EMBEDDING (1)

• It was suggested first in [1] that nothing really imposed that the vectors of the reconstructed attractor should be composed of equally spaced samples. Actually, what matters is the time span \((m-1)\tau\) covered by the vectors.

• But the approach proposed consists in selecting terms in a linear AR model and retaining those as vector components in the embedding.
EMBEDDING (2)

• In [2] this idea was somewhat amplified. Let us consider now embedded vectors

\[ X_{n-1} = [x(n-i_1), x(n-i_2), \ldots, x(n-i_{\text{max}})]^T \]

with all \( i_k < d \), \( d \) chosen beforehand, and one tries to predict sample \( x(n) \) using some \( G[X_{n-1}] \).

• Selection of the indices amounts to selecting a bit string \( a \) of length \( d \).

• The idea is to select the best possible \( G \) and \( a \) using an MDL approach.
• Recall that the coding length of data + model is:
  \[ L(x, \theta, a) = - \log P(x|\theta, a) + L(\theta) + L(a) \]
  where \( \theta \) is the model parameter vector.

• The log likelihood of the data \( \{x(n)\}, n = 1, \ldots, N \), must include the log likelihood of the initial conditions \( X^{(0)} = [x(d), x(2), \ldots, x(1)]^T \):
  \[ - \log P(x|\theta, a) = - \log P(x|\theta, a, X^{(0)}) - \log P(X^{(0)}) \]
EMBEDDING (4)

- If it is assumed that the errors and the initial conditions are independent and normally distributed, one gets:

\[ L(x, \theta, a) = - \log P(x|\mathcal{N}(0, \sigma^2)) - \log P(X^{(0)}|\mathcal{N}(0, \sigma_X^2)) + L(\theta) + L(a) \]

- Now, in addition to determining the best set of indices, one must also select the best model.
The idea proposed in [2] consists in using the deterministic prediction scheme as the model. Any imbedded vector can be used for prediction. This approach is simple to implement, and it is robust. It does not require any parameter, thus $L(\theta) = 0$. But it is only a pseudo-model, since prediction of $x(n)$ may imply $X_k$, with $k > n$. 
• After some manipulations, and suppression of constant or negligible terms, the MDL criterion is:

\[
\text{MDL}(a) \approx \frac{d}{2} \log \left[ \frac{1}{d} \sum_{i=1}^{d} (x(i) - \bar{x})^2 \right] + \frac{N - d}{2} \log \left[ \frac{1}{N - d} \sum_{i=d+1}^{N} e(i)^2 \right] + d
\]
• Examining all possible combinations of indices up to a maximum value \(d\) means estimating \(2^d\) models. This is not feasible.

• What is proposed in [2] is a sequential selection procedure, i.e.:

  - the first term is the one giving the smallest MDL and kept.

  - All possible 2nd terms are tested and the best one kept if the MDL is smaller …
Unfortunately, this sequential selection does not work as well as for term selection in pseudo-linear models, which amounts to defining the best subspace to project to.

Why not apply a genetic algorithm? Each possible embedding is coded as a bit string, with a ‘1’ if the component is selected, ‘0’ otherwise.
EMBEDDING (9)

- Lorenz system, $x$ coordinate

Sequential and GA:
Same model
Indices 1, 2
(maximum 15)
• Sunspot time series

Sequential
Indices 1, 2, 5
MDL: 839.02

GA
Indices 1, 2, 3, 5, 11
MDL: 837.5

(maximum 15)
• Surrogate on sunspot time series

Sequential
Indices 1, 3, 12, 13
MDL: 931.62

GA
Indices 1, 3, 12, 13
MDL: 931.62

(maximum 15)
EMBEDDING (12)

- RR Intervals (between heartbeats), sampling 2 Hz.

Sequential Indices 1, 3, 6
MDL: 1623.8

GA
Indices 1, 3, 12, 15
MDL: 1597.8

(maximum 15)
• Increase in model size with noise (Henon map, 300 samples)

<table>
<thead>
<tr>
<th>snr (dB)</th>
<th>terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>15</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>10</td>
<td>1, 2, 3, 8</td>
</tr>
<tr>
<td>5</td>
<td>1, 2, 3, 4, 7, 8, 12, 15</td>
</tr>
<tr>
<td>0</td>
<td>1, 2, 3, 5, 7, 9, 10, 11, 13, 14</td>
</tr>
</tbody>
</table>
A technique to assess the interdependence between two multivariate time series, related to deterministic prediction, has been proposed [3,4].

Of course, those multivariate time series $X = \{X_n\}$ and $Y = \{Y_n\}$ may result from the embedding of two signals $\{x(k)\}$ and $\{y(k)\}$.

The idea is to look for the possible existence of (well behaved) functions $G(.)$ and $F(.)$ such that

$$X_n = G(Y_n) \quad \text{and} \quad Y_n = F(X_n)$$
• It is to be noted that interdependence does not imply causation. The relationship between the two series may be due to the influence of a third one on both of them.

• The principle of the test is simple and robust: if indeed $G(.)$ and $F(.)$ exist and are reasonably continuous, then if $Y_n$ is close to $Y_k$, then $X_n = G(Y_n)$ should be close to $X_k = G(Y_k)$.

• Closeness is quantified with respect to a deterministic prediction of $X_n$ using closest neighbors in $\{X_k\}$. 
• Let $r_{n,j}$ and $s_{n,j}$, $j = 1, \ldots, k$, denote the indices of the $k$ nearest neighbors of $X_n$ and $Y_n$ respectively.

• The mean squared distance between $X_n$ and its closest neighbors is:

$$R_{n}^{(k)}(X) = \frac{1}{k} \sum_{j=1}^{k} \| X_n - X_{r_{n,j}} \|^2$$

This quantity should be small for close enough neighbors.
• The conditional mean squared distance with respect to \( \{Y_k\} \) is:

\[
R_{n}^{(k)}(X|Y) = \frac{1}{k} \sum_{j=1}^{k} \|X_n - X_{s_{n,j}}\|^2
\]

• One can define the same quantities for \( Y_n \):

\[
R_{n}^{(k)}(Y) = \frac{1}{k} \sum_{j=1}^{k} \|Y_n - Y_{s_{n,j}}\|^2 \quad R_{n}^{(k)}(Y|X) = \frac{1}{k} \sum_{j=1}^{k} \|Y_n - Y_{r_{n,j}}\|^2
\]
• If the point set \( \{ Y_k \} \) has an average squared radius \( R(X) \):

\[ R_n^{(k)}(X)/R(X) \approx (k/N)^2/m \ll 1 \text{ for } k \ll N \]

with \( m \) the dimension and \( N \) the number of vectors.

• If \( X \) and \( Y \) are indeed interdependent, then the same relationship should hold for the conditional mean square distance.
A measure of local interdependence is:

\[ S_n^{(k)}(X \mid Y) = \frac{R_n^{(k)}(X)}{R_n^{(k)}(X \mid Y)} \quad 0 < S_n^{(k)}(X) \leq 1 \]

And the global measure is:

\[ S^{(k)}(X \mid Y) = \frac{1}{N} \sum_{n=1}^{N} S_n^{(k)}(X \mid Y) \quad 0 < S^{(k)}(X \mid Y) \leq 1 \]

This global measure should be close to 0 when there is no interdependence, and close to 1 for strong interdependence.
INTERDEPENDENCE (7)

• To assess the significance of $S^{(k)}(X|Y)$, one can generate several surrogate data $Y^{(u)}$ for $Y$, compute $S^{(k)}(X|Y^{(u)})$, and perform a rank test.

• If there is indeed a significant interdependence, it is possible to test if it is linked to phase only by generating bivariate surrogate data $\{X^{(b)}, Y^{(b)}\}$, compute $S^{(k)}(X^{(b)}|Y^{(b)})$, and perform a rank test.

• Those bivariate surrogate data are generated by applying the same phase randomization to both signals.
• Recurrence plots (RP) were initially introduced to display the recurrence of patterns and possible non stationarities in imbedded time series [5,6].

• Apart from producing nice pictures, several parameters can be extracted from RP to characterize dynamical processes.

• RP have also been generalized to the analysis of the interdependence between two time series.
• The principle of RP is quite simple: given a multivariate time series \( \{X_n\} \) \( n = 1, \ldots, N \), possibly obtained through the embedding of a signal, an RP is an \( N \times N \) array in which a dot is placed at location \((i, j)\) if \( ||X_i - X_j|| < \delta \), \( \delta \) a predetermined small number.

• Typically, \( \delta \) is a fraction of the sum of standard deviations of the vector components (sstd).

• Of course, RPs are symmetrical with respect to the main diagonal.
• Periodicities (recurrences) are expressed by diagonal lines in RPs. RP-based analysis is mostly based on the characterization of these lines.

• Of course, an RP is dependent upon the threshold $\delta$, but it is also quite dependent upon vector dimension $m$.

• What matters is to conserve these values through the various experiments performed.
• Example: sinusoid with period 20, $m = 1$, $\delta = 0.05$ sstd.
Example: sinusoid with period 20, $m = 2$, $\delta = 0.05$ sstd.
• Example: Lorenz, \([x \ y \ z] \), \(\delta = 0.05\ \text{sstd.}\)
Parameters extracted:

- **REC** proportion of recurrence (black) points. Higher for periodic dynamics.

- **DET** proportion of recurrence points in diagonal lines of length at least 2. Higher for deterministic dynamics.

- **ENT** entropy of line lengths on the diagonal. Higher for more complex dynamics.
• DIV inverse of the length of the longest diagonal line. Higher for larger maximum Lyapunov exponent.

• TREND slope $a$ of the regression line:

$$ R_j = b + ad_j + \varepsilon_j $$

with $R_j$ the proportion of recurrence points on diagonal at distance $d_j$ from the main diagonal. Quantifies the decrease in recurrence with respect to time difference.
Clearly, significance of the values of these parameters in terms of determinism or nonlinearity must be assessed using surrogate data.

Henon

surrogate

REC=0.03 DET=0.86 DIV=0.09 trend=-1.5e-5

REC=0.008 DET=0.24 DIV=0.33 trend=-1.7e-5
• It is also possible to generate RP with two time series (same range, same dimension).
• Example: x and y of Lorenz
REFERENCES


