# Developing a data-driven emulator of an ocean model


Rachel Furner

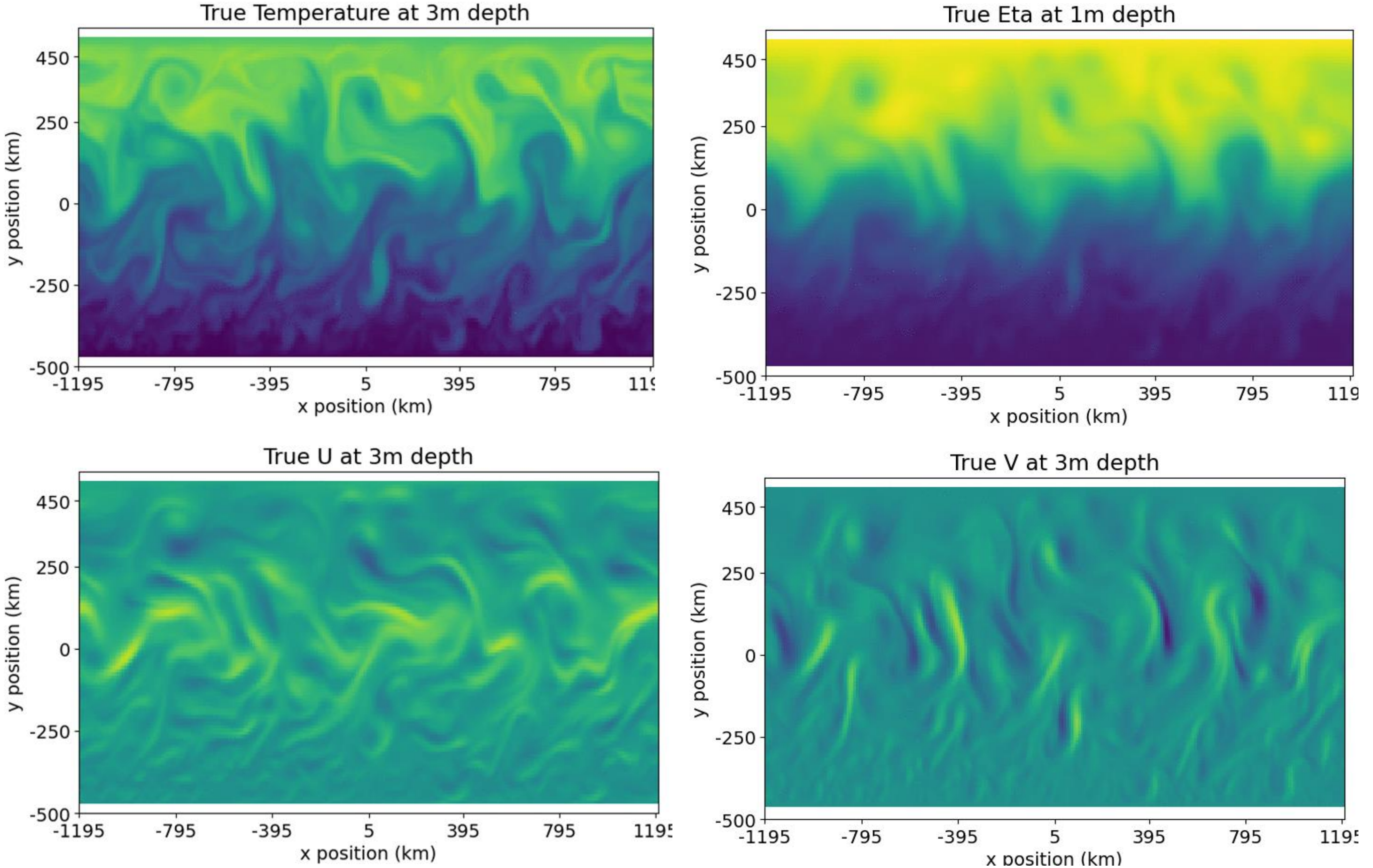
**Supervisors:** Peter Haynes (DAMTP), Brooks Paige (UCL), Dan Jones (BAS),

Dave Munday (BAS), Emily Shuckburgh (Computer Science)

UNIVERSITY OF CAMBRIDGE

British Antarctic Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

# My problem - trying to model a channel of water with jets

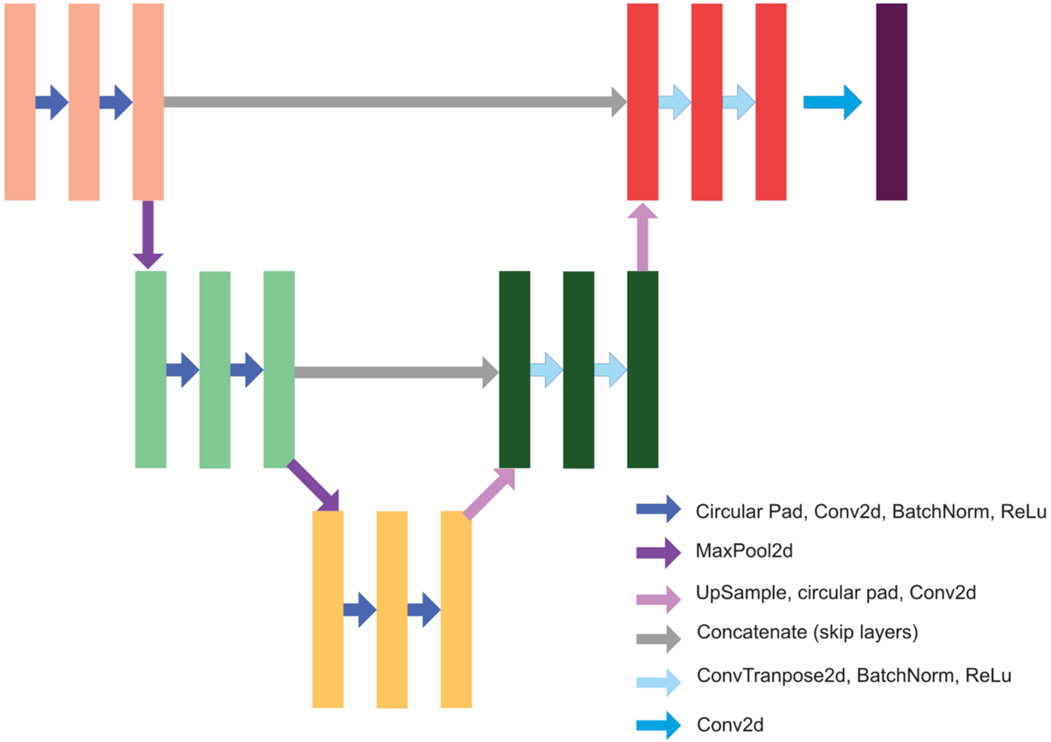



We run an MITgcm configuration of a simple ocean channel, with wind forcing & periodic boundaries.

Can we create a data-driven model which learns from the output of the process based simulator, to predict the flow one time step (one day) ahead?

Can we iterate that model to forecast further out in time?

# Approach with a U-Net style network




Input and output shapes: 96 x 240 ‘pixels’ per channel,115 channels (each variable at each of 38 depth levels)

Combination of convolutional, max pooling, upsampling, & convtranspose layers

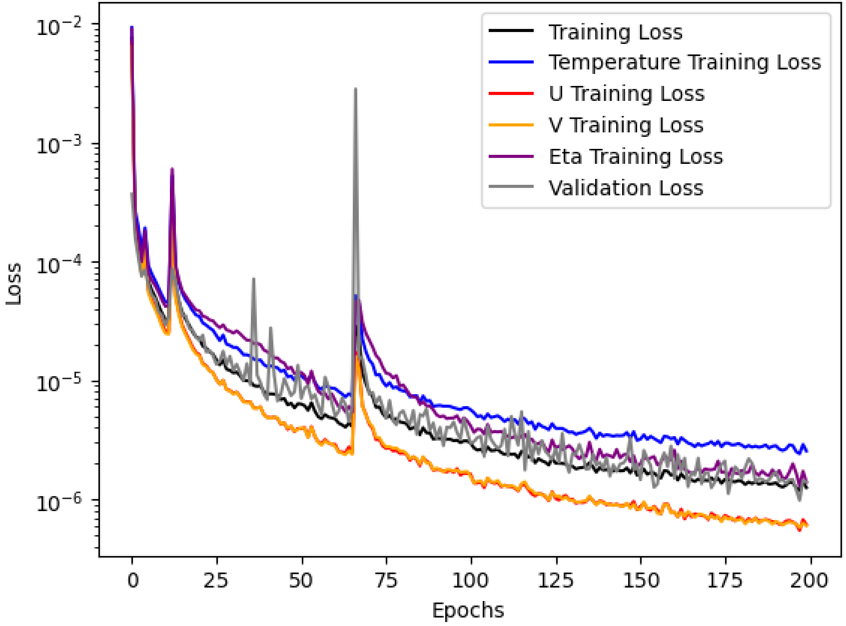
Reduce resolution to 17 by ~70 and increase number of channels to 1024

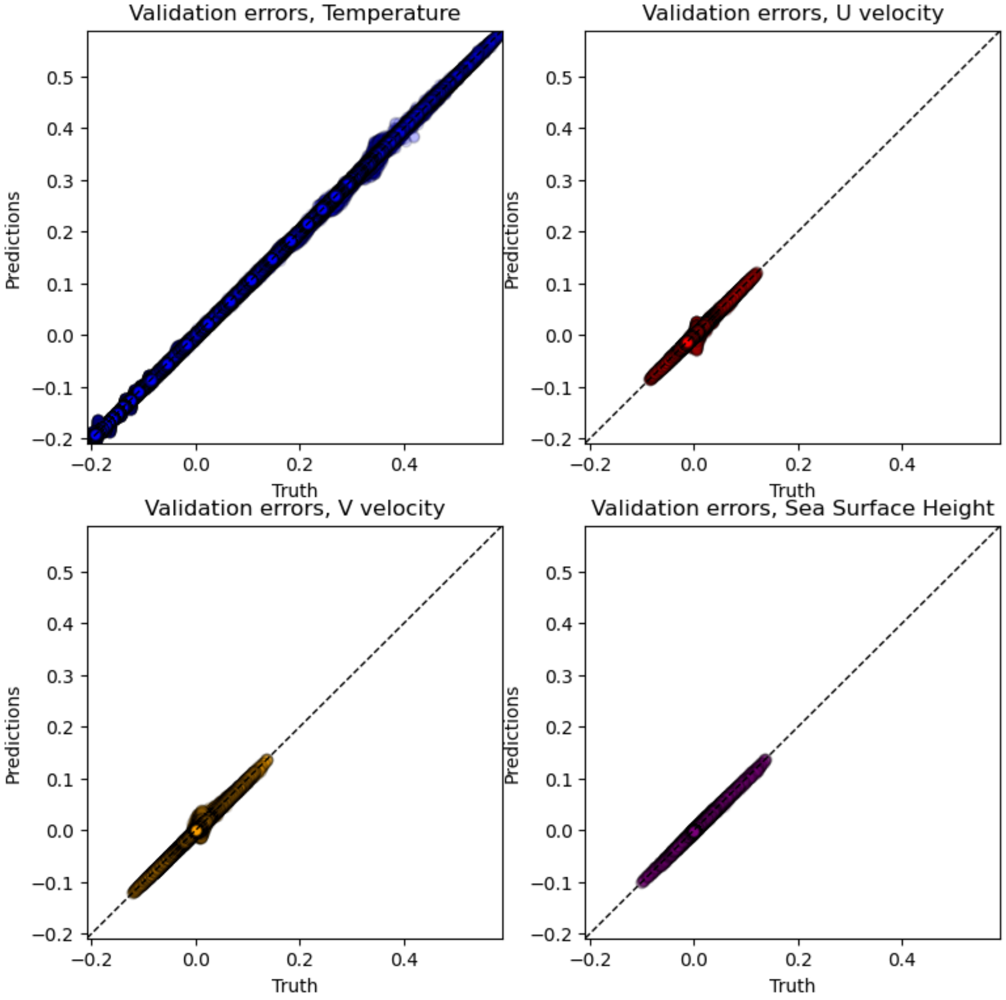
We apply circular padding at East & West boundaries. North & South boundaries aren’t padded.

Use Adam optimiser, learning rate of 0.001

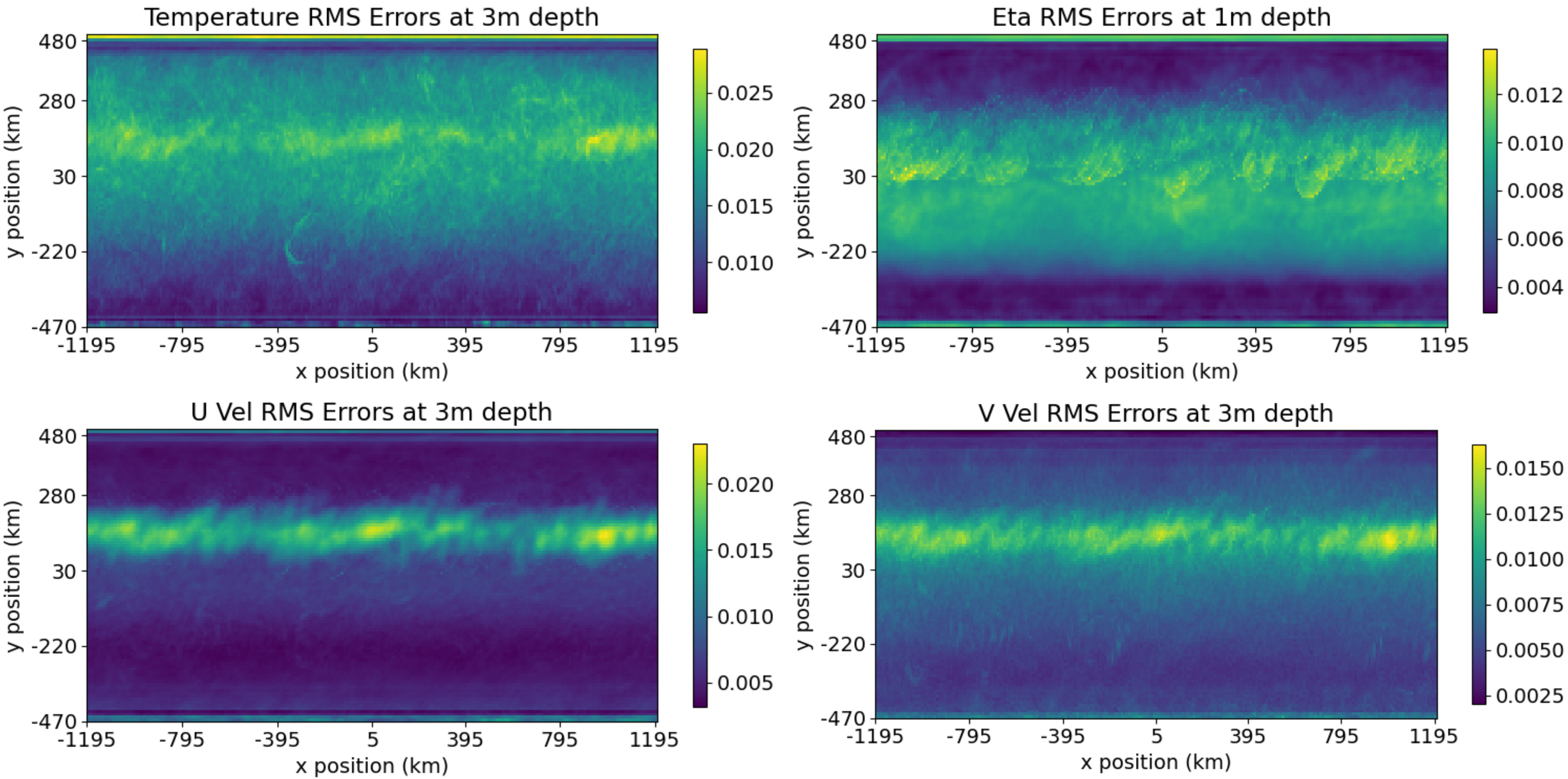
2700 training samples, 540 validation samples

# Training and validation errors over a single prediction

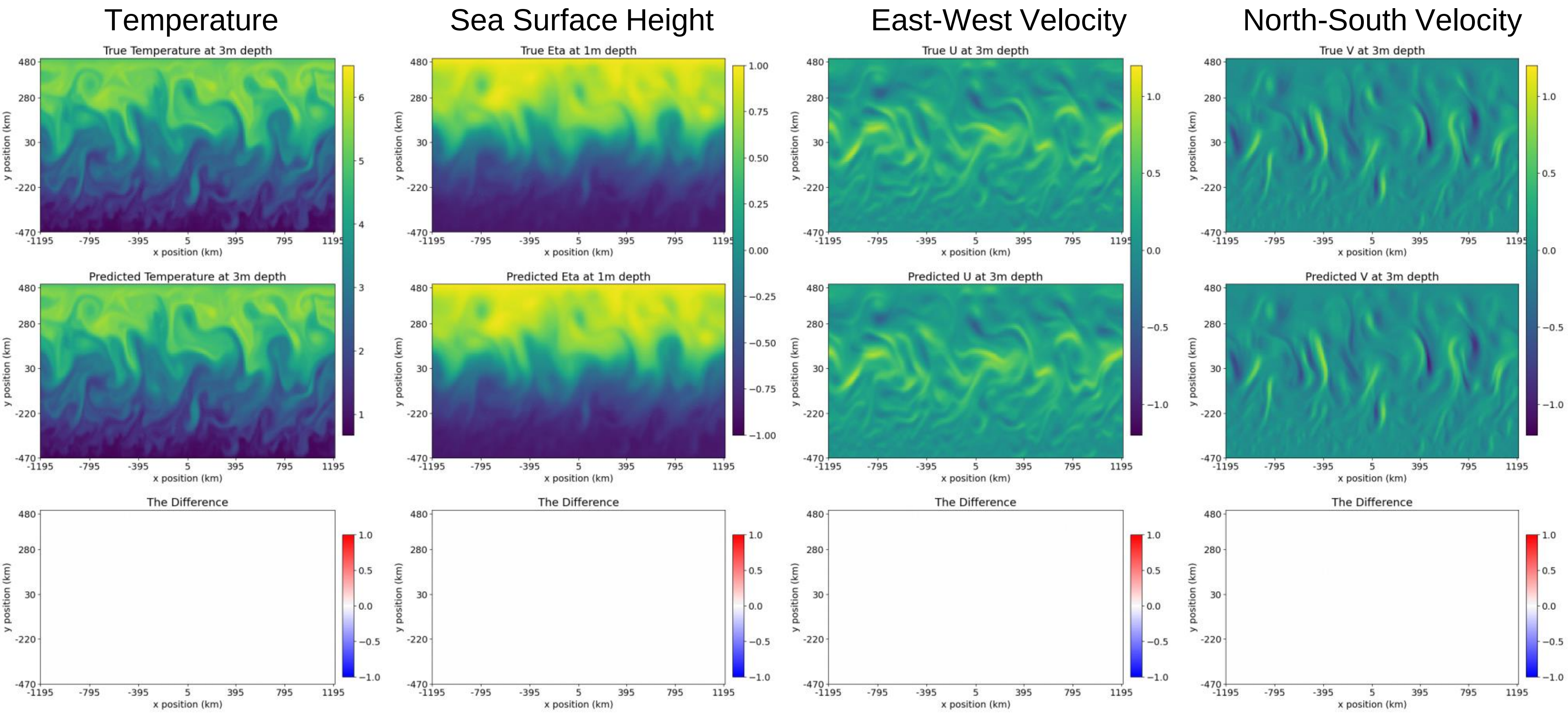


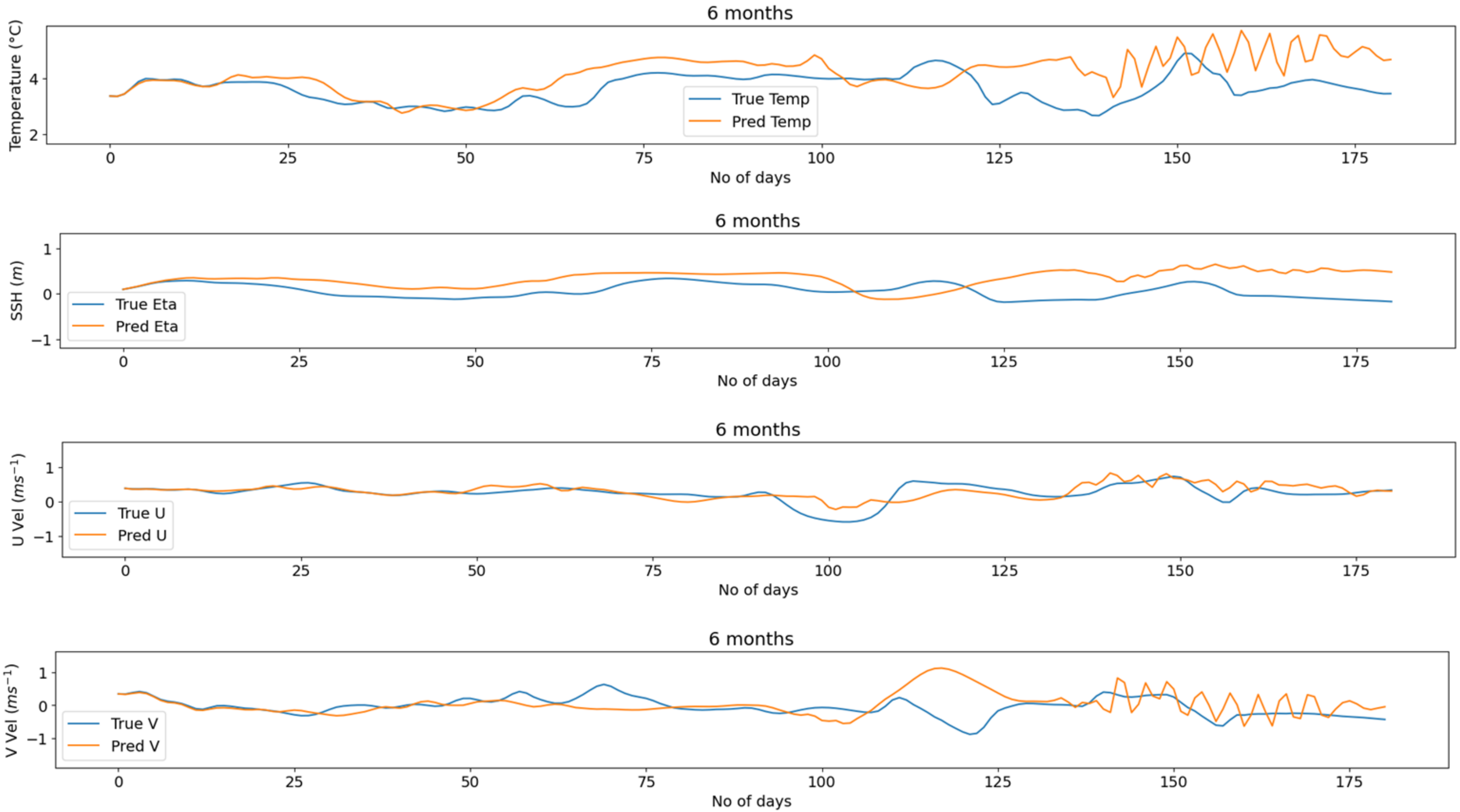
# RMS errors just below the surface, averaged over validation dataset



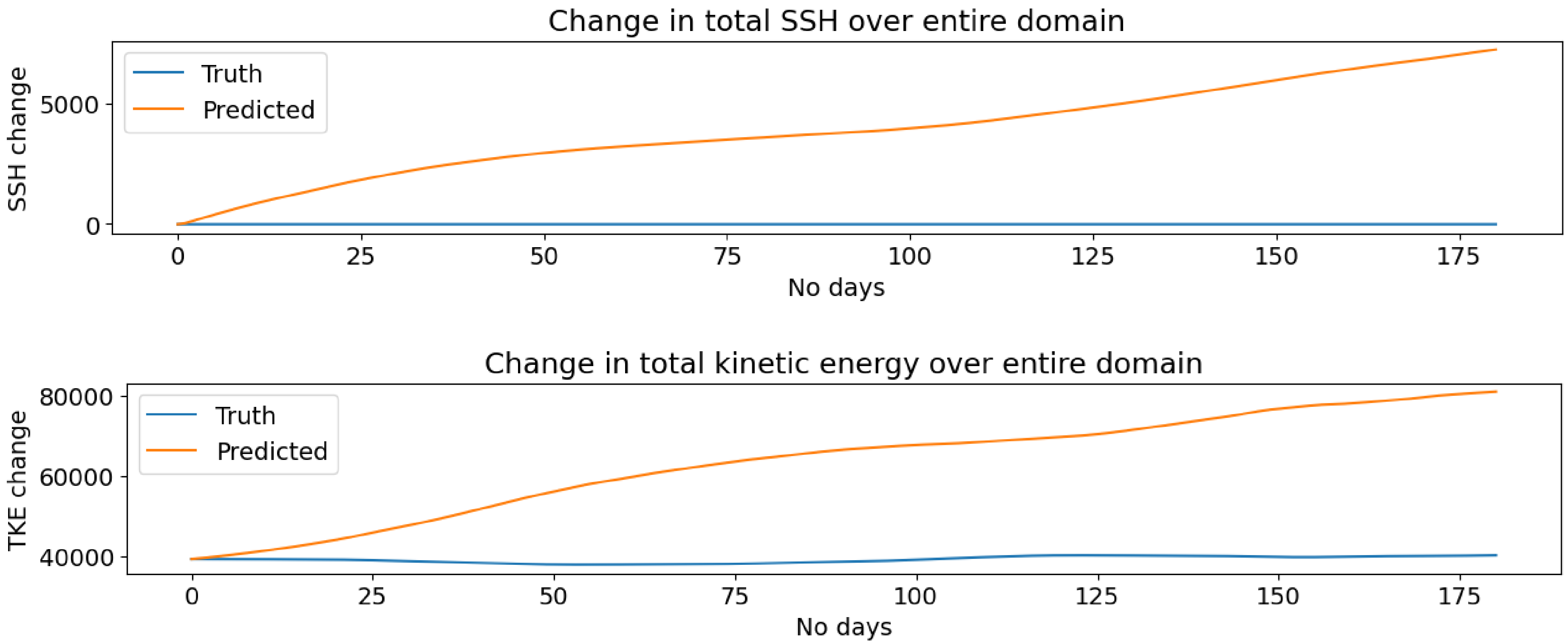
# Results from iterating this model for 6 months



# 6 month timeseries at a point near the surface in the centre of the domain

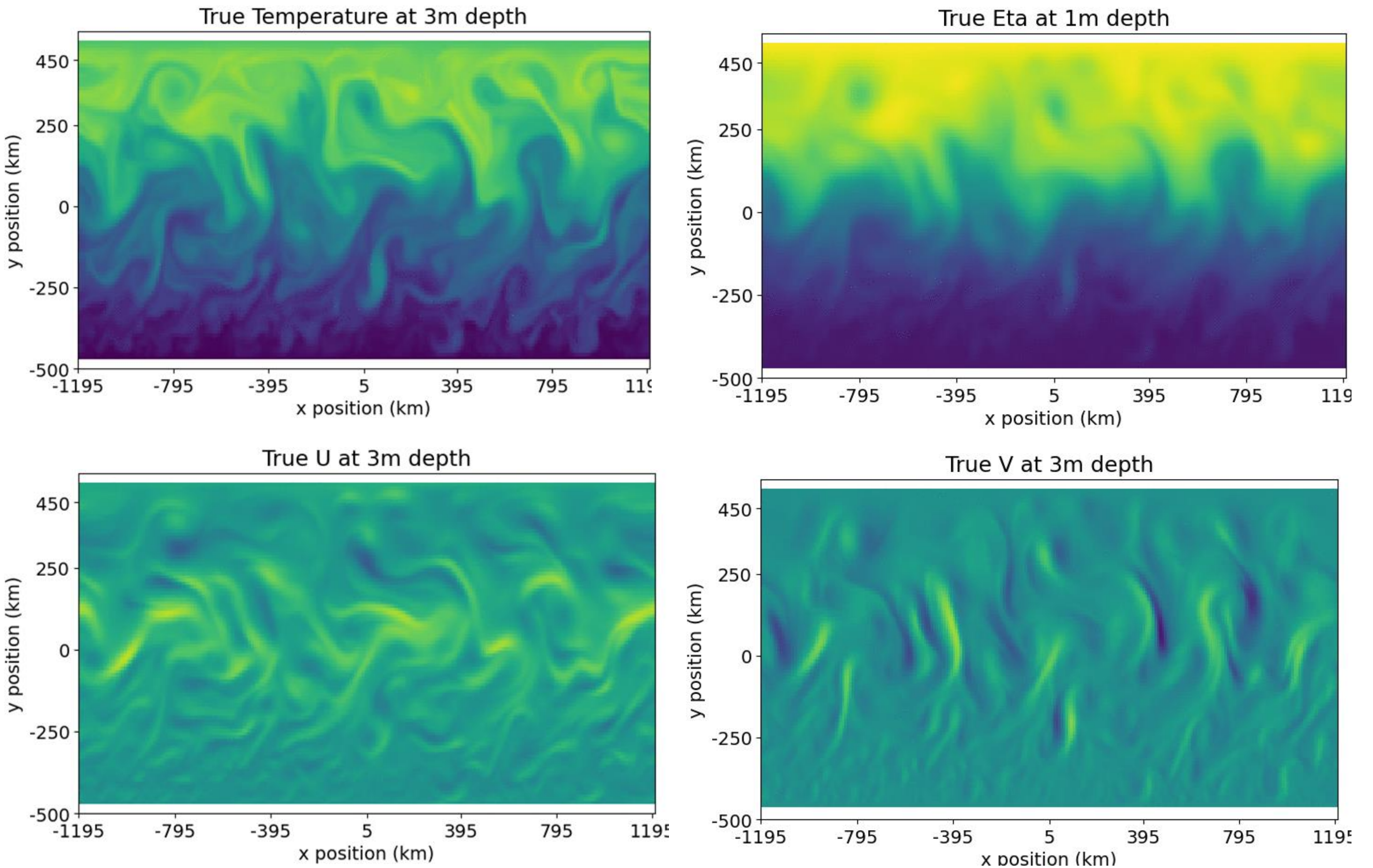


# Conservation of mass and energy



# But what about land?

# Including land in the domain

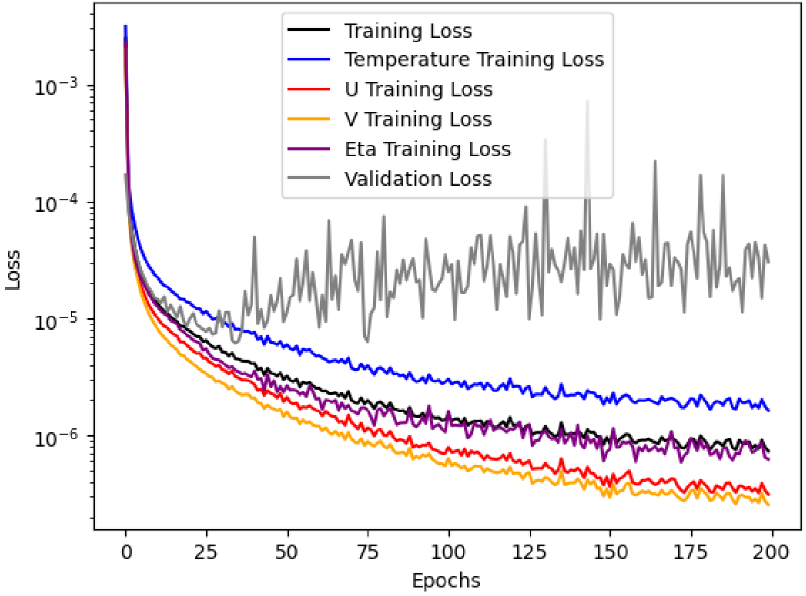


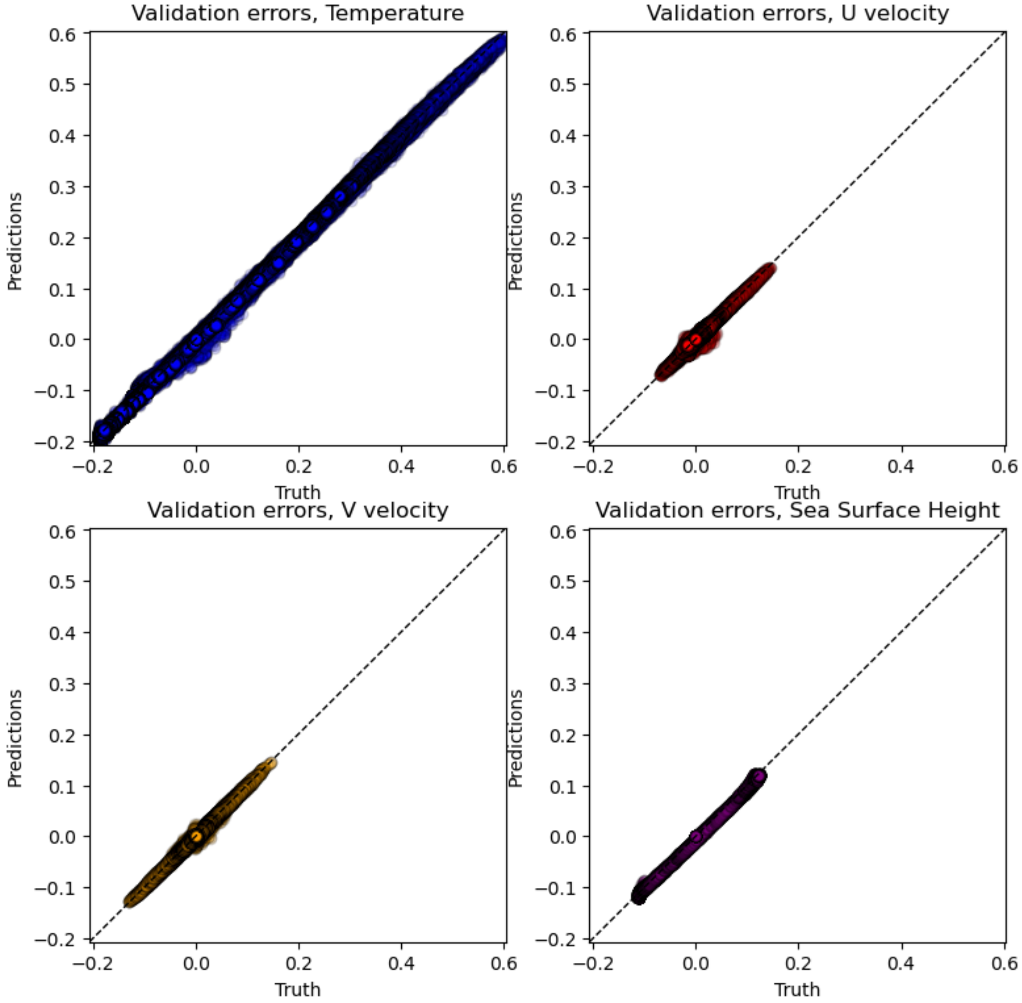

We retrain a new model - this time on the entire domain, including the land at the North and South

We add mask channels to the network inputs

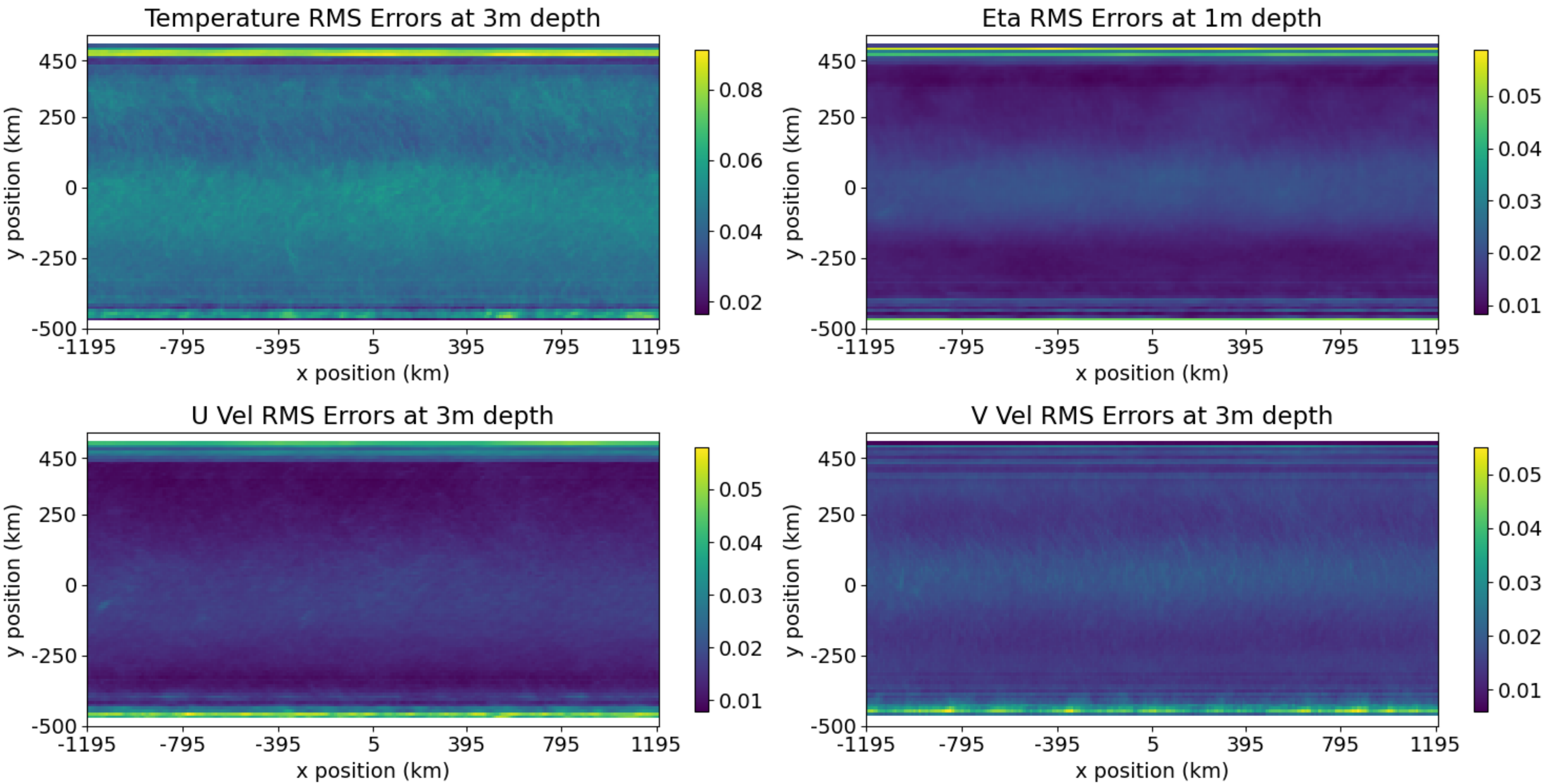
We calculate the loss function over ocean points only, as we don't care what the model does over land (we mask output before using)

# Training loss and training errors over a single prediction



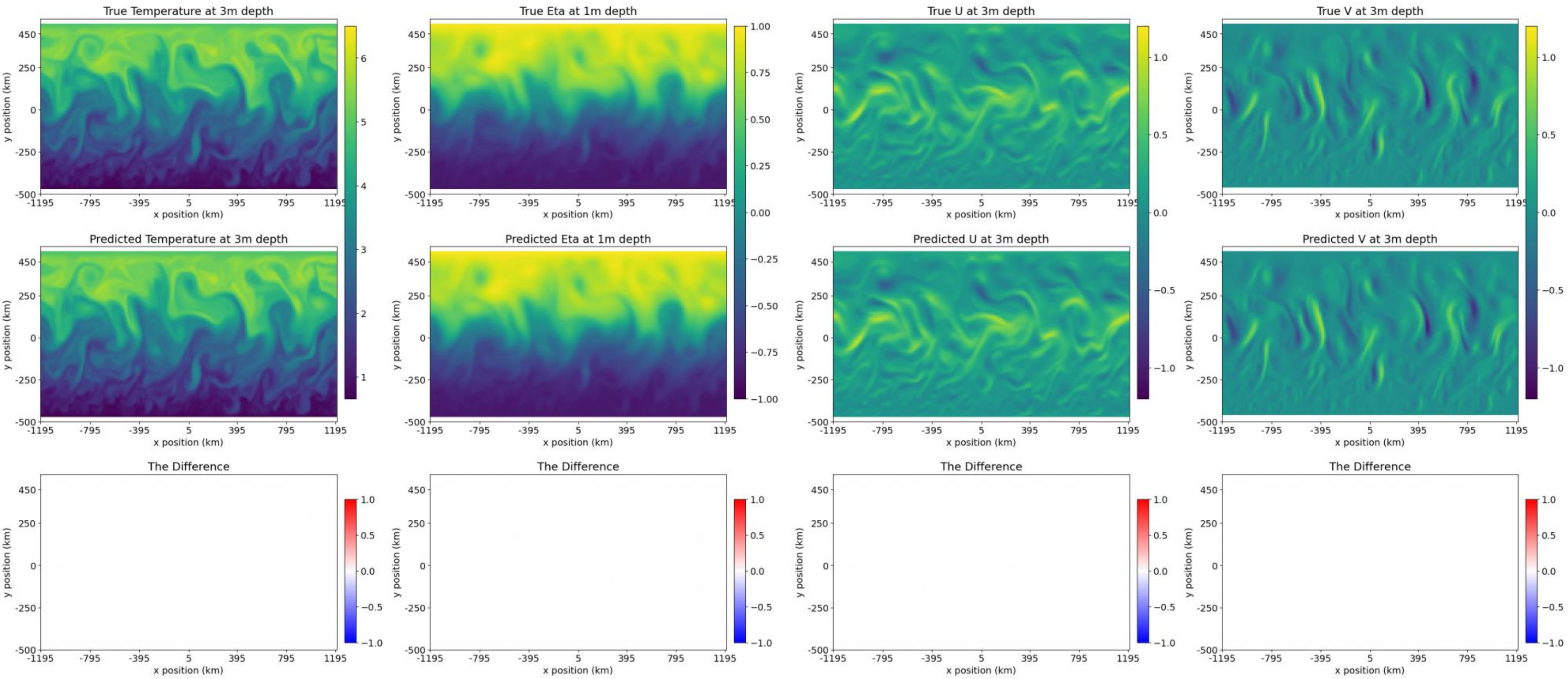



# RMS errors just below the surface, averaged over validation dataset

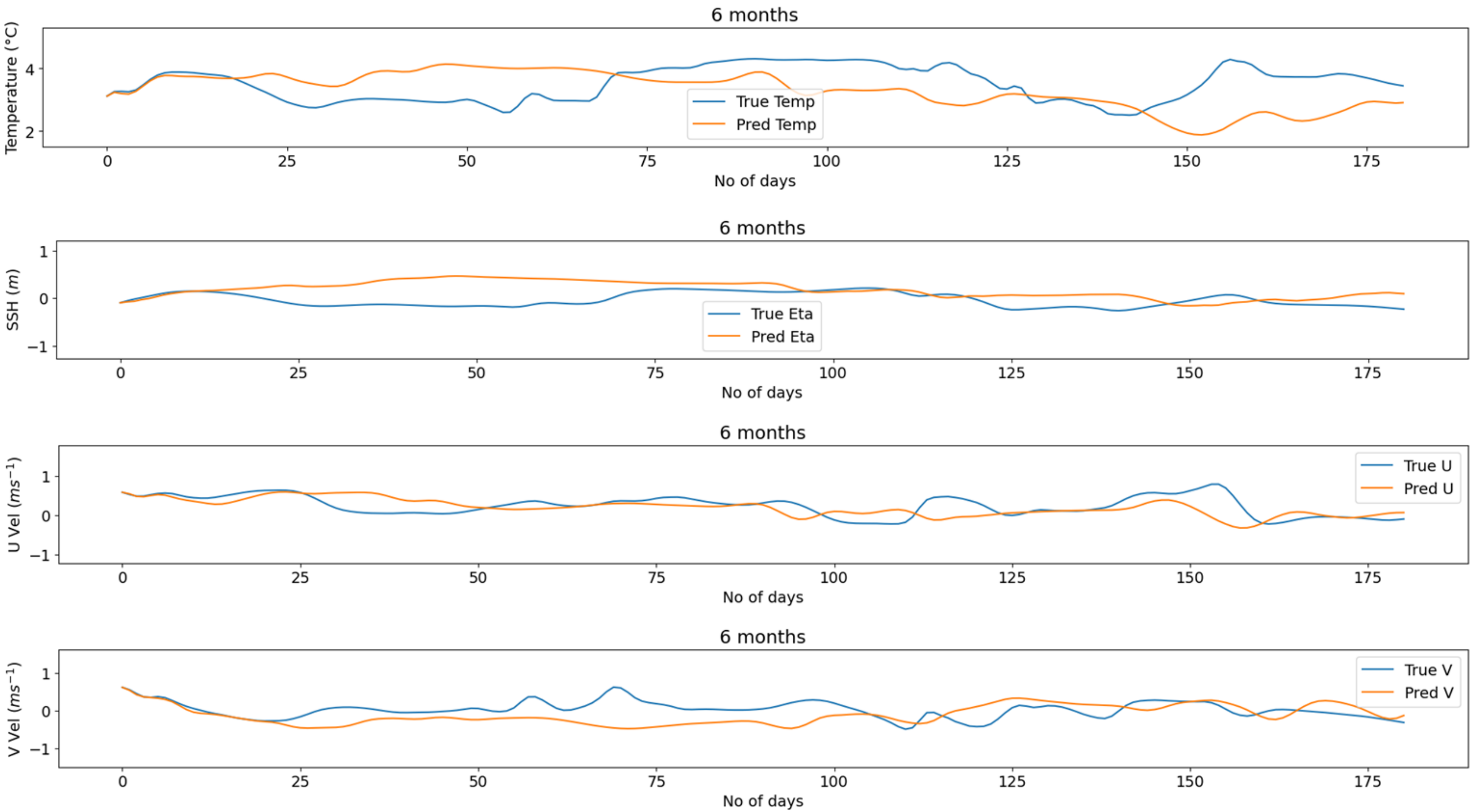


# Results from iterating this model for 6 months

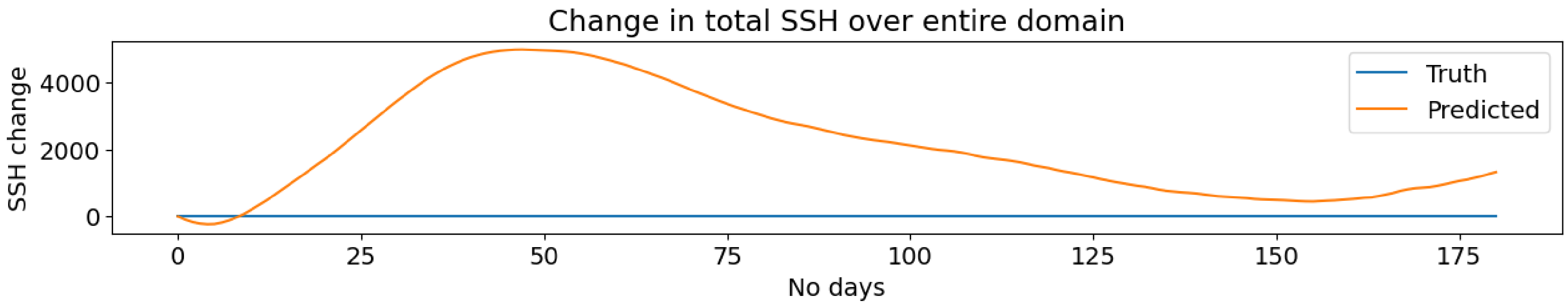
Temperature | Sea Surface Height | East-West Velocity | North-South Velocity

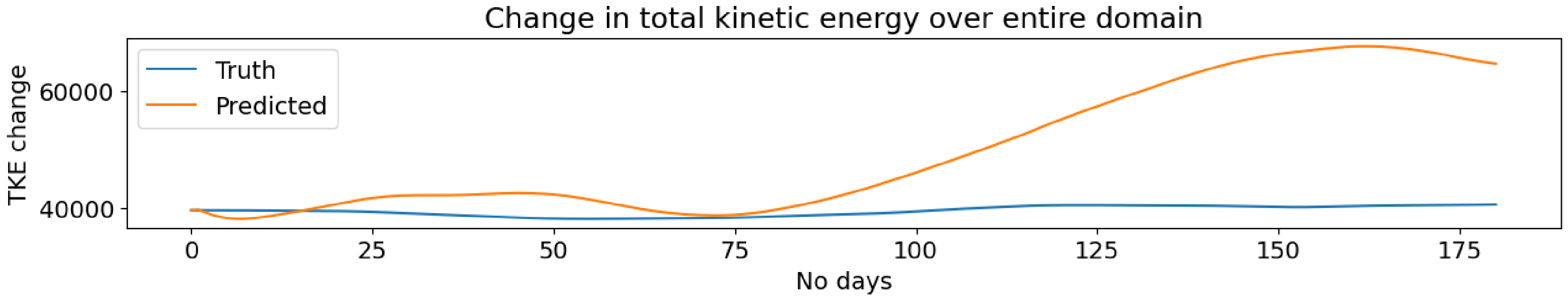


# 6 month timeseries at a point near the surface in the centre of the domain



# Conservation of mass and energy



# Summary and further challenges

- Results from this idealised channel configuration show promise for predicting change over a single time step, when we have an ocean only configuration
- Iterating the model works well for short forecast periods, but leads to instability over time. More consideration needs to be given to how to best iterate these sorts of systems:
    - Consider loss function we are using (which emphasises one-time step error)
    - More sophisticated iterations methods possible
    - Add physical constraints to the model - especially as we see key metrics are not conserved
- Moving to a configuration which includes land proves more difficult:
    - Overfitting becomes an issue
    - Model struggles near land-sea boundary
    - But the model seems more stable when iterated, but at the cost of accuracy
- Moving to a more realistic ocean poses further additional challenges:
    - The ocean is not a cuboid - there are different depths, as well as more complicated coastlines and islands
    - The dynamics varies across space, perhaps additional input such as the latitude (for the coriolis effect) is needed, and/or consideration of how we scale the domain