Estimating the horizon of predictability in time-series predictions using inductive modelling tools

Josefina López[a], François E. Cellier[b]* and Gabriela Cembrano[c]

[a]Software Department, Technical University of Catalonia (UPC), C/Colom 11, Terrassa 08222, Spain; [b]Department of Computer Science, ETH Zurich, CH-8092 Zurich, Switzerland; [c]Institute of Robotics & Industrial Informatics, Technical University of Catalonia (UPC), LLorens i Artigas 4–6, Barcelona 08028, Spain

This paper deals with the assessment of how far into the future a time series can be safely predicted using inductive modelling and extrapolation techniques. Three different time series are used to demonstrate the viability of the approaches presented in the paper: one time series representing the water demand of the city of Barcelona, another characterizing the water demand of a section of the city of Rotterdam, and a third describing weather data for the city of Tucson. Fuzzy inductive reasoning (FIR) is used to predict future values of these time series on the basis of their own past. FIR predictions come with two different built-in measures of confidence that can be used to obtain a quantitative estimate of how far into the future a time series can be predicted.

Keywords: inductive modelling; time-series prediction; fuzzy inductive reasoning; estimation of predictability horizon

1. Introduction

In a companion paper (Cellier *et al.* 2010), the problem of estimating the forecasting error in time-series predictions was discussed. It was shown that, especially in soft science simulation, it is important to estimate the error of a prediction together with the prediction itself, since it cannot be expected of the users that they would be able to assess the reliability of the simulation results. Scepticism must be instilled in the simulation software, rather than demanding it of its users.

In other forecasting methods, it is customary to provide a prediction value accompanied by a confidence range or error envelope, based on a computation of the mean square error of the series prediction in question. Assuming that the prediction errors are normally distributed around the 0 value and based on the characteristics of the normal distribution, it is possible to provide, for example, 95% confidence ranges, using the prediction value $\pm 2s$, s being an estimation of the standard deviation of the error distribution around 0. Such a confidence estimation is informative, but it has limitations. On the one hand, it only provides a global estimation of the expected error,

and on the other, it relies entirely on the normality assumption that may or may not hold.

The method proposed in this work computes a *local* estimation of the confidence of each individual prediction and makes no previous assumption on the error distribution of the prediction. Additionally, since successive predictions of a several-step-ahead forecast are not totally accurate, errors are likely to accumulate during iterative predictions of future values of a time series. It is thus of much interest to the user of such a tool to be able to assess the quality of predictions made not only locally, but also as a function of time, i.e. the user should be able to obtain a (generally decaying) function of *accumulated confidence* in progressive predictions. During the first step of a multi-step prediction, the predicted value depends entirely on measurement data, and is, therefore, more likely to be accurate than in subsequent steps, when the predictions depend on previously predicted data points that are by themselves associated with a degree of uncertainty already.

There exist many applications for such a technology. For example, model predictive control uses predictions of future values of measurement data to provide the controller with an early warning if the system is about to leave the zone of safe operation. The earlier such a warning can be provided, the more time the controller has to prevent this situation from ever happening.

Yet, there are two types of errors that can occur in such predictions: (i) the predictor foresees that the system will leave its operating zone, although in reality, this would not happen and (ii) the predictor does not foresee any problems, although they do occur. Both types of errors can degrade the achievable performance of the controller. The first error type will make the controller overly conservative, preventing it from making use of the full operating zone. The second error type may lead to either instability or plant shutdown.

Both error types are closely related to the horizon of predictability. As the accuracy of forecasts in a multi-step prediction decreases over time, the likelihood of making either type of error grows. Hence, assessing the likelihood of these errors to occur is synonymous with being able to assess the horizon of predictability of each measurement signal used in the predictive control scheme.

The paper introduces measures for estimating the horizon of predictability. It then calculates the prediction errors made when forecasting three separate time series over multiple steps, and shows the strong positive correlation between the prediction error on the one hand and the estimated horizon of predictability on the other.

2.   Accumulated confidence measures in time-series prediction

As was shown in Cellier *et al.* (2010), the local prediction error can be estimated using either *proximity* or *similarity* measures. Both types of estimators lead to satisfactory results when used together with a fuzzy inductive reasoning (FIR) algorithm for the time-series prediction, although the similarity measure is usually preferred as it is slightly more sensitive than the proximity measure.

Both measures only account for uncertainty stemming from a single step of prediction, i.e. they assume that the data on which the prediction is based are totally accurate, i.e. they measure the *local* uncertainty associated with a single prediction, but not the *accumulated* uncertainty resulting from multiple predictions, whose premises are themselves uncertain already.

Either measure can easily be extended to become an estimator of accumulated confidence. The reader may remember that an FIR model of a time-series predictor

is characterized by a single-column optimal mask, e.g.

$$
\begin{array}{cc}
t\backslash x & y \\
t - 5\delta t & \\
t - 4\delta t & \\
t - 3\delta t & \\
t - 2\delta t & \\
t - \delta t & \\
t &
\end{array}
\begin{pmatrix}
-1 \\
0 \\
0 \\
-2 \\
-3 \\
+1
\end{pmatrix}
\tag{1}
$$

denoting the equation

$$
y(t) = \tilde{f}\big(y(t - 5\delta t), y(t - 2\delta t), y(t - \delta t)\big), \tag{2}
$$

where $\tilde{f}$ denotes a function specified through a finite state machine, rather than being provided in the form of an analytical expression.

Negative mask elements denote mask inputs (m-inputs), whereas the $+1$ element, which will always show up in the last row, denotes the mask output (m-output).

For the above mask, it makes sense to define the accumulated confidence in the prediction of $y(t)$ as follows:

$$
c_a(t) = c_l(t) \cdot \frac{1}{3} \cdot (c_a(t - 5\delta t) + c_a(t - 2\delta t) + c_a(t - \delta t)), \tag{3}
$$

i.e. the accumulated confidence in the prediction of $y(t)$, called $c_a(t)$, is defined as the product of the local confidence in that prediction, $c_l(t)$, with the average accumulated confidence in the m-inputs. Clearly, both the local and accumulated confidence values of measured data points are 1.0; and, therefore, the accumulated confidence of the first prediction step is always equal to the local confidence, computed using either the proximity or the similarity measure, but at later times, the accumulated confidence is always lower than the local confidence. The accumulated confidence is usually decaying over time, although it is not necessarily a monotonically decreasing function.

The multiplication of the local confidence of the m-output with the average accumulated confidence of the m-inputs is only correct, in a strict sense, if subsequent values of y can be assumed to be uncorrelated, which, of course, is never the case. However, from a practical standpoint, the measure works exceedingly well, as shall be demonstrated by means of three separate examples.

Of course, the proposed approach to estimate the accumulated confidence in predictions made is not limited to time series. For example, given a system with two inputs and three outputs, characterized by the following optimal mask

$$
\begin{array}{cccccc}
t\backslash x & u_1 & u_2 & y_1 & y_2 & y_3 \\
t - 2\delta t & & & & & \\
t - \delta t & & & & & \\
t & & & & &
\end{array}
\begin{pmatrix}
-1 & 0 & -2 & 0 & 0 \\
0 & -3 & 0 & 0 & -4 \\
0 & 0 & +1 & 0 & 0
\end{pmatrix}
\tag{4}
$$

denoting that

$$y_1(t) = \tilde{f}\big(u_1(t - 2\delta t), y_1(t - 2\delta t), u_2(t - \delta t), y_3(t - \delta t)\big) \qquad (5)$$

would lead to the following expression of accumulated confidence:

$$c_a(y_1(t)) = c_l(y_1(t))\big(0.5 + 0.25 \cdot c_a(y_1(t - 2\delta t)) + 0.25 \cdot c_a(y_3(t - \delta t))\big). \qquad (6)$$

Since the input variables are always measured and, therefore, assumed to be accurate, the accumulated confidence values associated with $u_1(t - 2\delta t)$ and $u_2(t - \delta t)$ are always assumed to be 1.0. This time, it was necessary to specify the names of the variables as arguments of the c_a and c_l functions, since multiple variables are contributing to the prediction.

3. Description of the simulation experiments

Matrix 7 shows how the computations were performed. It shows an excerpt of the time series. The first column denotes the true measurement data. At each sampling point, a multi-step prediction was performed, the results of which are written to the right of the last measurement data point used in the prediction. The first argument denotes the time instant for which the prediction is computed, whereas the second argument denotes the number of prediction steps used to reach the prediction. Across one of the anti-diagonals, values are marked in bold type for illustration. They all refer to the same time point, yet values further to the right and top are less accurate, because they were obtained using a longer prediction path (second argument).

$$Y = \begin{pmatrix}
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
y(t - 4\delta t) & y(t - 3\delta t, 1) & y(t - 2\delta t, 2) & y(t - \delta t, 3) & y(t, 4) & \cdots \\
y(t - 3\delta t) & y(t - 2\delta t, 1) & y(t - \delta t, 2) & y(t, 3) & y(t + \delta t, 4) & \cdots \\
y(t - 2\delta t) & y(t - \delta t, 1) & y(t, 2) & y(t + \delta t, 3) & y(t + 2\delta t, 4) & \cdots \\
y(t - \delta t) & y(t, 1) & y(t + \delta t, 2) & y(t + 2\delta t, 3) & y(t + 3\delta t, 4) & \cdots \\
y(t) & y(t + \delta t, 1) & y(t + 2\delta t, 2) & y(t + 3\delta t, 3) & \mathbf{y(t + 4\delta t, 4)} & \cdots \\
y(t + \delta t) & y(t + 2\delta t, 1) & y(t + 3\delta t, 2) & \mathbf{y(t + 4\delta t, 3)} & y(t + 5\delta t, 4) & \cdots \\
y(t + 2\delta t) & y(t + 3\delta t, 1) & \mathbf{y(t + 4\delta t, 2)} & y(t + 5\delta t, 3) & y(t + 6\delta t, 4) & \cdots \\
y(t + 3\delta t) & \mathbf{y(t + 4\delta t, 1)} & y(t + 5\delta t, 2) & y(t + 6\delta t, 3) & y(t + 7\delta t, 4) & \cdots \\
\mathbf{y(t + 4\delta t)} & y(t + 5\delta t, 1) & y(t + 6\delta t, 2) & y(t + 7\delta t, 3) & y(t + 8\delta t, 4) & \cdots \\
y(t + 5\delta t) & y(t + 6\delta t, 1) & y(t + 7\delta t, 2) & y(t + 8\delta t, 3) & y(t + 9\delta t, 4) & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{pmatrix}. \qquad (7)$$

The above matrix can now be used in different ways. Horizontal rows indicate individual multi-step predictions starting from the time shown in the first column that represents the last measurement data point. Each new data point contains more sources of error than the previous one, because it is built on a longer prediction history. Vertical columns show long-term prediction cycles, whereby the measurement data lag behind the prediction

by a fixed number of steps. Columns further to the left should, on average, be more accurate than columns further to the right, because the prediction history leading to them is shorter. The first column is 100% accurate, since it represents the measurement data. Finally, values in anti-diagonals represent the same time instant estimated using longer and longer prediction histories.

The *accumulated confidence* values associated with each prediction were stored in a second matrix of identical dimensions. All values in column 1 are 1.0. Values further to the right are likely to be smaller than those further to the left. The average value of each column

$$c_a[j] = \frac{1}{n} \sum_{i=1}^{n} c_a[i,j] \tag{8}$$

is the average accumulated confidence associated with a prediction that is based on measurement data that are lagging j sampling intervals behind.

It was decided to define the *prediction error* in the same fashion as proposed in Cellier *et al.* (2010):

$$M = \max(\max(Y)), \tag{9}$$

$$m = \min(\min(Y)), \tag{10}$$

$$Yn = \frac{Y - m}{M - m}, \tag{11}$$

$$\text{err}_{\text{abs}}[i,j] = |Yn[i,j] - Yn[i,0]|, \tag{12}$$

$$\text{simty}_i[i,j] = \frac{\min(Yn[i,j], Yn[i,0])}{\max(Yn[i,j], Yn[i,0], \varepsilon)}, \tag{13}$$

$$\text{err}_{\text{sim}} = 1.0 - \text{simty}, \tag{14}$$

$$\text{err} = \frac{\text{err}_{\text{abs}} + \text{err}_{\text{sim}}}{2}, \tag{15}$$

i.e. all the data in the prediction matrix, Matrix 7, are first normalized together to the range [0.0,1.0]. Then, the absolute error of each prediction is computed by calculating the absolute difference between each value of the normalized prediction matrix and the value stored in the same anti-diagonal (representing the same time instant) in column 1. The matrix of absolute errors exhibits a first column of 0.0 values. Then, a dissimilarity error is computed. Also the matrix of dissimilarity errors shows a first column of 0.0 values. Finally, the error matrix is computed as the average between the absolute and dissimilarity error matrices.

The average value of each column of the error matrix

$$\text{err}[j] = \frac{1}{n} \sum_{i=1}^{n} \text{err}[i,j] \tag{16}$$

represents the average error associated with a prediction made over j sampling intervals.

4.   Simulation results

Three separate time series were used to investigate the effectiveness of the proposed accumulated confidence measures as indirect statistical estimators for the prediction error to be expected.

The first time series represents the water demand of an area of the city of Barcelona (Aigües de Barcelona 1987). The measurement data are shown in Figure 1. The time series is only mildly stochastic. The autocorrelation of this time series is shown in Figure 2. Even by naked eye, it is quite easy to discern a strong weekly cycle. One-and-a-half years' worth of daily measurements, from January 1985 to July 1986, were available to generate the model. This is one of the time series that had been used in Cellier *et al.* (2010) and López *et al.* (1996).

The second time series represents the water demand of the city of Rotterdam (Europoort 1996). The measurement data are shown in Figure 3. The behaviour of this time series is considerably more stochastic than the previous one. However, there is still quite a bit of autocorrelation contained in this time series, as shown in Figure 4. Also this autocorrelation function shows a weekly cycle, though the peaks decay much more rapidly than in the case of the Barcelona series. Luckily, more measurement data were available for Rotterdam, namely 10 years' worth of daily measurements, from January 1986 to December 1995.

The third time series represents the temperature of the city of Tucson (NOAA records for the city of Tucson). The measurement data are shown in Figure 5. The data exhibit
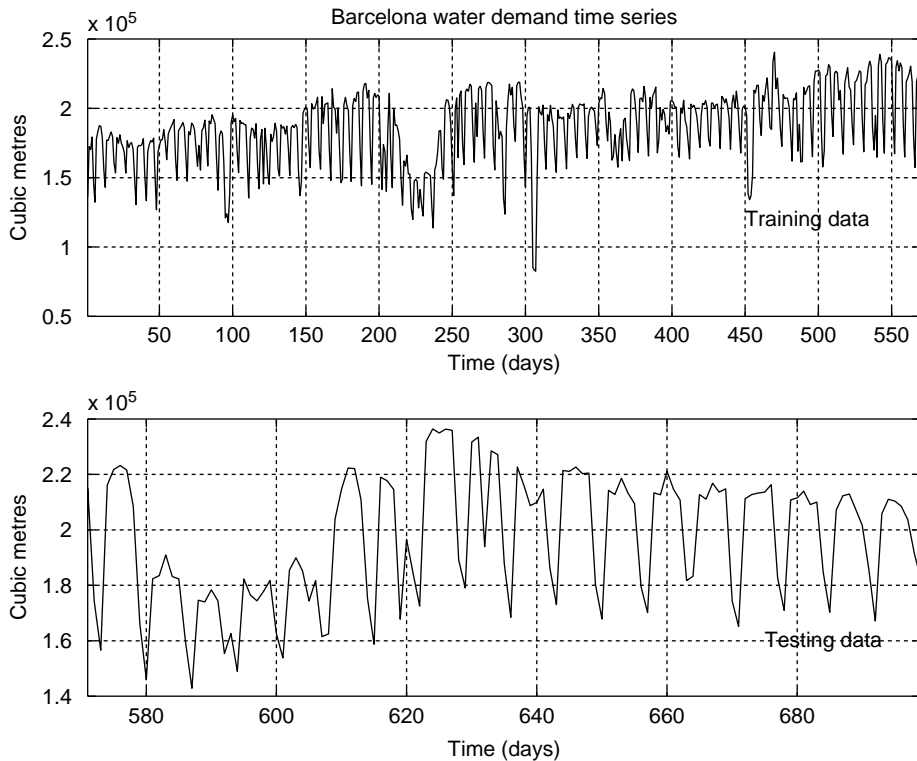

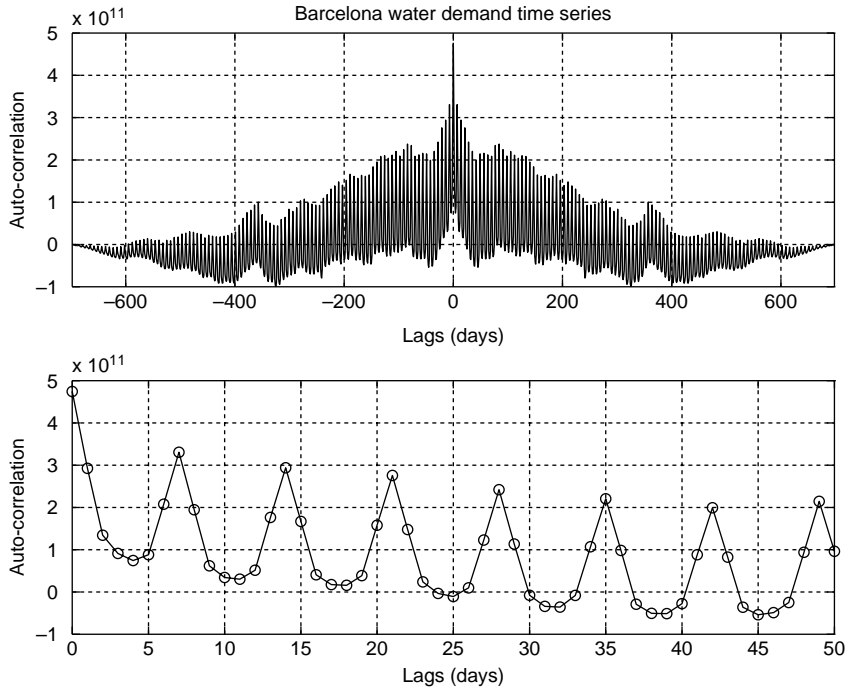Figure 1.   Barcelona water demand – training and testing data.

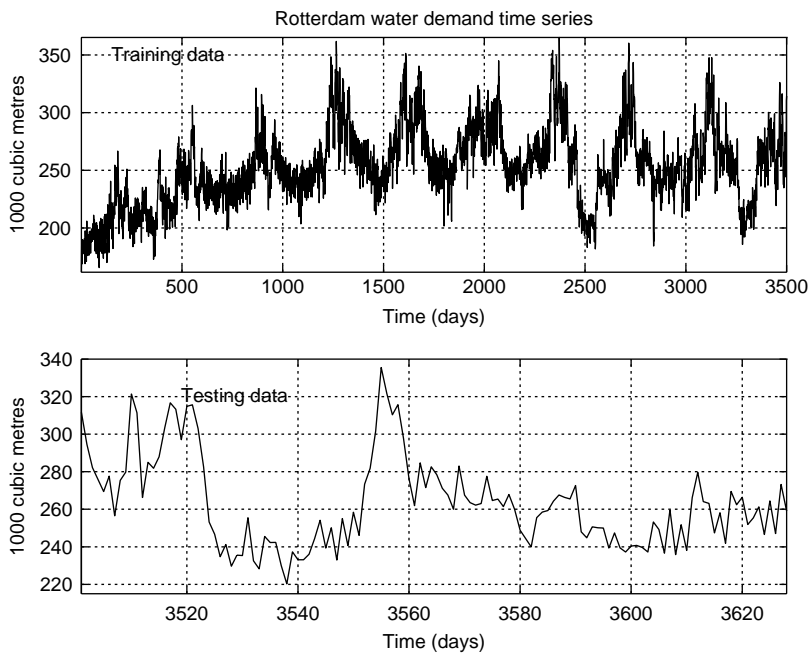Figure 2.   Barcelona water demand − autocorrelation functions.

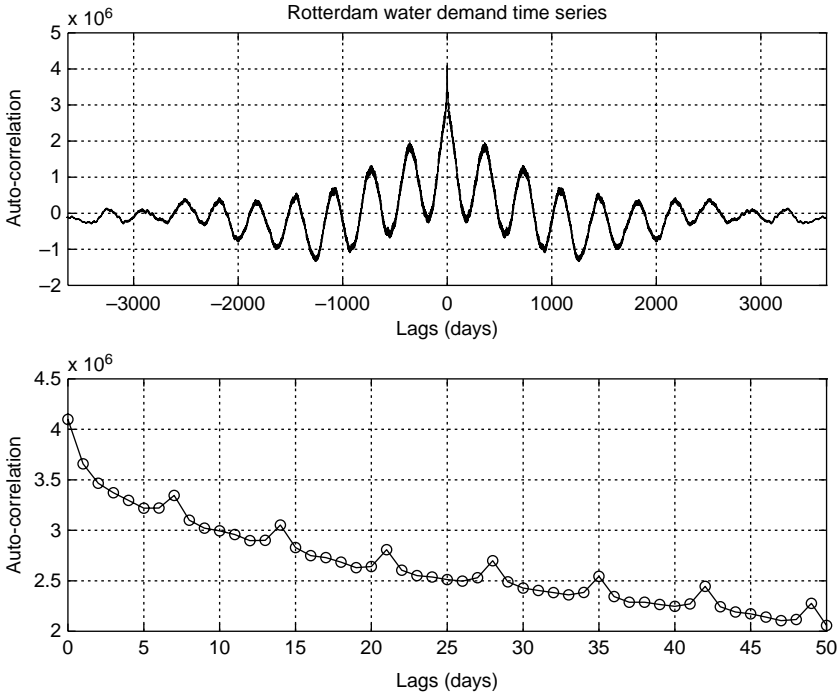Figure 3.   Rotterdam water demand − training and testing data.

Figure 4.   Rotterdam water demand – autocorrelation functions.

a strong daily cycle. The autocorrelation function is given in Figure 6. Hourly measurements were available for the entire year 1995.

   The three time series, as well as all the programs leading to the results discussed in subsequent sections, can be obtained from the World Wide Web at the URL: http://www.inf.ethz.ch/ ~ fcellier/Pubs/FIR/ConfHorz.html.

4.1   *Water demand of the city of Barcelona*

The optimal FIR model for this time series was found to be

$$
\begin{array}{c c}
 & y \\
t - 14\delta t & \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ -2 \\ 0 \\ 0 \\ 0 \\ -3 \\ +1 \end{pmatrix}
\end{array}
$$

$$
\begin{aligned}
t - 14\delta t &\quad -1 \\
t - 13\delta t &\quad 0 \\
\cdots &\quad 0 \\
t - 8\delta t &\quad 0 \\
t - 7\delta t &\quad -2 \\
t - 6\delta t &\quad 0 \\
\cdots &\quad 0 \\
t - 2\delta t &\quad 0 \\
t - \delta t &\quad -3 \\
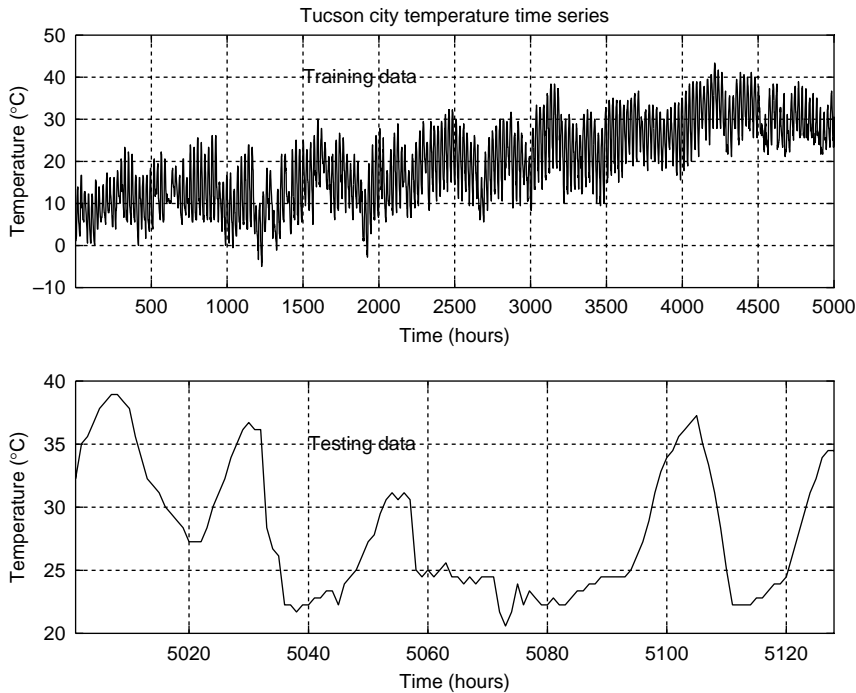t &\quad +1
\end{aligned}
\tag{17}
$$

Figure 5.   Tucson weather data – training and testing data.

This result is quite reasonable. Due to the strong weekly cycle inherent in this time series, FIR concludes that the most useful data points to predict today's water demand are yesterday's water demand, last week's water demand and the water demand 2 weeks ago.

Five hundred and seventy days (from 1 January 1985 to 24 July 1986) were used as training data, whereas 128 days (from 25 July 1986 to 29 November 1986) were used as testing data. Thanks to the strong autocorrelation of this time series, 570 data points were sufficient to derive a model exhibiting fairly good short-term prediction capabilities.

Figure 7 shows the accumulated confidence obtained when predicting over multiple days starting from the first day of prediction. As expected, the accumulated confidence decays rapidly, although not monotonically. The accumulated confidence does not have to decay monotonically, because the accumulated confidence function

$$c_a(t) = c_l(t) \cdot \frac{1}{3}(c_a(t - \delta t) + c_a(t - 7\delta t) + c_a(t - 14\delta t)) \tag{18}$$

depends each time on different past confidence values.

A prediction matrix with 16 columns was constructed, i.e. at each time instant, a multi-step prediction over 15 days was performed. The average error err[j] and the average accumulated confidence $c_a[j]$ are plotted in Figure 8. As a gauge, the error is compared to that of the *trivial prediction* that operates on the simple hypothesis that the predicted water use is equal to the current one. On average, the FIR predictor exhibits an error that is approximately 30% below that of the trivial predictor, except for days 7 and 14, for which the trivial predictor makes decent predictions due to the high autocorrelation for these days.
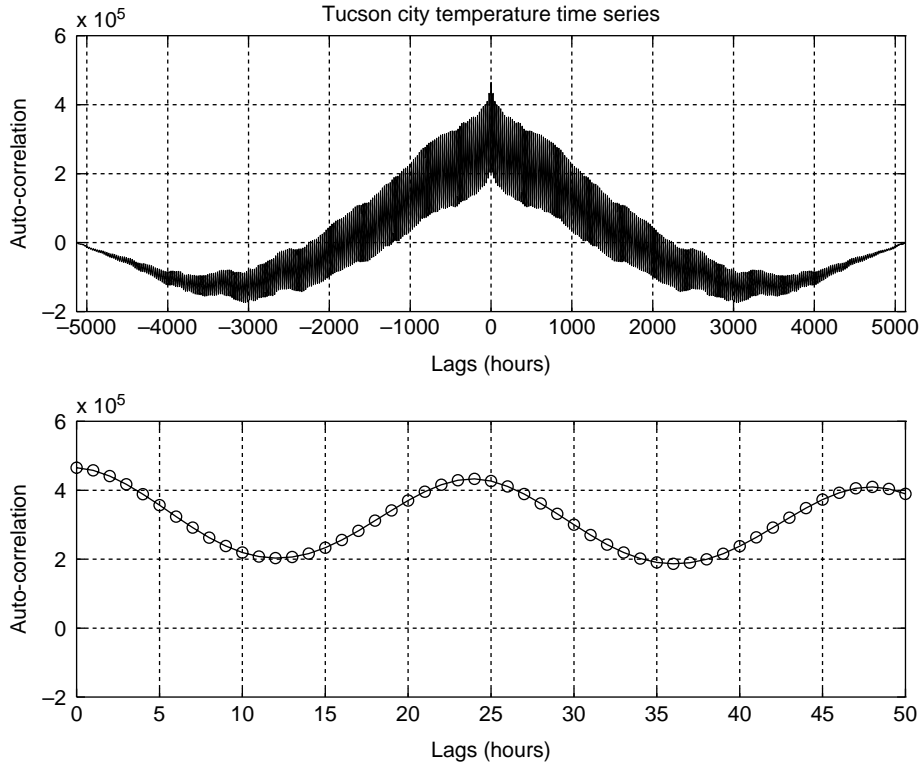
Figure 6. Tucson weather data – autocorrelation functions.

As already indicated in Cellier *et al.* (2010), the similarity measure is slightly more sensitive than the proximity measure, and, therefore, the accumulated confidence decays more rapidly.

The gradients of both confidence measures are slightly steeper during the 8th day of prediction than during the previous days. This makes sense. During the first day, the prediction depends on measured data only. During days 2–7, one of the past data points used in the prediction, the value $y(t - \delta t)$ is already contaminated by previous predictions, whereas the other two data points, $y(t - 7\delta t)$ and $y(t - 14\delta t)$, are measurement data. As of day 8, a second of the past data points used in the prediction gets contaminated, and, therefore, the accumulated confidence drops down further. As of day 15, even the third past data point gets contaminated, leading to a further decay in accumulated confidence.

The correlation between the averaged error, err[j], and the function $1.0 - c_a[j]$ is indeed very strong. The reader may notice that also the error grows more rapidly during the first day, and then again during the 8th and 9th days.

Figure 9 compares the 1-day prediction, the 8-day prediction, and the 15-day prediction with the measurement data. The reduction in forecasting quality is quite noticeable, yet, even a 2-week forecast is still somewhat meaningful.

Earlier models of the same time series using a Box–Jenkins approach and a neural network methodology, respectively, were published in Quevedo *et al.* (1988) and Griñó (1992).
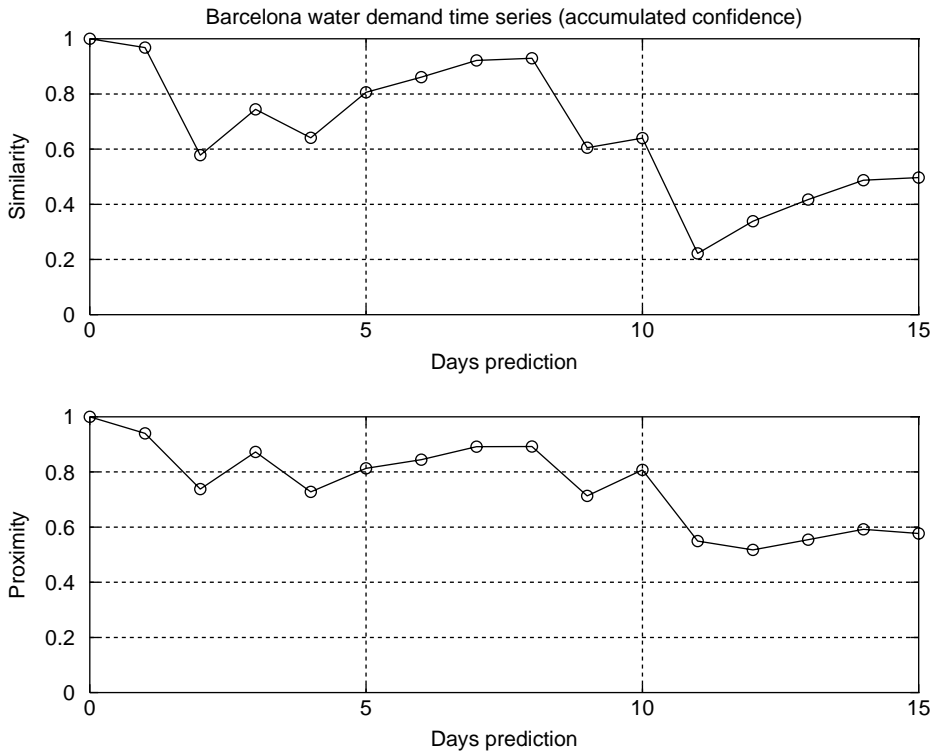
Figure 7. Barcelona water demand prediction – accumulated confidence values.

4.2 *Water demand of the city of Rotterdam*

Due to the more stochastic nature of this time series, more data points were needed for model identification. Of the available 10 years of data, 9.5 years of data (corresponding to 3500 data points) were used as training data (i.e. for model identification), whereas the remaining 0.5 years' worth of data were used as testing data (i.e. for model validation).

FIR found the following optimal mask:

$$
\begin{array}{c}
 & y \\
t - 7\delta t \\
t - 6\delta t \\
t - 5\delta t \\
t - 4\delta t \\
t - 3\delta t \\
t - 2\delta t \\
t - \delta t \\
t
\end{array}
\begin{pmatrix}
-1 \\
0 \\
0 \\
0 \\
-2 \\
0 \\
-3 \\
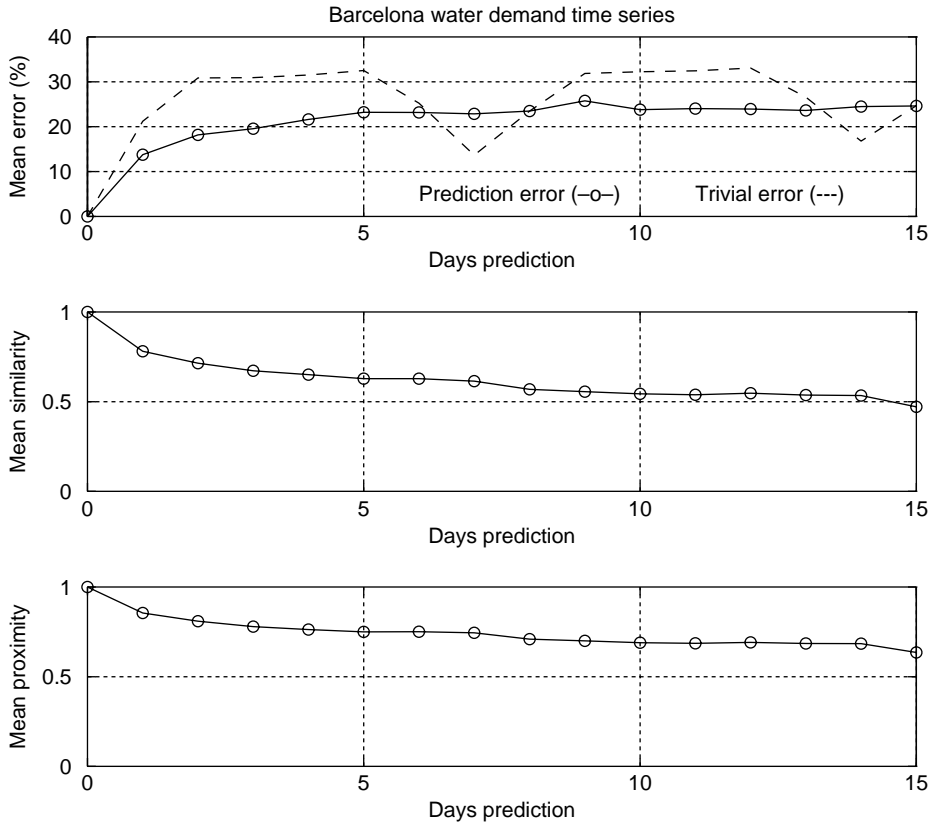+1
\end{pmatrix} . \tag{19}
$$

Figure 8.   Barcelona water demand predictions – mean errors and confidence values.

Again, the model that FIR proposes is quite reasonable. Because of the more rapid decay of the autocorrelation function, the data point $y(t - 14\delta t)$ is less relevant than in the previous case. Instead, FIR chose to also use the data point $y(t - 3\delta t)$ for the prediction.

Figure 10 shows the averaged error err[j] and the two averaged accumulated confidence functions $c_a[j]$ as functions of the number of sampling periods, j, that the measurement data lag behind the prediction. Just as in the case of the Barcelona series, the errors are compared to those of the trivial predictor. It turns out that the series is so highly stochastic that, over a short time span, FIR cannot beat the trivial predictor, i.e. it essentially does not predict anything.

Figure 11 compares the 1-day prediction, the 8-day prediction and the 15-day prediction to the measurement data. The 1-day prediction looks quite decent to the naked eye, but so would the trivial prediction. It can be seen that the hills and valleys lag more and more behind as the number of prediction days is increased. This is because they are not really being predicted, they are only remembered.

An earlier model for the prediction of the water demand of the city of Rotterdam (or, more precisely, its Berenplaat region) was published in Baggelaar (1992). It used a Box–Jenkins approach to make a 1-day prediction of the water demand in that region.
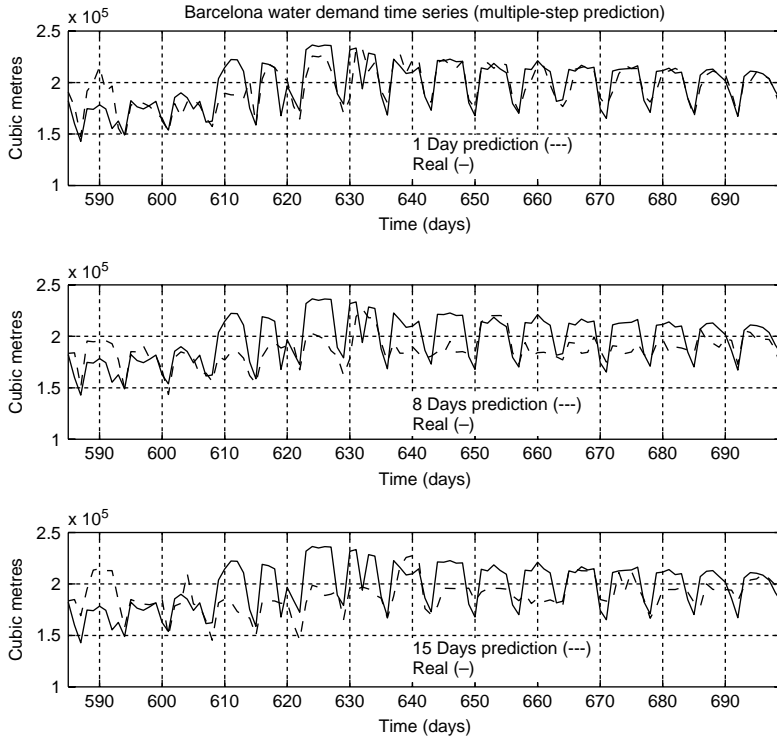
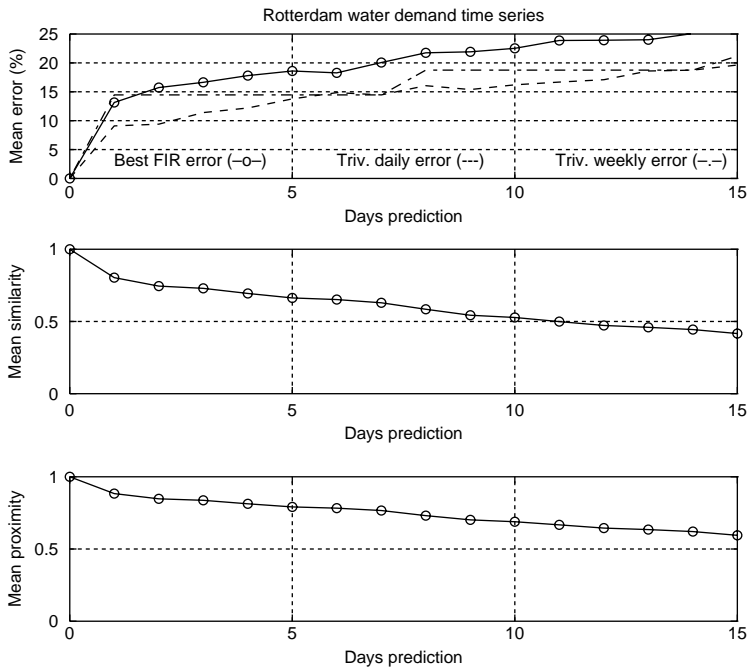Figure 9.   Barcelona water demand – predictions.

Figure 10.   Rotterdam water demand predictions – mean errors and confidence values.

4.3 *Tucson weather prediction*

The third time series is of particular interest, because there exists a rich literature about weather prediction and the (usually quantitative) models used for it. It is well established that a prediction over about 5 days is feasible from local data, whereas a longer term prediction will not work due to the chaotic nature of the underlying physical system. It is to be expected that the simple FIR model used in this paper will do a much poorer job than the sophisticated partial differential equation models discussed in the open literature, as it only takes into account previous ambient temperature values, ignoring other important factors such as cloud cover, humidity, sky radiation and the effective temperature of the night sky to mention just a few of the more important influencing factors, quantities that are being taken into account by the more sophisticated quantitative models. Yet, the investigation is of interest in order to better understand the relative importance of first-order vs. second-order effects, i.e. to be able to assess to what extent the temperature is determined by its own past.

Five thousand of the available data points were used as training data, whereas another 1000 data points were used for prediction. The optimal model proposed by FIR is the following:

$$
\begin{array}{cc}
 & y \\
t - 21\delta t & \begin{pmatrix} -1 \\ \end{pmatrix} \\
t - 20\delta t & 0 \\
\cdots & 0 \\
t - 2\delta t & 0 \\
t - \delta t & -2 \\
t & +1 \\
\end{array}
. \tag{20}
$$

The optimal mask here is of complexity 3, i.e. makes use of one less mask input than in the case of the previous two time series. The best mask of complexity 4 is

$$
\begin{array}{cc}
 & y \\
t - 32\delta t & \begin{pmatrix} -1 \\ \end{pmatrix} \\
t - 31\delta t & 0 \\
\cdots & 0 \\
t - 21\delta t & 0 \\
t - 20\delta t & -2 \\
t - 19\delta t & 0 \\
\cdots & 0 \\
t - 2\delta t & 0 \\
t - \delta t & -3 \\
t & +1 \\
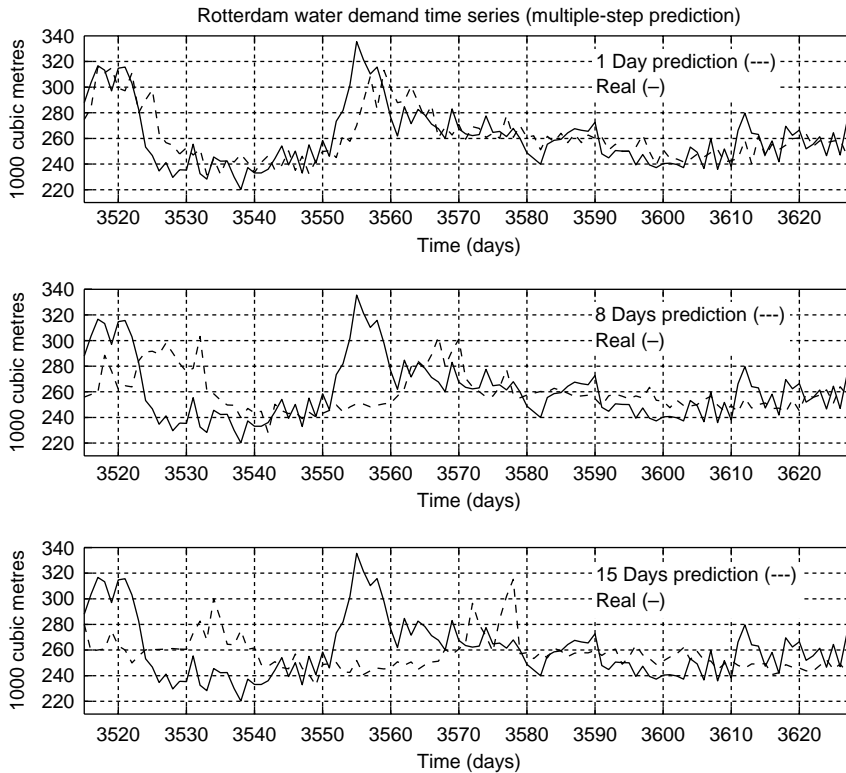\end{array}
. \tag{21}
$$

Figure 11.   Rotterdam water demand – predictions.

Surprisingly, FIR decided that it is better to base the prediction on data values that are somewhat off in time from the expected sampling points of $t - 24\delta t$ and $t - 48\delta t$, i.e. although the autocorrelation function (Figure 6) has its peaks at $t - 24\delta t$ and $t - 48\delta t$, FIR does not consider these to be the best sampling points to use for its prediction.

This time, a prediction matrix with 51 columns was chosen, i.e. at each time instant, a multi-step prediction over 50 steps was performed. The averaged errors, err[j], and the averaged accumulated confidence values, $c_a[j]$, are shown in Figure 12. Superimposed with the averaged errors are the error values obtained for the trivial prediction. FIR cannot beat the trivial predictor on short-term predictions due to the slow time constant of the temperature. Yet for predictions of more than about 4 h, FIR starts to outperform the trivial predictor, and does so consistently until 25 h. Thereafter, FIR gets lost in its own computations, and starts producing noise instead of predicting the desired signal.

The confidence measures (lower two curves of Figure 12) show dramatically the effect of data values being contaminated by previous predictions. Up to 21 h of prediction, only one of the two data points is contaminated by a previous prediction, whereas the other data point is still clean. After 42 h, the next iteration comes to play, as there are now data being used that are doubly contaminated, i.e. each of the data points in use depends on predictions that themselves already depend on previously made predictions.

Figure 13 compares the 1-h prediction the 24-h prediction, and the 48-h prediction with the trivial predictions made over the same time horizons, and with the measurement
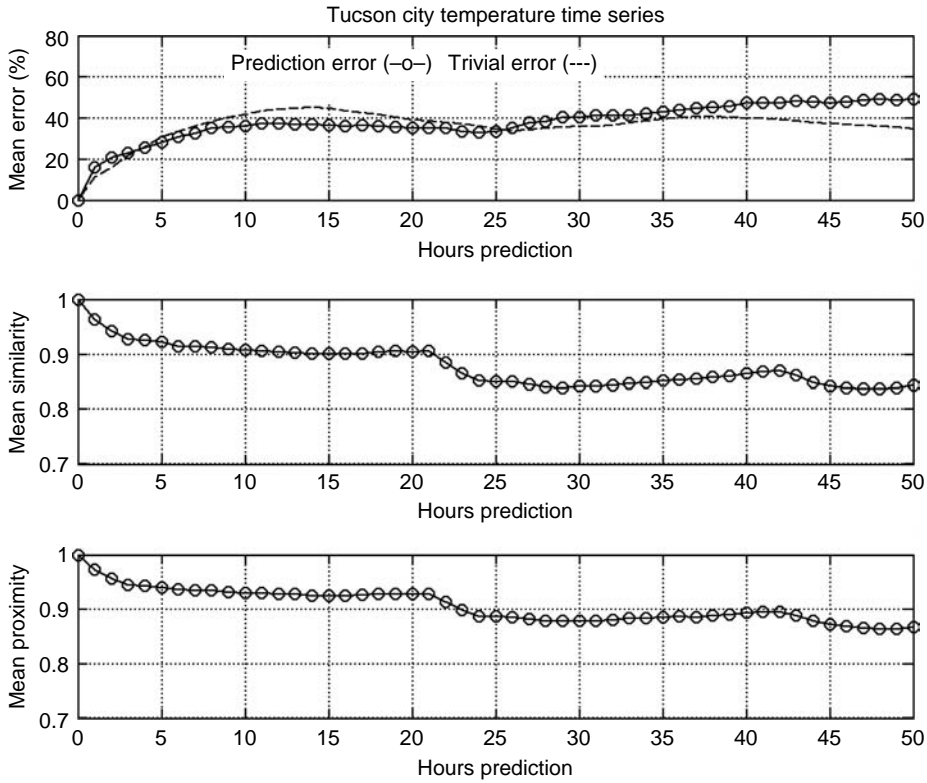
Figure 12.　Tucson weather predictions: mean errors and confidence values.

data. Clearly, the 1-h FIR prediction is considerably better than the 1-h trivial prediction. However, the two 24-h predictions are of comparable quality, and the same holds for the 48-h predictions.

Was FIR correct in its assessment of which sampling points to use? To answer this question, a second simulation was performed using the mask

$$
\begin{array}{cc}
 & y \\
t - 24\delta t & \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ -2 \\ +1 \end{pmatrix} \\
t - 23\delta t & \\
\cdots & \\
t - 2\delta t & \\
t - \delta t & \\
t &
\end{array}. \tag{22}
$$

The results of this simulation are plotted in Figure 14. The errors are consistently higher than those found with the previous mask, and the confidence values are consistently lower. Evidently, FIR was correct in its assessment. The Shannon entropy measure that FIR uses to determine its optimal mask is considerably a better estimator of information content than the (linear) autocorrelation function.
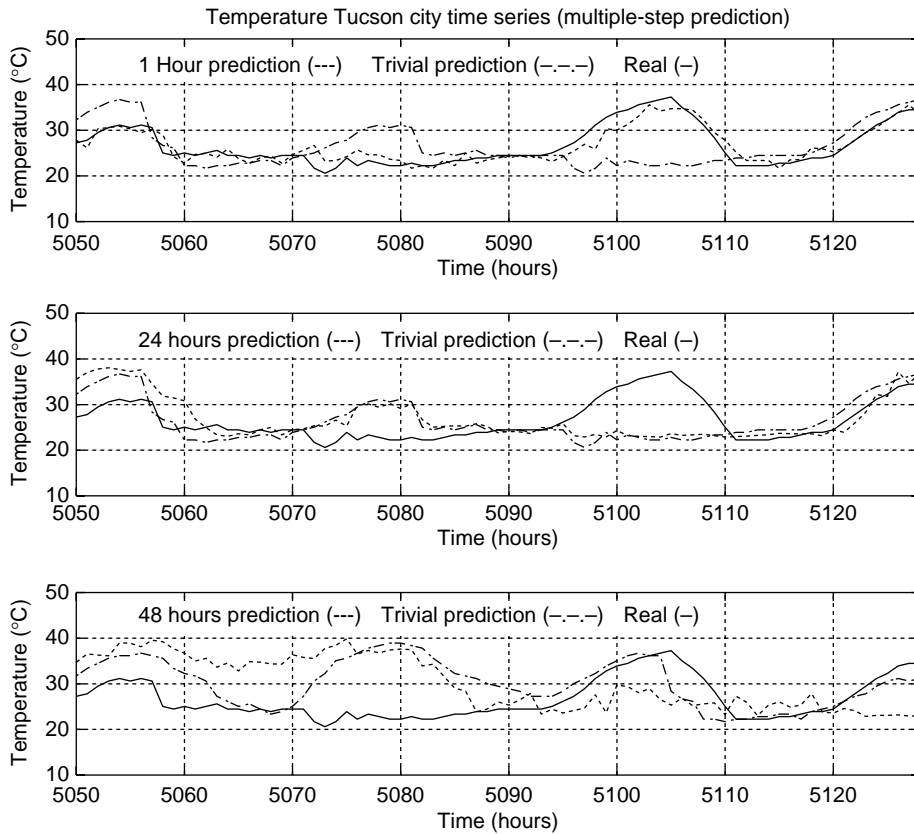
Figure 13.   Tucson weather – predictions.

5.   The predictability horizon

How can the information presented up to this point be used to determine the horizon of predictability? Clearly, the predictability horizon is not fixed, but depends on the magnitude of errors that can be tolerated. The more accurate the results have to be, the smaller will be the horizon of predictability.

The curves presented in Figures 8, 10, and 12 can be used in different ways. If the desired time span of a prediction is predetermined, such as in a predictive control scheme, the information provided in these figures can be used to determine how large the average errors will be that result from such a demand. On the other hand, if the tolerated average errors are given, the graphs can be used to determine the maximum time span of the prediction, i.e. the horizon of predictability.

Unfortunately, the errors cannot be computed at the time the predictions are being made, but only at the time for which the predictions are being made, which is too late. However, confidence values can be estimated at the time when the prediction is being made.

For any given example, the average errors can be correlated with the corresponding confidence values. Considering the example of the Tucson weather prediction, Figure 12 shows that a prediction over 5 h corresponds, on average, to an error of 30% (using the error formula proposed in this paper). It also corresponds to an average similarity value of 0.92. This information can be turned around. A prediction can be said to be acceptable if its
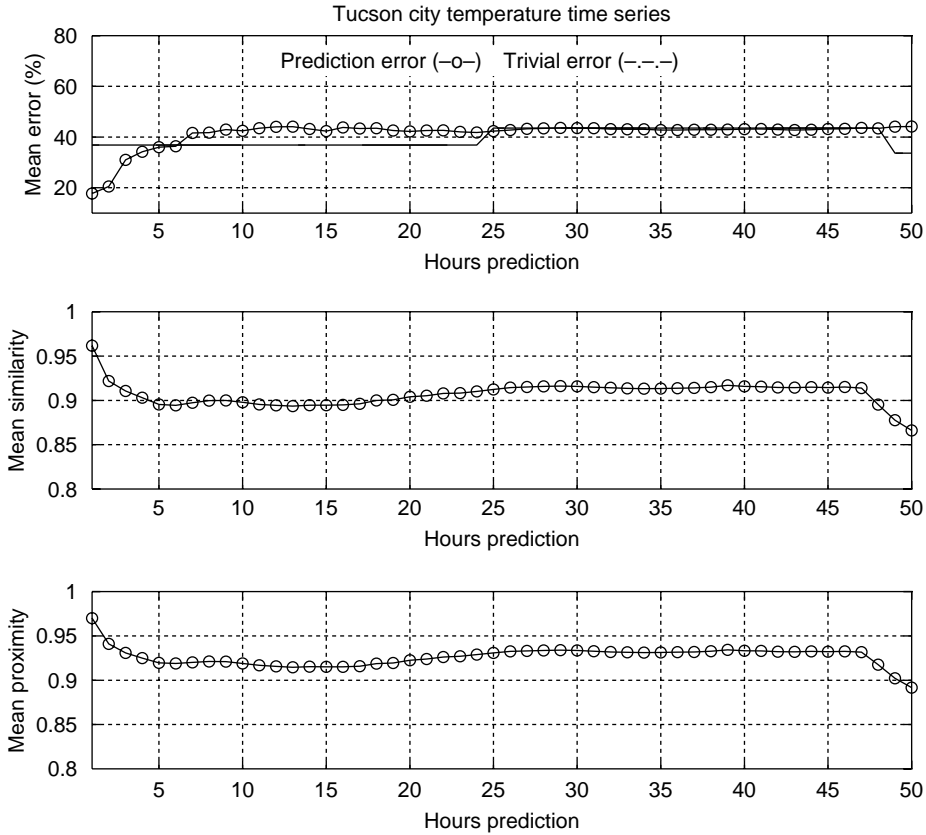
Figure 14.   Tucson weather predictions – mean errors and confidence values using a suboptimal mask.

accompanying similarity confidence value is above 0.92. Otherwise, the prediction must be rejected as being too speculative.

6.   Conclusions

This paper has shown a systematic means to evaluate the horizon of predictability of a time series making use of local as well as accumulated confidence measures that can be obtained in parallel with the prediction of the time series itself using the FIR methodology.

FIR was applied to three different time series representing the water demands of regions of the cities of Barcelona and Rotterdam, and the ambient temperature of the city of Tucson, respectively.

The predictions obtained were not breathtaking. FIR performed rather well on the Barcelona series that exhibits an almost periodic behaviour. It outperformed the trivial predictor only by a moderate amount in the cases of the two more stochastic time series. Even the Barcelona series could have been predicted almost as well using a slightly different trivial predictor. Because of the high weekly correlation, it would make sense to predict today's water demand to be the same as 7 days ago rather than 1 day ago. Once this possibility has been recognized, the predictions might be even further improved by somewhat more elaborate (less trivial) predictors. For example, it might make sense

to predict today's water demand as a weighted sum of the water demands of the last week. The factors could for instance be taken from the autocorrelation function, normalizing their sum to a value of 1.0. Alternatively, one could use FIR to calculate the qualities of the masks of complexity 2 that have their input 1, 2, ... days back, and use the normalized mask qualities as the mixing factors. Finally, one could use a neural network to train these factors such that the forecasting error over the training data is minimized. There are as many options for developing reasonable predictors as one may dream of.

The reader should not expect any miracles, neither of FIR nor of any other prediction technique. A time-series predictor can only exploit the information that it is provided with. For example, if a random number generator is used to produce a 'time series', neither FIR nor any other technique will ever be able to predict anything but noise, as does the trivial predictor. Thus, if FIR outperforms the trivial predictor only by a small amount for a given series, this may be because there is very little signal underneath the noise.

What if another technique can predict a series considerably better than FIR? Even such a result would not be truly surprising. For example in Weigend and Gershenfeld (1994), a prediction of the Lorenz attractor was published that was amazingly accurate (Wan 1994). Yet, if the three parameters of the Lorenz attractor system are only slightly modified and the advocated approach is repeated, the method does not provide a particularly good estimation any longer. Wan had simply been lucky – after all, someone always wins the lottery. Only if another technique can be found that consistently outperforms FIR on a large number of different time series, one would have to conclude that FIR does not exploit all the information that it is being presented with.

The authors are convinced that FIR actually does its job, and does it rather well. There are many indications that FIR indeed exploits all the information that it is being given, and it does so in a rather robust fashion, i.e. it does not require any fine-tuning in order to make decent predictions. Yet, the same holds for many other reasonable prediction techniques, i.e. there are many ways in which the available information can be exploited to lead to equally reasonable predictions. What makes FIR unique is not its prediction performance. It is the robust and flexible fashion in which it deals with a variety of different model identification problems, its elegant way to cope with multiple correlated time series (not shown in this article), and its compelling way of computing a local estimate of the confidence it has in its own prediction.

Acknowledgements

Notes on contributors

Josefina López-Herrera obtained her PhD degree in Advanced Automation and Robotics from the Polytechnical University of Catalonia, Barcelona, Spain, in 1999. From 1988 until 2001, Dr López worked as a project leader in several informatics companies in Europe. She is currently Assistant Professor of Software Engineering at the Polytechnical University of Catalonia, Terrassa, Spain. Her research interests concern fuzzy systems, inductive reasoning and software design for advanced computational science.

François E. Cellier received his BS degree in Electrical Engineering in 1972, his MS degree in Automatic Control in 1973, and his PhD degree in Technical Sciences in 1979, all from the Swiss Federal Institute of Technology (ETH) Zurich. Dr Cellier worked at the University of Arizona as Professor of Electrical and Computer Engineering from 1984 until 2005. He then returned to his home country of Switzerland, where he is now working at his alma mater of ETH Zurich. Dr Cellier's main scientific interests concern modelling and simulation methodologies, and the design of advanced software systems for simulation, computer-aided modelling, and computer-aided design. Dr Cellier has authored or co-authored more than 250 technical publications, and he has edited several books. He published a textbook on Continuous System Modeling in 1991 and a second textbook on Continuous System Simulation in 2006, both with Springer-Verlag, New York.

Gabriela Cembrano received her MS and PhD degrees from the Polytechnical University of Catalonia (UPC) in 1984 and 1988, respectively. Since 1989, she has been a tenured researcher of the Spanish National Research Council (CSIC) at the Institute of Robotics and Industrial Informatics. Her main research area is control engineering and she has been involved in industrial projects on modelling and optimal control of water supply, distribution, and urban drainage systems since 1985. Currently, Dr Cembrano is also a member of CETAQUA – a water technology centre funded jointly by the water company Agbar, by UPC, and by CSIC – as head of one of four main research lines of the centre. She has taken part in several Spanish and European research projects in the field of advanced control and especially its application in water systems. She is now the main researcher of project ITACA (Integration of Advanced Techniques for Modelling, Supervision and Control in the Water Cycle) funded by the Spanish Ministry of Science and Education. She has published and co-authored various journal and conference papers in this field.

References

Aigües de Barcelona, S., 1987. Water demand data for the city of Barcelona. Technical report, Passeig de Sant Joan 39, Barcelona, Spain.

Baggelaar, P., 1992. Voorspelling van het Dagverbruik in het Voorzieningsgebied van de Berenplaat. *H_2O*, 25 (3), 71–74.

Cellier, F.E., López, J., Nebot, A. and Cembrano, G., 2010. Confidence measures for predictions in fuzzy inductive reasoning. *International journal of general systems*, 39 (8), 839–853.

Europoort, W., 1996. Water demand data for the city of Rotterdam. Technical report, P.O. Box 59.999, 3000 RA Rotterdam, The Netherlands.

Griñó, R., 1992. Neural network for univariate time series forecasting and their application to water demand prediction. *Neural network world*, 2 (5), 437–450.

López, J., Cembrano, G. and Cellier, F.E., 1996. Time series prediction using fuzzy inductive reasoning. *Proceedings ESM'96, European Simulation MultiConference*. Ghent: SCS Europe BvbA, 765–770.

Quevedo, J., Cembrano, G., Valls, A. and Serra, J., 1988. *Time series modelling of water demand – a study on short-term and long-term predictions*. Vol. 1. New York: John Wiley, 268–288.

Wan, E.A., 1994. Time series prediction by using a connectionist network with internal delay lines. *In*: A.S. Weigend and N.A. Gershenfeld, eds. *Time series prediction: forecasting the future and understanding the past*. Reading, MA: Addison-Wesley, 195–217.

Weigend, A.S. and Gershenfeld, N.A., eds, 1994. *Time series prediction: forecasting the future and understanding the past*. Reading, MA: Addison-Wesley.