Assignment 1

MAST30034

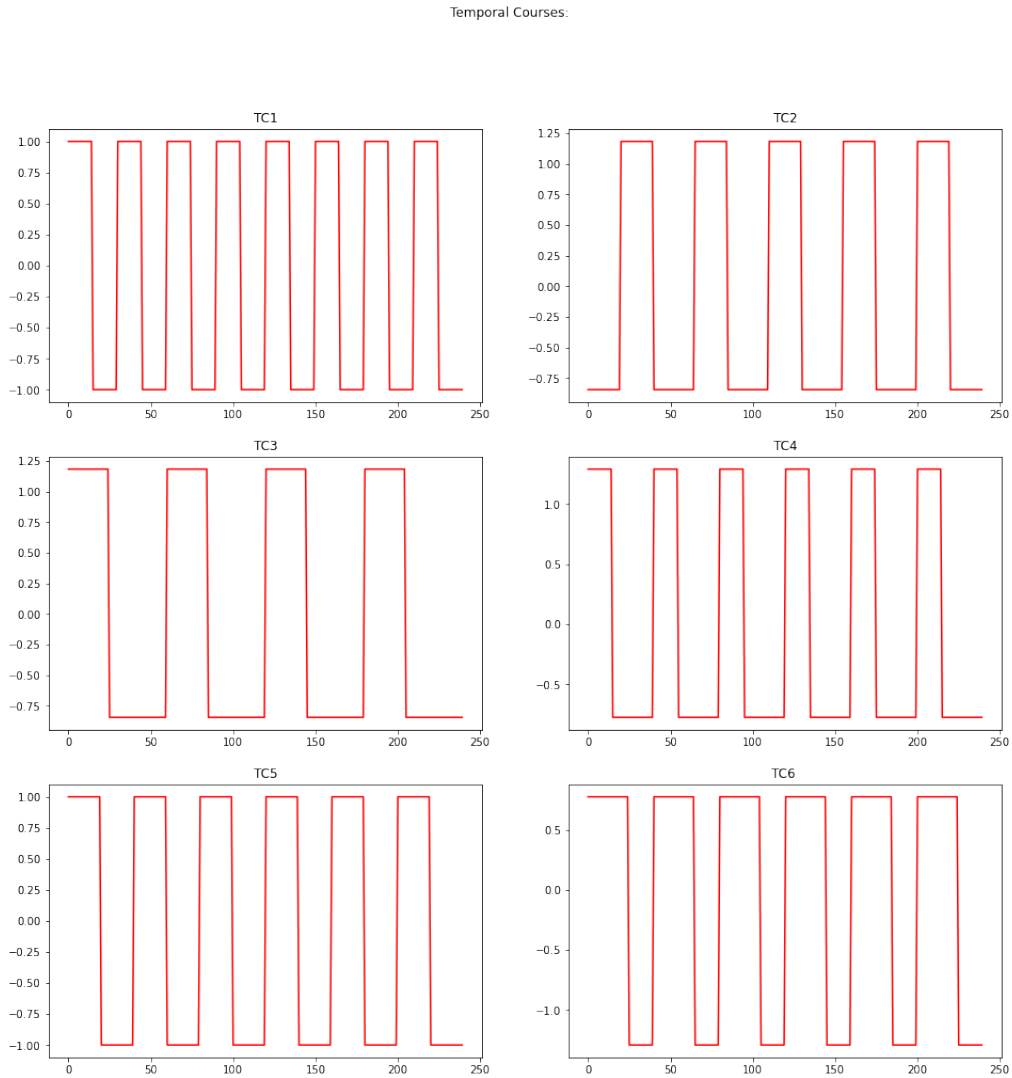
Junzhi Ning
Student ID: 1086241

September 4, 2021

Note that the precise results obtained from this analysis might not be exactly consistent with codes due to the fact that the random noise is added to this Synthetic dataset.

Question 1

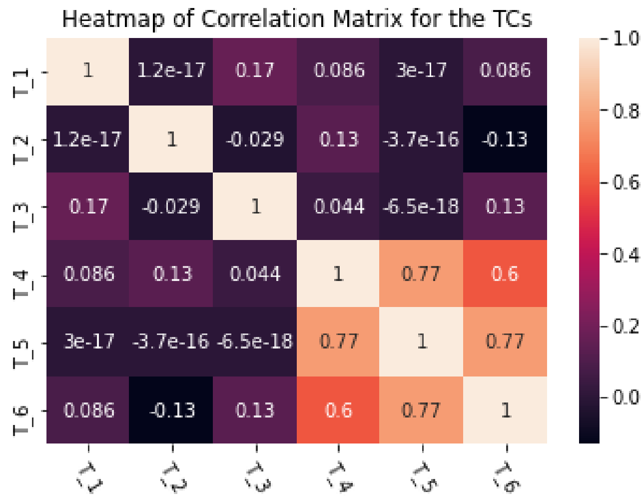
1.



Why not normalize (divide by l-2 norm) the TCs instead of standardizing it?

Dividing by l-2 will not center the Time sources at 0 and the vertical scale of the time courses will be negligible within a very small interval. It makes the comparisons between different TCs harder. In later questions, if we apply the PCA, centering at 0 for time courses is therefore needed to make them bias free.

2.

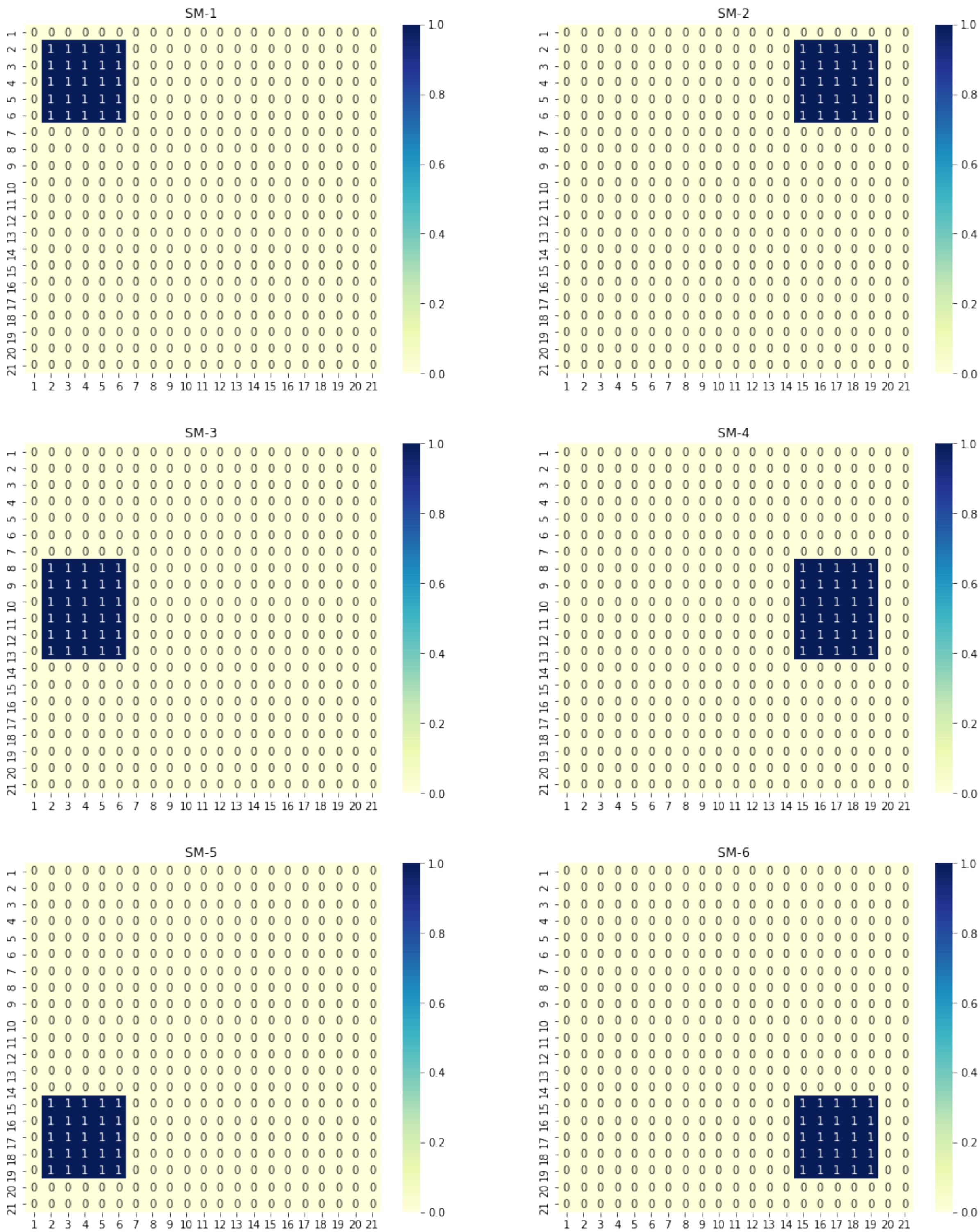


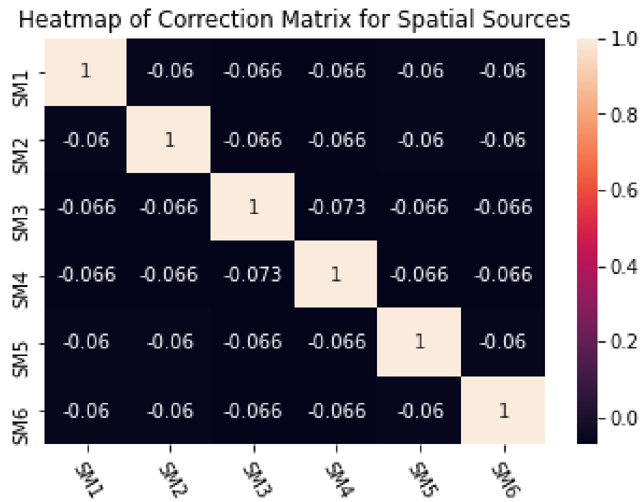
Show its plot, and can you tell visually which two TCs are highly correlated? If not, can you tell this from CM?

From the above heat-map, TC 5 and TC 6 are highly correlated with a positive correlation of 0.77.

3.

Spatial Maps Visualisation:

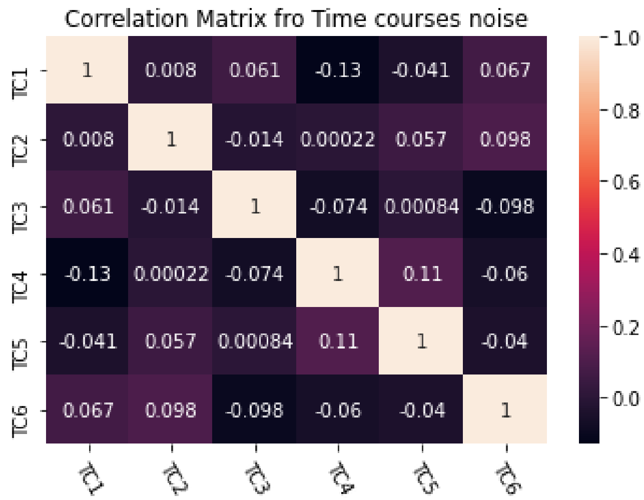


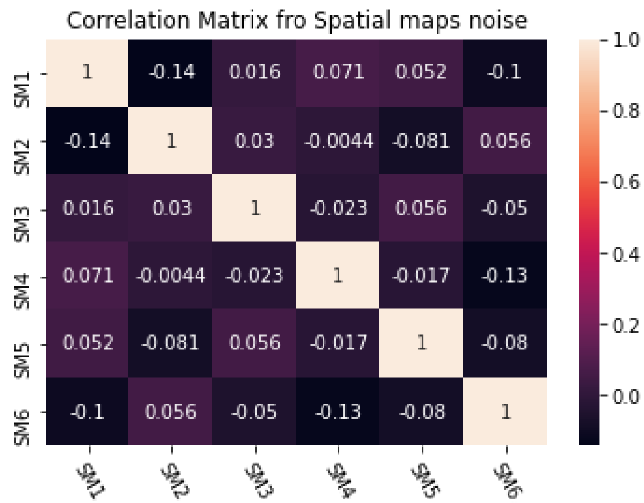


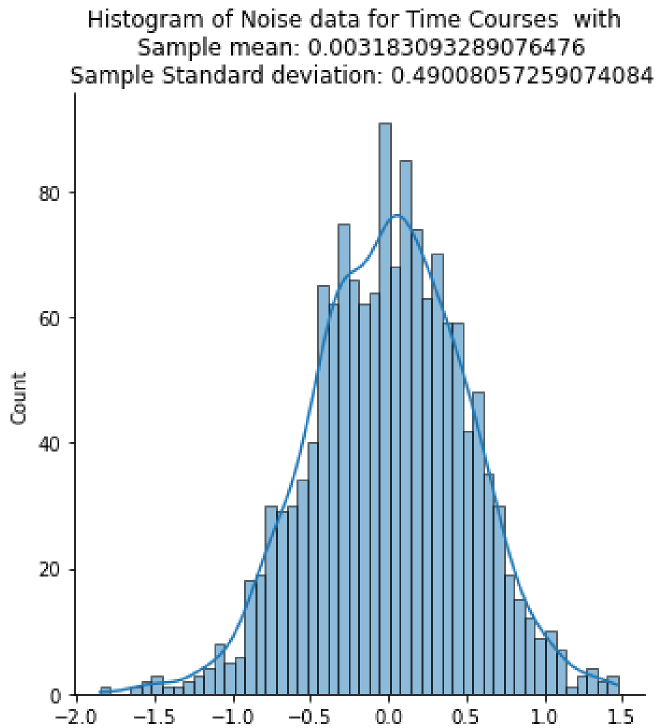
Using CM show if these 6 vectored SMs are independent? For our particular case, why standardization of SMs like TCs is not important?

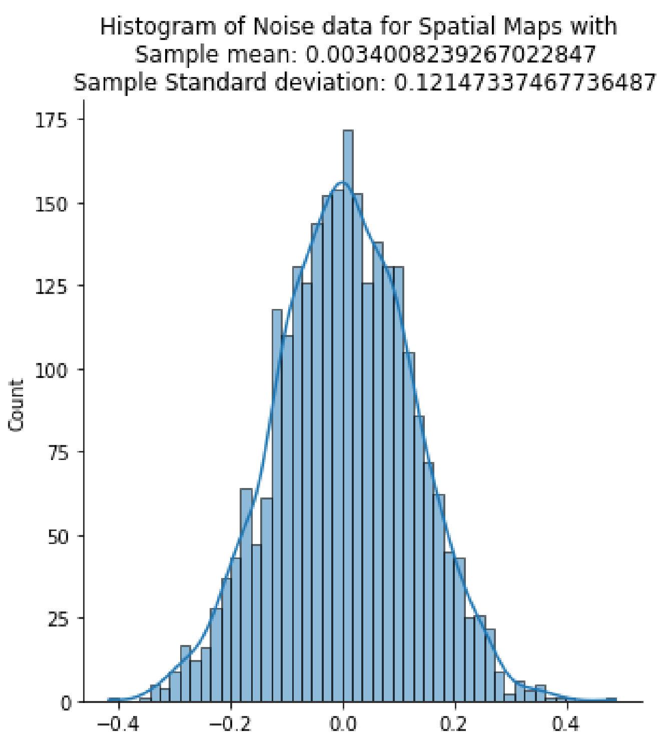
From CM above, it seems to suggest that the SMs are independent of each other. All values on those slices lie in the same range of values, it either takes 0 or 1 and the quantity of ones for each spatial maps are almost the same(5×6 or 5×5), so the differences in standardized SPs are small across six spatial maps and it is not necessary to standardize them.

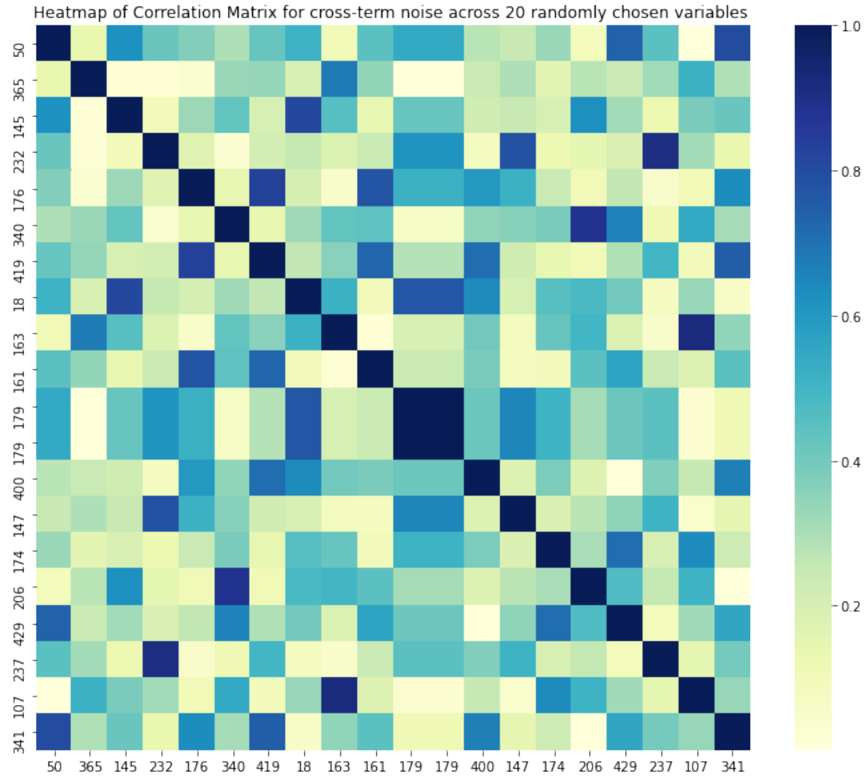
4.


Correlation Matrix fro Time courses noise
TC1
TC2
TC3
TC4
TC5
TC6
1 0.008 0.061 -0.13 -0.041 0.067
0.008 1 -0.014 0.00022 0.057 0.098
0.061 -0.014 1 -0.074 0.00084 -0.098
-0.13 0.00022 -0.074 1 0.11 -0.06
-0.041 0.057 0.00084 0.11 1 -0.04
0.067 0.098 -0.098 -0.06 -0.04 1
1.0
0.8
0.6
0.4
0.2
0.0


Correlation Matrix fro Spatial maps noise
SM1
SM2
SM3
SM4
SM5
SM6
1 -0.14 0.016 0.071 0.052 -0.1
-0.14 1 0.03 -0.0044 -0.081 0.056
0.016 0.03 1 -0.023 0.056 -0.05
0.071 -0.0044 -0.023 1 -0.017 -0.13
0.052 -0.081 0.056 -0.017 1 -0.08
-0.1 0.056 -0.05 -0.13 -0.08 1
1.0
0.8
0.6
0.4
0.2
0.0


Histogram of Noise data for Time Courses with
Sample mean: 0.003183093289076476
Sample Standard deviation: 0.49008057259074084
Count
0
20
40
60
80
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5

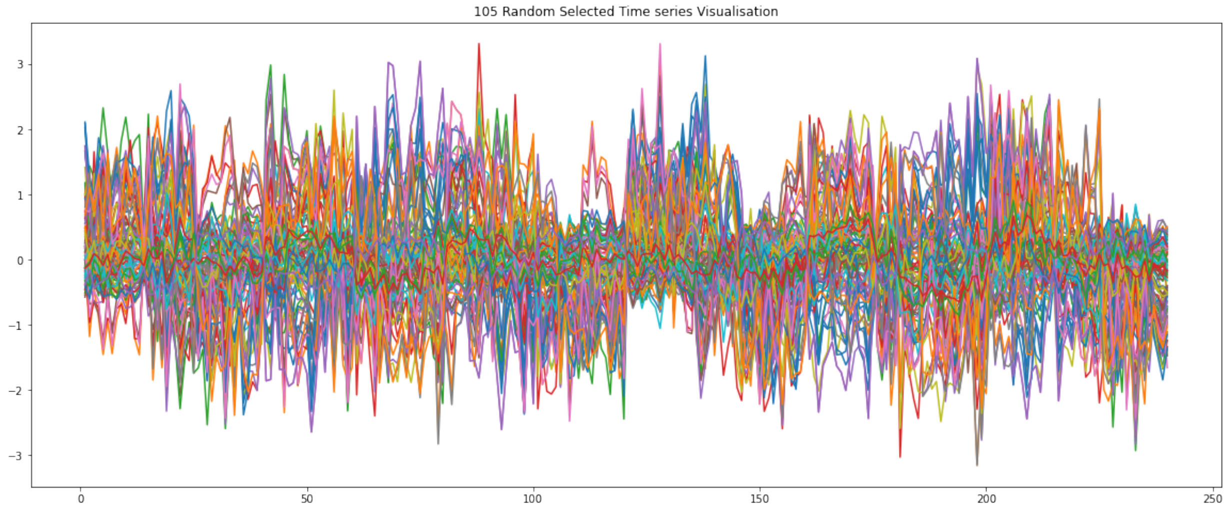

Histogram of Noise data for Spatial Maps with
Sample mean: 0.0034008239267022847
Sample Standard deviation: 0.12147337467736487
Count
0
25
50
75
100
125
150
175
−0.4
−0.2
0.0
0.2
0.4

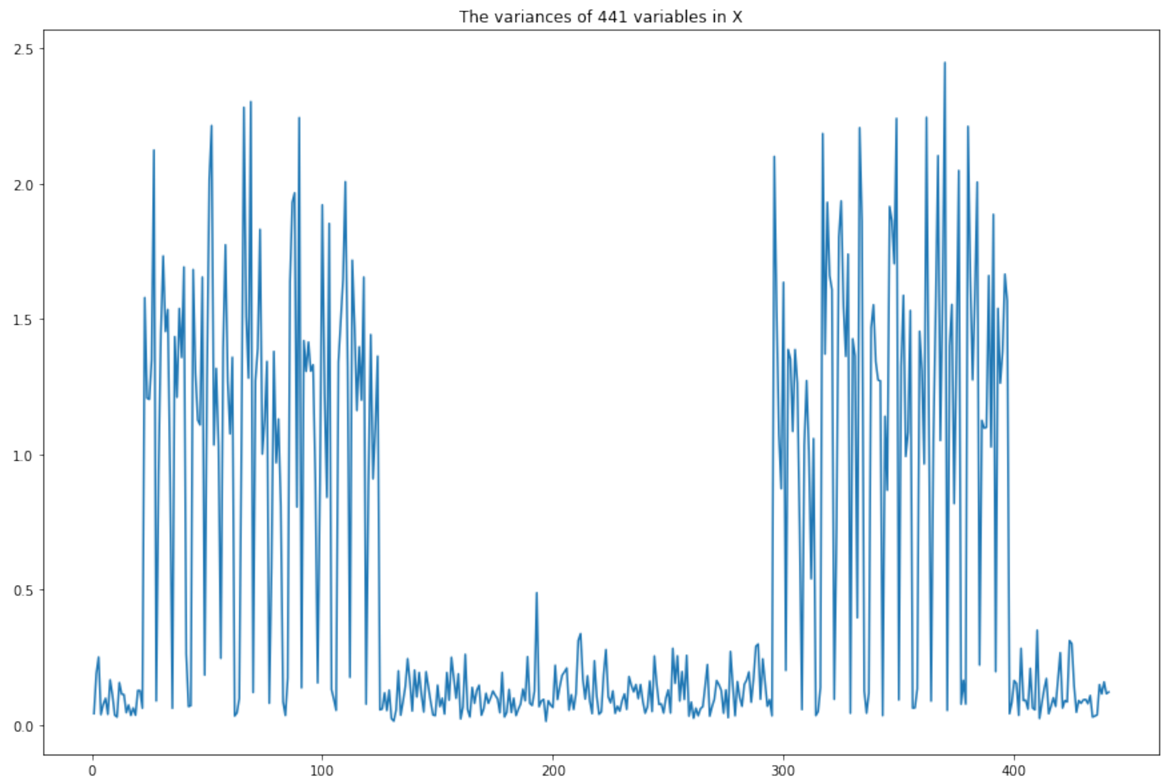


Using a 6 $\times$ 6 CM for each noise type (spatial and temporal) can you show if they are correlated across sources? Also plot the histogram of both noise sources to see if they have a normal distribution? Does this normal distribution fulfils the mean and variance= 1.96 criteria relating to 0.25, 0.015, and zero mean? Is there product t s correlated across V number of variables?

The empirical distributions of both noise sources are almost following the normal distributions $N(0, 0.5)$ *and* $N(0, 0.015)$.
From the Heat-map of the correlation matrix, we can see that $\Gamma_s\Gamma_t$ are moderately correlated across the 20 randomly chosen variables. This corresponds to the fact that $\Gamma_s\Gamma_t$ is generated by the matrix multiplication between Γ_s and Γ_t.

5.





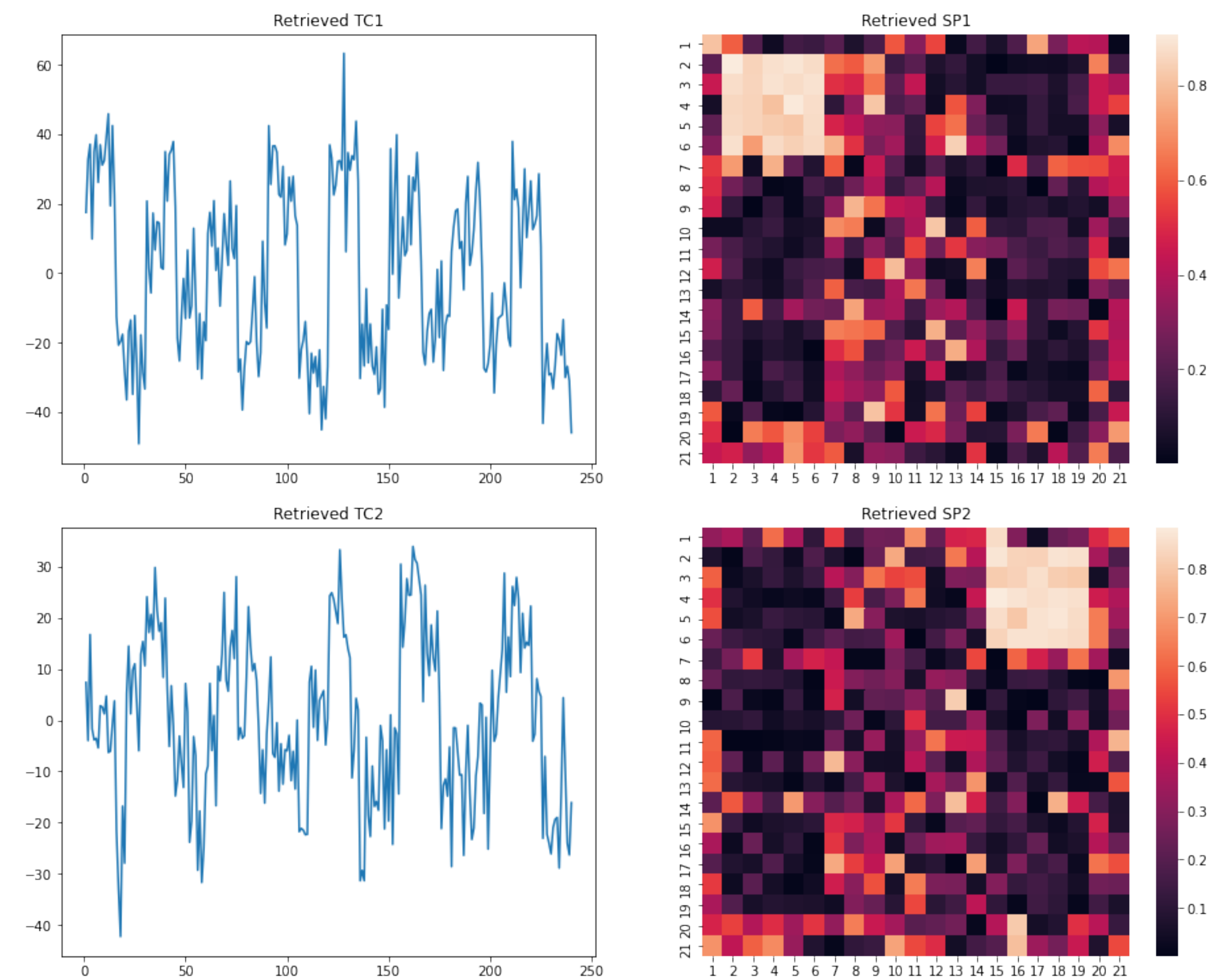
Can these products $TC \times \; s$ and $\; t \times SM$ exist, If yes what happened to them because if we keep them then we cannot fit our model onto (1)? Plot at least 100 randomly selected time-series from X as few of them are shown in Figure 2 (left). Also plot variance of all 441 variables on a separate plot. What information does this plot give you?

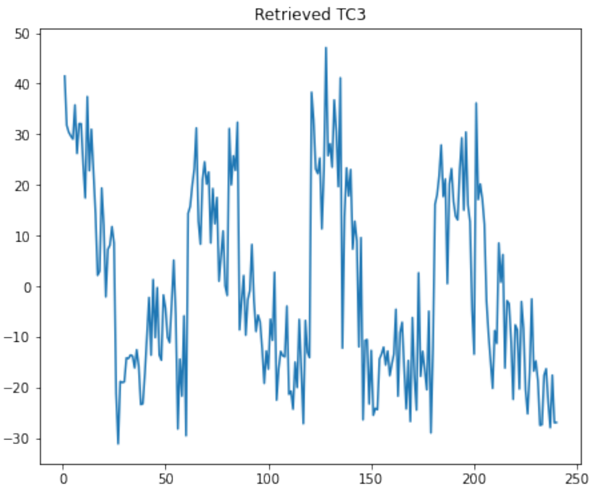
These terms can only exist when creating the synthetic dataset, but when we are trying to fit the model described in the equation (1), we ignore the existence of these cross terms and the effect of these terms are fitted within $\hat{A}$ in the model, this can be shown in the later plots of retrieved spatial maps where the noise data are very conspicuous other than the pixel values of 1 and 0 in the Original Spatial Maps.

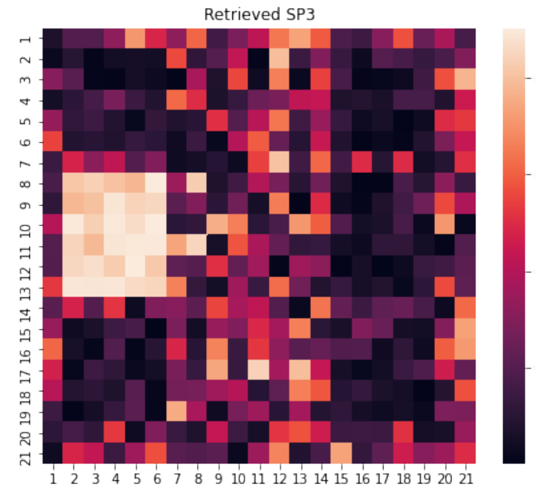
The plot of variances of 441 variables in X tells us where the pixel values have the large variations over the temporal sources and spatial maps, it might be useful to provide some information about the values in A when fitting model. Due to the fact that X is a synthetic dataset, we know how exactly X is created and how the spatial maps(SMs) looks like. To be more specific, if we reshape the spatial maps vertically while generating the synthetic X, then pixel values with higher variance shown in this plot are correspond to pixel value of 1 from at least one spatial map in SMs.Although the real coefficient matrix will be usually unknown to us in a real world situation, the distribution of the variances of the X can still be crucial for analysts to understand the underlying structure of the coefficient matrix in a fitted model.

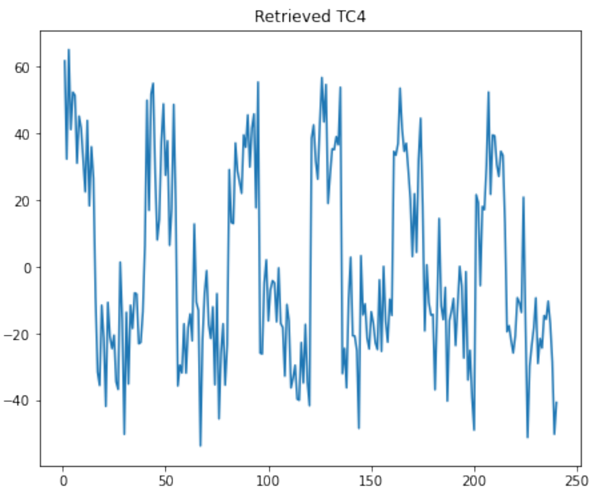
Question 2

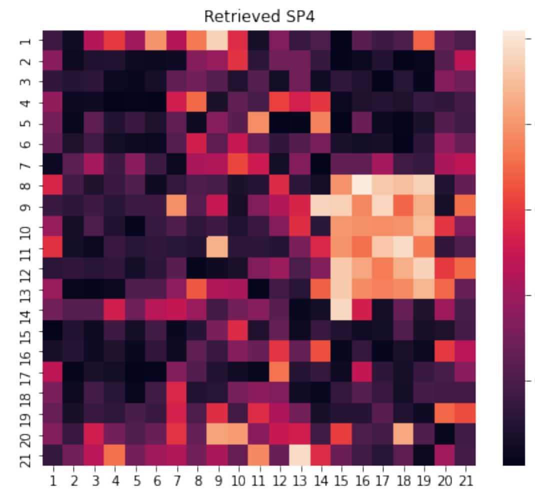
1.

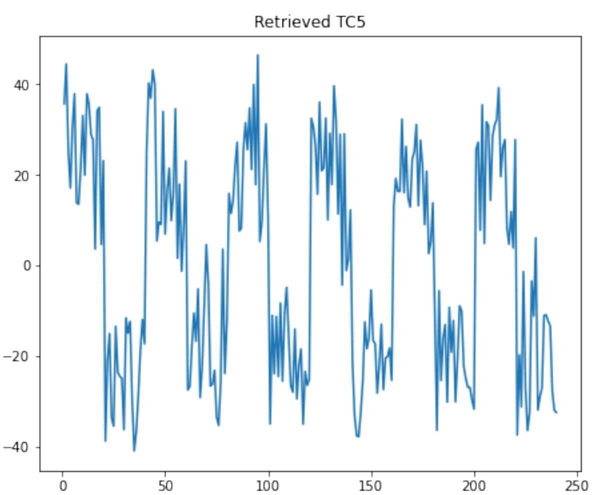


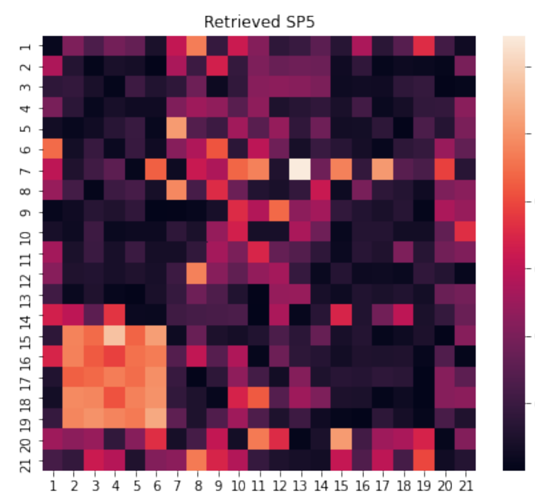

Retrieved TC3

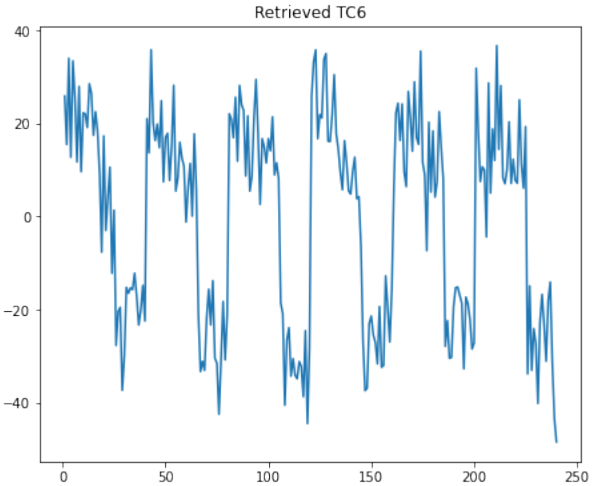

Retrieved SP3

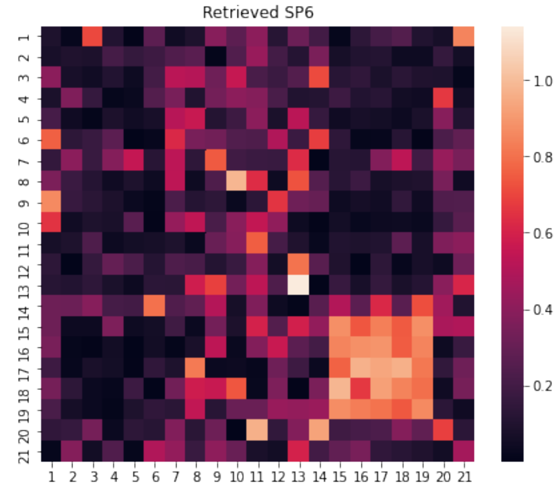

Retrieved TC4

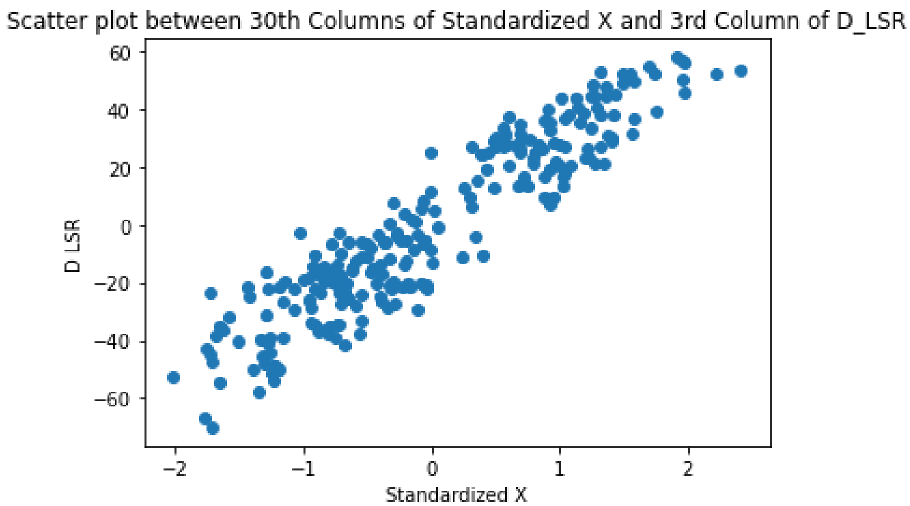

Retrieved SP4

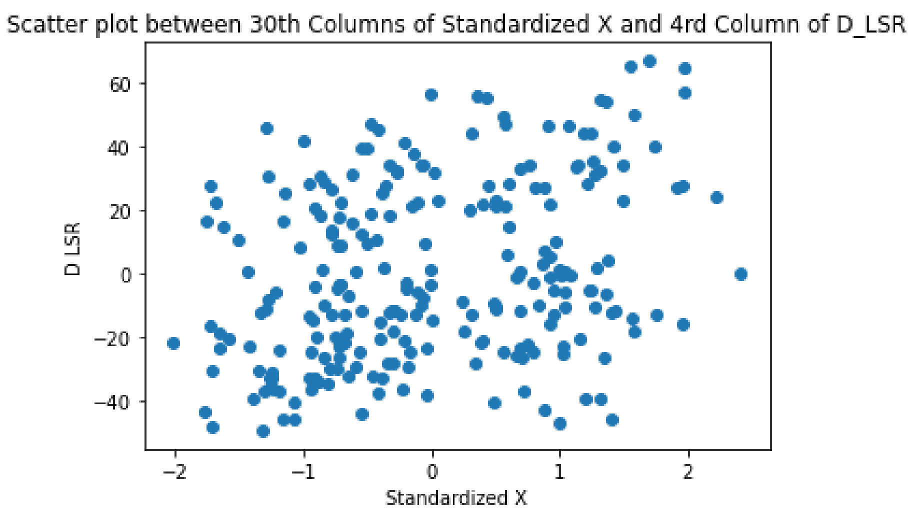

Retrieved TC5


Retrieved SP5


Retrieved TC6


Retrieved SP6


Scatter plot between 30th Columns of Standardized X and 3rd Column of D_LSR
D LSR
Standardized X


Scatter plot between 30th Columns of Standardized X and 4rd Column of D_LSR
D LSR
Standardized X

Plot six retrieved sources using A LSR and D LSR side by side as shown in Figure 2 (right) for one of the retrieved sources. Do a scatter plot between 3rd column of D LSR and 30th column of standardized X, you will find a linear relationship between them, why this does not exist between 4th column of D LSR and same column of X

Despite the presence of the noise data in D_{LSR} and A_{LST}, the third temporal source is still only main time course that contributed to the value of $30th$ element on the standardised X as the variances of the noise data added to TCs and SMs are relatively negligible and the only third spatial map turns out to be 1(excludes the noise) among 6 spatial maps at this pixel value, which ensures the effect from the the third temporal source being preserved after taking the dot product in the matrix multiplication.

2.

Calculate the sum of these two correlation vectors. If you have carefully selected the value of λ between 0 and 1 you must end up with $\sum c_{TRR} > \sum c_{TLSR}$, and remember $\hat{\lambda} = V$. Also, for $= 1000$, plot first vector from A_{RR} and the corresponding vector from A_{LSR} , Do you find all values in a^1_{RR} shrinking towards zero?

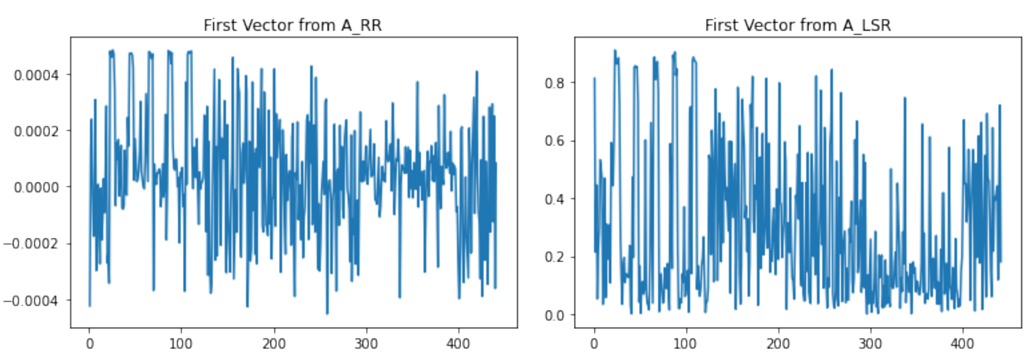
The λ value chosen is 0.43.
For this particular set of noise dataset we have:

Correlation Vectors between TCs and D_{RR}(left), D_{LSR}(right)

[0.8616225013750377,	[0.8422421825543195,
0.8454727193345866,	0.8925746166868138,
0.8582591980048191,	0.8665610224712175,
0.9057104600868033,	0.805690952031377,
0.8986793221863054,	0.8910959164314679,
0.902958843497664]	0.8168075947252129]

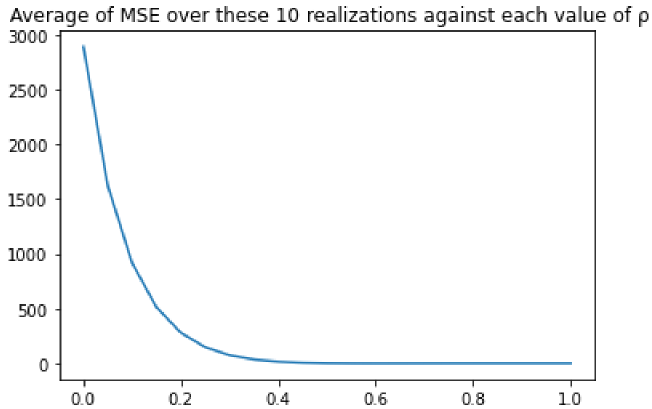
$$\Rightarrow \text{Clearly, } \sum c_{TRR} \approx 5.27270 > 5.11497 \approx \sum c_{TLSR}$$

For $\lambda = 1000$,

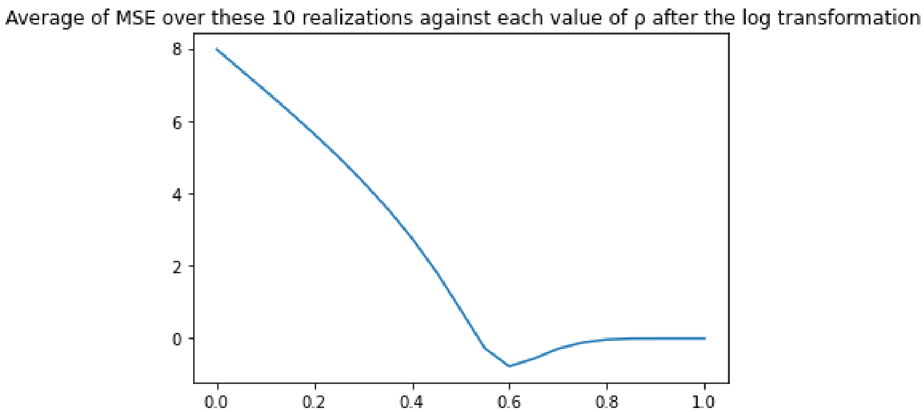


The range of first vector in A_{RR} is between -0.0005 to 0.0005. In comparison to that, the range of the first vector in A_{LSR} is between 0 and 0.1. Not hard to see that the values in a^1_{RR} are shrinking to zero.

3.



To find out the turning point more clearer around the values of 0, we can apply the log transformation of MSE curve to mark the turning point.



Then plot average of MSE over these 10 realizations against each value of . At what value of do you find the minimum MSE? Is it okay to select this value?

The optimal ρ value found in this case is 0.600 which minimizes the MSE over 10 realizations with respect to different noise data-sets, this result, in general, should be quite reliable as it is obtained upon 10 times of different realizations of noise sources. However, to obtain more reliable result of ρ, we can use the cross-validation technique to evaluate the MSE at different ρ values.

4.

The results of correlation vectors between each estimates of LR and RR:
For LR, c_{SLR}(left) and c_{TLR}(right)

```
[0.9514347210809001,
 0.9721467555740937,
 0.933131346168836,
 0.8957992247558692,
 0.7414606154664495,
 0.849186787250824]
```

```
[0.8912324394603541,
 0.8889213327513134,
 0.9040361649875314,
 0.9169654855944724,
 0.9247408601769325,
 0.898626339085396]
```

For RR, c_{SRR}(left) and c_{TRR}(right)

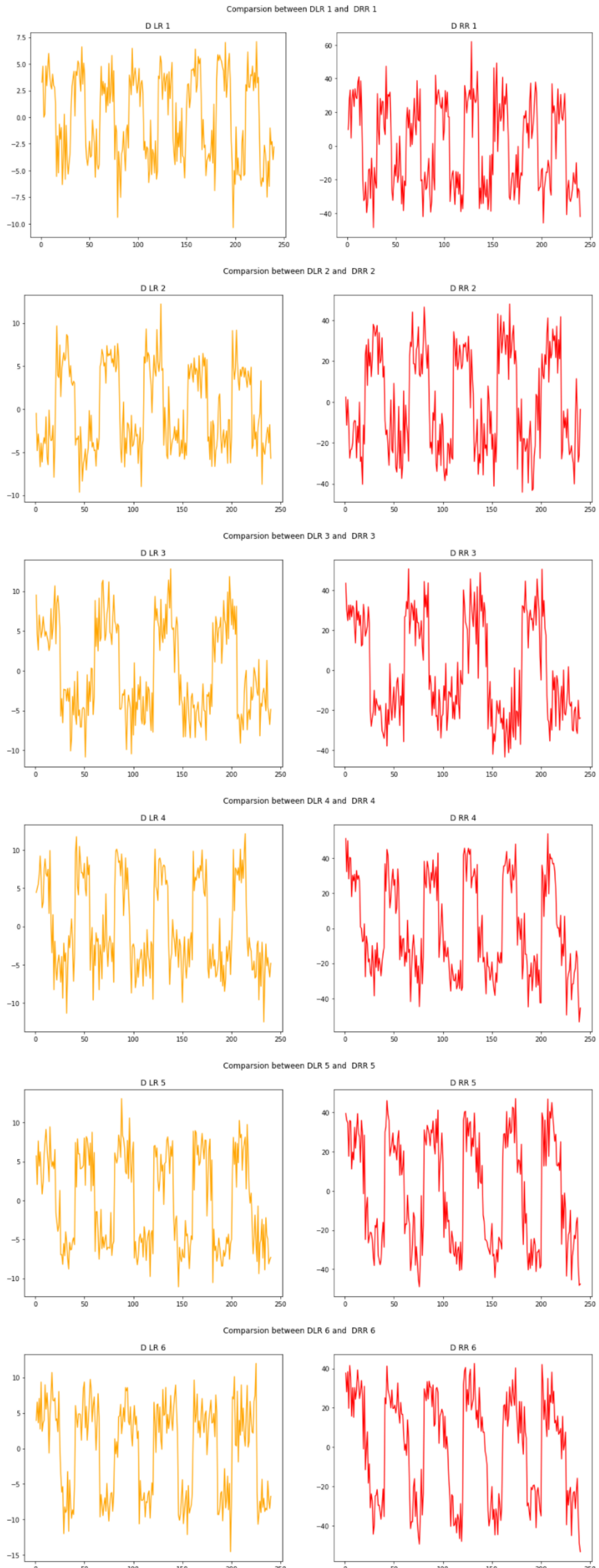
```
[0.5556779143508057,
 0.5310060136119317,
 0.5656032968340876,
 0.5430887105102842,
 0.44862746239100354,
 0.4906154397208619]
```

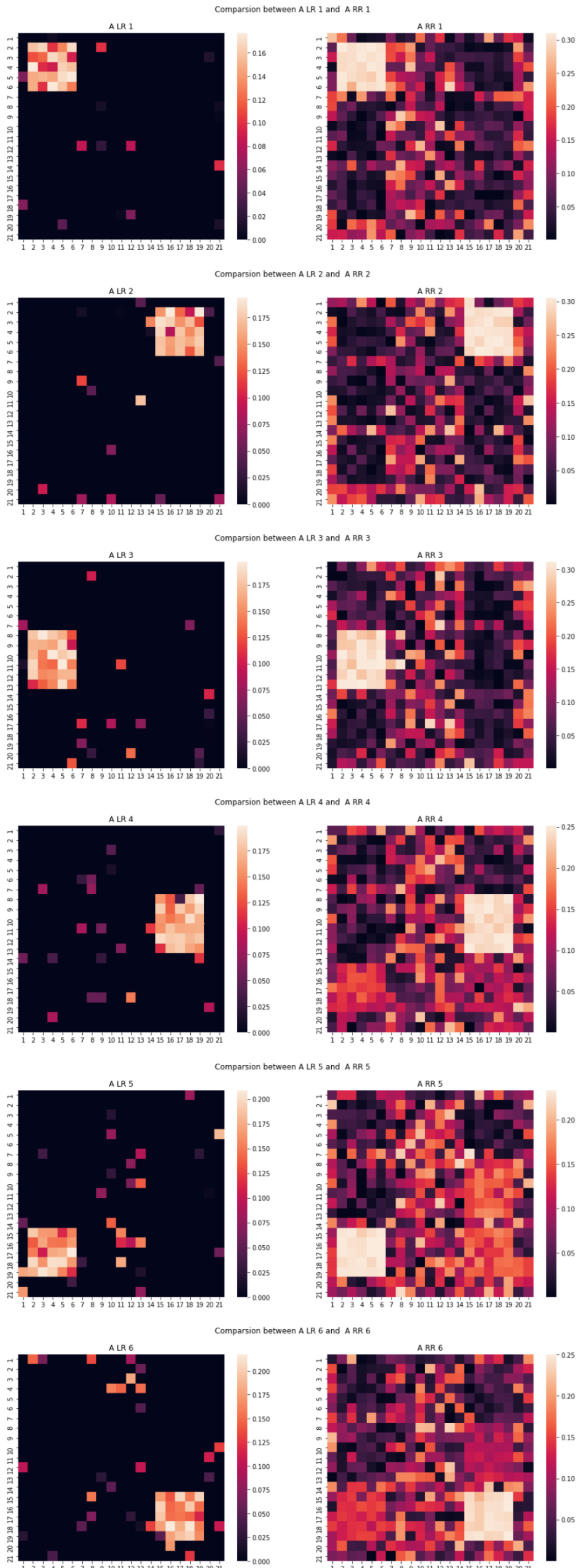
```
[0.8616225013750377,
 0.8454727193345866,
 0.8582591980048191,
 0.9057104600868033,
 0.8986793221863054,
 0.902958843497664]
```

Calculate the sum of these four correlation vectors. If you have carefully selected the value of you must end up with $\sum c_{TLR} > \sum c_{TRR}$ and $\sum c_{SLR} > c_S RR$. Plot side by side in form of 4 columns estimates of D and A for both RR and LR to know the difference visually. You will see a major difference in estimates of A in terms of false positives. Can you mention the reason behind this difference?

Clearly, by the inspection, the sum of c_{SLR} is greater than the sum of c_{SRR}.
For Time courses,

$$\sum c_{TRR} \approx 5.2727 < \sum c_{TLR} \approx 5.4245$$


Comparsion between DLR 1 and DRR 1
D LR 1
D RR 1
Comparsion between DLR 2 and DRR 2
D LR 2
D RR 2
Comparsion between DLR 3 and DRR 3
D LR 3
D RR 3
Comparsion between DLR 4 and DRR 4
D LR 4
D RR 4
Comparsion between DLR 5 and DRR 5
D LR 5
D RR 5
Comparsion between DLR 6 and DRR 6
D LR 6
D RR 6


Comparsion between A LR 1 and A RR 1
A LR 1
A RR 1
Comparsion between A LR 2 and A RR 2
A LR 2
A RR 2
Comparsion between A LR 3 and A RR 3
A LR 3
A RR 3
Comparsion between A LR 4 and A RR 4
A LR 4
A RR 4
Comparsion between A LR 5 and A RR 5
A LR 5
A RR 5
Comparsion between A LR 6 and A RR 6
A LR 6
A RR 6

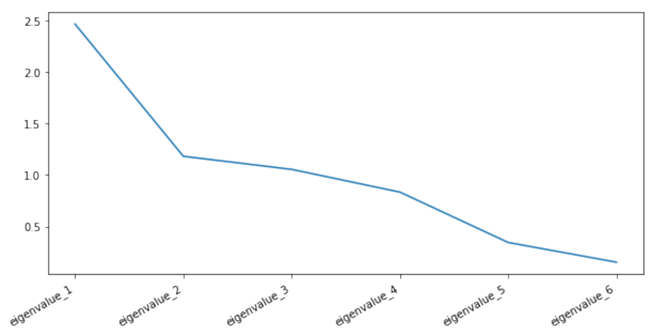
In estimated A_{LR}, many pixel values are now equal to zero and the retrieved Spatial Map are much clearer as the noise data we once added to the X has been filtered out to zero mainly due to the properties of the Lasso Regression allowing the coefficients equal to zero, unlike the Ridge Regression. The direct result of this is that the number of false positive has been greatly reduced on the retrieved spatial maps.

5.

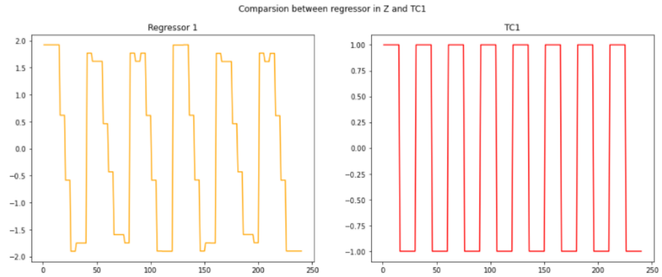
Plot their eigen values. For which PC the eigen-value is the smallest? Plot the regressors in Z and source TCs side by side. Did you notice deteriorated shape of PCs? Why the shape of TCs has been lost? Now keeping all components in Z apply lasso regression on X using = 0.001 and then Plot the results of D PCR and A PCR side by side (note that A PCR = B and your regressors are in Z (PCs of the TCs)). Did you notice the inferior performance of PCR compared to the other three regression models? Why is that so?

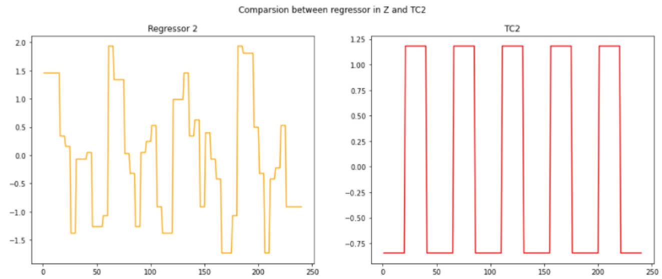
The smallest eigenvalue we here is correspond to the last PC = [0.07547725 -0.00182036 0.07644906 0.55928169 0.59912548 0.56276821], which has the value of 0.15379429. From the plots of regressors in Z, the shapes of PCs are very distorted and unclear compared to the original TCs. One possible for this shape lost is due to the procedure of Principle Component Analysis which only captures the maximum variation across different TCs.
Below are Principal components and eigenvalues:

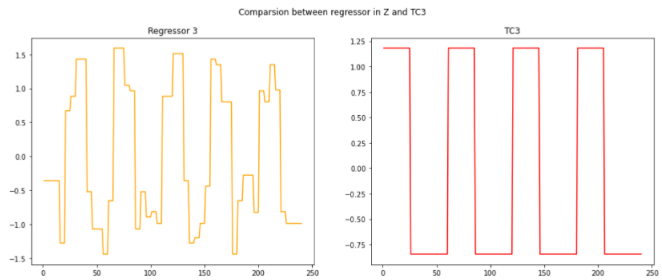
```
[ 0.07547725 -0.00182036  0.07644906  0.55928169  0.59912548  0.56276821]
[ 0.65622983 -0.23060954  0.69510789 -0.09419415 -0.14242208  0.06204901]
[ 0.27383063  0.9228611   0.08033206  0.17961053 -0.03861215 -0.18204424]
[-0.69491023  0.15452469  0.70146127 -0.02395617 -0.00205921  0.02440991]
[ 0.02352637  0.26694917 -0.0822161  -0.68756474  0.02130854  0.66949671]
[-0.07230956  0.00189375 -0.07563011  0.41563487 -0.78664821  0.4443873 ]
```

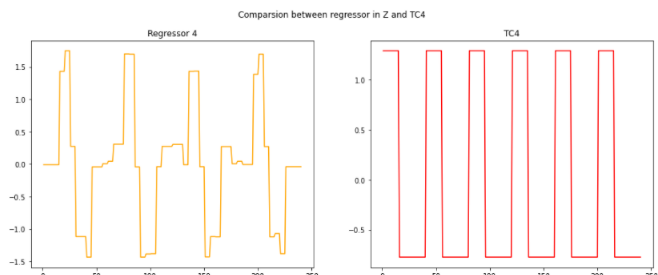


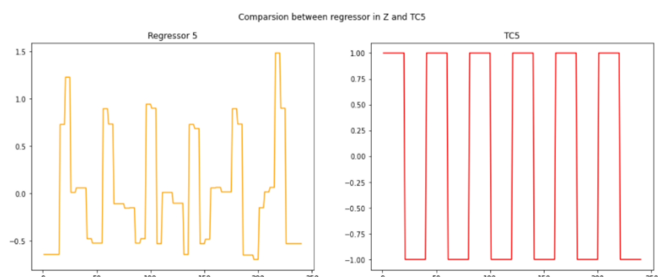
Eigenvalues = $[2.46092846, 1.1797447, 1.05321103, 0.83278942, 0.34463671, 0.15379429]$

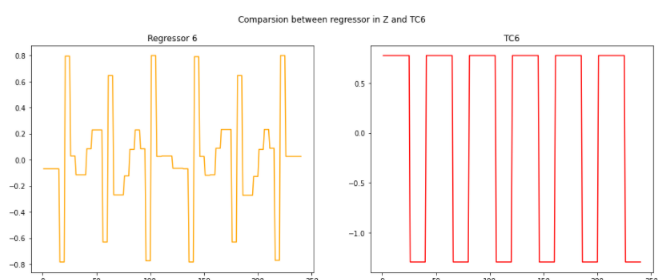

Comparsion between regressor in Z and TC1
Regressor 1
TC1


Comparsion between regressor in Z and TC2
Regressor 2
TC2

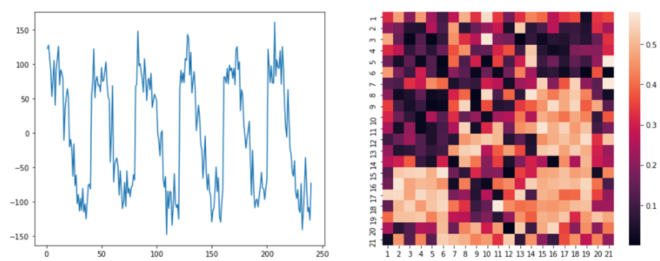

Comparsion between regressor in Z and TC3
Regressor 3
TC3


Comparsion between regressor in Z and TC4
Regressor 4
TC4

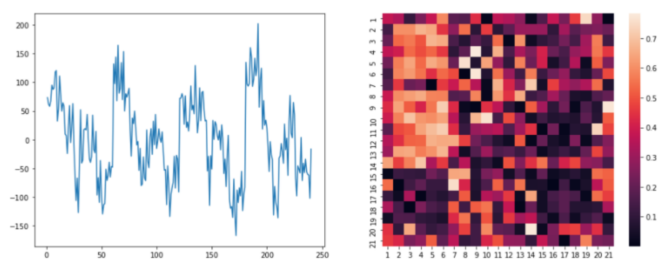

Comparsion between regressor in Z and TC5
Regressor 5
TC5


Comparsion between regressor in Z and TC6
Regressor 6
TC6

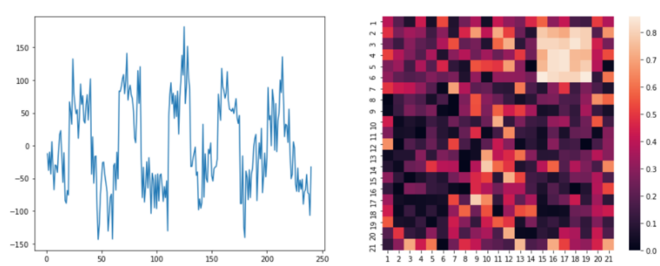
Time Courses and Spatial Maps retrieved from PCR SM1



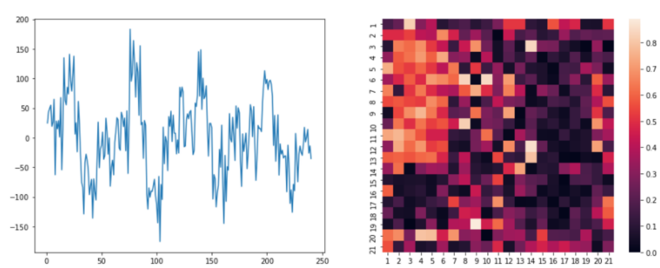
Time Courses and Spatial Maps retrieved from PCR SM2



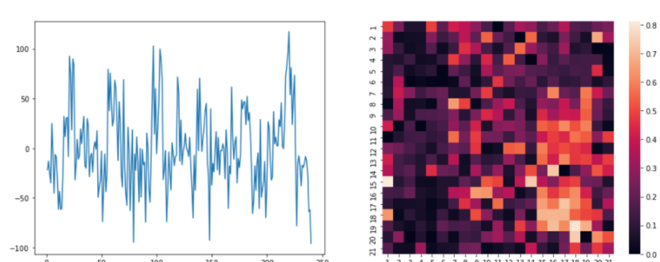
Time Courses and Spatial Maps retrieved from PCR SM3



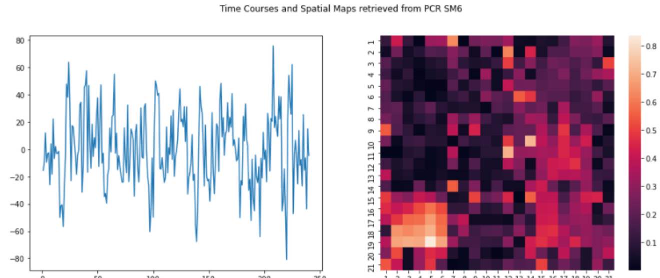
Time Courses and Spatial Maps retrieved from PCR SM4



Time Courses and Spatial Maps retrieved from PCR SM5



Time Courses and Spatial Maps retrieved from PCR SM6



The D_{PCR} and A_{PCR} are far worse than the estimates using RR and LR. It is also noted that the last three estimated time courses that have almost illegible shape compared of that of original TCs. One potential reason for this is that PCA are good at constructing the PCs that are orthogonal to each other which ensures the capture of the maximum variation in the dataset for different dimensions, but the last three time series, as seen from the previous question, are positively correlated, so the resultant principal components found by PCA might not be in a good quality to recover the shape of original temporal courses.