2004 International Symposium on Nonlinear Theory and its Applications (NOLTA2004)
Fukuoka, Japan, Nov. 29 - Dec. 3, 2004

Looking for nonlinearity in the dynamics of surface wind using surrogate data

Yoshito Hirata[1], Hideyuki Suzuki[1], Kazuyuki Aihara[2,1,3], Rikiya Abe[4], Koichi Kanie[4], Taiji Yamada[5], and Jun Takahashi[5]

[1] Department of Mathematical Informatics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[2] Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
[3] ERATO Aihara Complexity Modelling Project, JST, 45-18 Oyama, Shibuya-ku, Tokyo 151-0065, Japan
[4] Electric Power Development Co., Ltd., 9-88, Chigasaki 1-Chome, Chigasaki, Kanagawa 253-0041, Japan
[5] R&D Team, AIHARA Electrical Engineering Co., Ltd. 2-16-8 Hama-cho, Funabashi, Chiba 273-0012, Japan
Email: yoshito@sat.t.u-tokyo.ac.jp, hideyuki@sat.t.u-tokyo.ac.jp, aihara@sat.t.u-tokyo.ac.jp, rikiya_abe@jpower.co.jp, koichi_kanie@jpower.co.jp, taiji@aihara.co.jp, junt@aihara.co.jp

Abstract—We try to show that the dynamics of surface wind is nonlinear. The dynamics of wind is nonstationary. Therefore, from a time series, we extract stationary segments and apply surrogate data analysis. While the result suggests that the dynamics is a linear Gaussian stochastic process, there might be still the possibility that the wind is a nonlinear process because we may have a problem in the sampling frequency.

1. Introduction

The dynamics of wind is believed to be chaotic or it has a low-dimensional attractor as air is a turbulent fluid. There is a lot of work [1, 2, 3, 4, 5, 6] trying to show that weather and climate systems have low-dimensional attractors using the Grassberger-Procaccia algorithm [7], while they may just be spurious results because of short data sets [8].

Nonlinear deterministic models are sometimes used for predicting time series of a turbulent fluid. Casdagli [9] showed that for weak turbulent data, the prediction error of a nonlinear deterministic model is smaller than that of a stochastic model, while they do not make any difference for fully developed turbulence. On the other hand, Ragwitz and Kantz [10] showed the wind velocity can be predicted well using nonlinear deterministic models when it changes fast, while a linear stochastic model predicts better when the variation is small.

In this communication, we attempt to show that the dynamics of wind is nonlinear using surrogate data analysis.

2. Data

The data set we used was obtained at Tomamae Winvilla Wind Farm at Tomamae, Hokkaido, Japan. The wind velocity was observed with 1/45 Hz on 1 November 2002 from 0:00 am for 8 hours. (Therefore the time series has 640 points.) Let $\{u_t\}_{t=1}^{640}$ be the time series. The time series is shown in Fig. 1.

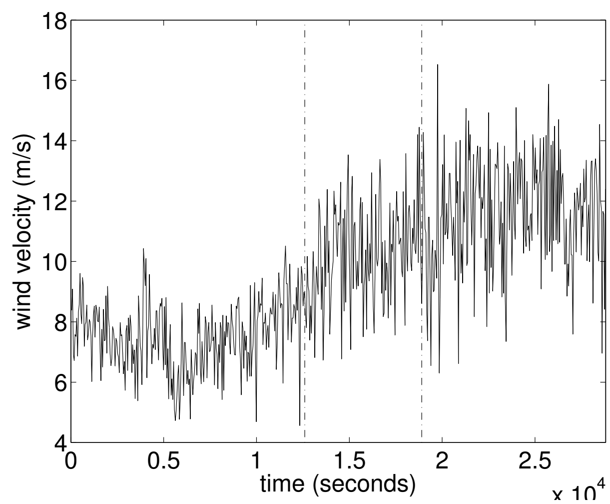

Figure 1: Time series of wind used in the analysis.

3. Stationarity test

Before applying surrogate data analysis, we need to check the stationarity of the data. Timmer [11] warned that a nonstationary data can be spuriously identified as a nonlinear deterministic system using surrogate data analysis even if it is not.

We employed the method of Kennel [12] for testing the stationarity of the data. If a data set is stationary, we expect that the nearest neighbors in a state space appear randomly in time except for the temporal neighbors. In Ref. [12], this property was implemented as a statistical hypothesis test.

First we applied the method to whole the set of the time series. Using Refs. [13, 14], we decided an embedding space as $(u_t, u_{t+2}, u_{t+4}, u_{t+6})$. After applying the method of Ref. [12], we obtain the statistic $z = 13.9528$. As it is bigger than 2.326, we rejected at 99% confidence level the null hypothesis that the data set is stationary.

To move on to the further studies, we looked the time series closely and decided to split it manually into the fol-

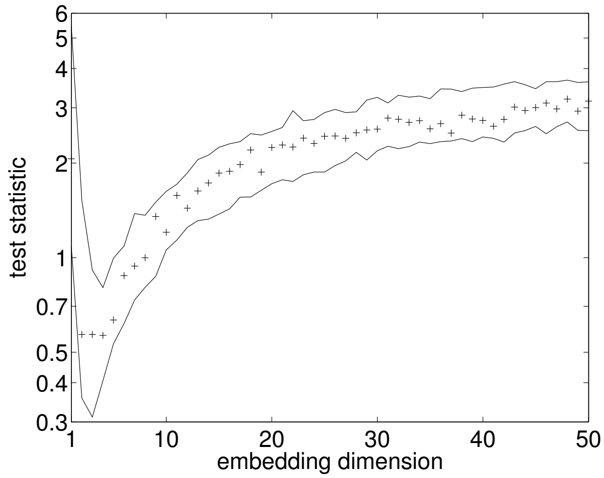

Figure 2: The result of surrogate data analysis for the segment $\{u_t\}_{t=1}^{280}$. The method of Ref. [16] was used to calculate the statistic. For each embedding dimension, solid lines show the minimum and the maximum of the statistics obtained from the surrogate data, and + shows the statistic obtained from the original segment.

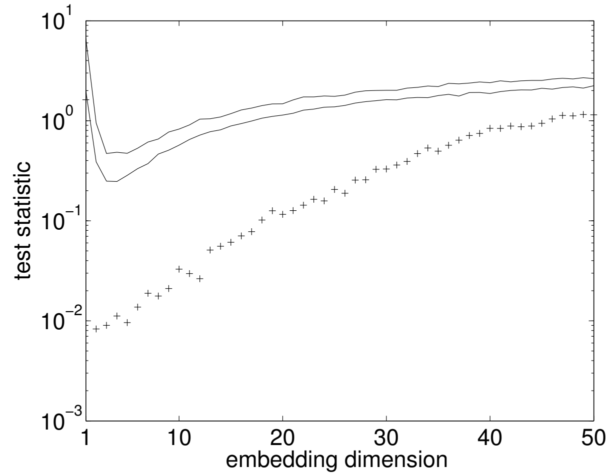

Figure 3: The result of surrogate data analysis for the Lorenz 63 model. The method of Ref. [16] was used to calculate test statistics. For each embedding dimension, solid lines show the minimum and the maximum of the statistics obtained for the surrogate data, and + shows the statistic obtained from the original time series.

lowing three segments: $\{u_t\}_{t=1}^{280}$, $\{u_t\}_{t=281}^{420}$, and $\{u_t\}_{t=421}^{640}$. For each segment, the methods of Refs. [13, 14] were applied and we obtained embedding spaces $(u_t, u_{t+2}, u_{t+4}, u_{t+6})$, $(u_t, u_{t+2}, u_{t+4}, u_{t+6})$, and $(u_t, u_{t+1}, u_{t+2}, u_{t+3})$, respectively. Then we used the method of Ref. [12] and obtained the statistics $z = 1.2210$, 1.9727, and -0.9454, respectively. As all of them are less than 2.326, each segment can be regarded as stationary.

4. Surrogate data analysis

For each segment, we generated surrogate data using Ref. [15]. Using the method of Ref. [15], we can generate surrogate data that preserve the distribution perfectly and the power spectrum approximately. Therefore, using the surrogate data we can test a hypothesis whether a data set is generated from a linear Gaussian stochastic process.

For the segment of $\{u_t\}_{t=1}^{280}$, we generated 199 sets of surrogate data. For each set of surrogate data and the original segment, we calculated the statistic of Ref. [16], which shows how deterministic the system could be. The result is shown in Fig. 2.

As in the Fig. 2 the statistics of the original segment are always between the minimum and the maximum of those of the surrogate data, the hypothesis cannot be rejected, i.e., the data set was possibly generated by a linear Gaussian stochastic process. (For the combined test, the p-value needs to be adjusted.)

We also applied the same test for the segments $\{u_t\}_{t=281}^{420}$ and $\{u_t\}_{t=421}^{640}$ and we obtained the similar results.

5. Is there still a possibility for nonlinearity?

The result of the surrogate data analysis suggested that the dynamics of wind is a linear Gaussian process. But, should we accept the result obediently? For example, there is a possibility that the test statistic of Ref. [16] is not sensitive to the form of nonlinearity present in the data. There is also a possibility that the data sets are too short. In this section, however, we look for another possibility: we observed the wind with a wrong sampling frequency.

We observed the wind with 1/45 Hz, which could be too long as Ragwitz and Kantz [10] observed the wind velocity with 8 Hz. Using two models, we will see how the result would change if we used a different sampling frequency.

For the first model, we use the Lorenz 63 model [17], which is defined as follows:

$$\begin{cases} \frac{dx}{dt} = -\sigma x + \sigma y \\ \frac{dy}{dt} = -xz + rx - y \\ \frac{dz}{dt} = xy - bz, \end{cases} \qquad (1)$$

where $(\sigma, r, b) = (10, 28, 8/3)$. We observed the variable x every 0.16 second in the time of the equation and obtained a scalar time series of 2 000 points.

First we tested the nonlinearity of the model using the observed data. We generated 199 sets of surrogate data using the method of Ref. [15]. We calculated the statistic of Ref. [16]. The result is shown in Fig. 3. As the statistic of the original data is below the minimum for the statistic of surrogate data for all the dimensions, the hypothesis is rejected.

Next we tested the nonlinearity of the model using a wrong time scale. We obtained another scalar time series

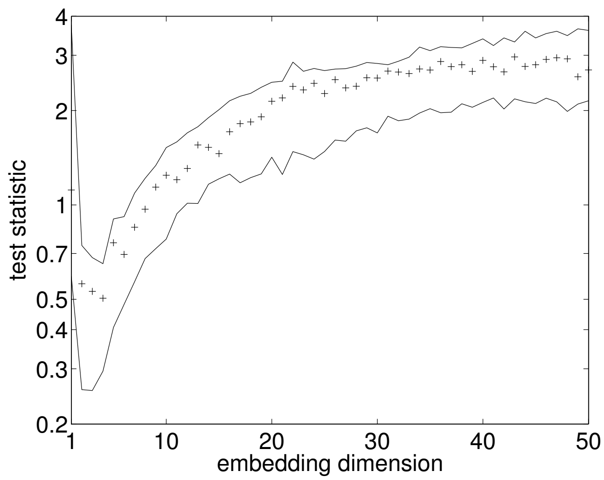

Figure 4: The result of surrogate data analysis for the time series of length 200 created by taking every 10 points of the scalar time series generated from the Lorenz 63 model. The statistic of Ref. [16] was calculated as a test statistic. For each embedding dimension, solid lines show the minimum and the maximum of the statistics obtained for the surrogate data, and + shows the statistic obtained from the original time series of length 200.

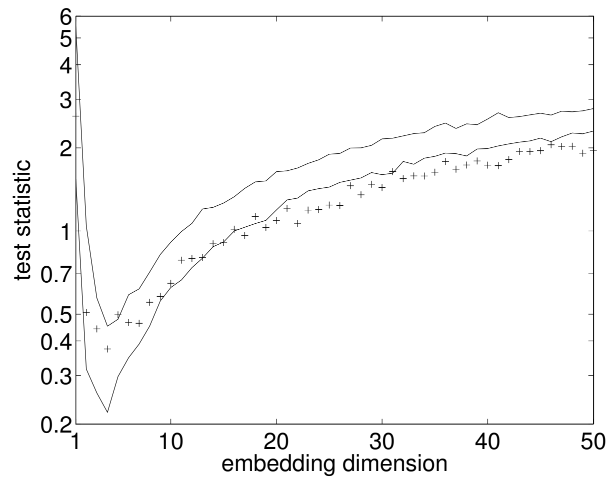

Figure 5: The result of surrogate data analysis for the time series generated from the Lorenz two-scale system. The statistic of Ref. [16] was calculated as a test statistic. For each embedding dimension, solid lines show the minimum and the maximum of the statistics obtained from the surrogate data, and + shows the statistic obtained from the original data.

of 200 points by taking every 10 points from the original scalar time series. Using the method of Ref. [15], we generated 199 sets of surrogate data. Using the statistic of Ref. [16], we obtained Fig. 4. As, for each embedding dimension, the statistic obtained for the data of length 200 is always between the minimum and the maximum for the surrogate data, the hypothesis cannot be rejected. It means that the dynamics of the Lorenz 63 model was identified as a linear Gaussian process, while it is not correct.

The second example is the Lorenz two-scale system [18, 19], a model of the atmosphere. The Lorenz two-scale system contains m slow large-scale variables x_i and $m \times n$ fast small-scale variables $y_{j,i}$. The equations are defined as follows:

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

$$\frac{dy_{j,i}}{dt} = cby_{j+1,i}(y_{j-1,i} - y_{j+2,i}) - cy_{j,i} + \frac{h_y c}{b} x_i, \quad (3)$$

where we enforced the following cyclic boundary conditions

$$x_{m+i} = x_i, y_{j+n,i} = y_{j,i+1}, y_{j-n,i} = y_{j,i-1}. \quad (4)$$

We set $m = 40$, $n = 5$, $F = 8$, $b = 10$, $c = 10$, $h_x = 1$, and $h_y = 1$. In this 240 dimensional model, we observed a fast small-scale variable $y_{1,1}$ every 0.05 second in the time of the equations, and obtained a scalar time series of length 2000. First we used the observed 2000 points data and did the surrogate data analysis. We generated 199 sets of surrogate data using the method of Ref. [15]. For each surrogate data and the original time series, the statistic of Ref. [16] was calculated. The result is shown in Fig. 5. In 33 embedding dimensions, the statistic of the original time series is smaller than the minimum of those of the surrogate data, and in the embedding dimension 5, it is bigger than the maximum. As 34 individual tests out of 50 are rejected, the hypothesis should be rejected in the combined test: the time series was identified as nonlinear.

Next we took every 10 points from the original time series and generated another time series of length 200. Using the method of Ref. [15], we generated 199 sets of its surrogate data. The statistic of Ref. [16] was calculated as shown in Fig. 6. As no individual test rejects the hypothesis, neither does the combined test.

These two examples show that even if a system is nonlinear, if it is observed with too low a frequency, it can be misclassified as a linear system.

6. Conclusion

We tested whether the dynamics of wind is nonlinear or not. The surrogate data analysis did not show evidence of deterministic nonlinearity. The argument using the two Lorenz models suggested that if we observe the wind with a higher sampling rate, we may find nonlinearity in its dynamics.

Acknowledgment

This study was supported by Industrial Technology Research Grant Program in 2003 from New Energy and In-

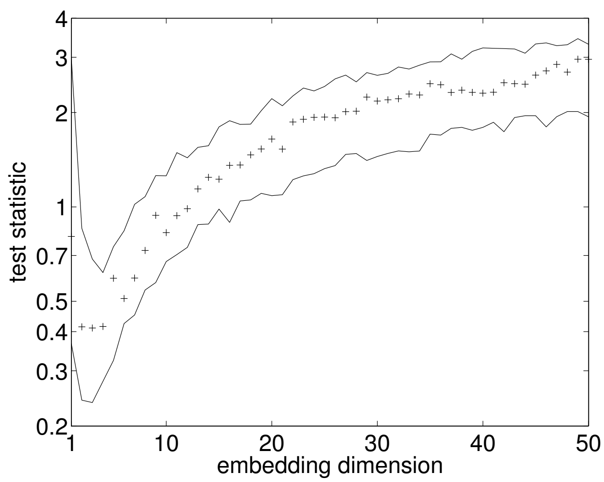

Figure 6: The result of surrogate data analysis for the time series of 200 points generated by taking every 10 points from the scalar time series generated from the Lorenz two-scale system. The statistic of Ref. [16] was calculated as a test statistic. For each embedding dimension, solid lines show the minimum and the maximum of the statistics obtained from the surrogate data, and + shows the statistic obtained from the original data of 200 points.

dustrial Technology Development Organization (NEDO) of Japan.

References

[1] K. Freadrich, “Estimating the dimensions of weather and climate attractors,” *J. Atmos. Sci.*, vol.43, pp.419–432, 1986.

[2] C. Nicolis and G. Nicolis, “Evidence for climatic attractors,” *Nature*, vol.326, pp.523–524,1987.

[3] A. A. Tsonis and J. B. Elsner, “The weather attractor over very short timescales,” *Nature*, vol.333, pp.545–547,1988.

[4] C. L. Keppenne and C. Nicolis, “Global properties and local structure of the weather attractor over western Europe,” *J. Atmos. Sci.*, vol.46, pp.2356–2370, 1989.

[5] M. B. Sharifi, K. P. Georgakakos, and I. Rodriguez-Iturbe, “Evidence of deterministic chaos in the pulse of storm rainfall,” *J. Atmos. Sci.*, vol.47, pp.888–893, 1990.

[6] G. Poveda-Jaramillo and C. Puente, “Strange attractors in atmospheric boundary layer turbulence,” *Boundary-Layer Meteorol.*, 64, pp.175–197, 1993.

[7] P. Grassberger and I. Procaccia, “Characterization of strange attractors,” *Phys. Rev. Lett.*, vol.50, pp.346–349, 1983.

[8] P. Grassberger, “Do climatic attractor exist?,” *Nature*, vol.323, pp.609–612, 1986.

[9] M. Casdagli, “Chaos and deterministic versus stochastic non-linear modelling,” *J. R. Statist. Soc. B*, vol.54, pp.303–328, 1991.

[10] M. Ragwitz and H. Kantz, “Detecting non-linear structure and predicting turbulent gusts in surface wind velocities,” *Europhys. Lett.*, vol.51, pp.595–601, 2000.

[11] J. Timmer, “Power of surrogate data testing with respect to nonstationarity,” *Phys. Rev. E*, vol.58, pp.5153–5156, 1998.

[12] M. B. Kennel, “Statistical test for dynamical nonstationarity in observed time-series data,” *Phys. Rev. E*, vol.56, pp.316–321, 1997.

[13] A. M. Fraser and H. L. Swinney, “Independent coordinates for strange attractors from mutual information,” *Phys. Rev. A*, vol.33, pp.1134–1140, 1986.

[14] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, “Determining embedding dimension for phase-space reconstruction using a geometrical construction,” *Phys. Rev. A*, vol.45, pp.3403–3411, 1992.

[15] T. Schreiber and A. Schmitz, “Improved surrogate data for nonlinearity tests,” *Phys. Rev. Lett.*, vol.77, pp.635–638, 1996.

[16] R. Wayland, D. Bromley, D. Pickett, and A. Passamante, “Recognizing determinism in a time series,” *Phys. Rev. Lett.*, vol.70, pp.580–582, 1993.

[17] E. N. Lorenz, “Deterministic non-periodic flow,” *J. Atmos. Sci.*, vol.20, pp.130–141, 1963.

[18] E. N. Lorenz, “Predictability — A problem partly solved,” in *Predictability* ECMWF, Seminar Proceedings, Shinfield Park, Reading, RG2 9AX, UK, 1995.

[19] J. A. Hansen and L. .A. Smith, “The role of operational constraints in selecting supplementary observations,” *J. Atmos. Sci.*, vol. 57, pp.2859–2871, 2000.