Neural Networks for the Retrieval of Methane from the Sentinel-5 Precursor Satellite

1st Rose Fenwick
School of Mathematics and Actuarial Science
University of Leicester
Leicester, United Kingdom
rof3@leicester.ac.uk

2nd Hartmut Boecsh
School of Physics and Astronomy and National Centre for Earth Observation
University of Leicester
Leicester, United Kingdom
hb100@leicester.ac.uk

3rd Ivan Tyukin
School of Mathematics and Actuarial Science
University of Leicester
and *Norwegian University of Science and Technology*
and *Saint-Petersburg State Electrotechnical University*

Leicester, LE1 7RH, UK
and Trondheim, Norway
and Saint-Petersburg, Russia
I.Tyukin@le.ac.uk

***Abstract*—Methane is the second most common anthropocentric greenhouse gas, it is therefore important to accurately and quickly get concentrations globally from satellite readings. Currently, data acquired by TROPOMI on board the Sentinel-5 Precursor satellite is transformed into Methane mixing ratio via a physics-based retrieval algorithm. This paper presents an alternative to the slow and complex algorithm: a neural network. A number of experiments are performed using a range of training sets and different network architectures. These experiments, not only help to chose the final network architecture but also allow discussion about the regional and seasonal correlation of methane mixing ratio in the data. These experiments conclude that there is some seasonal and regional correlation in the data which must be taken into account during training. Finally, a single hidden layer network,with 256 nodes in the hidden layer, is trained over 2500 epochs giving a respectable, but improvable, 1.36% error. This error is a slight improvement on the original experiment and the results of this network are analysed with areas of improvement suggested. Particularly, improvement of the extreme data points, which are highlighted as being the worst predicted by the network. This work presents the ground work for using neural networks to replace lengthy retrieval systems, not just for methane but other gases commonly investigated in a similar way.**

***Index Terms*—Neural network applications, Remote sensing, Satellites**

I. Preliminaries and Problem Statement

A. Sentinel-5 Precursor and TROPOMI

The Sentinel-5 Precursor (S5P) satellite was launched in October 2017 and aboard was the Tropospheric Monitoring Instrument (TROPOMI). TROPOMI 'measures reflected sunlight in the ultraviolet, visible, near-infrared, and shortwave infrared spectral range.' [1] Orbiting around $14\frac{1}{2}$ times a day, an abundance of data is collected and almost daily global coverage is achieved.

There are three different levels of data that are of relevance, the first Level 0 is what is measured by the instruments, Level 1b is a calibrated version of the Level 0 data which is available and the Level 2 geophysical data is inferred from Level 1b data. This research works with the Level 1b and Level 2 data available at the time of writing.

The TROPOMI instrument was successfully calibrated and the instrument tested over a six month period concluding in April 2018, at which point the data was made available for use. This means a full year of data, from May 2018 to April 2019, is available for use. However, there is data missing for June, August and November 2018 and so it is not possible to include data from those months.

B. CH_4 Retrieval

Methane mixing ratios (XCH_4), which are an average of the methane concentrations across the altitudes in the atmosphere where methane has been recorded, is the measure of methane delivered as a product from the satellite data collected. It is, among other variables, provided in the Level 2 data. In order to transform the Level 1 satellite data into methane mixing ratio a retrieval algorithm, known as RemoTeC-S5P, is used. This algorithm is fully detailed in the Methane ATBD file [2] provided as part of the S5P and TROPOMI documentation.

As with all retrieval algorithms, the aim is to infer an atmospheric state vector $x \in \mathbb{R}^d$ from a measurement vector $y \in \mathbb{R}^n$ using a forward model $y = f(x, b)$, where $b \in \mathbb{R}^p$ is the vector of model parameters which are not retrieved. The state vector includes the Methane mixing ratios along with other variables calculated during the retrieval process.The algorithm focuses on specific wavelengths where methane is found, 2305-2385nm, along with a number of inputs which must be calculated or inferred from the data first.

In this work, these retrieved values will be used as truth when training the networks. There may be the possibility to use actual recorded methane values in future work for validation,

The project was supported by the Leicester Institute for Space and Earth Observation (LISEO)

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

but the coverage of the latter type of data is not enough to use in training.

C. Problem Statement

Methane is the 'second most important anthropogenic greenhouse gas after carbon dioxide' [1], and so it is important to accurately and quickly obtain concentrations globally once the S5P satellite has performed its measurements. The amount of data produced by the satellite is vast and the time it takes for the retrieval of methane can be large. The data needs to be processed and prior information calculated or inferred before the retrieval can take place. There is also some physical assumed knowledge used within the retrieval process based on known or assumed relationships between the Level 1 data and Methane.

A possible way to overcome these constraints is to infer methane concentrations from Level 1 data directly with the framework of Machine Learning (ML). As opposed to conventional statistical modelling, no prior assumptions on statistical properties of the data are imposed in ML. This makes the approach particularly attractive for this particular application. As ML models, neural networks are excellent candidates for methane retrieval as the abundance of data is a benefit and, once the network is trained it will be much faster than the current method. Using a neural network has the added bonus that it will infer connections from the data rather than using assumed knowledge.

This paper addresses the question; is it possible to replicate the methane retrieval process using a neural network with results that are reasonable and acceptable to the Earth observation community.

II. Data Processing

A. Pre-Processing

The data, described in Section I, is accessed using the computer cluster JASMIN , the initial processing is performed here and involves matching the L1 and L2 files and creating one merged file for each set of these. Once the initial processing is completed it is transferred to another machine (local to Leicester), all the files are then merged to create a master file and further processing occurs.

1) Selection: It is important that the data used in this works has both global and seasonal coverage, to ensure any trends can be learned by the network(s). As the satellite records near daily global coverage, with the right selection criteria, it is easy to ensure global coverage. As seasonality trends would be expected to be seen across months rather than days, it should be enough to select one days worth of data per month. Unfortunately, there are some data gaps so to ensure full annual coverage two days worth of data per month will be selected. As the data gaps often mean either the beginning or the end of the month is missing a day from near the start, the 5^{th}, and a day from towards the end, the 28^{th}, were selected for these experiments. Checks were made to ensure that at least one of the two days were present for every month that data has been collected for.

2) Filtering: The data is provided with a quality filter flag, a continuous quality descriptor varying between 0 (no data) and 1 (full quality data). The documentation [2] recommends that 'values' < 0.5 are ignored to ensure the best quality data is used. This recommendation was followed and all data with a quality filter flag of < 0.5 was ignored in the work. It was also necessary to filter further as methane is only detected in wavelengths 2305-2385nm, therefore any data outside of these wavelengths were removed. There are also a number of variables in the data which are either no useful or provided with the Level 2 data and these were also removed for training. Date, time, latitude and longitude are also removed as they may infer geography by region rather than globally. The variables left after all these have been removed are:

- A number of wavelengths (approx. 1000)
- Radiance's corresponding to each wavelength
- Solar azimuth angle
- Solar zenith angle
- Viewing azimuth angle
- Viewing zenith angle

The filtered version of this data is shown in figure 1.

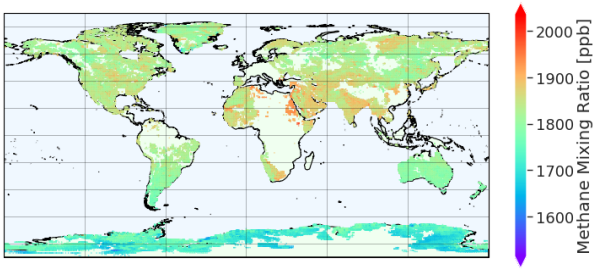

Fig. 1. Data after Quality filter

B. Data Splits

Once the data is processed, selected and filtered as described, three independent sets were randomly sampled; Training (60%), Test (25%) and Validation (15%). These sets are used in some of the experiments described in Section III. Figure 2 (pannels a-c) show the three sets, they have similar global coverage and do not show any obvious extreme differences in methane concentration.

Figure IV-Ad shows the distribution of CH4 in each of these three sets, again these show the sets are distributed similarly for methane. Finally the similarities in these sets can be seen in table I which shows some statistics on each of the sets, confirming the geographic coverage of all these sets are the same and the methane ranges are similar.

Two other sets were created to be used in the experiments described in section III. These were not randomly sampled and will be used to test how well the networks generalise

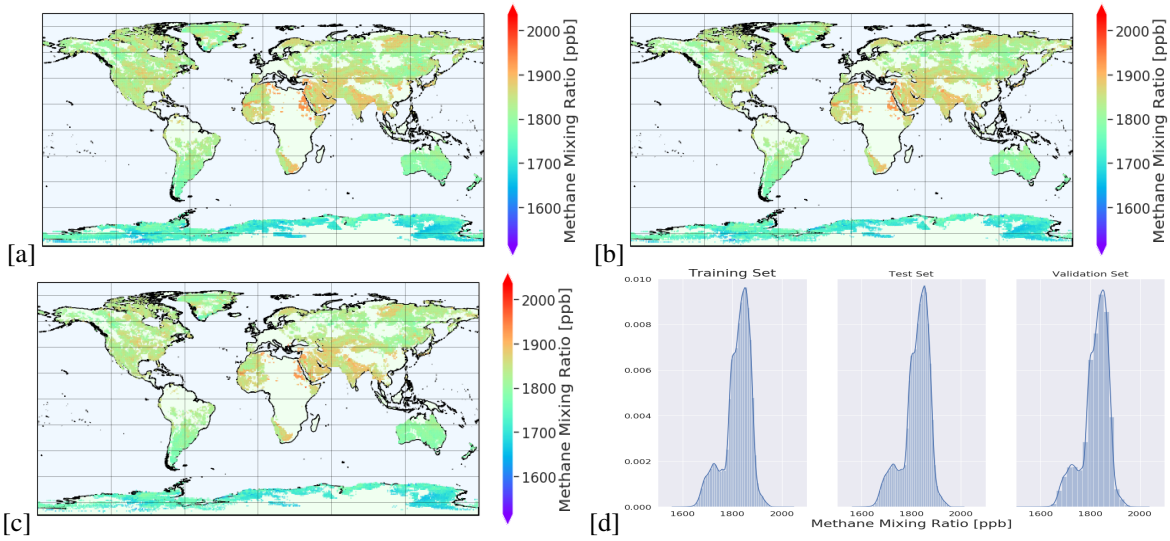

Fig. 2. Plots for the randomly selected sets; a. Training Set, b. Test Set, c. Validation Set, d. Set distributions

TABLE I
SET STATISTICS

Metric	Training	Test	Validation
Records	429,278 (60%)	179,489 (25%)	106,810(15%)
Latitude Range	[-89,83]	[-89,83]	[-89,83]
Longitude Range	[-179,179]	[-179,179]	[-179,179]
Methane Range	[1568,2036]	[1575,2000]	[1514,2023]
Mean Methane	1820	1820	1820

over region and season.

The first of these sets was created using location, data points in Central Europe (Latitudes between [37,67] and Longitudes between [-27,51]) are used for training and testing with the remaining data being used in the validation set. This set will show how well the networks trained generalise over region.Figure 3 shows how the data is split across the three sets.

The second of these sets is split based on the data, 2018 (June - December) is used in training and testing with 2019 being used for validation. This set will show how well the networks trained generalise over season. Figure 4 show the data for each of the sets in this split.

III. NETWORK SELECTION

A. Methodology

A number of experiments were performed in order to find the best network architecture and to test how well the networks generalise. To find the best network architecture a number of different versions were tested, increasing the number of hidden layers and varying the number of nodes in these layers, table II shows the different architectures tested.

To help test how well the networks generalise three different training, test and validation sets were created using different criteria, these are the sets described in section II. The first sets were created randomly (Random), , the second set (Date based) includes 2018 data for training and testing, with 2019 used in validation, and the third set (Regional) includes Central Europe only in training and test. A fourth selection (Random*), is a variation of the first sets (Random) but with the latitude and longitude information included, this has been removed from the other sets.

B. Network Architectures

A number of different architectures were used initially, these are described in table II. In order that comparisons between the experiments could be made an number of training parameters were kept the same. The following did not change for any of the experiments in table II:

- Input Layer Activation : RELU
- Hidden Layers Activation : RELU
- Output Layer Activation : Linear
- Optimiser : Adam
- Learning Rate : 0.01
- Decay : 0
- Loss Function (to optimise) : Mean Squared Error
- Other Metrics Recorded: Mean Absolute Error, Mean Absolute Percentage Error
- Batch Size : 500
- Epochs: 100

The experiments can be paired up by the set split used (1&2, 3&4, 5&6, 7&8). The only difference within the pairs is the number of hidden layers in the network that is trained. This helps to conclude how many hidden layers should be used in the final network.

IV. RESULTS

A. Experiments

The top line results from the experiments are shown in table III, for each network both the root mean square error (RMSE) and the percentage error is given for each of the experiments

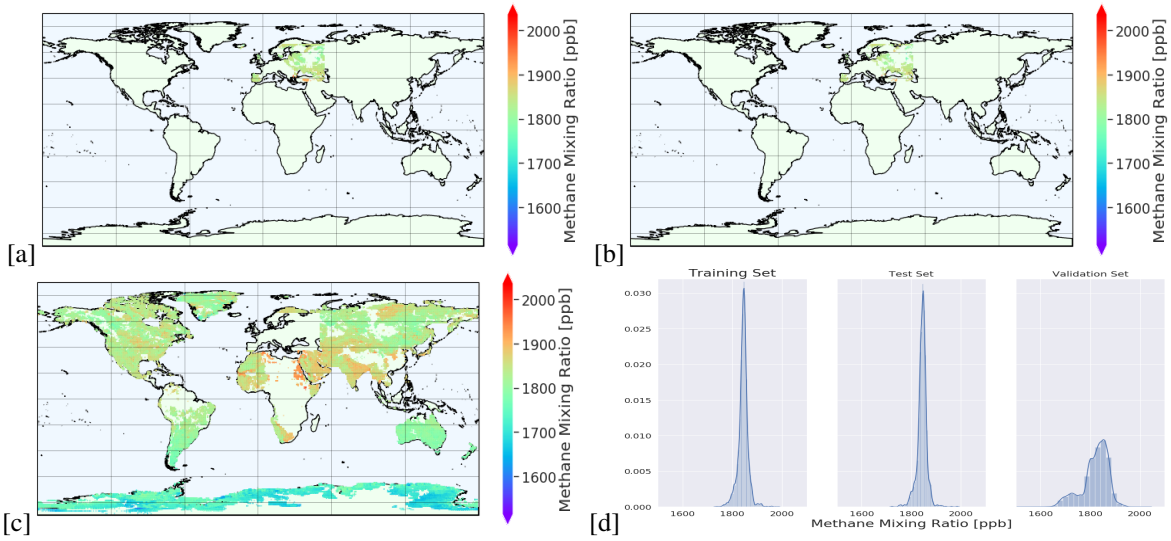

Fig. 3. Plots for the geographically selected sets; a. Training Set (Just Europe), b. Test Set (Just Europe) c. Validation Set (No Europe), d. Set distributions

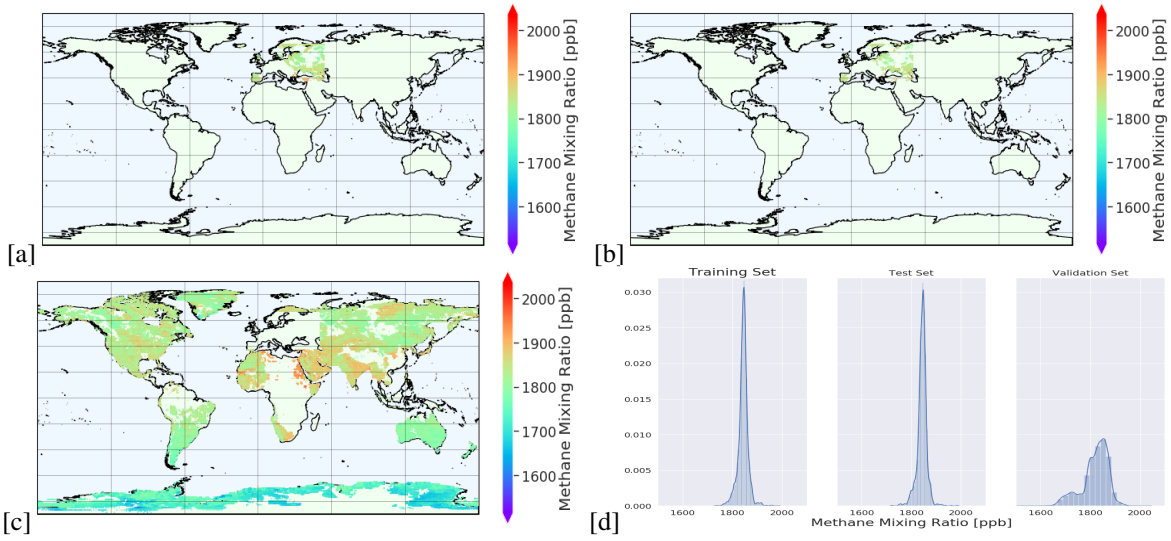

Fig. 4. Plots for the date selected sets; a. Training Set (2018), b. Test Set (2018) c. Validation Set (2019), d. Set distributions

TABLE II
EXPERIMENTS

Experiment	Sets Used	Hidden Layers	Nodes
1	Random	1	[256]
2	Random	3	[512,256,128]
3	Date Based	1	[256]
4	Date Based	3	[512,256,128]
5	Regional	1	[256]
6	Regional	3	[512,256,128]
7	Random*	1	[256]
8	Random*	3	[512,256,128]

in table II.

Only the pair trained on the regional sets (5&6) showed better results for the three hidden layer network and even these results are not considerably better than the single layer network. Furthermore, the single layer networks are faster to train, so without giving improved results there is no reason to further train a larger network.

Experiments (3&4) show the highest validation errors which suggests that there is some overfitting on the training set. This means that, as this was the date based training set, there is some correlation between the time of year and the methane concentration which the network cannot learn without data from across the entire year. This is not unexpected as there is some known seasonality in methane concentration.

Although experiments (5&6) performed much better in training and testing than experiments (3&4), the validation results do not greatly improve. These were the experiments using the regional set split and so this result shows that there are some regional differences in the data. The fourth pair of experiments (7&8) use longitude and latitude which gives some regional context.

The best results are seen in experiment 7, this indicates that training over the whole data is import and that latitude and longitude add some important information. This however, may not be the best way to include the geography of the regions.
Looking at the network training curves, figure 5, it is clear that all of the networks tested in the experiments have not been trained for enough epochs. Each of the networks in the experiments require further training for the best results, however, as already discussed the three hidden layer networks are not significantly better so those will not be trained further. The date based (3&4) and regional based (5&6) training sets will also not be trained further, the experiments were designed to look at how well the network can generalise when given limited data and show overfitting. Ruling out these experiments leaves experiment 1 and experiment 7, of these experiment 7 achieves much lower errors and so this is the network that will be trained further.

TABLE III
EXPERIMENT RESULTS, ROOT MEAN SQUARED ERROR (PERCENTAGE ERROR) FOR EACH SET

Experiment	Training	Test	Validation
1	40.88 (2.15%)	41.12 (2.16%)	41.17 (2.15%)
2	42.05 (2.23%)	42.31 (2.24%)	42.34 (2.23%)
3	38.17 (2.01%)	38.42 (2.01%)	46.53 (2.52%)
4	36.51 (1.92%)	36.76 (1.92%)	46.73 (2.53%)
5	16.07 (0.74%)	16.27 (0.74%)	43.76 (2.36%)
6	16.03 (0.74%)	16.53 (0.45%	43.79 (2.36%)
7	27.20 (1.36%)	27.29 (1.36%)	27.49 (1.36%)
8	43.01 (2.28%)	43.26 (2.29%)	43.28 (2.28%)

B. Final Network

The final network trained was a single layer network with 256 nodes using the random training, test and validation sets, and including the Latitude and Longitude information, as in experiment 7. This network was trained over 2500 epochs, but no other changes were made to the architecture. As table IV shows the results do not improve much following the further training, the root mean squared error is slightly improved but the absolute percentage error does not. Figure 6 shows the training curve for the 2500 epochs and although there is a lack of much change it is clear that the training could be extended further as the network is not yet overfitting the training set.

TABLE IV
FINAL RESULTS, ROOT MEAN SQUARED ERROR (PERCENTAGE ERROR) FOR EACH SET

Experiment	Training	Test	Validation
Final	27.18 (1.36%)	27.26 (1.36%)	27.46 (1.36%)

C. Validation

Comparing the plots for the actual and predicted methane mixing ratio in figure 7, the network results are similar to the actual values however there are fewer high and low points on the predicted plot. This suggests that the network is not particularly good at predicting the extreme values, something that could pose a problem as it is important to accurately predict these extremes. One of the reasons for this error could be a lack of extreme points in the whole data sets, and therefore the training set.
The density correlation plot in figure 9 shows the predicted against the actual methane mixing ratios for the validation set. It is clear that there is an excellent correlation particularly in the middle region of the data, between 1700 and 1900ppbs, where there is the most data. Outside of this range both the volume of data and the correlation is reduced, confirming that the network is less accurate at the more extreme values as we saw on the comparison plots in figure 7.
The distribution of errors shown in Figure 10 show that the majority of errors are below 5% and above -5% and that a there are a very small number of errors outside of this region. Table V shows some error statistics, the errors range from -21% to 9%, but with a mean of 1%. The absolute percentage error shows a mean error of 1%. Although errors of 1% are excellent the larger errors at -21% and 9% could be improved.

TABLE V
VALIDATION STATISTICS

Metric	Difference	% Difference	Abs % Difference
Mean	0	0%	1%
Median	-1.18	0%	1%
Max	186	9%	21%
Min	-216	-21 %	0 %

Further analysis of the validation sets has been performed by looking at the best and worst predictions made by the network. Figure 8 shows the location, and actual concentration values, of the best and the worst predictions. The best points are, as expected in the middle range of the values of methane mixing ratio seen in the data, whereas the worst points are mostly the higher or lower value concentrations.
However, even in the worst 10% the majority of the absolute percentage errors are less than 7.5% and in the range -10% to 10%. This indicates that the points where the network perfoms the worse occur very rarely.
Overall the network is a good predictor for the majority of data but some work is needed to improve the results, particularly in the more extreme valued regions.

V. DISCUSSION

The final results presented in section IV are of sufficient accuracy to provide a suitable argument for using a neural network to replicate the retrieval of methane mixing ratio from satellite data. The final network achieves an excellent mean error of 1.36% and provides a strong base for further work on this problem.
The earth observation science community require results of higher accuracy in the extreme regions. These regions are particularly important as these are the areas where a fast retrieval can be the most useful.
The experiments presented regional differences in the data which can be partially accounted for with the inclusion of latitude and longitude. However, as these only act as a proxy

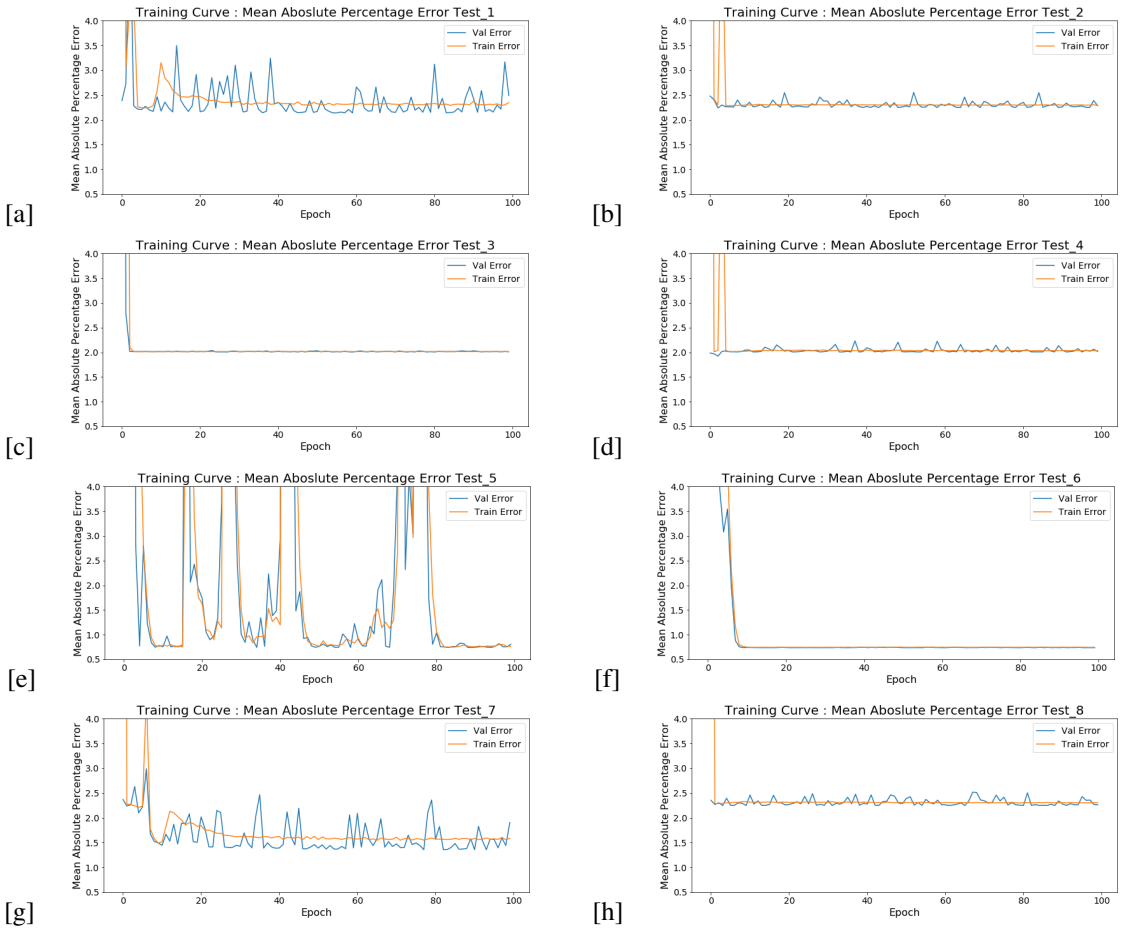

Fig. 5. Training Curves for the tested networks

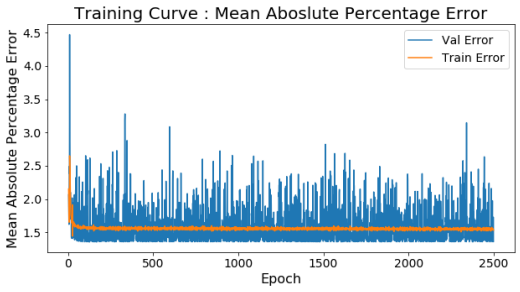

Fig. 6. Training curve for final network over 2500 epochs

for the geographical nature of different regions, they are not ideal for inclusion. The geography of a region may be similar to that of another region located further away where latitude and longitude would not account for the similarities in these cases. One way to rectify this is to replace latitude and longitude with some other geographical information, for example altitude. This may further improve the network results, but could remove the benefit of only using data directly from the satellite (L1) if any information needs to be inferred or calculated.

The biggest inadequacy with the final neural network is that it cannot predict the methane mixing ratio values of the more extreme data points. This is largely due to the fact that they rarely occur in the data and so are not prevalent in the training set. Therefore, the network has less examples to use and does not accurately learn about these points. As only a subset of the available data was used in these experiments it is possible to widen this data set and ensure that more of the extreme data points are present in the training set.

The final network training curve shows that there is not yet overfitting in the training set, however as the results are not improved by increasing the number of training epochs it is possible that the network has found a local minimum and so the results may not improve with further training. If this is the case then a larger batch size can be used to try and avoid this, and improve the results.

Overall the results presented serve as sufficient motivation for using neural networks to replicate the retrieval process for the methane mixing ratio, and potentially other gases, but further progress is needed to produce results at the necessary standard to be viable.

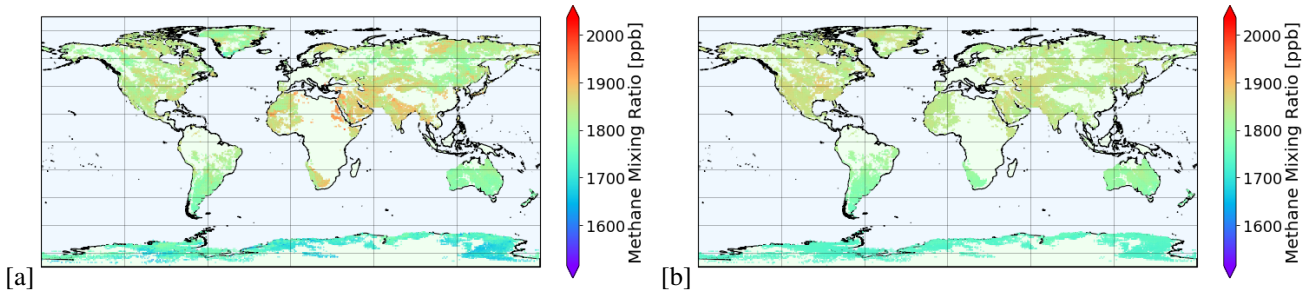

Fig. 7. Plots of a. Actual Data and b. Predicated Data

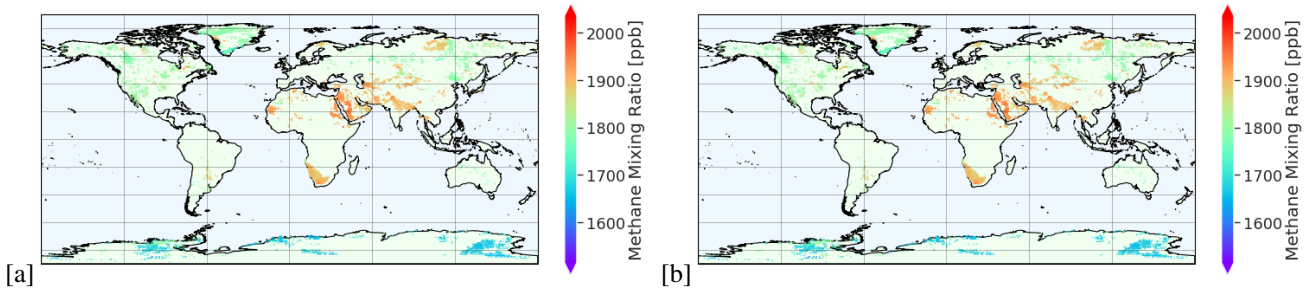

Fig. 8. Plots of the a. Best 10% of predictions and b. Worst 10% of predictions

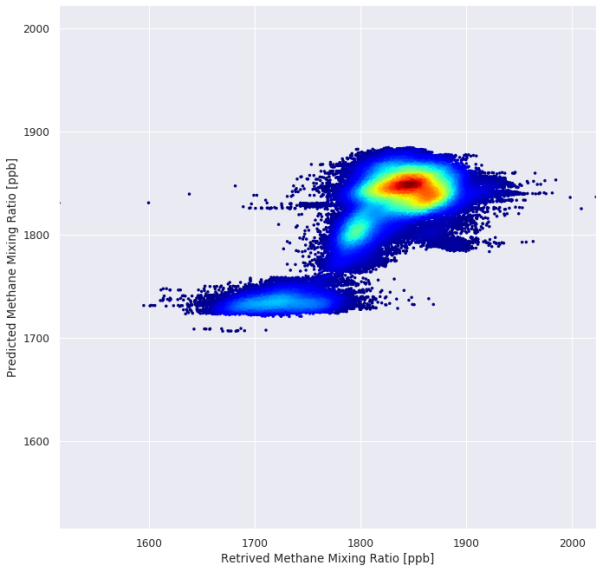

Fig. 9. Actual Vs Predicated Correlation Plot

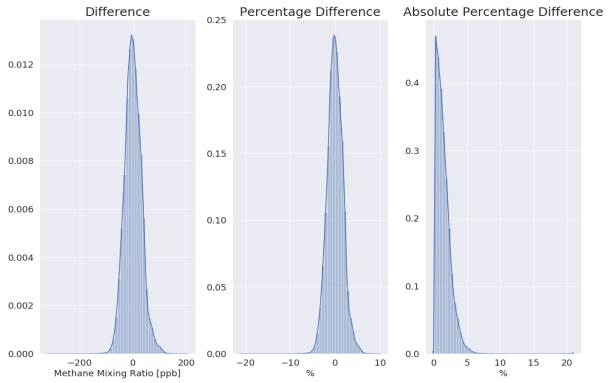

Fig. 10. Error Distribution

VI. Further Work

Further experiments to improve the results can be performed the first and most simple is to experiment with batch size ,decay rate and learning rate to help to avoid local minimums in the data. Latitude and longitude will be replaced with a less regional measure, such as altitude, to help account for the effect of geography on the methane mixing ratio. The training set will be enhanced to include more of the extreme values in an attempt to improve the networks ability to predict these. The network will be trained further, over more epochs until overfitting is show to find the optimum stopping point.
After the network has been tweaked and the best possible results achieved then some further work to enhance the results will be applied, error correction techniques and sensitivity analysis will be looked at as possible method of improvement. Finally, as a single hidden layer network has been used the computing power needed to produce results, once trained, is low. This allows for the possibility of implementation aboard a satellite, meaning near real time retrieval could be possible, which would be a substantial improvement and is an exciting prospect for future work.

VII. Acknowledgements

1) This research used the MA-GPU Computing Facility in the department of Mathematics at the University of Leicester.

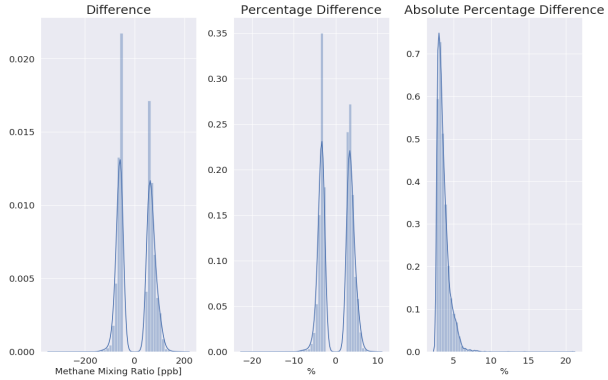

Fig. 11. Error Distribution for the worst areas

2) This research used the JASMIN Computing Cluster.

REFERENCES

[1] Hu, H., Landgraf, J., Detmers, R., Borsdorff, T., Aan de Brugh, J., Aben, I., et al. (2018). Toward global mapping of methane with TROPOMI: First results and intersatellite comparison to GOSAT. Geophysical Research Letters, 45, 3682– 3689. https://doi.org/10.1002/2018GL077259

[2] Atbd for methane. http://www.tropomi.eu/sites/default/files/files/publicSentinel-5P-TROPOMI-ATBD-Methane-retrieval.pdf