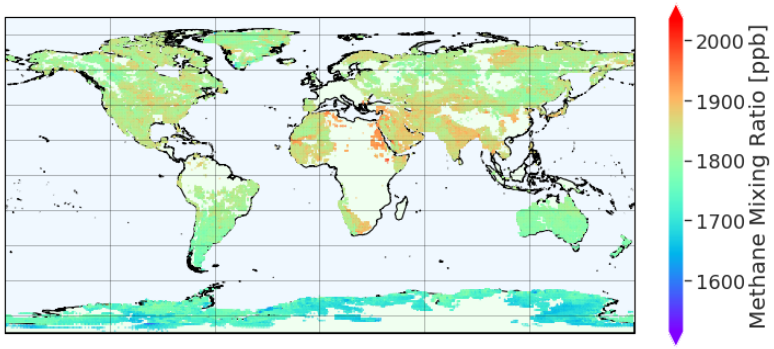
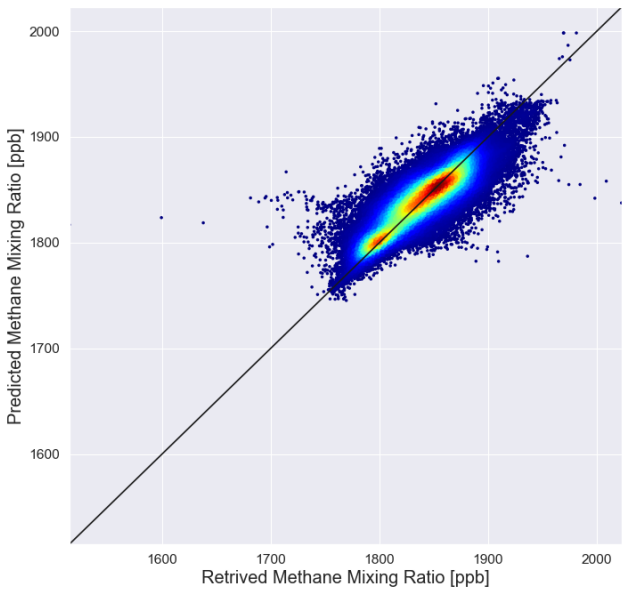
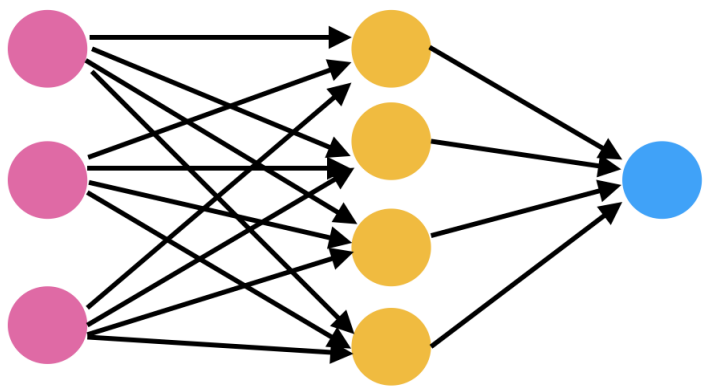
Machine Learning Techniques for the Retrieval of Methane from the Sentinel-5 Precursor Satellite

Rose Fenwick, Prof. Hartmut Boesch, Prof. Ivan Tyukin, Dr. Robert Parker





The Sentinel-5 Precursor satellite generates massive amounts of data. The retrieval of Methane from this data is computationally expensive and time consuming.



A Neural Network can be trained to reproduce this process, saving computational expense and time.

With thanks to the S5P team, and the team at SRON.

Why Use Machine Learning?

This data is both massive and has an unknown distribution, making it an excellent application of Machine Learning.
Neural Networks are scale-able and computationally efficient, unlike the current retrieval methods.
Computationally expensive methods are costly and slow, so a neural network solution would be very beneficial.

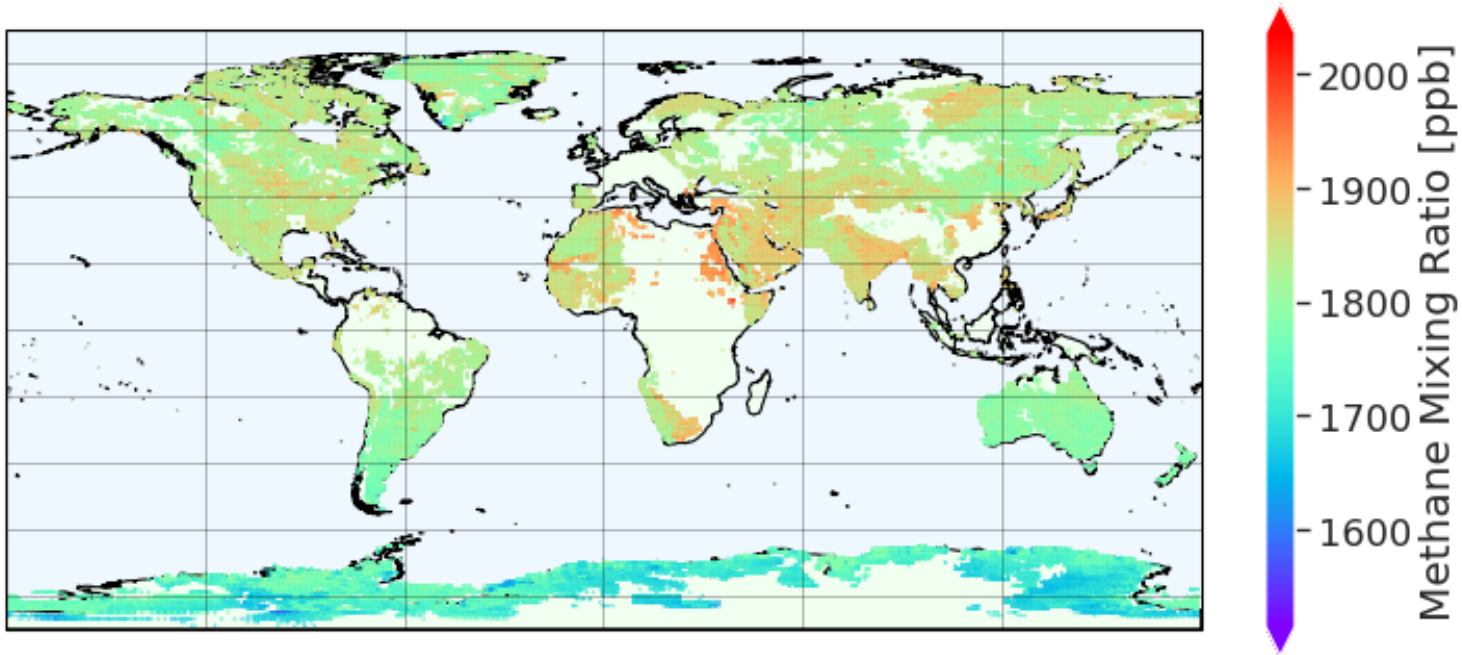


Figure 1: Global Coverage

As the input data is on different scales, the data used was standardised using $S = \frac{X-\mu}{\sigma}$ to remove this bias.

Neural Network

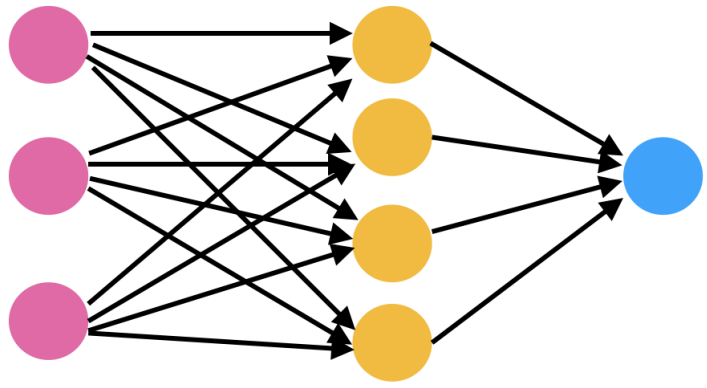


Figure 2: Neural Network Design

- Training set : 429,278 data points
- Test set : 179,489 data points
- Inputs include : Spectral Data, Solar Zenith Angle, Solar Azimuth Angle
- Target : SRON bias corrected CH_4

- 2 hidden layers
- Nodes : [128,256]
- Epochs : 2500
- Loss : Root Mean Squared Error

Results

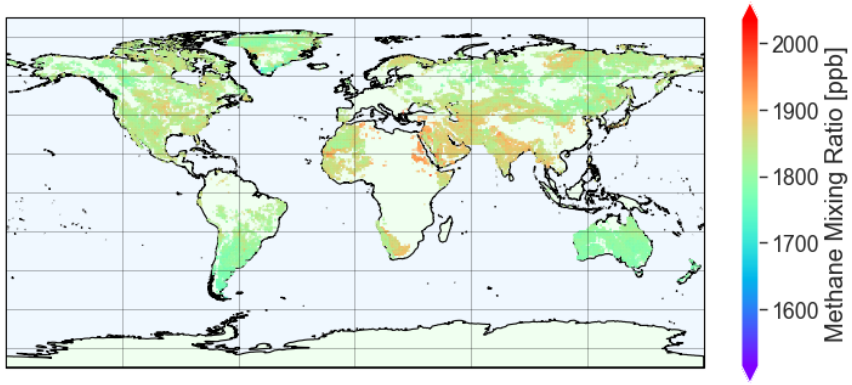


Figure 3: Validation : Actual

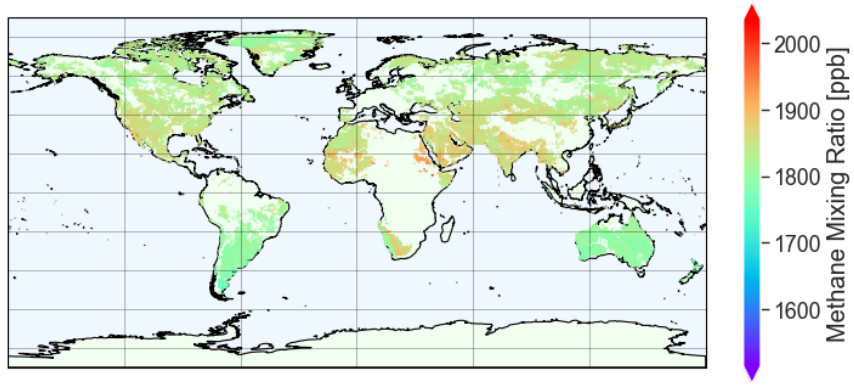


Figure 4: Validation : Predicted

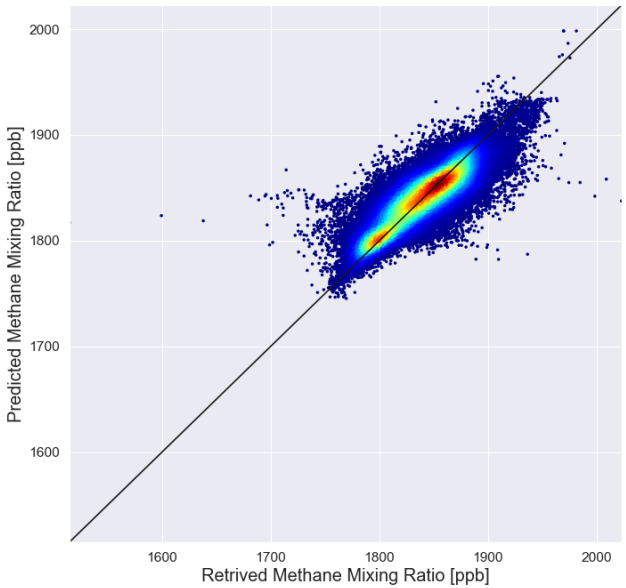


Figure 5: Actual Vs Predicted Density Plot

Set	Result
Training	15.97 (0.73%)
Test	16.07 (0.73%)
Validation	16.23 (0.73%)

Table 1: Results, Root Mean Squared Error (Percentage Error)

Feature Selection

PCA is applied in order to reduce the dimension of the data, reducing the dimension in this way allows for faster training whilst retaining the majority of the variability in the data.

Three principle components were selected to transform the data, since these explain 99.51% of the variance in the data (Figure 6).

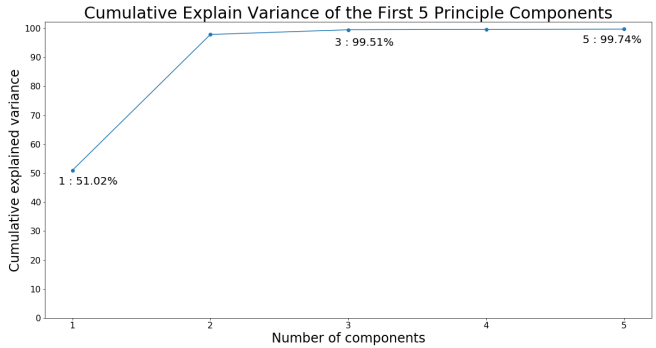


Figure 6: Explained Variance of PCs

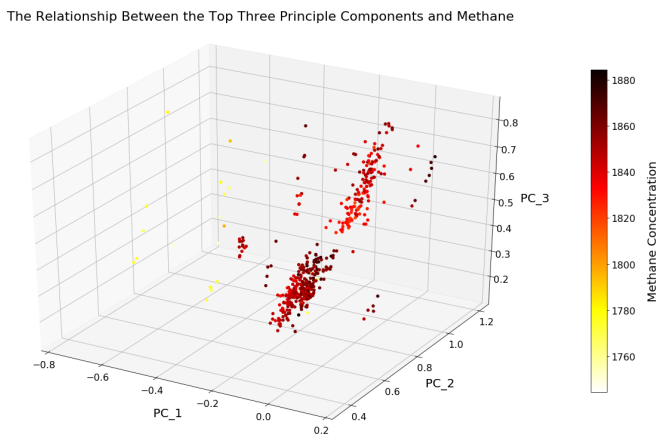


Figure 7: Top 3 Principle Components

Data of dimension 3 can be plotted in a meaningful way, Figure 7 shows the transformed data along with methane concentration.

There is a clear pattern of methane concentration within the three principle components.