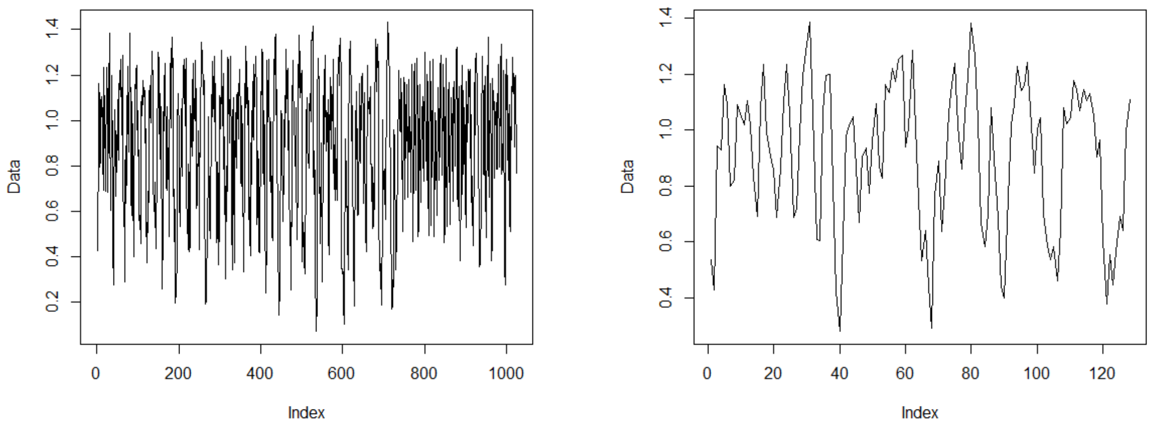
UNVEILING THE MYSTERY OF NON-LINEAR DYNAMICS

Candidate Number: 41524



(a) Overview of the Data

(b) Detailed Plot on the First 128 Points

Figure 1: Glance Look at the Data

	Value	5% CI	95% CI
Mean	0.8875	0.8729	0.9022
Standard Deviation	0.2915	0.2829	0.3002
Skewness	-0.4928	-0.5770	-0.4110
Kurtosis	2.3562	2.2185	2.5103

Notes: The CIs are calculated by bootstraping 3000 times.

Table 1: Basic Statistics of the Data

0.1 Overview of the Time Series Data

The time series data has 1024 data points, with 996 unique data. The quantisation rate is therefore 1.03. It looks stationary, without any linear or quadratic trend(Fig.1a). The maximum value of the data is 1.43, while the minimum value is 0.07. When zooming into the plot(Fig.1b), there are different patterns in the data. For the first 40 points, the data is oscillating very quickly. For the latter part, the data shows some periodicity. As is shown in Tab.1, the data is left-skewed but not heavy-tailed. Therefore, the data is unlikely to be financial data.

The delay plot(Fig.2) for delay $= 1$ shows linear relationship between the data and its lags. Such relationship becomes unclear as the delay increase(delay $= 3, 6$) until delay=9, when quadratic relationship appears in the delay plot. The data is not generated from a i.i.d process.

The auto correlation(ACF) and partial auto correlation(PACF) plots(Fig.3) confirm our findings above. The ACF plot exhibits a sine-like shape and it's significant up to lag=25. The PACF value is significant when lag $= 1, 8, 9, 10, 20$. These also convince that the data is far

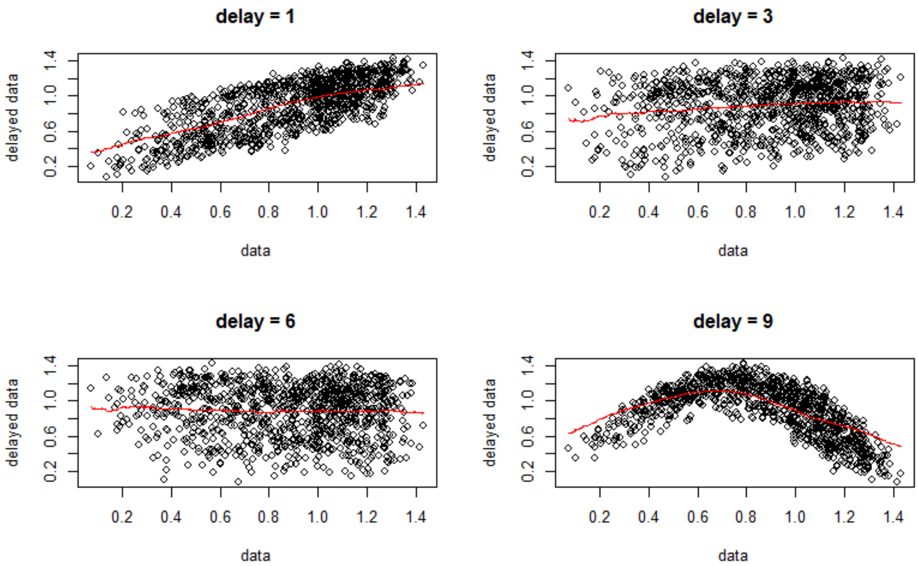


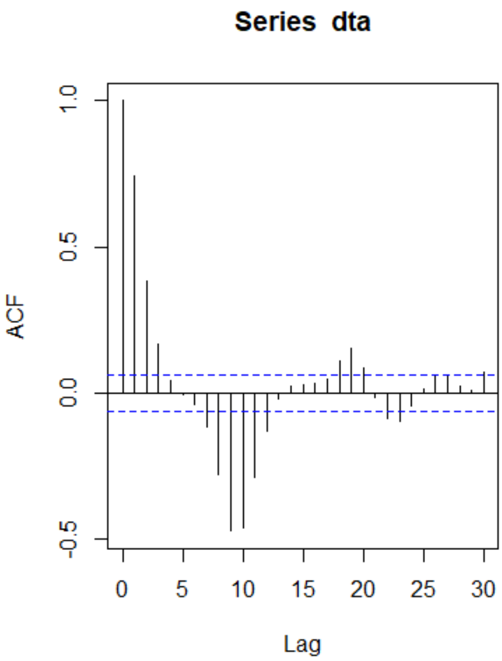
Figure 2: Delay Plots of the Data

from independent. The mutual information plot(Fig.4a) shows that a cycle of length 9 is likely to exist behind the data. Similar patterns appear in the state time separation plot(Fig.4b).

The PCA decomposition(Fig.5) gives hope to unveil such non-linear dynamics. The first principle component can explain for most of the deviation. Therefore, some linear transformation of lagged variable may be helpful in analysis.

The periodogram(Fig.6) also shows that there are some periods rooted inside the data we analyse. If the data is independent, we would observe spectrum density equal over all frequency. The spectrum is high when period equals 9(i.e. frequency equals 0.11). This corresponds to the above findings. What notable is that lower frequency even has higher spectrum density, which means that some longer period may exist behind the data.

To conclude, the data is not generated from any chaos free or i.i.d process. There are evidences that chaos as well as correlations exist in the data. The process is not likely to be one-dimensional. If so we would not observe linear correlation when lag equals 1 and quadratic correlation when lag equals 9. Therefore, the data is likely to be high-dimensional data's projection onto one-dimension map. According to Taken's Theorem, such chaotic but deterministic dynamic system can be reconstructed by delay embedding. In following section we propose two different models to reconstruct such dynamics.



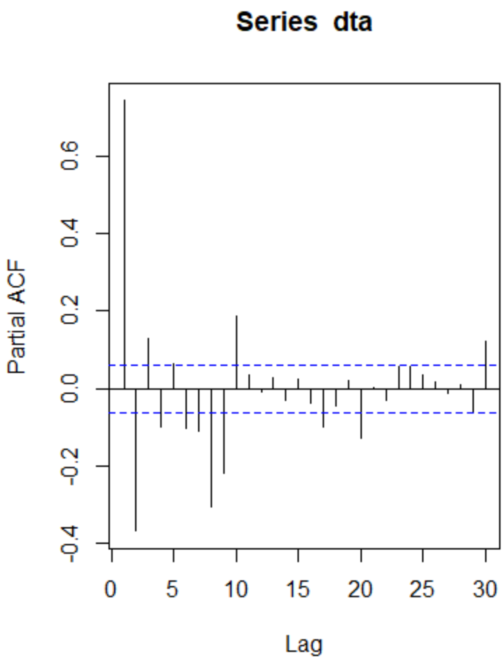
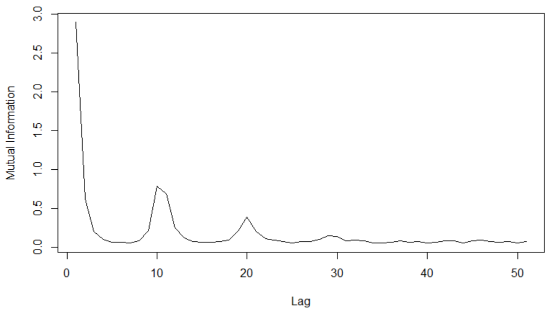
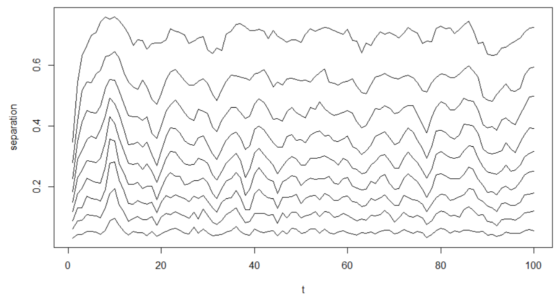


Figure 3: ACF and PACF of the Data



(a) Mutual Information Plot



(b) State Time Separation Plot

Figure 4: Mutual Information and State Time Separation Plot

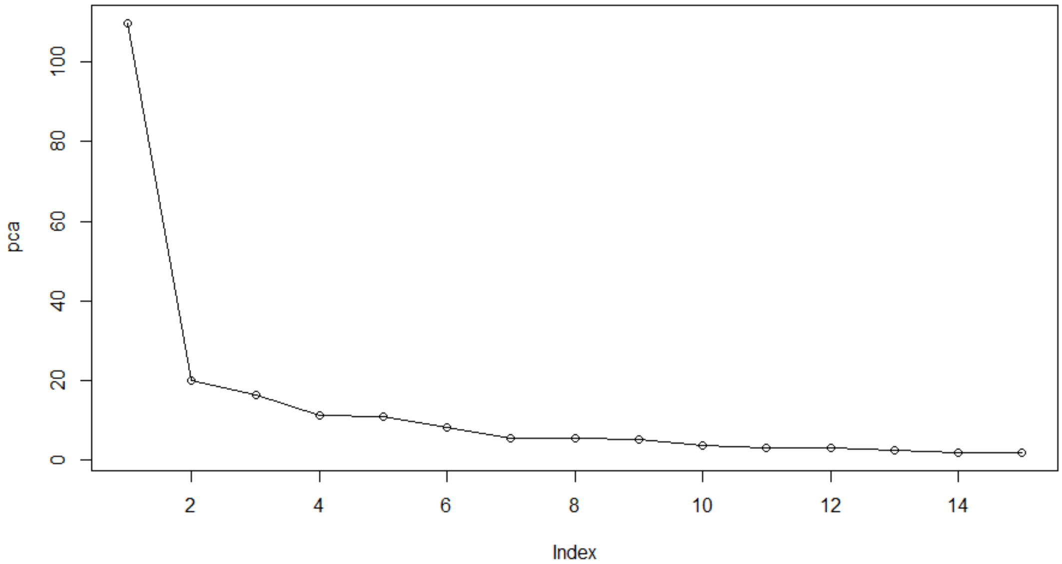


Figure 5: Principle Component Analysis

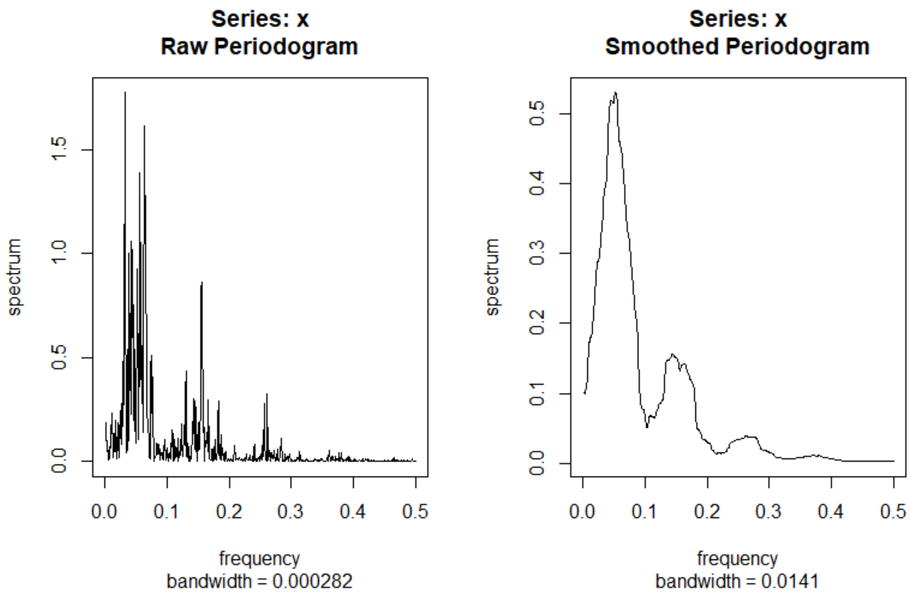


Figure 6: Periodogram of the Data

0.2 Model Fitting

0.2.1 Separation of Training Set and Test Set

Before fitting models to the data, we should separate the data into training set and test set so that we can evaluate the performance of models. We separate the data by half-half principle(i.e. the first half of data is training set, the second half is the test set).

0.2.2 Introduction and Simple Comparisons of Models

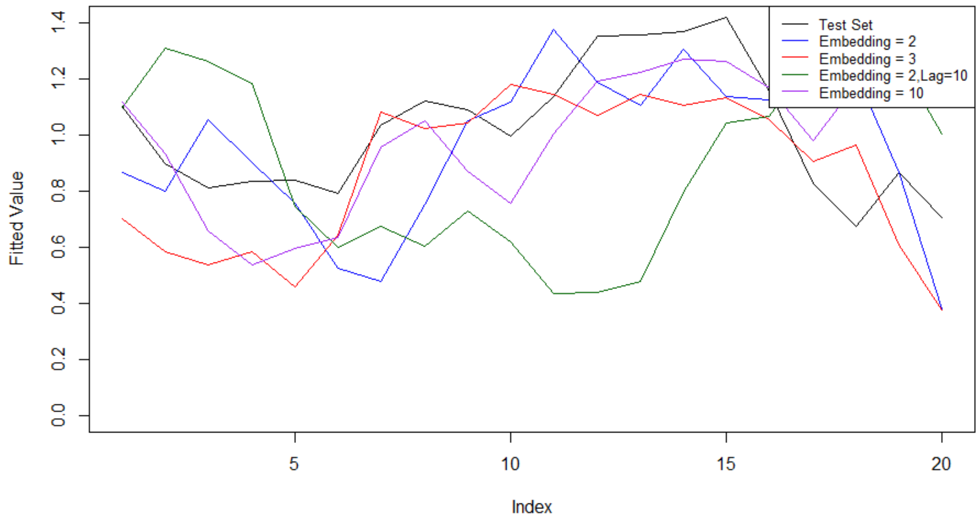
Local Analogue Model. Local analogue model can model non-linear relationships in the data. No assumption about how the data evolve over time is needed. The model is done by looking through its past and make prediction based on that. We try four different local analogue models, with different embedding dimensions and different lags, to model the training data. The estimation is simply a k step prediction up to 20. For the four models we fit, the one with embedding dimension equals 10 has the smallest RMS(Root Mean Square Error), 0.24. This is reasonable because only models with embedding dimension no smaller than 10 can pick up the linear and quadratic correlations between data and its lags(Fig.7a).

Local Linear Model. Local linear model is similar to local analogue model in a way that they both build up estimation on its nearest neighbors. Rather than simple take nearest neighbor's future as its estimate, local linear use linear regression to its neighbors, then calculate prediction on the linear fit. Similar model settings are used and the number of neighbours is simply 32. The model with embedding dimension equals 10 also stands out among all, with RMS 0.16(Fig.7b). In later analysis, only local analogue model and local linear model with embedding dimension equals 10 are used for analysis.

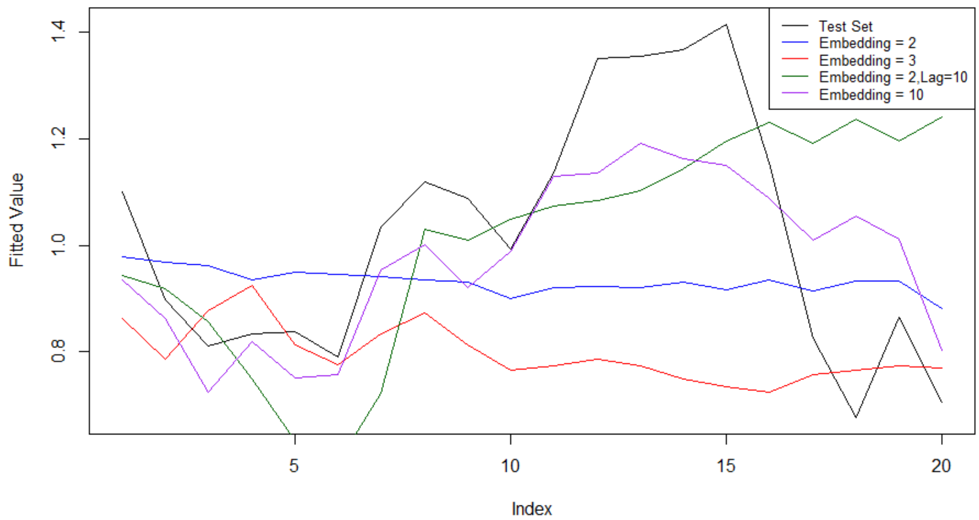
The reason why local methods are applied is that the data is likely generated from projection of high-dimensional data, as explained before. According to Taken's Theorem, such projection can be reconstructed by applying such local methods for delay embedding. Therefore, such local methods are better than any parametric methods(e.g. ARIMA model).

0.2.3 Determining Prediction Length

To determine to which lead time local methods are useful, we compare the performance of these methods with mean indicators. We vary the value of lead time from 1 to 40. For each lead time, we then loop through the test set and calculate k step predictions of each data point. Then RMS can be calculated by averaging the squared error in the test set. The number of neighbors in local linear model is 32. We then compare the RMS of models with that of mean estimators. If the models are useful, we would expect RMS smaller than that of mean estimator. If not, simple estimate by its mean is better. As we can see in Fig.8, the RMS of



(a) First 20 Estimations of Local Analogue



(b) First 20 Estimations of Local Linear

Figure 7: Estimation of Local Analogue and Local Linear

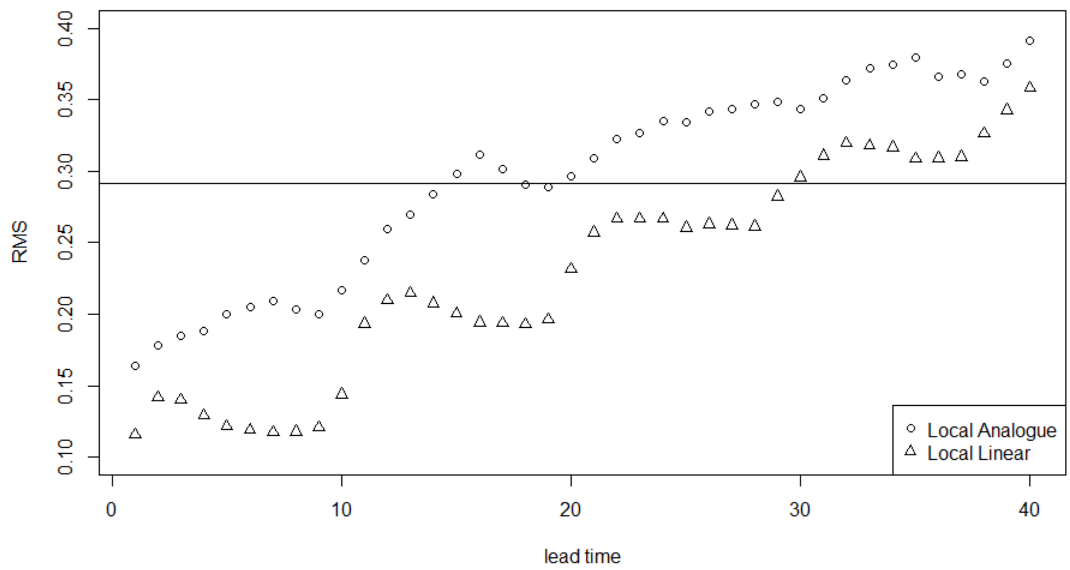


Figure 8: RMS Comparison

both models increase as the lead time getting larger. Overall, the two local models have lower RMS than mean predictor when lead time is smaller than 20, which means that for lead time larger than 20, the two models provide no better estimate than mean estimator. Therefore, in later analysis, we constraint lead time to 20.

0.3 Ensembles Making and Probabilistic Forecasting

0.3.1 Choosing Reasonable Scalar for Ensembles

In order to make probabilistic forecast, we should first estimate ensembles for each models. The ensembles of local analogue model is simply taken by future value of its n nearest neighbours(n is the number of ensembles). For local linear models, ensembles are obtained by changing the initial condition by small amount(e.g. add normal noise to initial condition). The scalar of such normal noise can greatly affect the quality of ensembles(i.e. $\Delta X = c\varepsilon$ where c is the scalar and ε follows i.i.d standard normal). Here we introduce talagrand diagram to measure the performance of ensembles. The scalar c is set to 0.03 and the lead time is 10(the choice of lead time is arbitrary, but different lead time will not affect talagrand diagram by much). As we can see in Fig.9(left local analogue, right local linear), the outcome fall uniformly in grids of ensembles. The upper one is estimated with ensemble size 32 and neighbour number 32. The lower one is with ensemble size 12 for local analogue, 28 for local linear as well as 16 for neighbours. The reasoned why we use such setting will be explained below.

0.3.2 Calculation of Ignorance Score

To calculate ignorance score, we first seperate the ensemble matrix also by half-half principle. The first half, the first half of outcome, is used for kernel density estimate. The bandwidth with highest probability is then chosen. Then estimate the second half of ensemble matrix and calculate ignorance score by $ign_i = -log_2(p_i)$. The ignorance score measures the probability of outcome in the estimated probability function. Then mean ignorance thus represents the performance of model fitted. The smaller mean ignorance is, the better model is capturing the dynamics.

0.3.3 Tuning Number of Neighbours

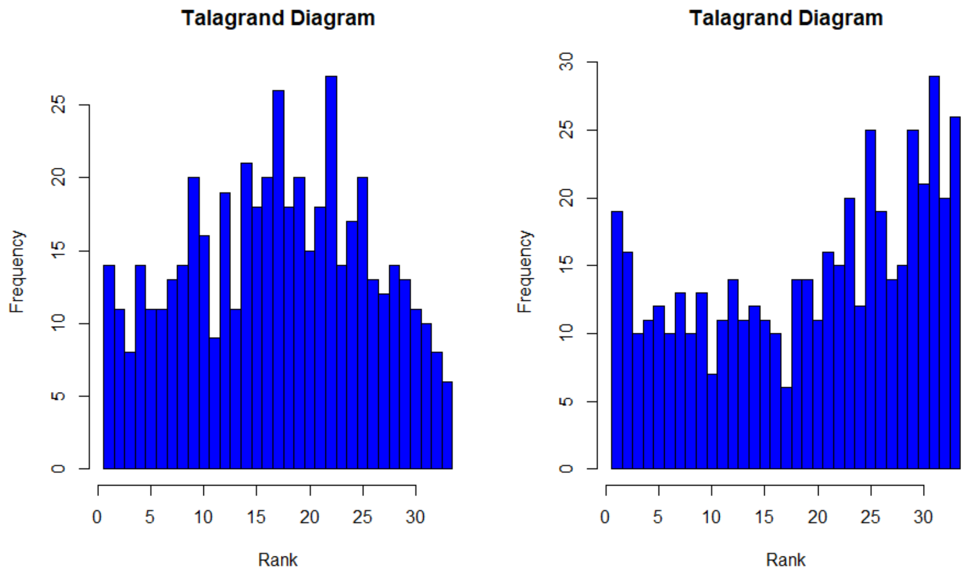
Changing number of neighbours of local linear model is a tradeoff between variance and bias. Too few neighbours has low bias but high variance. Too many neighbours, however, has low variance but high bias because the neighbours included in later may not be correlated. Therefore, we need to calculate the best number of neighbours that optimize probabilistic forecast. The models are fitted with ensemble dimension equals 32 and lead time equals 10. Mean ignorance is calculated as to determine best number of neighbours. In Fig.10, mean ignorance is minimized when number of neighbours is 16.

0.3.4 Tuning Number of Ensembles

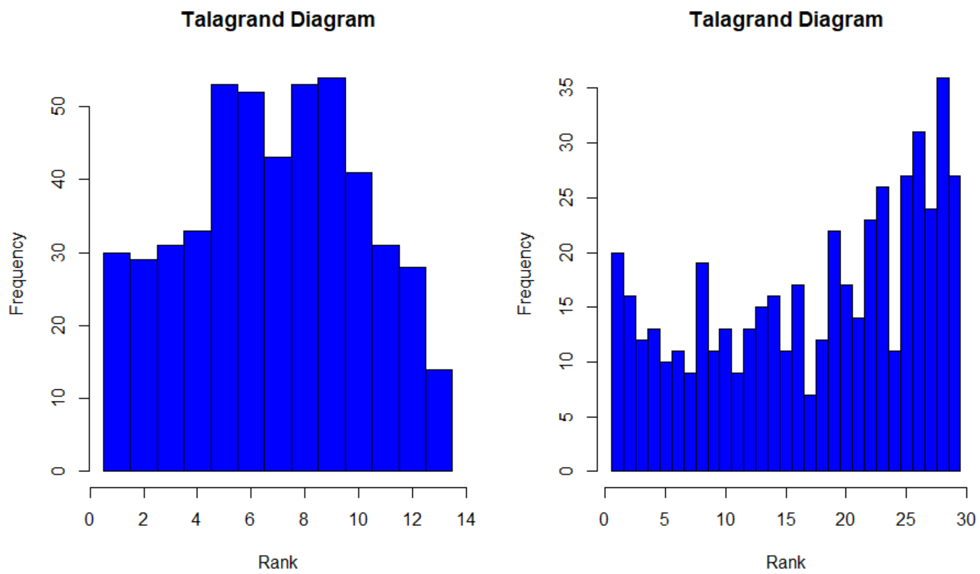
Different number of ensembles will also affect our probabilistic forecast of local analogue and local linear. This is also similar to variance-bias tradeoff. Few ensembles will be more of representatives of true underlying distribution, but will make kernel density estimate unstable, and vice versa. Here models with number of neighbours equals 16, lead time equals 10 and varying number of ensembles are fitted, from where mean ignorance is derived. As is shown in Fig.11, local analogue model reaches its minimum when ensemble size equals 12, local linear model reach its minimum with ensemble size 28.

0.3.5 Prediction Performance for Different Lead Time

After tuning parameter described above, we thus obtain the best models for forecasting. The best local analogue model has ensemble size 12. The best local linear model has ensemble size 28, number of neighbour 16 and scalar $c = 0.03$. Then we can compare the performance of these two models in sense of probabilistic forecast(i.e. mean ignorance) rather than deterministic forecast(i.e. RMS). The result is shown in Fig.12. Local linear model always perform better



(a) Talagrand Diagram of Default Model Setting



(b) Talagrand Diagram of Final Model Setting

Figure 9: Talagrand Diagram of Local Analogue and Local Linear

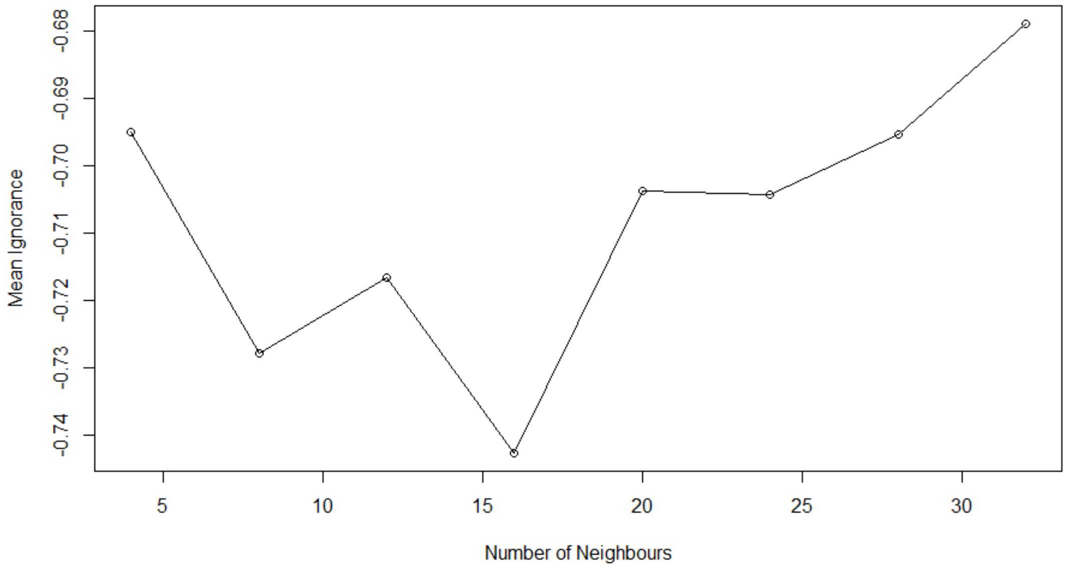


Figure 10: Mean Ignorance with Different Number of Neighbours

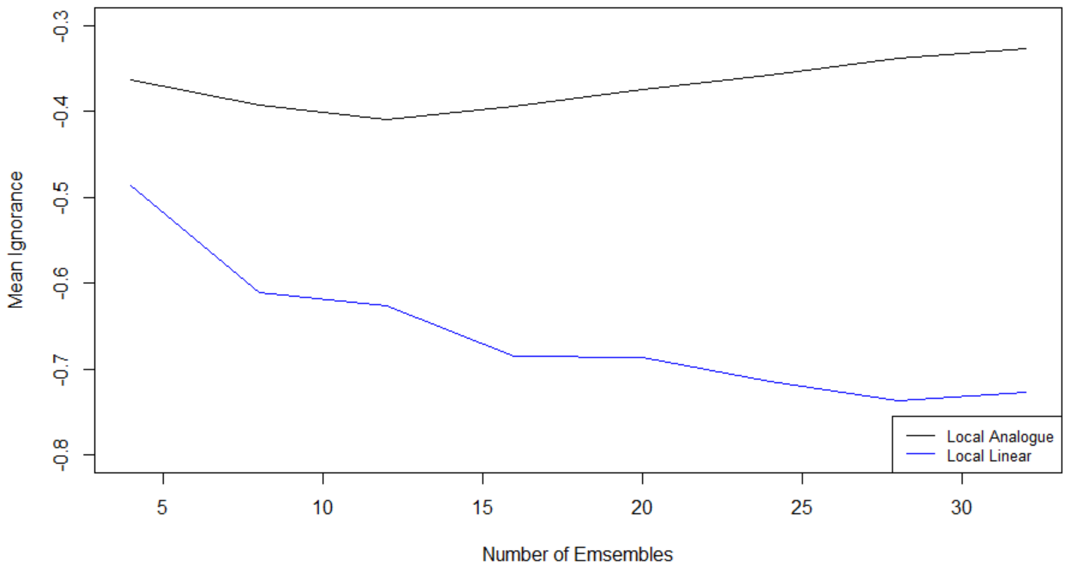


Figure 11: Mean Ignorance with Different Number of Ensembles

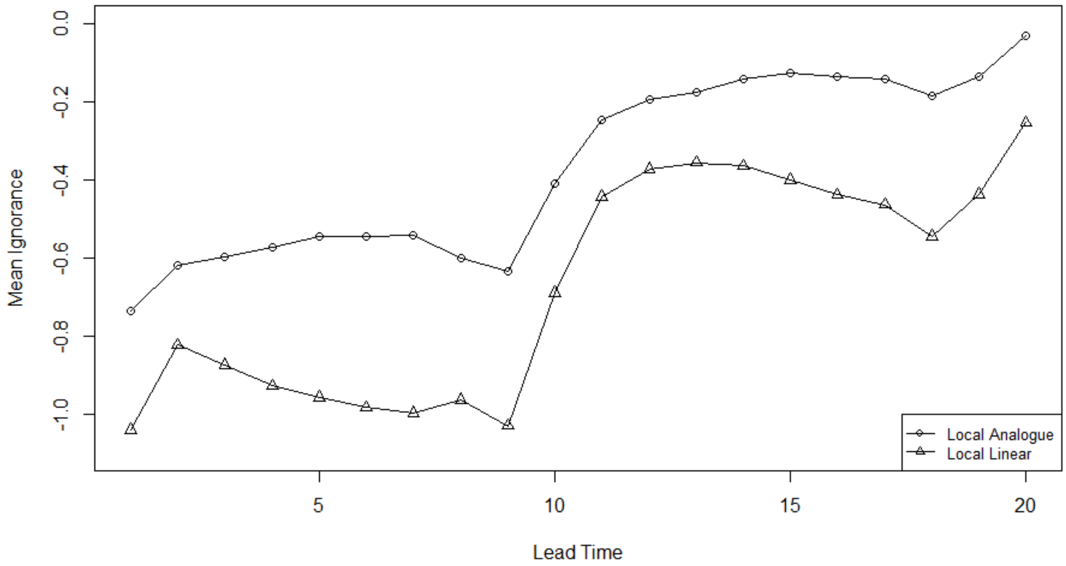


Figure 12: Mean Ignorance with Different Lead Time

than local analogue model when lead time is smaller than 20. The forecast performances of two models are stable when lead time smaller or equal to 9. When lead time exceed 9, the mean ignorance score increase sharply. And similarly we can observe a sharp increase of ignorance score when lead time exceed 9, this may correspond to our finding that there may exist a period of length 9.

0.3.6 Prediction Performance for Different Initial Conditions

The fact that linear model outperforms local analogue model does not necessarily means that local linear model is the preferred model in all circumstances. With our final parametric setting, when lead time is 10, there are 74 out of 246 data(test set of ensembles) for which local analogue model has lower ignorance score than local linear model. We then differentiate the ignorance score of the two models and filter the 20 points that have most extreme difference of each side. Then we mark these data in delay plot to determine which region does local analogue model performs better, and vice versa. The result is shown in Fig.13. The big filled point in the figure is the gravity center of the most extreme data. As is shown in the plot, when delay equals 1, the position is not that informative, with the two centers lie close to each other. However, what notable is that when delay equals 9, the local analogue seems to outperform in the right of delay plot. Which means that if have a data point to estimate its 10 step forecast, if the former 9th data is larger, then local analogue is better way to characterized such dynamics. When the former 9th data is smaller, then local linear is better than local analogue. Therefore, we can combine the two models to make forecast rather simply taking local linear model as out rule-of-thumb.

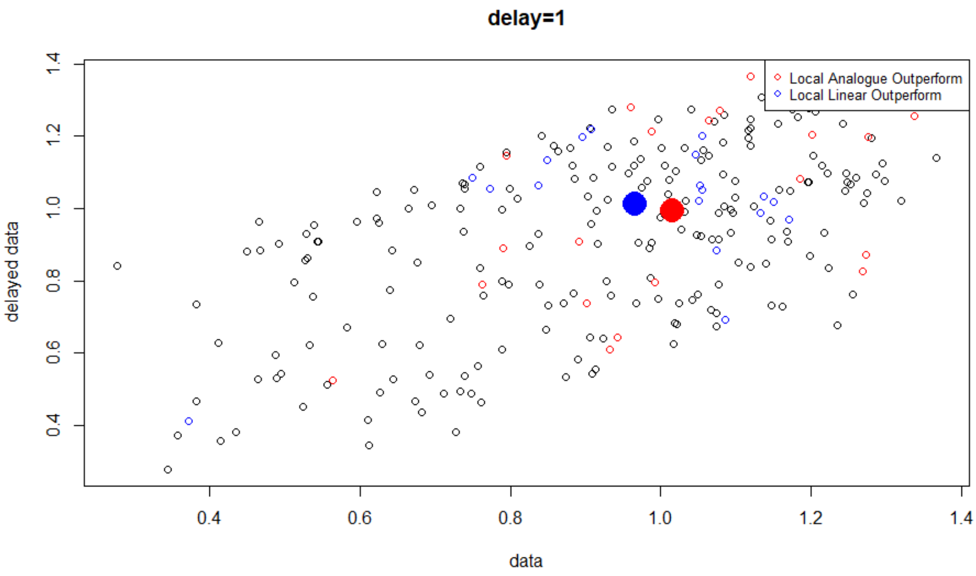
0.4 Summary

The data we use in this project is likely derived from projection of high-dimensional data to one-dimension, rather than from any stochastic, linear or low-dimensional data. A period of length 9 is outstanding no matter in delay plot, mutual information, state time separation plot or periodogram. Though there may be some longer period indicated by periodogram, we don't investigate on such longer periods.

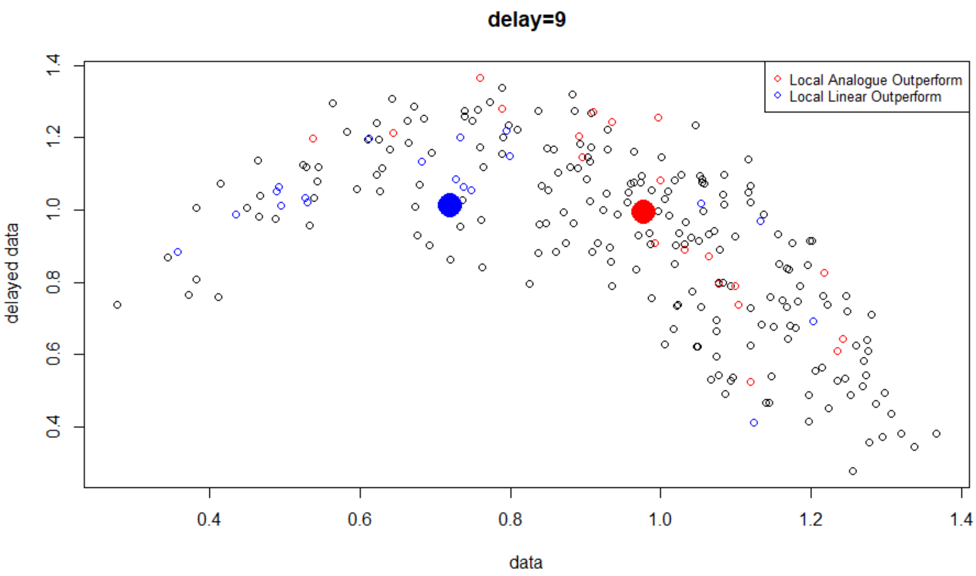
The reconstruction of high-dimensional chaotic system can be done by delay embedding, according to Taken's Theorem. Further we use local analogue model as well as local linear model to make forecast. As the deterministic forecast(i.e. RMS) is not stable for all time, we turn to making ensembles and comparing their ignorance score. Parameter can be tuned with mean ignorance score. The best parameter is thus obtained by minimizing the mean ignorance score.

For different lead time, local linear model always outperform local analogue model in term of mean ignorance score. However, this is not always the case if we don't simply average ignorance scores of all outcomes. This turns out that local analogue model can outperform local linear model in some specific region of initial condition. It is thus suggested that combination of these two models is needed to improve the forecast.

If more computing power is available, we can then tune parameter with smaller steps and get more precise best parameter. Models with more embedding dimensions, more neighbours are thus allowed for estimation. Another possible approach is to reverse the data and implement similar approach to see if these conclusions still hold in this circumstances. Last but not least, we can further what we have done in section 3.6 and change the value of lead time. This will suggest for each lead time, what combination of two models are most appropriate, which is very helpful for improving our forecast.



(a) Delay Plot with Delay = 1



(b) Delay Plot with Delay = 9

Figure 13: Talagrand Diagram of Local Analogue and Local Linear