Using deep learning to predict time-dependent temperature field for 2D mantle convection -CERI 8102-

R.C. Lam

Spring 2022

Abstract

Mantle convection is the fundamental driving agent behind many geological processes. Numerical models are frequently used to investigate this complex, temperature-dependent process but recently deep learning applications in geoscience have gained traction. In this paper, we assess the capabilities of deep learning to recognize patterns and make predictions for a simplistic case featuring a rising spherical anomaly in a homogeneous mantle. Specifically, we use a combined CNN-LSTM neural network. The CNN layers extract features, and LSTMs are well-known for their suitability for time-series forecasting. We evaluated the neural network using data generated by the numerical solver ASPECT. The 3124 models generated each contained a homogeneous mantle and a spherical anomaly described by five varying parameters: reference mantle viscosity, temperature gradient across the anomaly ΔT, anomaly radius (size), anomaly radial location, anomaly angular location. Using those parameters and the time series temperature-field data extracted from the simulations, the neural network was able to predict the temperature fields for a later time step with an R2 score of 0.76 and 0.0013 mean squared error.

1 Introduction

We as a scientific community still have so much to learn about mantle properties and convection (i.e. dynamics, rheology, composition). Not only will further study of this enigmatic system give insights into the basic physics and evolution of our planet but it can also give insights plate tectonics, volcanic activity, and more. Constraining mantle properties and initial conditions can also improve non-unique interpretations for geologic structures. That is, it can help narrow the field of possible causes for some observed effect.

The reason why we still have so much left to learn about mantle convection is because of the complicated physics that govern it. Numerical solvers are required to solve coupled nonlinear partial differential equations and can be very computationally expensive. However, in recent decades, neural

networks (NNs) have gained popularity across all disciplines, including geoscience. NNs are useful for handling large amounts of data, particularly for multivariate problems that aren't fully understood. Convolutional neural networks (CNNs) are useful reducing data size and for extracting data features without needing to be told what exactly to extract. Long short-term memory (LSTM) networks are a special sub-category of recurrent neural networks (RNNs). They are well-suited for time-series prediction predictions without the drawbacks of standard RNNs.

Neural networks have been used in a number of different geodynamic problems. Atkins (2017) used an artificial NN to map convection simulation patterns to mantle characteristics which she later used to interpret geophysical interpretations, such as quantifying the probability of different non-unique interpretations. Gillooly et al. (2019) attempted to use machine learning to capture how tectonic plates are organized and interpolate plate boundaries for poorly-constrained areas. Agarwal et al. (2020) used neural networks to build a forward surrogate model for a Mars-like planet capable of predicting 1D temperature profiles with an accuracy of 99.7%. Similar to Agarwal et al. (2020), the objective of the experiment presented in this paper is also a forward temperature prediction problem. However his study used machine learning to predicted convection sequences for a given set of parameters, whereas this study will use a numerical solver to simulate convection sequences and use that as input for a neural network. In this paper I will discuss how I set up and generated my models, the architecture of my neural network, and my results.

2 Method

2.1 Model set-up

We considered mantle convection for a 2D cylindrical domain with a regular-spaced grid. The grid resolution was 32 radial layers by 96 cells per layer with a 4x initial mesh refinement, giving a total of 12288 data points. This ended up being too fine, so we cleaned the data and reduced the data points to 3201 (see Figure 1). We assumed free-slip boundary conditions for all sides (no flux across boundaries and zero tangential force), shear heating, and a radially-constant gravity. Total run time was 9 Ma with outputs every 0.5 Ma. Computations were done using the ASPECT (Advanced Solver for Problems in Earth's ConvecTion) code version 2.3.0 (see Kronbichler et al. (2012); Heister et al. (2017); Bangerth et al. (2021a,b)). We used the 'simple' material model provided by ASPECT. This model considers density ρ and viscosity η as functions of temperature and pressure, and it considers all other variables constant. Neglecting the c term, the simple material model solves for ρ and η using:

$$\rho(p, T, c) = (1 - \alpha(T - T_0))\rho_0 + \Delta\rho c_0 \tag{1}$$

$$\eta(p, T, c) = \eta_0 e^{-\beta(T-T_0)/T_0} \tag{2}$$

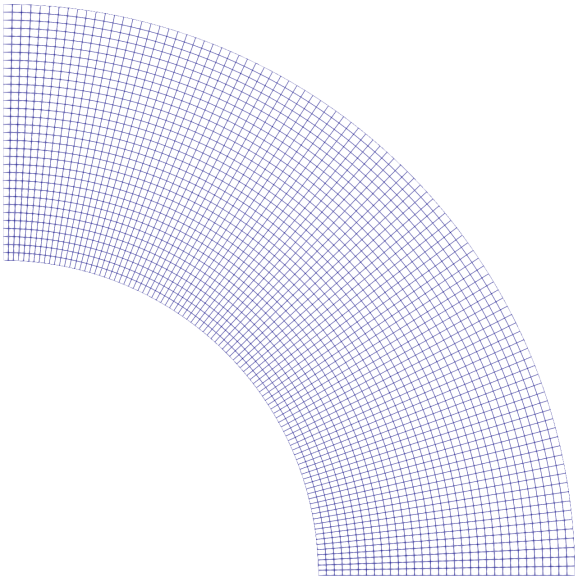


Figure 1: 2D quarter-shell (cylindrical) domain with grid resolution of 32 x 96, regular-spaced cells. Inner radius = 3473 km, Outer radius = 6301 km, Opening angle = 90. Models ran for a total of 9 Ma with outputs every 0.5 Ma. Model assumed shear heating, free slip boundary conditions, gravity constant along radial.

2.2 Problem design and data generation

We designed our simplified mantle convection to be a single layer where the reference state contains a hot spherical anomaly in an otherwise uniform mantle. This type of toy problem is frequently used to study mantle problems and simulations. Our own set-up was largely inspired by the experiments in Liu and Gurnis (2008). As in their study, our anomalies are also hottest in the center with a Gaussian temperature profile across the diameter. We defined the initial temperature condition as:

$$T(x) = \begin{cases} T_m + (T_m \Delta T)\exp(r_d^2/r_c^2), & \text{if } r_d \leq r_c \\ T_m, & \text{otherwise} \end{cases} \tag{3}$$

We generated 3124 models with varying combinations of the parameters given in Table 1. The reference viscosity(η_0) describes the mantle while the other four parameters describe the anomaly (temperature gradient ΔT, size r_c, and location r_d, θ). We converted the anomaly location from Polar to Cartesian coordinates ($x_0 = r_d * cos(\theta)$; $y_0 = r_d * sin(\theta)$) to match the models' coordinate systems. See Figure 2a through 2f for histograms of all parameter distribution. Figures 2a, 2b, and 2e all show a uniform distribution. Figures 2c and 2d are less regular due to the conversion but show a good variety in anomaly location. Figure 2f shows a very skewed data set where most points are reference mantle temperature (3974 °K) which is expected. This skewness does it make it difficult to discern temperature distribution, so Figure 3 shows a smaller window of figure 2f. There does seem to be some Gaussian behavior in the variety of the points with higher temperatures (4100-4400 K) before rapidly increasing

to the ambient temperature. There is a substantial difference in the amount of representation for data points at the ambient temperature to non-ambient temperatures. Nevertheless, these few non-ambient temperature data points are the most integral feature in our data set.

All six parameters (η_0, Δ T, r_d, x_0, y_0, T) were scaled using a MinMaxScaler with feature range constrained from 0 to 1 before being fed into the NN. We also included the point coordinate data (all points, not just anomaly) into our NN input array, so the compiled input had the form (models, time steps, data points, parameters). Each of our models had a total of 18 time steps, so the NN input data contained the information described above for the first 17 time steps. Our target labels were the temperature distribution data for the last time step.

Parameters	units					
Viscosity (η_0)	Pa s	1e21	1.4e21	1.8e21	2e21	2.2e21
ΔT	K	0.099	0.098	0.097	0.096	0.095
Radial distance (r_d)	km	4222	4554.5	4887	52195	5552
Angular coordinate (θ)	°	15	30	45	60	75
Anomaly radius (r_c)	km	700	650	600	550	500

Table 1: Parameters

2.3 Neural Network

We built a neural network containing 12 layers: two convolutional, two pooling, one time distributed flattening layer, four LSTM layers, a repeat vector, one dropout layer, and one fully connected dense output layer. See Table 2 for the model summary. The dropout layer had a 5% probability of dropping parameters. We used ReLu activation function for all layers where applicable, and we optimized the model using the Adam function with a learning rate of 0.0002; loss function was evaluated using a mean-squared error metric.

We split the 3124 model configurations 70-30 into training and testing data (training data set having 2186 models, testing having 938). Training occurred over 100 epochs with a batch size of 16 with shuffling. Training data was further split with 30% allocated for validation.

3 Results

Figure 4 shows the loss and accuracy curves for the training and validation data. The loss curve (figure 4a) converges after approximately 40 epochs and is quite low. The accuracy curve (figure 4b) is nearly zero with some spikes of improvement that are so small as to be inconsequential. The model has an R-squared score of 0.76 and mean-squared error of 0.0013 (see Table 3).

Figure 5 shows the true and predicted temperature distribution for all 3201 points for each of the 938 testing data. Figure 6 shows the true and predicted temperature distribution for a single testing model.

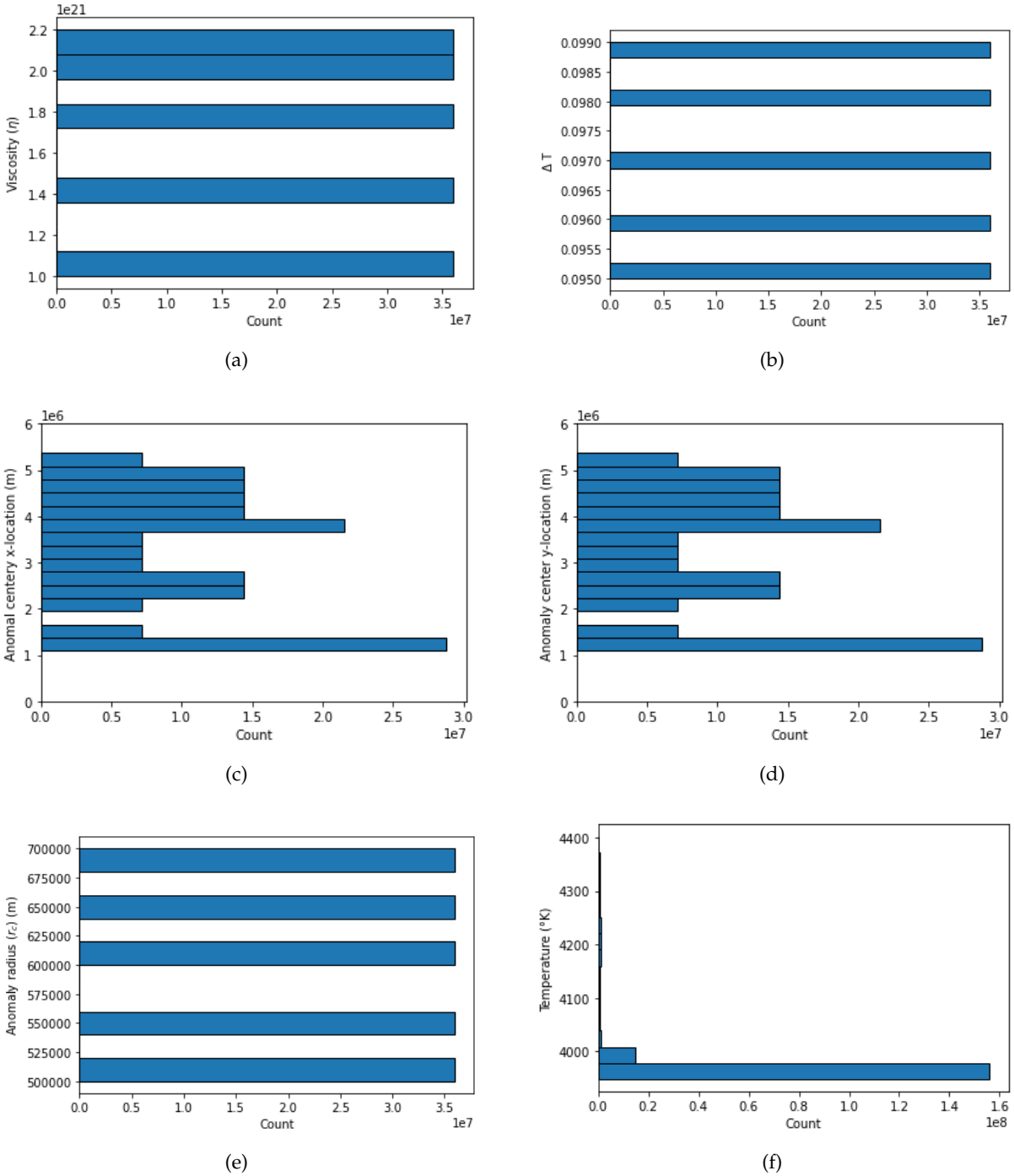


Figure 2: Parameter distribution across entire dataset. **(a)** reference viscosity **(b)** Δ T: determines maximum temperature in center of anomaly **(c)** Center of anomaly (calculated from parameters r_d and θ) x-direction **(d)** Center of anomaly (calculated from parameters r_d and θ) y-direction **(e)** Anomaly radius **(f)** Original (not scaled) temperature distribution for all configuration data points

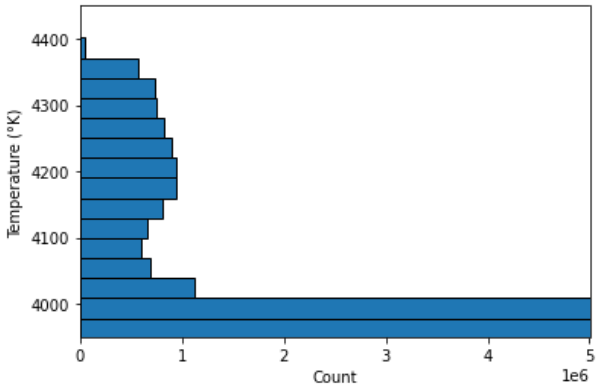


Figure 3: Close up Figure 2f: temperature distribution across data set

Layer Name	(type)	Output shape	Parameters
conv_1	(Conv2D)	(None, 6, 1067, 32)	16416
pool_1	(MaxPooling2D)	(None, 3, 533, 32)	0
conv_2	(Conv2D)	(None, 2, 267, 64)	51264
pool_2	(MaxPooling2D)	(None, 1, 133, 64)	0
time_dist_1	(TimeDistributed)	(None, 1, 8512)	0
lstm_1	(LSTM)	(None, 1, 128)	4424192
lstm_2	(LSTM)	(None, 64)	49408
repeat_vector_1	(RepeatVector)	(None, 18, 64)	0
dropout_1	(Dropout)	(None, 18, 64)	0
lstm_3	(LSTM)	(None, 18, 128)	98816
lstm_4	(LSTM)	(None, 128)	131584
fc_1	(Dense)	(None, 3201)	412929

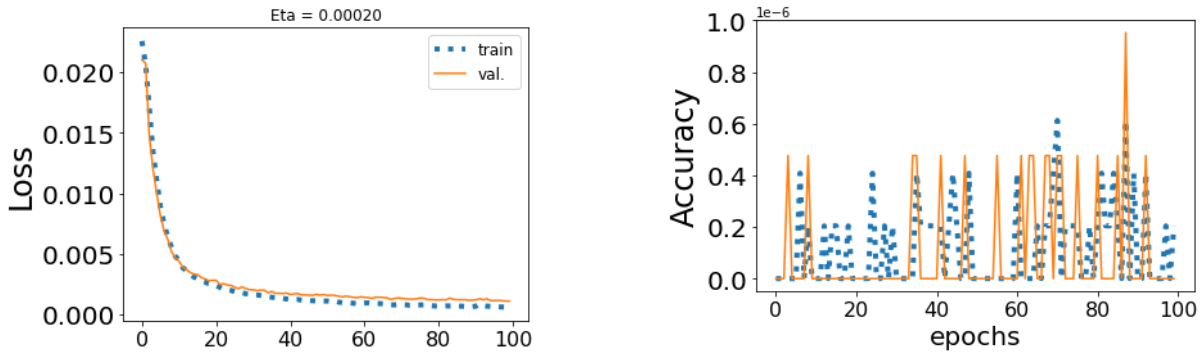
Total params:	5,184,609
Trainable params:	5,184,609
Non-trainable param:	0

Table 2: Summary of neural network architecture

Here, we used testing model configuration 100. The misfit for this singular model is shown in Figure 7. This error displayed on the histogram (figure 7a) has a normal distribution with no obvious skewness.

4 Discussion

Overall, the model performs moderately well with an R2 score of 0.76 but requires further improvements. The combination of a low loss with a low accuracy score implies that the model has a significant amount of small errors. It might be possible to improve this by changing the spatial discretization to be more coarse or reducing the time step from 500,000 years to something smaller. Looking at figures 5 and 6 shows that the NN was able to generally learn the major features. We can see from figure 7b that the model has some trouble predicting the spatial spread and magnitude of temperature from the



(a) (b)

Figure 4: **(a)** Loss curve for testing data. **(b)** Accuracy curve for test data. Dotted blue line represents training data; solid orange line is validation.

R2	0.763
MSE	0.001251
RMSE	0.03291

Table 3: Neural network performance evaluation

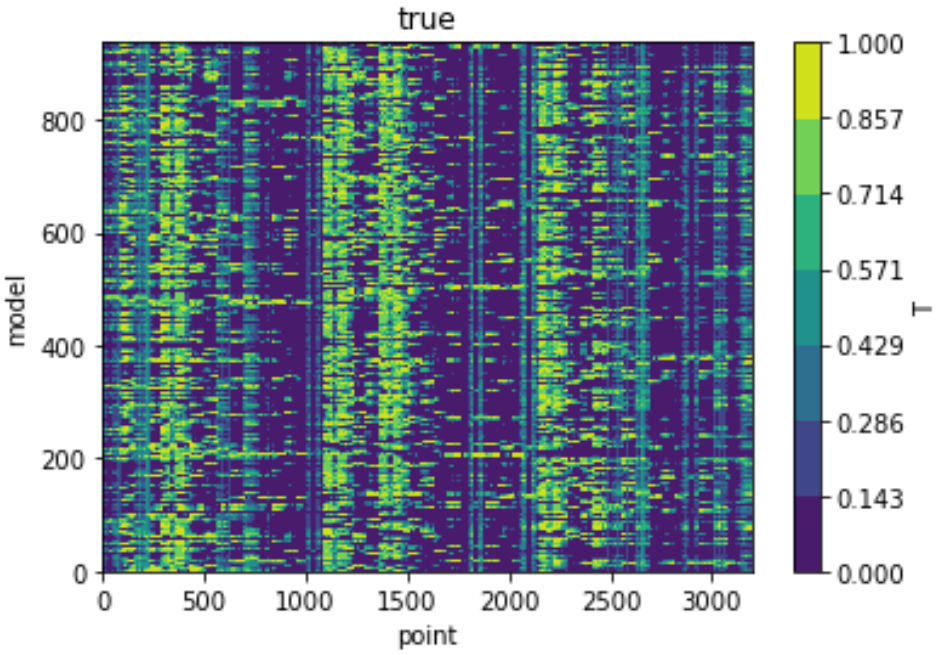
relatively high concentration of absolute misfit near the outer edges of the anomaly with increasingly less misfit towards the anomaly center. As far as forward modeling goes, the result is underwhelming and requires some refinement.

Future plans for this project include creating a more complicated mantle simulation (e.g. multiple layer with thermal boundary layers, creating a sinking anomaly, possibly having multiple anomalies) and predicting temperature distributions at previous time steps.

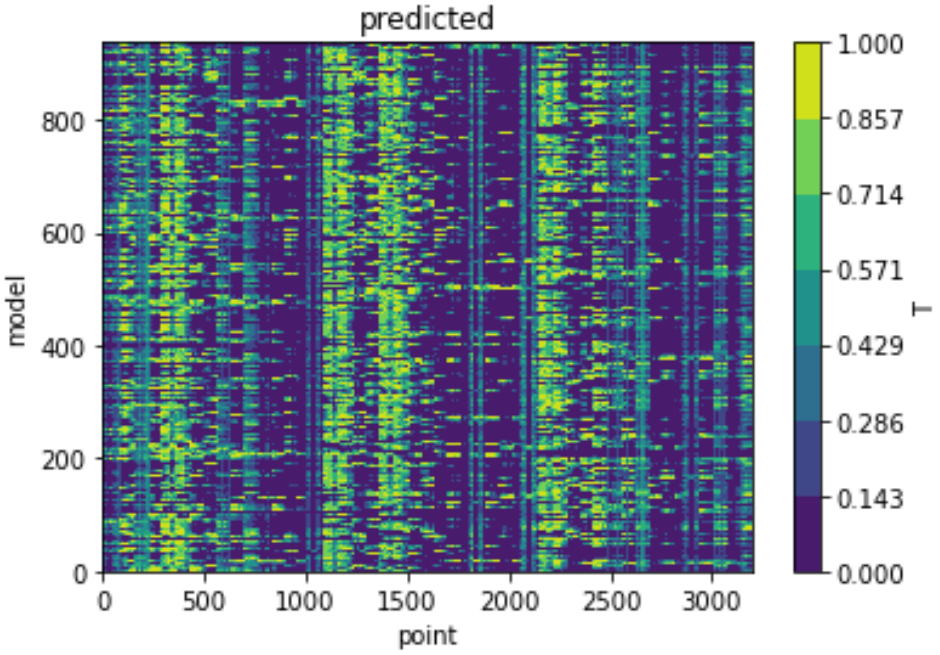
5 Acknowledgments

The codes used to generate data and build the NN are located at https://github.com/rclam/LG_simple although the data files for features and target labels are not included as they are too large to upload to GitHub.

We thank the Computational Infrastructure for Geodynamics (geodynamics.org) which is funded by the National Science Foundation under award EAR-0949446 and EAR-1550901 for supporting the development of ASPECT.

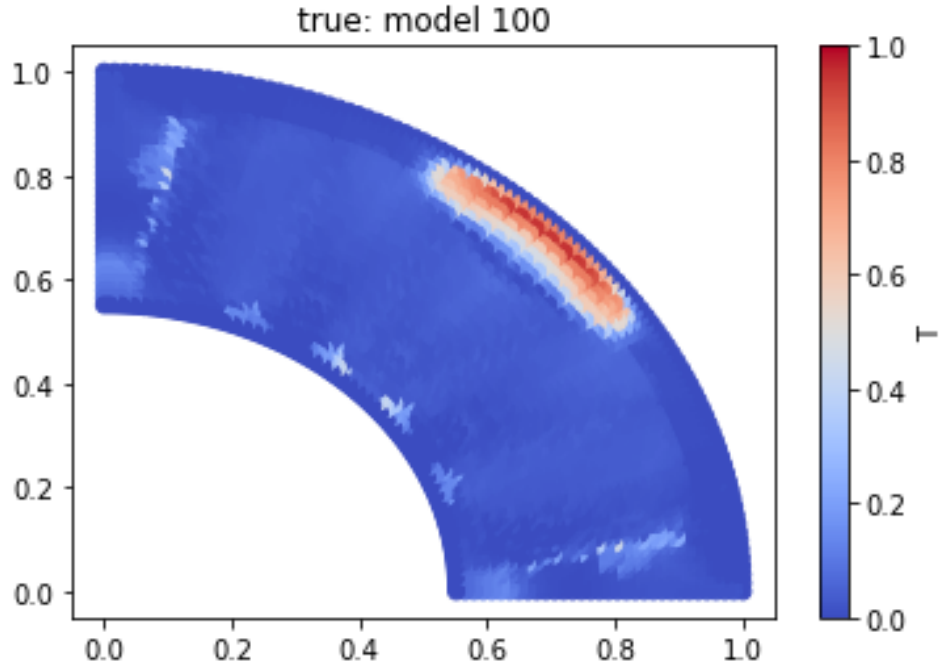


(a)

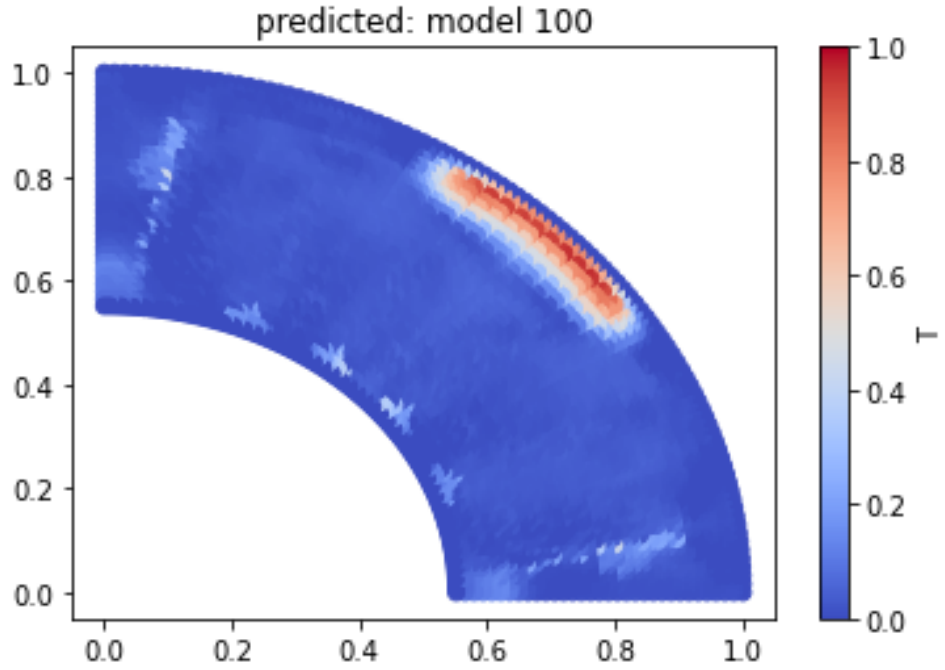


(b)

Figure 5: Comparing all test data scaled temperature distributions at 9 Ma. **(a)** True scaled temperature distributions for cleaned (reduced to 3201 data points) test data **(b)** Predicted scaled temperature distributions.

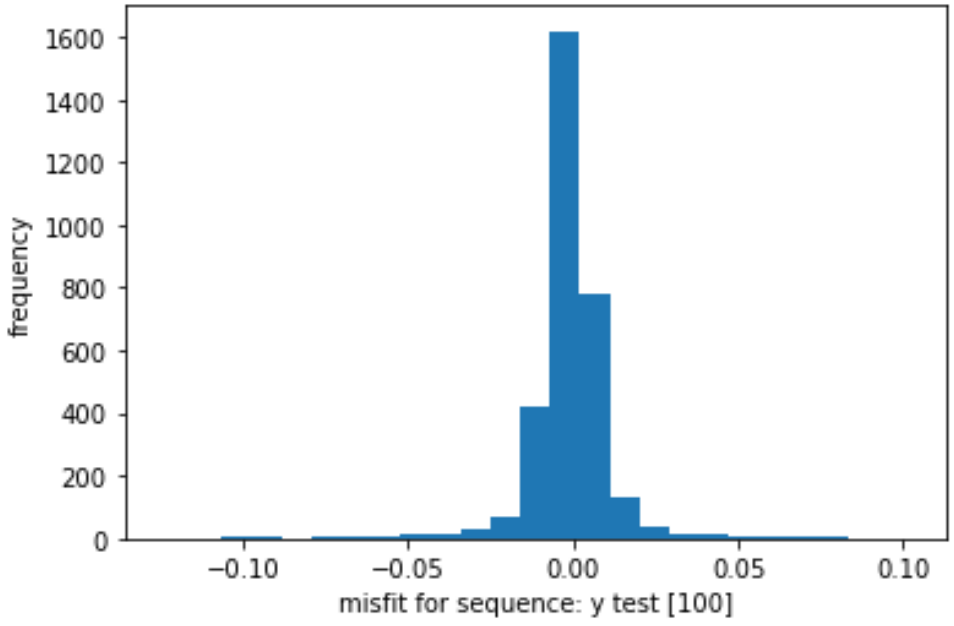


(a)

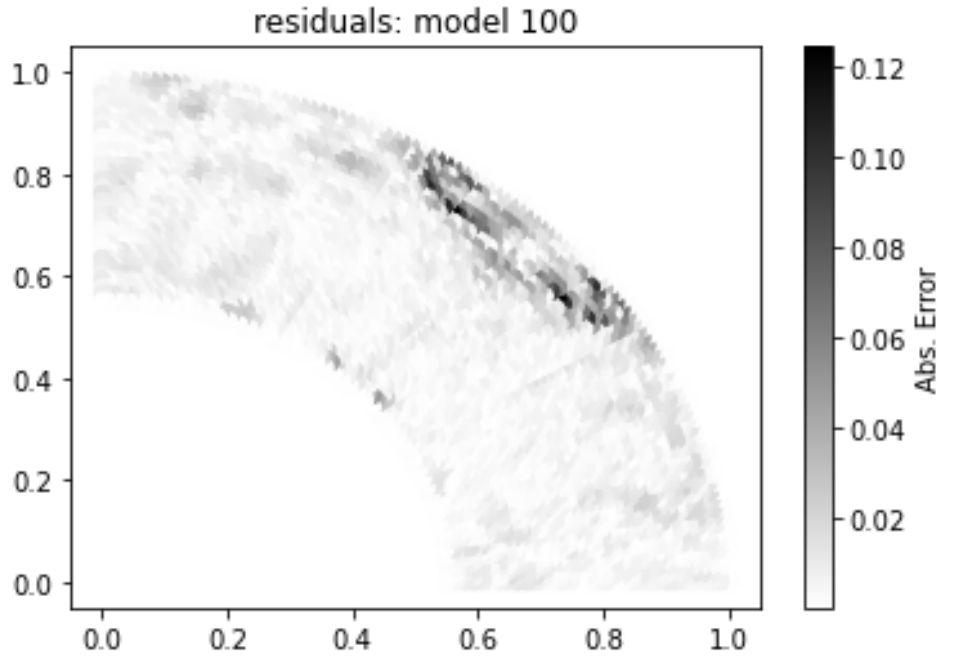


(b)

Figure 6: Comparing testing model 100's scaled temperature distributions at 9 Ma. **(a)** True scaled temperature distribution for test data model **(b)** Predicted scaled temperature distribution for test data model



(a)



(b)

Figure 7: Misfit for test data model 100 **(a)** Histogram of misfit **(b)** Misfit for each data point

References

Agarwal, S., Tosi, N., Breuer, D., Padovan, S., Kessel, P., and Montavon, G. (2020). A machine-learning-based surrogate model of Mars' thermal evolution. *Geophysical Journal International*, 222(3):1656–1670.

Atkins, S. (2017). *Finding the patterns in mantle convection*. PhD thesis, UU Dept. of Earth Sciences.

Bangerth, W., Dannberg, J., Fraters, M., Gassmoeller, R., Glerum, A., Heister, T., and Naliboff, J. (2021a). Aspect v2.3.0.

Bangerth, W., Dannberg, J., Fraters, M., Gassmoeller, R., Glerum, A., Heister, T., and Naliboff, J. (2021b). ASPECT: Advanced Solver for Problems in Earth's ConvecTion, User Manual. doi:10.6084/m9.figshare.4865333.

Gillooly, T., Coltice, N., and Wolf, C. (2019). An anticipation experiment for plate tectonics. *Tectonics*, 38.

Heister, T., Dannberg, J., Gassmöller, R., and Bangerth, W. (2017). High accuracy mantle convection simulation through modern numerical methods. II: Realistic models and problems. *Geophysical Journal International*, 210(2):833–851.

Kronbichler, M., Heister, T., and Bangerth, W. (2012). High accuracy mantle convection simulation through modern numerical methods. *Geophysical Journal International*, 191:12–29.

Liu, L. and Gurnis, M. (2008). Simultaneous inversion of mantle properties and initial conditions using an adjoint of mantle convection. *Journal of Geophysical Research*, 113(B8):B08405.