Machine Learning for Earth and Space Observation

Fenwick,R.O., University of Leicester, LISEO

Introduction

In this poster I present the first steps towards replicating the retrieval scheme for methane from satellite data using neural networks. Replicating retrievals using neural networks has been previously, for example O_3 [4] , however these cases tend to use simulated data, whereas I will be using actual satellite data and methane from retrievals. I have chosen to look as Methane as it is the 'second most important anthropogenic greenhouse gas after carbon dioxide' [3] and so is an interesting and important focus point. It should also be possible to further validate my results using ground truth data, however this cannot be used for training a network as there is not enough coverage.

I intend to extend my work by implementing error correction methods after training my network. This will be particularly useful as the methane data currently available contains known biases and error correction, when this bias is identified, will be less computationally expensive than retaining with updated data. Error correction has been implemented in [5],[2] in order to avoid the expense of retraining legacy AI systems which are making mistakes.

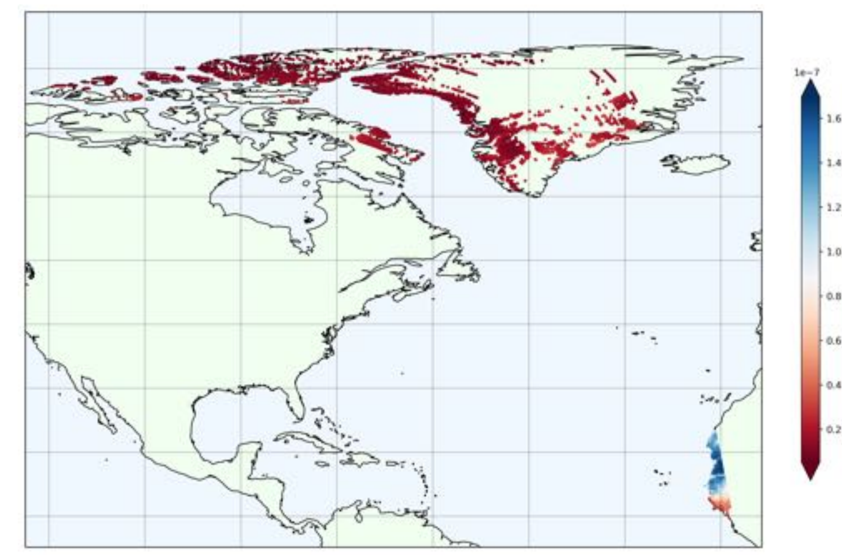

Figure 1: Radiance Data coverage used in Training and Validation

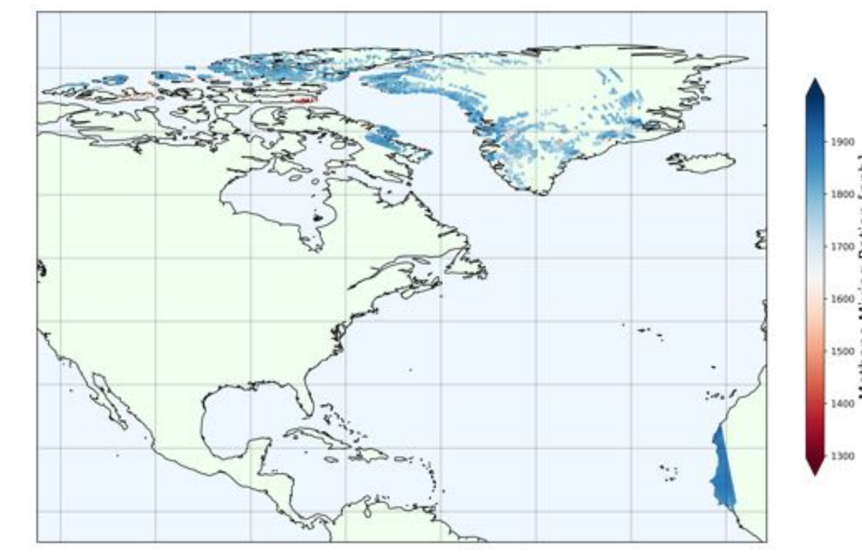

Figure 2: Methane Data coverage used in Training and Validation

TROPOMI and S5P

The TROPOspheric Monitoring Instrument (TROPOMI) is the satellite instrument on board the Copernicus Sentinel-5 Precursor (s5p) satellite. It is designed to have a lifetime of 7 years and orbit approximately 14 times a day. Each full orbit produces 139 Gb of data.

There are three levels of data product; the first L0 is not available and contains raw satellite data, the second and third level of data are those of interest to me. The second level of data product, L1B, contains 'geo-located and radiometrically corrected top of the atmosphere Earth radiances in all spectral bands, as well as solar irradiances' [1] and data from this Level will form the input for my neural network. The third data level, L2, contains the retrieved methane which will be used as the output of my neural network for training.

For this poster I have presented work using just one orbit worth of data. The methane (L2) data is not complete, as in some places the retrieval was not deemed good enough to publish. I have therefore excluded the missing data from both the L1B and L2 data in training and validation. The L1B data file for this orbit contains a number of different variables, in my training set I have included the 480 wavelengths and radiances as well as latitude and longitude. There is other data in the file which I may include in future network training to improve my results.

Figure 1 shows the remaining data points for one of the radiance bands, figure 2 shows the corresponding methane data. It is clear that the coverage for this one orbit is not global and, although sufficient for this proof of concept work, is not a sufficient sample set for my final network.

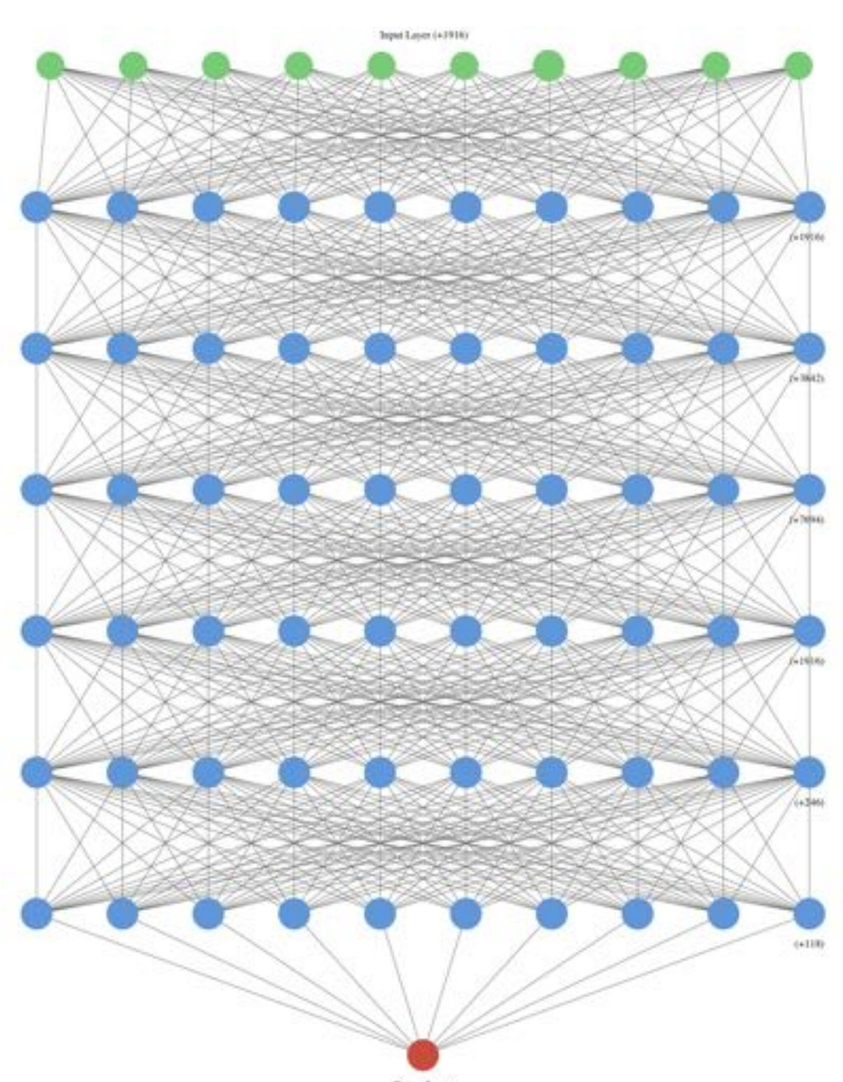

Figure 3: Network Architecture

Neural Network

I have designed a neural network, trained on the data described above and shown in figures 1 and 2. The data was split into three, 25% was set aside for validation, the remaining 75% was split 70 : 30 into training and test data. The network consists of 6 hidden layers as well as an input and output layer. All of the layers use the RELU activation function other than the output layer, which uses Linear activation. The Network was trained using the Adam optimiser over 500 epochs.

The network architecture can be seen in figure 3, it shows the input layer (green), hidden layers (blue) and the output layer (red) as well as the number of nodes. The network was built and trained using Keras in python.

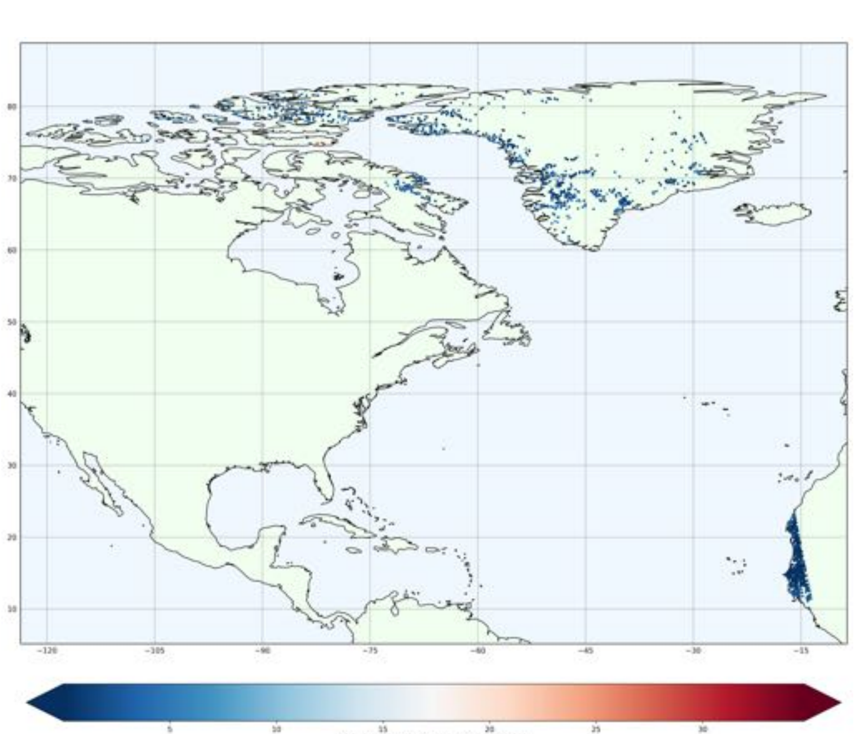

Figure 4: Percentage Differences between retrieval and network

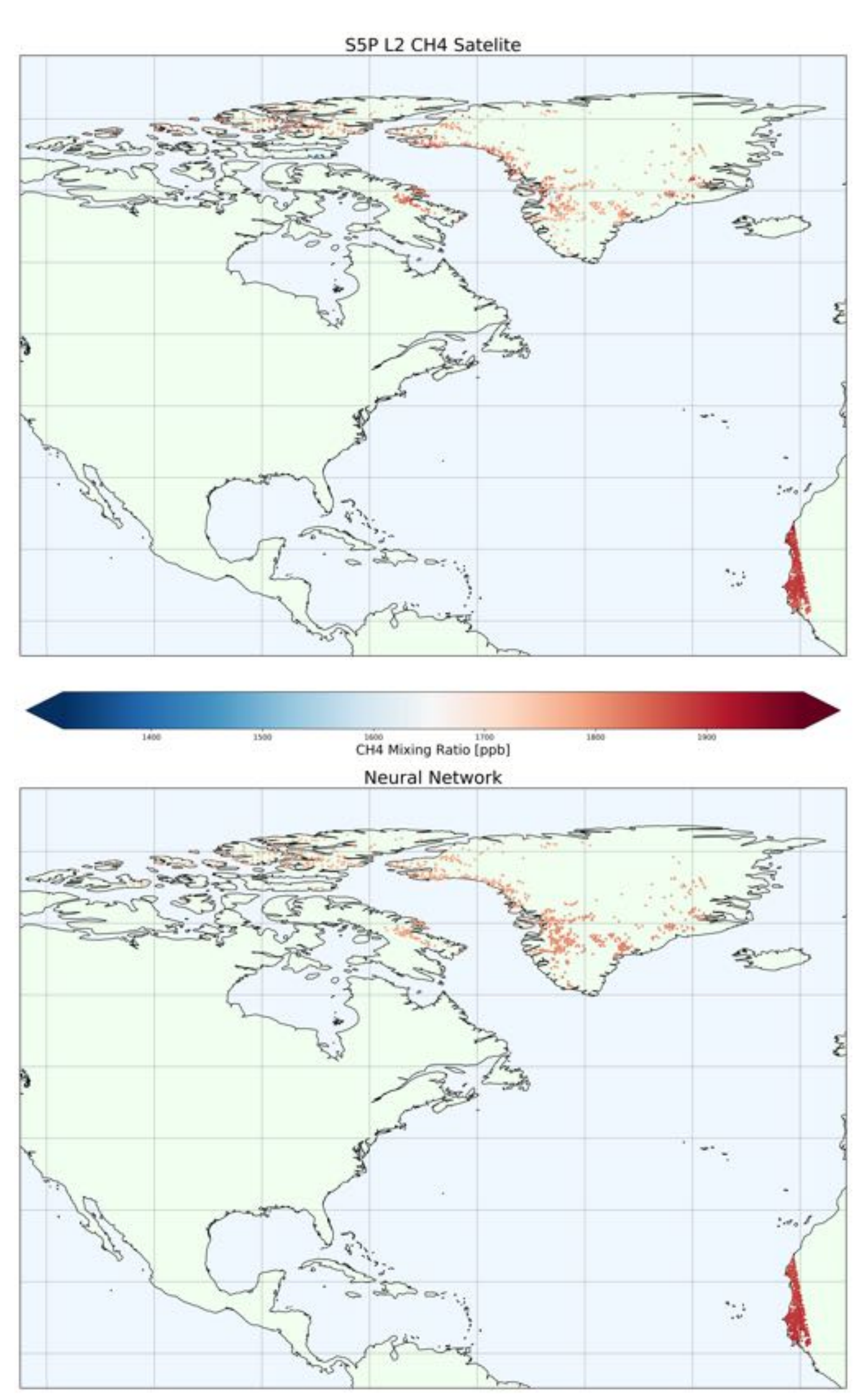

Figure 5: Retrieved Methane (Top) and Network Methane (Bottom)

Results

The results from this network are fairly good, with a mean percentage error of 2.18%. Figure 5 shows the retrieved methane (top) and the corresponding network values (bottom), and the network results look similar to the retrieved Methane. Looking at figure 4, it is clearer where the differences are, the larger errors occur mostly when the actual methane values were lower. In figure 8 I have zoomed into the area with the most errors for a clearer view, and comparing this with the same region of figure 2 shows, more clearly, that it is the lower methane values which are showing the highest percentage differences. The distribution of errors is interesting, figure 7 shows the majority of errors are under 5%. However, there are a few much larger errors, some higher than 30%. Having looking in further detail at these results, I suggest that the reason for these higher errors is the limited number of lower valued methane data points in this set. I expect that when I train on a larger, better constructed, sample of the data then I can further improve this result.

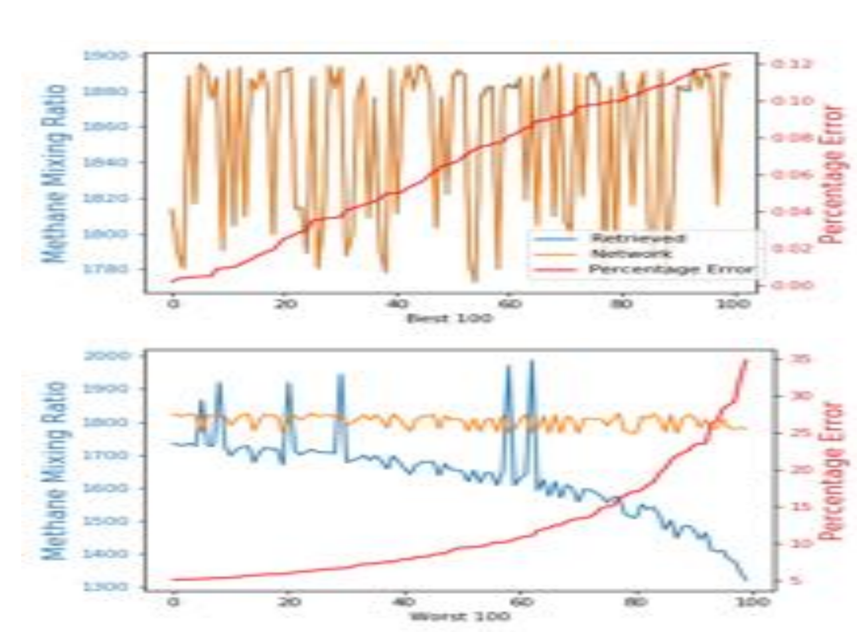

Figure 6: Best (Top) and Worst (Bottom) network results

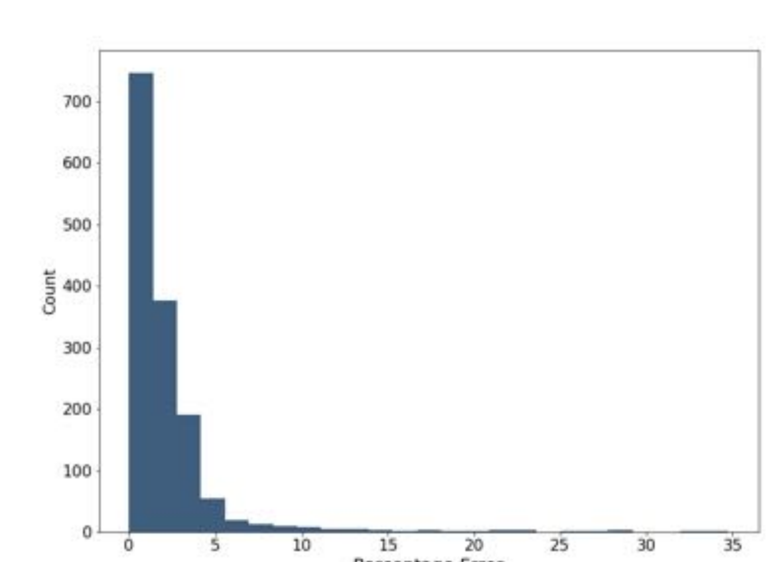

Figure 7: Distribution of Errors

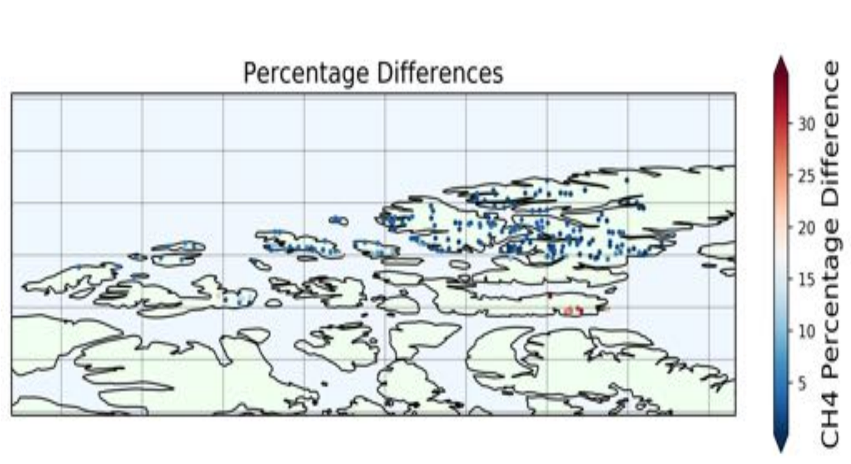

Figure 8: Subsection of Percentage Differences between retrieval and network

Further Work

This poster presents a sufficient proof of concept for the neural network I will train on a much larger sample of the data. The next step will be to build a sample of the entire available data which is representative both in terms of the time span it covers and the global coverage of the data. This sample data will form the training and test sets for my neural network. Once this sample has been built I will develop the structure of the network. I expect to have to make changes to the basic network I have presented in this poster, however it will be a good basis to start from. Once the network achieves acceptable results I will move on to looking at the error correction methods and attempt to further improve the results.

References

[1] Tropomi ,http://www.tropomi.eu/.

[2] A.N. Gorban, Aleksandr Golubkov, Bogdan Grechuk, Evgeny Mirkes, and Ivan Tyukin. Correction of ai systems by linear discriminants: Probabilistic foundations.

[3] Haili Hu, Jochen Landgraf, Rob Detmers, Tobias Borsdorff, Joost Aan de Brugh, Ilse Aben, AndrÃl Butz, and Otto Hasekamp. Toward global mapping of methane with tropomi: First results and intersatellite comparison to gosat.