Complex Systems **8** (1994) 25–39

Analysis and Forecasting of Time Series by Averaged Scalar Products of Flow Vectors

C. Santa Cruz
R. Huerta
Instituto de Ingeniería del Conocimiento,
Universidad Autónoma de Madrid,
Canto Blanco, Mod. C-XVI, P.4, 28049 Madrid, Spain

J.R. Dorronsoro
Vicente López
Departamento de Ingeniería Informatica,
Universidad Autónoma de Madrid,
Canto Blanco, 28049 Madrid, Spain

Abstract. The relationship between the quality of state space reconstruction and the accuracy in time series forecasting is analyzed. The averaged scalar product of the dynamical system flow vectors has been used to give a degree of determinism to the selected state space reconstruction. This value helps distinguish between those regions of the state space where predictions will be accurate and those where they are not. A time series measured in an industrial environment where noise is present is used as an example. It is shown that prediction methods used to estimate future values play a less important role than a good reconstruction of the state space itself.

1. Introduction

Much work has recently been done in nonlinear time series prediction [1–4]. However, most of this effort has been focused on time series originating from dynamical systems where there is a well-defined, although complex and often analytic, underlying model. For such systems, a great number of different techniques have been developed to tackle the prediction problem. These techniques include state space reconstruction of chaotic systems [5–7] as well as various methods to approximate future values from either local [2, 8] or global models [1, 9].

State space reconstruction is aimed at obtaining a trajectory in the state space leading to a deterministic reconstruction of the time series. This is usually accomplished using Takens' theorem [10], which states that if the underlying dynamical system has dimension d, the reconstruction, usually called

an embedding, can be carried out using delay vectors in an m-dimensional space ($m \geq 2d + 1$). Unlike the behavior of irregular time series, the state space usually demonstrates simplicity and regularity. Vector fields of the state space are approximated to estimate future values. The quality of the predictions strongly depends on the quality of the state space reconstruction. Furthermore, any lack of accuracy or defect in the state space reconstruction has consequences for the prediction of time series. Therefore, a measure of the quality of a state space reconstruction gives an estimate of the goodness of the forecasting model. The averaged scalar product P of the dynamical system flow vectors [5] has recently been used to determine optimal state space reconstruction by maximizing the P-value as a function of both the state space dimension and the time delay. The P-value can be used to qualify the "degree of determinism" of state space reconstruction; that is, it provides a measure of how parallel the local flow vectors are throughout the state space. It has been shown [5] that for time series originating from ordinary differential equations the P-value can be raised close to unity. This value indicates that the neighboring flow vectors in the reconstructed state space are parallel to each other. On the other hand, neighboring flows for time series contaminated with noise or originating from ill-defined systems will be far from parallel. This leads to lower P-values. This means that the systems are less deterministic than the former, since the flow of points along the state space are not so well-defined. Hence, the prediction will not be as accurate.

This work highlights the existing relationship between the characteristics of state space reconstruction of time series and the quality of its predictions in light of results derived from the P-value. Some of the conclusions drawn from state space analysis apply to both, substantiating the results obtained in the prediction and devising new prediction algorithms.

As an example, we will use the state space reconstruction and time series prediction of a variable measured in an industrial environment. This time series corresponds to an important temperature measured in a petrochemical plant. Forecasting is one of the techniques that has been integrated into the HINT project (ESPRIT) for intelligent control. The series is sampled every five minutes and the time history over three months (25,000 data points) is used. Figure 1 illustrates 20 hours of the actual time series and the filtered signal obtained after removing the high frequences using a low-pass filter. The filtered signal will be used throughout this paper.

Since the time series is measured in an industrial setting where a great number of uncontrolled factors are no doubt present, the underlying model is not well-defined. This lack of accuracy in the definition of the system leads to various problems in both state space reconstruction and the quality of the prediction that would otherwise not appear in situations where a well-defined model existed.

The article is structured as follows. Section 2 describes problems that arise when performing state space reconstructions of real time series. Section 3 analyses the forecasting of time series using results obtained in the previous section, and finally, section 4 summarizes the conclusions of this paper.

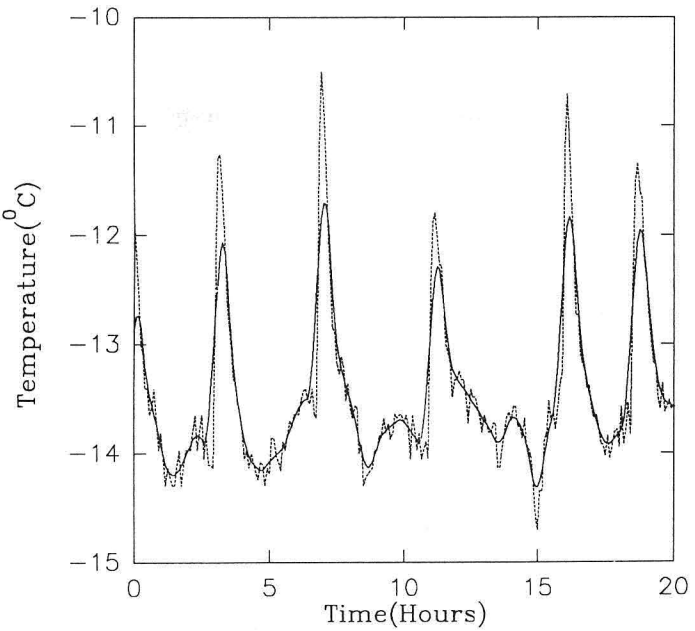

Figure 1: Variation of the original time series (dashed line) versus a filtered time series (solid line) as a function of time. Temperature is measured in Celsius and is sampled every five minutes.

2. State space reconstruction

State space reconstruction is necessary before applying forecasting methods. This analysis provides a framework in which to make state space reconstructions, which enables us to approximate the flow vectors.

State space reconstruction starts with a univariate time series x_t such as temperature values. This series is transformed into a vector $\mathbf{x}_t$. A map that transforms x_t into $\mathbf{x}_t$ must be chosen in such a way that points and flows maintain their individuality in the reconstruction: that is, a deterministic and smooth reconstruction is desired. The most accessible and robust method to perform state space reconstruction is the delay-coordinate map [6] given by $\mathbf{x}_t = (x_t, x_{t-\tau}, \ldots, x_{t-(m-1)\tau})$ where τ is the time delay and m is the state space dimension. Hence, the map is parameterized by two variables, τ and m that are chosen so that a deterministic reconstruction is obtained. Takens [10] showed that if d is the dimension of the underlying dynamical system, an embedding of the original time series x_t can be obtained in a m-dimensional space where $m \geq 2d+1$ and τ is arbitrary. However, the presence of noise in any experimental measure makes it difficult to realize the result of the theorem. In practice, we are faced with an optimization problem that pursues the ideal $\{\tau, m\}$-parameters to obtain the best system reconstruction. The state space points are rotated using the principal component axis and

each coordinate is scaled to the standard deviation of the new axis [5, 11]. The new state space points are called $\mathbf{z}_t$.

The usual method of determining the dimension of the state space reconstruction is by means of the correlation dimension D_2. This dimension is a lower limit for the box counting dimension D_0. According to Mañé [12], $m > 2D_0 + 1$ is a good estimate, but Sauer et al. [7] suggest that $m > D_0$ is typically sufficient (see Ott [13] for a review). We calculate the correlation dimension in rotated and scaled coordinates $\mathbf{z}_t$ as follows:

$$D_2 = \lim_{\epsilon \to 0} \frac{\log_2 C(\epsilon)}{\log_2 \epsilon}, \tag{1}$$

where $C(\epsilon)$ is the correlation integral

$$C(\epsilon) = \lim_{N \to \infty} \frac{1}{N^2} \sum_{i \neq j}^{N} \theta(\epsilon - \|\mathbf{z}_i - \mathbf{z}_j\|) \tag{2}$$

and $\theta(\cdot)$ is the Heaviside function, N is the number of points in the time series, and $\|\cdot\|$ is the Euclidean norm.

Experimentally, D_2 is obtained by studying the convergence of $\log_2 C(\epsilon)$ versus $\log_2 \epsilon$ as a function of m. In Figure 2, $\log_2 C(\epsilon)$ is shown as a function of both $\log_2 \epsilon$ and m. The plateau observed in the figure is due to the limited number of points within a sphere of small radius for large dimensions. When the ϵ-value is increased, the number of points within the sphere also increases and a more accurate value of $C(\epsilon)$ is obtained. In Figure 3, the slopes of various $\log_2 C(\epsilon)$ versus $\log_2 \epsilon$ curves are plotted as a function of m and $\log_2 \epsilon$. We see from this figure that there is no clear convergence to a constant slope. The only thing we can be sure of is that $D_2 > 8$, that is, we need at least eight dimensions for the state space reconstruction ($m > 8$).

In conclusion, this analysis does not yield solid evidence with which to model the system. It should be noted that the correlation dimension D_2 is an indirect measure of the embedding dimension and does not take into account the dynamics of the time series, but only the points. Moreover, this time series corresponds to an ill-defined system. These problems account for the unreliable estimate obtained for the underlying dynamical system of this series.

On the other hand, a deterministic reconstruction can be obtained using the averaged scalar product P. This value gives a measure of the number of self-crossings and tells how smooth the reconstruction is. A value close to unity implies a highly deterministic and smooth reconstruction, whereas a value close to zero implies a completely random reconstruction. The P-value is calculated by averaging the scalar product between the flow vector of a trajectory point and the normalized flow vectors inside a ball of radius ϵ for all points of the trajectory. The P-value is calculated as follows (see [5] for a more detailed description).

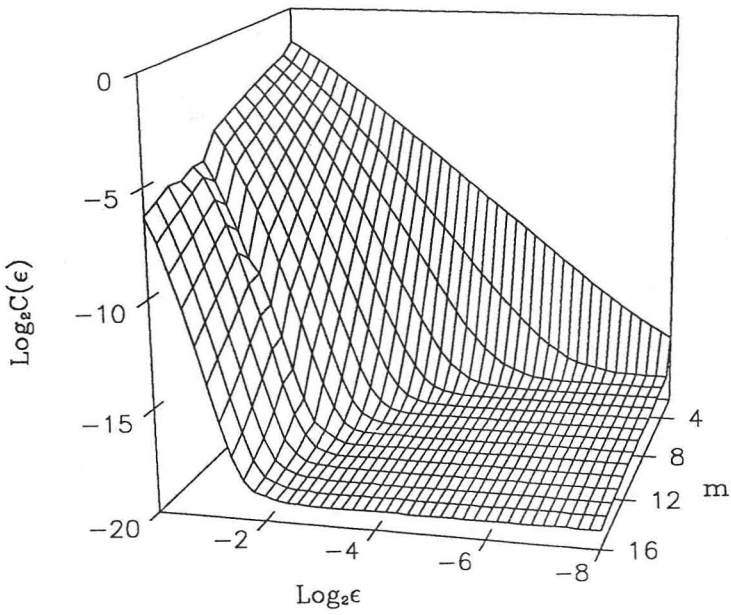

Figure 2: Three-dimensional plot of the variation of $\log_2 C(\epsilon)$ as a function of both $\log_2 \epsilon$ and the state space dimension m.

Define $O(\mathbf{z}_i)$ as the set of neighboring points belonging to a closed ball of dimension m and radius ϵ, that is,

$$O(\mathbf{z}_i) = \{\mathbf{z}_j : \|\mathbf{z}_i - \mathbf{z}_j\| \leq \epsilon, j = 1, \ldots, N, j \neq i\} \tag{3}$$

where μ_i is the number of points in $O(\mathbf{z}_i)$. The normalized flow vectors at each point are calculated as

$$\mathbf{f}(\mathbf{z}_j) = \frac{\mathbf{z}_{j+1} - \mathbf{z}_j}{\|\mathbf{z}_{j+1} - \mathbf{z}_j\|} \tag{4}$$

and the mean flow vector value $\mathbf{V}_i$ inside the set $O(\mathbf{z}_i)$ is given by

$$\mathbf{V}_i = \frac{1}{\mu_i} \sum_{\mathbf{z}_j \in O(\mathbf{z}_i)} \mathbf{f}(\mathbf{z}_j). \tag{5}$$

Finally, the scalar product is averaged along the trajectory using the expression

$$P = \frac{1}{N} \sum_{i=1}^{N} \mathbf{f}(\mathbf{z}_i) \cdot \mathbf{V}_i. \tag{6}$$

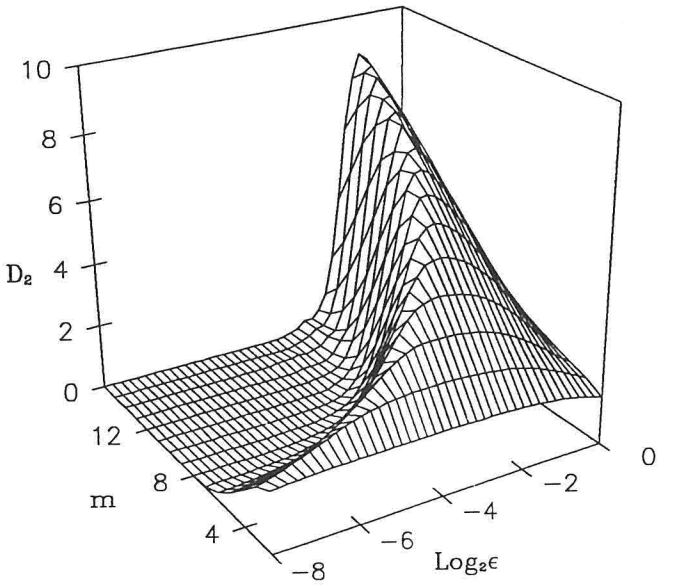

Figure 3: Three-dimensional plot of the variation of the slopes of the various curves of Figure 2 as a function of both $\log_2 \epsilon$ and the state space dimension m.

Think of the above calculations as a trip along the trajectory within a tube of radius ϵ, averaging the scalar product with all the flows inside this tube.

Since this time series has some isolated points for either small radius ϵ or high dimension m, we arbitrarily fix a default value of $\mathbf{f}(\mathbf{z}_i) \cdot \mathbf{V}_i = 0.2$ when no average is performed due to the lack of points within the sphere. This serves to discriminate between the spurious results due to isolated points and the correct P-values. If P converges to 0.2 as m increases, we must increase ϵ in such a way that the number of points does not vanish for large m.

Figure 4 illustrates the variation of P for the original time series as a function of both m and τ. A value of $\epsilon = 0.2$ is used. (Recall that a P-value close to unity means that the system is highly deterministic, while a P-value close to zero indicates a random system.) A maximum P-value ($P = 0.55$) is obtained for $m = 4$ and $\tau = 1$. From this result, we conclude that the time series is not too deterministic. This P-value contrasts with those obtained for systems originating from ordinary differential equations where P-values close to unity have been obtained [5]. The reason for this is twofold. On the one hand, there may be noise (high frequencies) coming from the sensor used to measure the time series. On the other hand, there are many uncontrolled factors that make the system ill-defined and give rise to random sequences in the time series.

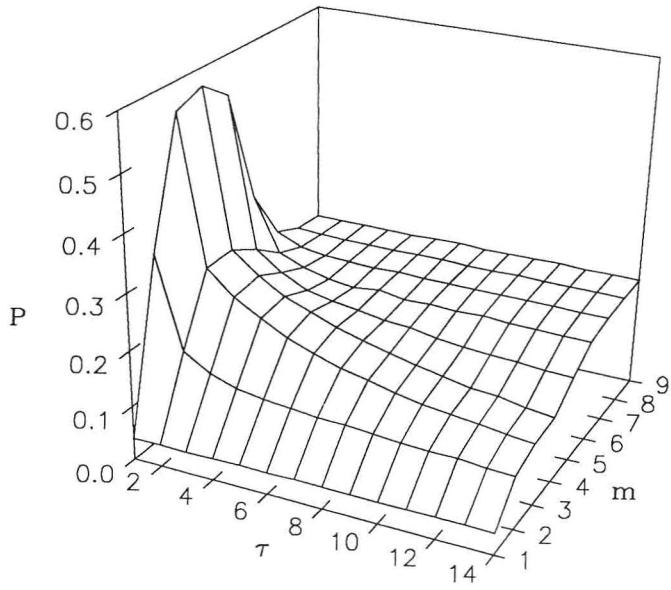

Figure 4: Three-dimensional plot of the variation of the averaged scalar products P of the original time series as a function of both the state space dimension m and time delay τ.

If we want to remove the noise due to the presence of high frequencies a low-pass filter may be used. The use of such a filter, however, has two disadvantages. First, the filtered signal obviously corresponds to a different system. Secondly, these filters introduce a delay in the signal [14] inversely proportional to the passband width that may be important in real time applications. Despite these disadvantages, a low-pass filter was used to remove high frequencies.

In Figure 5, we display the P-value for the filtered time series as a function of τ for different m-values. A value of $\epsilon = 0.2$ was used in the calculation of P. The maximum value is obtained for $m = 4$ and $\tau = 1$ or 3. Although the maximum P-value is nearly the same for both the original and filtered time series, the average P-value is much larger in the filtered series than in the original one.

A careful inspection of the reconstruction gives better insight into what is happening in this time series. Figure 6 shows a plot of the second coordinate as a function of the third for a 4-dimensional state space reconstruction. The coordinate order is the one given by the principal component transformation. From this figure we conclude that there are two well-defined regions. The first one is an inner region corresponding to a sphere of radius $r \approx 1.0$ centered at the origin that is characterized by a great number of crossing trajectories. This region corresponds to the mean value of the time series and the excessive number of crossing trajectories might be due to the pres-

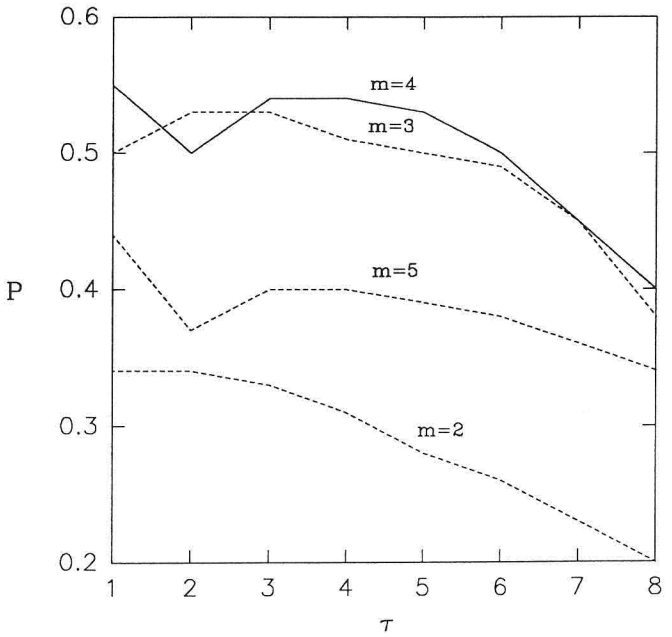

Figure 5: Variation of the P-value of the filtered signal as a function of τ for different state space dimensions, ranging from $m = 2$ to $m = 5$.

ence of a higher-dimensional movement or to random fluctuations around the mean, that is, noise. The second region, the outer one, is characterized by a low level of crossing trajectories. This observation makes clear the need to perform two different estimates of P, one inside and another outside the sphere. In Figure 7, one can see that in the outer region the P-value reaches its maximum value for $m = 3$ while in the inner region a value of $m = 6$ is obtained. Unlike the inner region, trajectories in the outer region are highly parallel leading to a more deterministic reconstruction. This implies that there is a more deterministic system in the outer region of the state space, suggesting the possibility that the system could be modeled with different dimensions depending on the region where the points are located.

3. Forecasting

In this section, we analyze the inherent prediction limitations of this particular series in light of the results obtained from the state space reconstruction. The quality of the prediction is expressed in terms of the following value:

$$E = \frac{\sum_t (x_t - \hat{x}_t)^2}{\sum_t (x_t - \bar{x}_t)^2}. \tag{7}$$

This ratio is the normalized mean-squared error whose value is unity when each of the predicted values correspond to the average. A ratio above unity

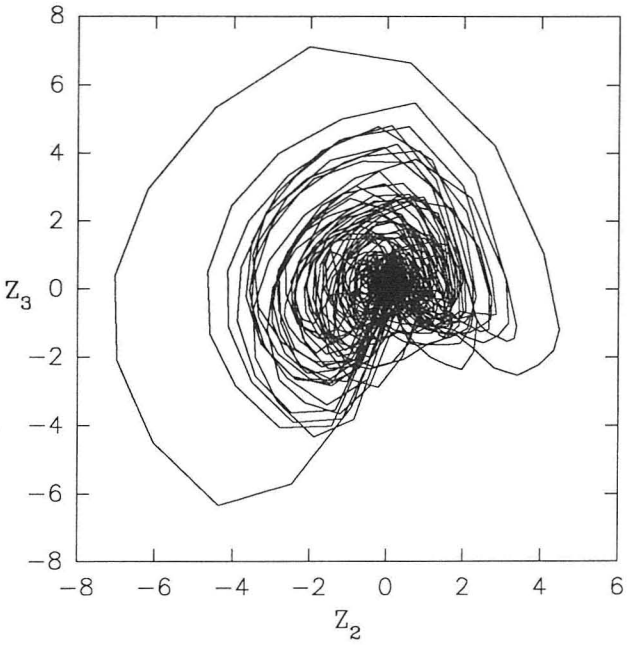

Figure 6: Variation of the second coordinate of the state space reconstruction z_2 as a function of the third coordinate z_3 for a Four-dimensional state space reconstruction obtained from the filtered time series.

corresponds to a prediction that is worse than the average; a ratio below unity is an improvement over the average.

For a given point in the time series x_t we wish to predict the next n points into the future $\hat{x}_{t+k}$, $k = 1, \ldots, n$, where n is the prediction horizon. These estimates are determined using the previous points in the time series x_i, $i = 1, \ldots, t$, which are then used to build the state space reconstruction. We use both the predicted values $\hat{x}_{t+k}$ and the actual values x_{t+k} to determine the E-value and thus to measure the performance of different models.

We see that the characteristics of the state space derived using the P-value (6), have their counterparts when examining the prediction behavior. These results highlight the relationship between the P-value and the prediction error E.

First of all, we determine the optimum state space dimension using the prediction error. In addition to the methods used in the previous section to determine the state space dimension, another possibility is to determine the performance of the reconstruction when estimating the time series prediction error [1]. In order to do this, we choose the simplest prediction method, that is, a local approach using linear predictors [2, 8].

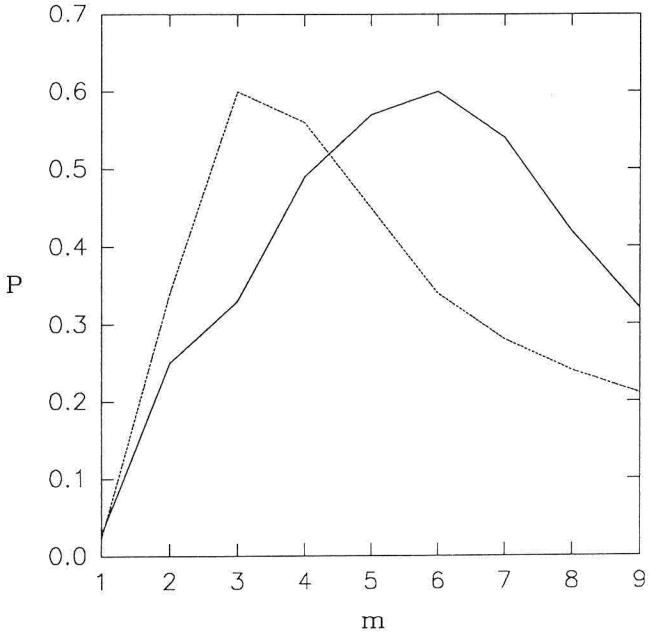

Figure 7: Variation of the P-value as a function of the state space dimension for both those points that lie outside a sphere of unit radius (solid line) and those that are inside (dashed line).

The procedure used to determine the prediction is described as follows. For each point of the time series, locate 50 nearby m-dimensional points in the state space. The prediction is obtained by fitting a linear predictor to these 50 points together with the points where they end up after one step into the future. The predicted value is now taken as the new point to be predicted and the procedure is repeated until the prediction horizon n is reached. The linear fit is obtained using singular value decomposition to avoid ill-conditioned situations.

Figure 8 shows the prediction error, calculated for a prediction horizon of $n = 1$, as a function of the state space dimension m. It is seen that the error decreases as m increases, to a minimum value at $m = 4$. That $m = 4$ is the optimum dimension for making predictions is in agreement with the results obtained in the previous section using averaged scalar products. The prediction error E decreases as the state space dimension is increased up to the point where self-intersections are avoided. For this dimension, the best local fit is obtained. However, when increasing the m-value beyond the optimum, even though self-crossing trajectories are thus avoided, points in the resulting high-dimensional space are too distant from each other to ensure a good local approximation.

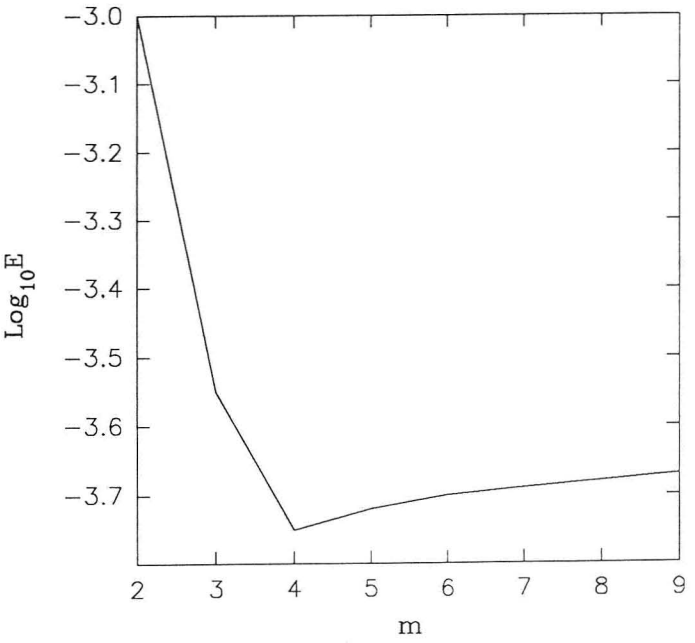

Figure 8: Variation of the prediction error E obtained for a prediction horizon of $n = 1$ as a function of the state space dimension m.

We now examine the consequences this reconstruction has on both short and long-term predictions.

We have used different prediction algorithms to estimate future values of the time series, including both global models (neural networks) and local methods that use both linear and nonlinear approximations. In all cases, the prediction errors were not significantly different. These results highlight the fact that the prediction algorithm was not as important as the state space reconstruction. Consequently, the method previously described was singled out. Although the optimal state space dimension was determined for a prediction horizon of $n = 1$, we assume that this dimension is not far from the optimum for larger prediction horizons.

Figure 9 illustrates short-term predictions ($n = 10$) for several time series compared with the actual values. The conclusions drawn from this figure are twofold. On the one hand, predicted values compare quite favorably with the actual values for those points of the time series that correspond to large peaks. On the other hand, predictions for those values close to the average $\overline{T} = -12.8°C$ are not so good. The reason for this is clear if one takes into account the state space reconstruction. It was shown that the core of the state space corresponds to higher-dimensional movement while the outer region is more deterministic (Figure 6). Therefore, it is difficult to obtain a deterministic reconstruction for the inner region of the state space and,

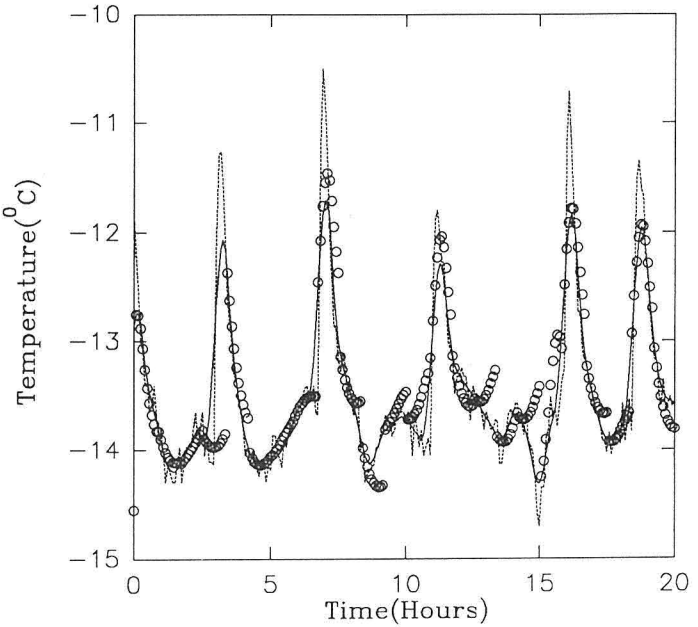

Figure 9: Variation of the original time series (dashed line) and the filtered time series (solid line) along with the predicted values for a prediction horizon of $n = 10$ (circles).

as a consequence, predictions are inaccurate. On the other hand, the outer region of the state space, which corresponds to the various peaks that appear in the time series, can be deterministically reconstructed and are therefore quite reliable. This result can be highlighted if we do not make predictions for those values that are located in the core of the state space. Taking into account the overall time series we find that $E = 0.092$. On the other hand, if only those values of the time series whose state space points lie outside a sphere of radius $r = 1.0$ are predicted, the error decreases to a value of $E = 0.074$. This means that the mean-square error decreases by a factor of 20% if the inner region of the state space (representing 30% of the total number of points) is ignored.

We now tackle the problem of long-term predictions. Figure 10 shows the predictions for $n = 100$ steps into the future compared with the original and filtered time series. In this figure, two complete 100-step predictions are shown. The predictions start at $t \approx 3$h and at $t \approx 11$h, respectively, for a duration of 100 points each. It is seen that both peaks are accurately predicted. These peaks correspond to closed orbits in the state space reconstruction (Figure 6). The remaining peaks of the figure are not accurately estimated due to the fact that when the prediction reaches the average value $\overline{T} = -12.8$°C after approximately 15 steps into the future (which corresponds

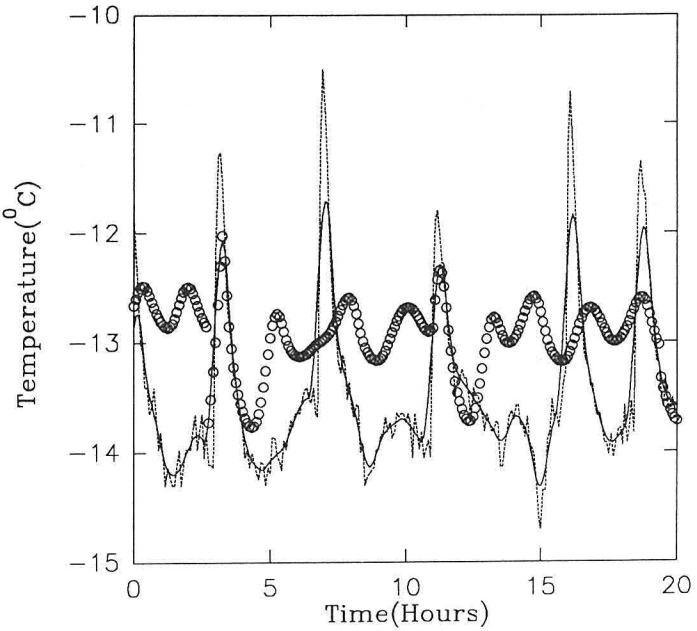

Figure 10: Variation of the original time series (dashed line), the filtered time series (solid line), and the predicted values with a prediction horizon of $n = 100$ (circles).

to entering the inner region of the state space) it can not escape and begins to oscillate. Thus, the prediction horizon for this system is limited by the time required to reach the inner region of the state space. Once there, due to the great number of crossing trajectories, there is no way out of the region, and hence the remaining peaks are poorly estimated.

Driven by the discovery that two well-defined regions with different dimensions are distinguished in the state space, we built a model that takes this into account. In this model, either a three or a six-dimensional state space is used depending on whether the state space point lies outside or inside the core region. The E-value obtained for this prediction model is only 5% better than the former. This slight difference is not significant and suggests that the inner region is mainly dominated by noise instead of being a movement embedded in a higher dimension. The model might be more successful for those systems characterized by a low presence of noise.

4. Discussion

The results obtained suggest that the P-value can be interpreted as a measure of the degree of determinism of the state space reconstruction. Furthermore, the P-value gives a measure of the quality of the reconstruction for differ-

ent regions of the state space that distinguishes between non-deterministic regions dominated by noise and highly deterministic regions.

It has been shown that the results obtained in the state space reconstruction have their counterpart in the prediction error method indicating the similarities between the local averaged scalar product value and the prediction error. This result suggests the possibility of using this value to estimate a local prediction error.

The procedure described in this paper outlines a method for modeling high-dimensional systems where due to the lack of points a bad estimate of D_2 is obtained. However, a degree of determinism and hence a degree of smoothness of the reconstruction may be obtained using the averaged scalar products of flow vectors.

In this work and for this particular time series, different prediction model including global models using neural networks, local models using both non-linear approximations such as neural networks and linear methods have been tested, leading in all cases to almost the same result. However, local models using linear approximations outperform the others in terms of execution time, along with the possibility to use it in adaptive real time systems. This result emphasizes the fact that a good reconstruction is critical, whereas the method used to model the local flow vectors is not so important.

5. Acknowledgments

Grateful acknowledgment is due to the ESPRIT project HINT E6447 for support of this work. One of us (R. H.) wants to thank to the M.E.C. for financial support.

References

[1] M. Casdagli, "Nonlinear Prediction of Chaotic Time Series," *Physica D*, **35** (1989) 335–356.

[2] J. D. Farmer and J. J. Sidorowich, "Predicting Chaotic Time Series," *Physical Review Letters*, **59** (1987) 845–848.

[3] J. D. Farmer and J. J. Sidorowich, "Exploiting Chaos to Predict the Future and Reduce Noise," in *Evolution, Learning and Cognition*, edited by Y. C. Lee (Singapore: World Scientific, 1988).

[4] J. P. Crutchfield and B. S. McNamara, "Equation of Motion from Data Series," *Complex Systems*, **1** (1987) 417–452.

[5] R. Huerta, C. Santa Cruz, J. R. Dorronsoro, and V. Löpez, "State Space Reconstruction Using Averaged Scalar Products of the Flow Vectors," *Physical Review E*, **49** (1994) 1962–1967.

[6] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. J. Shaw, "Geometry from Time Series," *Physical Review Letters*, **45** (1980) 712–716.

[7] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedeology," *Journal of Statistical Physics*, **65** (1991) 579–616.

[8] T. Sauer, "Time Series Prediction by Using Delay Coordinates Embedding," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld (Reading, MA: Addison Wesley, 1993).

[9] V. López, R. Huerta and J. R. Dorronsoro, "Recurrent and Feedforward Polynomial Modeling of Coupled Time Series," *Neural Computation*, **5** (1993) 795–811.

[10] F. Takens, "Detecting Strange Attractors in Fluid Turbulence," in *Dynamical Systems and Turbulence*, edited by D. Rand and L. S. Young (Berlin: Springer, 1981).

[11] D. S. Broomhead and G. P. King, "Extracting Qualitative Dynamics from Experimental Data," *Physica D*, **20** (1986) 217–236.

[12] R. Mañé, "On the Dimension of the Compact Invariant Sets of Certain Nonlinear Maps," in *Dynamical Systems and Turbulence*, edited by D. Rand and L. S. Young (New York: Springer, 1980).

[13] E. Ott, "Chaos in Dynamical Systems" (Cambridge: Cambridge University Press, 1993).

[14] A. V. Oppenheim and R. W. Schafer, "Discrete-time Signal Processing," Prentice Hall Signal Processing Series (Englewood Cliffs, NJ: Prentice Hall, 1989).