

Modelling the meridional heat transport at the Top of the Atmosphere, using a Feed Forward Neural Network and Gaussian Process Regression.

Christina Kappatou,
Department of Geosciences, UiO,
christina.kappatou@geo.uio.no

May 2023

Abstract

In a non-analytical effort to assess the meridional heat transport at the Top of the Atmosphere (TOA), I estimated the training error using a sinusoidal fit, with a Feed Forward Neural Network (FFNN). For the NN, 5 different activation functions were combined with 6 different ways to tune the learning rate for the stochastic gradient descent. With the results leading to fairly large training errors, even for the optimum combination of a sigmoid activation function with a constant learning rate, I moved to approaching the heat flux curve differently, this time employing Gaussian Process regression with 3 different percentages of train/test data. As expected, decreasing the amount of train data points leads to high train errors and large confidence intervals, rendering the curve approach less reliable.

1 Introduction

In his paper [Stone, 1978] uses a very simple equilibrium to calculate the meridional heat transport at the Top of the Atmosphere. Using the following equation:

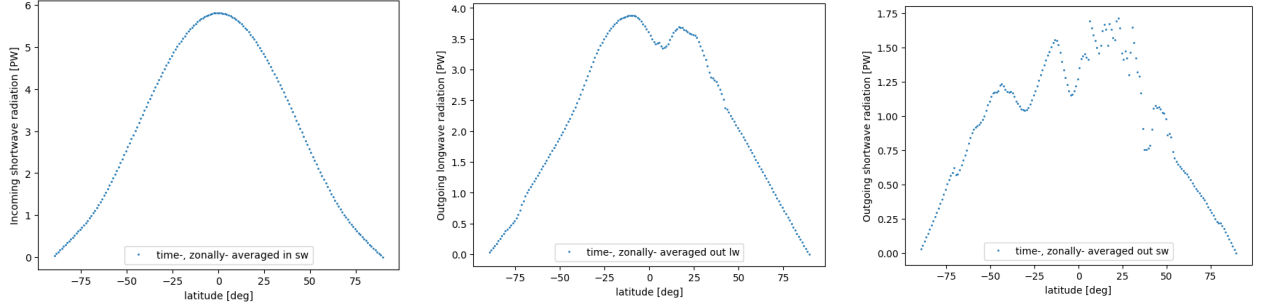
$$\frac{dF}{d\phi} = 2\pi \cdot R^2 \cdot \cos\phi \cdot [Q \cdot (1 - \alpha) - I], \quad (1)$$

he states that, when the planet is in equilibrium, the total energy flux across a latitude belt is given by the difference of the incoming solar radiation (Q) and the outgoing reflected solar radiation ($\alpha \cdot Q$) from the clouds and snow/ ice (expressed by the albedo, α) and the outgoing thermal radiation (I), caused by emissions.

However, because in eqn.(1), all Q , α and I are latitude dependent, it is impossible to find an analytical expression for the heat flux just by integrating. Yet, having all of the three values in a longitude, latitude grid, one can try approaching the total heat flux numerically. In this paper, I tried using both a Feed Forward Neural Network (FFNN) algorithm to estimate the training error between the predicted and the actual data values and a simple Gaussian Process Regression to approach the heat transport curve.

2 Data treatment

To calculate the total heat transport, I used the dataset from [NASA CERES EBAF](#) for the years 2000-2022. Integrating over longitude and over time, gives Figs.1. The total heat transport (Fig.2) comes from subtracting the values of the outgoing shortwave (Fig.1c) and longwave (Fig.1b) radiation from the incoming shortwave radiation (Fig.1a).



(a) Time-, zonally- averaged incoming shortwave radiation from the years 2000-2022. Data taken from [NASA CERES EBAF](#).

(b) Time-, zonally- averaged outgoing longwave radiation from the years 2000-2022. Data taken from [NASA CERES EBAF](#).

(c) Time-, zonally- averaged outgoing shortwave radiation from the years 2000-2022. Data taken from [NASA CERES EBAF](#).

Figure 1: Heat transport components at the TOA (Top of the Atmosphere).

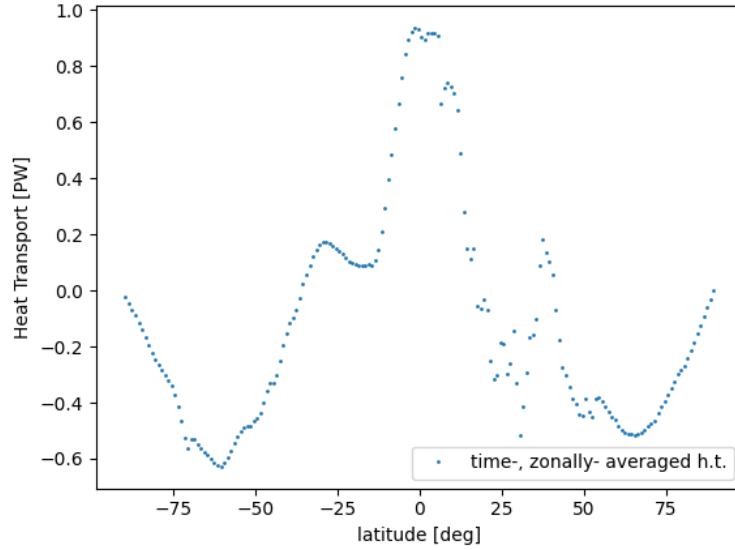


Figure 2: The total heat transport, expressed as the difference of the outgoing shortwave and longwave radiation from the incoming shortwave radiation at the TOA.

3 The Neural Network

Following the basic architecture of a Feed Forward Neural Network (FFNN) for regression, as described by [Mehta et al., 2019], pg. 46-47, I run the code of M.Hjorth-Jensen, with minor adjustments. The entire code, along with the dataset treatment can be found at my [GitHub page](#).

For the Feed Forward stage, I combined 5 of the most common activation functions (identity, sigmoid, softmax, RELU and LRELU) with 6 different ways (constant γ , momentum Stochastic Gradient Descent (SGD), AdaGrad, Adagrad with momentum, RMS-propagation and ADAM, as explained in my previous Project 1) to tune the learning rate in the Gradient Descent stage.

For the SGD, I used 1 minibatch, downgrading it into the plain Gradient descent case with 1000 epochs. The design matrix included $\sin(x)$, instead of plain x , with x being the latitude values, in order to approach the heat flux as a polynomial of up to 3rd degree of latitude sinuses.

The train error depicted in Fig.3 is calculated, using the MSE as described in my previous project 1, as a cost function. Although its values are small, they are not *relatively* small; one shouldn't forget that the total heat transport is calculated in PW and, from Fig.2 one can see that its values don't get higher than 1 PW. Therefore an error of the order of 0.2 PW is important.

From Fig. 3, one can see that the best scheduler-activation function combination is using a sigmoid function

with a constant learning rate, γ for the GD. However, given that the train error value is still quite high (0.166 PW), the Neural Network will have to be better tuned to the dataset or another method to be tried altogether.

In the next section, I am investigating approaching the curve of the total heat flux using Gaussian Process regression.

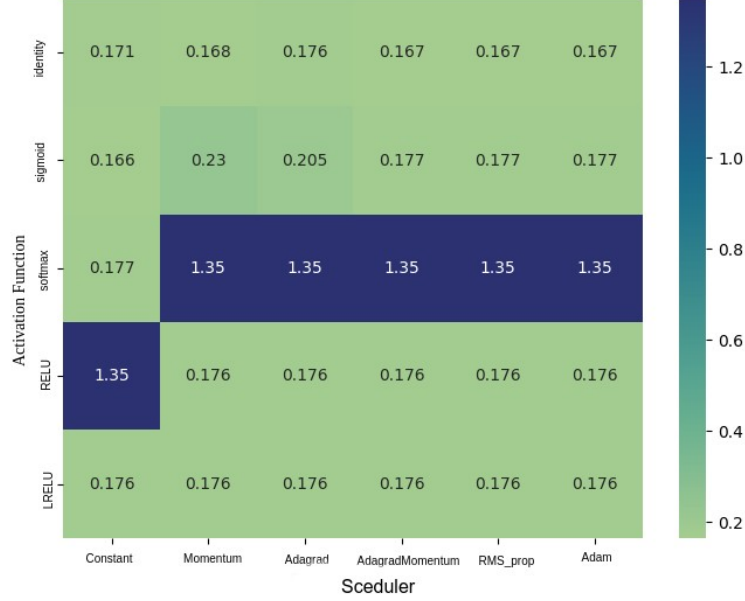


Figure 3: The training error for each combination of activation function and learning rate tuning method (scheduler).

4 The Gaussian Process Regression

Following the code, written by [Dubourg, Vanderplas, Metzen and Lemaitre](#), I used GaussianProcessRegression from Scikit-learn, using a radial basis function (RBF) kernel. Three different percentages of train/test datasets were applied and their effect on the final fitting was investigated.

As can be seen from Figs. 4, having less training data, gradually increases both the deviation of the mean prediction from the actual curve and the 95% confidence interval.

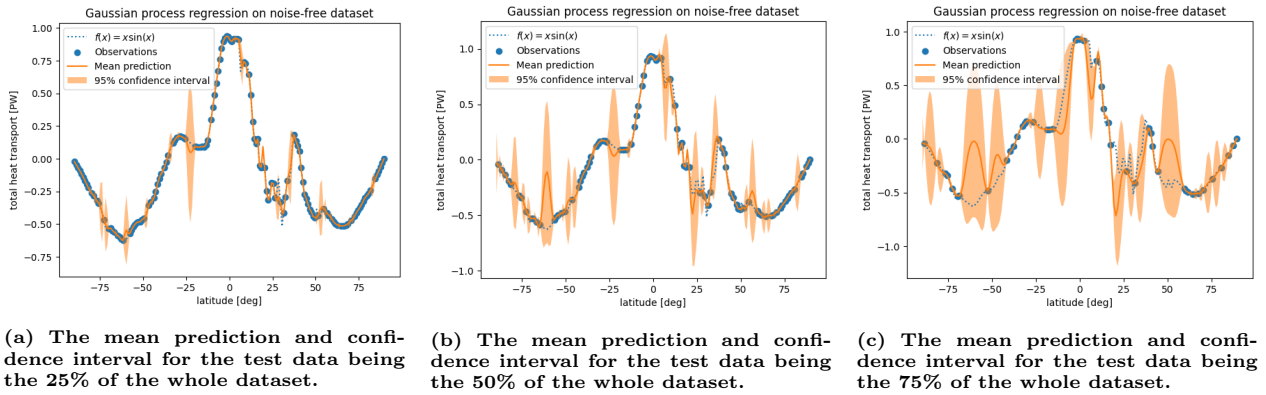


Figure 4: The mean prediction and confidence interval varying with the size of the test/train data.

5 Conclusions and Discussion

A Feed Forward Neural Network regression was applied to a dataset of the total heat transport at the Top of the Atmosphere, in order to settle on a sinusoidal expression. Although I coupled 5 different activation functions with 6 different ways of tuning the learning parameter, γ in the gradient descent part, the train error was significant, even for the best combination, that is using a sigmoid activation function, along with a constant learning rate. For this reason, I moved to simply approaching the curve of the heat flux itself, without aiming for an analytical expression. For this, I used Gaussian Process regression, varying the train data size. The size of the train data was chosen to be equal to 75% of the entire dataset, since this was the value with the least visible train error and the narrowest confidence bands.

6 Future work

Although the tight time frame of the present project did not allow for any original scripting, I am still very interested in writing my own code for a FFNN regression. The dataset itself has also to be further evaluated by my supervisor, Joe Lacasce, and me. The plots do not look like Stone's sinusoidal results and it is still unclear if I treated the data the same as he did with the ones from [Oort and Haar, 1976]. In any case, I would like to work on a longer time-series, since 22 years (from 2000 to 2022) are probably not enough to assess a general trend of the heat transports.

Overall, the need for better understanding of the NN principles is more that clear to me now. I can see it's possible to integrate them into my research more constructively and I will try investigate that route during the next months.

References

- [Mehta et al., 2019] Mehta, P., Bukov, M., Wang, C.-H., Day, A. G., Richardson, C., Fisher, C. K., and Schwab, D. J. (2019). A high-bias, low-variance introduction to machine learning for physicists. *Physics Reports*, 810:1–124.
- [Oort and Haar, 1976] Oort, A. H. and Haar, T. H. V. (1976). On the observed annual cycle in the ocean-atmosphere heat balance over the northern hemisphere. *Journal of Physical Oceanography*, 6:781–800.
- [Stone, 1978] Stone, P. H. (1978). Constraints on dynamical transports of energy on a spherical planet. *Dynamics of Atmospheres and Oceans*, 2(2):123–139.