Int J Biometeorol
DOI 10.1007/s00484-005-0004-9

ORIGINAL ARTICLE

J.-J. Delaunay · R. Konishi · C. Seymour

Analysis of cedar pollen time series: no evidence of low-dimensional chaotic behavior

Received: 17 January 2005 / Revised: 21 July 2005 / Accepted: 27 July 2005

Abstract Much of the current interest in pollen time series analysis is motivated by the possibility that pollen series arise from low-dimensional chaotic systems. If this is the case, short-range prediction using nonlinear modeling is justified and would produce high-quality forecasts that could be useful in providing pollen alerts to allergy sufferers. To date, contradictory reports about the characterization of the dynamics of pollen series can be found in the literature. Pollen series have been alternatively described as featuring and not featuring deterministic chaotic behavior. We showed that the choice of test for detection of deterministic chaos in pollen series is difficult because pollen series exhibit $1/f^{\alpha}$ power spectra. This is a characteristic that is also produced by colored noise series, which mimic deterministic chaos in most tests. We proposed to apply the Ikeguchi–Aihara test to properly detect the presence of deterministic chaos in pollen series. We examined the dynamics of cedar (*Cryptomeria japonica*) hourly pollen series by means of the Ikeguchi–Aihara test and concluded that these pollen series cannot be described as low-dimensional deterministic chaos. Therefore, the application of low-dimensional chaotic deterministic models to the prediction of short-range pollen concentration will not result in high-accuracy pollen forecasts even though these models may provide useful forecasts for certain applications. We believe that our conclusion can be generalized to pollen series from other wind-pollinated plant species, as wind speed, the forcing parameter of the pollen emission and transport, is best described as a nondeterministic series that originates in the high dimensionality of the atmosphere.

Keywords Aerobiology · Cedar pollen · Chaos · Forecasting · Time series

J.-J. Delaunay (✉) · R. Konishi
Department of Mechanical Engineering (8-323),
The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku,
113-8656 Tokyo, Japan
e-mail: jean@mech.t.u-tokyo.ac.jp
Tel.: +81-3-58413017
Fax: +81-3-58418818

C. Seymour
The University of British Columbia,
2329 West Mall,
Vancouver, BC, Canada, V6T 1Z4

Introduction

Allergies are on the rise in modern society, as revealed by the increasing prevalence of asthma. In Japan, it has been reported that about 20% of the population suffers from allergies, from which one in five patients has developed asthma. While the causes of allergies are still being debated, the medical community is clear in their advice that the most effective way to prevent wheezing and asthma from developing is to avoid inhaling airborne allergens. In Japan, one of the main sources of airborne allergens is Japanese cedar (*Cryptomeria japonica*) pollen. Efforts to provide cedar pollen forecasts to allergy sufferers have resulted in the development of real-time forecasting systems (Kawashima and Takahashi 1995, 1999; Delaunay et al. 2002). These systems are based on the modeling of physical processes such as the flowering, emission, and transport of airborne pollen and can typically provide 2-day ahead pollen forecasts using meteorological forecasts from mesoscale models as their inputs. A few media companies are starting to integrate pollen forecasts generated by these systems into their weather reports. It is usually found that short-range (a few hours of lead time) pollen forecasts produced by these transport models differ significantly from observed pollen series collected by pollen samplers. In this respect, time series analysis of measured pollen concentration may provide a means to improve short-range prediction and/or to better understand the underlying dynamics of pollen series. Bianchi et al. (1992) and Arizmendi et al. (1993) used nonlinear time series techniques to analyze 2-h pollen series and concluded that their series exhibited low-dimensional chaotic behavior. Examining cedar pollen series with the correlation integral method (Grassberger and Procaccia 1983), the largest Lyapunov exponent method (Wolf et al. 1985), and Casdagli's test (1991), we found no convincing evidence of low-dimensional chaotic behavior though the pollen series were found to have some degree of determinism as revealed by Casdagli's test. One of the problems when using the above tests to detect deterministic chaos arises from the existence of stochastic processes such as colored noise that can masquerade chaotic behavior in these tests. There is a need to analyze pollen series with a test that can distinguish colored noise from deterministic chaos. At the time we first published our results, we were unaware of the work of Ikeguchi and Aihara (1997), which addresses the distinction of deterministic chaos from colored noise. In their test, the prediction performance of a nonlinear model is evaluated using two types of correlation coefficients (the standard linear correlation coefficient and a difference correlation coefficient) between predicted and observed values. The predictability of the time series can be inferred from the values of the two correlation coefficients and thus the dynamics underlying the series classified as deterministic chaos or a stochastic process. Here, we report our analysis of cedar pollen series using Ikeguchi and Aihara's test and discuss the predictability of wind-pollinated pollen series.

Materials and methods

In this study, we analyzed cedar pollen concentration time series recorded with a pollen counter of the type KH3000 from the Yamato Corporation (Yamato 1998). The KH3000 is an automatic pollen counter that measures the scattered light intensities of airborne particles from two light beams, resulting in information about the size and the shape of the sampled particles. Spherical particles with a diameter in the range of 28–35 μm are counted as cedar pollen grains.

Pollen series from two locations in Japan, namely Yamagata and Takao, were examined (Fig. 1). In the Yamagata location, the instrument was placed on the roof of the Murayama Public Health Center of the Yamagata Prefecture and operated for the period 30 March 2001–21 April 2001. In the Takao location, the instrument was placed on the roof of the Forestry and Forest Products Research Institute of Japan (Hachioji, Tokyo) and operated for two consecutive cedar pollen seasons, namely 2000 and 2001, for the period 1 February–31 March. The Yamagata pollen series represents the dynamics of a pollen series collected at a site distant (5–10 km) from major cedar plantations whereas the Takao series represent that of a site close (<3 km) to major cedar plantations.

The Ikeguchi–Aihara test was introduced to distinguish stochastic noise with a $1/f^{\alpha}$ power spectrum from low-dimensional deterministic chaos. The test compares the correlation coefficient between observations and one-step-ahead predictions with a difference correlation coefficient. Colored noise series generated by stochastic systems

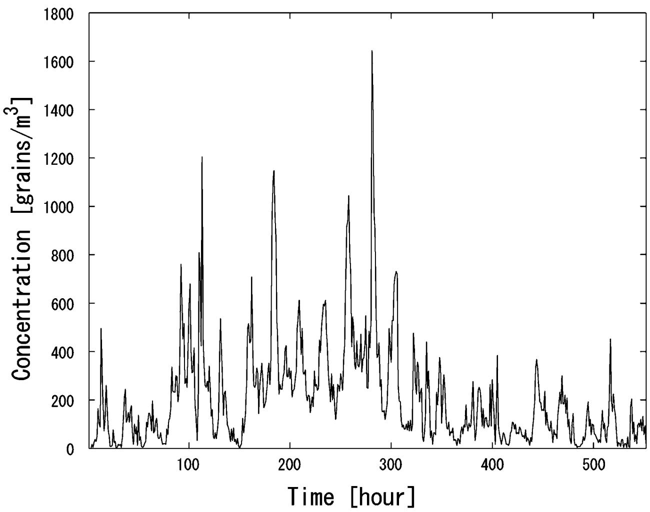

Fig. 1 Pollen time series recorded with a KH3000 sampler at Yamagata for the period 30 March 2001–21 April 2001. Hourly pollen concentrations are plotted as a function of time

exhibit high values for the correlation coefficient but low values for the difference correlation coefficient. This contrasts with series produced by low-dimensional chaotic systems (e.g., Henon, Ikeda, and Bernoulli maps) for which both the correlation coefficient and the difference correlation coefficient have high values. The predictions are obtained using a nonlinear prediction scheme such as the *k*-nearest neighbor scheme. We used the local constant approximation (zeroth order) of the *k*-nearest neighbor scheme. First, state vectors are formed from the *N*-point time series $\{x_1, x_2,..., x_N\}$ using the time delay reconstruction method formalized by Takens (1981):

$$X_t(m) = \left(x_t, x_{t-\tau}, ..., x_{t-(m-1)\tau}\right), \quad (1)$$

where m is the embedding dimension and τ is the delay time (we used τ=1).

The k nearest neighbors $\{X_{k_1}(m), X_{k_2}(m), ..., X_{k_k}(m)\}$ of the state vector $X_t(m)$ are then selected from the collection of the state vectors. Finally, the one-step-ahead prediction is computed using:

$$\hat{x}_{t+1} = \frac{1}{k}\sum_{i=1}^{k} x_{k_i+1}, \quad (2)$$

Note that the prediction can also be computed using a weighted function of the distance between the state vector and its nearest neighbors; here we used a simple average or local constant approximation. In their paper, Ikeguchi and Aihara divide the series in two sets, the first half of the series being used as a database to predict the second half of the series. Here we used a slightly different method for the data division, where the complete series, with the exception of the part that is to be predicted, is used. This data division is useful in increasing the amount of data available for the test in the case of series with a low number of data points, such as in our pollen series. In this method, the database has to be redefined for each new prediction.

From Eq. 2, the predicted series $\{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_P\}$ can be extracted and compared with the observed series $\{x_1, x_2, \ldots, x_P\}$, where the number P of available predicted values depends on the embedding dimension and on the number of neighbors used. The observed and predicted series are compared using two correlation coefficients: the standard linear correlation coefficient and a difference correlation coefficient. The standard linear correlation coefficient, denoted r_1, is:

$$r_1 = \frac{\sum (x_i - \bar{x})\left(\hat{x}_i - \bar{\hat{x}}\right)}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum \left(\hat{x}_i - \bar{\hat{x}}\right)^2}}, \quad (3)$$

where $\bar{x}$ and $\bar{\hat{x}}$ are the averages of $\{x_i\}$ and $\{\hat{x}_i\}$.

The difference correlation coefficient, denoted r_2, is the correlation coefficient between the differences $\Delta x_i = x_{i+1} - x_i$ and $\Delta \hat{x}_i = \hat{x}_{i+1} - x_i$:

$$r_2 = \frac{\sum \left(\Delta x_i - \overline{\Delta x}\right)\left(\Delta \hat{x}_i - \overline{\Delta \hat{x}}\right)}{\sqrt{\sum \left(\Delta x_i - \overline{\Delta x}\right)^2}\sqrt{\sum \left(\Delta \hat{x}_i - \overline{\Delta \hat{x}}\right)^2}}, \quad (4)$$

We applied the Ikeguchi–Aihara test on a Bernouilli series, a red noise series, and the pollen series. The Bernoulli series was produced by numerically iterating a modified one-dimensional Bernoulli map, as described in Aizawa and Kohyama (1984). The values we used for the parameters of the modified Bernoulli map were $B = 3$, $\varepsilon = 10^{-13}$. The red noise series was generated by integrating the Uhlenbeck-Ornstein (1930) linear differential equation, which combines a deterministic exponential decay with additive normally distributed noise. The Bernoulli and red noise series were chosen because their power spectra are similar to that of the pollen series, and they provide a means to compare our pollen series with series originating from both a low-dimensional deterministic system and a stochastic system.

Results and discussion

Figure 2 shows the power spectra of the Bernoulli, red noise, and pollen series. The three series exhibit $1/f^{\alpha}$ power-law spectrum with $\alpha \simeq 1.5$, 2.0, and 1.5 for the Bernoulli, red noise, and pollen series, respectively. According to Fig. 3 of Ikeguchi and Aihara (1997), large differences between the correlation coefficient and the difference correlation coefficient are obtained for colored noise series having α in the range 1.5–3.0 whereas for

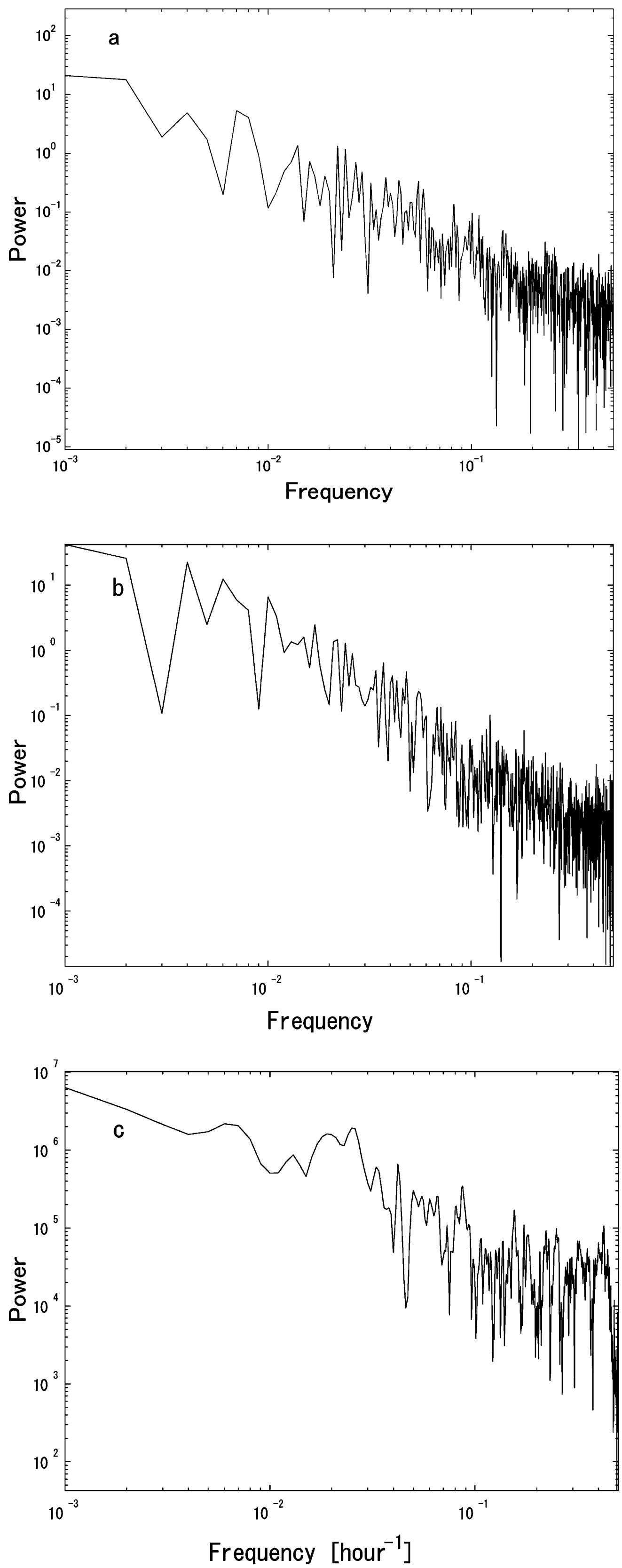

Fig. 2 Power spectra of **a** the Bernoulli series, **b** the red noise series, and **c** the Takao 2001 pollen series

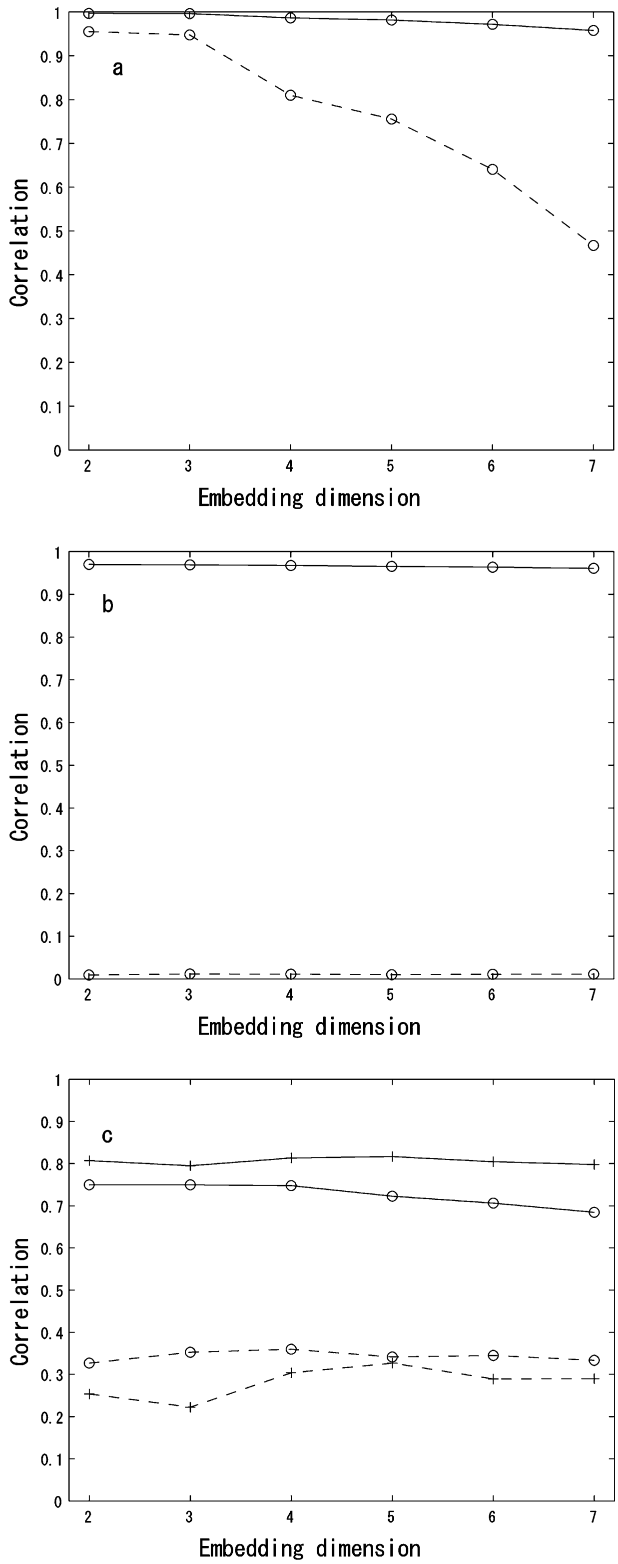

Fig. 3 Ikeguchi and Aihara's test showing r_1 (*solid line*) and r_2 (*dashed line*) as a function of the embedding dimension m for **a** the Bernoulli series, **b** the red noise series, and **c** the pollen series. For the Bernoulli and the red noise series, the number of points N=1,000 and k=1. For the red noise series, the correlation results were obtained by averaging over 100 trials. For the pollen series, k=10; both Yamagata (*crosses*) and Takao (*circles*) for 2001 are shown. The Takao pollen series for 2000 exhibits the same trend as that of the 2001 series and is not shown here

deterministic series, both the correlation coefficient and the difference correlation coefficient have high values. Thus the test results make a clear distinction between colored noise and deterministic series having power-law spectra with α in the range 1.5–3.0. Our pollen series was found to be a good candidate for the application of the Ikeguchi–Aihara test, as it had $\alpha \simeq 1.5$.

The correlation coefficient and the difference correlation coefficient are shown as a function of the embedding dimension in Fig. 3 for the Bernoulli, red noise, and pollen series. The number of data points N was fixed to 1,000 and k to 1 for the Bernoulli and the red noise series. For the red noise series, averaging over 100 trials was used. For the pollen series, the number k of nearest neighbors was fixed to 10, as it was found in our previous study (Delaunay et al. 2004) that $k \simeq 10$ resulted in improved prediction performance. In the graph of the Ikeguchi–Aihara test, both r_1 and r_2 have values close to 1 for the Bernoulli series whereas only r_1 show high values for the red noise series, with r_2 close to 0. The differences in behavior of r_1 and r_2 provide a criterion to discriminate deterministic chaotic series from red noise series. Note that in the case of the Bernoulli series, both r_1 and r_2 decrease with the embedding m, a tendency that is explained by the small fractal dimension of the series (<1 according to Aizawa and Kohyama 1984) and therefore the small value of the ideal embedding. The pollen series have high values for r_1 and low values for r_2. In the pollen series, the difference between r_1 and r_2 is large and argues against the presence of deterministic chaos. Further comparison between known deterministic series and stochastic series from real-world data can be found in Fig. 7 of Ikeguchi and Aihara's publication. The behavior of our pollen series does not resemble that of known deterministic chaotic series such that of the squid axon response, the laser, and the measles data, for which both r_1 and r_2 have high values. However, the observed value of r_2 for the pollen series is higher than that of known stochastic series such as stock exchange index and electroencephalography data, which may be evidence for a small degree of determinism in the pollen series.

From these results, it can be concluded that the dynamics underlying the hourly variations in cedar pollen concentration series is not tied to a low-dimensional chaotic system even though the series exhibits a small degree of determinism. This validates the results of our first report on short-range pollen forecasting (Delaunay et al. 2004) and contradicts a study by Bianchi et al. (1992) claiming low-dimensional chaotic behavior for their pollen series. In the latter study, evidence of chaos in the pollen series was obtained from estimating the correlation dimension, an approximate value for the dimension of the strange attractor of a low-dimensional chaotic series. The correlation dimension was computed using a modified Grassberger and Procaccia (1983) algorithm and was found to be a small noninteger value, which was interpreted as evidence for low-dimensional chaos. Unfortunately, the correlation dimension is also known to converge to finite values for stochastic series such as colored noise characterized by a power-law spectrum (Osborne and Provenzale 1989). For colored noise series, the computed correlation dimension varies with the exponent of the power-law spectrum. Thus, the observation of a small finite value for the correlation dimension is not sufficient to draw conclusions about the low-dimensional chaotic behavior of a time series, as colored noise can mimic the behavior of low-dimensional chaotic series in this particular test.

One consequence of our findings, which indicate pollen series do not derive from low-dimensional chaos, is that one cannot expect to obtain high-accuracy results when forecasting short-range cedar pollen concentration using low-dimensional deterministic chaotic models such as local nonlinear and artificial neural network models. The performance of these models applied to the forecast of 1-h-ahead cedar pollen concentrations is detailed in Delaunay et al. (2004). Both local nonlinear and artificial neural network models were found to have similar prediction performance; these models were effective in predicting small to medium pollen variations but failed to reproduce pollen bursts. We concluded that our pollen series exhibit a small degree of determinism but cannot be described as originating from a low-dimensional dynamics because nonlinear models failed to accurately reproduce intermittent pollen bursts, which is an essential characteristic of pollen series. The results of the Ikeguchi–Aihara test validated our previous conclusion. However, Arizmendi et al. (1993), applying artificial neural network models to the prediction of 2-h-ahead pollen concentration, report almost perfect agreement between predicted and observed values (see Figs. 2–4 of their report). We do not believe their results to be reasonable in light of the lack of evidence supporting low-dimensional chaotic behavior of pollen series and would like to suggest that the authors test their pollen series for the presence of deterministic chaos using a test such as the Ikeguchi–Aihara test.

We believe that our conclusion on the dynamics of cedar pollen series should generalize to pollen series of wind-pollinated plants because for these species, the pollen emission is chiefly governed by variations in local surface wind speed, and the dynamics of local surface wind speed is best described in terms of a high-dimensional nonlinear system, not a low-dimensional deterministic system.

Finally, the procedure outlined in this report is general and therefore should be useful in detecting deterministic chaos in time series of other environmental parameters.

Acknowledgements We are very grateful to Toshitaka Yokoyama of the Forestry and Forest Products Research Institute of Japan and Yuichi Takahashi of the Yamagata Prefecture Institute of Public Health for providing the pollen series.

References

Aizawa Y, Kohyama T (1984) Asymptotically non-stationary chaos. Prog Theor Phys 71:847–850
Arizmendi CM, Sanchez JR, Ramos NE, Ramos GI (1993) Time series predictions with neural nets: application to airborne pollen forecasting. Int J Biometeorol 37:139–144
Bianchi MM, Arizmendi CM, Sanchez JR (1992) Detection of chaos: new approach to atmospheric pollen time-series analysis. Int J Biometeorol 36:172–175
Casdagli M (1991) Chaos and deterministic versus stochastic nonlinear modelling. J R Stat Soc Ser B Stat Methodol 54:303–328
Delaunay J-J, Fedra K, Kubat M (2002) Cedar pollen forecasting in the Kanto region. Arch Complex Environ Stud 14:34
Delaunay J-J, Seymour C, Fouillet V (2004) Investigation of short-range cedar pollen forecasting. Phys Rev E 70 (doi 066214)
Grassberger P, Procaccia I (1983) Characterization of strange attractors. Phys Rev Lett 50:346–349
Ikeguchi T, Aihara K (1997) Difference correlation can distinguish deterministic chaos from $1/f^{\alpha}$ type colored noise. Phys Rev E 55:2530–2538
Kawashima S, Takahashi Y (1995) Modeling and simulation of mesoscale dispersion processes for airborne cedar pollen. Grana 34:142–150
Kawashima S, Takahashi Y (1999) An improved simulation of mesoscale dispersion of airborne cedar pollen using a flowering-time map. Grana 38:316–324
Osborne AR, Provenzale A (1989) Finite correlation dimension for stochastic systems with power-law spectra. Physica D 35:357–381
Takens F (1981) Detecting strange attractors in fluid turbulence. In: Rand DA and Young LS (eds) Dynamical systems and turbulence, Springer, Berlin Heidelberg New York
Uhlenbeck GE, Ornstein LS (1930) On the theory of the Brownian motion. Phys Rev 36:823–841
Wolf A, Swift JB, Swinney HL, Vastano JA (1985) Determining Lyapunov exponent from a time series. Physica D 16:285–317
Yamato Corporation (1998) Japanese patent number 10–318908