Y. Okabe and T. Yamane
Nagoya Math. J.
Vol. 152 (1998), 175–201

THE THEORY OF KM_2O-LANGEVIN EQUATIONS AND ITS APPLICATIONS TO DATA ANALYSIS (III): DETERMINISTIC ANALYSIS

YASUNORI OKABE AND TOSHIYUKI YAMANE

Abstract. A technique is given for detecting deterministic dynamics in time series. Some stochastic difference equations, called KM_2O-Langevin equations, are extracted directly from given data. We can find deterministic dynamics in time series by evaluating the magnitude of innovation part of the KM_2O-Langevin equations. We can further find chaotic dynamics in time series by predicting it from the viewpoint of the theory of KM_2O-Langevin equations.

We apply our method to the data of measles and chicken pox, which are also treated by G.Sugihara and R.M.May in [1]. The result of numerical experiments indicates that there seem to exist some deterministic dynamics in both time series. It also suggests, however, that the data of measles seems to be chaotic while that of chicken pox not, which corresponds to the result of G.Sugihara and R.M.May.

§1. Introduction

There are a lot of systems in the world, whose behavior as a whole is never understandable if we only view their components separately. Such systems, called complex systems, arise from a variety of origins of complexity such as stochastic structure, deterministic chaos and so on. This feature of complex systems has the result that a priori parametric statistical models (e.g. ARMA model, linear regression model) may fail to catch the underlying structure arising from the complex systems which lies behind the data.

Therefore, we must check the validity of the preconditions that is assumed before data analysis. We call such an approach toward data analysis *a qualitative approach* in contrast to quantitative approaches such as parametric statistical models. One of the authors has presented a precondition-

Received July 7, 1997.

This research was partially supported by Grant-in-Aid for Science Research (B) No. 07459007 and Grant-in-Aid for Exploratory Research No. 08874007, the Ministry of Education, Science, Sports and Culture, Japan and by Promotion Work for Creative Software, Information-Technology Promotion Agency, Japan.

free method of qualitative approach to time series based on the theory of KM_2O-Langevin equations [3], [4], [5], [6], [7], [8], [9].

The theory of KM_2O-Langevin equations is a new method of time series analysis that originates from the study of the fluctuation-dissipation theorem which is thought to be one of the principles of nonequilibrium statistical physics. The main feature of this theory is that it requires no prior information about data and does not define any parametric models before data analysis, but extracts explanatory models in the form of stochastic difference equations directly from data.

The most fundamental qualitative property of this theory is stationarity of time series. Testing stationarity of given time series is possible using Test(S), which is established in [4]. We can go to the next step of analysis only for data whose stationarity is assured by Test(S). Given two stationary time series, we can discuss whether they are related to each other, or in other words, whether there exists some causality between them by using Test(CS) in [7].

The contents of this paper are the following. First, in section 2 and section 3, we briefly review the theory of KM_2O-Langevin equations and its application to causality analysis which is developed in [3],[6]. Secondly, in section 3, we give a definition of nonlinear causality and give a method of checking it based on the Bootstrap method [10]. The results described here are the extension of [7]. Thirdly, in section 4, we give a method of checking whether there exist some deterministic dynamics in a stationary time series by applying Test(CS), which is called Test(D). The importance to check determinism in time series is argued in [1], and we can give an answer to this question using Test(D). Fourthly, in section 4, we give a method of testing whether the underlying dynamics in time series are chaotic from the viewpoint of prediction analysis in [5], [8]. Finally, in section 5, in order to make the analysis in [9] more precise, we apply our method to the measles data and chickenpox data, which are thought to originate from some complex systems. It is reported in [1] that the measles data appear chaotic, while the chickenpox data show a seasonal cycle with additive noise. We wish to take up the more basic question of whether these time series can really be thought to come from deterministic dynamic systems before a discussion of whether they are chaotic or not. The result of numerical experiments indicates that there seem to exist some deterministic dynamics in both time series. It also suggests, however, that the measles data seem

to be chaotic while the chickenpox data are not, which corresponds to the result of [1].

The authors would like to express their gratitude to Mr. A. Kaneko for his helpful support in computer experiments.

§2. A short review of the theory of KM_2O-Langevin equations

2.1. KM_2O-Langevin equation

Let $\mathbf{X} = (X(n); |n| \leq N)$ be any $\mathbf{R}^d$-valued stochastic process on a probability space $(\Omega, \mathcal{B}, P)$. We introduce *the forward KM_2O-Langevin fluctuation flow* $\nu_+ = (\nu_+(n); 0 \leq n \leq N)$ associated with $\mathbf{X}$ by projecting $X(n)$ on the vector space $\mathbf{M}_0^{n-1}(\mathbf{X})$ which is spanned by $\{X_j(k);\ 0 \leq k \leq n-1,\ 1 \leq j \leq d\}$:

$$\nu_+(0) \equiv X(0) \tag{1}$$

$$\nu_+(n) \equiv X(n) - P_{\mathbf{M}_0^{n-1}(\mathbf{X})}X(n) \quad (1 \leq n \leq N), \tag{2}$$

where $P_{\mathbf{M}_0^{n-1}(\mathbf{X})}$ denotes for the projection operator to $\mathbf{M}_0^{n-1}(\mathbf{X})$.

In the same way, we introduce *the backward KM_2O-Langevin fluctuation flow* $\nu_- = (\nu_-(\ell); -N \leq \ell \leq 0)$ by

$$\nu_-(0) \equiv X(0) \tag{3}$$

$$\nu_-(\ell) \equiv X(\ell) - P_{\mathbf{M}_{\ell+1}^{0}(\mathbf{X})}X(\ell), \quad (-N \leq \ell \leq -1). \tag{4}$$

The forward and backward KM_2O-Langevin fluctuation flows are sometimes called innovation processes of $\mathbf{X}$.

We assume the following independence condition for $\mathbf{X}$:

$$\begin{cases} \{X_j(n); 0 \leq n \leq N-1, 1 \leq j \leq d\} \text{ is} \\ \qquad \text{linearly independent in } \mathbf{M}_0^{N-1}(\mathbf{X}), \\ \{X_j(-n); 0 \leq n \leq N-1, 1 \leq j \leq d\} \text{ is} \\ \qquad \text{linearly independent in } \mathbf{M}_{-N+1}^{0}(\mathbf{X}). \end{cases} \tag{5}$$

Then, we can derive the stochastic difference equation that describes the time evolution of $\mathbf{X}$ in the following way:

$$X(\pm n) = -\sum_{k=1}^{n-1} \gamma_\pm(n,k)X(\pm k) - \delta_\pm(n)X(0) + \nu_\pm(\pm n) \quad (0 \leq n \leq N), \tag{6}$$

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

where $\gamma_{\pm}(n,k)$ and $\delta_{\pm}(n)$ are $d \times d$ matrices which are determined uniquely from $\mathbf{X}$. We call this equation *KM_2O-Langevin equation.* For each $n(0 \leq n \leq N)$, we denote by $V_{\pm}(n)$ the inner product of $\nu_{\pm}(\pm n)$ themselves:

$$V_{\pm}(n) \equiv E[\nu_{\pm}(\pm n)^t \nu_{\pm}(\pm n)]. \tag{7}$$

It is to be noted that

$$E[\nu_{\pm}(\pm n)^t \nu_{\pm}(\pm m)] = \delta_{nm} V_{\pm}(n) \quad (0 \leq m, n \leq N). \tag{8}$$

We put

$$\begin{aligned} \mathcal{LM}(\mathbf{X}) \equiv \{&\gamma_{\pm}(n,k), \delta_{\pm}(n), V_{\pm}(m); \\ &1 \leq n \leq N,\ 0 \leq k \leq n-1,\ 0 \leq m \leq N\} \end{aligned} \tag{9}$$

and call the system $\mathcal{LM}(\mathbf{X})$ *KM_2O-Langevin matrix* associated with $\mathbf{X}$. Note that $\gamma_{\pm}(n,0) = \delta_{\pm}(n)$.

In this paper, we treat only local and weakly stationary processes which are defined as follows.

DEFINITION 2.1. $\mathbf{X}$ is called a local and weakly stationary process if and only if $\mathbf{X}$ satisfies the following two properties:

(i) $E[X(n)] = 0 \quad (|n| \leq N)$,

(ii) The covariance matrix of $X(n)$ and $X(m)$ depends only on the difference of times n and m, that is, there exists a matrix function, called *correlation matrix function*, R: $\{-2N, \ldots, 0, \ldots, 2N\} \to M(d; \mathbf{R})$ such that

$$E[X(n)^t X(m)] = R(n-m) \quad (-N \leq n, m \leq N). \tag{10}$$

2.2. The fluctuation-dissipation theorem for local and weakly stationary process

For a local and weakly stationary process $\mathbf{X}$, there exist some special relations among the elements of KM_2O-Langevin matrix, called *the fluctuation-dissipation theorem.* The fluctuation-dissipation theorem is stated as follows.

THEOREM 2.1. ([3], [4]) *Among the elements of KM_2O-Langevin matrix $\mathcal{LM}(\mathbf{X})$ associated with a local and weakly stationary process $\mathbf{X}$, the following relations hold.*

$$V_{\pm}(0) = R(0), \tag{11}$$

$$\delta_{\pm}(1) = -R(\pm 1) R(0)^{-1} \tag{12}$$

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

and for $n(1 \leq n \leq N)$

$$V_{\pm}(n) = (I - \delta_{\pm}(n)\delta_{\mp}(n))V_{\pm}(n-1), \tag{13}$$

$$\gamma_{\pm}(n,k) = \gamma_{\pm}(n-1,k-1) + \delta_{\pm}(n)\gamma_{\mp}(n-1,n-k-1), \tag{14}$$

$$\delta_{\pm}(n) = -\left\{R(\pm n) + \sum_{k=0}^{n-2} \gamma_{\pm}(n-1,k)R(\pm(k+1))\right\}V_{\mp}(n-1)^{-1}. \tag{15}$$

What has to be noticed is that the fluctuation-dissipation theorem describes the algorithm to compute the KM_2O-Langevin matrix $\mathcal{LM}(\mathbf{X})$ from the correlation matrix function R.

For a given d-dimensional data $\mathcal{X} = (\mathcal{X}(n); 0 \leq n \leq N)$, we adopt *the sample mean* $\mu^{\mathcal{X}}$ *and the sample correlation matrix function* $R^{\mathcal{X}} = (R^{\mathcal{X}}(n); |n| \leq N)$ defined by

$$\mu^{\mathcal{X}} \equiv \frac{1}{N+1}\sum_{n=0}^{N} \mathcal{X}(n), \tag{16}$$

$$R^{\mathcal{X}}(n) \equiv \frac{1}{N+1}\sum_{m=0}^{N-n} (\mathcal{X}(n+m) - \mu^{\mathcal{X}})^t(\mathcal{X}(m) - \mu^{\mathcal{X}}), \quad (0 \leq n \leq N) \tag{17}$$

$$R^{\mathcal{X}}(-n) \equiv {}^tR^{\mathcal{X}}(n), \qquad (0 \leq n \leq N) \tag{18}$$

as an estimator of the correlation matrix function R.

2.3. Nonlinear KM_2O-Langevin equation

Here, we deal with an $\mathbf{R}$-valued strictly stationary process $\mathbf{X} = (X(n); n \in \mathbf{Z})$ that satisfies the following three conditons:

(H1) $\mathbf{X}$ is bounded, that is, there exists a real number $c > 0$ such that for all $n \in \mathbf{Z}$, $|X(n)(\omega)| \leq c$ holds for a.s. $\omega \in \Omega$.

(H2) For any $k \in \mathbf{N}$ and $n_j \in \mathbf{Z}$ $(1 \leq j \leq k)$, $n_1 < n_2 < \cdots < n_k$, the support of the probability distribution of the random variable ${}^t(X(n_1), \ldots, X(n_k))$ has a positive Lebesgue measure.

(H3) $E[X(n)] = 0$, $(n \in \mathbf{Z})$.

From the theory of KM_2O-Langevin equations, we can take in the nonlinearity of time series in the form of polynomials. This is based on the theory of nonlinear prediction analysis for strictly stationary processes by Masani and Wiener [2]. The practical computable algorithm of that theory

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

is obtained in [8]. Exactly, we adopt the group of polynomials shown by Table 1 as nonlinear transformation.

0	$X(n)$
1	$X(n)^2$
2	$X(n)^3$
3	$X(n) \cdot X(n-1)$
4	$X(n)^4$
5	$X(n)^2 \cdot X(n-1)$
6	$X(n) \cdot X(n-2)$
7	$X(n)^5$
8	$X(n)^3 \cdot X(n-1)$
9	$X(n)^2 \cdot X(n-2)$

10	$X(n) \cdot X(n-1)^2$
11	$X(n) \cdot X(n-3)$
12	$X(n)^6$
13	$X(n)^4 \cdot X(n-1)$
14	$X(n)^3 \cdot X(n-2)$
15	$X(n)^2 \cdot X(n-1)^2$
16	$X(n)^2 \cdot X(n-3)$
17	$X(n) \cdot X(n-1) \cdot X(n-2)$
18	$X(n) \cdot X(n-4)$

Table 1: Nonlinear transformations

These polynomials are part of infinite number of polynomials that form the basis of closed subspace $\mathbf{N}^n_{-\infty}(\mathbf{X})$ of Hilbert space $L^2(\Omega, \mathcal{B}, P)$,

$$\mathbf{N}^n_{-\infty}(\mathbf{X}) \equiv \{Y \in L^2(\Omega, \mathcal{B}, P); \quad Y \text{ is a } \mathcal{B}^n_{-\infty}(\mathbf{X})\text{-measurable random variable}\}, \tag{19}$$

where

$$\mathcal{B}^n_{-\infty}(\mathbf{X}) \equiv \sigma(\{X(m);\ m \le n,\ 1 \le j \le d\}) \quad (n \in \mathbf{Z}). \tag{20}$$

We define the subset $\mathbf{K}^0_{-\infty}(\mathbf{X})$ of $L^2(\Omega, \mathcal{B}, P)$ by

$$\mathbf{K}^0_{-\infty}(\mathbf{X}) \equiv \left\{ \prod_{k=0}^{m} X(n-k)^{p_k} - E\left(\prod_{k=0}^{m} X(n-k)^{p_k}\right); \quad m \in \mathbf{N}^*,\ n \le 0,\ p_0 \in \mathbf{N},\ p_k \in \mathbf{N}^* (1 \le k \le m) \right\}. \tag{21}$$

We can obtain the following relation between $\mathbf{N}^0_{-\infty}(\mathbf{X})$ and $\mathbf{K}^0_{-\infty}(\mathbf{X})$:

$$P_{\mathbf{N}^0_{-\infty}(\mathbf{X})} X(n) = P_{[\mathbf{K}^0_{-\infty}(\mathbf{X})]} X(n) \quad (n \in \mathbf{N}). \tag{22}$$

Next, we parametrize all the elements of $\mathbf{K}^0_{-\infty}(\mathbf{X})$ as follows.

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

Step 1: We define the subspace Λ of sequences whose terms are non-negative integer by

$$\Lambda \equiv \{\boldsymbol{p} = (p_0, p_1, p_2, \ldots);\ p_0 \geq 1,\ p_i \geq 0\ (i \geq 1), \text{ and } p_i = 0 \text{ without finite number of exceptions}\}. \tag{23}$$

For each $\boldsymbol{p} \in \Lambda$, we introduce a one-dimensional strictly stationary process $\boldsymbol{\varphi}_{\boldsymbol{p}} = (\varphi_{\boldsymbol{p}}(n); n \in \mathbf{Z})$ by

$$\varphi_{\boldsymbol{p}}(n) \equiv \prod_{k=0}^{\infty} X(n-k)^{p_k} \tag{24}$$

and define the subset G of stochastic processes as

$$G \equiv \{\boldsymbol{\varphi}_{\boldsymbol{p}};\ \boldsymbol{p} \in \Lambda\}. \tag{25}$$

Step 2: For all $q \in \mathbf{N}$, we define $\Lambda_q (\subset \Lambda)$ and $G_q (\subset G)$ by

$$\Lambda_q \equiv \left\{ \boldsymbol{p} \in \Lambda;\ \sum_{k=0}^{\infty} (k+1) p_k = q \right\}, \tag{26}$$

$$G_q \equiv \{\boldsymbol{\varphi}_{\boldsymbol{p}};\ \boldsymbol{p} \in \Lambda_q\}. \tag{27}$$

We can express Λ and G as the direct sum of Λ_q, $q = 1, 2, \ldots$ and G_q, $q = 1, 2, \ldots$, respectively:

$$\Lambda = \sum_{q \in \mathbf{N}} \Lambda_q, \tag{28}$$

$$G = \sum_{q \in \mathbf{N}} G_q. \tag{29}$$

Step 3: We introduce the lexicographical order into Λ. For any $\boldsymbol{p}, \boldsymbol{p}' \in \Lambda$, we arrange them by comparing $q = \sum (k+1) p_k$, $q' = \sum (k+1) p'_k$. That is, if $q > q'$, we denote $\boldsymbol{p} > \boldsymbol{p}'$. If $q = q'$, we arrange them by comparing $\{p_k\}$ and $\{p'_k\}$ according to the lexicograpical order.

Step 4: In the same way, we arrange all the elements of G as follows

$$G = \{\boldsymbol{\varphi}_j;\ j = 0, 1, 2, \ldots\}. \tag{30}$$

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

The polynomials shown in Table 1 are the first nineteen elements of G_6 that are arranged according to the above order. Note that $\varphi_0 = \mathbf{X}$. Roughly speaking, these polynomials are arranged in the order of similarity to the original strictly stationary process $\mathbf{X} = (X(n); n \in \mathbf{Z})$.

We define a $d+1$-dimensional stochastic process $\mathbf{X}^{(d)}_{\text{nonl}} = (X^{(d)}_{\text{nonl}}(n); n \in \mathbf{Z})$ by

$$X^{(d)}_{\text{nonl}}(n) = \begin{pmatrix} \varphi_0(n) \\ \varphi_1(n) - E[\varphi_1(n)] \\ \vdots \\ \varphi_d(n) - E[\varphi_d(n)] \end{pmatrix}. \tag{31}$$

The following theorem shows that the closed subspace $\mathbf{K}^0_{-\infty}(\mathbf{X})$ can be approximated by the linear subspaces $\mathbf{M}^0_{-N}(\mathbf{X}^{(d)}_{\text{nonl}})$ $(N \in \mathbf{N}^*, d \in \mathbf{N})$.

THEOREM 2.2. ([8]) *The following relation holds:*

$$\left[\bigcup_{N=0}^{\infty}\bigcup_{d=1}^{\infty} \mathbf{M}^0_{-N}(\mathbf{X}^{(d)}_{\text{nonl}})\right] = [\mathbf{K}^0_{-\infty}(\mathbf{X})]. \tag{32}$$

Given a one-dimensional time series $\mathcal{X} = (\mathcal{X}(n); 0 \leq n \leq N)$, we constitute various kinds of multi-dimensional time series, for example,

$$\mathcal{X}^{(3)}_{\text{nonl}} \equiv ({}^t(\mathcal{X}(n), \mathcal{X}(n)^2, \mathcal{X}(n)\mathcal{X}(n-1)); 1 \leq n \leq N). \tag{33}$$

We represent the above nonlinear transformation by the symbol of $(0,1,3)$-*type*. Now, we can derive a KM_2O-Langevin equation from the transformed time series, which we call *nonlinear KM_2O-Langevin equation.*

§3. Causality analysis

In this section we give a short review of the causality analysis in the theory of KM_2O-Langevin equations and its applications to real data [7]. We also give a more precise criterion for Test(CS) in view of statistical data analysis.

3.1. The definition of causality

Let $(\Omega, \mathcal{B}, P)$ be a probability space. Suppose that we have two $\mathbf{R}^d$-valued stochastic process $\mathbf{X} = (X(n); -\infty < n \leq r)$ and $\mathbf{R}$-valued stochastic process $\mathbf{Y} = (Y(n); -\infty < n \leq r)$ on $(\Omega, \mathcal{B}, P)$. We define the causality as follows.

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

DEFINITION 3.1. There exists causality that $\mathbf{X}$ is cause and $\mathbf{Y}$ is result if for all $n(-\infty < n \leq r)$, there exists a Borel function F_n with infinite number of variables such that

$$Y(n) = F_n(X(n), X(n-1), \ldots) \quad \text{a.s.} \tag{34}$$

We denote the above relation as

$$\mathbf{X} \xrightarrow{(C)} \mathbf{Y}. \tag{35}$$

If each component of $\mathbf{X}$ and $\mathbf{Y}$ is square integrable, then the definition of causality is equivalent to

$$\mathbf{N}^n_{-\infty}(\mathbf{Y}) \subset \mathbf{N}^n_{-\infty}(\mathbf{X}) \qquad (-\infty < \forall n \leq r). \tag{36}$$

3.2. Linear causality

We define the linear causality as a special case of causality defined above, that is, F_n in (34) is linear. The linear causality that $\mathbf{X}$ is cause and $\mathbf{Y}$ is result is equivalent to

$$\mathbf{M}^n_{-\infty}(\mathbf{Y}) \subset \mathbf{M}^n_{-\infty}(\mathbf{X}) \qquad (-\infty < \forall n \leq r). \tag{37}$$

We denote the above relation as

$$\mathbf{X} \xrightarrow{(LC)} \mathbf{Y}. \tag{38}$$

The algorithm to check the linear causality is given in the theory of KM_2O-Langevin equations. Let $\mathbf{X} = (X(n); -\infty < n \leq r)$ and $\mathbf{Y} = (Y(n); -\infty < n \leq r)$ be $\mathbf{R}^d$-valued and $\mathbf{R}$-valued stochastic processes on $(\Omega, \mathcal{B}, P)$, respectively. We assume that

$$\mathbf{U} = (U(n); n \in \mathbf{Z}) \equiv ({}^t(Y(n), {}^tX(n)); n \in \mathbf{Z}) \tag{39}$$

is a $d+1$ dimensional weakly stationary process with mean 0. We define three correlation matrix functions R_1, R_2, R_3 by

$$R_1(n) \equiv E(X(n){}^tX(0)) \in M(d, d; \mathbf{R}), \tag{40}$$

$$R_2(n) \equiv E(Y(n){}^tX(0)) \in M(1, d; \mathbf{R}), \tag{41}$$

$$R_3(n) \equiv E(Y(n)Y(0)) \in M(1, 1; \mathbf{R}) \tag{42}$$

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

and define $C_*(\mathbf{Y}|\mathbf{X}) = C_n(\mathbf{Y}|\mathbf{X}): \mathbf{N}^* \longrightarrow [0, \infty)$ by

$$C_n(\mathbf{Y}|\mathbf{X}) \equiv \frac{\{E[(P_{\mathbf{M}_0^n(\mathbf{X})}Y(n))^2]\}^{1/2}}{\sqrt{R_3(0)}}, \tag{43}$$

which we call *causality function.* The following theorem gives us a method of checking linear causality quantitatively.

THEOREM 3.1. ([7]) *The following two statements are equivalent:*

(i) $\mathbf{X} \xrightarrow{(LC)} \mathbf{Y}$,
(ii) $C_n(\mathbf{Y}|\mathbf{X}) \nearrow 1$ *as* $n \to \infty$.

Taking the above theorem into account, we define

$$C_\infty(\mathbf{Y}|\mathbf{X}) \equiv \lim_{n\to\infty} C_n(\mathbf{Y}|\mathbf{X}) \tag{44}$$

and call it *causality value* from $\mathbf{X}$(cause) to $\mathbf{Y}$(result).

THEOREM 3.2. ([7]) *The causality value is calculated from the KM_2O-Langevin matrix $\mathcal{LM}(\mathbf{X})$ associated with $\mathbf{X}$ as follows: for all n $(0 \le n < \infty)$*

$$C_n(\mathbf{Y}|\mathbf{X}) = \frac{\{\sum_{k=0}^n C(n,k)V_+(k)\ {}^tC(n,k)\}^{1/2}}{\sqrt{R_3(0)}}, \tag{45}$$

where $1 \times d$ matrices $C(n,k)$, $(0 \le k \le n)$ are given by

$$C(n,k) = \begin{cases} R_2(n)R_1(0)^{-1} & (k=0), \\ \{R_2(n-k) + \sum_{l=0}^{k-1} R_2(n-l)\,{}^t\gamma_+(k,l)\}V_+(k)^{-1} & (k \ge 1). \end{cases} \tag{46}$$

3.3. Nonlinear causality

As a generalization of the results in [7], we consider the general case of causality, that is, F_n in the definition of causality (34) is nonlinear here.

The fundamental point of the method of checking nonlinear causality is to approximate F_n by a linear combination of polynomials. Let $\mathbf{X} = (X(n); n \in \mathbf{Z})$ and $\mathbf{Y} = (Y(n); n \in \mathbf{Z})$ be two $\mathbf{R}$-valued stochastic processes. Suppose that $\mathbf{X}$ is a strictly stationary process that satisfies the conditions (H1) and (H2) in section 2.4 and $\mathbf{U} = ({}^t(Y(n), X(n)); n \in \mathbf{Z})$ is a strictly stationary process. As a refinement of Theorem 3.1, we obtain the following theorem.

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

THEOREM 3.3. *These two statements are equivalent:*

(i) $\mathbf{X} \xrightarrow{(C)} \mathbf{Y}$,

(ii) $C_\infty(\mathbf{Y}|\mathbf{X}^{(d)}_{\text{nonl}})/\sqrt{R_3(0)} \nearrow 1$ *as* $d \to \infty$.

Proof. From the definition of nonlinear causality (34), $\mathbf{X} \xrightarrow{(C)} \mathbf{Y}$ is equivalent to

$$P_{\mathbf{N}^0_{-\infty}(\mathbf{X})}Y(0) = Y(0),$$

that is,

$$E[\{P_{\mathbf{N}^0_{-\infty}(\mathbf{X})}Y(0)\}^2] = E[(Y(0))^2]. \tag{47}$$

From the stationarity of $\mathbf{U}$,

$$\begin{aligned} C_N(\mathbf{Y}|\mathbf{X}^{(d)}_{\text{nonl}}) &= E[\{P_{\mathbf{M}^N_0(\mathbf{X}^{(d)}_{\text{nonl}})}Y(N)\}^2] \\ &= E[\{P_{\mathbf{M}^0_{-N}(\mathbf{X}^{(d)}_{\text{nonl}})}Y(0)\}^2]. \end{aligned}$$

This value increases monotonically with respect to N and d. Thus, we obtain

$$\begin{aligned} \lim_{d,N\to\infty} C_N(\mathbf{Y}|\mathbf{X}^{(d)}_{\text{nonl}}) &= \lim_{d\to\infty}\Big(\lim_{N\to\infty} C_N(\mathbf{Y}|\mathbf{X}^{(d)}_{\text{nonl}})\Big) \\ &= \lim_{d\to\infty} C_\infty(\mathbf{Y}|\mathbf{X}^{(d)}_{\text{nonl}}). \end{aligned}$$

Applying Theorem 2.2, we obtain

$$E[\{P_{\mathbf{M}^0_{-N}(\mathbf{X}^{(d)}_{\text{nonl}})}Y(0)\}^2] \nearrow E[\{P_{\mathbf{K}^0_{-\infty}(\mathbf{X})}Y(0)\}^2] \text{ as } d,\ N \to \infty. \tag{48}$$

Since the mean vector of $Y(0)$ is 0, we can get from (22)

$$P_{\mathbf{N}^0_{-\infty}(\mathbf{X})}Y(0) = P_{[\mathbf{K}^0_{-\infty}(\mathbf{X})]}Y(0). \tag{49}$$

Therefore, by (47), (48), (49), we can complete the proof. □

Thus, we can check the nonlinear causality approximately by reducing it to the linear causality between $\mathbf{X}^{(d)}_{\text{nonl}}$(cause) and $\mathbf{Y}$(result).

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

3.4. Data analysis: Test(CS)

Let $\mathbf{X}$ and $\mathbf{Y}$ be d-dimensional and one-dimensional stochastic process, respectively, and $\mathcal{X} = \{\mathcal{X}(n);\ 0 \leq n \leq N\}$ and $\mathcal{Y} = \{\mathcal{Y}(n);\ 0 \leq n \leq N\}$ be their corresponding time series. We assume that the $(d+1)$-dimensional data $({}^t(\mathcal{Y}(n), {}^t\mathcal{X}(n)); 0 \leq n \leq N)$ passes Test(S) and can be regarded as a realization of $(d+1)$-dimensional stationary process ${}^t(Y(n), {}^tX(n))$. We discuss here an application of causality analysis to actual data analysis, which we call *Test(CS)*.

We can estimate the causal function $C_n(\mathbf{Y}|\mathbf{X})$ from data sets using the algolithm of the fluctuation-dissipation theorem and Theorem 3.2. We denote this estimator for the causal function by $\hat{C}_n(\mathbf{Y}|\mathbf{X})$ or $\hat{C}_n(\mathcal{Y}|\mathcal{X})$. These estimator are obtained by replacing the correlation matrix functions in (46) with the sample correlation matrix function defined by (17) and (18).

In data analysis, however, we find some difficulties that come from the fact that the number of available data is finite. First, the number of reliable sample correlation matrix function in (17), (18) is limitted. From the empirical knowledge in time series analysis, the domain of reliable sample correlation matrix function of $({}^t(\mathcal{Y}(n), {}^t\mathcal{X}(n)); 0 \leq n \leq N)$ is limitted to $\{-M, \ldots, 0, \ldots, M\}$, where

$$M \equiv [3\sqrt{N+1}/(d+1)] - 1 \ (< N). \tag{50}$$

Secondly, we cannot calculate the limit in the definition of causality value (44). Instead, we replace $n \to \infty$ with $n = M$, and adopt $\hat{C}_M(\mathcal{Y}|\mathcal{X})$ as an approximate value to the causality value $\hat{C}_\infty(\mathcal{Y}|\mathcal{X})$. We call $\hat{C}_M(\mathcal{Y}|\mathcal{X})$ *sample causality value.* Given two $\mathbf{R}$-valued time series $\mathcal{Y}$ and $\mathcal{X}$, it should be noted that we must check the stationarity of $({}^t(\mathcal{Y}(n), {}^t\mathcal{X}^{(d)}_{\text{nonl}}(n)); 0 \leq n \leq \tilde{N})$ as well as $\mathcal{X}$ when we calculate $\hat{C}_{\tilde{M}}(\mathcal{Y}|\mathcal{X}^{(d)}_{\text{nonl}})$, where $\tilde{M} = [3\sqrt{\tilde{N}+1}/d+1] - 1$.

In actual data analysis, in order to check nonlinear causality, we calculate all sample causality values between $\mathcal{Y}$ and possible nonlinear transformations $\mathcal{X}^{(d)}_{\text{nonl}}$, and find the maximum sample causality value.

Actually, because of finiteness of available data, we can only approximate the underlying dynamics by nonlinear transformations. Therefore, in actual data analysis, the sample causality value $\hat{C}_M(\mathcal{Y}|\mathcal{X}^{(d)}_{\text{nonl}})$ does not reach 1 even if there exists a causal relation between given two time series.

Therefore, it is necessary to interpret the sample causality value according to some criteria.

We must give up the idea of requiring a sample causality value to be exact 1 even when there exists a causal relation. Instead, we check if sample causality value is large enough for the two time series to have a causal relation. In the following subsections, we describe some ideas of checking this criteria.

3.5. Selection of effective variables

It gives us useful information to plot a sample causality function against time. If the sample causality function is observed to saturate at some time n in the plot, it is considered that the dynamics depends only on n steps before, and the shortage of sample causality value toward 1, that is, $1 - \hat{C}_n(\mathcal{Y}|\mathcal{X})$, is due to the fluctuation.

We can find an integer $n = M'$ at which the sample causality function $\hat{C}_n(\mathcal{Y}|\mathcal{X})$ exceeds the 95% of $\hat{C}_M(\mathcal{Y}|\mathcal{X})$ for the first time. We adopt $\hat{C}_{M'}(\mathcal{Y}|\mathcal{X})$ as a new sample causality value insted of that of $n = M$, and rewrite M' as M.

3.6. Estimation of the confidence interval of a sample causality value

In this subsection, we give a method of estimating the confidence interval of sample causality value based on the Bootstrap method [10], which has been frequently used as a nonparametric method in statistical data analysis.

(i) It follows from the definition of the causality function (43) and the strictly stationarity of $\mathbf{U} = ({}^t(Y(n), {}^tX(n)); n \in \mathbf{Z})$ that the causality value $\hat{C}_M(\mathbf{Y}|\mathbf{X})$ depends on only the joint probability of $(Y(n), X(n), \ldots, X(n-M))$. The estimation of this joint probability distribution is given by the following empirical distribution

$$\hat{\rho}(A) = \frac{1}{N-M+1} \sum_{n=M}^{N} \delta_{(\mathcal{Y}(n),\mathcal{X}(n),\cdots,\mathcal{X}(n-M))}(A), \qquad A \in \mathcal{B}(\mathbf{R}^{d(M+1)+1}), \tag{51}$$

where δ_z is the delta measure whose support is centered at z.

(ii) Next, we choose $(d(M+1)+1)$-dimensional $(N-M+1)$ samplings with displacement from the empirical distribution $\hat{\rho}$. The concrete procedure of choosing samples is that we select randomly an integer n from

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

$\{M, \ldots, N\}$ and adopt the corresponding data set $(\mathcal{Y}(n), \mathcal{X}(n), \ldots, \mathcal{X}(n-M))$ as a new sample.

(iii) Since these samples are considered to be $\overset{\text{i.i.d}}{\sim} \hat{\rho}$, we can compute the quantities which might be called bootstrapped sample causality value. We repeat this procedure B times, and obtain the sequence of sample causality values $\hat{C}_1, \ldots, \hat{C}_B$.

(iv) We can make a statistical inference of sample causality value $\hat{C}_M(\mathbf{Y}|\mathbf{X})$ by making the emprical distiribution of $\hat{C}_1, \ldots, \hat{C}_B$. For example, we can obtain a confidence interval of $\hat{C}_M(\mathbf{Y}|\mathbf{X})$ by approximating the distribution of $\hat{C}_M(\mathbf{Y}|\mathbf{X}) - C_M(\mathbf{Y}|\mathbf{X})$ by the empirical distiribution of $\hat{C}_i - \hat{C}_M(\mathcal{Y}|\mathcal{X})$, $i = 1, \ldots, B$.

3.7. Criterion for the causal relations between two time series

In this subsection, we give a method of testing the null hypothesis that there are no causal relations between two time series.

The fundamental idea is that we compare the confidence interval of sample causality value with the distribution of sample causality values under the condition that there are no causal relations, and if these two distributions are quite different from each other, we can reject the null hypothesis. If the causal time series $\mathcal{X}$ is replaced by a physical random sequence $\mathcal{R}$, there would be no causal relations between $\mathcal{Y}$ and $\mathcal{R}$. In this case, we can make the distribution of sample causality values by replacing the physical random sequence $\mathcal{R}$ one after another.

Let $F_{1-2\alpha}$ be the upper $100(1-2\alpha)\%$ point of the distribution of sample causality values between $\mathcal{Y}$ and $\mathcal{R}$ and $L_{1-2\alpha}$ be the lower bound of the confidence interval of sample causality value in question. Comparing these two quantities, we can reject the null hypothesis if

$$L_{1-2\alpha} > F_{1-2\alpha}. \tag{52}$$

§4. Finding underlying dynamics in time series

In this section, we describe a procedure to investigate the deterministic dynamics of time series.

4.1. Finding deterministic dynamics in time series: Test(D)

Suppose that we have a one-dimensional time series $\mathcal{X} = (\mathcal{X}(n); 0 \leq n \leq N)$ which is judged to be stationary by Test(S). The data $\mathcal{X}$ can be regraded as a realization of the underlying stationary stochastic process

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

$\mathbf{X} = (X(n); 0 \leq n \leq N)$. We define another stationary stochastic process $\mathbf{X}_{+1} = (X_{+1}(n); 0 \leq n \leq N-1)$ by

$$X_{+1}(n) \equiv X(n+1), \quad (0 \leq n \leq N-1) \tag{53}$$

that is, we obtain $\mathbf{X}_{+1}$ by shifting time domain by one step into future. $\mathbf{X}$ is called *deterministic* if there exists causality that $\mathbf{X}$ is cause and $\mathbf{X}_{+1}$ is result:

$$\mathbf{X} \xrightarrow{(C)} \mathbf{X}_{+1}. \tag{54}$$

We call the above method of checking determinism in time series *Test(D).*

We should notice that the determinism of time series defined here does not necessarily imply that it comes from a deterministic dynamic system, such as chaos. According to the definition of causality value (44), it means that the fluctuation flow occupies only small part of (nonlinear) KM_2O-Langevin equation derived from $\mathcal{X}$ or $\mathcal{X}^{(d)}_{\text{nonl}}$. Therefore, even a time series that comes from AR model can be judged to be deterministic by our method.

4.2. Finding chaotic dynamics in time series

Let $\mathcal{X} = (\mathcal{X}(n); 0 \leq n \leq N)$ be a one-dimensional time series which is judged to be deterministic by Test(D). Our concern here is to consider whether the dynamics behind this time series is chaotic.

On the basis of deterministic time series prediction, it is argued in [1] that we can distinguish chaos from uncorrelated measurement error by time series prediction. Specifically, if prediction precision falls off as prediction step becomes large, though near future of time series can be predicted with sufficiently high precision, the dynamics is thought to be chaotic. Whereas, if the prediction precision is independent of prediction step, the randomness is due to measurement error.

In the same way, we can investigate the underlying chaotic dynamics from the viewpoint of prediction analysis based on the theory of KM_2O-Langevin equations as follows. First, we find the nonlinear KM_2O-Langevin equation that approximates the time evolution of $\mathcal{X}$ the most precisely. Secondly, we divide the time series into two parts, one is for reference for prediction, and another is for comparison of predicted data with real data. We fix the prediction step, and repeat the prediction of $\mathcal{X}$ using the nonlinear KM_2O-Langevin equation. In order to evaluate the quality of forecasts,

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

two error estimators are introduced. One is the normalized mean square root error $C(T)$ defined by

$$C(T) = \frac{\langle (x_{\text{pred}}(T) - \langle x_{\text{pred}}(T)\rangle)(x_{\text{org}} - \langle x_{\text{org}}\rangle)\rangle}{[\langle (x_{\text{pred}}(T) - \langle x_{\text{pred}}(T)\rangle)^2\rangle \langle (x_{\text{org}} - \langle x_{\text{org}}\rangle)^2\rangle]^{1/2}}, \tag{55}$$

where, T is a prediction step, x_{org} and $x_{\text{pred}}(T)$ denote the the original data and the predicted data with prediction step T, respectively. The other is the normalized centered correlation coefficient $E(T)$ between predicted and real data defined by

$$E(T) \equiv \frac{\langle [x_{\text{pred}}(T) - x_{\text{org}}]^2\rangle^{1/2}}{\langle (x_{\text{org}} - \langle x_{\text{org}}\rangle)^2\rangle^{1/2}}. \tag{56}$$

It follows that $C(T) = 1$ or $E(T) = 0$ means that the prediction is perfect, while $C(T) = 0$ or $E(T) = 1$ means that prediction is quite useless. Finally, we plot the above two quantities against prediction step T, and find the relation between prediction precision and prediction step.

§5. The result of numerical experiments

In this section, we apply our method to time series from the natural world. We treat the measles data and chickenpox data. The original data are not judged to be stationary by Test(S). Therefore, in order to extract stationarity, we transform them before data analysis as follows:

$$\begin{aligned}
\text{Difference:}\quad & \mathcal{V}(n) = \mathcal{U}(n) - \mathcal{U}(n-1), \quad (1 \le n \le N). \\
& \Downarrow \\
\text{Standardization:}\quad & \mathcal{W}(n) = \frac{1}{\sqrt{R^{\mathcal{V}}(0)}}(\mathcal{V}(n) - \mu^{\mathcal{V}}). \\
& \Downarrow \\
\text{Arctangent:}\quad & \mathcal{X}(n) = \arctan \mathcal{W}(n), \quad (0 \le n \le N). \\
& \Downarrow \\
\text{Nonlinear transformation:}\quad & \mathcal{Y}(n) = \mathcal{X}^{(d)}_{\text{nonl}}(n), \quad (0 \le n \le \tilde{N}). \\
& \Downarrow \\
\text{Standardization:}\quad & \mathcal{Z}(n) = \begin{pmatrix} \sqrt{R^{\mathcal{Y}}_{11}(0)^{-1}} & & 0 \\ & \ddots & \\ 0 & & \sqrt{R^{\mathcal{Y}}_{dd}(0)^{-1}} \end{pmatrix} (\mathcal{Y}(n) - \mu^{\mathcal{Y}}), \\
& \qquad\qquad (0 \le n \le \tilde{N}).
\end{aligned}$$

We do not refer to Test(S), but all the time series which we deal with here can be checked to be stationary by Test(S)([4]).

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

As is clear from (50) that gives the number of reliable sample correlation matrix function, a nonlinear transformation with large dimension d overshortens the domain of sample correlation matrix function $R^{\mathcal{X}}$. Therefore, we use only nonlinear KM_2O-Langevin equations whose dimensions are smaller than three.

We repeat the bootstrap procedure $B = 1000$ times for both measles data and chickenpox data.

5.1. Measles

Here, we deal with the time series of monthly cases of measles in New York City from 1928 to 1963. The original data are shown in Figure 1.

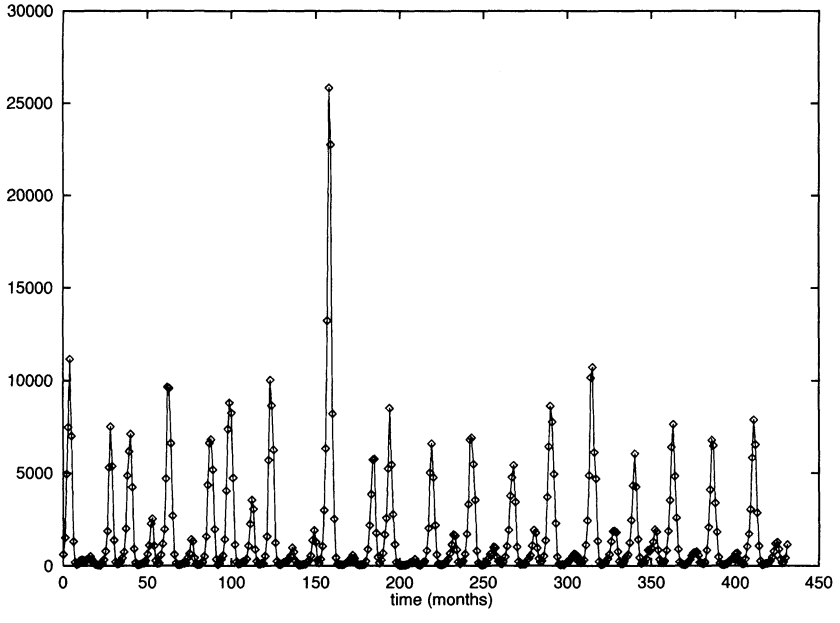

Figure 1: Data of measles

The data have a conspicuous peak around $n = 160$. This suggests that the inner structure of time series changes around $n = 160$. In fact, the whole time series that contains data around $n = 160$ cannot be judged to be stationary for the above transformations. Therefore, we split the time series around $n = 160$ into two parts; one is from $n = 0$ to $n = 150$ (I) and the other is after $n = 160$ (II), and analyse them separately. Both (I) and (II) are judged to be stationary by Test(S) under the above transformations.

First, we perform Test(D) for (I) and for (II). The maximum sample

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

causality values for dimension 1, 2 and 3 are shown in Table 2. Both in (I) and in (II), the nonlinear KM_2O-Langevin equations give larger sample causality values than the linear KM_2O-Langevin equation. This shows that nonlinearity has an important role in the dynamics of measles data.

	dimension	maximum sample causality value	type
data of measles (I)	1	0.818	(0)
	2	0.870	(0, 17)
	3	0.887	(0, 6, 17)
data of measles (II)	1	0.846	(0)
	2	0.891	(0, 18)
	3	0.914	(0, 3, 18)

Table 2: Nonlinearity of data of measles

In Figure 2, we show the plot of the sample causality function of $(0, 6, 17)$-type for (I) and that of $(0, 3, 18)$-type for (II). The plots of the sample causality functions show the tendency to saturate both for (I) and for (II).

We can select the orders of causal analysis $M = 1$ for (I) and $M = 2$ for (II).

In Figure 3 and in Figure 4, we show the distributions of the bootstrapped sample causality values and those where physical random sequences were used as causal data.

data	measles (I)	measles (II)
nonlinear transformation	(0, 6, 17)	(0, 3, 18)
order	$M = 1$	$M = 2$
sample causality value	0.8588	0.8785
confidence interval (confidence coefficient = 0.90)	[0.8086, 0.8879]	[0.8521, 0.8983]
90% point of the distribution for random sequences	0.4150	0.3550

Table 3: Test(D) for the data of measles

We can conclude that both time series are deterministic based on Table 3.

Secondly, we show the result of predictions using three-dimensional nonlinear KM_2O-Langevin equations. The prediction precision falls off both

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

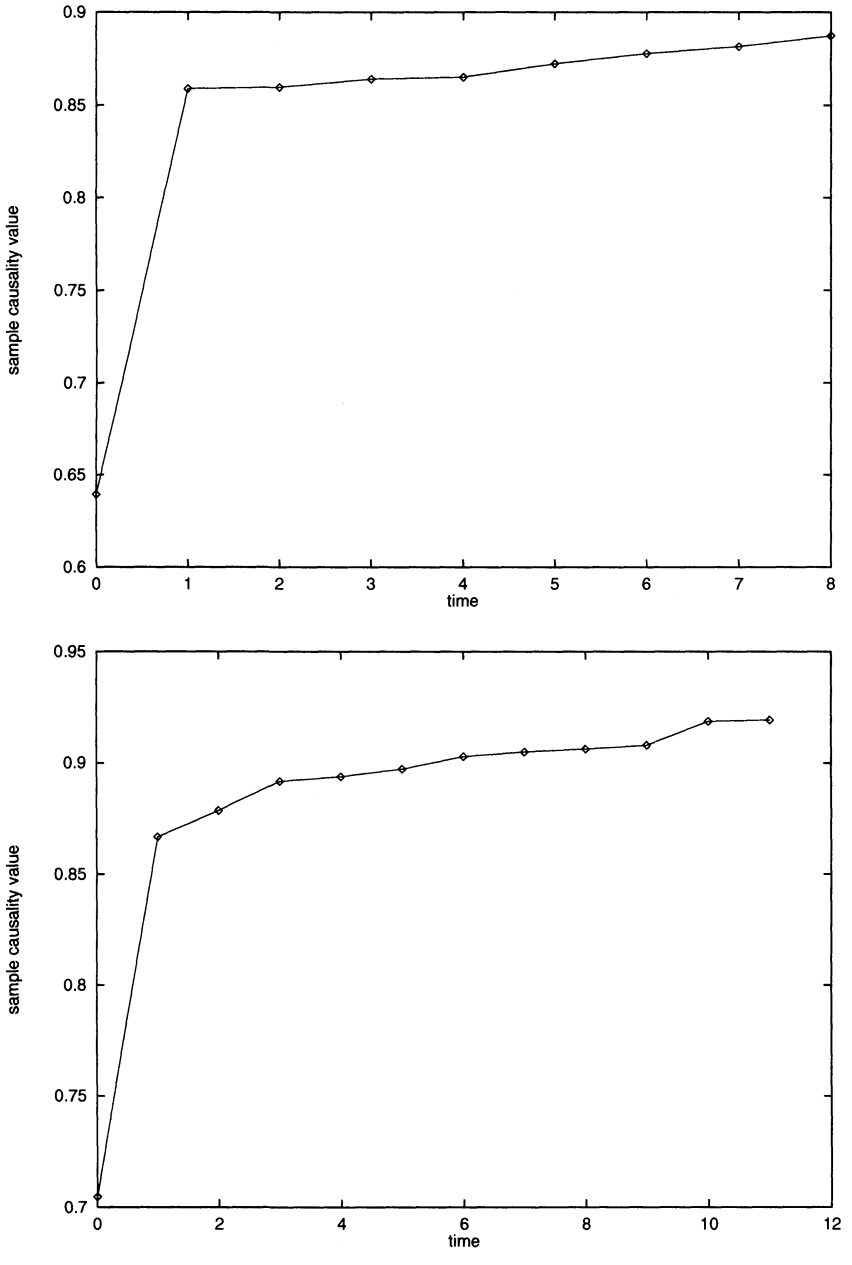

Figure 2: Sample causality functions for measles data (I) (Top) and (II) (Bottom)

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

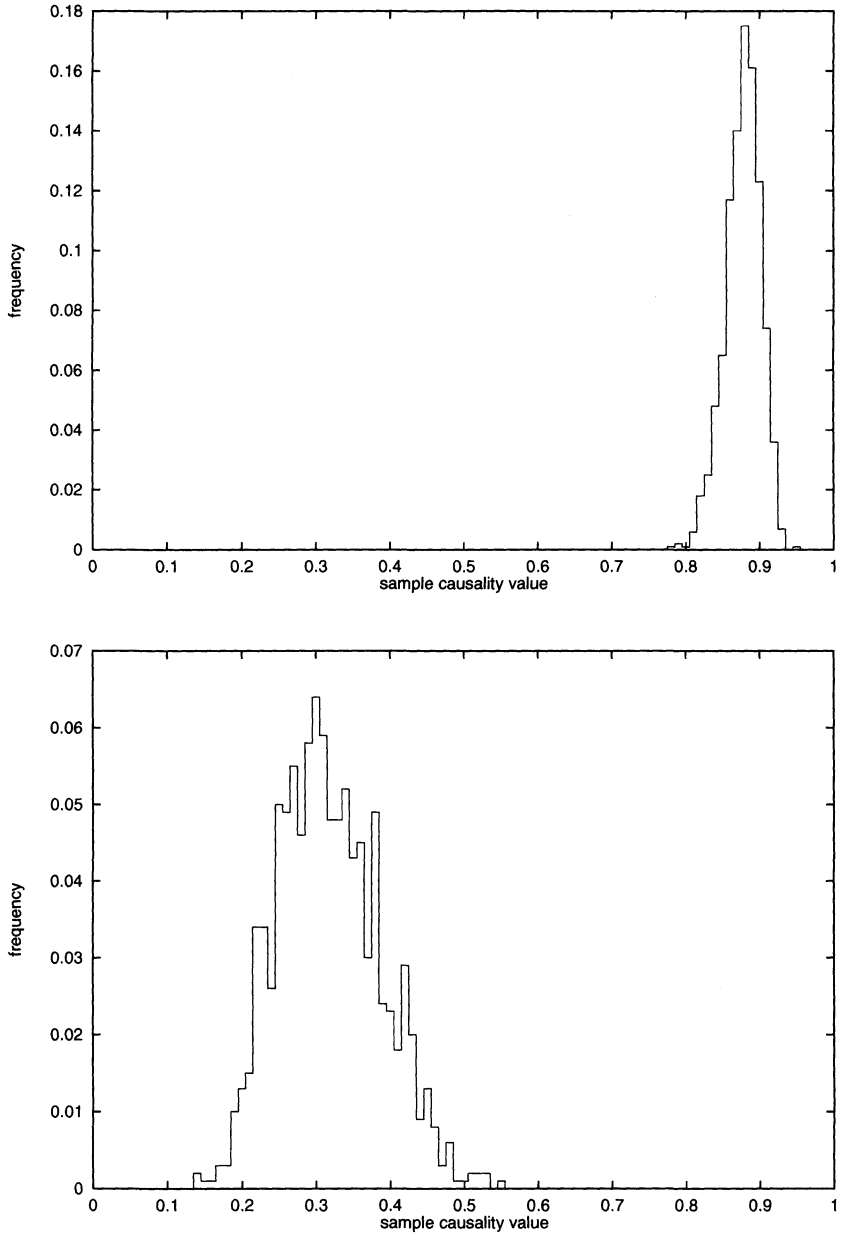

Figure 3: The distribution of the bootstrapped sample causality values (Top) and that of sample causality value between measles data and random sequences (Bottom): (I)

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

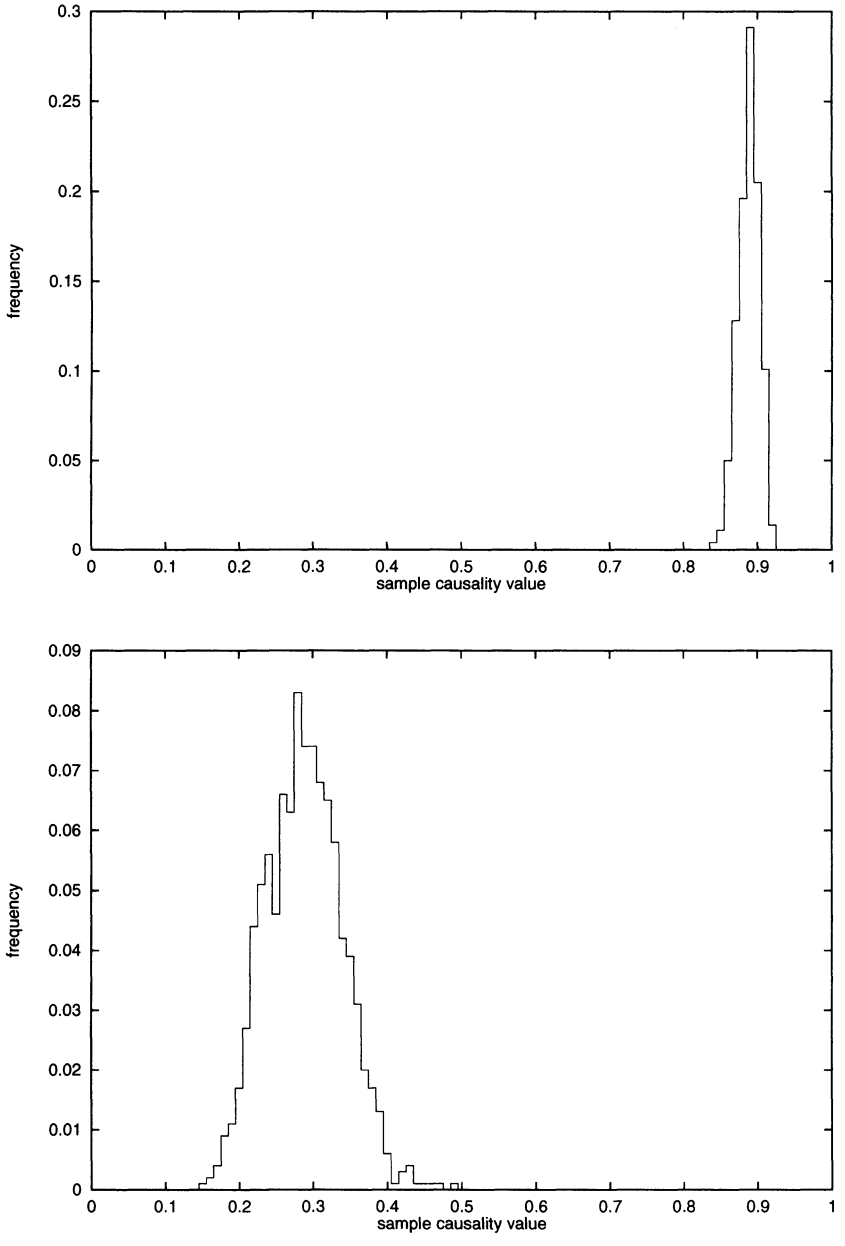

Figure 4: The distribution of the bootstrapped sample causality values (Top) and that of sample causality value between measles data and random sequences (Bottom): (II)

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

in (I) and in (II) as the prediction step increases. Therefore, the dynamics both in (I) and in (II) are thought to be chaotic.

In conclusion, we can assert that the data of measles is not only deterministic but also chaotic. Comparing the result of experiments for (I) and for (II), we can infer that the fluctuation flow of (I) occupies the larger part of the dynamics than that of (II), though the fundamental structure of dynamics seems to be unchanged throughout (I) and (II).

5.2. Chickenpox

Here, we deal with the time series of monthly cases of chickenpox in New York City from 1949 to 1972. The original data are shown in Figure 6.

We repeat the same procedure which was performed for the measles data.

First, we perform Test(D). The maximum sample causality values for dimension 1, 2 and 3 are shown in Table 4. In contranst with measles data, the nonlinear KM_2O-Langevin equations give almost the same sample causality values as the linear KM_2O-Langevin equation. This shows that nonlinearity has no important meaning in the dynamics of chickenpox data.

	dimension	maximum sample causality value	type
data of chickenpox	1	0.861	(0)
	2	0.873	(0, 13)
	3	0.881	(0, 8, 18)

Table 4: Nonlinearity of data of chickenpox

We show in Figure 7 the plot of the sample causality function for the linear KM_2O-Langevin equation. The plot of the sample causality function shows the tendency to saturate.

We can select the orders of causal analysis $M = 8$ for the data of chickenpox.

In Figure 8, we show the distribution of the bootstrapped sample causality values and that where physical random sequences were used as causal data.

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

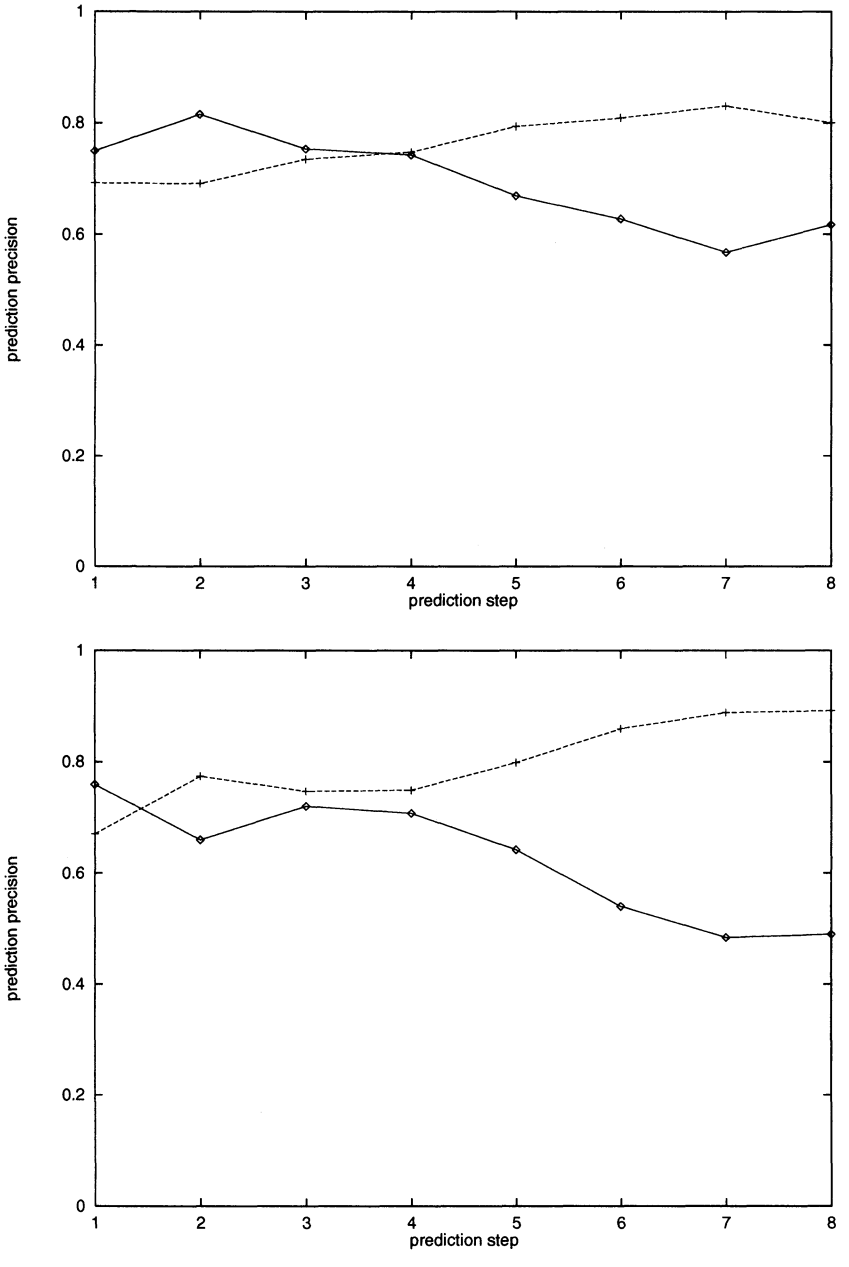

Figure 5: Prediction precision for measles data (I) (Top) and (II) (Bottom). The solid line and the dashed line stand for $C(T)$ and $E(T)$, respectively.

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

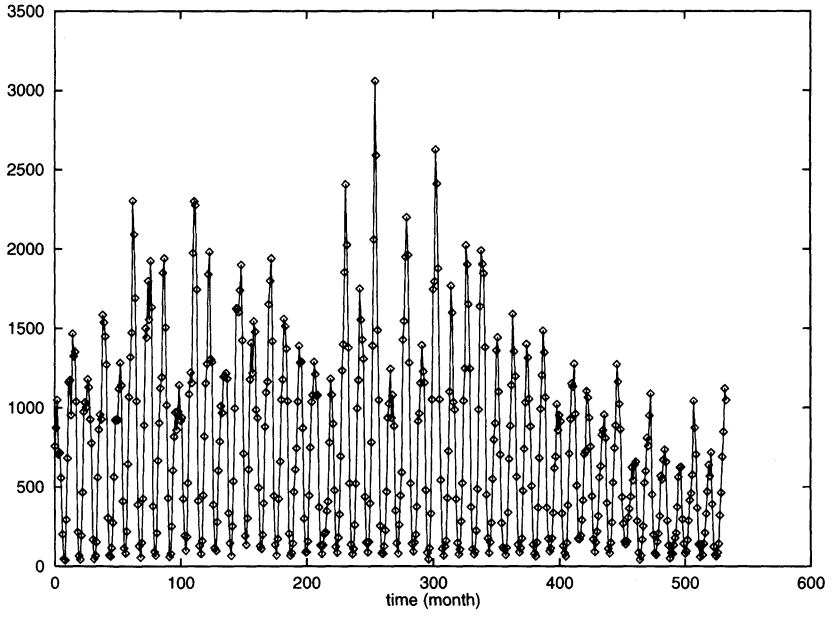

Figure 6: Data of chickenpox

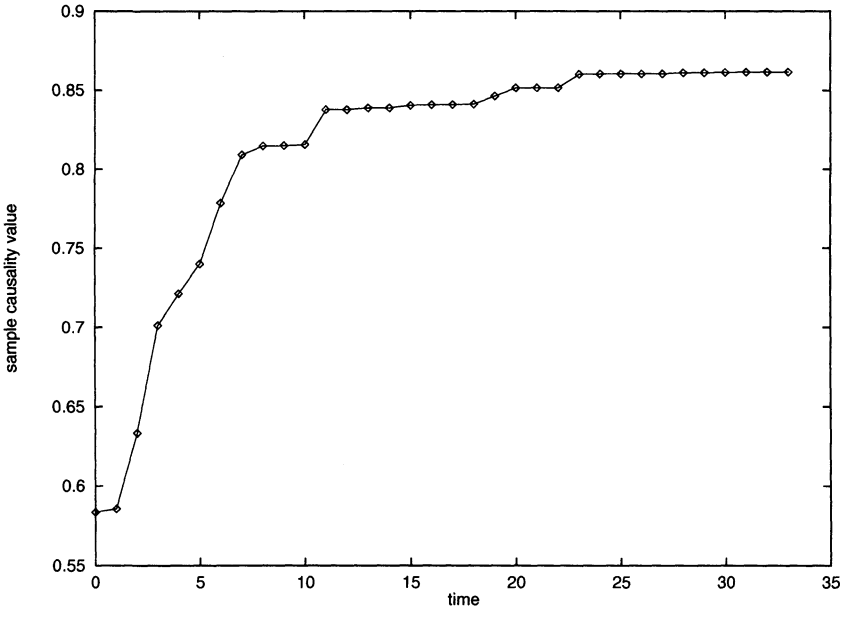

Figure 7: Sample causality function for chickenpox data

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

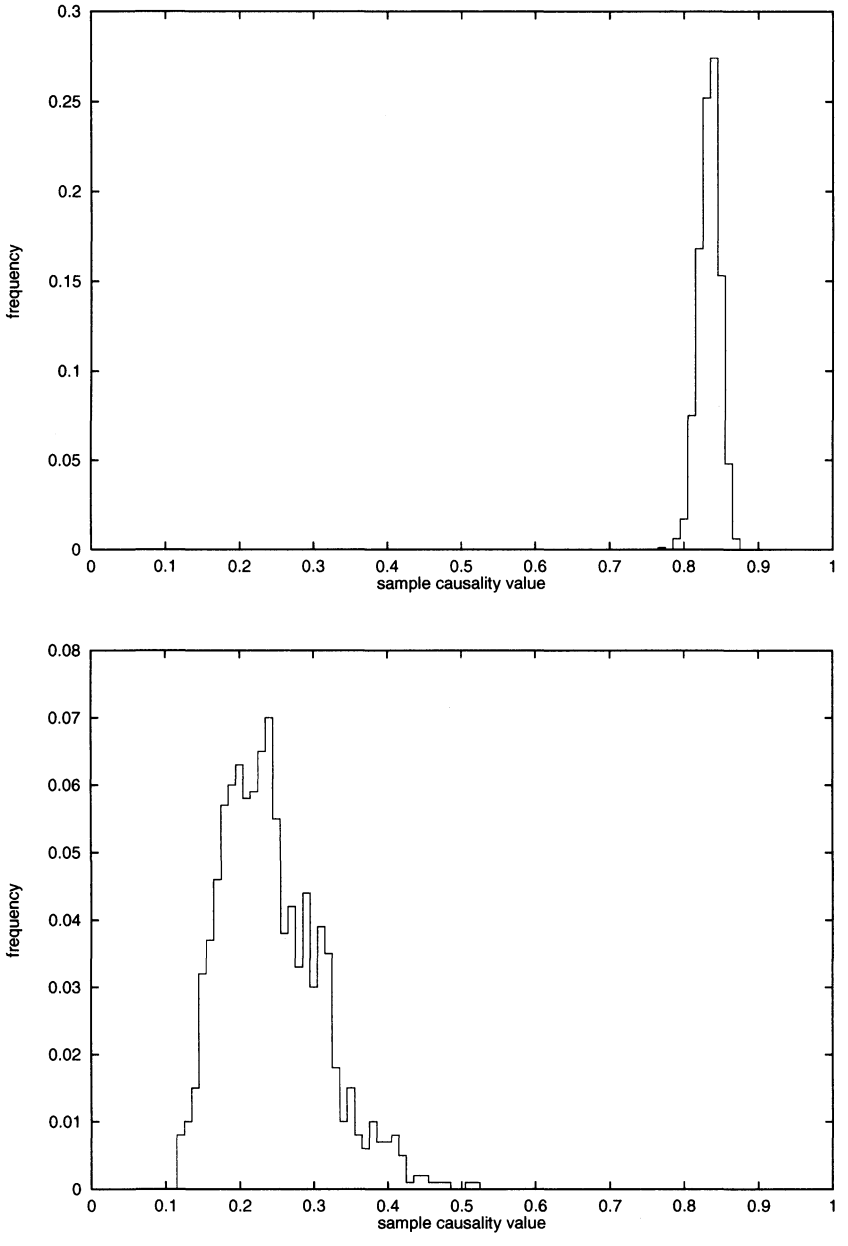

Figure 8: The distribution of the bootstrapped sample causality values (Top) and that of sample causality value between chickenpox data and random sequences (Bottom)

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

data	chickenpox
nonlinear transformation	(0)
order	$M = 8$
sample causality value	0.8146
confidence interval(confidence coefficient = 0.90)	[0.7789, 0.8227]
90% point of the distribution for random sequences	0.3350

Table 5: Test (D) for the data of chickenpox

We can conclude that time series of chickenpox is deterministic based on Table 5.

Secondly, we show the result of predictions using linear KM_2O-Langevin equation. The prediction precision is independent of the prediction step. Therefore, the dynamics in chickenpox data is thought to be governed not by chaos but by stochastic error that can be well described by linear models.

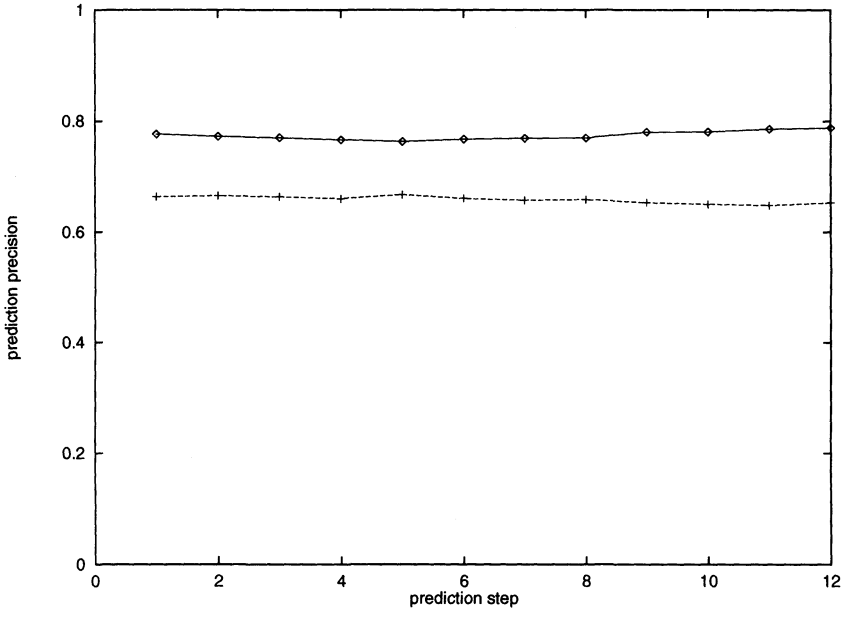

Figure 9: Prediction precision for chickenpox data. The solid line means correlation, and the dashed line means mean square root error.

In conclusion, we can assert that the chickenpox data are deterministic but its randomness is stochastic that can be treated by linear models.

https://doi.org/10.1017/S002776300000684X Published online by Cambridge University Press

REFERENCES

[1] G. Sugihara and R. M. May, *Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series*, Nature, **344** (1990), 734–741.

[2] P. Masani and N. Wiener, Non-linear prediction, Probability and Statistics,, The Harald Cramér, (V. Grenander, ed.), John Wiley, 1959, pp. 190–212.

[3] Y. Okabe, On a stochastic difference equation for the multi-dimensional weakly stationary process with discrete time, Prospect of Algebraic Analysis, (M. Kashiwara and T. Kawai, eds.), Academic Press, Tokyo, 1988, pp. 601–645.

[4] Y. Okabe and Y. Nakano, *The theory of KM_2O-Langevin equations and its applications to data analysis* (I)*: Stationary analysis*, Hokkaido Math. J, **20** (1991), 45–90.

[5] Y. Okabe, *Application of the theory of KM_2O-Langevin equations to the linear prediction problem for the multi-dimensional weakly stationary process*, J. Math. Soc. of Japan, **45** (1993), 277–294.

[6] Y. Okabe, *Langevin equations and causal analysis*, SUGAKU Exposition, Amer. Math. Soc. Transl., **161** (1994), 19–50.

[7] Y. Okabe and A. Inoue, *The theory of KM_2O-Langevin equations and its applications to data analysis* (II)*: causal analysis* (1), Nagoya Math J., **134** (1994), 1–28.

[8] Y. Okabe and T. Ootsuka, *Applications of the theory of KM_2O-Langevin equations to the nonlinear prediction problem for the one dimensional strictly stationary time series*, J. Math. Soc. Japan, **47** (1995), 349–367.

[9] Y. Okabe, *Non-linear time series analysis based upon the fluctuation-dissipation theorem*, Nonlinear Analysis, Theory, Method & Applications, **30** (1997), 2249–.

[10] B. Efron, *Bootstrap methods: Another look at the jackknife*, The Annals of Statistics, **7** (1979), 1–28.

Yasunori Okabe
Department of Mathematical Engineering and Information Physics
Graduate School and Faculty of Engineering
University of Tokyo
Bunkyo-ku, Tokyo 113
Japan

Toshiyuki Yamane
Department of Mathematical Engineering and Information Physics
Graduate School and Faculty of Engineering
University of Tokyo
Bunkyo-ku, Tokyo 113
Japan