Hokkaido Mathematical Journal Vol. 29 (2000) p. 601–657

On a non-linear prediction analysis for multi-dimensional stochastic processes with its applications to data analysis

Yasunori OKABE* and Akihito KANEKO

(Received November 4, 1999)

Abstract. Recently, Matsuura-Okabe solved the prediction problem for one-dimensional stochastic processes which had remained to be solved for forty years after Masani-Wiener's work. In this paper, we shall develop a non-linear prediction analysis for multi-dimensional local stochastic processes based upon the theory of KM_2O-Langevin equations, which gives a refinement of Okabe-Yamane's work for causal problems and Matsuura-Okabe's work for prediction problems that have been investigated for one-dimensional stochastic processes. Moreover, we apply our results to concrete time series which concern the increase problem of earth temperature.

Key words: non-linear prediction analysis, non-linear causal analysis, KM_2O-Langevin equation, fluctuation-dissipation theorem, meteorological phenomena.

1. Introduction

The non-linear prediction problem for stochastic processes has a long history. Masani-Wiener ([1]) have given a prediction formula for calculating the non-linear predictor for the one-dimensional strictly stationary process satisfying the boundedness and the non-degenerateness. As stated in [1], their formula lacks a workable and computable algorithm. It has remained to weaken these conditions and give certain workable and computable algorithms for calculating the non-linear predictor.

Under the same conditions as in [1], Okabe-Ootsuka ([7]) have given a workable and computable algorithm for calculating the non-linear predictor by using the theory of KM_2O-Langevin equations for non-degenerate flows in inner product spaces. Based upon the fluctuation-dissipation principle behind this theory, one of the authors has proposed a method for detecting non-linear information behind a given time series data by using three kinds

---

1991 Mathematics Subject Classification : Primary 60G25; Secondary 60G12, 82C05.

*Partially supported by Grand-in-Aid for Science Research No. 10554001 and Grand-in-Aid for Exploratory Research No. 08874007, the Ministry of Education, Science, Sports and Culture, Japan.

of tests: the first one is Test(S) introduced in [5] which judges whether a given data has stationarity; the second one is Test(CS) introduced in [6] and [8] which judges whether there exists a causal direction between two given data; third one is Test(D) introduced in [8] which judges whether a given data has deterministic property.

However, it has still remained to weaken both the boundedness and the non-degenerate properties. Recently, Matsuura-Okabe ([12]) succeeded in replacing the boundedness by an exponential integrability condition by using a theory of polynomial functionals developed in Dobrushin-Minlos ([2]) and then in removing the non-degenerateness by developing a theory of $\mathrm{KM_2O}$-Langevin equations for degenerate flows in inner product spaces.

Let $\mathbf{X} = (X(n); 0 \leq n \leq N)$ be a one-dimensional stochastic process defined on a probability space $(\Omega, \mathcal{B}, P)$ such that all random variables $X(n)$ $(0 \leq n \leq N)$ are square integrable. By the local non-linear predictor for the stochastic process $\mathbf{X}$, we mean the conditional expectation of the future random variable $X(n+p)$ conditioned by the past sub σ-field $\mathcal{B}_0^n(\mathbf{X})$ $(0 \leq n \leq N-1, 1 \leq p \leq N-n)$:

$$E(X(n+p) \mid \mathcal{B}_0^n(\mathbf{X})), \tag{1.1}$$

where for each n $(0 \leq n \leq N)$, $\mathcal{B}_0^n(\mathbf{X})$ stands for the smallest σ-field with respect to which all random variables $X(k)$ $(0 \leq k \leq n)$ are measurable. By regarding each random variable $X(n)$ $(0 \leq n \leq N)$ as an element of the real Hilbert space $L^2(\Omega, \mathcal{B}, P)$, we find that the local non-linear predictor accords with the vector obtained by projecting the vector $X(n+p)$ on the closed subspace $\mathbf{N}_0^n(\mathbf{X}) \equiv L^2(\Omega, \mathcal{B}_0^n(\mathbf{X}), P)$:

$$P_{\mathbf{N}_0^n(\mathbf{X})} X(n+p), \tag{1.2}$$

where $P_{\mathbf{N}_0^n(\mathbf{X})}$ stands for the projection operator from $L^2(\Omega, \mathcal{B}, P)$ onto $\mathbf{N}_0^n(\mathbf{X})$.

The purpose of this paper is to extend the results of [12] to the case for multi-dimensional stochastic processes. A practical reason is as follows: It is an important problem to find a cause-and-effect relationship. There exist various kinds of data in both natural and social science that seem to influence each other. Therefore, when we intend to predict the future of a target data, it is important for us to find another causal data that influences the target data and then predict the future of the target data together with the causal data. The causal analysis developed by one of the authors is restricted to the one-dimensional data. Therefore, we have to develop a causal analysis for multi-dimensional data.

Let $\mathbf{X} = (X(n); 0 \leq n \leq N)$ and $\mathbf{Y} = (Y(n); 0 \leq n \leq N)$ be a one-dimensional stochastic process and a d-dimensional stochastic process defined on a probability space $(\Omega, \mathcal{B}, P)$, respectively, such that all random variables $X(n), Y(n)$ $(0 \leq n \leq N)$ are square integrable, where d is a positive integer. The local non-linear predictor of the stochastic process $\mathbf{X}$ together with the information of the stochastic process $\mathbf{Y}$ is given by

$$P_{\mathbf{N}_0^n(\mathbf{X},\mathbf{Y})} X(n+p), \tag{1.3}$$

where

$$\mathcal{B}_0^n(\mathbf{X},\mathbf{Y}) \equiv \sigma(X(k), Y_j(k); 0 \leq k \leq n, 1 \leq j \leq d) \tag{1.4}$$

$$\mathbf{N}_0^n(\mathbf{X},\mathbf{Y}) \equiv \{f \in L^2(\Omega, \mathcal{B}, P); f \text{ is } \mathcal{B}_0^n(\mathbf{X},\mathbf{Y})\text{-measurable}\}. \tag{1.5}$$

We define a $(d+1)$-dimensional stochastic process $\mathbf{Z} = (Z(n); 0 \leq n \leq N)$ by

$$Z(n) \equiv {}^t(X(n), Y_1(n), \ldots, Y_d(n)). \tag{1.6}$$

Since it follows that

$$\mathbf{N}_0^n(\mathbf{X},\mathbf{Y}) = \mathbf{N}_0^n(\mathbf{Z}), \tag{1.7}$$

we find that

$$P_{\mathbf{N}_0^n(\mathbf{X},\mathbf{Y})} X(n+p) = \text{the first component of } P_{\mathbf{N}_0^n(\mathbf{Z})} Z(n+p), \tag{1.8}$$

which implies that the local non-linear predictor of the one-dimensional stochastic process $\mathbf{X}$ together with the information of the d-dimensional stochastic process $\mathbf{Y}$ is reduced to the local non-linear predictor of the $(d+1)$-dimensional stochastic process $\mathbf{Z}$. This is one of the theoretical reasons why we generalize the results of [12] to the case of multi-dimensional stochastic processes.

Now we shall state the contents of this paper. In Section 2, we shall generalize the results of [12] to the case of multi-dimensional local stochastic processes satisfying the same conditions as in [12]. We first introduce a non-linear information space and construct its generator as a nest system of multi-dimensional local stochastic processes.

Section 3 treats a causal analysis for the multi-dimensional local stochastic processes. We have investigated in [6] and [8] a causal analysis based upon the non-linear information spaces for the global stochastic processes. We shall define a notion of local causality based upon the results of Section 2. In particular, we shall give a notion of local non-linear weak causality that can be introduced for the multi-dimensional stochastic processes. Moreover, we shall characterize these causality quantitatively in terms of the causal functions.

In Section 4, we shall deal with the multi-dimensional local stochastic processes satisfying the same conditions as in [12]. In association with a nest system of multi-dimensional stochastic processes constructed in Section 2, we shall introduce the minimal KM_2O-Langevin dissipation matrix functions and derive the KM_2O-Langevin equations describing their time evolutions. By introducing the prediction matrix functions that can be given in terms of the minimal KM_2O-Langevin dissipation matrix functions, we shall give a workable and computable algorithm for calculating the local non-linear predictor, which is a multi-dimensional version of the results of [12].

We give in Section 3 a quantitative characterization of non-linear causality between two local stochastic processes by using a nest system of multi-dimensional local stochastic processes. Generally speaking, it is impossible to check whether there exists a non-linear causality between two stochastic processes by taking a certain finite step procedure. However, if its causality were a non-linear causality of finite rank in some sense, then it might be possible to check its non-linear causality of finite rank. Thus, it is important and effective in time series analysis to find a method for checking whether there exists a non-linear causality of finite rank behind given two time series data, by taking a finite step procedure. For that reason, as a continuation of Section 2, we shall introduce in Section 5 a partial non-linear information space for the multi-dimensional local stochastic processes. Moreover, we shall deal with the stochastic processes with weakly stationary property and give certain algorithms for calculating the minimal KM_2O-dissipation matrix functions from the covariance matrix functions, which will be found to be equivalent to the so-called fluctuation-dissipation theorem.

In Section 6, we shall develop a causal analysis and prediction analysis based upon the results of Section 5 and give some prediction formulas for calculating the non-linear predictor by using causality.

We shall develop in Section 7 a non-linear time series analysis which bridges the gap between theory and phenomena. As we have determined a criterion of Test(CS) by using so many physical random number sequences which have not any causal relation with a target data, it becomes a negative assertion to say that there exist causal relations between given two time series data. In order to make the statement a positive assertion, we shall first give a refinement of Test(CS) by using some data obtained by shifting the target. Next, we shall give a method of model selection by using the maximal sample causal value associated with two given data. Finally, we shall give a certain prediction formula based upon causal analysis. In Section 8, we shall deal with three kinds of data related to meteorological phenomena: the first one is SST (sea surface temperature at 0 degree in latitude and 100 degree west in longitude); the second one is LIT (air temperature observed at Lima Callao Airport); the third one is LIP (air pressure observed at the same place). By applying the method in Section 7 to these data, we shall detect some information such that there exists a non-linear causal relation from SST to LIT and both the data LIT and LIP affect each other. We shall find that it is effective to predict the future of LIP by using not only the past information of LIP but also the past information of LIT. These confirm that it is effective to develop the prediction analysis together with the causal analysis for multi-dimensional stochastic processes.

2. Non-linear information space

We shall generalize the results of [12] to the case of multi-dimensional stochastic processes.

[2.1] (Local non-linear information space) Let d be a positive integer and $\mathbf{Z} = (Z(n); l \leq n \leq r)$ be an $\mathbf{R}^d$-valued stochastic process defined on a probability space $(\Omega, \mathcal{B}, P)$ satisfying the following two conditions (E) and (M):

(E) For any n $(l \leq n \leq r)$, there exists a positive constant $\lambda_0(n)$ such that for any real number λ $(|\lambda| \leq \lambda_0(n))$ and integer j $(1 \leq j \leq d)$,

$$E(\exp\{\lambda Z_j(n)\}) < \infty;$$

(M) $E(Z_j(n)) = 0 \quad (l \leq n \leq r, 1 \leq j \leq d).$

For any n_1, n_2 $(l \leq n_1 \leq n_2 \leq r)$, we define closed subspaces $\mathbf{M}_{n_1}^{n_2}(\mathbf{Z})$

and $\mathbf{N}_{n_1}^{n_2}(\mathbf{Z})$ of $L^2(\Omega, \mathcal{B}, P)$ by

$$\mathbf{M}_{n_1}^{n_2}(\mathbf{Z}) \equiv [\{Z_j(m); n_1 \le m \le n_2, 1 \le j \le d\}] \tag{2.1}$$

$$\mathbf{N}_{n_1}^{n_2}(\mathbf{Z}) \equiv \{Y \in L^2(\Omega, \mathcal{B}, P); Y \text{ is } \mathcal{B}_{n_1}^{n_2}(\mathbf{Z})\text{-measurable}\}, \tag{2.2}$$

where for any subset S of $L^2(\Omega, \mathcal{B}, P)$, we denote by $[S]$ the closed subspace of $L^2(\Omega, \mathcal{B}, P)$ which is generated by all elements in S and by $\mathcal{B}_{n_1}^{n_2}(\mathbf{Z})$ the smallest σ-field with respect to which all random variables $Z_j(k)$ $(n_1 \le k \le n_2, 1 \le j \le d)$ are measurable. We call $\mathbf{M}_{n_1}^{n_2}(\mathbf{Z})$ (resp. $\mathbf{N}_{n_1}^{n_2}(\mathbf{Z})$) a local linear (resp. non-linear) information space associated with the stochastic process $\mathbf{Z}$.

As noted in Dobrushin and Minlos [2], it follows from condition (E) that

Lemma 2.1

(i) *For any integers* n, j $(l \le n \le r, 1 \le j \le d)$,
$Z_j(n) \in \bigcap_{1 \le p < \infty} L^p(\Omega, \mathcal{B}, P)$.

(ii) *For any integers* n, p_k, j_k $(l \le n \le r,\ p_k \in \mathbf{N}^*,\ 1 \le j_k \le d,\ l \le k \le n)$,

$$Z_{j_l}(l)^{p_l} Z_{j_{l+1}}(l+1)^{p_{l+1}} \cdots Z_{j_n}(n)^{p_n} \in \mathbf{N}_l^n(\mathbf{Z}).$$

For each n $(l \le n \le r)$, we define a subset $\mathbf{F}_l^n(\mathbf{Z})$ of $\mathbf{N}_l^n(\mathbf{Z})$ by

$$\mathbf{F}_l^n(\mathbf{Z}) \equiv \left\{ \prod_{k=0}^{n-l} \prod_{j=1}^{d} Z_j(n-k)^{p_{k,j}} - E\left(\prod_{k=0}^{n-l} \prod_{j=1}^{d} Z_j(n-k)^{p_{k,j}} \right); \right.$$
$$\left. p_{k,j} \in \mathbf{N}^*, \sum_{j=1}^{d} p_{0,j} > 0 \right\}. \tag{2.3}$$

By virtue of condition (E), similarly as in Theorem 8.1 of [12], we can apply Proposition 2.1 in [2] to the family $\{Z_j(n); l \le n \le r, 1 \le j \le d\}$ of random variables to show that

Theorem 2.1 $\mathbf{N}_l^n(\mathbf{Z}) = [\{1\}] \oplus [\bigcup_{m=l}^{n} \mathbf{F}_l^m(\mathbf{Z})]$ $\quad (l \le n \le r)$.

[2.2] (Generator of non-linear information space) In this subsection, we shall construct a generator of non-linear information space by introducing a nest system of multi-dimensional stochastic processes.

[2.2.1] (Parameter) We parameterize the set $\bigcup_{m=l}^{r} \mathbf{F}_l^m(\mathbf{Z})$. We put

$\mathbf{M} \equiv (\mathbf{N}^*)^d$:

$$\mathbf{M} \equiv (\mathbf{N}^*)^d = \left\{p = {}^t(p_1, p_2, \ldots, p_d);\ p_j \in \mathbf{N}^*\ (1 \leq j \leq d)\right\}. \quad (2.4)$$

We define a subset Λ_{loc} of the product space $\mathbf{M}^{r-l+1}$ by

$$\Lambda_{loc} \equiv \{\mathbf{p} = (p^0, p^1, \ldots, p^{r-l}) \in \mathbf{M}^{r-l+1}; |p^0| \geq 1\}, \quad (2.5)$$

where for any element $p = {}^t(p_1, p_2, \ldots, p_d)$ of $\mathbf{M}$, we define the norm of p by

$$|p| \equiv \sum_{j=1}^{d} p_j. \quad (2.6)$$

Moreover, we use the following notation: for any vector $v = {}^t(v_1, v_2, \ldots, v_d)$ of $\mathbf{R}^\mathbf{d}$ and multi-index $\alpha = {}^t(\alpha_1, \alpha_2, \ldots, \alpha_d)$ of $(\mathbf{N}^*)^d$, we define a real number $v^\alpha \in \mathbf{R}$ by

$$v^\alpha \equiv \prod_{j=1}^{d} v_j^{\alpha_j}. \quad (2.7)$$

For any $\mathbf{p} = (p^0, p^1, \ldots, p^{r-l}) \in \Lambda_{loc}$, we define a one-dimensional stochastic process $\boldsymbol{\varphi}_\mathbf{p}(\mathbf{Z}) = (\varphi_\mathbf{p}(\mathbf{Z})(n); l + \sigma(\mathbf{p}) \leq n \leq r)$ by

$$\varphi_\mathbf{p}(\mathbf{Z})(n) \equiv \prod_{k=0}^{\sigma(\mathbf{p})} Z(n-k)^{p^k} \quad (l + \sigma(\mathbf{p}) \leq n \leq r), \quad (2.8)$$

where

$$\sigma(\mathbf{p}) \equiv \max\{k \in \{0, 1, \ldots, r-l\}; |p^k| > 0\} \quad (2.9)$$

and we denote by $G_{loc}(\mathbf{Z})$ the set of these stochastic processes:

$$G_{loc}(\mathbf{Z}) \equiv \{\boldsymbol{\varphi}_\mathbf{p}(\mathbf{Z}); \mathbf{p} \in \Lambda_{loc}\}. \quad (2.10)$$

It is to be noted that for each $l \leq n \leq r$,

$$\mathbf{F}_l^n(\mathbf{Z}) = \bigcup_{\mathbf{p} \in \Lambda_{loc}} \{\varphi_\mathbf{p}(\mathbf{Z})(n) - E(\varphi_\mathbf{p}(\mathbf{Z})(n)); l + \sigma(\mathbf{p}) \leq n\}. \quad (2.11)$$

To classify the non-linear information space $\mathbf{N}_l^r(\mathbf{Z})$, we define for any

$q \in \mathbf{N}$ a subset $\Lambda_{loc}(q)$ of Λ_{loc} and a subset $G_{loc}(\mathbf{Z})(q)$ of $G_{loc}(\mathbf{Z})$ by

$$\Lambda_{loc}(q) \equiv \left\{\mathbf{p} \in \Lambda_{loc}; \sum_{k=0}^{r-l}(k+1)|p^k| = q\right\} \tag{2.12}$$

$$G_{loc}(\mathbf{Z})(q) \equiv \{\varphi_{\mathbf{p}}(\mathbf{Z}); \mathbf{p} \in \Lambda_{loc}(q)\}. \tag{2.13}$$

It is to be noted that

$$G_{loc}(\mathbf{Z}) = \bigcup_{q \in N} G_{loc}(\mathbf{Z})(q) \qquad \text{(direct sum)}. \tag{2.14}$$

[2.2.2] (Lexicographical order) We first introduce a lexicographical order in the set Λ_{loc}. Let $\mathbf{p}$, $\mathbf{p}'$ be any fixed elements of Λ_{loc}. There exist $q, q' \in \mathbf{N}$ such that $\mathbf{p} \in \Lambda_{loc}(q)$, $\mathbf{p}' \in \Lambda_{loc}(q')$. We say that $\mathbf{p}$ precedes $\mathbf{p}'$ if $q < q'$ or if $q = q'$ and the j_0th component of p^{k_0} is larger than the j_0th component of p'^{k_0}, where k_0, j_0 are given by

$$k_0 \equiv \min\{0 \leq k \leq r-l; p^k \neq (p')^k\} \tag{2.15}$$

$$j_0 \equiv \min\{1 \leq j \leq d; \text{the } j\text{th component of } p^{k_0} \neq \text{the } j\text{th component of } (p')^{k_0}\}. \tag{2.16}$$

By using a one-to-one correspondence between $G_{loc}(\mathbf{Z})$ and Λ_{loc}, we can introduce an order into $G_{loc}(\mathbf{Z})$ according to the lexicographical order in Λ_{loc} and parameterize the set $G_{loc}(\mathbf{Z})$ as follows:

$$G_{loc}(\mathbf{Z}) = \{\varphi_j(\mathbf{Z}); j \in \mathbf{N}^*\}. \tag{2.17}$$

Since there exists for each $j \in \mathbf{N}^*$ a unique element $\mathbf{p}_j$ of the set Λ_{loc} such that $\varphi_j(\mathbf{Z}) = \varphi_{\mathbf{p}_j}(\mathbf{Z})$, we can define an integer $\sigma(j) \equiv \sigma(\mathbf{p}_j)$ and represent the stochastic process $\varphi_j(\mathbf{Z}) = (\varphi_j(\mathbf{Z})(n); l + \sigma(j) \leq n \leq r)$ as

$$\varphi_j(\mathbf{Z})(n) \equiv \varphi_{\mathbf{p}_j}(\mathbf{Z})(n) \qquad (l + \sigma(j) \leq n \leq r). \tag{2.18}$$

We define a natural number d_q by

$$d_q \equiv \left(\text{the number of elements in } \bigcup_{s=1}^{q} G_{loc}(\mathbf{Z})(s)\right) - 1 \tag{2.19}$$

and then we can see that

$$G_{loc}(\mathbf{Z})(q) = \{\varphi_{d_{q-1}+1}(\mathbf{Z}), \varphi_{d_{q-1}+2}(\mathbf{Z}), \ldots \varphi_{d_q}(\mathbf{Z})\}. \tag{2.20}$$

The concrete forms of elements of $G_{loc}(\mathbf{Z})(q)$ are stated in (A.1) of Appendix. In particular, the cases $d=1$, $q=6$ and $d=2$, $q=6$ are shown in Tables A1 and A2 of Appendix, respectively.

[2.2.3] (Nest system) Let us fix any $q \in \mathbf{N}$. For any integer j $(0 \le j \le d_q)$, we define a one-dimensional stochastic process $\boldsymbol{\phi}_j(\mathbf{Z}) = (\phi_j(\mathbf{Z})(n); l + \sigma(j) \le n \le r)$ by

$$\phi_j(\mathbf{Z})(n) \equiv \varphi_j(\mathbf{Z})(n) - E(\varphi_j(\mathbf{Z})(n)). \tag{2.21}$$

It is to be noted that the time parameter space of stochastic process $\boldsymbol{\phi}_j(\mathbf{Z})$ depends upon j. By enlarging this time parameter space to the common space $\{l, l+1, \dots, r\}$, we define a one-dimensional stochastic process $\tilde{\boldsymbol{\phi}}_j(\tilde{\mathbf{Z}}) = (\tilde{\phi}_j(\tilde{\mathbf{Z}})(n); l \le n \le r)$ by

$$\tilde{\phi}_j(\tilde{\mathbf{Z}})(n) \equiv \begin{cases} 0 & (l \le n < l + \sigma(j)) \\ \tilde{\varphi}_j(\mathbf{Z})(n) - E(\tilde{\varphi}_j(\mathbf{Z})(n)) & (l + \sigma(j) \le n \le r). \end{cases} \tag{2.22}$$

We call the set of these stochastic processes $\tilde{\boldsymbol{\phi}}_j(\tilde{\mathbf{Z}})$ the class of stochastic processes obtained by non-linear transformations of rank q and denote it by $\mathcal{T}^{(q)}(\mathbf{Z})$:

$$\mathcal{T}^{(q)}(\mathbf{Z}) \equiv \{\tilde{\boldsymbol{\phi}}_j(\mathbf{Z}); 0 \le j \le d_q\}. \tag{2.23}$$

Now, we define a (d_q+1)-dimensional stochastic process $\tilde{\mathbf{Z}}^{(q)} = (\tilde{Z}^{(q)}(n); l \le n \le r)$ and a $(d_{q+1} - d_q)$-dimensional stochastic process $\tilde{\mathbf{W}}^{(q+1)} = (\tilde{W}^{(q)}(n); l \le n \le r)$ by

$$\tilde{Z}^{(q)}(n) \equiv {}^t(\tilde{\boldsymbol{\phi}}_0(\mathbf{Z})(n), \tilde{\boldsymbol{\phi}}_1(\mathbf{Z})(n), \dots, \tilde{\boldsymbol{\phi}}_{d_q}(\mathbf{Z})(n)) \tag{2.24}$$

$$\tilde{W}^{(q+1)}(n) \equiv {}^t(\tilde{\boldsymbol{\phi}}_{d_q+1}(\mathbf{Z})(n), \tilde{\boldsymbol{\phi}}_{d_q+2}(\mathbf{Z})(n), \dots, \tilde{\boldsymbol{\phi}}_{d_{q+1}}(\mathbf{Z})(n)). \tag{2.25}$$

Concerning the relation among these stochastic processes $\tilde{\mathbf{Z}}^{(q)}$, $\tilde{\mathbf{W}}^{(q)}$ and the original stochastic process $\mathbf{Z}$, we can see from Theorem 2.1 that

Theorem 2.2

(i) $\tilde{\mathbf{Z}}^{(1)} = \mathbf{Z}$.

(ii) *The system* $\{\tilde{\mathbf{Z}}^{(q)}; q \in \mathbf{N}\}$ *has a nest structure, that is,*

$$\tilde{Z}^{(q+1)}(n) = \begin{pmatrix} \tilde{Z}^{(q)}(n) \\ \tilde{W}^{(q+1)}(n) \end{pmatrix} \quad (q \in \mathbf{N}, l \le n \le r).$$

(iii) $$\left[\bigcup_{l\le m\le n} \mathbf{F}_l^m(\mathbf{Z})\right] = \left[\bigcup_{q=1}^{\infty} \mathbf{M}_l^n(\tilde{\mathbf{Z}}^{(q)})\right] \quad (l \le n \le r).$$

(iv) $$\mathbf{N}_l^n(\mathbf{Z}) = [\{1\}] \oplus \left[\bigcup_{q=1}^{\infty} \mathbf{M}_l^n(\tilde{\mathbf{Z}}^{(q)})\right] \quad (l \le n \le r).$$

3. Non-linear causality

The purpose of this section is to give a refinement and a generalization of the notion of causality and determinism for one-dimensional global stochastic processes investigated in [6] and [8] to the case of multi-dimensional local stochastic processes.

[3.1] (Causality) In this subsection, we investigate a non-linear causal problem which gives a refinement of the results of [8].

Let $\mathbf{X} = (X(n); l \le n \le r)$ be a one-dimensional stochastic process and $\mathbf{Y} = (Y(n); l \le n \le r)$ be a d-dimensional stochastic process defined on a probability space $(\Omega, \mathcal{B}, P)$ such that all random variables $X(m)$ and $Y_j(n)$ $(l \le m, n \le r, 1 \le j \le d)$ are square integrable, where d is a natural integer, l and r are integers and $Y_j(n)$ stands for the jth component of $Y(n)$.

[3.1.1] (Linear causality) We say that there exists a linear causality from $\mathbf{Y}$ to $\mathbf{X}$ if there exist an integer M_0 $(l \le M_0 \le r)$ and a linear function $L_n : \mathbf{R}^{(n-l+1)d} \to \mathbf{R}$ such that

$$X(n) = L_n(Y(n), Y(n-1), \dots, Y(l)) \quad (M_0 \le n \le r). \tag{3.1}$$

We denote this relation by

$$\mathbf{Y} \overset{(LC)}{\longrightarrow} \mathbf{X}. \tag{3.2}$$

The qualitative definition of linear causality can be stated quantitatively as follows:

Theorem 3.1 *The following three properties are equivalent:*

(i) $\mathbf{Y} \overset{(LC)}{\longrightarrow} \mathbf{X}$;

(ii) *There exists an integer M_0 $(l \le M_0 \le r)$ such that*

$$\mathbf{M}_l^n(\mathbf{X}) \subset \mathbf{M}_l^n(\mathbf{Y}) \quad (M_0 \le n \le r);$$

(iii) *There exists an integer* M_0 $(l \leq M_0 \leq r)$ *such that*

$$||P_{\mathbf{M}_l^n(\mathbf{Y})}X(n)|| = ||X(n)|| \quad (M_0 \leq n \leq r),$$

where (u, v) *and* $||w||$ *stand for the inner product of the vectors* u, v *and the norm of the vector* w *in* $L^2(\Omega, \mathcal{B}, P)$, *respectively.*

[3.1.2] (Non-linear causality) We say that there exists a non-linear causality from $\mathbf{Y}$ to $\mathbf{X}$ if there exist an integer M_0 $(l \leq M_0 \leq r)$ and a Borel-measurable function $F_n : \mathbf{R}^{(n-l+1)d} \to \mathbf{R}$ such that

$$X(n) = F_n(Y(n), Y(n-1), \ldots, Y(l)) \quad (M_0 \leq n \leq r). \tag{3.3}$$

We denote this relation by

$$\mathbf{Y} \overset{(C)}{\longrightarrow} \mathbf{X}. \tag{3.4}$$

From now on, we assume that the stochastic process $\mathbf{Y}$ satisfies conditions (E) and (M). By applying Theorem 2.2 to $\mathbf{Y}$, we can characterize a qualitative definition of non-linear causality quantitatively in terms of linear causality as follows:

Theorem 3.2 *The following three properties are equivalent*:

(i) $\mathbf{Y} \overset{(C)}{\longrightarrow} \mathbf{X}$;

(ii) *There exists an integer* M_0 $(l \leq M_0 \leq r)$ *such that*

$$\mathbf{N}_l^n(\mathbf{X}) \subset \mathbf{N}_l^n(\mathbf{Y}) \quad (M_0 \leq n \leq r);$$

(iii) *There exists* M_0 $(l \leq M_0 \leq r)$ *such that*

$$\lim_{q \to \infty} ||P_{\mathbf{M}_l^n(\tilde{\mathbf{Y}}^{(q)})}X(n)|| = ||X(n)|| \quad (M_0 \leq n \leq r).$$

[3.1.3] (Non-linear weak causality) We note that the definition of non-linear causality from $\mathbf{Y}$ to $\mathbf{X}$ implies that the information at the time n of the stochastic process $\mathbf{X}$ can be determined by the one of the past until the time n of the stochastic process $\mathbf{Y}$. As stated in Section 1, it is effective to use the information of the past of $\mathbf{X}$ in the description of causality relation between $\mathbf{X}$ and $\mathbf{Y}$. We shall give a notion of non-linear weak causality.

We say that there exists a non-linear weak causality from $\mathbf{Y}$ to $\mathbf{X}$ if there exist an integer M_0 $(l \leq M_0 \leq r-1)$ and a Borel-measurable function

$G_n : \mathbf{R}^{n-l+1+(n-l+2)d} \to \mathbf{R}$ such that for any n ($M_0 \le n \le r-1$),

$$\begin{aligned} &X(n+1) \\ &\quad = G_n(X(n), X(n-1), \ldots, X(l), Y(n+1), Y(n), \ldots, Y(l)) \end{aligned} \tag{3.5}$$

and we denote this relation by

$$\mathbf{Y} \overset{(WC)}{\longrightarrow} \mathbf{X}. \tag{3.6}$$

Immediately from definition, we have

Theorem 3.3

$$\mathbf{Y} \overset{(C)}{\longrightarrow} \mathbf{X} \implies \mathbf{Y} \overset{(WC)}{\longrightarrow} \mathbf{X}.$$

In order to characterize the non-linear weak causality quantitatively, we define two stochastic processes $\mathbf{X}_{+1} = (X_{+1}(n); l \le n \le r-1)$ and $\mathbf{Y}_{+1} = (Y_{+1}(n); l \le n \le r-1)$ by

$$X_{+1}(n) \equiv X(n+1) \quad (l \le n \le r-1) \tag{3.7}$$

$$Y_{+1}(n) \equiv Y(n+1) \quad (l \le n \le r-1) \tag{3.8}$$

and a $(d+1)$-dimensional stochastic process $\mathbf{Z} = (Z(n); l \le n \le r-1)$ by

$$Z(n) \equiv {}^t(X(n), {}^tY_{+1}(n)) = {}^t(X(n), {}^tY(n+1)). \tag{3.9}$$

Moreover, we assume that $\mathbf{X}$ together with $\mathbf{Y}$ satisfies conditions (E) and (M). By applying Theorem 3.2 to $\mathbf{Z}$, we have

Theorem 3.4 *The following properties are equivalent*:

(i) $\mathbf{Y} \overset{(WC)}{\longrightarrow} \mathbf{X}$;

(ii) $\mathbf{Z} \overset{(C)}{\longrightarrow} \mathbf{X}_{+1}$;

(iii) *There exists an integer* M_0 ($l \le M_0 \le r-1$) *such that*

$$\lim_{q\to\infty} ||P_{\mathbf{M}_l^n(\mathbf{Z}^{(q)})} X(n+1)|| = ||X(n+1)|| \quad (M_0 \le n \le r-1).$$

[3.1.4] (Non-instantaneous and non-linear weak causality) Even if we could find a stochastic process $\mathbf{Y}$ such that there exists a non-linear causality or a non-linear weak causality from $\mathbf{Y}$ to $\mathbf{X}$, it is only the information of the past of $\mathbf{X}$ and $\mathbf{Y}$ for us to be able to use in predicting the future of $\mathbf{X}$.

For this reason, we shall give a notion of non-instantaneous and non-linear weak causality which is weaker than that of non-linear weak causality. We say that there exists a non-instantaneous and non-linear weak causality from $\mathbf{Y}$ to $\mathbf{X}$ if there exist an integer M_0 $(l \leq M_0 \leq r-1)$ and a Borel-measurable function $H_n : \mathbf{R}^{(n-l+1)(d+1)} \to \mathbf{R}$ such that for any n $(M_0 \leq n \leq r-1)$,

$$\begin{aligned} &X(n+1) \\ &\quad = H_n(X(n), X(n-1), \ldots, X(l), Y(n), Y(n-1), \ldots, Y(l)) \end{aligned} \tag{3.10}$$

and we denote this relation by

$$\mathbf{Y} \overset{(WC^-)}{\longrightarrow} \mathbf{X}. \tag{3.11}$$

In order to characterize the non-instantaneous and non-linear weak causality quantitatively, we define a $(d+1)$-dimensional stochastic process $\mathbf{W} = (W(n); l \leq n \leq r)$ by

$$W(n) \equiv {}^t(X(n), {}^tY(n)). \tag{3.12}$$

Moreover, we assume that $\mathbf{X}$ together with $\mathbf{Y}$ satisfies conditions (E) and (M). By applying Theorem 3.2 to $\mathbf{W}$, we have

Theorem 3.5 *The following properties are equivalent*:

(i) $\mathbf{Y} \overset{(WC^-)}{\longrightarrow} \mathbf{X}$;

(ii) $\mathbf{W} \overset{(C)}{\longrightarrow} \mathbf{X}_{+1}$;

(iii) *There exists an integer* M_0 $(l \leq M_0 \leq r-1)$ *such that*

$$\lim_{q\to\infty} ||P_{\mathbf{M}_l^n(\mathbf{W}^{(q)})} X(n+1)|| = ||X(n+1)|| \quad (M_0 \leq n \leq r-1).$$

[3.2] (Determinism) In this subsection, we investigate a non-linear deterministic problem which gives a refinement of the results of [8].

Let $\mathbf{X} = (X(n); l \leq n \leq r)$ be a one-dimensional stochastic process defined on a probability space $(\Omega, \mathcal{B}, P)$ such that all random variables $X(n)$ $(l \leq n \leq r)$ are square integrable. As in (3.7), we define a stochastic process $\mathbf{X}_{+1} = (X_{+1}(n); l \leq n \leq r-1)$ by

$$X_{+1}(n) \equiv X(n+1) \quad (l \leq n \leq r-1). \tag{3.13}$$

Moreover, we restrict the time domain of the stochastic process $\mathbf{X}$ to the subset $\{l, l+1, \ldots, r-1\}$ and denote it by $\mathbf{X}^{(l,r-1)}$. We say that $\mathbf{X}$ has determinism if and only if

$$\mathbf{X}^{(l,r-1)} \xrightarrow{(C)} \mathbf{X}_{+1}, \tag{3.14}$$

that is, there exist an integer M_0 ($l \le M_0 \le r-1$) and a Borel measurable function $I_n : \mathbf{R}^{n-l+1} \to \mathbf{R}$ such that

$$X(n+1) = I_n(X(n), X(n-1), \ldots, X(l)) \quad (M_0 \le n \le r-1). \tag{3.15}$$

4. Non-linear prediction formula

In [12], we have developed a theory of KM$_2$O-Langevin equations for degenerate flows in an inner product space and applied it to the non-linear prediction problem for one-dimensional stochastic processes. In its course, we have constructed a nest system consisting of multi-dimensional stochastic processes and obtained the prediction formulas for calculating the linear predictor for them.

In this section, we shall return to the same situation as in Section 2 and give a prediction formula for the nest system $\{\tilde{\mathbf{Z}}^{(q)}; q \in \mathbf{N}\}$ constructed in (2.24).

[4.1] (KM$_2$O-Langevin equation for $\tilde{\mathbf{Z}}^{(q)}$) We shall apply the theory of KM$_2$O-Langevin equations to the degenerate stochastic processes $\tilde{\mathbf{Z}}^{(q)}$. For any $q \in \mathbf{N}$, we define a (d_q+1)-dimensional stochastic process $\boldsymbol{\nu}_+(\tilde{\mathbf{Z}}^{(q)}) = (\nu_+(\tilde{\mathbf{Z}}^{(q)})(n); 0 \le n \le r-l)$, to be called a KM$_2$O-Langevin fluctuation flow associated with $\tilde{\mathbf{Z}}^{(q)}$, by

$$\nu_+(\tilde{\mathbf{Z}}^{(q)})(n) \equiv \begin{cases} \tilde{Z}^{(q)}(l) & (n=0) \\ \tilde{Z}^{(q)}(l+n) - P_{\mathbf{M}_l^{l+n-1}(\tilde{\mathbf{Z}}^{(q)})}\tilde{Z}^{(q)}(l+n) & (0 < n \le r-l) \end{cases} \tag{4.1}$$

and define a matrix function $V_+(\tilde{\mathbf{Z}}^{(q)})(n)$ $(0 \le n \le r-l)$, to be called a KM$_2$O-Langevin fluctuation matrix function associated with $\tilde{\mathbf{Z}}^{(q)}$, by

$$V_+(\tilde{\mathbf{Z}}^{(q)})(n) \equiv E\Big(\nu_+(\tilde{\mathbf{Z}}^{(q)})(n)\, {}^t\nu_+(\tilde{\mathbf{Z}}^{(q)})(n)\Big) \quad (0 \le n \le r-l). \tag{4.2}$$

It is to be noted that the set $\{\tilde{Z}_j^{(q)}(n); l \le n \le r, 1 \le j \le d_q + 1\}$ is degenerate, that is, not linearly independent in $L^2(\Omega, \mathcal{B}, P)$. However, we can apply Theorem 4.4 in [12] to the degenerate flow $\tilde{\mathbf{Z}}^{(q)}$ to find that there exists a unique minimal $\mathrm{KM_2O}$-Langevin dissipation matrix function $\gamma_+^0(\tilde{\mathbf{Z}}^{(q)}) = (\gamma_+^0(\tilde{\mathbf{Z}}^{(q)})(n,k); 0 \le k < n \le r-l)$ such that

$$\tilde{Z}^{(q)}(l+n) = -\sum_{k=0}^{n-1} \gamma_+^0(\tilde{\mathbf{Z}}^{(q)})(n,k)\tilde{Z}^{(q)}(l+k) + \nu_+(\tilde{\mathbf{Z}}^{(q)})(n) \qquad (0 \le n \le r-l). \tag{4.3}$$

We call equation (4.3) a $\mathrm{KM_2O}$-Langevin equation associated with $\tilde{\mathbf{Z}}^{(q)}$.

[4.2] (Prediction formula for the linear predictor of $\tilde{\mathbf{Z}}^{(q)}$) According to the algorithm in (6.1) of [12], we define a prediction matrix function $Q_+(\tilde{\mathbf{Z}}^{(q)}) = (Q_+(\tilde{\mathbf{Z}}^{(q)})(n,m;k); 0 \le k \le m < n \le r-l)$ by

$$Q_+(\tilde{\mathbf{Z}}^{(q)})(n,m;k) \equiv \begin{cases} -\gamma_+^0(\tilde{\mathbf{Z}}^{(q)})(m+1,k) & (n = m+1) \\ -\displaystyle\sum_{j=m+1}^{n-1} \gamma_+^0(\tilde{\mathbf{Z}}^{(q)})(n,j)Q_+(\tilde{\mathbf{Z}}^{(q)})(j,m;k) \\ \qquad - \gamma_+^0(\tilde{\mathbf{Z}}^{(q)})(n,k) & (n > m+1). \end{cases} \tag{4.4}$$

By applying Theorem 6.1 of [12] to the degenerate flow $\tilde{\mathbf{Z}}^{(q)}$, we have

Lemma 4.1 *For any integers n, p $(l \le n \le r-1, 1 \le p \le r-n)$,*

$$P_{\mathbf{M}_l^n(\tilde{\mathbf{Z}}^{(q)})}\tilde{Z}^{(q)}(n+p) = \sum_{k=0}^{n-l} Q_+(\tilde{\mathbf{Z}}^{(q)})(n-l+p, n-l; k)\tilde{Z}^{(q)}(l+k).$$

[4.3] (Prediction formula for the non-linear predictor of $\mathbf{Z}$) Now we are in a position to state the main theorem of this section. By virtue of Theorem 2.2, as in Theorem 9.1 of [12], we can let q tend to ∞ in Lemma 4.1 to obtain

Theorem 4.1 (Multi-dimensional local non-linear predictor) *For any integers n, p $(l \le n \le r-1, 1 \le p \le r-n)$,*

$$P_{\mathbf{N}_l^n(\mathbf{Z})} Z(n+p) = \text{the first } d \text{ components of } \lim_{q\to\infty} \sum_{k=0}^{n-l} Q_+(\tilde{\mathbf{Z}}^{(q)})(n-l+p, n-l; k) \tilde{Z}^{(q)}(l+k).$$

This theorem implies that the multi-dimensional local non-linear predictor of $\mathbf{Z}$ can be expressed as a limit of local linear predictors of $\tilde{\mathbf{Z}}^{(q)} (q \in \mathbf{N})$.

5. Partial non-linear information space and stationarity

As a continuation of the previous section, we shall deal with a d-dimensional stochastic process $\mathbf{Z} = (Z(n); l \le n \le r)$ satisfying conditions (E) and (M).

[5.1] (Partial non-linear information space) We have introduced in (2.23) a class $\mathcal{T}^{(q)}(\mathbf{Z})$ of non-linear transformations of rank q which consists of one-dimensional stochastic processes. From a theoretical point of view, there exists an unsatisfactory point that even if the stochastic process $\mathbf{Z}$ has strict stationarity, any stochastic process $\tilde{\mathbf{Z}}^{(q)}$ $(q \in \mathbf{N})$ does not have weak stationarity, though the stochastic process $\phi_j(\mathbf{Z})$ $(0 \le j \le d_q)$ has strict stationarity. On the other hand, for the practical aim of applying our results to non-linear time series analysis, we have to pay attention to the usage of non-linear transformations of $\mathbf{Z}$.

For that reason, for any fixed natural numbers q, D such that $D \le d_q$, we define a space $\mathbf{J}^{(q,D)}$ of multi-indices by

$$\mathbf{J}^{(q,D)} \equiv \{J = (j_1, j_2, \ldots, j_D); \; 0 \le j_1 < j_2 < \cdots < j_D \le d_q, j_k \in \mathbf{N}^*\}. \tag{5.1}$$

For each element $J = (j_1, j_2, \ldots, j_D)$ of $\mathbf{J}^{(q,D)}$, we define a D-dimensional stochastic process $\mathbf{Z}_J = (Z_J(n); l + \sigma(J) \le n \le r)$ by

$$Z_J(n) \equiv {}^t(\phi_{j_1}(\mathbf{Z})(n), \phi_{j_2}(\mathbf{Z})(n), \ldots, \phi_{j_D}(\mathbf{Z})(n)), \tag{5.2}$$

where $\sigma(J)$ is given by

$$\sigma(J) \equiv \max\{\sigma(j_k); 1 \le k \le D\}. \tag{5.3}$$

We denote by $\mathcal{T}^{(q,D)}(\mathbf{Z})$ the system of these stochastic processes:

$$\mathcal{T}^{(q,D)}(\mathbf{Z}) \equiv \{\mathbf{Z}_J; J \in \mathbf{J}^{(q,D)}\} \tag{5.4}$$

and call this system the partial non-linear information space of rank (q, D) associated with the stochastic process $\mathbf{Z}$.

[5.2] (Stationarity and Fluctuation-Dissipation Theorem) As noted in [5.1], we know that if the stochastic process $\mathbf{Z}$ has strict stationarity, then all the elements of $\mathcal{T}^{(q,D)}(\mathbf{Z})$ have strict stationarity and so weak stationarity. On the other hand, it is important to check weak stationarity of time series in a non-linear time series analysis, which will be done in Sections 7 and 8 based upon the so-called fluctuation-dissipation theorem.

For this reason, we shall deal with the elements of $\mathcal{T}^{(q,D)}(\mathbf{Z})$ satisfying weak stationarity; for any fixed $J \in \mathbf{J}^{(q,D)}$, we say that a stochastic process $\mathbf{Z}_J$ has weak stationarity if there exists a covariance matrix function $R(\mathbf{Z}_J)$: $\{-(r-l-\sigma(J)), -(r-l-\sigma(J))+1, \ldots, r-l-\sigma(J)\} \to M(D; \mathbf{R})$ such that

$$E(Z_J(m)^t Z_J(n)) = R(\mathbf{Z}_J)(m-n) \quad (0 \leq m, n \leq r-l-\sigma(J)). \tag{5.5}$$

We denote the subset of such stochastic processes by $\mathcal{ST}^{(q,D)}(\mathbf{Z})$:

$$\mathcal{ST}^{(q,D)}(\mathbf{Z}) \equiv \left\{ \mathbf{Z}_J \in \mathcal{T}^{(q,D)}(\mathbf{Z}); \mathbf{Z}_J \text{ has weak stationarity} \right\}. \tag{5.6}$$

We shall introduce two kinds of KM$_2$O-Langevin fluctuation flows associated with the stochastic process $\mathbf{Z}_J$. As in (4.1), for each element J of $\mathbf{J}^{(q,D)}$, we define a forward KM$_2$O-Langevin fluctuation flow $\boldsymbol{\nu}_+(\mathbf{Z}_J) = (\nu_+(\mathbf{Z}_J)(n); 0 \leq n \leq r-l-\sigma(J))$ by

$$\begin{aligned} &\nu_+(\mathbf{Z}_J)(n) \\ &\equiv \begin{cases} Z_J(l+\sigma(J)) & (n=0) \\ Z_J(l+\sigma(J)+n) & \\ \quad - P_{\mathbf{M}_{l+\sigma(J)}^{l+\sigma(J)+n-1}(\mathbf{Z}_J)} Z_J(l+\sigma(J)+n) & (0 < n \leq r-l-\sigma(J)) \end{cases} \end{aligned} \tag{5.7}$$

and define a forward KM$_2$O-Langevin fluctuation matrix function $V_+(\mathbf{Z}_J) = (V_+(\mathbf{Z}_J)(n); 0 \leq n \leq r-l-\sigma(J))$ associated with $\mathbf{Z}_J$ by

$$V_+(\mathbf{Z}_J)(n) \equiv E(\nu_+(\mathbf{Z}_J)(n)\, {}^t\nu_+(\mathbf{Z}_J)(n)) \quad (0 \leq n \leq r-l-\sigma(J)). \tag{5.8}$$

Moreover, by applying Theorem 4.4 in [12] to the flow $\mathbf{Z}_J$, we find that there exists a unique minimal forward KM$_2$O-Langevin dissipation matrix function $\gamma_+^0(\mathbf{Z}_J) = (\gamma_+^0(\mathbf{Z}_J)(n,k); 0 \le k < n \le r-l-\sigma(J))$ such that for any n $(0 \le n \le r-l-\sigma(J))$,

$$\begin{aligned} & Z_J(l+\sigma(J)+n) \\ & \quad = -\sum_{k=0}^{n-1} \gamma_+^0(\mathbf{Z}_J)(n,k) Z_J(l+\sigma(J)+k) + \nu_+(\mathbf{Z}_J)(n). \quad (5.9) \end{aligned}$$

We call equation (5.9) a forward KM$_2$O-Langevin equation associated with $\mathbf{Z}_J$.

On the other hand, we define a backward KM$_2$O-Langevin fluctuation flow $\boldsymbol{\nu}_-(\mathbf{Z}_J) = (\nu_-(\mathbf{Z}_J)(n); -r+l+\sigma(J) \le n \le 0)$ by

$$\begin{aligned} & \nu_-(\mathbf{Z}_J)(n) \\ & \equiv \begin{cases} Z_J(r) & (n=0) \\ Z_J(r+n) - P_{\mathbf{M}_{r+n+1}^r(\mathbf{Z}_J)} Z_J(r+n) & (-r+l+\sigma(J) \le n < 0). \end{cases} \end{aligned} \quad (5.10)$$

As in (5.8), we define a backward KM$_2$O-Langevin fluctuation matrix function $V_-(\mathbf{Z}_J) = (V_-(\mathbf{Z}_J)(n); 0 \le n \le r-l-\sigma(J))$ associated with $\mathbf{Z}_J$ by

$$\begin{aligned} V_-(\mathbf{Z}_J)(n) \equiv\ & E(\nu_-(\mathbf{Z}_J)(-n)\, {}^t\nu_-(\mathbf{Z}_J)(-n)) \\ & (0 \le n \le r-l-\sigma(J)). \quad (5.11) \end{aligned}$$

Moreover, by applying Theorem 4.4 in [12] to the stochastic process $\mathbf{Z}_J$, we find that there exists a unique minimal backward KM$_2$O-Langevin dissipation matrix function $\gamma_-^0(\mathbf{Z}_J) = (\gamma_-^0(\mathbf{Z}_J)(n,k); 0 \le k < n \le r-l-\sigma(J))$ such that

$$\begin{aligned} Z_J(r-n) =\ & -\sum_{k=0}^{n-1} \gamma_-^0(\mathbf{Z}_J)(n,k) Z_J(r-k) + \nu_-(\mathbf{Z}_J)(-n) \\ & (0 \le n \le r-l-\sigma(J)). \quad (5.12) \end{aligned}$$

We call equation (5.12) a backward KM$_2$O-Langevin equation associated with $\mathbf{Z}_J$.

In particular, we define a minimal forward (resp. backward) KM$_2$O-Langevin partial correlation matrix function $\delta_+(\mathbf{Z}_J) = (\delta_+(\mathbf{Z}_J)(n); 1 \le n \le$

$r-l-\sigma(J))$ (resp. $\delta_-(\mathbf{Z}_J)=(\delta_-(\mathbf{Z}_J)(n);1\le n\le r-l-\sigma(J)))$ by

$$\delta_\pm(\mathbf{Z}_J)(n)\equiv\gamma_\pm(\mathbf{Z}_J)(n,0). \tag{5.13}$$

We define a system $\mathcal{LM}(\mathbf{Z}_J)$, to be called a $\mathrm{KM_2O}$-Langevin matrix associated with $\mathbf{Z}_J$, by

$$\begin{aligned}\mathcal{LM}(\mathbf{Z}_J) \equiv \{&\gamma^0_\pm(\mathbf{Z}_J)(n,k),\delta^0_\pm(\mathbf{Z}_J)(n),V_\pm(\mathbf{Z}_J)(m);\\ &0\le k<n\le r-l-\sigma(J),\,0\le m\le r-l-\sigma(J)\}.\end{aligned} \tag{5.14}$$

From Theorem 5.3 in [12], we have

Theorem 5.1 $\mathbf{Z}_J$ *has weak stationarity if and only if the following properties hold*:

(i) $\gamma^0_\pm(\mathbf{Z}_J)(n,k)=\gamma^0_\pm(\mathbf{Z}_J)(n-1,k-1)+\delta^0_\pm(\mathbf{Z}_J)(n)\gamma^0_\mp(n-1,n-1-k)$ $(1\le k<n\le r-l-\sigma(J))$;

(ii) $V_\pm(\mathbf{Z}_J)(n)=(I-\delta^0_\pm(\mathbf{Z}_J)(n)\delta^0_\mp(\mathbf{Z}_J)(n))V_\pm(\mathbf{Z}_J)(n-1)$ $(1\le n\le r-l-\sigma(J))$;

(iii) $V_+(\mathbf{Z}_J)(n)^t\delta^0_-(\mathbf{Z}_J)(n+1)=\delta^0_+(\mathbf{Z}_J)(n+1)V_-(\mathbf{Z}_J)(n)$ $(0\le n\le r-l-\sigma(J)-1)$.

We call the algorithms (i), (ii) and (iii) in Theorem 5.1 the fluctuation-dissipation theorem ((FDT)) ([9], [12]).

Moreover, there exists an algorithm, to be called (PAC), by which the $\mathrm{KM_2O}$-Langevin partial correlation matrix functions can be calculated from the covariance matrix function $R(\mathbf{Z}_J)$. To explain it, we shall introduce an additional noise flow which was used in [12]. We define for each positive number w a stochastic process $\mathbf{Z}^w_J=(Z^w_J(n);l+\sigma(J)\le n\le r)$ by

$$Z^w_J(n)\equiv Z_J(n)+w\xi(n), \tag{5.15}$$

where $\boldsymbol{\xi}=(\xi(n);l+\sigma(J)\le n\le r)$ is an additional noise flow for the stochastic process $\mathbf{Z}_J$, that is, a D-dimensional stochastic process satisfying the following properties:

$$\boldsymbol{\xi}\text{ is a white noise, that is, }(\xi(m),{}^t\xi(n))=\delta_{mn}I_D \quad (l+\sigma(J)\le m,\,n\le r) \tag{5.16}$$

$$(\xi(m),{}^tZ_J(n))=0\quad(l+\sigma(J)\le m,\,n\le r). \tag{5.17}$$

We note that the covariance matrix function $R(\mathbf{Z}_J^w) = (R(\mathbf{Z}_J^w)(n); |n| \le r - l - \sigma(J))$ is given by

$$R(\mathbf{Z}_J^w)(n) = R(\mathbf{Z}_J)(n) + w^2\delta_{n0}I_D \quad (|n| \le r - l - \sigma(J)). \tag{5.18}$$

From Theorem 6.1 in [9], we have

Theorem 5.2

(i) $\delta_+(\mathbf{Z}_J^w)(n)$

$$= -\Big\{R(\mathbf{Z}_J^w)(n) + \sum_{k=0}^{n-2}\gamma_+(\mathbf{Z}_J^w)(n-1,k)R(\mathbf{Z}_J^w)(k+1)\Big\}V_-(\mathbf{Z}_J^w)(n-1)^{-1}.$$

(ii) $\delta_-(\mathbf{Z}_J^w)(n)$

$$= -\Big\{R(\mathbf{Z}_J^w)(-n) + \sum_{k=0}^{n-2}\gamma_-(\mathbf{Z}_J^w)(n-1,k)R(\mathbf{Z}_J^w)(-k-1)\Big\}V_+(\mathbf{Z}_J^w)(n-1)^{-1}.$$

The algorithms (i) and (ii) in Theorem 5.2 are called (PAC).

[5.3] We shall investigate the algorithms (FDT) and (PAC) in Theorems 5.1 and 5.2 from the viewpoint of calculating the $\mathrm{KM_2O}$-Langevin matrix from the covariance matrix function $R(\mathbf{Z}_J)$. For that purpose, we define for each n $(0 \le n \le r - l - \sigma(J))$ a subsystem $\mathcal{LM}(\mathbf{Z}_J; n)$ of the $\mathrm{KM_2O}$-Langevin matrix $\mathcal{LM}(\mathbf{Z}_J)$ by

$$\begin{aligned}\mathcal{LM}(\mathbf{Z}_J; n) \equiv \{&\gamma_\pm^0(\mathbf{Z}_J)(m,k), \delta_\pm^0(\mathbf{Z}_J)(m), V_\pm(\mathbf{Z}_J)(j);\\ &0 \le k < m \le n, 0 \le j \le n\}.\end{aligned} \tag{5.19}$$

It follows from (FDT) in Theorems 5.1 that for each n $(1 \le n \le r - l - \sigma(J))$, the matrices $\gamma_\pm^0(\mathbf{Z}_J)(n,k)$ $(1 \le k < n)$ and $V_\pm(\mathbf{Z}_J)(n)$ can be calculated from the matrices $\delta_\pm^0(\mathbf{Z}_J)(n)$. Therefore, we have only to obtain an algorithm by which the matrices $\delta_\pm^0(\mathbf{Z}_J)(n)$ can be calculated from the system $\mathcal{LM}(\mathbf{Z}_J; n-1)$ and the matrices $R(\mathbf{Z}_J)(m)$ $(0 \le m \le n)$. This can be done according to (PAC) in Theorem 5.2 as follows: by Theorems 4.1 and 4.5 in [12],

$$\lim_{w\to 0}\delta_+(\mathbf{Z}_J^w)(n) = \delta_+^0(\mathbf{Z}_J)(n) \tag{5.20}$$

$$\lim_{w\to 0}\delta_-(\mathbf{Z}_J^w)(n) = \delta_-^0(\mathbf{Z}_J)(n). \tag{5.21}$$

From (PAC), (5.20) and (5.21), we can obtain a method by which $\delta_\pm^0(\mathbf{Z}_J)(n)$ can be calculated from the covariance matrix function $R(\mathbf{Z}_J)$.

Thus we have obtained a computational method for calculating the KM_2O-Langevin matrix from the covariance matrix function. For this reason, we define the KM_2O-Langevin matrix $\mathcal{LM}(\mathbf{Z}_J)$ by

$$\begin{aligned}\mathcal{LM}(\mathbf{Z}_J) &\equiv \mathcal{LM}(R(\mathbf{Z}_J))\\ &\equiv \{\gamma_{\pm}^{0}(R(\mathbf{Z}_J))(n,k), \delta_{\pm}^{0}(R(\mathbf{Z}_J))(n), V_{\pm}(R(\mathbf{Z}_J))(m);\\ &\qquad 0 \le k < n \le r-l-\sigma(J), 0 \le m \le r-l-\sigma(J)\}.\end{aligned} \tag{5.22}$$

6. Non-linear causality of finite rank and non-linear prediction with causality

We shall return to the same situation as in Section 3 and apply the results of the previous section to both the causality problem and the non-linear prediction problem.

Let $\mathbf{X} = (X(n); l \le n \le r)$ be a one-dimensional stochastic process and $\mathbf{Y} = (Y(n); l \le n \le r)$ be a d-dimensional stochastic process on a probability space $(\Omega, \mathcal{B}, P)$ such that both the processes satisfy conditions (E) and (M), where d is a natural integer and l and r are integers.

[6.1] (Determinism and causality with partial non-linear information) Corresponding to [3.2], [3.1.2], [3.1.3] and [3.1.4], we shall give a definition of non-linear determinism of finite rank and three kinds of definitions of non-linear causality of finite rank.

[6.1.1] (Non-linear determinism of finite rank) Let us fix any natural numbers q, D $(1 \le D \le d_q)$. We say that $\mathbf{X}$ has a non-linear determinism of rank (q, D) if there exists an element J of $\mathbf{J}^{(q,D)}$ such that $\mathbf{X}_J^{(l+\sigma(J),r-1)} \xrightarrow{(LC)} \mathbf{X}_{+1}^{(l+\sigma(J),r-1)}$. Immediately from the definition of determinism in [3.2], we have

Theorem 6.1 *Let us fix any natural numbers q, D $(1 \le D \le d_q)$. If $\mathbf{X}$ has a non-linear determinism of rank (q, D), then $\mathbf{X}$ has determinism.*

[6.1.2] (Non-linear causality of finite rank) Let us fix any natural numbers q, D $(1 \le D \le d_q)$. We say that there exists a non-linear causality of rank (q, D) from $\mathbf{Y}$ to $\mathbf{X}$ if there exists an element J of $\mathbf{J}^{(q,D)}$ such that $\mathbf{Y}_J \xrightarrow{(LC)} \mathbf{X}^{(l+\sigma(J),r)}$. Immediately from the definition of non-linear causality in [3.1.2], we have

Theorem 6.2 *Let us fix any natural numbers q, D $(1 \le D \le d_q)$.*

(i) *If there exists a natural number q such that $\tilde{\mathbf{Y}}^{(q)} \overset{(LC)}{\longrightarrow} \mathbf{X}$, then $\mathbf{Y} \overset{(C)}{\longrightarrow} \mathbf{X}$.*

(ii) *If there exists a non-linear causality of rank (q, D) from $\mathbf{Y}$ to $\mathbf{X}$, then $\mathbf{Y} \overset{(C)}{\longrightarrow} \mathbf{X}$.*

[6.1.3] (Non-linear weak causality of finite rank) To give a definition of non-linear weak causality of finite rank (cf. [3.1.3]), we define a $(d+1)$-dimensional stochastic process $\mathbf{Z} = (Z(n); l \le n \le r-1)$ by

$$Z(n) \equiv {}^t(X(n), {}^tY_{+1}(n)) \qquad (l \le n \le r-1). \tag{6.1}$$

Let us fix any natural numbers q, D $(1 \le D \le (d+1)_q)$. We say that there exists a non-linear weak causality of rank (q, D) from $\mathbf{Y}$ to $\mathbf{X}$ if there exists an element J of $\mathbf{J}^{(q,D)}$ such that $\mathbf{Z}_J^{(l+\sigma(J), r-1)} \overset{(LC)}{\longrightarrow} \mathbf{X}_{+1}^{(l+\sigma(J), r-1)}$. Then, we have

Theorem 6.3 *Let us fix any natural numbers q, D $(1 \le D \le (d+1)_q)$. If there exists a non-linear weak causality of rank (q, D) from $\mathbf{Y}$ to $\mathbf{X}$, then $\mathbf{Y} \overset{(WC)}{\longrightarrow} \mathbf{X}$.*

[6.1.4] (Non-instantaneous and non-linear weak causality of finite rank) Concerning the definition of non-instantaneous and non-linear weak causality in [3.1.4], we define a $(d+1)$-dimensional stochastic process $\mathbf{W} = (W(n); l \le n \le r)$ by

$$W(n) \equiv {}^t(X(n), Y_1(n), Y_2(n), \ldots, Y_d(n)) \quad (l \le n \le r). \tag{6.2}$$

Let us fix any natural numbers q, D $(1 \le D \le (d+1)_q)$. We say that there exists a non-instantaneous and non-linear weak causality of rank (q, D) from $\mathbf{Y}$ to $\mathbf{X}$ if there exists an element J of $\mathbf{J}^{(q,D)}$ such that $\mathbf{W}_J^{(l+\sigma(J), r-1)} \overset{(LC)}{\longrightarrow} \mathbf{X}_{+1}^{(l+\sigma(J), r-1)}$. Then, we have

Theorem 6.4 *Let us fix any natural numbers q, D $(1 \le D \le (d+1)_q)$. If there exists a non-instantaneous and non-linear weak causality of rank (q, D) from $\mathbf{Y}$ to $\mathbf{X}$, then $\mathbf{Y} \overset{(WC^-)}{\longrightarrow} \mathbf{X}$.*

[6.2] (Non-linear predictor with determinism and causality) We shall give four kinds of prediction formulas for calculating the non-linear predictor.

[6.2.1] (Non-linear predictor with non-linear determinism of finite rank) Corresponding to Theorem 6.1, we have the following prediction formula:

Theorem 6.5 *We assume that there exist two natural numbers q, D ($D \le q$) and an element $J = (j_1, j_2, \ldots, j_D)$ of $\mathbf{J}^{(q,D)}$ such that $\mathbf{X}_J^{(l+\sigma(J),r-1)} \overset{(LC)}{\longrightarrow} \mathbf{X}_{+1}^{(l+\sigma(J),r-1)}$.*

(i) *If $j_1 = 0$, then there exists an M_0 ($l + \sigma(J) \le M_0 \le r-1$) such that for any n, p ($l + M_0 \le n \le r-1, 1 \le p \le r-1-n$),*

$$P_{\mathbf{M}^n_{l+\sigma(J)}(\mathbf{X}_J)} X(n+p) = \textit{the first component of}$$
$$\sum_{k=0}^{n-l-\sigma(J)} Q_+(\mathbf{X}_J)(n-l-\sigma(J)+p, n-l-\sigma(J); k) X_J(l+\sigma(J)+k).$$

(ii) *If $j_1 \ne 0$, then there exists an M_0 ($l + \sigma(J) \le M_0 \le r-1$) such that for any n, p ($l + M_0 \le n \le r-1, 1 \le p \le r-1-n$),*

$$P_{\mathbf{M}^n_{l+\sigma(J)}(\mathbf{S})} X(n+p) = \textit{the first component of}$$
$$\sum_{k=0}^{n-l-\sigma(J)} Q_+(\mathbf{S})(n-l-\sigma(J)+p, n-l-\sigma(J); k) S(l+\sigma(J)+k),$$

where $\mathbf{S} = (S(n); l+\sigma(J) \le n \le r-1)$ is the $(D+1)$-dimensional stochastic process defined by

$$S(n) \equiv {}^t(X(n), {}^tX_J(n)). \tag{6.3}$$

[6.2.2] (Non-linear predictor with non-linear causality of finite rank) Corresponding to Theorem 6.2, we have the following prediction formula:

Theorem 6.6 *We assume that there exist two natural numbers q, D ($D \le q$) and an element $J = (j_1, j_2, \ldots, j_D)$ of $\mathbf{J}^{(q,D)}$ such that $\mathbf{Y}_J \overset{(LC)}{\longrightarrow} \mathbf{X}^{(\sigma(J),r)}$. Then, for any n, p ($l + M_0 \le n \le r-1, 1 \le p \le r-1-n$),*

$$P_{\mathbf{M}^n_{l+\sigma(J)}(\mathbf{T})} X(n+p) = \textit{the first component of}$$
$$\sum_{k=0}^{n-l-\sigma(J)} Q_+(\mathbf{T})(n-l-\sigma(J)+p, n-l-\sigma(J); k) T(l+\sigma(J)+k),$$

where $\mathbf{T} = (T(n); l+\sigma(J) \le n \le r-1)$ is the $(D+1)$-dimensional stochastic *process defined by*

$$T(n) \equiv {}^t(X(n), {}^tY_J(n+1)). \tag{6.4}$$

[6.2.3] (Non-linear predictor with non-linear weak causality of finite rank) Corresponding to Theorem 6.3, we have the following prediction formula:

Theorem 6.7 *We assume that there exist two natural numbers q, D ($D \leq (d+1)_q$) and an element $J = (j_1, j_2, \ldots, j_D)$ of $\mathbf{J}^{(q,D)}$ such that $\mathbf{Z}_J^{(l+\sigma(J),r-1)} \stackrel{(LC)}{\longrightarrow} \mathbf{X}_{+1}^{(l+\sigma(J),r-1)}$.*

(i) *If $j_1 = 0$, then there exists an M_0 ($l + \sigma(J) \leq M_0 \leq r-1$) such that for any n, p ($l + M_0 \leq n \leq r-1, 1 \leq p \leq r-1-n$),*

$$P_{\mathbf{M}^n_{l+\sigma(J)}(\mathbf{Z}_J)^{(l+\sigma(J),r-1)}} X(n+p) = \textit{the first component of}$$
$$\sum_{k=0}^{n-l-\sigma(J)} Q_+(\mathbf{Z}_J^{(l+\sigma(J),r-1)})(n-l-\sigma(J)+p, n-l-\sigma(J); k) \\ Z_J(l+\sigma(J)+k).$$

(ii) *If $j_1 \neq 0$, then there exists an M_0 ($l + \sigma(J) \leq M_0 \leq r-1$) such that for any n, p ($l + M_0 \leq n \leq r-1, 1 \leq p \leq r-1-n$),*

$$P_{\mathbf{M}^n_{l+\sigma(J)}(\mathbf{U}^{(l+\sigma(J),r-1)})} X(n+p) = \textit{the first component of}$$
$$\sum_{k=0}^{n-l-\sigma(J)} Q_+(\mathbf{U}^{(l+\sigma(J),r-1)})(n-l-\sigma(J)+p, n-l-\sigma(J); k) \\ U(l+\sigma(J)+k),$$

where $\mathbf{U} = (U(n); \sigma(J) \leq n \leq r-1)$ is the $(D+1)$-dimensional stochastic process defined by

$$U(n) \equiv {}^t(X(n), {}^tZ_J(n)). \tag{6.5}$$

[6.2.4] (Non-linear predictor with non-instantaneous and non-linear weak causality of finite rank) Corresponding to Theorem 6.4, we have the following prediction formula:

Theorem 6.8 *We assume that there exist natural numbers q, D ($D \leq (d+1)_q$) and an element $J = (j_1, j_2, \ldots, j_D)$ of $\mathbf{J}^{(q,D)}$ such that $\mathbf{W}_J^{(l+\sigma(J),r-1)} \stackrel{(LC)}{\longrightarrow} \mathbf{X}_{+1}^{(l+\sigma(J),r-1)}$.*

(i) *If $j_1 = 0$, then there exists an M_0 $(l + \sigma(J) \le M_0 \le r)$ such that for any n, p $(l + M_0 \le n \le r, 1 \le p \le r-1)$,*

$$P_{\mathbf{M}^n_{l+\sigma(J)}(\mathbf{W}_J^{(l+\sigma(J),r-1)})}X(n+p) = \textit{the first component of}$$
$$\sum_{k=0}^{n-l-\sigma(J)} Q_+(\mathbf{W}_J^{(l+\sigma(J),r-1)})(n-l-\sigma(J)+p, n-l-\sigma(J); k) \\ W_J(l+\sigma(J)+k).$$

(ii) *If $j_1 \neq 0$, then there exists an M_0 $(l + \sigma(J) \le M_0 \le r)$ such that for any n, p $(l + M_0 \le n \le r, 1 \le p \le r-1-n)$,*

$$P_{\mathbf{M}^n_{l+\sigma(J)}(\mathbf{V})}X(n+p) = \textit{the first component of}$$
$$\sum_{k=0}^{n-l-\sigma(J)} Q_+(\mathbf{V})(n-l-\sigma(J)+p, n-l-\sigma(J); k)V(l+\sigma(J)+k),$$

where $\mathbf{V} = (V(n); l+\sigma(J) \le n \le r-1)$ is the $(D+1)$-dimensional stochastic process defined by

$$V(n) \equiv {}^t(X(n), {}^tW_J(n)). \tag{6.6}$$

7. Non-linear time series analysis

We first briefly recall Test(S) introduced in [5]. Next, we shall give a refinement of Test(CS) and Test(D) introduced in [6] and [8], respectively. Thirdly, we shall give a method of selecting a certain model behind a given data. Finally, we shall give a prediction formula by taking account of causality.

[7.1] (Test(S)) Let us be given any d-dimensional data $\mathcal{Z} = (\mathcal{Z}(n); 0 \le n \le N)$.

[7.1.1] We shall briefly recall Test(S).

[Step 1] We define a sample mean vector $\mu^{\mathcal{Z}}$ and a sample covariance matrix function $R^{\mathcal{Z}} = ((R^{\mathcal{Z}}_{jk}(n))_{1\le j,k\le d}; |n| \le N)$ by

$$\mu^{\mathcal{Z}} \equiv \frac{1}{N+1}\sum_{n=0}^{N}\mathcal{Z}(n) \tag{7.1}$$

$$\begin{cases} R^{\mathcal{Z}}_{jk}(n) \equiv \dfrac{1}{N+1}\displaystyle\sum_{m=0}^{N-n}(\mathcal{Z}_j(n+m)-\mu^{\mathcal{Z}}_j)(\mathcal{Z}_k(m)-\mu^{\mathcal{Z}}_k) \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (0 \le n \le N) \\ R^{\mathcal{Z}}_{jk}(-n) \equiv R^{\mathcal{Z}}_{kj}(n) \qquad\qquad\qquad\qquad (0 \le n \le N). \end{cases} \tag{7.2}$$

[Step 2] We define the normalized data $\tilde{\mathcal{Z}} = (\tilde{\mathcal{Z}}(n); 0 \le n \le N)$ by

$$\tilde{\mathcal{Z}}_j(n) \equiv R^{\mathcal{Z}}_{jj}(0)^{-1/2}(\mathcal{Z}(n) - \mu^{\mathcal{Z}}_j) \quad (1 \le j \le d, 0 \le n \le N). \tag{7.3}$$

It follows that $\mu^{\tilde{\mathcal{Z}}} = 0$ and $R^{\tilde{\mathcal{Z}}}(0) = I_d$. We know from an empirical rule in data analysis [3] that the maximal number M such that the matrix function $(R^{\tilde{\mathcal{Z}}}(n); |n| \le M)$ is reliable is given by

$$M \equiv [3\sqrt{N+1}/d] - 1, \tag{7.4}$$

where $[x]$ stands for the integer part of a real number x.

[Step 3] By applying the algorithms in Section 5 to the sample correlation matrix function $R^{\tilde{\mathcal{Z}}} = (R^{\tilde{\mathcal{Z}}}(n); |n| \le M)$, we can construct the sample KM$_2$O-Langevin matrix $\{\gamma^0_{\pm}(R^{\tilde{\mathcal{Z}}})(n,k), \delta^0_{\pm}(R^{\tilde{\mathcal{Z}}})(m), V_{\pm}(l); 0 \le k < n \le M, 1 \le m \le M, 0 \le l \le M\}$.

[Step 4] Let us fix any integer i $(0 \le i \le N - M)$. We derive a d-dimensional data $\boldsymbol{\nu}_i = (\nu_i(n); 0 \le n \le M)$, to be called the sample KM$_2$O-Langevin fluctuation data, by

$$\begin{cases} \nu_i(0) \equiv \tilde{\mathcal{Z}}(i) \\ \nu_i(n) \equiv \tilde{\mathcal{Z}}(i+n) + \displaystyle\sum_{k=0}^{n-1} \gamma^0_+(R^{\tilde{\mathcal{Z}}})(n,k)\tilde{\mathcal{Z}}(i+k) \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (1 \le n \le M). \end{cases} \tag{7.5}$$

By taking $M+1$ elements $W_+(n)$ of $GL(d;\mathbf{R})$ such that

$$V_+(n) = W_+(n)^t W_+(n) \quad (0 \le n \le M), \tag{7.6}$$

we proceed to a standardization $\boldsymbol{\xi}_{+i} = (\xi_{+i}(n); 0 \le n \le M)$ of $\boldsymbol{\nu}_i$:

$$\xi_{+i}(n) \equiv W_+(n)^{-1}\nu_i(n) \quad (0 \le n \le M) \tag{7.7}$$

and construct a one-dimensional data $\boldsymbol{\xi}_i = (\xi_i(n); 0 \le n \le d(M+1)-1)$

by

$$\xi_i(n) \equiv (\xi_{+i})_j(m), \quad n = dm + j - 1 \quad (1 \le j \le d, 0 \le m \le M). \tag{7.8}$$

[Step 5] By [9], $\tilde{\mathcal{Z}}$ has stationarity, that is, it is a realization of a local and weak stationary process if and only if for any integer i ($0 \le i \le N-M$), ξ_i is a realization of a standard white noise. According to this important principle, we have proposed Test(S) in [5] by running i ($0 \le i \le N-M$) and taking a statistical reasoning. By using Test(S), we shall investigate in the next section whether a given time series data has stationarity.

[Step 6] By taking account of the prediction analysis that will be discussed in subsection [7.4], we put $i = N - M$ in (7.5) to obtain the following equation:

$$\begin{cases} \tilde{\mathcal{Z}}(N-M) = \nu_{N-M}(0) \\ \tilde{\mathcal{Z}}(N-M+n) = -\displaystyle\sum_{k=0}^{n-1} \gamma_+^0(R^{\tilde{\mathcal{Z}}})(n,k)\tilde{\mathcal{Z}}(N-M+k) \\ \qquad\qquad\qquad\qquad + \nu_{N-M}(n) \quad (1 \le n \le M). \end{cases} \tag{7.9}$$

We call it a sample forward $\mathrm{KM_2O}$-Langevin equation associated with the d-dimensional data $\mathcal{Z}$.

[7.1.2] We have constructed in Section 2 a generating system $\{\boldsymbol{\varphi}_j(\mathbf{Z}); j \in \mathbf{N}^*\}$ of the non-linear information space $\mathbf{N}_l^n(\mathbf{Z})$ for a given d-dimensional local stochastic process $\mathbf{Z}$.

By applying the class $\{\boldsymbol{\varphi}_j; j \in \mathbf{N}^*\}$ of these non-linear transformations to the given data $\tilde{\mathbf{Z}}$, we define for any $q \in \mathbf{N}$ a class $\mathcal{T}^{(q)}(\tilde{\mathcal{Z}})$ of non-linear information of rank q by

$$\mathcal{T}^{(q)}(\tilde{\mathcal{Z}}) \equiv \{\boldsymbol{\varphi}_j(\tilde{\mathcal{Z}}); 0 \le j \le d_q\}. \tag{7.10}$$

In particular, in a similar way to (5.4), for any natural numbers q, D such that $D \le d_q$, we define a class $\mathcal{T}^{(q,D)}(\tilde{\mathcal{Z}})$ of non-linear information of rank (q, D) by

$$\mathcal{T}^{(q,D)}(\tilde{\mathcal{Z}}) \equiv \{\tilde{\mathcal{Z}}_J; J \in \mathbf{J}^{(q,D)}\}, \tag{7.11}$$

where for each element $J = (j_1, j_2, \ldots, j_D)$ of $\mathbf{J}^{(q,D)}$, $\tilde{\mathcal{Z}}_J = (\tilde{\mathcal{Z}}_J(n); \sigma(J) \le$

$n \leq N)$ is the D-dimensional data defined by

$$\tilde{\mathcal{Z}}_J(n) \equiv {}^t(\phi_{j_1}(\tilde{\mathcal{Z}})(n), \phi_{j_2}(\tilde{\mathcal{Z}})(n), \ldots, \phi_{j_D}(\tilde{\mathcal{Z}}(n)). \tag{7.12}$$

By applying Test(S) to these data $\tilde{\mathcal{Z}}_J$, we can examine whether $\tilde{\mathcal{Z}}_J$ has stationarity, that is, whether the original data $\mathcal{Z}$ has stationarity of certain non-linear type.

[7.2] (Test(CS)) We shall briefly recall Test(CS) introduced in [6] and [8] and propose a new Test(CS)-2 as its refinement.

[7.2.1] Let $\mathcal{X} = (\mathcal{X}(n); 0 \leq n \leq N)$ and $\mathcal{Y} = (\mathcal{Y}(n); 0 \leq n \leq N)$ be a one-dimensional and a d-dimensional data, respectively. Moreover, we assume not only that both the data $\mathcal{X}$ and $\mathcal{Y}$ have stationarity, but also that the $(d+1)$-dimensional data $\mathcal{U} \equiv {}^t(\mathcal{X}, {}^t\mathcal{Y})$ has stationarity. We denote by $R^{\mathcal{X}\mathcal{Y}}$ the sample mutual covariance matrix function of $\mathcal{X}$ and $\mathcal{Y}$, that is,

$$\begin{cases} R^{\mathcal{X}\mathcal{Y}}(n) \equiv \dfrac{1}{N+1} \displaystyle\sum_{m=0}^{N-n} (\mathcal{X}(n+m) - \mu^{\mathcal{X}})(\mathcal{Y}(m) - \mu^{\mathcal{Y}}) \\ \hfill (0 \leq n \leq N) \\ R^{\mathcal{X}\mathcal{Y}}(-n) \equiv {}^t R^{\mathcal{X}\mathcal{Y}}(n) \hfill (0 \leq n \leq N). \end{cases} \tag{7.13}$$

By using the sample KM$_2$O-Langevin matrix associated with the data $\mathcal{Y}$ and the matrix functions $R^{\mathcal{Y}}$, $R^{\mathcal{X}\mathcal{Y}}$, according to the definition in [6] and [8], we can define the sample causality function $C_*(\mathcal{X}|\mathcal{Y}) : \{0, 1, \ldots, M\} \to [0, 1]$ by

$$C_n(\mathcal{X}|\mathcal{Y}) \equiv \left\{ \sum_{k=0}^{n} C(n,k) V_+(k) {}^t C(n,k) \right\}^{1/2}, \tag{7.14}$$

where

$$M \equiv [3\sqrt{N+1}/(d+1)] - 1 \tag{7.15}$$

$$C(n,k) = \begin{cases} R^{\mathcal{X}\mathcal{Y}}(n) R^{\mathcal{Y}}(0)^{-1} \qquad (k = 0) \\ \left\{ R^{\mathcal{X}\mathcal{Y}}(n-k) + \displaystyle\sum_{l=0}^{k-1} R^{\mathcal{X}\mathcal{Y}}(n-l) {}^t \gamma_+(k,l) \right\} V_+(k)^{-1} \\ \hfill (1 \leq k \leq M). \end{cases} \tag{7.16}$$

Noting the stationarity of the data $\mathcal{U}$ implies that the sample causality function $C_*(\mathcal{X}|\mathcal{Y})$ is increasing, we called in [6] and [8] the value of the sample causality function at $n = M$ the sample causal value from $\mathcal{Y}$ to $\mathcal{X}$. Moreover, we gave a qualitative test to judge whether there exists a linear causality from $\mathcal{Y}$ to $\mathcal{X}$, to be denoted by

$$\mathcal{Y} \overset{(LC)}{\longrightarrow} \mathcal{X}, \tag{7.17}$$

by comparing the sample causal value from $\mathcal{Y}$ to $\mathcal{X}$ with the statistical distribution of the sample causal values from 1000 kinds of random number sequences to $\mathcal{X}$ and called it Test(CS).

[7.2.2] Since we gave the above criterion of Test(CS) by using so many physical random number sequences which do not have any causal relation with a target data, it becomes a negative assertion to say that there exist causal relations between given two data. In order to make the statement a positive assertion, we shall give in this paper a new method for judging qualitatively whether there exists a linear causality from $\mathcal{Y}$ to $\mathcal{X}$ by using the shifted data of $\mathcal{X}$ which affect $\mathcal{X}$ always as a cause. We define for each i $(1 \leq i \leq N)$, $\mathcal{X}^{(0,N-i)} = (\mathcal{X}^{(0,N-i)}(n); 0 \leq n \leq N-i)$ and $\mathcal{X}_{+i} = (\mathcal{X}_{+i}(n); 0 \leq n \leq N-i)$ by

$$\mathcal{X}^{(0,N-i)}(n) \equiv \mathcal{X}(n) \qquad (0 \leq n \leq N-i) \tag{7.18}$$

$$\mathcal{X}_{+i}(n) \equiv \mathcal{X}(n+i) \qquad (0 \leq n \leq N-i). \tag{7.19}$$

We can see that for each i $(1 \leq i \leq N)$, there exists a linear causality from $\mathcal{X}_{+i}$ to $\mathcal{X}^{(0,N-i)}$ if and only if $i \leq M(i)$, where $M(i) \equiv [3\sqrt{N-i+1}/2] - 1$. The maximal value sh_c of i $(1 \leq i \leq N)$ satisfying the inequality $i \leq M(i)$ is given by

$$sh_c \equiv [(-17 + 3\sqrt{16N+41})/8]. \tag{7.20}$$

This calculation is due to Mr. Masaya Matsuura.

Test(CS)-2: By running i $(1 \leq i \leq sh_c)$, we first obtain a distribution of the sample causal values $C_{M(i)}(\mathcal{X}^{(0,N-i)} \mid \mathcal{X}_{+i})$ from $\mathcal{X}_{+i}$ to $\mathcal{X}^{(0,N-i)}$. We judge that there exists a linear causality from $\mathcal{Y}$ to $\mathcal{X}$ if the sample causal value $C_M(\mathcal{X}|\mathcal{Y})$ from $\mathcal{Y}$ to $\mathcal{X}$ lies inside 90% from the top of the distribution obtained above.

[7.2.3] (Test(D)) Let $\mathcal{X}$ be a one-dimensional data which passes Test(S). We define two data $\mathcal{X}^{(0,N-1)} = (\mathcal{X}^{(0,N-1)}(n); 0 \leq n \leq N-1)$,

$\mathcal{X}_{+1} = (\mathcal{X}_{+1}(n); 0 \le N-1)$ by

$$\mathcal{X}^{(0,N-1)}(n) \equiv \mathcal{X}(n) \quad (0 \le n \le N-1) \tag{7.21}$$

$$\mathcal{X}_{+1}(n) \equiv \mathcal{X}(n+1) \quad (0 \le n \le N-1). \tag{7.22}$$

Corresponding to [6.1.1], by applying the results in [7.1.2] to the data $\mathcal{X}^{(0,N-1)}$, for any natural numbers q, D such that $D \le d_q$, we can define a class $\mathcal{T}^{(q,D)}(\widetilde{\mathcal{X}^{(0,N-1)}})$ of non-linear information of rank (q, D) by

$$\mathcal{T}^{(q,D)}(\widetilde{\mathcal{X}^{(0,N-1)}}) \equiv \{(\widetilde{\mathcal{X}^{(0,N-1)}})_J; J \in \mathbf{J}^{(q,D)}\} \tag{7.23}$$

and a subset $\mathcal{ST}^{(q,D)}(\widetilde{\mathcal{X}^{(0,N-1)}})$ of the set $\mathcal{T}^{(q,D)}(\widetilde{\mathcal{X}^{(0,N-1)}})$ by

$$\begin{aligned} &\mathcal{ST}^{(q,D)}(\widetilde{\mathcal{X}^{(0,N-1)}}) \\ &\quad \equiv \{(\widetilde{\mathcal{X}^{(0,N-1)}})_J \in \mathcal{T}^{(q,D)}(\widetilde{\mathcal{X}^{(0,N-1)}}); \text{ both the data } (\widetilde{\mathcal{X}^{(0,N-1)}})_J \\ &\qquad\qquad \text{and } {}^t(\mathcal{X}, {}^t(\widetilde{\mathcal{X}^{(0,N-1)}})_J) \text{ pass Test(S)}\}. \end{aligned} \tag{7.24}$$

By applying Test(CS)-2 to any element $(\widetilde{\mathcal{X}^{(0,N-1)}})_J$ of $\mathcal{ST}^{(q,D)}(\widetilde{\mathcal{X}^{(0,N-1)}})$, we can judge qualitatively whether there exists a linear causality from $(\widetilde{\mathcal{X}^{(0,N-1)}})_J$ to $\mathcal{X}_{+1}$, to be denoted by

$$(\widetilde{\mathcal{X}^{(0,N-1)}})_J \xrightarrow{(LC)} \mathcal{X}_{+1}. \tag{7.25}$$

We say that $\mathcal{X}$ has a non-linear determinism of rank (q, D) if (7.25) holds for certain element $(\widetilde{\mathcal{X}^{(0,N-1)}})_J$ of $\mathcal{ST}^{(q,D)}(\widetilde{\mathcal{X}^{(0,N-1)}})$.

[7.2.4] Let $\mathcal{X} = (\mathcal{X}(n); 0 \le n \le N)$ and $\mathcal{Y} = (\mathcal{Y}(n); 0 \le n \le N)$ be a one-dimensional data and a d-dimensional data, respectively. Moreover, we assume that $\mathcal{X}$ passes Test(S).

Corresponding to [6.1.2], by applying the results in [7.1.2] to the data $\mathcal{Y}$, for any natural numbers q, D such that $D \le d_q$, we can define a class $\mathcal{T}^{(q,D)}(\tilde{\mathcal{Y}})$ of non-linear information of rank (q, D) by

$$\mathcal{T}^{(q,D)}(\tilde{\mathcal{Y}}) \equiv \{\tilde{\mathcal{Y}}_J; J \in \mathbf{J}^{(q,D)}\} \tag{7.26}$$

and a subset $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{Y}})$ of the set $\mathcal{T}^{(q,D)}(\tilde{\mathcal{Y}})$ by

$$\mathcal{ST}^{(q,D)}(\tilde{\mathcal{Y}}) \equiv \{\tilde{\mathcal{Y}}_J \in \mathcal{T}^{(q,D)}(\tilde{\mathcal{Y}}); \text{ both the data } \tilde{\mathcal{Y}}_J \text{ and } {}^t(\mathcal{X}, {}^t\tilde{\mathcal{Y}}_J) \text{ pass Test(S)}\}. \tag{7.27}$$

By applying Test(CS)-2 to any element $\tilde{\mathcal{Y}}_J$ of $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{Y}})$, we can judge qualitatively whether there exists a linear causality from $\tilde{\mathcal{Y}}_J$ to $\mathcal{X}$, to be denoted by

$$\tilde{\mathcal{Y}}_J \overset{(LC)}{\longrightarrow} \mathcal{X}. \tag{7.28}$$

We say that there exists a non-linear causal relation of rank (q, D) from $\mathcal{Y}$ to $\mathcal{X}$ if (7.28) holds for certain element $\tilde{\mathcal{Y}}_J$ of $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{Y}})$.

[7.2.5] Under the same situation as in [7.2.4], we define two data $\mathcal{Y}_{+1} = (\mathcal{Y}_{+1}(n); 0 \leq N-1)$ and $\mathcal{Z} = (\mathcal{Z}(n); 0 \leq N-1)$ by

$$\mathcal{Y}_{+1}(n) \equiv \mathcal{Y}(n+1) \quad (0 \leq n \leq N-1) \tag{7.29}$$

$$\mathcal{Z}(n) \equiv {}^t(\mathcal{X}(n), {}^t\mathcal{Y}_{+1}(n)) = {}^t(\mathcal{X}(n), {}^t\mathcal{Y}(n+1)) \quad (0 \leq n \leq N-1). \tag{7.30}$$

Corresponding to [6.1.3], by applying the results in [7.1.2] to the data $\mathcal{Z}$, for any natural numbers q, D such that $D \leq (d+1)_q$, we can define a class $\mathcal{T}^{(q,D)}(\tilde{\mathcal{Z}})$ of non-linear information of rank (q, D) by

$$\mathcal{T}^{(q,D)}(\tilde{\mathcal{Z}}) \equiv \{\tilde{\mathcal{Z}}_J; J \in \mathbf{J}^{(q,D)}\} \tag{7.31}$$

and a subset $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{Z}})$ of the set $\mathcal{T}^{(q,D)}(\tilde{\mathcal{Z}})$ by

$$\mathcal{ST}^{(q,D)}(\tilde{\mathcal{Z}}) \equiv \{\tilde{\mathcal{Z}}_J \in \mathcal{T}^{(q,D)}(\tilde{\mathcal{Z}});\ \text{both the data } \tilde{\mathcal{Z}}_J \text{ and } {}^t(\mathcal{X}, {}^t\tilde{\mathcal{Z}}_J) \text{ pass Test(S)}\}. \tag{7.32}$$

By applying Test(CS)-2 to any element $\tilde{\mathcal{Z}}_J$ of $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{Z}})$, we can judge qualitatively whether there exists a linear causality from $\tilde{\mathcal{Z}}_J$ to $\mathcal{X}$, to be denoted by

$$\tilde{\mathcal{Z}}_J \overset{(LC)}{\longrightarrow} \mathcal{X}. \tag{7.33}$$

We say that there exists a non-linear weak causal relation of rank (q, D) from $\mathcal{Y}$ to $\mathcal{X}$ if (7.33) holds for certain element $\tilde{\mathcal{Z}}_J$ of $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{Z}})$.

[7.2.6] Under the same situation as in [7.2.4], we define a $(d+1)$-dimensional data $\mathcal{W} = (\mathcal{W}(n); 0 \leq n \leq N-1)$ by

$$\mathcal{W}(n) \equiv {}^t(\mathcal{X}(n), {}^t\mathcal{Y}(n)). \tag{7.34}$$

Corresponding to [6.1.4], by applying the results in [7.1.2] to the data

$\mathcal{W}$, for any natural numbers q, D such that $D \le (d+1)_q$, we can define a class $\mathcal{T}^{(q,D)}(\tilde{\mathcal{W}})$ of non-linear information of rank (q, D) by

$$\mathcal{T}^{(q,D)}(\tilde{\mathcal{W}}) \equiv \{\tilde{\mathcal{W}}_J; J \in \mathbf{J}^{(q,D)}\} \tag{7.35}$$

and a subset $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{W}})$ of the set $\mathcal{T}^{(q,D)}(\tilde{\mathcal{W}})$ by

$$\mathcal{ST}^{(q,D)}(\tilde{\mathcal{W}}) \equiv \{\tilde{\mathcal{W}}_J \in \mathcal{T}^{(q,D)}(\tilde{\mathcal{W}}); \text{ both the data } \tilde{\mathcal{W}}_J \text{ and } {}^t(\mathcal{X}, {}^t\tilde{\mathcal{W}}_J) \text{ pass Test(S)}\}. \tag{7.36}$$

By applying Test(CS)-2 to any element $\tilde{\mathcal{W}}_J$ of $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{W}})$, we can judge qualitatively whether there exists a linear causality from $\tilde{\mathcal{W}}_J$ to $\mathcal{X}$, to be denoted by

$$\tilde{\mathcal{W}}_J \overset{(LC)}{\longrightarrow} \mathcal{X}. \tag{7.37}$$

We say that there exists a non-instantaneous and non-linear weak causal relation of rank (q, D) from $\mathcal{Y}$ to $\mathcal{X}$ if (7.37) holds for certain element $\tilde{\mathcal{W}}_J$ of $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{W}})$.

[7.3] (Sample causal values and model selection) In this subsection, we shall give a method of model selection by obtaining the maximal sample causal values concerning the non-linear determinism of finite rank stated in [7.2.3] and three kinds of non-linear causality of finite rank stated in [7.2.4], [7.2.5] and [7.2.6]. Let us fix any natural number q.

[7.3.1] Under the same situation as in [7.2.3], for each natural number D $(1 \le D \le d_q)$, we choose the element $(\widetilde{\mathcal{X}^{(0,N-1)}})_{J^{(1)}} = ((\widetilde{\mathcal{X}^{(0,N-1)}})_{J^{(1)}}(n); \sigma(J^{(1)}) \le n \le N-1)$ of $\mathcal{ST}^{(q,D)}(\widetilde{\mathcal{X}^{(0,N-1)}})$ in (7.24) such that the sample causal value from $(\widetilde{\mathcal{X}^{(0,N-1)}})_{J^{(1)}}$ to $\tilde{\mathcal{X}}_{+1}$ is the largest in the set $\mathcal{ST}^{(q,D)}(\widetilde{\mathcal{X}^{(0,N-1)}})$, where D is restricted so as to satisfy the inequality $[3\sqrt{N-\sigma(J^{(1)})}/(D+1)] - 1 \ge 10$.

By taking account of Theorem 6.5 for the non-linear determinism of rank (q, D), we define a $D^{(1)}$-dimensional data $\mathcal{X}^{(1)} = (\mathcal{X}^{(1)}(n); \sigma(J^{(1)}) \le n \le N)$ by

$$\mathcal{X}^{(1)}(n) \equiv \begin{cases} \tilde{\mathcal{X}}_{J^{(1)}}(n) & \text{if the first element of } J^{(1)} = 0 \\ {}^t(\tilde{\mathcal{X}}(n), {}^t\tilde{\mathcal{X}}_{J^{(1)}}(n)) & \text{if the first element of } J^{(1)} \ne 0, \end{cases} \tag{7.38}$$

where $D^{(1)}$ is given by

$$D^{(1)} \equiv \begin{cases} D & \text{if the first element of } J^{(1)} = 0 \\ D+1 & \text{if the first element of } J^{(1)} \neq 0. \end{cases} \tag{7.39}$$

As a model selection, we adopt the sample forward KM$_2$O-Langevin equation associated with the $D^{(1)}$-dimensional data $\tilde{\mathcal{X}}^{(1)}$.

[7.3.2] Under same situation as in [7.2.4], for each natural number D $(1 \leq D \leq d_q)$, we choose the element $\tilde{\mathcal{Y}}_{J^{(2)}} = ((\tilde{\mathcal{Y}}_{J^{(2)}}(n); \sigma(J^{(2)}) \leq n \leq N)$ of $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{Y}})$ in (7.27) such that the sample causal value from $\tilde{\mathcal{Y}}_{J^{(2)}}$ to $\tilde{\mathcal{X}}$ is the largest in the set $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{Y}})$, where D is restricted so as to satisfy the inequality $[3\sqrt{N - \sigma(J^{(2)}) + 1}/(D+1)] - 1 \geq 10$.

By taking account of Theorem 6.6 for the non-linear causality of rank (q, D), we define a $D^{(2)}$-dimensional data $\mathcal{X}^{(2)} = (\mathcal{X}^{(2)}(n); \sigma(J^{(2)}) \leq n \leq N)$ by

$$\mathcal{X}^{(2)} \equiv {}^t(\tilde{\mathcal{X}}, {}^t\tilde{\mathcal{Y}}_{J^{(2)}}), \tag{7.40}$$

where $D^{(2)}$ is given by

$$D^{(2)} \equiv D + 1. \tag{7.41}$$

As a model selection, we adopt the sample forward KM$_2$O-Langevin equation associated with the $D^{(2)}$-dimensional data $\mathcal{X}^{(2)}$.

[7.3.3] Under same situation as in [7.2.5], for each natural number D $(1 \leq D \leq (d+1)_q)$, we choose the element $\tilde{\mathcal{Z}}_{J^{(3)}} = ((\tilde{\mathcal{Z}})_{J^{(3)}}(n); \sigma(J^{(3)}) \leq n \leq N-1)$ of $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{Z}})$ in (7.32) such that the sample causal value from $\tilde{\mathcal{Z}}_{J^{(3)}}$ to $\tilde{\mathcal{X}}_{+1}$ is the largest in the set $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{Z}})$, where D is restricted so as to satisfy the inequality $[3\sqrt{N - \sigma(J^{(3)})}/(D+1)] - 1 \geq 10$.

By applying the linear prediction formula in [5] to the d-dimensional data $\mathcal{Y} = (\mathcal{Y}(n); 0 \leq n \leq N)$, we obtain the value $\mathcal{Y}(N+1)$ of the one-step future of $\mathcal{Y}$, which will be explained in the subsequent subsection [7.4.1]. Therefore, by taking account of Theorem 6.7 for the non-linear weak causality of rank (q, D), we can define a $D^{(3)}$-dimensional data $\mathcal{X}^{(3)} = (\mathcal{X}^{(3)}(n); \sigma(J^{(3)}) \leq n \leq N)$ by

$$\mathcal{X}^{(3)}(n) \equiv \begin{cases} (\tilde{\mathcal{Z}})_{J^{(3)}}(n) & \text{if the first element of } J^{(3)} = 0 \\ {}^t(\tilde{\mathcal{X}}(n), {}^t(\tilde{\mathcal{Z}})_{J^{(3)}}(n)) & \text{if the first element of } J^{(3)} \neq 0, \end{cases} \tag{7.42}$$

where $D^{(3)}$ is given by

$$D^{(3)} \equiv \begin{cases} D & \text{if the first element of } J^{(3)} = 0 \\ D+1 & \text{if the first element of } J^{(3)} \neq 0. \end{cases} \tag{7.43}$$

As a model selection, we adopt the sample forward KM_2O-Langevin equation associated with the $D^{(3)}$-dimensional data $\mathcal{X}^{(3)}$.

[7.3.4] Under same situation as in [7.2.6], for each natural number D $(1 \leq D \leq (d+1)_q)$, we choose the element $\tilde{\mathcal{W}}_{J^{(4)}} = ((\tilde{\mathcal{W}})_{J^{(4)}}(n); \sigma(J^{(4)}) \leq n \leq N-1)$ of $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{W}})$ in (7.36) such that the sample causal value from $\tilde{\mathcal{W}}_{J^{(4)}}$ to $\tilde{\mathcal{X}}$ is the largest in the set $\mathcal{ST}^{(q,D)}(\tilde{\mathcal{W}})$, where D is restricted so as to satisfy the inequality $[3\sqrt{N-\sigma(J^{(4)})}/(D+1)] - 1 \geq 10$.

By taking account of Theorem 6.8 for the non-instantaneous and non-linear weak causality of rank (q, D), we define a $D^{(4)}$-dimensional data $\mathcal{X}^{(4)} = (\mathcal{X}^{(4)}(n); \sigma(J^{(4)}) \leq n \leq N)$ by

$$\mathcal{X}^{(4)}(n) \equiv \begin{cases} (\tilde{\mathcal{W}})_{J^{(4)}}(n) & \text{if the first element of } J^{(4)} = 0 \\ {}^t(\tilde{\mathcal{X}}(n), {}^t(\tilde{\mathcal{W}})_{J^{(4)}}(n)) & \text{if the first element of } J^{(4)} \neq 0, \end{cases} \tag{7.44}$$

where $D^{(4)}$ is given by

$$D^{(4)} \equiv \begin{cases} D & \text{if the first element of } J^{(4)} = 0 \\ D+1 & \text{if the first element of } J^{(4)} \neq 0. \end{cases} \tag{7.45}$$

As a model selection, we adopt the sample forward KM_2O-Langevin equation associated with the $D^{(4)}$-dimensional data $\mathcal{X}^{(4)}$.

[7.3.5] **Model Selection** We choose the number j_0 $(1 \leq j_0 \leq 4)$ such that the sample causal value obtained in subsection $[7.3.j_0]$ is the largest among the maximal sample causal values obtained in four subsections $[7.3.j]$ $(1 \leq j \leq 4)$.

As a model for the original data $\mathcal{X}$, we adopt the sample forward KM_2O-Langevin equation associateed with the $D^{(j_0)}$-dimensional data $\mathcal{X}^{(j_0)}$.

[7.4] (Prediction Formulas) We shall give four kinds of prediction formulas based upon the models selected in [7.3.1], [7.3.2], [7.3.3] and [7.3.4].

[7.4.1] (Linear predictor) Let $\mathcal{Z} = (\mathcal{Z}(n); 0 \leq n \leq N)$ be a D-dimensional data which passes Test(S). Therefore, there exists a D-dimensional stochastic process $\mathbf{Z} = (Z(n); 0 \le n \le M)$ defined on a probablity space $(\Omega, \mathcal{B}, P)$ satisfying weak stationarity whose realizaion is equal to the data $\tilde{\mathcal{Z}}_{N-M} = (\tilde{\mathcal{Z}}(N-M+n); 0 \le n \le M)$. In [5], we have given a linear prediction formula for calculating the linear predictor $\overrightarrow{\mathcal{Z}}_M^{(L;1)}(N+1)$ of the one-step future of the data $\tilde{\mathcal{Z}}$. The strategy is to assume that the stochastic process $\mathbf{Z}$ would keep weak stationarity until the one-step future. By applying the prediction formula (7.1) in [5] to the data $\mathcal{Z}_{N-M}$, we have

$$\overrightarrow{\mathcal{Z}}_M^{(L;1)}(N+1) = H_M^{(L;1)}(\mathcal{Z}(N), \mathcal{Z}(N-1), \dots, \mathcal{Z}(N-M+1)), \tag{7.46}$$

where $H_M^{(L;1)} = H_M^{(L;1)}(z_N, z_{N-1}, \dots, z_{N-M+1})$ is the $\mathbf{R}^D$-valued function defined by

$$\begin{aligned} &H_M^{(L;1)}(z_n, z_{N-1}, \dots, z_{N-M+1}) \\ &\equiv \mu^{\mathcal{Z}} - \sum_{k=0}^{M-1} \begin{pmatrix} \sqrt{R_{11}^{\mathcal{Z}}(0)} & & 0 \\ & \ddots & \\ 0 & & \sqrt{R_{dd}^{\mathcal{Z}}(0)} \end{pmatrix} \gamma_+(\tilde{\mathcal{Z}})(M,k) \\ &\quad \cdot \begin{pmatrix} \sqrt{R_{11}^{\mathcal{Z}}(0)^{-1}} & & 0 \\ & \ddots & \\ 0 & & \sqrt{R_{dd}^{\mathcal{Z}}(0)^{-1}} \end{pmatrix} (z_{N-M+k+1} - \mu^{\mathcal{Z}}). \end{aligned} \tag{7.47}$$

Moreover, for any $p \in \mathbf{N}$, the linear predictor $\overrightarrow{\mathcal{Z}}_M^{(L;1)}(N+p)$ of the p-step future of the data $\mathcal{Z}$ is given by

$$\begin{aligned} &\overrightarrow{\mathcal{Z}}_M^{(L;1)}(N+p) \\ &\quad \equiv H_M^{(L;1)}(\overrightarrow{\mathcal{Z}}_M^{(L;1)}(N+p-1), \dots, \overrightarrow{\mathcal{Z}}_M^{(L;1)}(N+p-M)), \end{aligned} \tag{7.48}$$

where for any k $(0 \le k \le N)$,

$$\overrightarrow{\mathcal{Z}}_M^{(L;1)}(k) \equiv \mathcal{Z}(k). \tag{7.49}$$

[7.4.2] We deal with the one-dimensional target data $\mathcal{X} = (\mathcal{X}(n); 0 \le$

$n \le N+L$) and the d-dimensional data $\mathcal{Y} = (\mathcal{Y}(n); 0 \le n \le N)$. For each j $(1 \le j \le 4)$, we apply the prediction formula (7.46) to the $D^{(j)}$-dimensional data $\mathcal{X}^{(j)} = (\mathcal{X}^{(j)}(\sigma(J^{(j)})+n); 0 \le n \le N-\sigma(J^{(j)}))$ to obtain the one-dimensional data $\hat{\mathcal{X}}^{(j)} = (\hat{\mathcal{X}}^{(j)}(n); N+1 \le n \le N+L)$ which is made from the non-linear predictors of the p-step futures $(1 \le p \le L)$ of the original data $\mathcal{X}$.

[7.4.3] (Multiple correlation coefficient and FPE) Let us fix any j $(1 \le j \le 4)$. We examine the fitness of the data $\hat{\mathcal{X}}^{(j)}$ with the rest $\mathcal{X}^{(N+1,N+L)}$ of the original data $\mathcal{X}$. For that purpose, we define a multiple correlation coefficient R between the data $\mathcal{X}^{(N+1,N+L)}$ and $\hat{\mathcal{X}^{(j)}}$ by

$$R \equiv \frac{\sum_{n=1}^{L}(\mathcal{X}(n+N)-\mu^{\mathcal{X}})(\hat{\mathcal{X}^{(j)}}(n+N)-\mu^{\hat{\mathcal{X}^{(j)}}})}{\sqrt{\sum_{n=1}^{L}(\mathcal{X}(n+N)-\mu^{\mathcal{X}})^2\sum_{n=1}^{L}(\hat{\mathcal{X}^{(j)}}(n+N)-\mu^{\hat{\mathcal{X}^{(j)}}})^2}}. \tag{7.50}$$

We consider the predictor is purposive when

$$R^2 \ge \frac{D^{(j)}F(D^{(j)}, L-D^{(j)}-1;\ 0.05)}{L-D^{(j)}-1+D^{(j)}F(D^{(j)}, L-D^{(j)}-1;\ 0.05)}, \tag{7.51}$$

where $F(*, \star; \alpha)$ is the F-distribution and α is the level which defines the critical region. We denote by $\mathrm{LB}(R^2)$ the right-hand side of (7.51).

Moreover, we can derive the sample forward $\mathrm{KM_2O}$-Langevin equation associated with the $D^{(j)}$-dimensional data $\mathcal{X}^{(j)} = (\mathcal{X}^{(j)}(\sigma(J^{(j)})+n); 0 \le n \le N-\sigma(J^{(j)}))$:

$$\begin{cases} \widetilde{\mathcal{X}^{(j)}}(N-M^{(j)}) = \nu_{N-\sigma(J^{(j)})-M^{(j)}}(0) \\ \widetilde{\mathcal{X}^{(j)}}(N-M^{(j)}+n) \\ \quad = -\displaystyle\sum_{k=0}^{n-1}\gamma_+^0(R^{\widetilde{\mathcal{X}^{(j)}}})(n,k)\widetilde{\mathcal{X}^{(j)}}(N-M^{(j)}+k) \\ \qquad\qquad + \nu_{N-\sigma(J^{(j)})-M^{(j)}}(n) \quad (1 \le n \le M^{(j)}), \end{cases} \tag{7.52}$$

where $M^{(j)}$ is the maximal reliable number of the sample covariance matrix function $R^{\widetilde{\mathcal{X}^{(j)}}}$ given by

$$M^{(j)} \equiv [3\sqrt{N-\sigma(J^j)+1}/D^{(j)}]-1. \tag{7.53}$$

According to [3], we define the final prediction error FPE for the $D^{(j)}$-dimensional data $\mathcal{X}^{(j)}$ by

$$\mathrm{FPE} \equiv \left(1 + \frac{M^{(j)}D^{(j)}+1}{N-\sigma(J^j)+1}\right)^{D^{(j)}} \left(1 - \frac{M^{(j)}D^{(j)}+1}{N-\sigma(J^j)+1}\right)^{-D^{(j)}} \det(V_+(\mathcal{X}^j)(M^{(j)})). \quad (7.54)$$

8. Data analysis

In this section, we shall apply the results of Section 7 to three kinds of data such as the sea surface temperature at 0 degree in latitude and 100 degree west in longitude, the air temperature observed at Lima Callao Airport and the air pressure observed at the same place. We call these data SST*, LIT and LIP†, respectively. These data are monthly average data of length 130 from May, 1985 to February, 1996 (Figures 8.1, 8.2 and 8.3).

[8.1] (Meteorological data) SST is observed by buoy floating about at 0n100w, which is considered to be strongly connected with El-Niño. El-Niño is related to not only unusual meteorological phenomena but also earth meteorological system. We examine the following two problems with deep interest:

(i) Does SST give a certain meteorological effect to LIT or LIP?
(ii) Do LIT and LIP affect meteorologically each other?

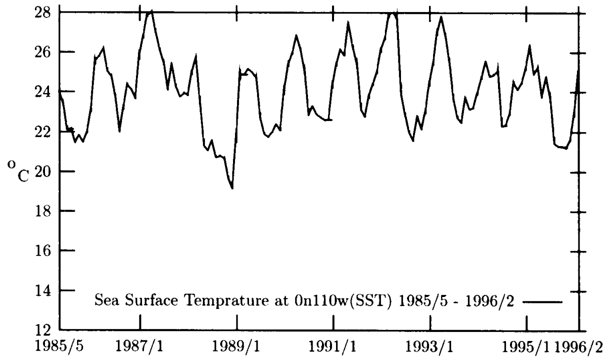

Fig. 8.1. Graph of SST

*SST can be obtained via anonymous ftp from ftp.pmel.noaa.gov

†CD-ROM containing LIT and LIP can be obtained from Japan Meteorological Agency

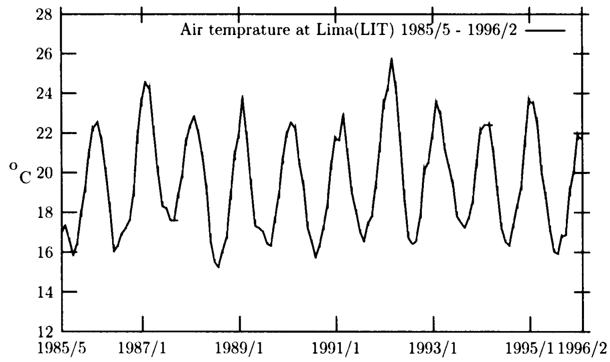

Fig. 8.2. Graph of LIT

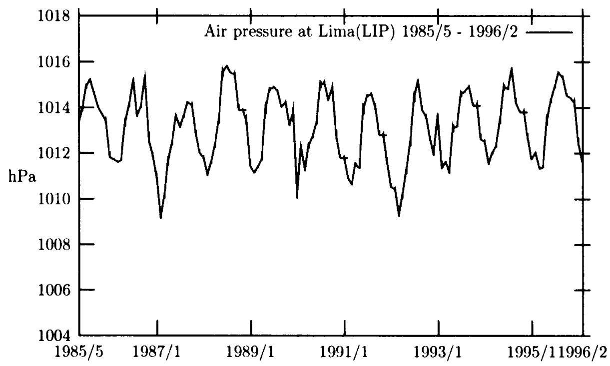

Fig. 8.3. Graph of LIP

To investigate these problems, we denote the data SST, LIT and LIP by $\mathcal{X}_1 = (\mathcal{X}_1(n); 0 \leq n \leq 129)$, $\mathcal{X}_2 = (\mathcal{X}_2(n); 0 \leq n \leq 129)$ and $\mathcal{X}_3 = (\mathcal{X}_3(n); 0 \leq n \leq 129)$, respectively. We perform stationary analysis, deterministic analysis and causal analysis among $\mathcal{X}_j^{(0,119)}$ $(1 \leq j \leq 3)$. Then we predict the future of these data by using the information of the result of causal analysis and then compare the predictor with the rest of data $\mathcal{X}_j^{(120,129)}$ $(1 \leq j \leq 3)$.

[8.2] (Analysis for LIT with SST) In this subsection, we examine whether there exists a causal relation from SST to LIT. For that purpose, we transform the data through the following procedure:

$$\text{Normalization} \to \text{Non-linear transformation} \to \text{Normalization}. \tag{8.1}$$

"Normalization" and "Non-linear transformation" are described in Section 7. Throughout this section, we treat the class of non-linear transformations of rank 6. For one-dimensional case and two-dimensional case, there are 19 transformations and 110 transformations, which are described in Tables A.1 and A.2 of Appendix, respectively.

[8.2.1] (Test(S)) We first check stationarity of LIT.

Table 8.1. Test(S) for one-dimensional non-linear transformations of LIT

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
○	○	○	○	×	×	○	×	×	○	×	○	×	×	×	×	○	○	○

Table 8.1 shows the result of the one-dimensional transformations of LIT. It says that $\phi(\tilde{\mathcal{X}}_2)_j$ passes Test(S) when $j = 0, 1, 2, 3, 6, 9, 11, 16, 17, 18$ and does not pass when $j = 4, 5, 7, 8, 10, 12, 13, 14, 15$. Table 8.2 shows the result of Test(S) for the two-dimensional non-linear transformations of LIT. For example, the pair ${}^t(\phi_j(\tilde{\mathcal{X}}_2), \phi_k(\tilde{\mathcal{X}}_2))$ passes Test(S) when $(j, k) = (0, 1)$ and does not pass when $(j, k) = (0, 2)$.

Table 8.2. Test(S) for two-dimensional non-linear transformations of LIT

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0	○	×	○	×	○	○	×	×	×	×	×	×	×	×	×	×	×	○
1		○	○	×	○	○	×	×	○	○	×	×	×	×	×	×	×	○
2			○	○	×	○	×	○	×	×	×	×	×	○	×	×	×	×
3				×	×	○	×	×	×	×	×	×	×	×	×	×	×	×
4					×	×	○	×	×	×	×	×	×	×	×	×	×	×
5						×	×	×	×	×	×	×	×	×	×	×	×	×
6							×	×	×	×	×	×	×	×	×	×	×	×
7								×	×	×	×	×	×	×	×	×	×	×
8									×	×	×	×	×	×	×	×	×	×
9										×	×	×	×	×	×	×	×	×
10											×	×	×	×	×	○	×	×
11												×	×	×	×	×	×	×
12													×	×	×	×	×	×
13														×	×	×	×	×
14															×	×	×	×
15																×	×	×
16																	×	×
17																		×

On the other hand, Table 8.3 (resp. Table 8.4) shows the result of Test(S) for the one (resp. two)-dimensional transformations of SST.

Table 8.3. Test(S) for one-dimensional non-linear transformations of SST

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
○	○	○	○	○	○	○	×	○	○	○	○	×	×	○	○	○	○	○

Table 8.4. Test(S) for two-dimensional non-linear transformations of SST

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0	○	○	○	○	○	○	○	×	○	○	○	×	○	×	×	○	○	○
1		○	○	×	○	○	×	×	×	○	○	×	×	○	×	×	×	○
2			○	○	○	×	×	○	○	×	×	○	○	×	×	○	×	×
3				×	×	○	×	×	×	×	×	×	×	×	×	×	×	×
4					×	×	○	○	×	×	×	○	×	×	×	×	×	×
5						×	×	×	○	×	×	×	×	×	×	×	×	×
6							×	×	×	×	×	×	×	×	×	×	×	×
7								×	×	×	×	×	×	×	×	×	×	×
8									×	×	×	×	×	○	×	×	×	×
9										○	×	×	×	×	×	×	×	×
10											×	×	×	×	×	×	×	×
11												×	×	×	×	×	×	×
12													×	×	×	×	×	×
13														×	×	×	×	×
14															×	×	×	×
15																×	×	×
16																	×	×
17																		×

It is to be noted that we have to apply Test(S) to the two dimensional data ${}^t(\mathcal{X}_1, \mathcal{X}_2)$.

[8.2.2] (Test(CS)) We shall check whether there exists a certain causal relation from SST to LIT. Table 8.5 shows the result of causal analysis from SST to LIT. Combining with the result of Test(CS)-2 for LIT (Table 8.8), we find that there does not exist either non-linear causality of rank (6,1) or that of rank (6,2) from SST to LIT. It is to be noted that sh_c is 14 when the length of data is 120.

Table 8.5. Test(CS) from SST to LIT

dimension	transformation	sample causal value
1	(0)	0.890167
2	(0,6)	0.895275

Table 8.6. Test(CS) of weak causality from SST to LIT

dimension	transformation	sample causal value
1	(0)	0.962792
2	(0,8)	0.969292

Table 8.7. Test(CS) of non-instantaneous weak causality from SST to LIT

dimension	transformation	sample causal value
1	(0)	0.962792
2	(0,8)	0.969292

On the other hand, Table 8.6 implies that there exist non-linear weak causality of rank (6,1) and rank (6,2) from SST to LIT, but the pair $(0,8)$ of non-linear transformation which has the maximal sample causal value depends upon only the non-linear transformations of LIT. It follows from Table 8.7 that the situation of the non-instantaneous and non-linear weak causality of finite rank from SST to LIT is the same.

However, it follows from Table 8.10 that the sample causal value of the pair $(0, 20)$ which depends upon the non-linear transformations of LIT and SST is the second. Therefore, we can say that there exists a non-linear weak

Table 8.8. Test(CS)-2 for LIT

1	0.995746	8	0.946087
2	0.987807	9	0.935442
3	0.972834	10	0.931748
4	0.954467	11	0.938913
5	0.949964	12	0.946281
6	0.951560	13	0.943704
7	0.950596	14	0.926623

Table 8.9. Test(CS)-2 for SST

1	0.998740	8	0.952032
2	0.997524	9	0.945965
3	0.988394	10	0.938858
4	0.982555	11	0.937387
5	0.975245	12	0.933308
6	0.969588	13	0.931515
7	0.959673	14	0.929517

Table 8.10. Transformations which pass Test(CS)-2

0.969292	(0,8)	0.966595	(0,22)
0.968371	(0,20)	0.966410	(0,49)
0.968357	(0,24)	0.966003	(0,7)
0.968317	(0,80)	0.965918	(0,3)
0.968125	(0,54)	0.965728	(0,12)
0.968007	(0,11)	0.964023	(0,4)
0.967913	(0,2)	0.959035	(0,56)
0.967727	(0,1)	0.955788	(0,92)
0.967579	(0,99)	0.953737	(0,93)
0.967415	(0,27)		

Table 8.11. Test(D) for LIT

dimension	transformation	sample causal value
1	(0)	0.962792
2	(0,3)	0.969292

Table 8.12. Test(D) for SST

dimension	transformation	sample causal value
1	(0)	0.897979
2	(0,11)	0.906990

causality of rank (6,2) from SST to LIT.

[8.2.3] (Test(D)) On the other hand, we see from Tables 8.8 and 8.11 that LIT has non-linear determinism of rank (6,1) and rank (6,2), but we find from Tables 8.9 and 8.12 that SST does not have either non-linear determinism of rank (6,1) or that of rank (6,2).

[8.2.4] (Prediction) We first predict SST whose predictors are shown in Figure 8.4.

The R for the predictors of SST and FPE for the models $(0), (0,11)$ in Figure 8.4 are shown in Table 8.13:

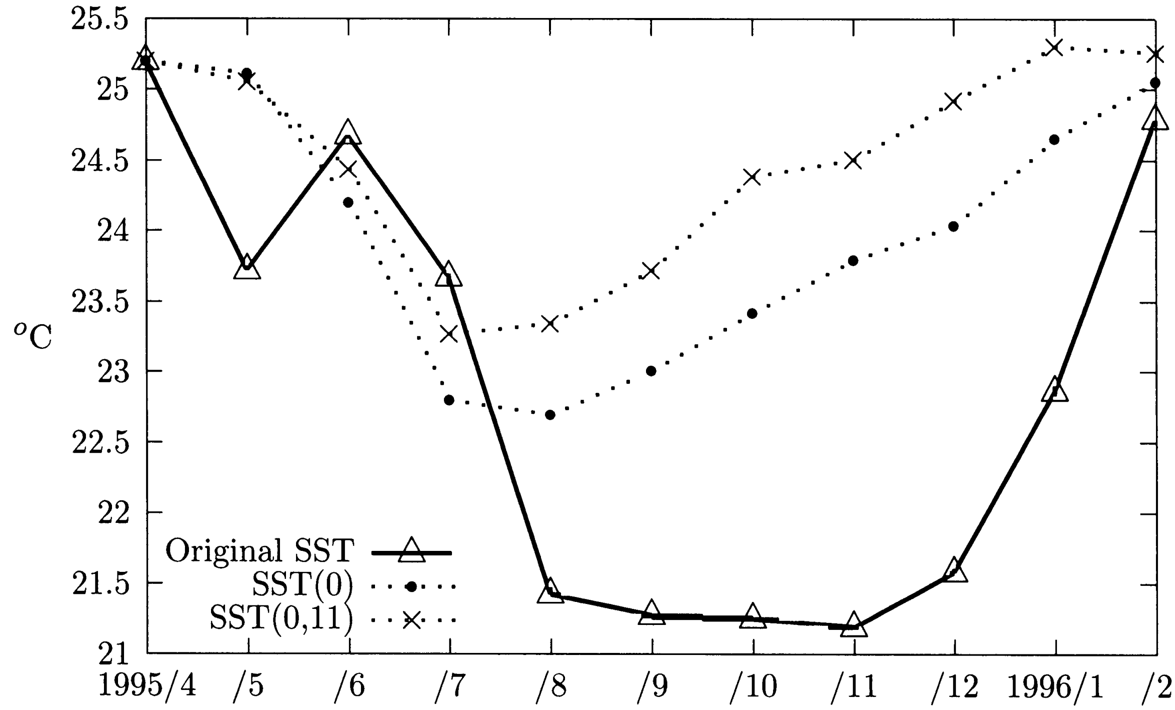

Fig. 8.4. Graph of predictor of SST

Table 8.13. Multiple correlation coefficient and FPE for SST

	sample causal value	Test(CS)-2	R^2	LB(R^2)	FPE
SST (0)	0.897979	×	0.6595423	0.399399	0.181017
SST (0, 11)	0.906990	×	0.8062431	0.575242	0.048086

Though the predictors in Figure 8.4 are purposive, they are not fitted to the original data. The reason seems to be that SST does not have either non-linear determinism of rank (6,1) or that of rank (6,2). We find that the sample causal value (resp. FPE) of the pair $(0, 1)$ is larger (resp. smaller) than the sample causal value (resp. FPE) of the pair (0).

Next, we show in Figure 8.5 the non-linear predictor of LIT using only LIT and both LIT and SST.

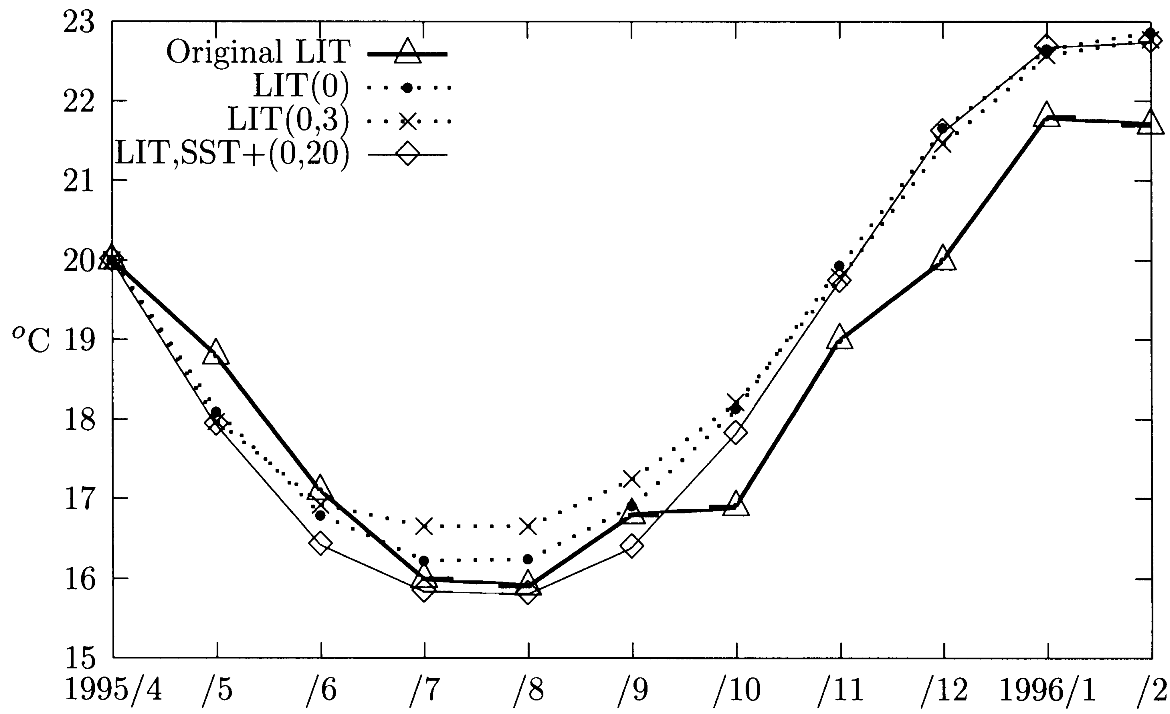

Fig. 8.5. Graph of predictor of LIT with SST

Table 8.14. Multiple correlation coefficient and FPE for LIT

	sample causal value	Test(CS)-2	R^2	LB(R^2)	FPE
LIT (0)	0.962792	○	0.9362284	0.399399	0.07359075
LIT (0, 3)	0.969292	○	0.9242081	0.575242	0.01142466
LIT,SST+ (0, 20)	0.968371	○	0.9407695	0.575243	0.01521758

The predictors are fitted well. The reason seems to be that LIT has non-linear determinism of rank (6,1) and rank (6,2).

[8.3] (Analysis for LIP with SST) In this subsection, we shall pursuit the same analysis for LIP and SST.

[8.3.1] (Test(S)) The results of Test(S) for LIP are shown in Tables 8.15 and 8.16 which tell us that there exist many pairs of non-linear transformations which pass Test(S) unlike the results of Test(S) for LIT and SST.

Table 8.15. Test(S) for one-dimensional non-linear transformations of LIP

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
○	○	○	○	×	○	○	×	×	○	○	○	×	×	×	○	○	○	○

Table 8.16. Test(S) for two-dimensional non-linear transformations of LIP

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0	○	×	○	○	○	○	×	○	○	○	○	×	×	○	○	○	○	○
1		○	○	×	×	○	○	×	○	×	○	×	×	×	×	○	○	○
2			○	○	×	○	×	○	×	○	○	○	×	○	×	○	×	○
3				○	○	○	○	○	○	○	○	○	×	○	○	○	○	○
4					×	○	×	×	○	○	○	×	○	×	×	×	○	○
5						○	×	×	○	○	○	×	○	×	×	○	○	○
6							○	○	○	○	○	○	○	○	○	○	○	○
7								○	○	○	○	×	×	○	×	○	×	○
8									×	×	○	×	×	×	×	×	×	○
9										○	○	○	○	×	×	○	○	○
10											○	×	×	×	×	○	○	×
11												○	○	○	○	○	○	○
12													○	×	×	○	×	○
13														×	×	○	○	×
14															×	○	×	×
15																○	×	×
16																	○	○
17																		○

[8.3.2] (Test(CS)) We show in Tables 8.17, 8.18 and 8.19 the results of Test(CS) from SST to LIP. From the results of Test(CS)-2 in Table 8.20, it follows that there does not exist either non-linear causality of rank (6,1) or that of rank (6,2) from SST to LIP.

Table 8.17. Test(CS) from SST to LIP

dimension	transformation	sample causal value
1	(0)	0.865109
2	(0,16)	0.869533

Table 8.18. Test(CS) of weak causality from SST to LIP

dimension	transformation	sample causal value
1	(0)	0.882594
2	(0,1)	0.901056

Table 8.19. Test(CS) of non-instantaneous weak causality from SST to LIP

dimension	transformation	sample causal value
1	(0)	0.882594
2	(0,8)	0.894097

Table 8.20. Test(CS)-2 for LIP

1	0.996510	8	0.948005
2	0.990698	9	0.942140
3	0.985276	10	0.938962
4	0.982224	11	0.937870
5	0.971999	12	0.933526
6	0.955616	13	0.904458
7	0.951329	14	0.892086

[8.3.3] (Test(D)) Table 8.21 shows the results of Test(D) for LIP. It follows from the results of Test(CS)-2 in Table 8.20 that LIP does not have either non-linear determinism of rank (6,1) or that of rank (6,2).

Table 8.21. Test(D) for LIP

dimension	transformation	sample causal value
1	(0)	0.882594
2	(0,3)	0.894097

[8.3.4] (Prediction) Figure 8.6 shows the non-linear predictor of LIP using only LIP and both LIP and SST. Though the predictors are purposive, the predictors are not so good. The reason seems to be that LIP does not have either non-linear determinism of rank (6,1) or that of rank (6,2) and there does not exist either non-linear causality of rank (6,1) or that of rank (6,2) from SST to LIP.

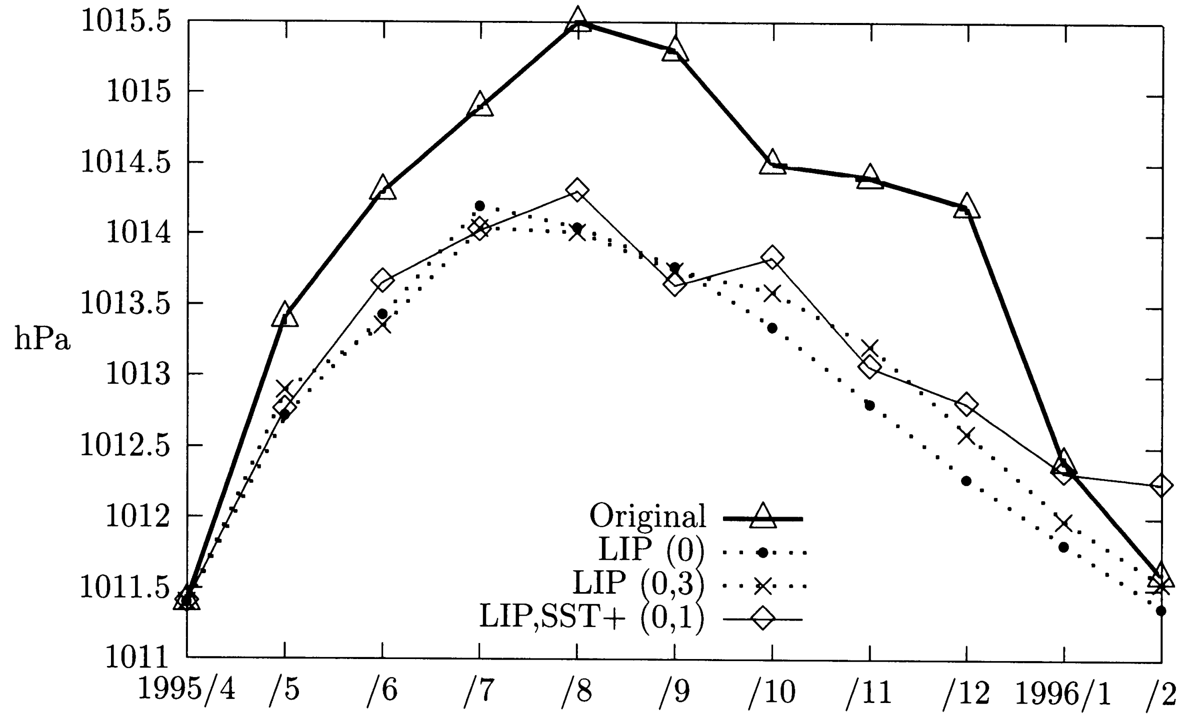

Fig. 8.6. Graph of predictor of LIP with SST

Table 8.22. Multiple correlation and FPE for LIP

	sample causal value	Test(CS)-2	R^2	LB(R^2)	FPE
LIP (0)	0.882594	×	0.833442	0.399399	0.205055
LIP (0, 3)	0.894097	×	0.897190	0.575242	0.068677

[8.4] (Analysis for LIP with LIT) We shall investigate whether there exist causal relations from LIT to LIP and predict LIP.

[8.4.1] (Test(S)) We have given in Tables 8.1 and 8.2, and Tables 8.15 and 8.16 the results of Test(S) for LIT and LIP, respectively. It is to be noted that we have to apply Test(S) to the two dimensional data ${}^t(\mathcal{X}_2, \mathcal{X}_3)$.

[8.4.2] (Test(CS))

We show in Tables 8.23, 8.24 and 8.25 the results of Test(CS) from LIT to LIP. It follows from the results of Test(CS)-2 in Table 8.20 that there exists a non-linear causality of rank (6,2) from LIT to LIP, but that there exist neither non-instantaneous and non-linear weak causality of rank (6,1) nor that of rank (6,2) from LIT to LIP.

Table 8.23. Test(CS) from LIT to LIP

dimension	transformation	sample causal value
1	(0)	0.929463
2	(0,3)	0.935441

Table 8.24. Test(CS) of weak causality from LIT to LIP

dimension	transformation	sample causal value
1	(1)	0.929542
2	(1,36)	0.936778

Table 8.25. Test(CS) of non-instantaneous weak causality from LIT to LIP

dimension	transformation	sample causal value
1	(1)	0.920850
2	(1,36)	0.928766

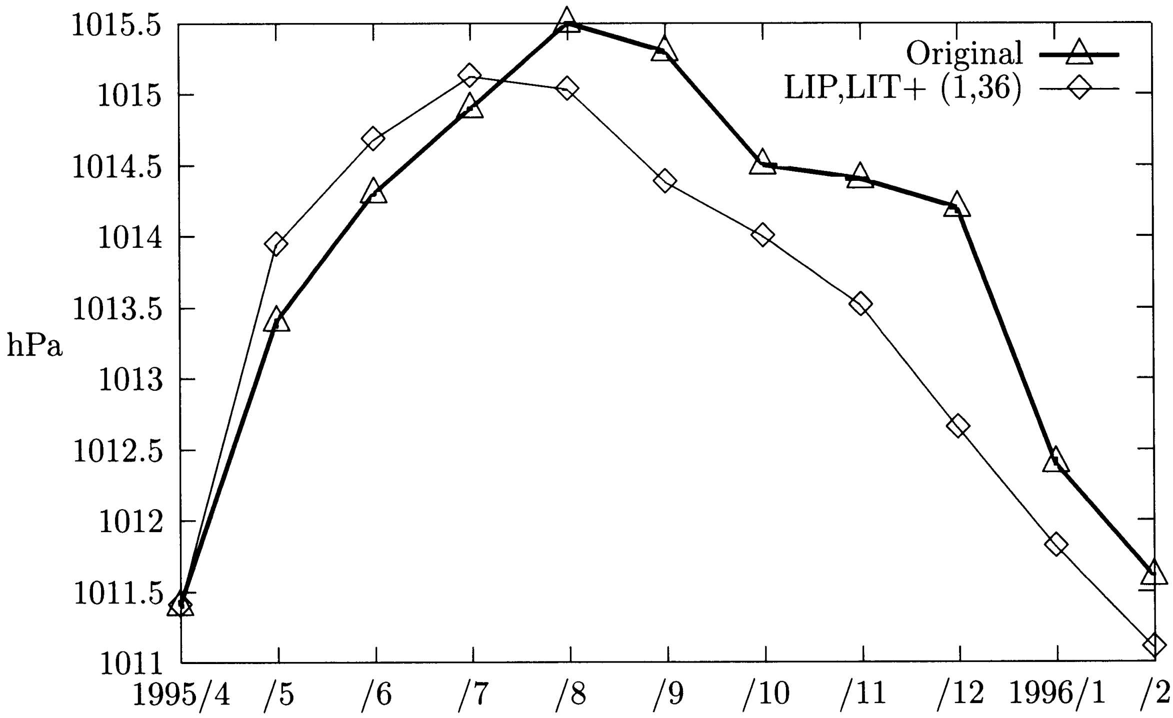

Fig. 8.7. Graph of predictor of LIP with LIT

[8.4.3] (Test(D)) We have shown in Tables 8.11 and 8.21 the results of Test(D) for LIT and LIP, respectively.

[8.4.4] (Prediction) Figure 8.7 shows the non-linear predictor of LIP using only LIP and both LIP and LIT. Though LIP does not have either non-linear determinism of rank (6,1) or that of rank (6,2), the predictors are so good. In Figure 8.6 and Table 8.22, we saw that the predictors of LIP are purposive, but they are not so good. These imply that it is effective to predict the future of LIP by using not only the past information of LIP but also the past information of LIT.

Table 8.26. Multiple correlation and FPE for LIP

	sample causal value	Test(CS)-2	R^2	LB(R^2)	FPE
LIP,LIT+ (1)	0.929542	×	0.757093	0.575242	0.007471
LIP,LIT+ (1,36)	0.936778	○	0.776213	0.704142	0.002329

[8.5] (Analysis for LIT with LIP) Conversely, we shall examine whether there exist causal relations from LIP to LIT and predict LIT.

[8.5.1] (Test(S)) As noted in [8.4.1], we have given in Tables 8.1 and 8.2, and Tables 8.15 and 8.16 the results of Test(S) for LIT and LIP, respectively.

[8.5.2] (Test(CS)) We show in Tables 8.27, 8.28 and 8.29 the results of Test(CS) from LIP to LIT. By the result of Test(CS)-2 for LIT in Table 8.8,

Table 8.27. Test(CS) from LIP to LIT

dimension	transformation	sample causal value
1	(0)	0.953450
2	(0,6)	0.962297

Table 8.28. Test(CS) of weak causality from LIP to LIT

dimension	transformation	sample causal value
1	(0)	0.962792
2	(0,1)	0.973069

Table 8.29. Test(CS) of non-instantaneous weak causality from LIP to LIT

dimension	transformation	sample causal value
1	(0)	0.962792
2	(0,52)	0.969401

we can assert that there exist three types of non-linear causal relations from LIP to LIT.

[8.5.3] (Test(D)) As noted in [8.4.3], we have shown in Tables 8.11 and 8.21 the results of Test(D) for LIT and LIP, respectively.

[8.5.4] (Prediction) Figure 8.8 shows the non-linear predictor of LIT using only LIT and both LIT and LIP.

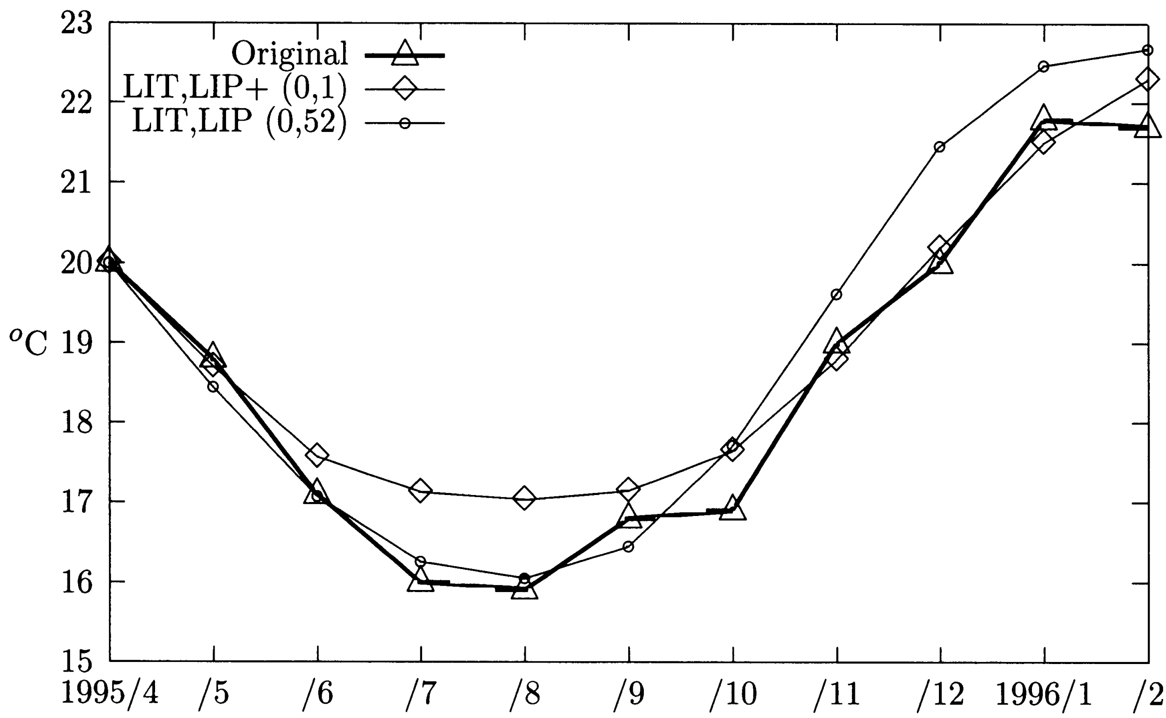

Fig. 8.8. Graph of predictor of LIT with LIP

Table 8.30. Multiple correlation and FPE for LIP

	sample causal value	Test(CS)-2	R^2	LB(R^2)	FPE
LIT,LIP+ (0,1)	0.973069	○	0.958742	0.575242	0.007635
LIT,LIP (0,52)	0.969401	○	0.963443	0.575242	0.021997

The predictors are so good. As stated in [8.2.4], we know that the reason is due to the determinism of LIT.

[8.6] (Analysis for SST with LIT or LIP) For completeness, we shall examine whether there exist causal relations from LIT to SST or from LIP to SST. It follows from Tables 8.31, 8.32, 8.33, 8.34, 8.35 and 8.36 that there do not exist either non-linear causal relations of rank (6,1) or that of rank (6,2) in the both directions from LIT to SST and from LIP to SST.

Table 8.31. Test(CS) from LIT to SST

dimension	transformation	sample causal value
1	(0)	0.897979
2	(0,11)	0.865614

Table 8.32. Test(CS) of weak causality from LIT to SST

dimension	transformation	sample causal value
1	(0)	0.897979
2	(0,97)	0.916923

Table 8.33. Test(CS) of non-instantaneous weak causality from LIT to SST

dimension	transformation	sample causal value
1	(0)	0.897979
2	(0,1)	0.916766

Table 8.34. Test(CS) from LIP to SST

dimension	transformation	sample causal value
1	(0)	0.849893
2	(0,3)	0.871012

Table 8.35. Test(CS) of weak causality from LIP to SST

dimension	transformation	sample causal value
1	(0)	0.897979
2	(0,1)	0.918709

Table 8.36. Test(CS) of non-instantaneous weak causality from LIP to SST

dimension	transformation	sample causal value
1	(0)	0.897979
2	(0,22)	0.913024

[8.7] (Conclusions) We show in Table 8.37 the results of the deterministic analysis for LIP, the causal analysis from SST or LIT to LIP and the prediction analysis for LIP with only LIP and with both SST and LIT.

We find that LIP does not have either non-linear determinism of rank (6,1) or that of rank (6,2) and that there does not exist non-linear causality of rank (6,1) or that of rank (6,2) from SST to LIP. However, we find that there exists non-linear causality of rank (6,2) from LIT to LIP.

Table 8.37. Sample causal value, multiple correlation coefficient and FPE for LIP

	sample causal value	Test(CS)-2	R^2	LB(R^2)	FPE
LIP (0)	0.882594	×	0.833442	0.399399	0.205055
LIP (0,3)	0.894097	×	0.897190	0.575242	0.068677
SST (0)	0.865109	×	0.890475	0.575242	0.024980
SST (0,16)	0.869533	×	0.536594	0.704142	0.007271
LIP,SST+ (0)	0.882594	×	0.833442	0.399399	0.205055
LIP,SST(0,1)	0.865109	×	0.890475	0.575242	0.024980
LIP,SST (0)	0.882594	×	0.833442	0.399399	0.205055
LIP,SST (0,8)	0.894097	×	0.897190	0.575242	0.068677
LIT (0)	0.929542	×	0.757093	0.575242	0.007471
LIT (0,3)	0.935441	○	0.753903	0.704142	0.001367
LIP,LIT+ (1)	0.929542	×	0.757093	0.575242	0.007471
LIP,LIT+ (1,36)	0.936778	○	0.776213	0.704142	0.002329
LIP,LIT (1)	0.920850	×	0.775106	0.575242	0.007692
LIP,LIT (1,36)	0.928766	×	0.780537	0.704142	0.001765

The data chosen by model selection is the three-dimensional one:

$$ {}^t(\phi_0(\mathcal{X}_3, (\mathcal{X}_2)_{+1}), \phi_1(\mathcal{X}_3, (\mathcal{X}_2)_{+1}), \phi_{36}(\mathcal{X}_3, (\mathcal{X}_2)_{+1}) \tag{8.2} $$

On the other hand, the data whose sample causal value is the second is the three-dimensional one:

$$ {}^t(\phi_0(\mathcal{X}_3), \phi_0((\mathcal{X}_2)_{+1}), \phi_3((\mathcal{X}_2)_{+1}) \tag{8.3} $$

The graph in Figure 8.7 of the non-linear predictor for the model (8.2) is purposive. The value of FPE for the model (8.2) is larger than that of FPE for the model (8.3). But, since their difference is about 0.002, it seems that the difference can be ignored.

Figure 8.9 shows the results of Test(CS) for SST, LIT and LIP.

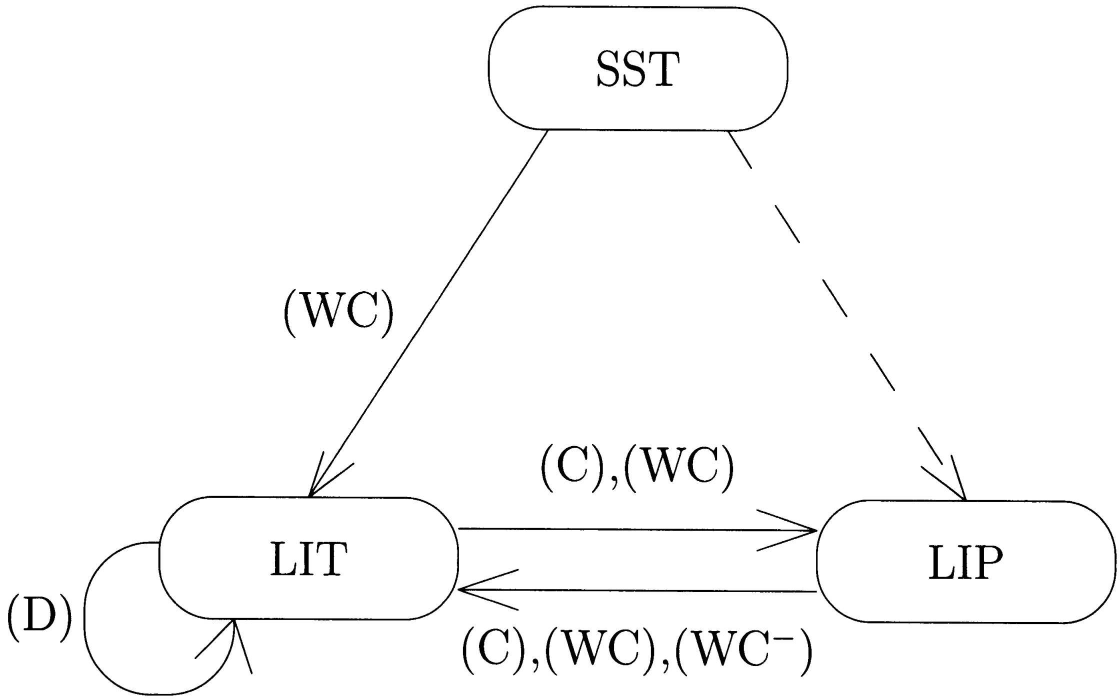

Fig. 8.9. Causal relations among SST, LIT and LIP

Acknowledgments The authors would like to thank the referee for helpful comments.

References

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

[2] Dobrushin R.L. and Minlos R.A., *Polynomials in linear random variables*, Russian Math. Surveys, **32:2** (1977), 71–127.

[3] Akaike H. and Nakagawa T., *Statistical Analysis and Control of Dynamic Systems.* Kluwer Academic Publishers, Tokyo, 1988.

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

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

[6] Okabe Y. and Inoue A., *The theory of KM_2O-Langevin equations and its applications to data analysis* (II): causal analysis (1). Nagoya Math. J. **47** (1994), 1–28.

[7] Okabe Y. and Ootsuka T., *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.

[8] Okabe Y. and Yamane T., *The theory of KM_2O-Langevin equations and its applications to data analysis* (III): deterministic analysis. Nagoya Math. J. **152** (1998), 175–201.

[9] Okabe Y., *On the theory of KM_2O-Langevin equations for stationary flows* (1): characterization theorem. J. Math. Soc. Japan **51** (1999), 817–841.

[10] Okabe Y., *On the theory of KM_2O-Langevin equations for stationary flows* (2): construction theorem. to appear in the special volume in honor of the 70th birthday of Professor Takeyuki Hida, 1999.

[11] Okabe Y. and Matsuura M., *On the theory of KM_2O-Langevin equations for stationary flows* (3): extension theorem. Hokkaido Math. J. **29** (2000), 369–382.

[12] Matsuura M. and Okabe Y., *On a non-linear prediction problem for one-dimensional stochastic processes.* to appear in Japanese Journal of Mathematics, 2001.

[13] Okabe Y., *Non-linear time series analysis based upon the fluctuation-dissipation theorem—with an aim to restoring mathematics as positive science, in preparation.* Japan Hyouronsha (in Japanese), 2000.

A. Appendix

[A.1] Example of non-linear transformations of rank q for d-dimensional data

$$\begin{cases} \varphi_{d_{q-1}+1}(\mathbf{Z}) = (Z_1(n)^q; l \leq n \leq r) \\ \varphi_{d_{q-1}+2}(\mathbf{Z}) = (Z_1(n)^{q-1}Z_2(n); l \leq n \leq r) \\ \varphi_{d_{q-1}+3}(\mathbf{Z}) = (Z_1(n)^{q-1}Z_3(n); l \leq n \leq r) \\ \vdots \\ \varphi_{d_{q-1}+d}(\mathbf{Z}) = (Z_1(n)^{q-1}Z_d(n); l \leq n \leq r) \\ \varphi_{d_{q-1}+d+1}(\mathbf{Z}) = (Z_1(n)^{q-2}Z_1(n)^2; l \leq n \leq r) \\ \varphi_{d_{q-1}+d+2}(\mathbf{Z}) = (Z_1(n)^{q-2}Z_1(n)Z_2(n); l \leq n \leq r) \\ \varphi_{d_{q-1}+d+3}(\mathbf{Z}) = (Z_1(n)^{q-2}Z_1(n)Z_3(n); l \leq n \leq r) \\ \varphi_{d_{q-1}+d+4}(\mathbf{Z}) = (Z_1(n)^{q-2}Z_1(n)Z_4(n); l \leq n \leq r) \\ \vdots \\ \varphi_{d_q-2}(\mathbf{Z}) = (Z_d(n)Z_{d-2}(n-q+2); l+q-2 \leq n \leq r) \\ \varphi_{d_q-1}(\mathbf{Z}) = (Z_d(n)Z_{d-1}(n-q+2); l+q-2 \leq n \leq r) \\ \varphi_{d_q}(\mathbf{Z}) = (Z_d(n)Z_d(n-q+2); l+q-2 \leq n \leq r) \end{cases} \tag{A.1}$$

[A.2] Non-linear transformations of rank 6 for one-dimensional data

Table A.1. One-dimensional non-linear transformations

0	$Z_1(n)$	10	$Z_1(n)Z_1(n-1)^2$
1	$Z_1(n)^2$	11	$Z_1(n)Z_1(n-3)$
2	$Z_1(n)^3$	12	$Z_1(n)^6$
3	$Z_1(n)Z_1(n-1)$	13	$Z_1(n)^4Z_1(n-1)$
4	$Z_1(n)^4$	14	$Z_1(n)^3Z_1(n-2)$
5	$Z_1(n)^2Z_1(n-1)$	15	$Z_1(n)^2Z_1(n-1)^2$
6	$Z_1(n)Z_1(n-2)$	16	$Z_1(n)^2Z_1(n-3)$
7	$Z_1(n)^5$	17	$Z_1(n)Z_1(n-1)Z_1(n-2)$
8	$Z_1(n)^3Z_1(n-1)$	18	$Z_1(n)Z_1(n-4)$
9	$Z_1(n)^2Z_1(n-2)$		

[A.3] Non-linear transformations of rank 6 for two-dimensional data

Table A.2. Two-dimensional non-linear transformations

0	$Z_1(n)$	37	$Z_1(n)^2Z_2(n-2)$	74	$Z_1(n)^2Z_1(n-1)Z_2(n-1)$
1	$Z_2(n)$	38	$Z_1(n)Z_2(n)^4$	75	$Z_1(n)^2Z_2(n-1)^2$
2	$Z_1(n)^2$	39	$Z_1(n)Z_2(n)^2Z_1(n-1)$	76	$Z_1(n)^2Z_1(n-3)$
3	$Z_1(n)Z_2(n)$	40	$Z_1(n)Z_2(n)^2Z_2(n-1)$	77	$Z_1(n)^2Z_2(n-3)$
4	$Z_2(n)^2$	41	$Z_1(n)Z_2(n)Z_1(n-2)$	78	$Z_1(n)Z_2(n)^5$
5	$Z_1(n)^3$	42	$Z_1(n)Z_2(n)Z_2(n-2)$	79	$Z_1(n)Z_2(n)^3Z_1(n-1)$
6	$Z_1(n)^2Z_2(n)$	43	$Z_1(n)Z_1(n-1)^2$	80	$Z_1(n)Z_2(n)^3Z_2(n-1)$
7	$Z_1(n)Z_2(n)^2$	44	$Z_1(n)Z_1(n-1)Z_2(n-1)$	81	$Z_1(n)Z_2(n)^2Z_1(n-2)$
8	$Z_1(n)Z_1(n-1)$	45	$Z_1(n)Z_2(n-1)^2$	82	$Z_1(n)Z_2(n)^2Z_2(n-2)$
9	$Z_1(n)Z_2(n-1)$	46	$Z_1(n)Z_1(n-3)$	83	$Z_1(n)Z_2(n)Z_1(n-1)^2$
10	$Z_2(n)^3$	47	$Z_1(n)Z_2(n-3)$	84	$Z_1(n)Z_2(n)Z_1(n-1)Z_2(n-1)$
11	$Z_2(n)Z_1(n-1)$	48	$Z_2(n)^5$	85	$Z_1(n)Z_2(n)Z_2(n-1)^2$
12	$Z_2(n)Z_2(n-1)$	49	$Z_2(n)^3Z_1(n-1)$	86	$Z_1(n)Z_2(n)Z_1(n-3)$
13	$Z_1(n)^4$	50	$Z_2(n)^3Z_2(n-1)$	87	$Z_1(n)Z_2(n)Z_2(n-3)$
14	$Z_1(n)^3Z_2(n)$	51	$Z_2(n)^2Z_1(n-2)$	88	$Z_1(n)Z_1(n-1)Z_1(n-2)$
15	$Z_1(n)^2Z_2(n)^2$	52	$Z_2(n)^2Z_2(n-2)$	89	$Z_1(n)Z_1(n-1)Z_2(n-2)$
16	$Z_1(n)^2Z_1(n-1)$	53	$Z_2(n)Z_1(n-1)^2$	90	$Z_1(n)Z_2(n-1)Z_1(n-2)$
17	$Z_1(n)^2Z_2(n-1)$	54	$Z_2(n)Z_1(n-1)Z_2(n-1)$	91	$Z_1(n)Z_2(n-1)Z_2(n-2)$
18	$Z_1(n)Z_2(n)^3$	55	$Z_2(n)Z_2(n-1)^2$	92	$Z_1(n)Z_1(n-4)$
19	$Z_1(n)Z_2(n)Z_1(n-1)$	56	$Z_2(n)Z_1(n-3)$	93	$Z_1(n)Z_2(n-4)$
20	$Z_1(n)Z_2(n)Z_2(n-1)$	57	$Z_2(n)Z_2(n-3)$	94	$Z_2(n)^6$
21	$Z_1(n)Z_1(n-2)$	58	$Z_1(n)^6$	95	$Z_2(n)^4Z_1(n-1)$
22	$Z_1(n)Z_2(n-2)$	59	$Z_1(n)^5Z_2(n)$	96	$Z_2(n)^4Z_2(n-1)$
23	$Z_2(n)^4$	60	$Z_1(n)^4Z_2(n)^2$	97	$Z_2(n)^3Z_1(n-2)$
24	$Z_2(n)^2Z_1(n-1)$	61	$Z_1(n)^4Z_1(n-1)$	98	$Z_2(n)^3Z_2(n-2)$
25	$Z_2(n)^2Z_2(n-1)$	62	$Z_1(n)^4Z_2(n-1)$	99	$Z_2(n)^2Z_1(n-1)^2$
26	$Z_2(n)Z_1(n-2)$	63	$Z_1(n)^3Z_2(n)^3$	100	$Z_2(n)^2Z_1(n-1)Z_2(n-1)$
27	$Z_2(n)Z_2(n-2)$	64	$Z_1(n)^3Z_2(n)Z_1(n-1)$	101	$Z_2(n)^2Z_2(n-1)^2$
28	$Z_1(n)^5$	65	$Z_1(n)^3Z_2(n)Z_2(n-1)$	102	$Z_2(n)^2Z_1(n-3)$
29	$Z_1(n)^4Z_2(n)$	66	$Z_1(n)^3Z_1(n-2)$	103	$Z_2(n)^2Z_2(n-3)$
30	$Z_1(n)^3Z_2(n)^2$	67	$Z_1(n)^3Z_2(n-2)$	104	$Z_2(n)Z_1(n-1)Z_1(n-2)$
31	$Z_1(n)^3Z_1(n-1)$	68	$Z_1(n)^2Z_2(n)^4$	105	$Z_2(n)Z_1(n-1)Z_2(n-2)$
32	$Z_1(n)^3Z_2(n-1)$	69	$Z_1(n)^2Z_2(n)^2Z_1(n-1)$	106	$Z_2(n)Z_2(n-1)Z_1(n-2)$
33	$Z_1(n)^2Z_2(n)^3$	70	$Z_1(n)^2Z_2(n)^2Z_2(n-1)$	107	$Z_2(n)Z_2(n-1)Z_2(n-2)$
34	$Z_1(n)^2Z_2(n)Z_1(n-1)$	71	$Z_1(n)^2Z_2(n)Z_1(n-2)$	108	$Z_2(n)Z_1(n-4)$
35	$Z_1(n)^2Z_2(n)Z_2(n-1)$	72	$Z_1(n)^2Z_2(n)Z_2(n-2)$	109	$Z_2(n)Z_2(n-4)$
36	$Z_1(n)^2Z_1(n-2)$	73	$Z_1(n)^2Z_1(n-1)^2$		

Yasunori Okabe
Department of Mathematical Engineering
and Information Physics
Graduate School and Faculty of Engineering
University of Tokyo
Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
E-mail: okabe@langewin2.t.u-tokyo.ac.jp

Akihito Kaneko
Department of Mathematics
Graduate School and Faculty of Science
Hokkaido University
Sapporo 060-0810, Japan
E-mail: a-kaneko@math.sci.hokudai.ac.jp