MAST30024 Assignment 1

Trung Chi Dang - 1069862

September 2021

1 Question 1

1.1

We don't use l-2 norm since this will make TCs became unequally important. L-2 norm introduces a regularisation term, hence, when dividing TC by l-2 norm, TC will lost its unit variance.

1.2

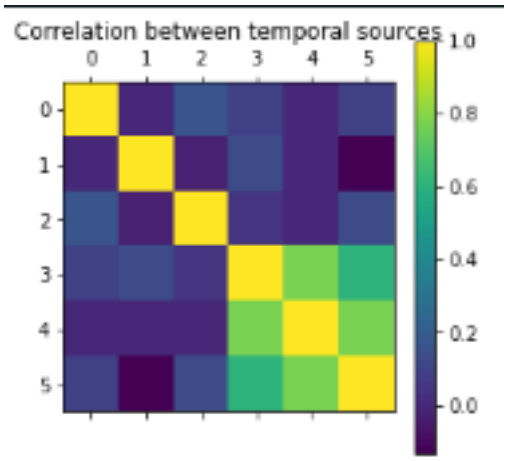


Figure 1: Correlation matrix

From Figure 1, we observe that there is high pairwise correlation between three temporal sources (third, fourth and fifth). It can not be easily differentiated which pair (either between third and fourth one or between fourth and fifth one) has higher correlation from the CM. But these two pair has highest correlation.

1.3

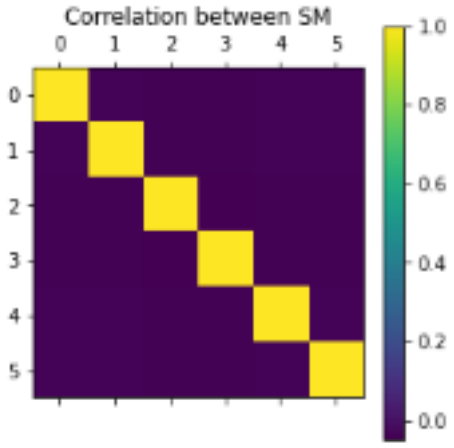
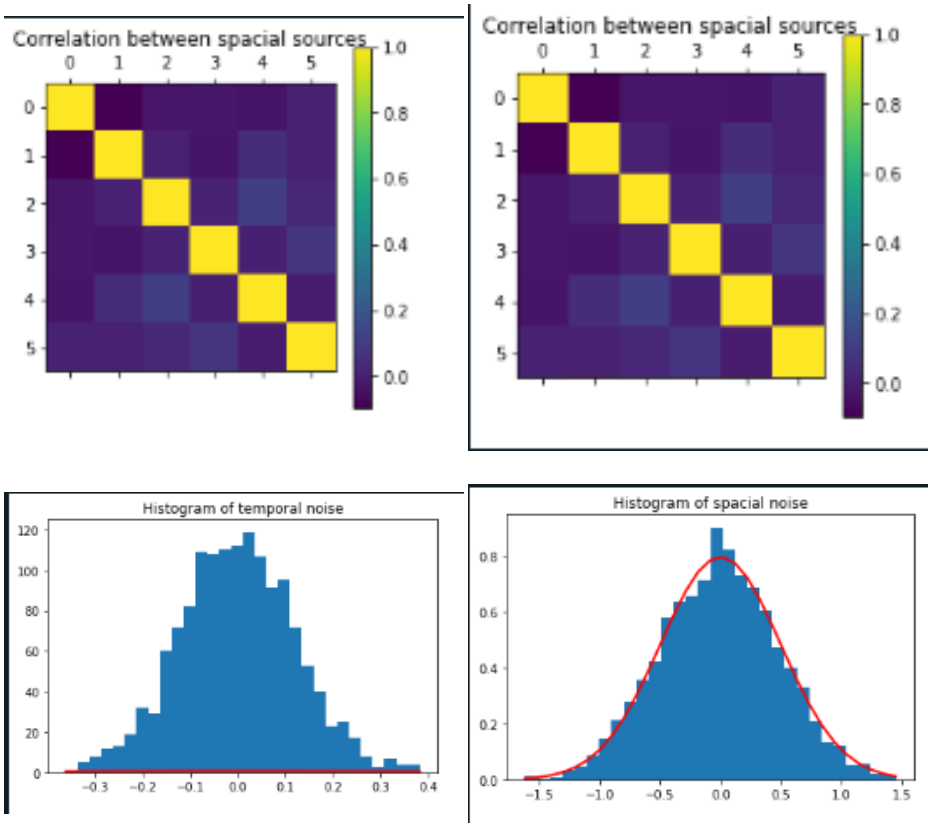


Figure 2: Correlation matrix

From Figure 4, it is observed that there is no correlation between these SM indicating independence among them.

1.4



From four above figure of temporal and spacial noises, we observe that these sources are not correlated and both type of noises follow normal distribution.

```
Percentage of value outside range for spacial noises:     3.8548752834467117
Mean of spacial noises: -0.009947798110354329
Percentage of value outside range for temporal noises:     4.722222222222222
Mean of temporal noises: -0.0009244047990367235
```

Figure 4: Output from notebook

Based on Python output above, these two distribution has approximately fulfilled both mean and variance condition of normal distribution. Despite some violation, this is a very small percentage and can not be fixed due to randomness.

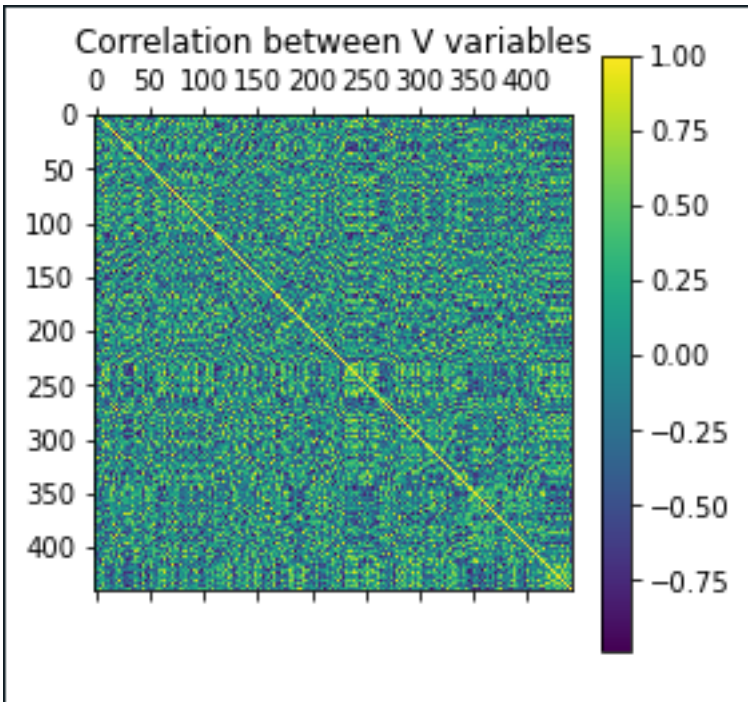


Figure 5: Output from notebook

From Figure 5, overall, across 441 variables, the color indicates there is high proportion of correlated variables

1.5


Plot of 100 random selected time series
3
2
1
0
−1
−2
−3
0
50
100
150
200
250

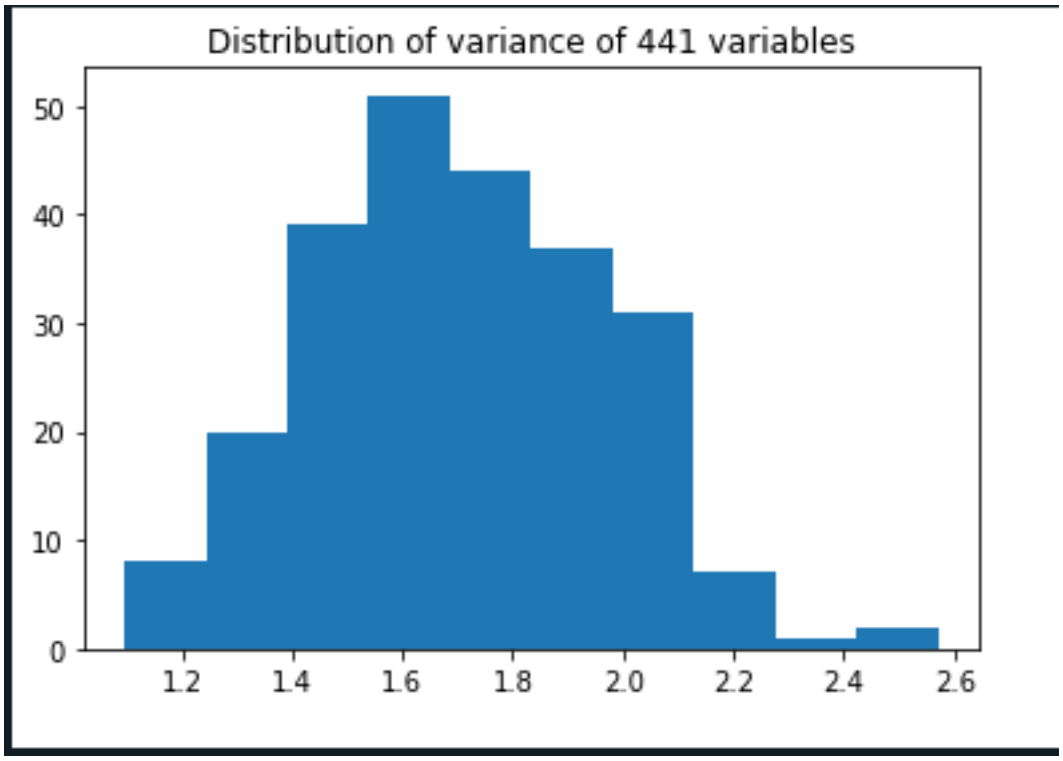


Figure 6: Output from notebook

Overall, distribution of variance among variables follows approximate normal distribution, ranging from 1.2 to 2.6. Therefore, it is necessary to standardise the dataset to make the variables equally importance.

2 Question 2