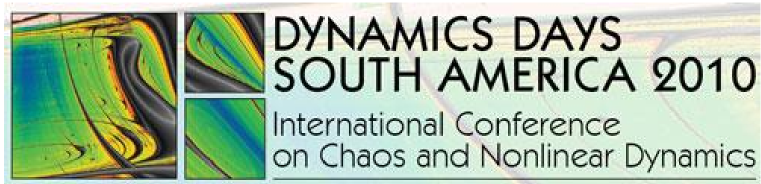

INPE – National Institute for Space Research
São José dos Campos – SP – Brazil – July 26-30, 2010

OPTIMAL IRREGULAR DELAY EMBEDDINGS

Lucas C. Uzal[1,2] *and Pablo F. Verdes*[1,3]

[1]CIFASIS - French Argentine International Center for Information and Systems Sciences UPCAM / UNR - CONICET, Rosario, Argentina
[2]uzal@cifasis-conicet.gov.ar
[3]verdes@cifasis-conicet.gov.ar

keywords: time series analysis, attractor reconstruction, embedding, chaotic dynamics, multiple time scales

In this paper we propose a criterion to select an optimal state-space reconstruction of the dynamics of a physical system from time series. The presented methodology is based on the minimization of a cost function L which is readily computable from the available observational data. Different reconstructions, whether multivariate or time-delayed univariate, regular or not, can be directly compared through L, and thereby the suitability of different embedding settings can be assessed. Optimal parameter values for the embedding dimension and time lag can be therefore simultaneously selected by a global optimization of the proposed cost function. An important advantage of the reported approach is given by its fully automatic and objective character, in contrast to, for example, the subjective practitioner-dependent choices on the location of the first local minimum of Mutual Information or the value of a threshold characterizing a negligible fraction of false nearest neighbours. In particular, in this work we focus on univariate time-delay embeddings and employ the proposed approach to explore the advantages of considering irregular embeddings, as opposed to homogeneous ones, for the reconstruction of dynamical systems.

The proposed cost function L is a local property of the reconstruction, which we then average over the attractor. On Fig. 1 we illustrate the local behaviour of L by plotting a two-dimensional projection of the Mackey-Glass attractor [1] as we reconstruct it in spaces of increasing dimension —from $m = 2$ (panel (a)) to $m = 4$ (panel (c)). As we can see in panel (a), L highlights regions of the attractor not yet unfolded for $m = 2$, i.e. regions where orbits with different dynamical evolutions overlap. These regions progressively vanish in higher dimensional reconstructions, as reflected by lower values of L in panels (b) and (c). As we detail below, L is built upon the concept of noise amplification introduced by Casdagli *et al.* [2].

Delay reconstruction is the fundamental first step of almost all nonlinear time series analysis methods such as the determination of attractor dimensions, Lyapunov exponents, and entropy, to name a few [3]. Takens regular delay-vector definition [4] can be generalized to the irregular case by considering different time delays between consecutive components. The embedding vector is then written as

$$\bar{x}(t) = (x(t), x(t-\tau_1), x(t-\tau_2), ..., x(t-\tau_{(m-1)})), \quad (1)$$

where $\{m, \tau_1, \tau_2, ..., \tau_{(m-1)}\}$ is a set of m parameters to be determined by the practitioner.

The quality of a reconstruction has been quantified by Casdagli *et al.* [2] in terms of its (observational) noise amplification effect when we want to estimate the state of the system. They define the *noise amplification* σ to locally measure this effect, a statistics which can potentially be computed from the observed time series. In principle, σ allows for an absolute comparison among reconstructions, which in turn enables an optimal embedding parameters selection. An obstacle, however, is given by the hypothesis of a full knowledge of the true generating dynamics. In [2] the Authors suggest to estimate σ after performing a data-driven approximation of the dynamical evolution law —a possibility that remained unexplored in the literature and we explicitly pursue in this work.

In this context, the definition of noise amplification (see [2] for details) is given by:

$$\sigma(T, \bar{x}) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \sqrt{Var(x(T)|B_\epsilon(\bar{x}))} \quad (2)$$

where $B_\epsilon(\bar{x})$ is a ball of size ϵ centered at $\bar{x}$.

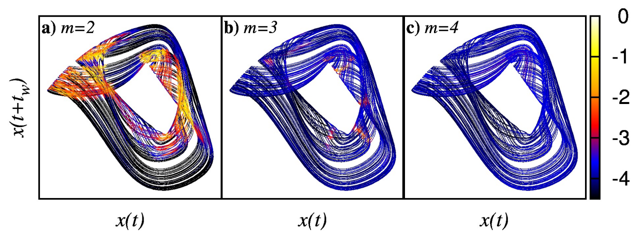

Figure 1 – Color-coded local loss function in a two-dimensional projection of m-dimensional homogeneous time-lag reconstructions of the Mackey-Glass attractor. The time window $t_w = 30$ is the same across panels —as the embedding dimension is increased from 2 to 4, new coordinates are introduced between t and $t + t_w$.

On one hand, $B_\epsilon(\bar{x})$ can be approximated by the set of k-nearest neighbours of $\bar{x}$, in such a way that the conditional variance $Var(x(T)|B_\epsilon(\bar{x}))$ can be re-written as

$$Var_k(x(T)|\mathcal{U}_k(\bar{x})) = \frac{1}{k+1}\sum_{\bar{x}'\in\mathcal{U}_k(\bar{x})}[x'(T)-u_k(\bar{x},T)]^2 \tag{3}$$

where $\bar{x}'$ is a near neighbour of $\bar{x}$, $x'(T)$ denotes the future value associated to it, and

$$u_k(\bar{x},T) = \frac{1}{k+1}\sum_{\bar{x}'\in\mathcal{U}_k(\bar{x})}x'(T). \tag{4}$$

is the average image of the neighbours. On the other hand, to achieve an estimation of $\sigma_k(\bar{x})$ it is necessary to compute a characteristic radius $\epsilon_k(\bar{x})$ of the set $\mathcal{U}_k(\bar{x})$, which we chose as the mean pairwise distance between elements in $\mathcal{U}_k(\bar{x})$. Finally, we introduce an additional normalization factor which measures irrelevance. This factor is sensitive to attractor stretching and allows, in particular, a direct comparison among possibly substantially different trial embedding dimensions. It can also be computed from the nearest k neighbours; however, space limitations prevent us from discussing further details in this short account.

We now report an exploration of the potential of this approach for the construction of irregular embeddings, i.e. the case where consecutive delayed coordinates are not equidistant. We briefly report a case study taken from Pecora *et al.*, who in [5] introduced an irregular embedding construction method and used it to analyse a time series of the x-coordinate of a quasiperiodic, multiple time-scale two-dimensional torus living in a three-dimensional space. Applying their greedy search algorithm, they arrived at a 4-dimensional reconstruction with delay times $\tau_1 = 8$, $\tau_2 = 67$, and $\tau_3 = 75$. The first delay $\tau_1 = 8$ captures the fast frequency, being exactly 1/4 of corresponding period (this fraction is the null autocorrelation lag for harmonic signals). However, $\tau_2 = 67$ slightly fails to capture the slow frequency (it should be $\tau_2 = 63$). We searched over the complete space of parameters $\{m, \tau_1, \tau_2, ..., \tau_{(m-1)}\}$ for the minimum of L. According to our methodology, the optimal irregular delay embedding is attained for parameter values $m = 4$, $\tau_1 = 8$, $\tau_2 = 63$, and $\tau_3 = 71$. In this solution τ_1 captures the fast and τ_2 the slow frequency of this time series. In order to have a clearer picture on how L captures the quality of the reconstruction, in Fig. 2 we show the values of L for the 4-dimensional reconstructions with delays restricted to the plane $\tau_3 = \tau_1 + \tau_2$ (both the solution of Pecora *et al.* and ours belong to this plane of time-symmetrical solutions). Fig. 2 gives a complete picture of all possible time-symmetrical solutions. According to bootstrapping experiments performed on our quality reconstruction measure L on independent sets of random samples, the irregular embedding found by Pecora *et al.* is (statistically) significantly better than the best possible regular embedding. Furthermore, the irregular embedding found by the approach here proposed is in turn significantly better, in a rigorous statistical sense, than the one found by Pecora.

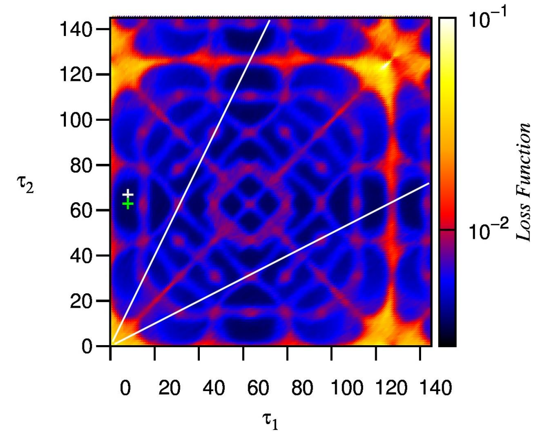

Figure 2 – The loss function L (codified by colors) as a function of the delays τ_1 and τ_2 of an irregular delay reconstruction of the torus time series. The embedding dimension is 4 and $\tau_3 = \tau_1 + \tau_2$ in order to have a time-symmetrical delay vector. The white cross shows the solution found in [5], namely $\tau_1 = 8$, $\tau_2 = 67$, $\tau_3 = 75$. The green cross indicates the solution found by our approach. Finally, the white straight lines indicate uniform delay embeddings.

We have also considered chaotic time series such as the Roessler and Lorenz systems, the Mackey-Glass equation, and experimental data from Chua's circuit. In all cases we have obtained promissing results which, due to space constraints, cannot analyse here but will be communicated to the Conference in the appropriate format.

References

[1] M. C. Mackey and L. Glass, Science **197**, 287 (1977).

[2] M. Casdagli *et al.*, Physica D **51**, 52 (1991).

[3] H. Kantz and T. Schreiber, Nonlinear Time Series Analysis, Cambridge Nonlinear Science Series 7 (Cambridge Univ. Press, Cambridge, 1997).

[4] F. Takens, in *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics **898**, 366, edited by D. A. Rand and L. S. Young (Springer, Berlin, 1981).

[5] L. M. Pecora *et al.*, Chaos **17**, 013110 (2007).