Coping with Nonstationarity by Overembedding

Rainer Hegger, Holger Kantz, Lorenzo Matassini, and Thomas Schreiber

Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Straße 38, D 01187 Dresden, Germany

(Received 26 July 1999; revised manuscript received 2 March 2000)

We discuss how nonstationarity in observed time series data due to pronounced fluctuations of system parameters can be resolved by making use of embedding techniques for scalar data. If a $D$-dimensional deterministic system is driven by $P$ slow time dependent parameters, a $(D + P)$-dimensional manifold has to be reconstructed from the scalar time series, which is done by an $m > 2(D + P)$-dimensional time delay embedding. We show that in this space essential aspects of determinism are restored. We demonstrate the validity of the idea heuristically, for numerical examples and for human speech data.

PACS numbers: 05.45.Tp, 43.72.+q, 87.19.Dd

Slight nonstationarity is ubiquitous in laboratory experiments. Quantities such as temperatures, pressures, or electric or magnetic field strengths are known to fluctuate at low amplitudes on long time scales. If really weak, such nonstationarity may be ignored for many purposes of data analysis. The situation is completely different in almost all field measurements: Usually, the system under consideration is coupled to external influences which change in time. Obvious examples are human electrocardiogram recordings, where, e.g., the average heart rate varies according to the physical activity of the body. Almost all known methods for data analysis rely on the assumption of stationarity. In a probabilistic description, this means that all transition probabilities from any state of the system to any other are invariant under time shift. In the framework of nonlinear time series analysis [1], aperiodic time evolution is interpreted in the spirit of underlying deterministic equations of motion with chaotic solutions. Stationarity in this framework means that the system which we observe through the time series creates an invariant measure which is fully sampled by the observations. It includes that all system parameters are kept constant during the measurement episode.

Nonstationarities of different origins can be identified by various methods; recent results include Refs. [2–6]. Once detected, the standard solution is to cut the time series into almost stationary segments and to characterize each segment independently. Since in many systems nonstationarity is an inherent property, this approach often is by far too crude. In many problems, the variability of the dynamical regimes due to nonstationarity might be the essential property. Casdagli [3] has suggested an alternative concept and presented a method to identify the fluctuations of some driving force by the help of recurrence plots. In this Letter, we consider nonautonomous deterministic systems, where the nonstationarity is due to fluctuations of parameters on time scales which are slow compared to the internal dynamics (however, permitting rare sudden changes). We show that all data analysis tasks related to the (implicit) knowledge of the equations of motion, such as predictions, noise reduction, or data classification [7], will be possible by overembedding.

Consider a deterministic dynamical system $\dot{x} = f_p(x)$, $x \in \Gamma \subset \mathbb{R}^D$ in a $D$-dimensional phase space $\Gamma$ or a discrete time version of this. Let the function $f_p$ depend on a set of parameters $p$. Given a measurement function $s(x)$, a time series is a set $\{s_n\}$, $n = 1, \ldots, N$ with $s_n = s(x(t = n\delta))$, where $\delta$ is the sampling interval. The well-known time delay embedding method [8,9] makes use of vectors $s_n$ reconstructed from the scalar observations $s_n$, $s_n = (s_n, s_{n-1}, \ldots, s_{n-(m-1)\delta})$, which represent the state of the system in the sense of uniqueness of the solution, if $m > 2D_f$, where $D_f$ is the fractal dimension of the attractor.

For nonautonomous systems, the embedding theorem is not valid, except for the case that the fluctuating driving terms are known [10,11], a situation not considered here. If unobserved driving terms are time periodic, this driving itself can be interpreted as the limit cycle solution of another deterministic system. Hence, the nonautonomous system can be modeled autonomously by two subsystems with a unidirectional coupling between them. For instance, a sinusoidal driving force, $p = a \cos(\omega t)$, can be created by $(\dot{y}_1, \dot{y}_2) = (y_2, -\omega^2 y_1)$. These two additional variables are confined to the one-dimensional limit cycle solution, and one can reconstruct these $D + 1$ degrees of freedom representing the whole system by a time delay embedding with $m > 2D + 2$. For example, for experimental data from a periodically driven electric resonance circuit, both autonomous four-dimensional maps and two-dimensional periodically driven maps were successfully constructed from the data [12].

When the driving term varies in a nondeterministic way, there is no way to rewrite the equations of motion in an autonomous form. However, in many applications the knowledge of the instantaneous equations of motion, i.e., the equations with the actual parameter settings, is sufficient. In such a situation, under the assumption that parameters vary on much longer time scales than those which rule the instantaneous dynamics, and have only rare additional sudden changes, overembedding solves the problem. As an illustration, we consider the Lozy map, $x_{n+1} = 1 - a|x_n| + bx_{n-1}$. If a time series of $x_n$ is recorded, a two-dimensional embedding permits a unique reconstruction.
of the above equation of motion from the data. If one parameter, say \(a\), changes slowly in time, a two-dimensional embedding will be insufficient, since among the neighbors of a delay vector, \(s_n = (x_n, x_{n-1})\), there are, in general, neighbors with different settings of \(a\) contained in the time series (at least in regions of the \(\mathbb{R}^2\) where the attractors for different \(a\) overlap), and uniqueness of the image is lost. However, the triple \((x_n, x_{n-1}, x_{n-2})\) determines \(a_{n-1}\) by

\[
a_{n-1} = \frac{x_n - 1 - bx_{n-2}}{|x_{n-1}|},
\]

(1)

and, under the assumption that \(a_n \approx a_{n-1}\), makes it possible to predict \(x_{n+1}\) as

\[
x_{n+1} = 1 - \frac{x_n - 1 - bx_{n-2}}{|x_{n-1}|} x_n - bx_{n-1},
\]

(2)

such that close neighbors of \((x_n, x_{n-1}, x_{n-2})\) can be used for the (local linear) reconstruction of this equation from data. Equation (1) proves that similar delay vectors with \(m = 3\) are necessarily related to similar values of \(a\). One can thus use the three-dimensional embedding for a mere selection of neighbors and then use these in an only two-dimensional embedding, if desirable. Equation (2), however, is a universal model of the dynamics of the Lozy map, valid for all quadruples generated with fixed \(a\), irrespective of the actual value of \(a\).

The above reasoning can obviously be generalized for more than one fluctuating parameter and to the situation where we cannot explicitly solve for the parameters. The accuracy to which we want an instantaneous deterministic equation of motion to hold delimited the tolerable rate of parameter variations. When we reconstruct equations of motion from observed data, a residual error remains even for stationary data, so that it appears to be sufficient for all practical purposes that the parameter variation per time step is below the percent level. When a sudden change of parameters occurs, there is, for a few time steps, no instantaneous deterministic rule in the time delay embedding space, and, in particular, we will generally find no neighbors for a delay vector covering the transition between two such different episodes. However, before and after this parameter jump, the above reasoning again applies.

For a numerical verification of the above scheme, we use data from the circle map with drifting parameters, \(x_{n+1} = \Omega(t) + x_n + k(t) \sin x_n\), where \(t\) is a slow time variable as compared to the discrete time \(n\). Figure 1 shows a two-dimensional time delay embedding of \(10^5\) data of the circle map with drifting \(k(t)\) and \(\Omega(t)\). A complex and nonunique superposition of the very different 1D graphs of the circle map for the different \(k\) and \(\Omega\) may be inferred. Moreover, some parameter settings lead to attracting fixed points. Hence, the nonstationary time series is also intermittent. A dimension analysis of these data, embedded in up to eight dimensions, is shown in Fig. 2 and confirms that indeed this cloud of points forms in good approximation a three-dimensional object. Because of the (trivial) linearity of the observable in the state variable \(x\) and in the parameters \(k\) and \(\Omega\) (because of the particular structure of the map), a full reconstruction of the dynamics is possible in three-dimensional delay embedding coordinates. The average one-step prediction errors, normalized to the variance of the data, decay from about 25% in an \((m = 1)\)-dimensional embedding space and \(\approx 12\%\) in \(m = 2\) to a saturation value of \(\approx 2\%\) for \(m \geq 3\) when being computed on the data underlying Fig. 1. The errors are computed by fitting local linear models [14] \(\hat{x}_{n+1} = a_n s_n + b_n\) to the data, where \(s_n\) are \(m\)-dimensional delay vectors of the time series \(\{x_n\}\), and \(a_n\) and \(b_n\) are determined by least squared minimization of the prediction error \(\hat{x}_{k+1} - x_{k+1}\) on a neighborhood of \(s_n\). The deterministic structure is obviously recovered in three and more dimensions, whereas here (two fluctuation parameters) predictability is reduced by the nonstationarity for \(m < 3\).
The heuristics above leads us to the following proposition: If a $D$-dimensional deterministic dynamical system depends on $P$ parameters with slow time dependence, then delay vectors of sufficient embedding dimension are approximately confined to a $(D + P)$-dimensional manifold, which can be reconstructed in an $[m > 2(D + P)]$-dimensional delay embedding space. “Approximately” means that the confinement is violated on length scales of the order of the standard deviation of the data times the average parameter change per time step.

The proposition claims that the way in which parameters vary is irrelevant for the embedding property. In particular, it holds irrespectively of whether the parameters change periodically, have a constant drift, or vary stochastically. The reason for this lies in the slowness of their change: When the time lag of the delay embedding is adjusted according to the time scales of the instantaneous dynamics, it is much too small for a reconstruction of the dynamics of the varying parameters. Delay vectors of solely a single parameter, e.g., $(a_n, a_{n-1}, \ldots, a_{n-m+1})$ of the Lozy map above, are located close to the diagonal in the embedding space because of the strong correlation of the successive $a$ values. Their “attractor” is not at all unfolded and they contribute in high precision only with a one-dimensional subspace (respectively, $P$ parameters with a $P$-dimensional manifold).

A direct comparison between the different approaches, namely, time slicing, ignoring nonstationarity, and over-embedding, is shown in Fig. 3 for an example which is optimal for the slicing method. A causal prediction of the next observation is given by the average of the images of the two closest past neighbors of a given point in embedding space. This scheme is applied to a signal created by the logistic equation $x_{n+1} = 1 - a_n x_n^2$ with $a_n$ alternating between 1.54 and 2 irregularly (upper panel of Fig. 3). The time series was cut into slices of constant $a_n$. Forecasts acting on the single segments suffer strongly from a lack of neighbors (upper curve in lower panel). When

![FIG. 3](image1.png)

**FIG. 3.** Lower panel: the averaged one-step forecast error (fce) of the nonstationary time series shown in the upper panel, caused by the irregular alternation of the parameter $a$ (uppermost curve) in the logistic equation. For further explanation, see text.

Ignoring the nonstationarity (i.e., still $m = 1$) the whole past serves as a data base, but the mis-specification of the model leads to considerable forecast errors (intermediate curve). Only with overembedding ($m = 3$), we can both exploit all the past and resolve the nonstationarity, thus reducing the forecast error by up to 1 order of magnitude (lower curve) when a history of more than 1000 points is available.

The usefulness of our concept of overembedding will now be illustrated for human speech signals. Human speech is highly nonstationary. Inside the logical units, called phonemes, a typical speech signal is almost periodic albeit unharmonic, so that it is reasonable to assume a limit cycle behavior on these episodes. The shape of the wave form changes smoothly and slowly inside a phoneme, but drastically from one phoneme to another (see Fig. 4).

One relevant application which requires the implicit knowledge of the instantaneous dynamics is noise reduction by a local projective algorithm. In stationary chaotic data with measurement noise, the nonlinear noise reduction scheme [15] locally identifies the manifold representing the equations of motion and projects the delay vectors onto this manifold. If the noise on the data is the main contribution to the deviations of the delay vector from this manifold, it is reduced by this projection. In a nonstationary setting, this can work only when we can identify neighbors of the current delay vectors representing the same dynamical state, which are used for the reconstruction of the local linear approximation of the unknown manifold. Figure 4 demonstrates clearly that only neighbors from the same phoneme as the current delay vector represent the same limit cycle. By overembedding, we thus want to distinguish between phonemes. This is found to be possible if a delay vector covers a time interval of about 5–10 ms (one full cycle of the waves of Fig. 4, in contrast to a stationary limit cycle signal, where one would cover only a quarter of a period), and that an embedding dimension of about 25 is required for a reasonable resolution of the dynamics inside such time intervals. Once identified, clean

![FIG. 4](image2.png)

**FIG. 4.** Time series of two different phonemes (top: “ll”, bottom: “j” in “jan”) from the same speaker.
delay vectors of a single phoneme are reasonably well confined to a manifold of about dimension 3. Figure 5 shows a section of a recurrence plot of suitably embedded speech data: The parallel line segments reflect the (slowly changing) period of the wave forms due to the fundamental pitch. More importantly, the gaps in these lines reflect the borders between different phonemes: in a “good” embedding, vectors from different phonemes are no neighbors, although they have the same fundamental period. Nonstationarity is thus resolved without explicit slicing of the time series, and the recurrence plot is a good tool for the verification of this fact.

The noise reduction based on the reconstruction of instantaneous deterministic features, including more technical details, is described elsewhere [16]. It yields a gain in signal-to-noise ratio of about 10–15 dB, and is superior to adaptive spectral filters which are widely used as benchmarks.

Overembedding, i.e., the use of embedding dimensions and embedding windows which are larger than optimal when considering only the instantaneous dynamics and thus would lead to additional redundancy inside a delay vector of a stationary series, solves the problem of nonstationarity in many applications. Data belonging to different parameter settings populate different regions of this embedding space and are thus distinguishable. The instantaneous dynamics can be extracted to be used for short time predictions or noise reduction. One can design methods for data classification to make the parameter changes evident, once one has the separation of delay vectors. Of course, when the parameter variation itself is nondeterministic, these changes cannot be predicted from the data, but in a continuous ongoing prediction the parameters entering our predictor would be implicitly adjusted to the actual values through the selection of the neighbors. For chaotic data, overembedding leads usually to larger interpoint distances on the attractor and thus does not exploit a small data base optimally. This can be seen in Fig. 3 in the lower curve on $i < 200$. However, the effect is overcompensated on longer time series. In summary, the reconstruction of extended state spaces should be the optimal way to deal with drifting parameters, since for dissipative systems it supplies the frame to exploit the combined effect of modified instantaneous rules together with the change of the invariant measure.