Time Series Forecasting using Wavelets with Predictor-Corrector Boundary Treatment

Peng Yu Anna Goldenberg Zhiqiang Bi

Center for Automated Learning and Discovery
Carnegie Mellon University
Pittsburgh PA, 15213

August 20, 2001

Abstract

Time series analysis has attracted a lot of attention due to its applicability in a wide range of fields. Various methods, such as linear and nonlinear models, and wavelets have been used for time-series analysis and predictions. We propose a new methodology that can be integrated with existing techniques to iteratively achieve better accuracy. We decompose the original signal into several components using stationary wavelet transform and forecast each component with a linear model. Predictions from all components are summed up to give the overall prediction. Our method, the predictor-corrector boundary treatment, is a new heuristic approach to solving the boundary problem. Rather than using an *adhoc* condition, we use the predicted values from the previous step for the boundary of the wavelet transform and then iterate. Preliminary experiments have shown that the algorithm improves its predictions with every iteration and is a promising methodology for time series analysis.

1 Introduction

In many fields of natural and social sciences, people are interested in predicting the future from the observed phenomena. More specifically, if we observe a sequence of n events, $y_1, \ldots, y_n$ at past times $t_1, \ldots, t_n$, how can we predict the events $y_{n+1}, \ldots, y_{n+m}$ that will happen at future times $t_{n+1}, \ldots, t_{n+m}$? Examples include studying the origin and the future evolution of the universe, forecasting the weather, predicting the stock price or economical growth, managing the real (and virtual) traffic, predicting the sales of various products in a supermarket, or even continuing a great composer's unfinished work. These problems can all be classified as time series analysis (TSA).

In this paper, we introduce a new method to do time series prediction and point out its application to real world problems. Our approach combines the state-of-the-art signal processing technique, the wavelet transformation, and linear models to perform the time series prediction. In wavelet transform, we apply a predictor-corrector method to treat the boundary condition, which will be described in detail in Section 3. Our algorithm has the following characteristics:

- Wavelet transform is used to decompose the original signal into several components in multiple scales. Because wavelets are able to represent a finite time sequence in multiple scales and wavelet transform is a non-parametric method, i.e. very few *adhoc* assumptions are made about the original signal, wavelets have proven to be a useful tool in time series analysis.

- A more careful treatment of boundary condition for wavelet transform is applied.

We believe the boundary condition treatment is important in time series analysis and prediction. By incorporating our innovative way of treating the boundary, we expect to improve the accuracy of the existing forecasting systems.

2 Prior Work

2.1 Linear Time Series Models

Time series are usually modeled as stochastic processes. Every observation in a stochastic process is a random variable, so a stochastic process can be defined as a collection of random variables ordered in time. The classical techniques for time series analysis are the linear time series models. A typical linear model is an autoregressive moving average model (ARMA)[1]. For an ARMA model of order p and q, or $ARMA(p, q)$, the event at time t can be expressed as

$$x_t = \sum_{p=1}^{M} a_p x_{t-p} + \sum_{q=0}^{N} b_q e_{t-q} \tag{1}$$

where e_t represents the noise and is a sequence of independent random variables drawn from a distribution with zero mean and constant variance. The coefficients a_p, b_q can be estimated using maximum likelihood estimation. In our testing algorithm, we use the ARMA model with q being 0, which is called the AR model.

In order to capture long term trend of a time series, a technique called "integration" [2] is also used. First, the original time series is transformed into its dth-order differences:

$$\begin{aligned} y_t^{(0)} &= x_t \\ y_t^{(d+1)} &= y_{t+1}^{(d)} - y_t^{(d)} \end{aligned}$$

Analysis and prediction can be done on dth order difference of the original signal, then they can be "integrated" d times to recover the prediction of the original signal. A model combining dth-order integration to the ARMA(m, n) is called ARIMA(m, d, n) model.

These linear models are very easy to understand and straightforward to implement. However, because they are over-simplified and have very limited expressing power, they can fail for even moderately nonlinear systems. Some examples of nonlinear systems for which the linear models fail are given in [3]. Therefore, for more accurate predictions, more complicated nonlinear models, such as neural networks, should be investigated.

2.2 Decomposition Methods

Linear and nonlinear techniques can be applied to the signal in time domain or in frequency domain. The basic idea is to first decompose the original signal into components and then to apply selected predicting method(s) to these individual components. High frequency components can be used to predict the near future while low frequency components can usually tell the long term trend. When they are treated separately, better results are expected. Fourier Transformation (FT) is usually a natural choice to decompose the signal into components. However, recent studies have shown [7] that they could be outperformed by another kind of transform, wavelet transform, because its intrinsic multi-resolution feature can automatically adjust the window size to resolve local information. Due to this nice property, wavelets have been successfully utilized in time series analysis. General reviews can be found in, Morettin[10],[11], Priestley[12] or Percival and Walden[13].

2.3 Stationary Wavelet Transform

When it comes to implementing the wavelet transform, a low-pass filter $\mathcal{H}$ and a high-pass filter $\mathcal{G}$ are needed. They are applied to the digital signal to yield a smooth and a detailed versions of the signal respectively. In case of the classical wavelet transform, a decimation is carried out so that only half of the coefficients of the detailed component are left at the current level and half of the coefficients of the smooth version are recursively processed using $\mathcal{H}$ and $\mathcal{G}$ for coarser resolution levels. Due to the decimation, the number of classical wavelet coefficients is halved with each move to a coarser level. The consequence is that the coarser level we are at, the less information is available to train the forecasting model, which may lead to the overall predicting inaccuracy. To overcome this problem, we are exploiting the stationary(redundant) wavelet transform. Below, we provide the definition of the stationary wavelet transform as given in [14].

The basic idea of the stationary wavelet transform is to fill in the gaps caused by decimation using redundant information obtained from the original signal. Assuming the original signal is $x(t)$ where $t = 1, 2, ..., n$, we define the smooth versions of $x(t)$ at different scales as

$$\begin{aligned} c_0(t) &= x(t) \\ c_i(t) &= \sum_{l=-\infty}^{\infty} h(l) c_{i-1}(t + 2^{i-1} l) \end{aligned} \tag{2}$$

where i goes from 1 to p and h is a low-pass filter with compact support, such as the B_3 spline, defined as $(1/16, 1/4, 3/8, 1/4, 1/16)$. Notice here the difference from the classical Discrete Wavelet Transform (DWT) is that the decimation is left out, which renders the components at different scales to be the same length. We define the component of $x(t)$ at level i as

$$d_i(t) = c_{i-1}(t) - c_i(t)$$

The set $\{d_1, d_2, ..., d_p, c_p\}$ represents the wavelet transform of the data up to the resolution level p. The term c_p is the residual. The inverse transform is given by

$$x(t) = c_p(t) + \sum_{i=1}^{p} d_i(t)$$

Some work has been done in direction of combining wavelets and predictive tools. For example, Aussem and Murtagh[14] have combined wavelets and neural networks. They decomposed the times series into components and let each component grow with a neural network, then combined predictions together to obtain an overall forecast. However, they have arbitrarily selected their boundary conditions. In this paper we propose a new method for treatment of the boundary condition. The boundary is proposed to be selected according to prediction made on the original signal by a predictive model, for example, an AR model, and then the decomposition-prediction steps are iterated. This methodology allows the use of a simple linear model rather than a black box technique, such as neural network, while achieving iteratively improved accuracy and is described in more detail in the next section.

3 Predictor-Corrector time series prediction using stationary wavelet transform and linear models

3.1 Boundary Condition

To understand the significance of the proposed algorithm, it is important to introduce the problem that arises at the boundary. Assuming that we use the B_3 spline filter, computing the last element of c_1 according to Eq. (2) requires the knowledge of "future" values $x(n+1)$, $x(n+2)$. Similarly, computing $c_2(n)$ requires $x(n+1)$, $x(n+2)$, ..., $x(n+6)$, and computing $c_i(n)$ would require $x(t)$ where $t = n+1, n+2, ..., n+(2^{i+1}-2)$. Since all these x values are in the "future", we have to provide some initial guesses for them. Typical boundary treatments are periodic, where $x(n+k) := x(k)$, and reflective treatment, where $x(n+k) := x(n-k+1)$, $k = 1, \ldots, m$ and m is the number of points to be predicted.

The boundary treatment plays a significant role in time series forecasting problems for the following reasons. First of all, the wavelet decomposition improves the forecasting quality by splitting the original signal into isolate components at different scales. The more "regular" is the scale that each component represents, the better the resulting prediction is. However, using *ad hoc* boundary condition such as periodic or reflective extensions does not guarantee that the tail part of each component really represents information with the same trend as represented by the rest of this component. Since forecasting extends the time series from the tail, boundary handling has significant influence on the continuation of the series. Based on this reasoning, we believe that a more careful handling of the boundary is necessary.

3.2 Methodology

We propose a Time Series Analysis with Wavelet Decomposition Using Predictor-Corrector Boundary Treatment — a combination of TSA, Wavelet Transform, Linear Models, and Predictor-Corrector treatment of boundary conditions. We use the idea of predictor-corrector approach from numerical analysis. Suppose we have time series, $y_1, \ldots, y_N$, and we want to predict $y_{N+1}, \ldots, y_{N+M}$. Our algorithm works as follows:

1. make initial prediction using linear models on the original time series and use the predicted values as the boundary condition for the wavelet decomposition.

2. use linear method to predict each component from y_{N+1} to y_{N+M} and sum up the components to obtain the overall prediction.

3. decompose $y_1, \ldots, y_N$ using wavelet transform again, but this time boundary is treated with predicted values obtained in step (2).

4. Iterate step (2) and (3) until pre-defined converging criterion is met, for example the criterion could be minimization of Root-Mean-Squared error (RMS).

To test goodness of fit, we can first split observed data sequence into two parts, one for training and one for testing. We then apply a prediction algorithm to grow the series to the same length as the testing sequence. A distance, L^2 for example, between the testing sequence and the predicted sequence is computed and can be taken as a score for testing the goodness of fit. Lower score represents better fit.

Intuitively, the end of the observed signal plays an important role in predicting the future, thus it has to be treated carefully. Our iterative scheme removes the arbitrariness of the standard periodic or reflective boundary condition in wavelet transform. It is reasonable to believe better results can be obtained.

4 Experiments

4.1 Datasets

To emphasize the importance of the project we would like to relate it to the real world data. One of the datasets that has become available relatively recently is the retail data. There is evidence of several clearly expressed periodicities, such as yearly, monthly, daily, etc. trends in the data. However, due to the high level of detail, the data still appears to be hard to predict for the standard Time Series prediction tools. We would specifically like to concentrate on prediction of the over-the-counter medication purchases. The results of such predictions can further be used in attempt to solve one of the most vital problems — the early detection of epidemics and bio-terrorism attacks. It has recently been proposed to use non-symptom specific data such as over-the-counter medication purchases, hits to medical websites, etc. for early detection systems. All previous research has concentrated on evaluating medical data obtained from emergency rooms and lab tests. Majority of the population however does not go to stores as soon as they feel sick, rather they go to stores to buy cold/flu medications or search the web for symptoms and early treatments. Preliminary research has shown that there is evidence to believe that the retail data gives an early indication of flu. It is easy to see the importance of accurate predictions of the medication purchases. If there is a belief that the predicted data has high accuracy then, when comparing it to the real-time data, any significant difference can be treated as a signal of abnormality with high confidence.

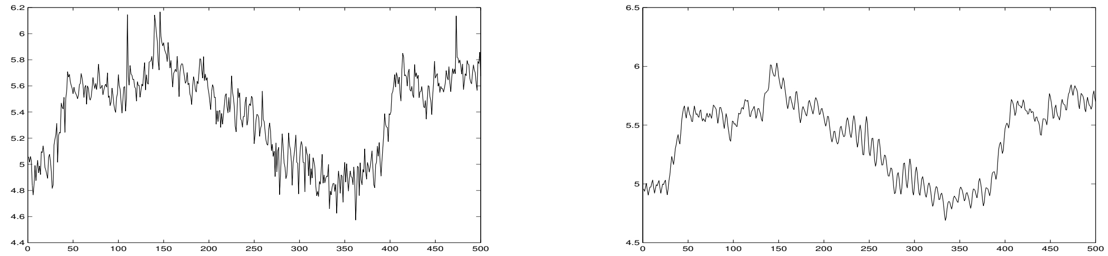
Specifically, we use 500 data points of cough/syrup subgroup of over-the-counter medications normalized sales counts and predict the next 30 days of sales.

Another dataset that we apply our predictor-corrector boundary treatment method to is the classical sunspot dataset which is often used as a benchmark in the forecasting literature (Tong [6], Weigend *et al* [4], etc). Sunspots appear as dark spots on the surface of the Sun. They typically last for several days, although very large ones may live for several weeks. They can be related to other solar activities such as magnetic field cycles, that in turn could influence the metereological conditions on Earth. The dataset exhibits periodicities but they are not regular enough to establish a clear pattern, thus it is beneficial to apply predictor-corrector method in order to achieve improvement on the accuracy of existing predictions. We use range-normalized yearly averages of the sunspot data from 1720 to 1984. Thirty data points from 1955 to 1984 were withheld for validation purposes.

4.2 Over-the-counter-medication data

We applied the proposed predictor-corrector scheme to a denoised version of an over-the-counter-medication data set. The denoised version is obtained by performing discrete cosine transform on the original data and filtering out the coefficients less than a predefined threshold 0.2. The original data and its denoised version are shown in figure 1. We chose the denoised version for time series forecasting because, on one hand, the denoised version is more predictable than the original noisy data so that the effect of the proposed predictor-corrector scheme could be clearly seen, and on the other hand, from practical point of view: we would like our predicting model to generate the deterministic pattern rather than noise.

The denoised series contains 500 values, from which we withheld the last 30 values for validation purposes and used the first 470 values to train linear autoregression models. We chose RMS as the



Original data Denoised data

Figure 1: Over-the-counter medication datasets

measurement of goodness of fit, i.e.

$$Error := \frac{\sqrt{\sum_{i=471}^{500}(predicted_i - original_i)^2}}{30}$$

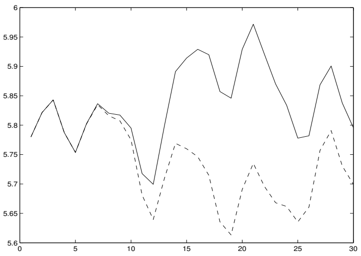
In order to see the effect of predictor-corrector scheme separate from other potential disturbance, we decided to use only AR models to do prediction on all wavelet components instead of adopting the hybrid method suggested in [14]. AR models were shown to suffice in this task partly because the denoised dataset is rather smooth and all its wavelet components are thus smooth as well. However, we think that for a more oscillating dataset ANN or FIR neural networks would be more appropriate. The procedure with which we carried out the predictor-corrector scheme is described below in detail.

As a starting point, we performed AR prediction on the denoised data. The order of AR model was chosen to be 25. The model was trained on the first 470 values of the denoised data and it produced 30 more values corresponding to data points from 471 to 500. We called this the 0th iteration, the one that is carried out before any wavelet decomposition is applied. The 0th iteration provides the initial boundary condition for the iterations that follow.

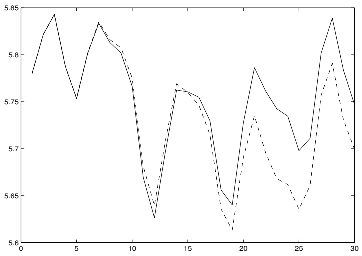
At the first iteration, the redundant wavelet transform was applied to yield 5 wavelet components of length 470, the last one being the residual. The values predicted in the 0th iteration were used as the boundary condition. Then, each wavelet component was passed to the AR model to make predictions for the next 30 values. The orders of the five AR models applied to the five wavelet components are respectively 25, 25, 25, 25 and 37. At the end of the first iteration, the predictions on the five wavelet components were summed up, to result in the overall prediction corresponding to the normalized over-the-counter medication counts $471 - 500$ after the first iteration.

The procedure was then applied recursively, the boundary was chosen to be the one obtained in the previous step, until convergence of the RMS was achieved. The overall predictions after several iterations are plotted against the denoised data in figure 2 where the solid line represents the prediction and dotted line represents the denoised original data. Only the predicted segment from 471 to 500 is shown. The error has converged approximately at the 80th iteration to 0.0243, as shown on figure 3. By applying predictor-corrector boundary treatment method we have achieved 82% improvement over the prediction made by the AR model at the 0th iteration.

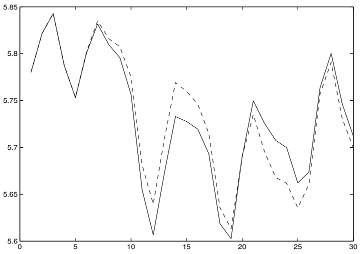
The wavelet components together with their predictions at iteration 0 and iteration 200 are shown in figure 4 respectively. From the graphs it is evident that the predictions on the last 2 components change dramatically over the 200 iterations while the predictions on the first 2 components do not change as much. This is not surprising since according to the definition of the redundant wavelet transform, the components with lower resolution are more influenced by the boundary treatment. When using predictor-corrector approach, the boundary values change



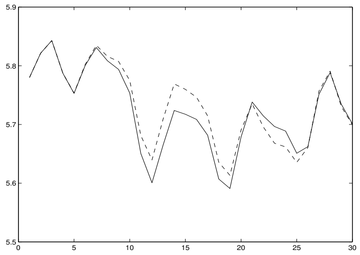
0th iteration



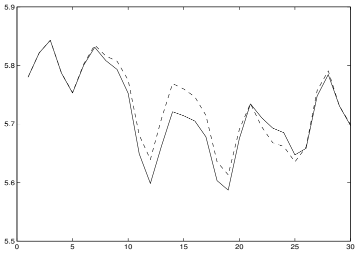
20th iteration



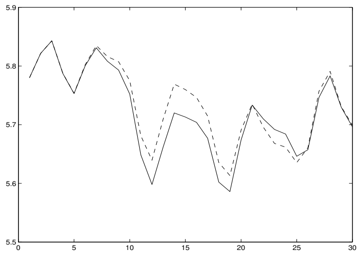
40th iteration



60th iteration



80th iteration



100th iteration

Figure 2: Predictor-corrector iterations

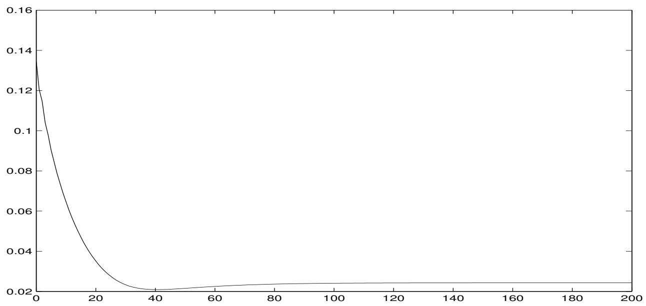


Figure 3: Errors up to 200 iterations

over iterations, therefore achieving an iterative improvement on the predictions, especially for the wavelet components with coarser resolutions.

Figure 3 provides a convenient test to determine when to stop the iteration in the experiment. However, in real life scenario, since the future values of the original series are not available to compare the prediction with, the error plot would become impossible. In this case, we design the stopping criterion to be as follows. We compare the successive iterations and if the change over the successive 2 iterations is smaller than some prescribed threshold, the iteration is terminated.

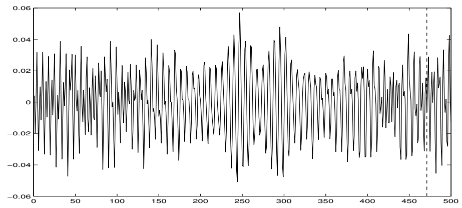
Finally, in order to study the importance of boundary conditions in wavelet transform in the context of time series forecasting, we applied the redundant wavelet transform to the denoised data using reflective boundary condition and periodic boundary conditions respectively. We also performed wavelet transform by taking advantage of the real values in the future(from 471 to 500) from the denoised data, which we call the "extensive" boundary condition. This boundary condition treatment is unrealistic, because the real values in the future are not known at the time we make prediction. However, extensive boundary condition provides us with a limit, i.e. the best wavelet decomposition we can get, to compare other boundary conditions, since the boundary is using the real values in the future. Residuals in the reflective and the periodic cases are plotted in figure 5 against the residual component from the extensive boundary treatment(dotted line). Only the values with indices from 400 to 470 are displayed. Under both, reflective and periodic, boundary conditions, the residues deviate from the correct trend significantly. So, it is expected that the prediction based on these residuals will be significantly off. In fact, experiments showed that the predictions on the first components obtained by adopting *adhoc* boundaries exploded due to the high irregularity in the tail. This suggests the necessity of a hybrid method combining Neural Nets with linear models as recommended in [14] when dealing with data of high irregularity.

To conclude this section, we would like to present a way to give an *apriori* estimate of the minimum prediction error that predictor-corrector can achieve without getting involved in the iterations. This is facilitated by the extensive boundary treatment described in the last paragraph. Extensive boundary treatment makes use of the real data values in the future, the values that predictor-corrector tries to approximate iteratively. Therefore, in principle, one should not expect predictor-corrector to ultimately do any better than the forecasting based on wavelet transform using the extensive boundary condition. Though in real life, the extensive boundary condition can not be applied due to the lack of knowledge about the future, this method can be readily utilized for analysis purposes. Applying this method to the over-the-counter-medication dataset yields the numeric error 0.013. Comparison of this error with the limiting error of the predictor-corrector iterations, 0.024, indicates that there is still room for improvement by tuning the parameters of AR or resorting to more advanced forecasting devices. The idea here may also be extended to the analysis of constraints on the time series for which predictor-corrector is helpful. This will be further explored in the future study.

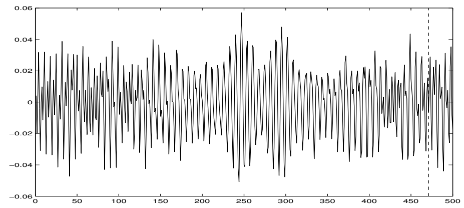
4.3 Sunspot data

In addition to the over-the-counter-medication dataset, we also applied predictor-corrector method to the well known sunspot dataset. In order to focus on the influence of the predictor-corrector scheme rather than the individual predicting devices, we insisted on using AR models instead of more sophisticated techniques. We had denoised the sunspot data using DCT first. The original sunspot and the denoised sunspot data using threshold 0.2 are shown in figure 6.

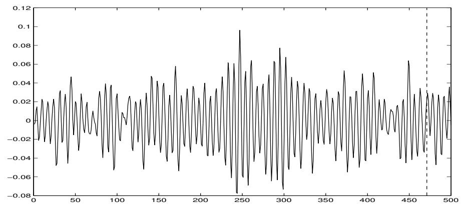
We withheld the last 30 values and used the first 235 values for training. The orders of AR for the 5 wavelet components are respectively 20, 20, 31, 45 and 41 which were chosen based



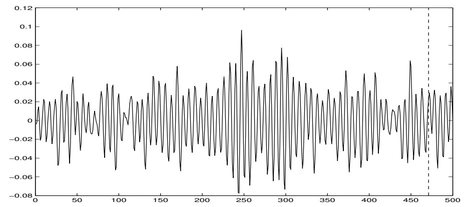
First component



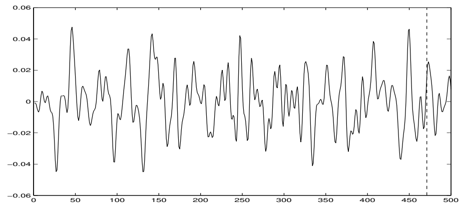
First component



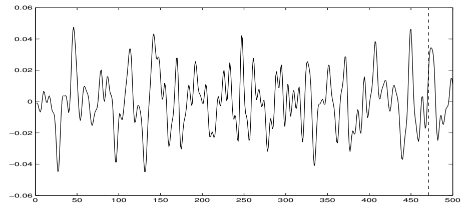
Second component



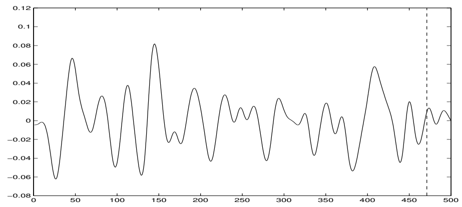
Second component



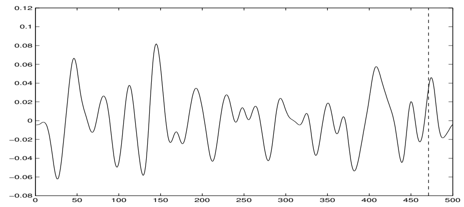
Third component



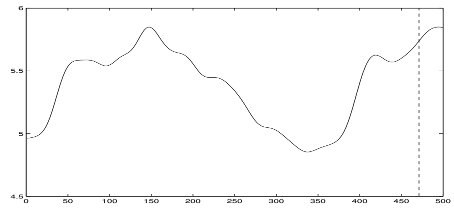
Third component



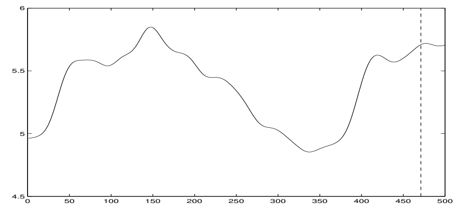
Fourth component



Fourth component

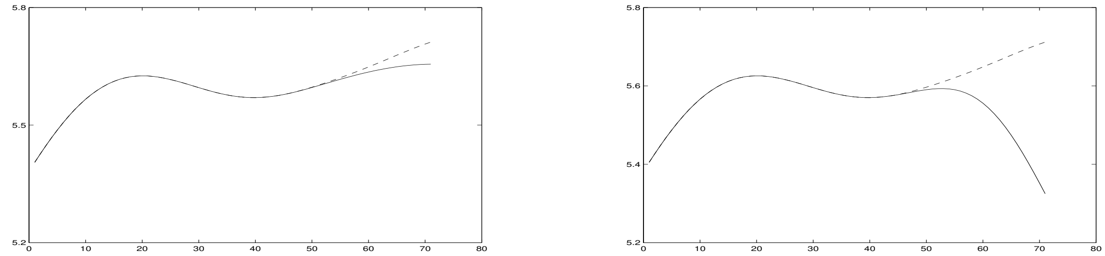


Residue



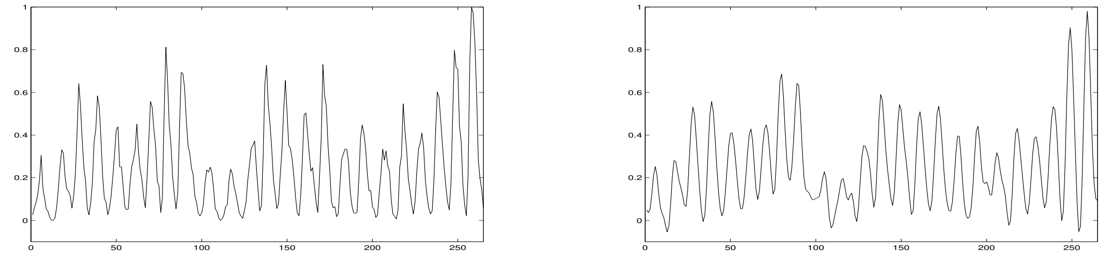
Residue

Figure 4: Five wavelet components and their predictions at the 0th(left-hand side) and the 200th(right-hand side) iterations



Residue by reflective boundary treatment Residue by periodic boundary treatment

Figure 5: Ad Hoc boundary conditions



Original sunspot data Denoised sunspot data

Figure 6: Sunspot datasets

on experiments. The error plot over 200 iterations is shown in figure 7 and overall predictions for several iterations are plotted against the denoised sunspot data in figure 8 with the solid line representing the prediction and the dotted line being the denoised sunspot data. One can see that the error curve is similar to that of the over-the-counter-medication dataset, while the way predictor-corrector improves the prediction over iterations differs from the previous case. Though for the over-the-counter-medication data it seems only the trend is gradually fixed by the predictor-corrector scheme, for the sunspot dataset three "spikes" are actually reshaped by predictor-corrector to approach the shape of the data being predicted. This suggests that the fourth wavelet component rather than the residue is being effectively corrected.

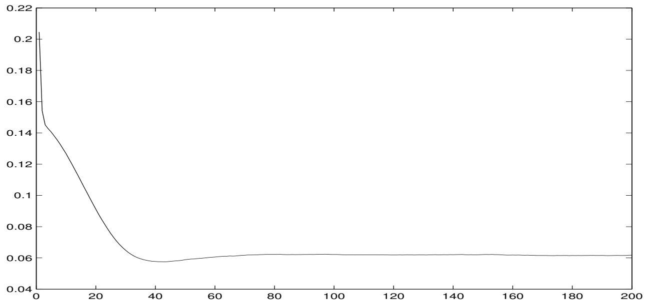


Figure 7: Errors up to 200 iterations

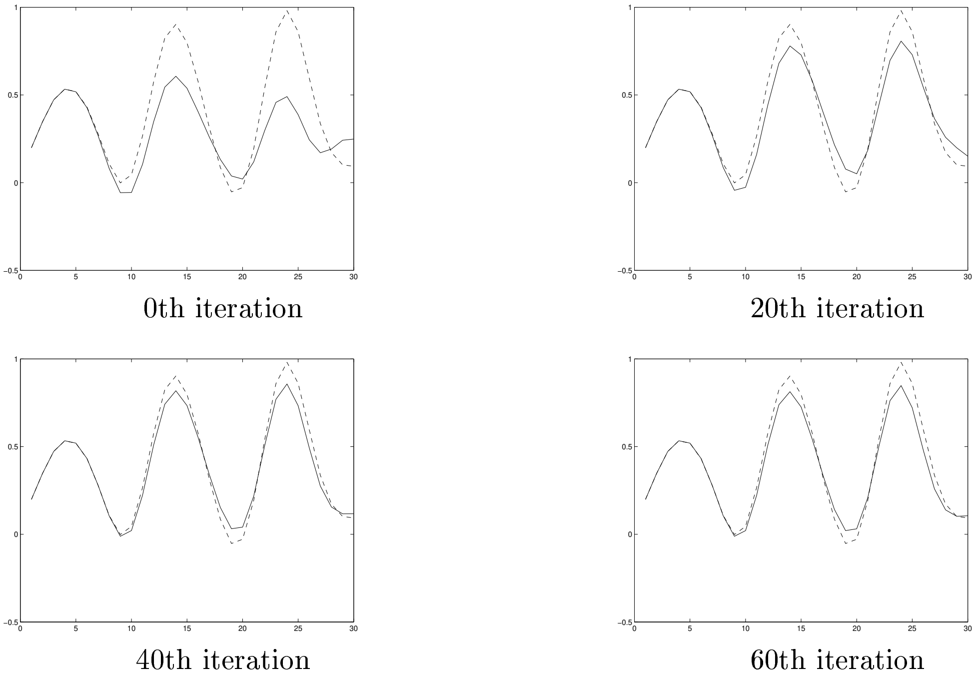


Figure 8: Predictor-corrector iterations

5 Summary

We have implemented a predictor-corrector boundary treatment algorithm that combines the heuristic methodology for boundary treatment with wavelet transform and AR models as described above. The algorithm was applied to two real world datasets, the over-the-counter-medication dataset and the sunspot dataset. In spite of the simplicity of the AR models, the results do converge to some fixed point with the error being much less than that of the initial prediction. This encouraging fact motivates the further study of combining this predictor-corrector scheme with more capable prediction techniques such as ARIMA, ANN, FIR nets to build a highly accurate and robust forecasting system. We intend to conduct a more rigorous search of the conditions when the methodology would be most helpful. More work on theoretical analysis of convergence of the predictor-corrector scheme is also planned.

6 Acknowledgements

We would like to thank David Deerfield and Laura F. McGinnis for the invaluable help with the over-the-counter medications dataset. We would also like to thank Christos Faloutsos and Rich Caruana for useful discussions.

References

[1] Yule, G. (1927) On a Method of Investigating Periodicity in Disturbed Series with Special Reference to Wolfer's Sunspot Numbers. Phil. Trans. Roy. Soc. London **A 226**:267-298.

[2] Box, G.E.P. and D.A. Pierce (1970), Distribution of residual autocorrelations in autoregressive integrated moving average time series models, Journal of the American Statistical Association, **65**, 1509-1526.

[3] Weigend, A.S. and Gershenfeld, N. A., Time Series Prediction, Forecasting the Future and Understanding the Past, A Proceedings Volume in the Santa Fe Institute Studies in the Sciences of Complexity, 16-17.

[4] Weigend, A.S., Rumelhart, D.E. and Huberman, B.A. (1990) Predicting the Future: A Connectionist Approach. International Journal of Neural Systems, 1, 195-220.

[5] Tong, H. and Lim, K.S. (1980) Threshold Autoregression, Limit Cycles and Cyclical Data. J. Roy. Stat. Soc. B **42**:245-292.

[6] Tong, F. (1990) Non-linear Time Series, Oxford: Clarendon Press.

[7] Daubechies, I. (1992) Ten Lectures on Wavelets. Philadelphia: SIAM

[8] Nason, G.P. and Silverman, B.W. (1995) The discrete wavelet transform is S. J. Comput. Graph. Statist., **3**, 163-191

[9] Nason, G.P. (1996) Wavelet shrinkage using cross-validation. J. R. Statist. Soc. B, **58**, 463-479

[10] Morettin, P.A. (1996) From Fourier to wavelet analysis of time series. In Proceedings in Computational Statistics, Pratt, A. (Ed.), 111-122, New York: Physica-Verlag.

[11] Morettin, P.A. (1997) Wavelets in Statistics. Resenhas, **3**, 211-272.

[12] Priestley, M.B.(1996) Wavelets and time-dependent spectral analysis. J. Time Series Anal., **17**, 85-104

[13] Percival, D. B. and Walden, A.T.(1999), Wavelet Methods for Time Series Analysis. Cambridge:Cambridge University Press.

[14] Aussem, A. and Murtagh, F. (1997) Combining Neural Network Forecasts on Wavelet-transformed Time Series. Connection Science Vol. 9, **1** 113-121.