Estimation of the size of the Corona Virus in Italy in the next ten days made by nonparametric time series prediction.

Sergio Conte[1], Ferda Kaleagasioglu[1,2], Alberto Foletti[3], Elio Conte[1].

[1]School of Advanced International Studies for Applied Theoretical and Non Linear Methodologies of Physics, Bari, Italy.

[2] Near East University , Medical Pharmacology.Turkey.

[3] Clinical Biophysics International Research Group, Lugano, Switzerland

Abstract : By using the methods of the non parametric time series prediction we obtain that in the next ten days we will reach the size of 8000-9000 contagions in Italy.

Introduction

In the note, the technique of nonparametric time series prediction is used for the estimation of the behavior of the coronavirus epidemic in Italy in the next ten day. The current prediction effected by such method is that the size of the epidemic will be about 8000-9000 cases of contagions in theoretical line, depending instead the actual size of the process from an increase o decrease of the prevention measures that are in act.

Materials and Methods

Obviously the great limit of the present work is in the restricted number of experimental data that we have at disposal. In fact, they are limited to only 11 days from February 20 to March 01.

The final objective of many time series analyses is prediction. It is often of interest to study the conditional means, conditional variances or complete conditional densities in some period, given the past of the process. When a point prediction is the final objective, an estimate of some conditional mean may be desired, while the conditional variances are needed if interval forecasts or assessments of future volatility are desired. In order to analyze the conditional mean non parametrically one may, for instance, start from a W. Hardle, H. Lutkephol and R. Chen model of the form X(t) = f(X(t-1), X(t-2), ...) + e(t) , where e(t), is a series of innovations which is independent of past X,. In this case f(.) represents the conditional expectation in period t, given past observations X(t-1), X(t-2), ... and it is the minimum mean squared error (MSE) 1-step predictor for X(t). In parametric time series analysis the function f(.) is chosen from some parametric class so that the specific candidate is obtained by specifying a fixed finite number of parameters. Nonparametric approaches on the other hand allow f(.) to be from some flexible class of functions and they approximate f(.) in such a way that the approximation precision increases with the sample size. For this purpose several different techniques and procedures are available. For instance, local approaches approximate f(.) in the neighborhood of any given argument by letting the neighborhood decrease and thereby increase the approximation precision with growing sample size. For this purpose the number of lagged X, used in the model, is usually limited. This is as a general tool. In particular, nonparametric modeling is used when we cannot make any assumptions about the functional form of the process that generated the observable time series. Instead of assuming a specific model and computing its coefficients, we derive the model from given data directly. This is done using *local models*. Instead of fitting one complex model with many coefficients to the entire data set, we fit many simple models to small portions of the data set. In effect, we use a model that changes its parameters adaptively depending on the geometry of the local neighborhood of the dynamical system.

The general methodology proceeds as it follows.

To predict point x_{n+1}, we determine the last known state of the system by vector **X** = $[x_n, x_{n-t}, x_{n-2t}, x_{n-(d-1)t}]$, where d is the embedding dimension and t is the time delay. We then search the time series to find k similar states that have occurred in the past. The similar states are determined by evaluating the distance between vector **X** and its neighbor vector **X'** in the d-dimensional state space. The idea is that if the observable signal was generated by some deterministic map $M(x_n, x_{n-t}, x_{n-2t}, x_{n-(d-1)t}) = x_{n+1}$, that map can be reconstructed from the data by looking at its behavior in the neighborhood of **X**. We find the approximation of M by fitting a (low order) polynomial which maps k nearest neighbors (similar states) of **X** onto their next immediate values. This map is used to predict x_{n+1}.

We used the software VRA of prof. Eugene Kononov , where one can construct such a model from a range of classes (nearest neighbor, locally constant, kernel regression, locally linear, locally weighted linear, and radial basis models).

To construct the actual model to generate predictions, we need to choose certain parameters (the model is nonparametric only in a sense of its global functional form), such as embedding dimension, time delay, the predictor.

Results

The obtained results are reported in Figures 1, 2, 3 and 4.

In Figure 1 and 2 we report the results for phase space reconstruction with the average Mutual Information and the use of False Nearest Neighbors.

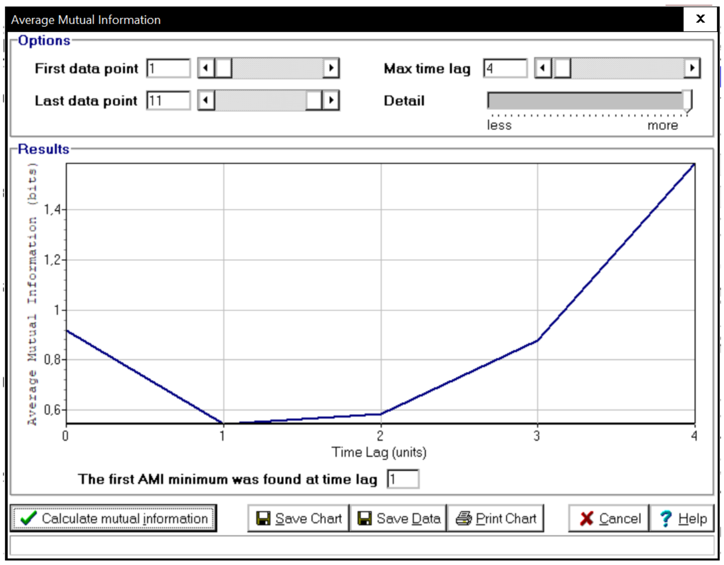

Fig. 1

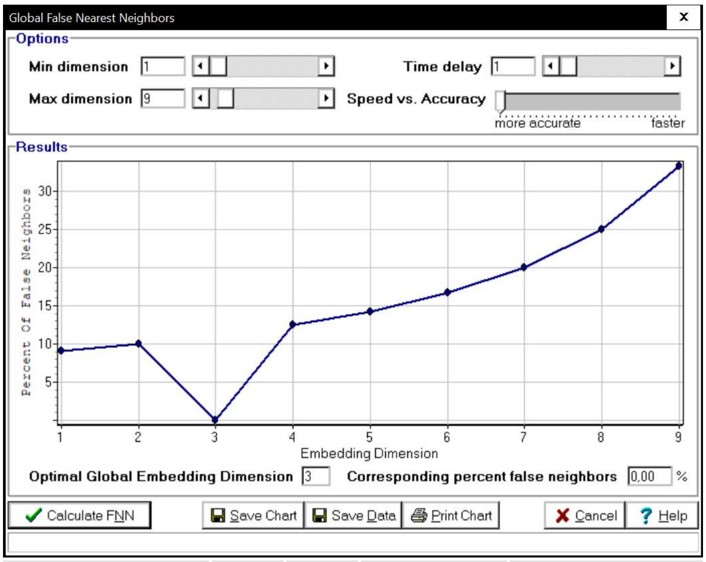

Fig. 2

In figures 3 and 4 we give the results of our analysis with the used parameters that are all specified in the figures.

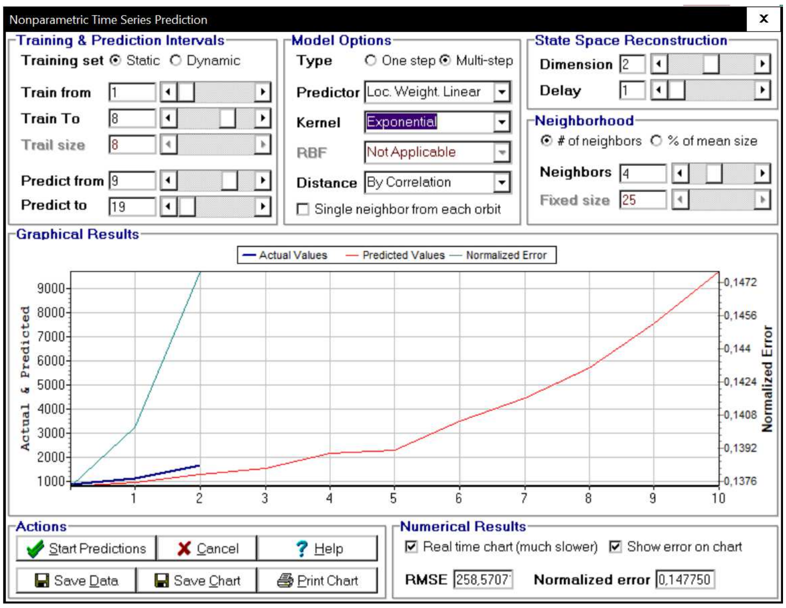

Fig. 3 Prediction with Loc. Weight Linear and Exponential Kernel

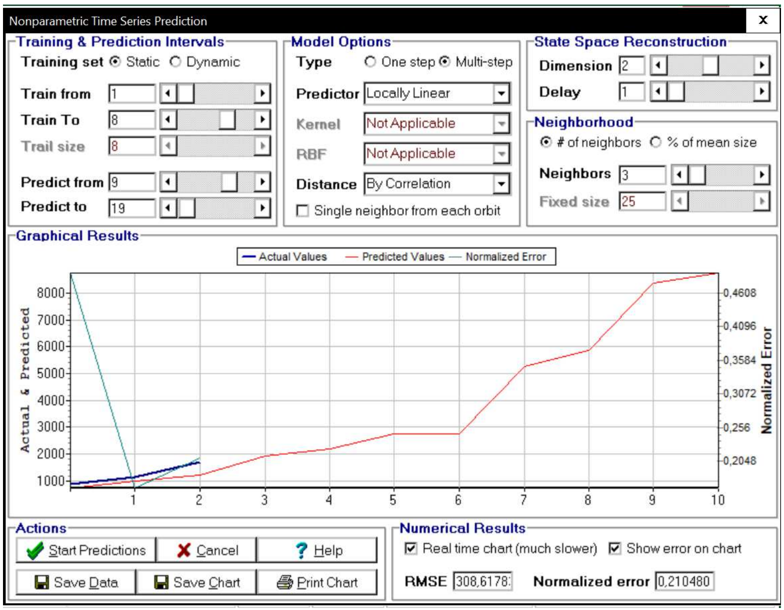

Fig. 4 Prediction with Locally Linear model.

References

Wolfgang Hardlel, Helmut Liitkepohl, Rong Chen, A review of non Parametric Time Series Analysis, International Statistical Review, 1997, 65, 49-72.