MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Wayland method can rate some datasets as less deterministic than their random shuffle surrogates

Yoshito Hirata, Shunsuke Horai, Hideyuki Suzuki, and Kazuyuki Aihara

METR 2005–05 February 2005

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.i.u-tokyo.ac.jp/mi/mi-e.htm

Wayland method can rate some datasets as less deterministic than their random shuffle surrogates

Yoshito Hirata[1], Shunsuke Horai[2], Hideyuki Suzuki[1], and Kazuyuki Aihara[3,1,2]

[1] Department of Mathematical Informatics, The University of Tokyo
[2] Aihara Complexity Modelling Project, ERATO, JST
[3] Institute of Industrial Science, The University of Tokyo
yoshito@sat.t.u-tokyo.ac.jp

February 14th, 2005

Abstract

We used the Wayland method, a method for detecting the determinism from a time series, in surrogate data analysis and found a puzzling phenomenon: random shuffle surrogates showed more "determinism" than the original data, which is generated from a purely deterministic system with small noise. We propose herewith that this phenomenon may occur because of observation noise, dynamical noise, or high dimensionality of the system.

Nonlinear time series analysis has greatly progressed for the last two decades. One of the main aims is to find evidence that a given time series is generated by deterministic chaos. There are several key features of deterministic chaos, one of them being determinism in spite of apparent randomness. For detecting determinism from a time series, there are several proposed methods [1, 2]. Among them, a simple and easy-to-use one, which is widely used, is the method of Wayland [2].

We report herewith using an example of wind data a seemingly puzzling phenomenon that the Wayland method regards some datasets as less deterministic than their randomly shuffled surrogates, and discuss some possible reasons for this.

First, we summarize the method proposed by Wayland *et al.* [2] for detecting determinism in a time series. Suppose that a scalar time series $\{s_t\}_{t=1}^{N}$ is given. Let τ be the time lag, and d, the embedding dimension. Then we form delay coordinates x_t by $(s_t, s_{t-\tau}, \cdots, s_{t-(d-1)\tau})^T$. We call $\{x_t\}_{t=(d-1)\tau+1}^{N}$ the experimental attractor.

Choose the number l of integers between $(d-1)\tau+1$ and $N-1$. Call this set of integers T. For x_t of $t \in T$, find the number k of the nearest neighbors from the experimental attractor as follows: Labeling the i-th nearest neighbor for x_t as $x_{n_i(t)}$ ($i \geq 1$), select the nearest neighbors so that $|n_i(t) - t| > \tau$. For notational convenience, we define that $n_0(t) = t$.

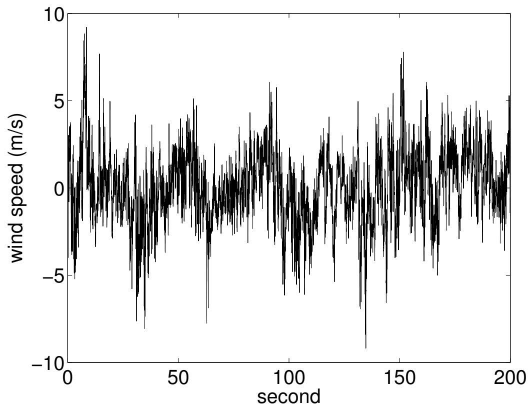

Figure 1: An example of the wind data

For each $x_{n_i(t)}$, we look at its image $x_{n_i(t)+m}$ and take the translation vector,

$$v_i(t) = x_{n_i(t)+m} - x_{n_i(t)}. \tag{1}$$

In this paper we use $m = 1$. To quantify this notion, let us define the average of the translation vectors $v(t)$ as follows:

$$v(t) = \frac{1}{k+1} \sum_{i=0}^{k} v_i(t). \tag{2}$$

Using these vectors, we define the translation error $e(t)$ as

$$e(t) = \frac{1}{k+1} \sum_{i=0}^{k} \frac{\|v_i(t) - v(t)\|^2}{\|v(t)\|^2}. \tag{3}$$

The translation error $e(t)$ yields the fractional spread in the displacements of $x_{n_i(t)}$ relative to the average displacement $v(t)$.

We find the median of $e(t)$ over T, and declare it to be a test statistic for determinism. Wayland *et al.* demonstrated that the test statistic is close to 0 if a time series is deterministic and the test statistic is about 1 if a time series is random. Here, we set $k = 4$ and $l = 100$, and choose τ using the first minimum value of the mutual information [3].

We applied surrogate data analysis with the Wayland statistic to a time series of wind. The wind was measured by an anemometer at 50Hz on the top of the building of Institute of Industrial Science, The University of Tokyo in Komaba, Tokyo, at 14:00 JST, on 18 August 2004, for 1 hour. A part of this time series is shown in Fig. 1. We confirmed using the method of Kennel [4] that this part of the time series is stationary.

For this data, we generated 39 randomly shuffled surrogates [5] and compared them with the original data by calculating the Wayland statistic. Figure 2 shows

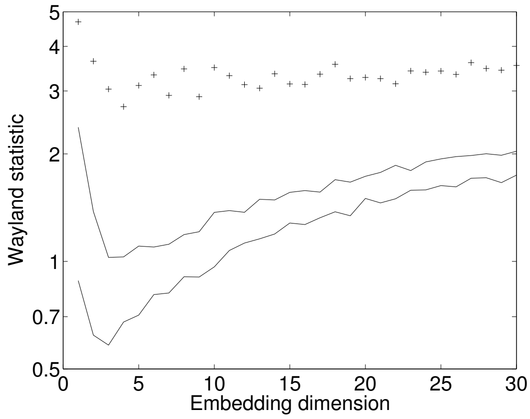

Figure 2: Result of surrogate data analysis. For each embedding dimension, the solid lines show the minimum and maximum Wayland statistics obtained from 39 randomly shuffled surrogates, and symbol $+$, Wayland statistic obtained from the original data.

the result. We observed that for each embedding dimension, the Wayland statistic obtained from the original data is greater than the maximum value of those from the 39 randomly shuffled surrogates. This means that the randomly shuffled surrogates are more deterministic than the original data, that should not be correct.

Judging from the nature of the wind, there are three possible reasons that caused this puzzling phenomenon: observation noise, dynamical noise, and the high dimensionality.

First, we tested whether observation noise caused the reverse rejection. We used the Lorenz'63 model [6], which is defined as

$$\left\{ \begin{array}{lcl} \frac{dx}{dt} & = & -ax + ay \\ \frac{dy}{dt} & = & -xz + bx - y \\ \frac{dz}{dt} & = & xy - cz, \end{array} \right. \tag{4}$$

where $(a, b, c) = (10, 28, 8/3)$. We used the fourth order Runge-Kutta method [7] to integrate the system with a step size of 0.001. During integration, we observed the x-coordinate every 0.01 unit time and obtained a scalar time series with a length of 10 000. We then added observational noise to the data, which follows Gaussian distribution $N(0, (0.1\sigma_o)^2)$ with a mean of 0 and a standard deviation of $0.1\sigma_o$, where σ_o is the standard deviation for the original clean data.

The result of the surrogate data analysis is shown in Fig. 3. Since in all the embedding dimensions the Wayland statistic of the original data is greater than those of the surrogates, observation noise could be the reason for the reverse rejection.

To explain the mechanism of the reverse rejection caused by observation noise, we argue that there is a scaling law between the Wayland statistic and the level of observation noise.

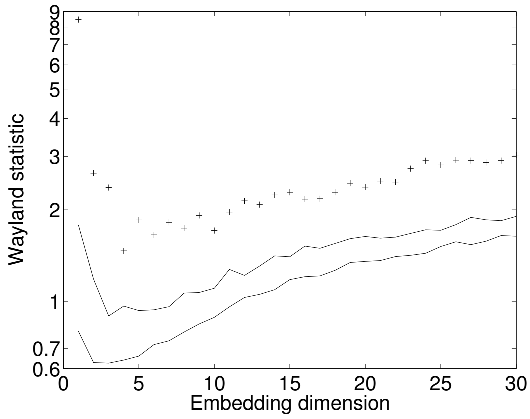

Figure 3: 39 randomly shuffled surrogates are generated for time series of Lorenz'63 model, contaminated by observation noise, and compared using Wayland statistic. The solid lines show the minimum and maximum Wayland statistics obtained from 39 surrogates, and symbol $+$, that of original data.

Suppose that $\{y_t\}_{t=1}^N$ is a clean deterministic scalar time series and contaminated by observational noise $\{\eta_t\}_{t=1}^N$. Therefore, now the observed time series $\{s_t\}_{t=1}^N$ has the relation $s_t = y_t + \eta_t$ for each t. We assume that for each t, the noise η_t follows Gaussian distribution $N(0, \sigma^2)$ with a mean of 0 and a standard deviation of σ.

Let

$$\tilde{v}_i(t) = \begin{pmatrix} y_{n_i(t)+1} - y_{n_i(t)} \\ y_{n_i(t)+1-\tau} - y_{n_i(t)-\tau} \\ \vdots \\ y_{n_i(t)+1-(d-1)\tau} - y_{n_i(t)-(d-1)\tau} \end{pmatrix}, \tag{5}$$

and

$$\tilde{v}(t) = \frac{1}{k+1} \sum_{i=0}^{k} \tilde{v}_i(t). \tag{6}$$

Letting $[\tilde{v}(t)]_j$ and $[v(t)]_j$ the j-th components of $\tilde{v}(t)$ and $v(t)$, respectively, we have $[v(t)]_{(j-1)} = s_{n_i(t)+1-j\tau} - s_{n_i(t)-j\tau} = (y_{n_i(t)+1-j\tau} - y_{n_i(t)-j\tau}) + (\eta_{n_i(t)+1-j\tau} - \eta_{n_i(t)-j\tau})$. Thus

$$v_i(t) = \tilde{v}_i(t) + \begin{pmatrix} \eta_{n_i(t)+1} - \eta_{n_i(t)} \\ \eta_{n_i(t)+1-\tau} - \eta_{n_i(t)-\tau} \\ \vdots \\ \eta_{n_i(t)+1-(d-1)\tau} - \eta_{n_i(t)-(d-1)\tau} \end{pmatrix}. \tag{7}$$

Hence, the average translation $v(t)$ is written as

$$v(t) = \tilde{v}(t) + \begin{pmatrix} \frac{\sum_{i=0}^{k}(\eta_{n_i(t)+1} - \eta_{n_i(t)})}{k+1} \\ \frac{\sum_{i=0}^{k}(\eta_{n_i(t)+1-\tau} - \eta_{n_i(t)-\tau})}{k+1} \\ \vdots \\ \frac{\sum_{i=0}^{k}(\eta_{n_i(t)+1-(d-1)\tau} - \eta_{n_i(t)-(d-1)\tau})}{k+1} \end{pmatrix}. \tag{8}$$

Assume that each $n_i(t)$ is separate from the others in terms of time. Then, for each j, the sum $\sum_{i=0}^{k}(\eta_{n_i(t)+1-j\tau} - \eta_{n_i(t)-j\tau})$ follows Gaussian distribution $N(0, 2(k+1)\sigma^2)$. Therefore $\frac{1}{k+1}\sum_{i=0}^{k}(\eta_{n_i(t)+1-j\tau} - \eta_{n_i(t)-j\tau})$ follows $N(0, \frac{2\sigma^2}{k+1})$. Because $2\sigma^2/(k+1)$ is small if k is large enough, we can approximate it to be

$$\frac{1}{k+1}\sum_{i=0}^{k}(\eta_{n_i(t)+1-j\tau} - \eta_{n_i(t)-j\tau}) \approx 0, \tag{9}$$

and then we have $v(t) \approx \tilde{v}(t)$.

Next we calculate $e(t)$. By definition, we have

$$\begin{aligned} & e(t) \\ \approx\ & \frac{1}{k+1}\sum_{i=0}^{k}\frac{\|v_i(t) - \tilde{v}(t)\|^2}{\|\tilde{v}(t)\|^2} \\ =\ & \frac{1}{(k+1)\|\tilde{v}(t)\|^2}\sum_{i=0}^{k}\sum_{j=0}^{d-1}\Big[([\tilde{v}_i(t)]_j - [\tilde{v}(t)]_j)^2 \\ & +2(\eta_{n_i(t)+1-j\tau} - \eta_{n_i(t)-j\tau})([\tilde{v}_i(t)]_j - [\tilde{v}(t)]_j) \\ & +(\eta_{n_i(t)+1-j\tau} - \eta_{n_i(t)-j\tau})^2\Big], \end{aligned}$$

where $[\tilde{v}_i(t)]_j$ denotes the j-th components of $\tilde{v}_i(t)$. Taking the mean of $e(t)$ over all η_t's, we have

$$E[e(t)] \approx \frac{\sum_{i=0}^{k}(\|\tilde{v}_i(t) - \tilde{v}(t)\|^2 + 2d\sigma^2)}{(k+1)\|\tilde{v}(t)\|^2}.$$

Letting

$$\tilde{e}(t) = \frac{1}{k+1}\sum_{i=0}^{k}\frac{\|\tilde{v}_i(t) - \tilde{v}(t)\|^2}{\|\tilde{v}(t)\|^2}, \tag{10}$$

we obtain

$$E[e(t)] \approx \tilde{e}(t) + \frac{2d\sigma^2}{\|\tilde{v}(t)\|^2}. \tag{11}$$

The translation error $\tilde{e}(t)$ can be different from the statistic we may obtain from a clean time series because nearest neighbors selected can be different due

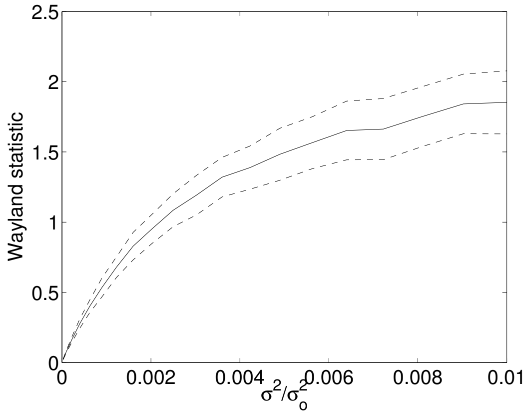

Figure 4: Relation between Wayland statistic and noise level σ of observation noise, in example of Lorenz'63 model. The solid line is mean, the broken lines are mean $\pm$ standard deviation of 100 realizations.

to observation noise. However, if the noise level is moderate, selected points are still the neighbors of x_t, and the value of $\tilde{e}(t)$ is expected to be close to the one we may obtain from a clean time series. Hence, we regard $\tilde{e}(t)$ as approximately equal to the statistic we may obtain from a clean time series, as far as the noise level is moderate.

As $\frac{d}{\|v(t)\|^2} > 0$, the average of $e(t)$ will increase if the variation of the noise increases. The formula Eq. (11) also suggests that we may avoid the increase of $E[e(t)]$ if we make $\|\tilde{v}(t)\|$ greater, or if we use $m > 1$.

To verify the formula in Eq. (11) numerically, we used a time series generated from the Lorenz'63 model. We observed the x-coordinate of the Lorenz'63 model every 0.01 unit time and obtained a time series with a length of 10 000. The tested noise levels σ are $0, 0.005\sigma_o, 0.01\sigma_o, \cdots, 0.1\sigma_0$ where σ_o is the standard deviation for the original time series. To find the Wayland statistic, we used $d = 3$. Figure 4 shows the result. In this example, we observed a linear relation between the Wayland statistic and σ^2, when σ^2 is small. When the noise level is high, the average of the Wayland statistic is greater than one.

Using the scaling law shown in Eq. (11), the reverse rejection can be explained as follows: When applied to data without temporal correlation, or randomly shuffled surrogates, the Wayland statistic tends to be about 1. On the other hand, because of the scaling law in Eq. (11), the Wayland statistic of the original data can be greater than 1.

For the second hypothesis, we tested whether dynamical noise could cause the reverse rejection. We added dynamical noise to the Lorenz'63 model in the following way:

$$\begin{cases} \frac{dx}{dt} &= -ax + ay + l_n\sigma_x\eta_x \\ \frac{dy}{dt} &= -xz + bx - y + l_n\sigma_y\eta_y \\ \frac{dz}{dt} &= xy - cz + l_n\sigma_z\eta_z, \end{cases} \quad (12)$$

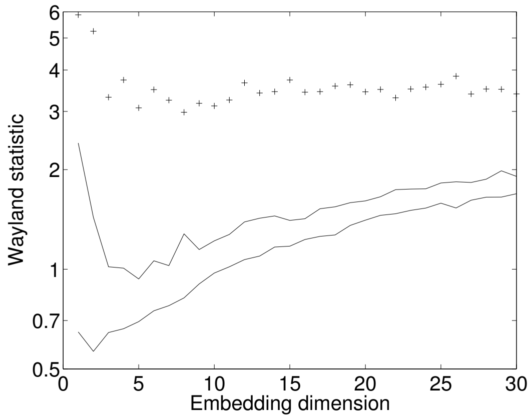

Figure 5: Result of surrogate data analysis for data generated from Lorenz'63 model, contaminated with dynamical noise. For each embedding dimension, symbol $+$ shows Wayland statistic of original data, and the solid lines, the minimum and maximum obtained from 39 surrogates.

where l_n is the level of dynamical noise, σ_x, σ_y, and σ_z are the standard deviation of clean data for each coordinate, and η_x, η_y, and η_z are the normal distribution $N(0, 1^2)$. We set $l_n = 100$, integrated the system using the fourth order Runge-Kutta method [7] with a step size 0.001 and observed the x-coordinate every 0.01 unit time to obtain a scalar time series with a length of 10 000. With this data, we generated 39 randomly shuffled surrogates and compared them to the original data using the Wayland statistic.

The result, shown in Fig. 5, also exhibits the reverse rejection. Therefore, dynamical noise can also be a reason for the reverse rejection.

We found numerically using the Lorenz'63 model that there is a scaling law between the Wayland statistic and the level of dynamical noise. We used the noise level $l_n = 10, 20, \cdots, 100$. For each noise level, we generated 20 time series with a length of 10 000 and calculated the Wayland statistic. With the results shown in Fig. 6, we observed a linear relation between the Wayland statistic and l_n^2 when the noise level is low.

As the scaling law holds, the reverse rejection caused by dynamical noise is similar to that of observation noise.

Lastly, we examined the third hypothesis: whether the high dimensionality of the system causes the reverse rejection. To test this hypothesis, we used Lorenz'96 model [8], which is a toy model of the atmosphere. In the Lorenz'96 model, there are the number I of large scale variables x_i and the number $I \times J$ of small scale variables $y_{j,i}$. Using these variables, the equations are defined as

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F - \frac{h_x c}{b} \sum_{j=1}^{J} y_{j,i} \tag{13}$$

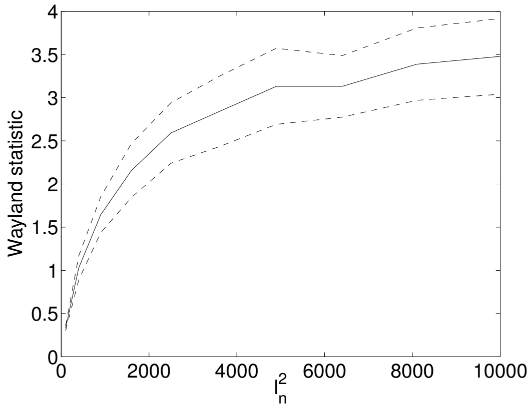

Figure 6: Relation between Wayland statistic and level of dynamical noise. For each noise level, the solid line shows mean, and the broken lines, mean $\pm$ standard deviation of 20 realizations.

$$\frac{dy_{j,i}}{dt} = cby_{j+1,i}(y_{j-1,i} - y_{j+2,i}) - cy_{j,i} + \frac{h_y c}{b} x_i, \tag{14}$$

where we used the following cyclic boundary conditions:

$$x_{I+i} = x_i, \quad y_{j+J,i} = y_{j,i+1}, \quad y_{j-J,i} = y_{j,i-1}. \tag{15}$$

We used $I = 40$, $J = 5$, $F = 8$, $b = 10$, $c = 10$, $h_x = 1$, and $h_y = 1$. We observed y_{11} every 0.05 unit time to obtain a scalar time series with a length of 20 000. Using this as the original data, we generated 39 randomly shuffled surrogates. The result, shown in Fig. 7, shows the reverse rejection. Hence, high dimensionality can cause the reverse rejection.

Given a dataset of finite length, we cannot distinguish a high dimensional system from a system with dynamical noise. As a result, the reverse rejection can occur for the same reason as that for the dynamical noise.

This study was partly supported by the Industrial Technology Research Grant Program in 2003, from the New Energy and Industrial Technology Development Organization (NEDO) of Japan.

References

[1] D. T. Kaplan and L. Glass, *Phys. Rev. Lett.* **68**, 427 (1992); D. T. Kaplan and L. Glass, *Physica D* **64**, 431 (1993); D. T. Kaplan, *Physica D* **73**, 38 (1994); L. W. Salvino and R. Cawley, *Phys. Rev. Lett.* **73**, 1091 (1994); G. J. Ortega and E. Louis, *Phys. Rev. Lett.* **81**, 4345 (1998); J. Bhattacharya and P. P. Kanjilal, *Physica D* **132**, 100 (1999); S. Horai, T. Yamada, and K. Aihara, *T. IEE Japan* **122**, 141(2002); M. Small and C. K. Tse, *IEEE Trans. Circuits Systems I Fund. Theory Appl.***50**, 663 (2003).

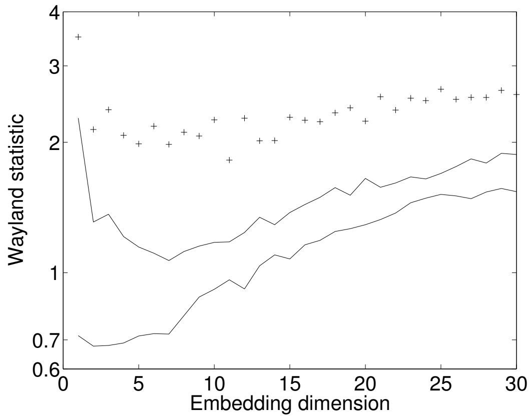

Figure 7: For data of Lorenz'96 model, we generated 39 randomly shuffled surrogates and compared them using Wayland statistic. For each embedding dimension, symbol $+$ shows Wayland statistic of original data, and the solid lines, the minimum and maximum of surrogates.

[2] R. Wayland, D. Bromley, D. Dickett, and A. Passamante, *Phys. Rev. Lett.* **70**, 580 (1993).

[3] A. M. Fraser and H. L. Swinney, *Phys. Rev. A* **33**, 1134 (1986).

[4] M. B. Kennel, *Phys. Rev. E***56**, 316 (1997).

[5] J. A. Scheinkman and B. LeBaron, *J. Business* **62**, 311 (1989).

[6] E. N. Lorenz, *J. Atmospheric Sci.* **20**, 130 (1963).

[7] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, (Cambridge University Press, Cambridge, UK, 1992).

[8] E. N. Lorenz, in *Proc. Seminar on Predictability*, **1**, 1, (ECMWF, Reading, Berkshire, UK, 1996); J. A. Hansen and L. A. Smith, *J. Atmospheric Sci.* **57** 2859 (2000).