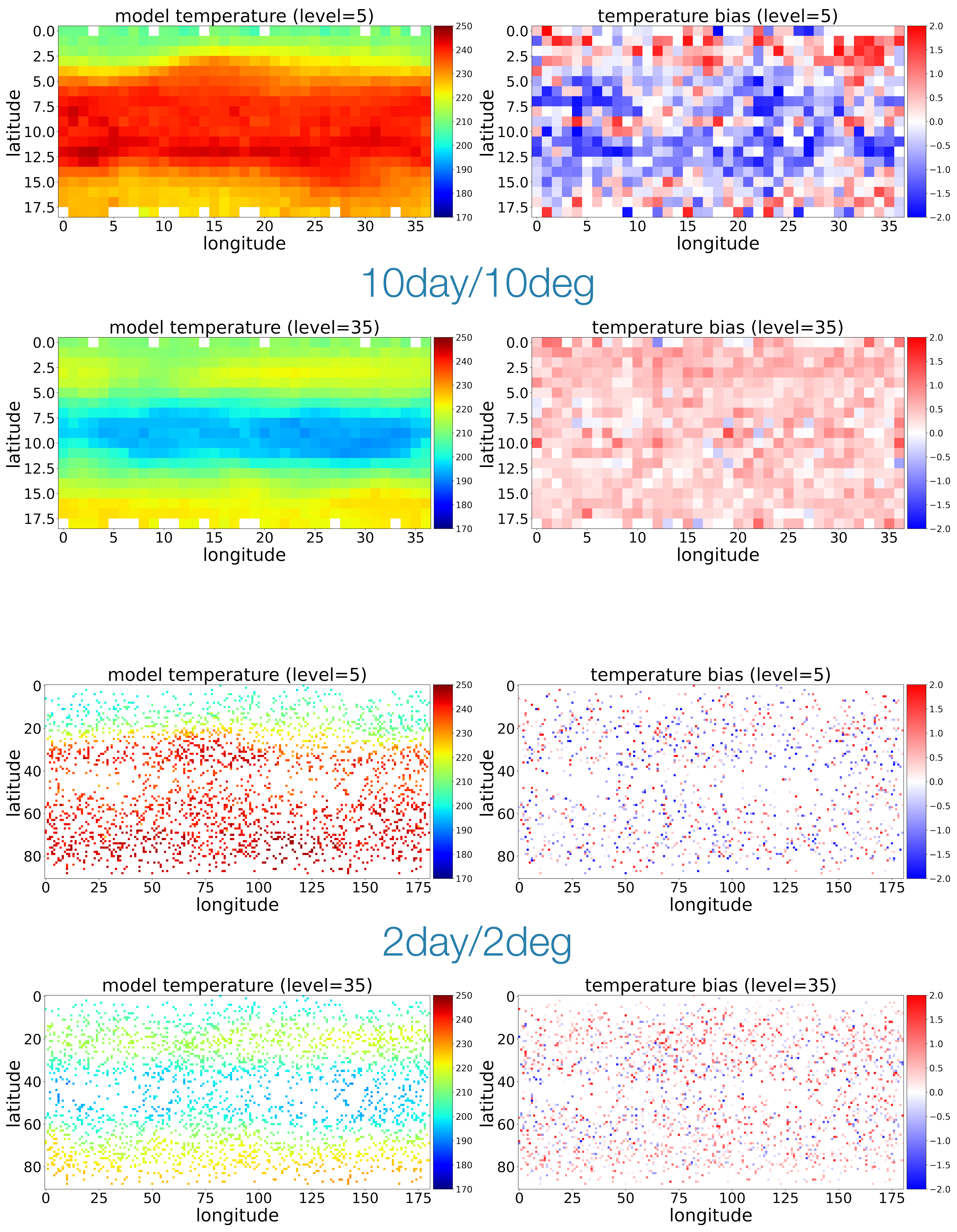
Challenges



- can a neural network learn and predict features of the bias?
- positive correction below level 30, negative above level 30
- Quasi-Biennial Oscillation (QBO)
- sudden stratospheric warming events (SSW)

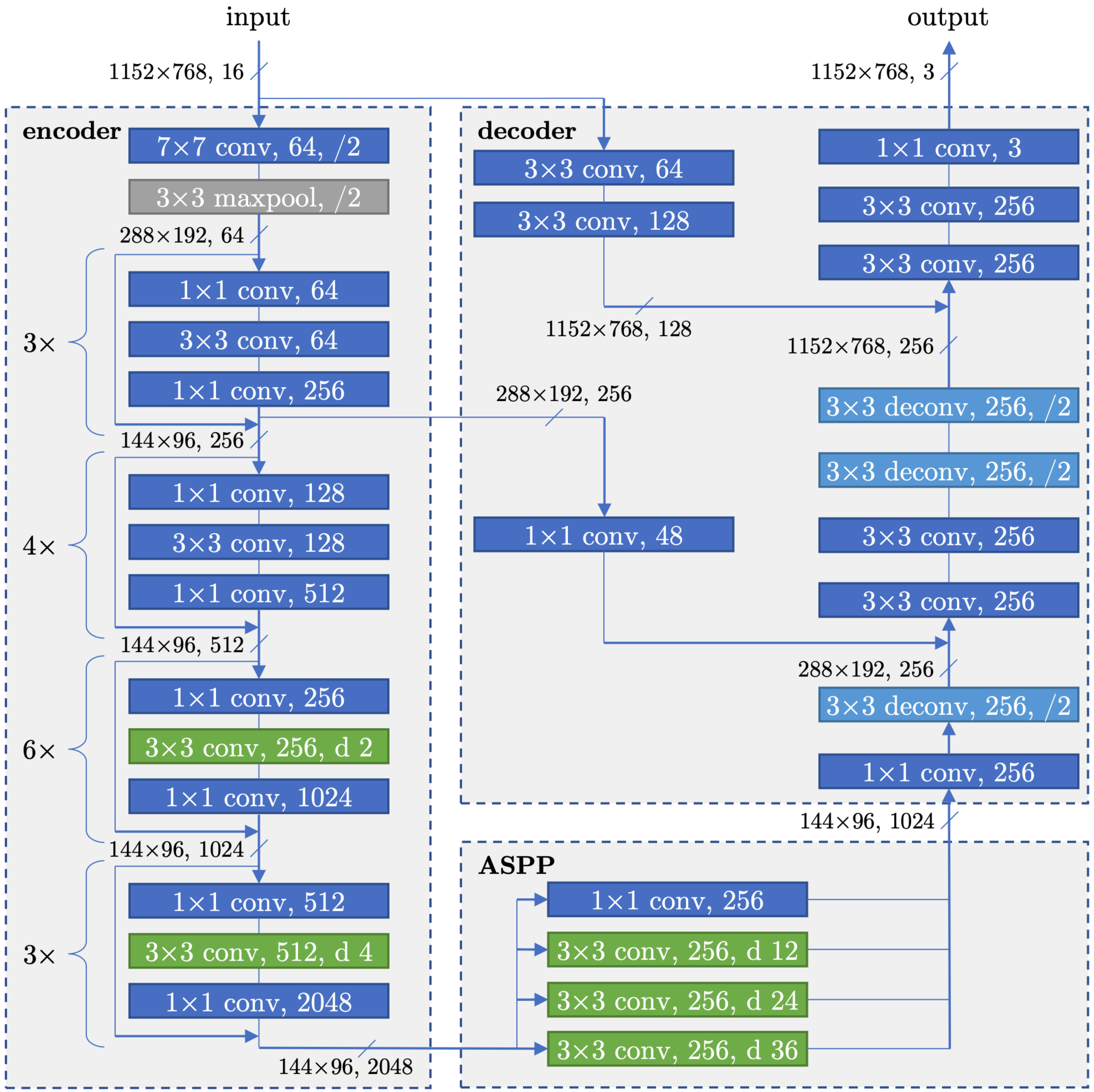
Dataset: Technical Details

- years 2008-2020, numpy format
- resolutions: 10day/10deg and 2day/2deg
- input, output: 19x37x45 or 91x181x45 in FP32
- masks 19x37x45 of 91x181x45 in INT
- average sparsity: ~97% and ~13% respectively
- data interpretation: 3D with 1 feature per pixel
- validation set: first 10 days of each month in 2019
- test set: 2020 and 2021
- training set: remaining data



Network Architecture

- based on DeepLabv3+ w/ Xception backend
- BatchNormalization or InstanceNormalization
- de-convolution decoder
- average pooling in upsampling for checkerboard artifact suppression



Kurth et al., doi:10.5555/3291656.3291724

Loss Function

- regression problem:
  - L1: good for penalizing small deviations as well as large ones similarly, but derivative not continuous in 0
  - L2: good for enhanced penalty of outliers, less sensitivity for small deviations, smooth everywhere
  - Smooth-L1: same as L2 inside [-1,1] and L1 outside, smooth everywhere
- weight pixels differently whether they are interpolated or real

$$L = \lambda_v \left\|(p - t) \circ m\right\| + \lambda_h \left\|(p - t) \circ (1 - m)\right\|$$

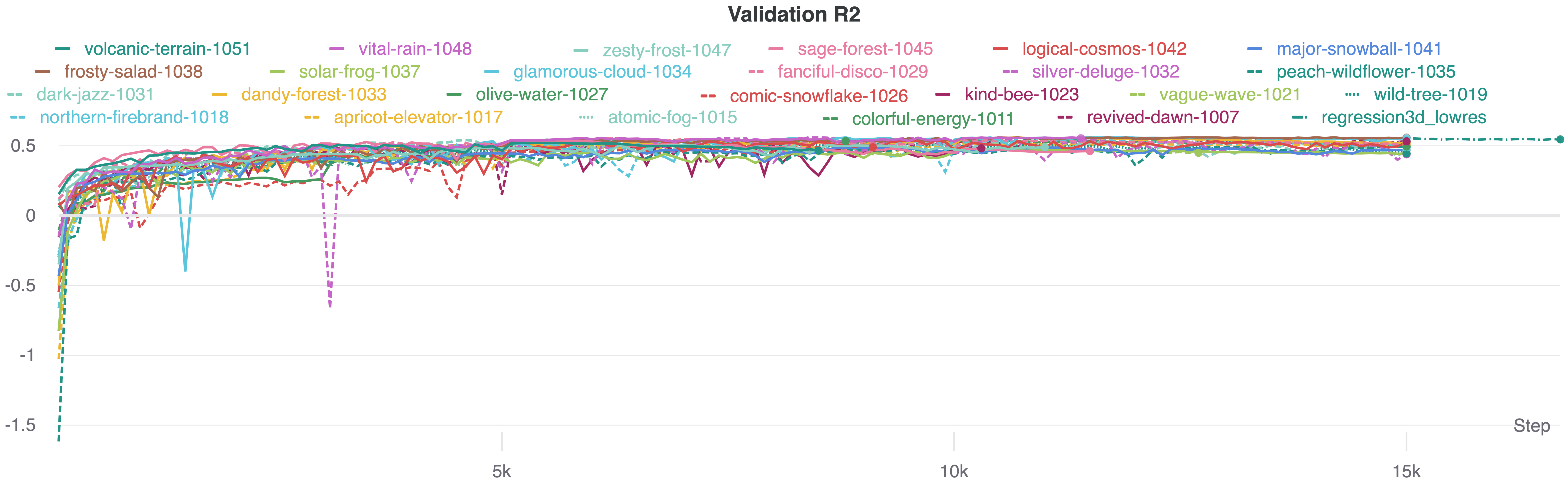
R2-Score

- can be used it to score the quality of the predictions w/ respect to the intrinsic variance
- R2<0: low accuracy
- R2~=0: predictions consistent with noise
- R2~=1: high accuracy

$$R2 = 1 - \frac{\sum_i (\mathbf{y}^i - \mathbf{f}^i)^2}{\sum_i (\mathbf{y}^i - \bar{\mathbf{y}})^2}$$

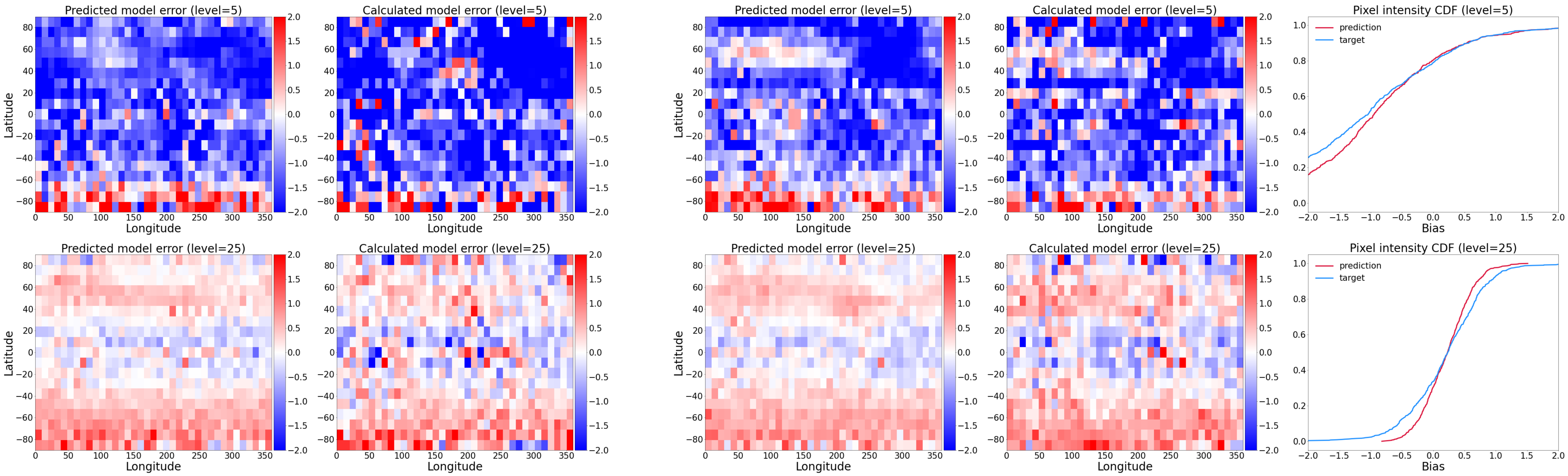
- **used as evaluation metric, not as training criterion**

Training Results: convergence



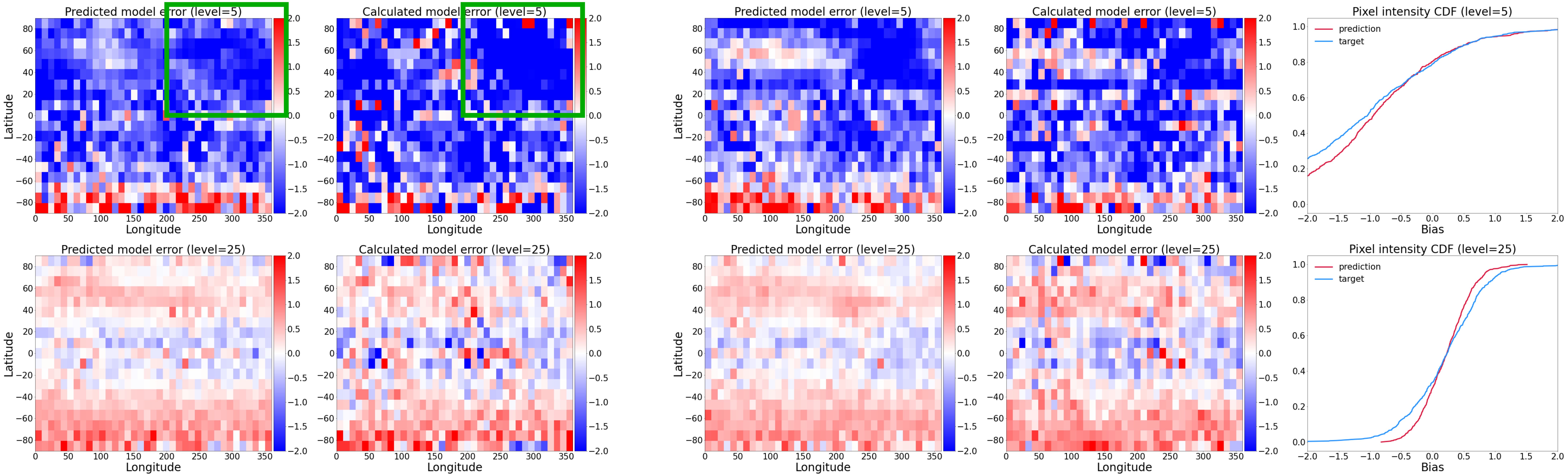
- stable learning for a wide range of hyper-parameters
- only light overfitting although only ~400 training samples available for 10days/10degrees data
- select networks with best performance and most stable convergence curves (qualitative)

Training Results: prediction quality 10days/10degrees



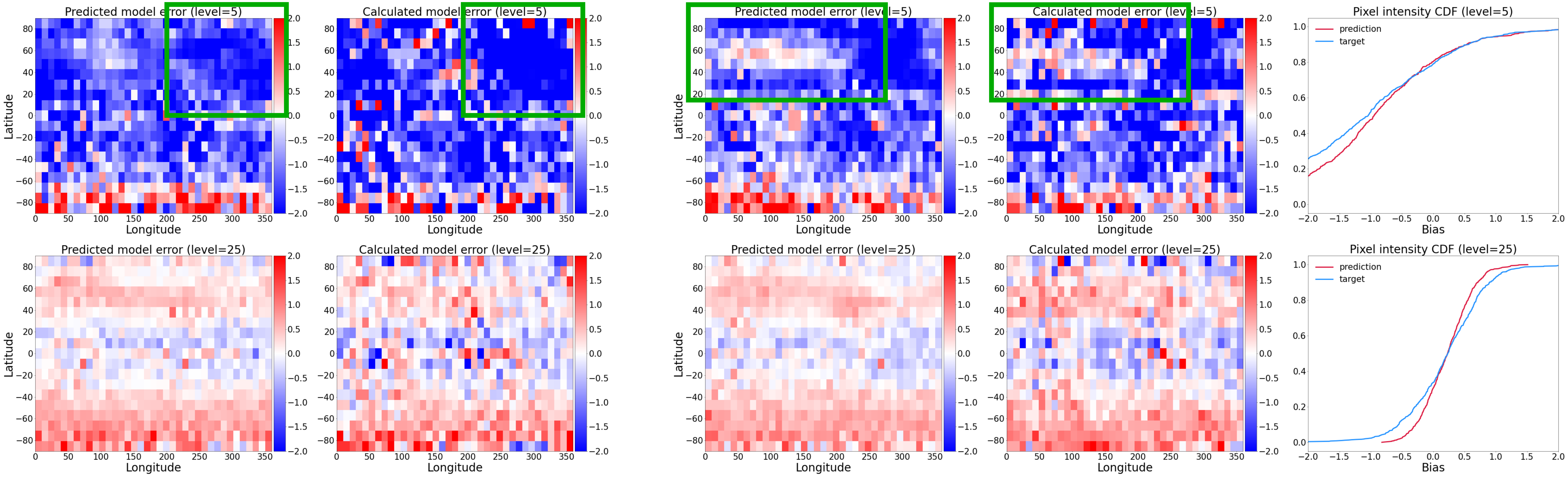
- NN can reproduce important features across levels
- scale of predicted bias matches computed bias

Training Results: prediction quality 10days/10degrees



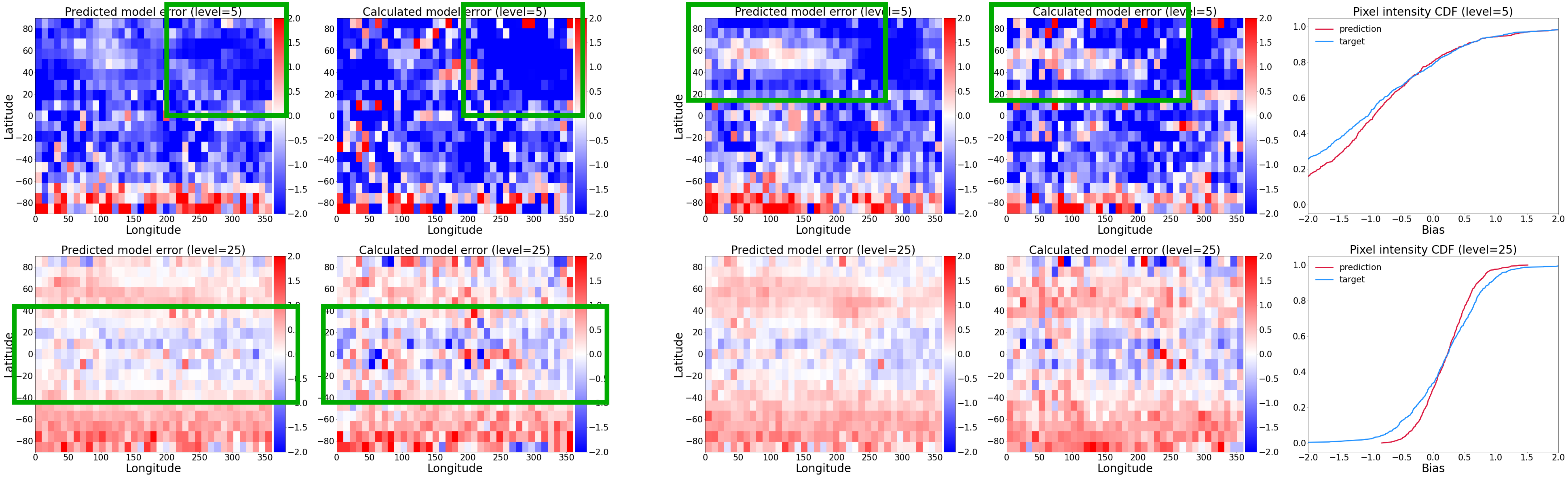
- NN can reproduce important features across levels
- scale of predicted bias matches computed bias

Training Results: prediction quality 10days/10degrees



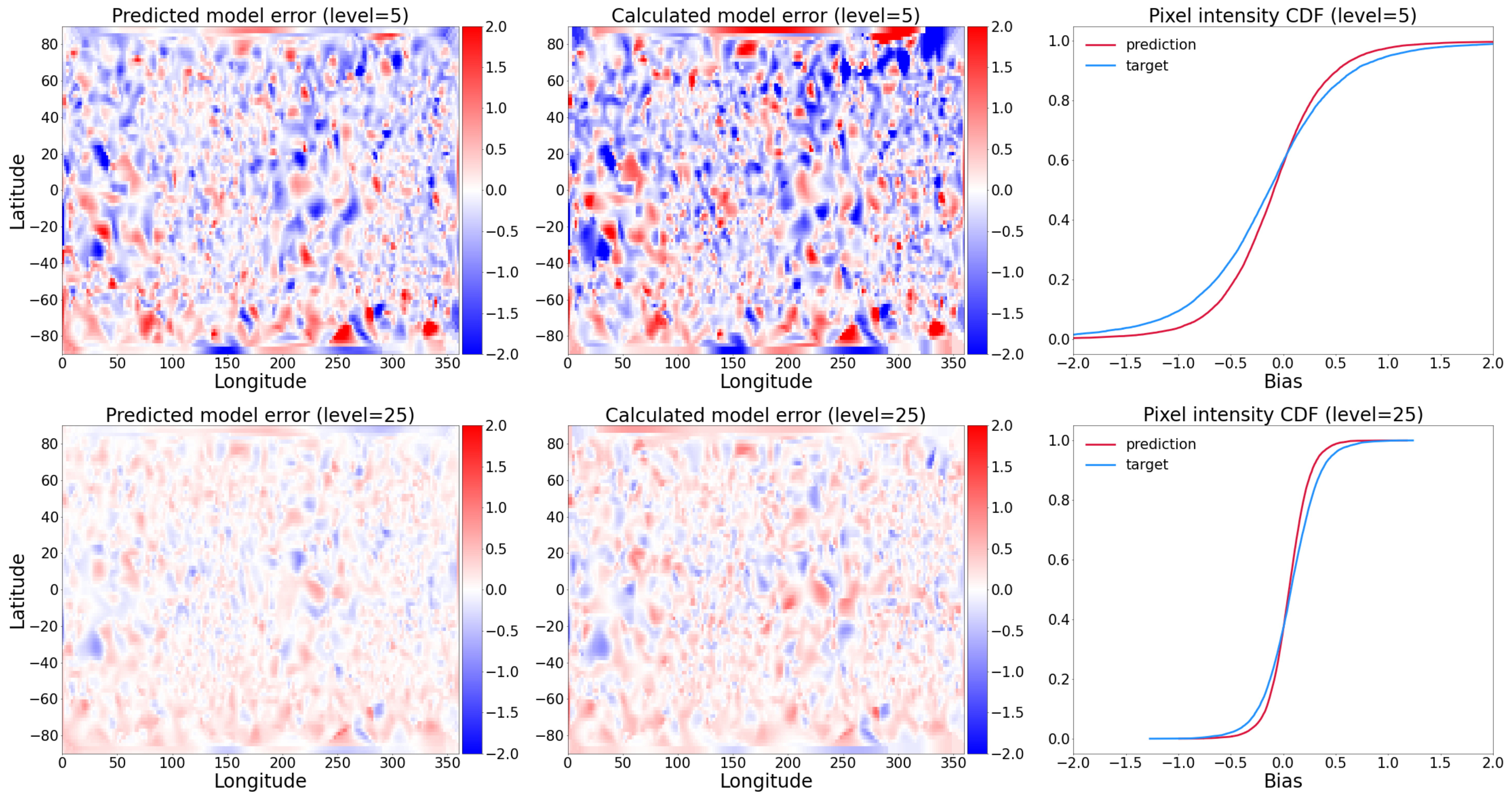
- NN can reproduce important features across levels
- scale of predicted bias matches computed bias

Training Results: prediction quality 10days/10degrees



- NN can reproduce important features across levels
- scale of predicted bias matches computed bias

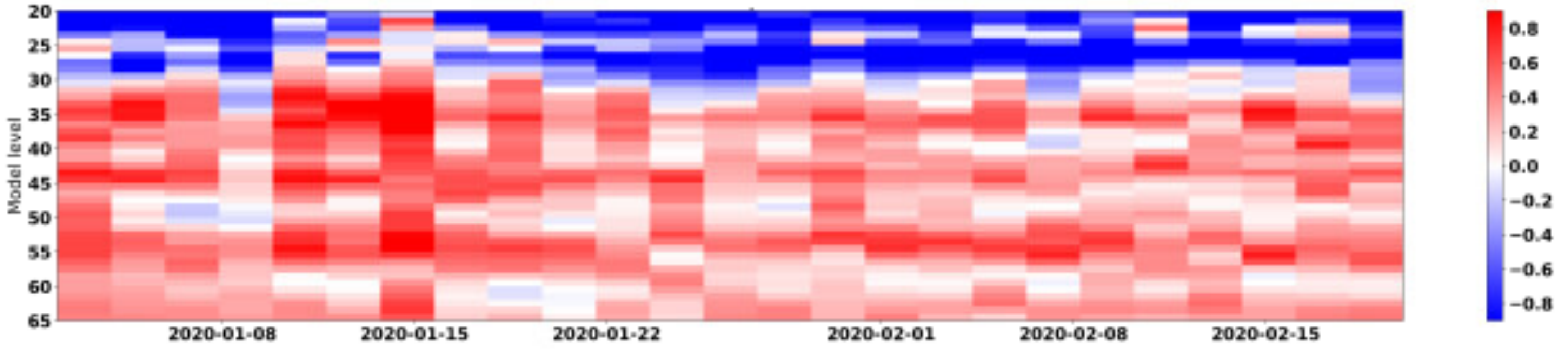
Training Results: prediction quality 2days/2degrees



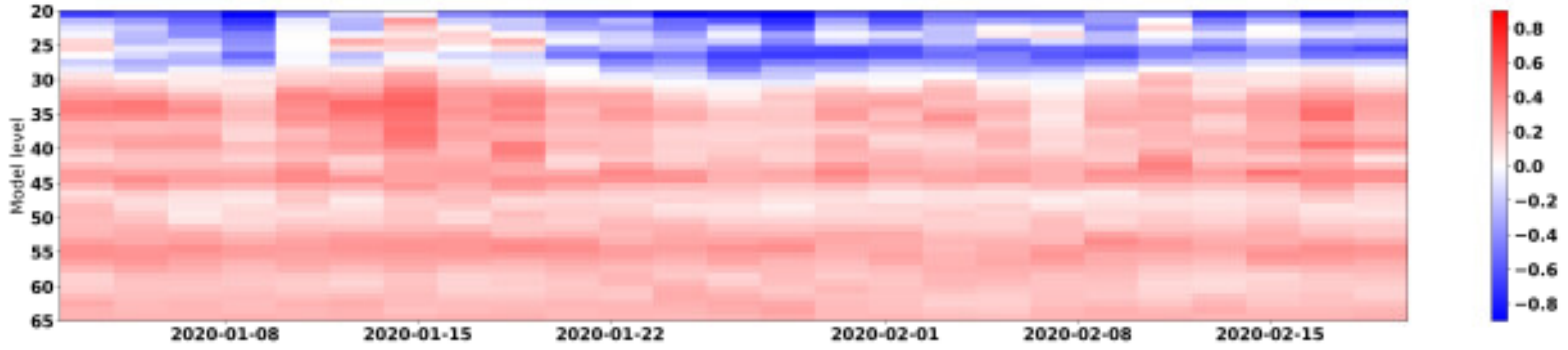
- noticeable features can be reproduced by the NN

Training Results: time series

prediction



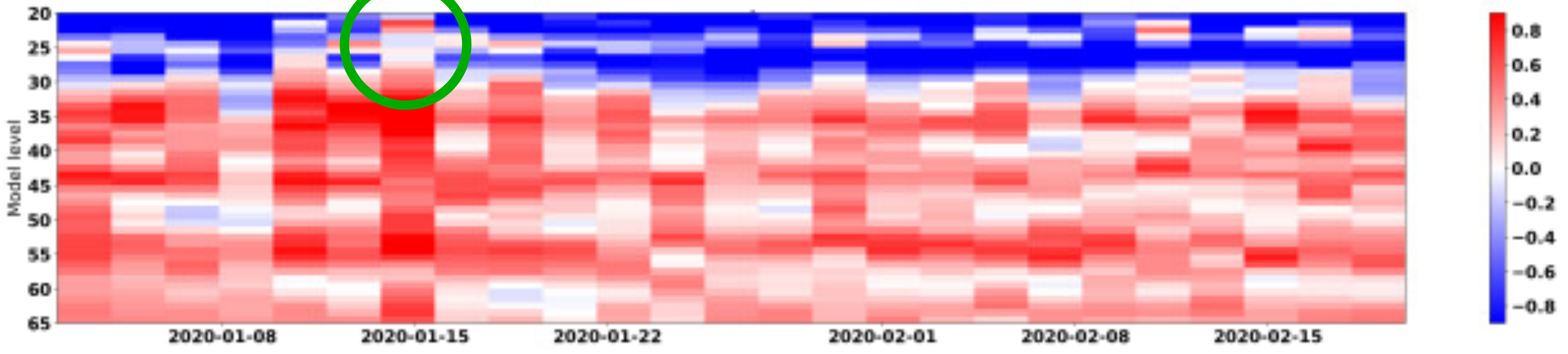
target



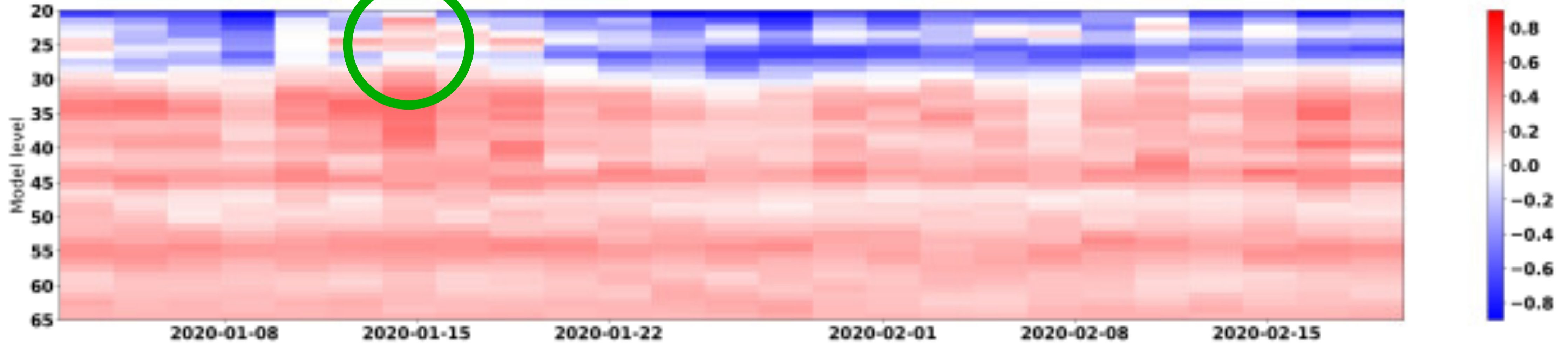
- NN picks up relevant features
- increased resolution would be great
- (overall magnitude difference: normalization issue in post-processing, corrected in future plots)

Training Results: time series

prediction

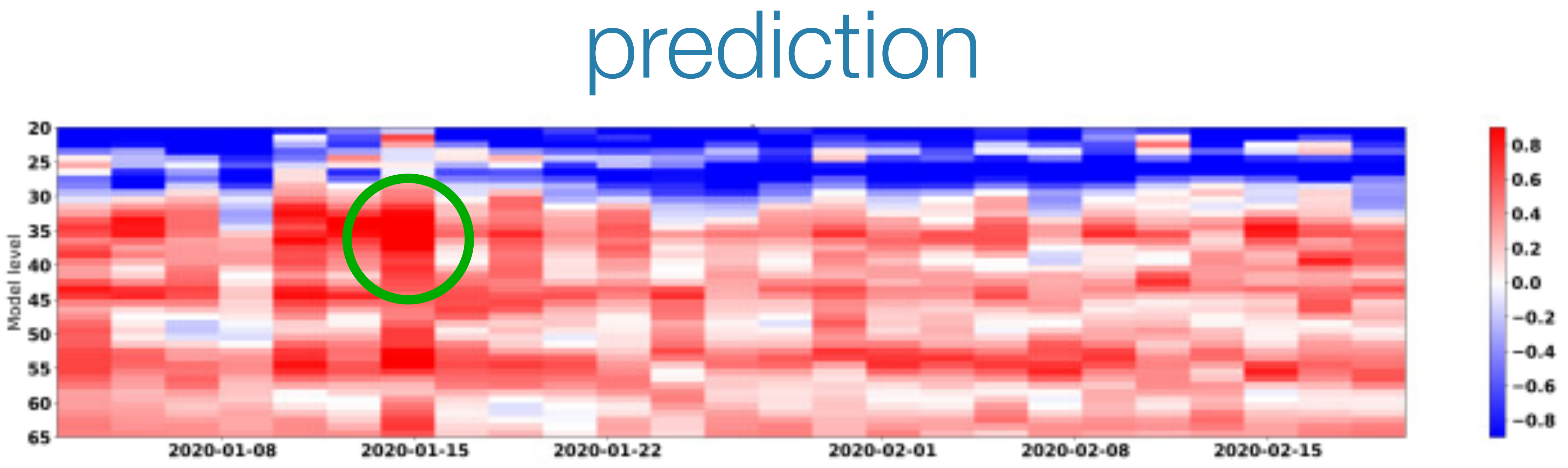


target



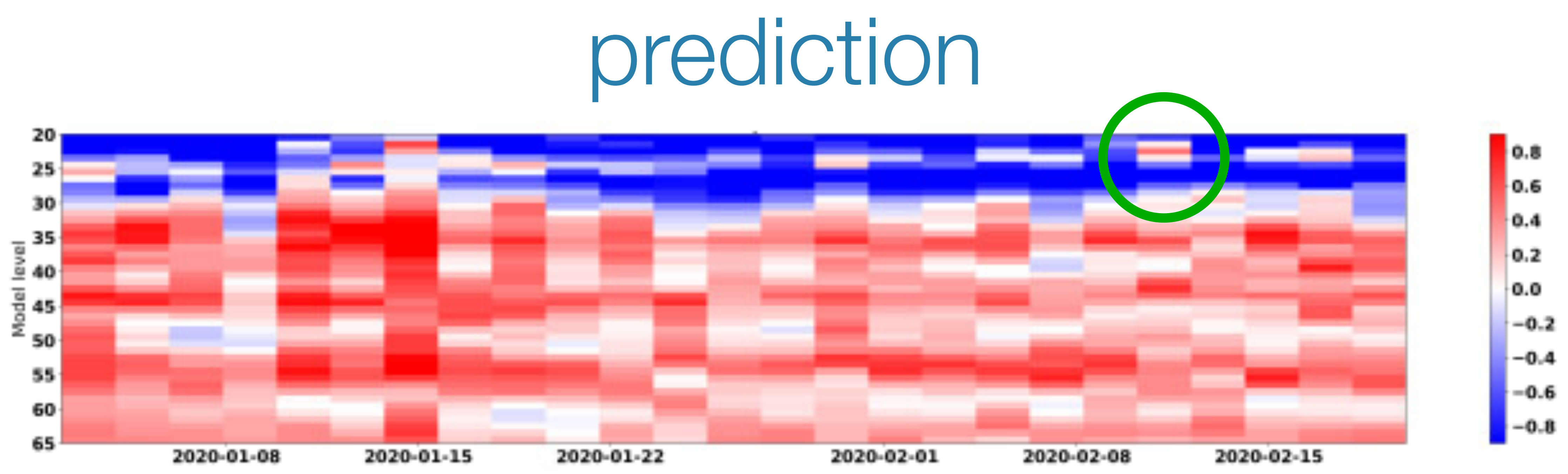
- NN picks up relevant features
- increased resolution would be great
- (overall magnitude difference: normalization issue in post-processing, corrected in future plots)

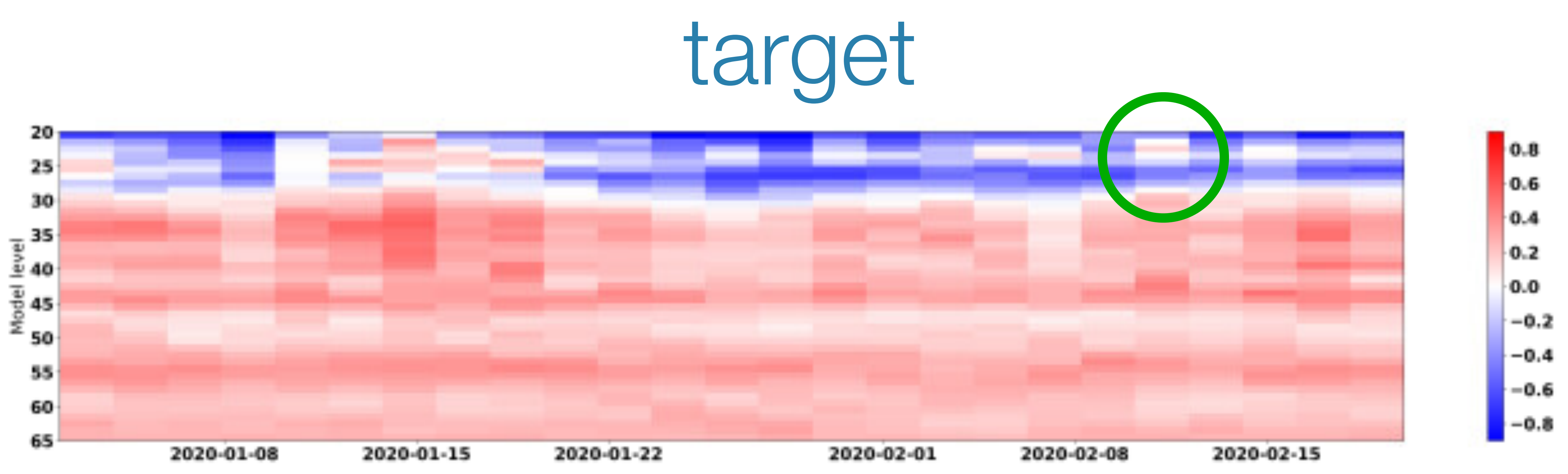
Training Results: time series



- NN picks up relevant features
- increased resolution would be great
- (overall magnitude difference: normalization issue in post-processing, corrected in future plots)

Training Results: time series

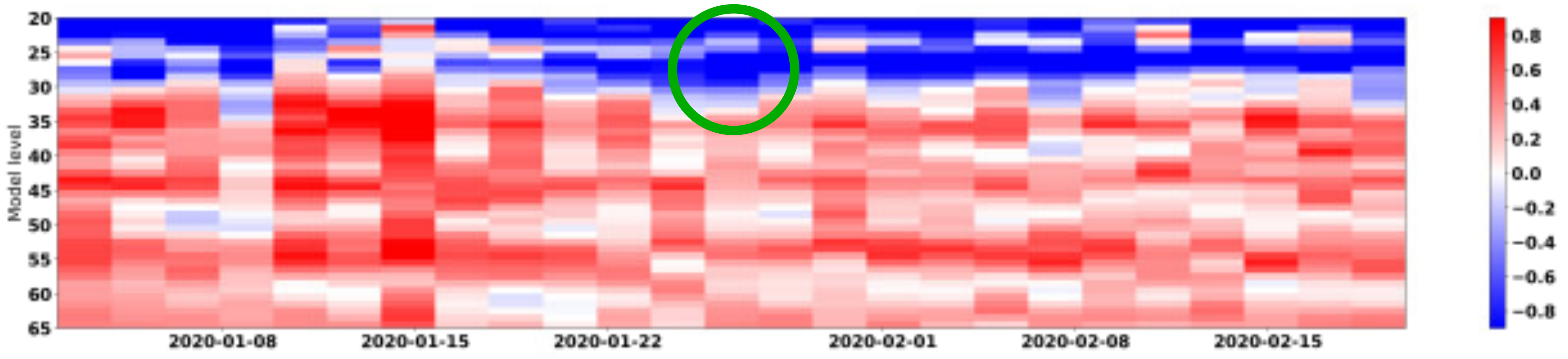




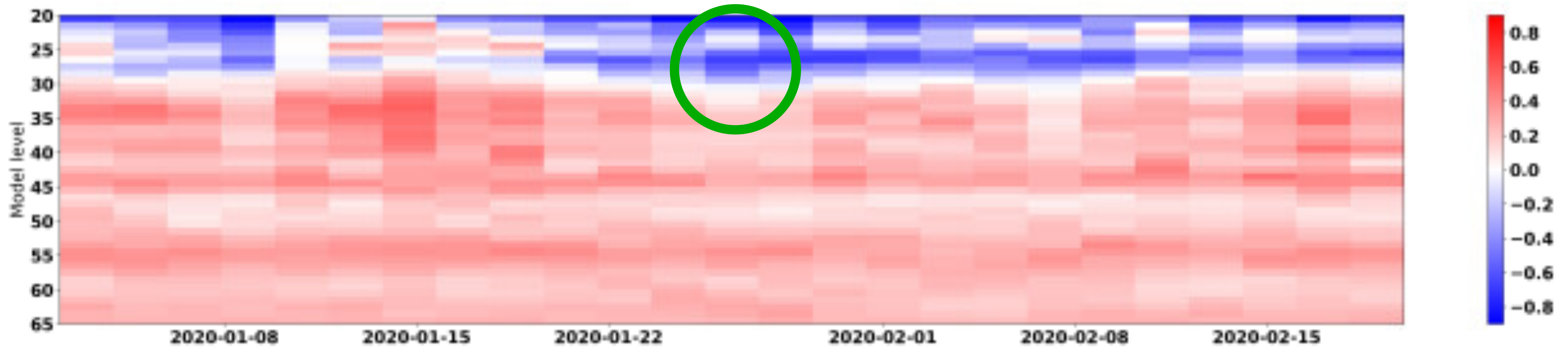
- NN picks up relevant features
- increased resolution would be great
- (overall magnitude difference: normalization issue in post-processing, corrected in future plots)

Training Results: time series

prediction



target



- NN picks up relevant features
- increased resolution would be great
- (overall magnitude difference: normalization issue in post-processing, corrected in future plots)

Summary and Outlook

- proposed neural network is learning model bias, including seasonal features
- bias could be corrected in DA initial conditions
- sensitivity studies could help to learn more about the properties of the model bias itself
- high resolution data more noisy, trade-off between resolution and noise
- sparsity: loss masking works, but for high resolution sparse to dense approaches (graph to image) might be preferable
- next step: investigate how to integrate the NN into the full DA pipeline

Thank You