Regular Paper

Scientific Discovery of Dynamic Models Based on Scale-type Constraints

Fuminori Adachi,† Takashi Washio† and Hiroshi Motoda†

This paper proposes a novel approach to discover dynamic laws and models represented by simultaneous time differential equations including hidden states from time series data measured in an objective process. This task has not been addressed in the past work though it is essentially important in scientific discovery since any behaviors of objective processes emerge in time evolution. The promising performance of the proposed approach is demonstrated through the analysis of synthetic data.

1. Introduction

Many approaches have been developed to identify numerical models of objective process dynamics from measurement data. For example, auto-regressive modeling [1] and artificial neural network [27] capture a relation among measurement variables reflecting dynamics underlying the measurement data. More advanced approaches explicitly introduce the notion of state transition process and measurement process in the framework of state space models. Majority approaches in system identification theory [15], linear and nonlinear model identification based on Hidden Marcov Model (HMM) [18),19] belong to this framework. All these approaches derive an "asymptotic model" of an objective process over a narrow range of its state. Their plausibility is based on the assumption that the characteristics of the objective process over the state range can be sufficiently well captured by the presumed structure of the adopted equations such as linear and/or logistic formulae and the search of the parameters representing the process behaviors. However, this assumption usually does not hold over a wide range of states in the objective process because the presumed structure is merely an approximation within the narrow range. Accordingly the conventional approaches usually do not identify the law equations to represent the first principles governing the objective process over a wide state range.

In contrast, the main goal of scientific law equation discovery is to discover the first principle based law equations from measurement data. BACON [14], FAHRENHEIT [11], ABACUS [5] and IDS [17], which belong to a family called BACON family, are well-known pioneering systems to discover scientific law equations. They search for a "*complete equation*" governing the data measured in a continuous process, where the complete equation is an equation constraining n quantities with $n-1$ degree of freedom ☆. They try to figure out an invariant and its associated relation between two quantities over a wide state range by bi-variate fitting under a given laboratory experiment where some quantities are actively controlled. The found bi-variate relations are successively composed into a complete equation relating multiple measurement quantities. However, one of the drawbacks of the BACON family is its low likeliness to discover the equations representing the first principles underlying the objective process, since they do not use any criteria to capture constraints induced by the first principles. To alleviate this difficulty, some systems introduced constraints of unit dimension to prune the meaningless solutions [5),11),12]. A problem of these approaches is its narrow applicability to only these cases that the units of quantities are clearly known. To further overcome this difficulty, a system called SDS [22] introduced "*scale-type constraints*" to limit the search space to mathematically admissible equations reflecting the first principles governing an objective process. This can be applied to many cases since the knowledge of scale-types is available in various domains.

† The Institute of Scientific and Industrial Research, Osaka University

☆ The equation ${x_1}^2 + {x_2}^2 + ... + {x_n}^2 = 0$ is not complete, since the values of all n quantities is 0, i.e., n quantities are constrained with no degree of freedom. On the other hand, $x_1 + x_2 + ... + x_n = 0$ is complete.

However, many processes such as large scale electric circuits are represented by "*simultaneous equations*". The model representation in form of simultaneous equations is essential to grasp the dependency structure among the multiple mechanisms in the processes [8),16)]. To address this issue, a system called SSF has been proposed [24)]. SSF discovers simultaneous equations governing an objective process based on identification of minimal complete subset of quantities measured under experimental environments. Another issue remained is to discover law equations from the data measured under "*passive observation*". Passive observation means that none of quantities are experimentally controllable due to some practical reasons. This issue has been addressed by extended SDS and SSF which introduced virtual experiments based on data sampling [25),26)]. Furthermore, the discovery of "*simultaneous time differential equations*" under passive observation has been addressed by LAGRANGE [4)]. It has been developed to discover simultaneous time differential equations reflecting dynamics of an objective process. It can discover the model equations from passively observed time series data based on an ILP-like generate and test reasoning on the objective process. Its extended version called LAGRAMGE [21)] introduced domain knowledge of the objective process to limit the search space and to enhance the plausibility of the discovered equations. A recently developed Process Modeling [13)] which takes a similar approach to LAGRAMGE can derive simultaneous time differential equations including hidden state variables by using rich domain specific knowledge.

However, scientists can develop good models based on their domain knowledge without using discovery systems in many cases. Accordingly, the main application domain of the discovery systems may not be the fields in which much domain knowledge is available. Instead, its application domain is the field to identify the simultaneous time differential equations reflecting the first principles under passive observation with "*little domain knowledge*". One such application is the discovery of "*hidden state variables*". Aforementioned approaches lack the ability to discover hidden state variables because they assume direct observation of all state variables in the objective process. Hidden states are the variables that cannot be observed directly. For example, consider a rocket having its mass M [kg] and producing its thrust by the fuel jet of m [kg/sec] and v [m/sec] in space. Then its dynamics is represented by the following three time differential equations.

$$\begin{cases} \dfrac{dV}{dt} = \dfrac{mv}{M}, \\ \dfrac{dX}{dt} = V, \\ \dfrac{dM}{dt} = -m, \end{cases} \tag{1}$$

where V [m/sec] and X [m] are the velocity and the position of the rocket. m and v are the parameters known from the design specification of the rocket. X and V can be measured from the outside of the rocket. But M is not observable unless the rocket has a specific mass sensor. In fact, the measurement of M for a real space rocket is so hard that it must be indirectly estimated from the measurement of X and V. In this case, M is called a hidden state variable since it is unobservable but has its independent dynamics represented by the third differential equation. Without any background domain knowledge, we do not know the number of hidden state variables, *i.e.*, the number of differential equations to be required to model the objective process. The identification of the hidden state variables from observed data is an essential task to discover the simultaneous time differential equations reflecting the first principles underlying the objective process.

In this paper, we propose a novel approach named SCALETRACK (SCALE-type and state TRACKing based discovery system) to discover law equations to represent a model of an objective process having the following features.

(1) The model consists of simultaneous time differential equations representing the dynamic behavior of an objective process.
(2) The model is not an asymptotic approximated model but a model representing the first principles governing the objective process.
(3) The model can contain hidden state variables and their governing differential equations.
(4) The model is discovered without using background domain knowledge specific to the objective process.
(5) The model is discovered from passively observed data.

In the rest of this paper, the basic problem

setting and the entire approach of SCALETRACK are outlined in Section 2. The details of SCALETRACK are described in Section 3, and the performance evaluations are shown in Section 4.

2. Outline

2.1 Basic Problem Setting

We adopt the following "*state space expression*" to model an objective processes and measurements without loss of generality.

$$\begin{cases} \dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t)) + \boldsymbol{v}(t), \\ \boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{w}(t), \end{cases} \quad (2)$$
$$(\boldsymbol{v}(t) \sim N(0, \Sigma_v), \boldsymbol{w}(t) \sim N(0, \Sigma_w)),$$

where the first equation is called a "*state equation*" and the second a "*measurement equation.*" $\boldsymbol{x}$ is called a "*state vector*", $\boldsymbol{f}(\boldsymbol{x})$ a "*state function*", $\boldsymbol{v}$ a "*process noise vector*", $\boldsymbol{y}$ a "*measurement vector*", $\boldsymbol{C}$ a "*measurement matrix*", $\boldsymbol{w}$ a "*measurement noise*" and t a "*time index*". $\boldsymbol{f}(\boldsymbol{x})$ is nonlinear in general, and any state transition of $\boldsymbol{x}$ can be represented by this formulation. While $\boldsymbol{C}$ is a linear transformation representing measurement facilities to derive the measurement variables in $\boldsymbol{y}$ from the state variables in $\boldsymbol{x}$, the facilities are artificial and linear in most cases. Thus, this does not reduce the generality of this expression. If $\boldsymbol{C}$ is a unit matrix, all state variables are directly observable through the measurement. If $\boldsymbol{C}$ is column full rank, the values of all state variables with the measurement noise can be estimated by solving the measurement equation with $\boldsymbol{x}$. Otherwise, some state variables cannot be estimated by the measurement equation only. Such state variables are called "*hidden state variables.*" The aforementioned mass of the rocket M is an example of the hidden state variable. The main reason to introduce the state space model in our approach is to systematically represent the hidden state variables. The power of this model to explicitly separate the state transition and the measurement enables a clear representation of the hidden state variables.

In practical setting of discovery, $\boldsymbol{f}(\boldsymbol{x})$ and some elements of $\boldsymbol{x}$ are initially unknown. We can know only subvector $\boldsymbol{x}'(\subseteq \boldsymbol{x})$ measured by artificial measurement facilities. Thus only a submatrix $\boldsymbol{C}'(\subseteq \boldsymbol{C})$ representing a relation between $\boldsymbol{x}'$ and $\boldsymbol{y}$ is initially known. So our proposing method should identify the correct dimension of $\boldsymbol{x}$ including hidden state variables based on the given measurement data at first. Subsequently, it searches plausible candidates of $\boldsymbol{f}(\boldsymbol{x})$ reflecting the first principles. More concretely speaking in the former rocket example, let a sub state vector be $\boldsymbol{x}'^T = (X, V)$ and a measurement vector $\boldsymbol{y}^T = (X, V)$ where M is a hidden state variable and sub measurement matrix $\boldsymbol{C}'$ a unit matrix. SCALETRACK takes a time series of $\boldsymbol{y}^T = (X, V)$, the scale-types of X, V and submatrix $\boldsymbol{C}'$ as inputs, and search the state equations Eq. (1) representing the rocket's dynamics without using any background domain specific knowledge such as a kinematics law equation.

Our framework is totally different from the model discovery approaches of LAGRAMGE and Process Modeling. Their approaches apply law equations given in the knowledge base of process models and background domain knowledge to build the models governing measurement data. The candidate hidden state variables are also explicity given in the knowledge base. They do not discover hidden state variables and law equations but models of measurement behaviors by usging the knowledge base. In contrast, the primary task of SCALETRACK is to discover hidden state variables and foundamental law equations from measurement data without using any domain specific knowledge base. In this regard, the task attacked by SCALETRACK belongs to a far difficult class in terms of lack of background knowledge and wide search space for variables and equations.

2.2 Outline of Approach

The outline of our proposing method is shown in **Fig. 1**. Given a set of measurement data and knowledge on scale-types of measurement variables, the dimension of $\boldsymbol{x}$ is identified through a statistical analysis called "*correlation dimension analysis*" [2]. For each element of $\boldsymbol{y}$, the locus of its temporal change is mapped to a phase space constructed by time-delayed values of the element, and the degree of freedom which is the dimension of $\boldsymbol{x}$ is estimated by calculating the sparseness of the locus in the phase space. Once the dimension of $\boldsymbol{x}$ is known, all possible combinations of scale-types of the elements in $\boldsymbol{x}$ are enumerated based on scale-type constraints from the known measurement submatrix $\boldsymbol{C}'$ and the scale-types of the elements in $\boldsymbol{y}$. Then for all combinations, the admissible candidate equations of $\boldsymbol{f}(\boldsymbol{x})$ are generated. Subsequently, the validity of the candidate is tested through a simulation based track-

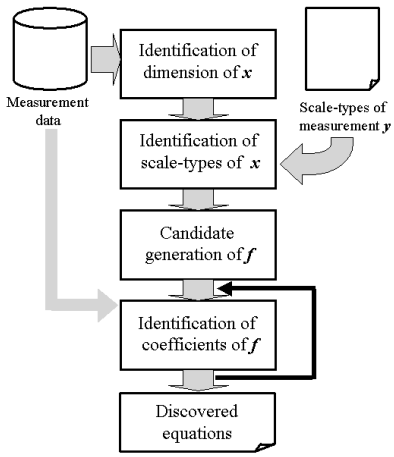

Fig. 1 Block diagram of approach.

ing method called "*Sequential Importance Sampling/Resampling Monte Carlo filter*(SIS/RMC filter)" on the given measurement data. Simulation based tracking is repeated for each candidate to optimize the coefficients in the candidate equations. Then, the combination of candidate $\boldsymbol{f}$ and its coefficients resulting in highly accurate tracking, in terms of "*mean square error* (MSE)", is selected as the discovered dynamic model of the objective process. Through these steps, SCALETRACK discovers the first principle based state space model of the objective process from passively observed data without detailed domain knowledge except for scale-types and measurement facilities. The details of the approach will be described in the following section.

3. Methods

3.1 Identification of Dimension of $\boldsymbol{x}$

When hidden states exist, some elements of state variables of $\boldsymbol{x}(t)$ in an objective process may not be observed. In practical situation, we do not know even the existence of the hidden state variables, in other words, the number of state variables. Accordingly, the dimension of $\boldsymbol{x}(t)$ must be estimated before the candidate equations are searched. To estimate the dimension of $\boldsymbol{x}(t)$, "*correlation dimension analysis*" is used in this work.

First, the time step τ_h is determined, where τ_h is the first time step that the autocorrelation function of an element $y^h(t)$ of $\boldsymbol{y}(t)$ becomes 0.

$$\tau_h = \arg \min_{\tau \in [1,N]} \left\{ \frac{1}{N} \sum_{i=1}^{N} (y^h(i) - y^h_{ave}) \right.$$
$$\left. \times (y^h(i+\tau) - y^h_{ave}) = 0 \right\},$$

where y^h_{ave} is the average value of $y^h(t) \in \boldsymbol{y}(t)(t \in [1, N])$ and N the number of time steps of measurement variables. τ_h is the time step that the local dependency among the observed states is vanished. Then the following vectors having the dimension of m are generated from time series data of $y^h(t)$ based on τ_h.

$$Y^h_1 = \left(y^h(1), y^h(1+\tau_h), \right.$$
$$\left. \ldots, y^h(1+(m-1)\tau_h)\right),$$
$$\vdots$$
$$Y^h_T = \left(y^n(T), y^h(T+\tau_h), \right.$$
$$\left. \ldots, y^h(T+(m-1)\tau_h)\right),$$

where $T+(m-1)\tau_h = N$. Given the dimension of state space $dim(\boldsymbol{x})$, $m \geq 2dim(\boldsymbol{x}) + 1$ is a sufficient condition to preserve the structure of the state space in reconstructed vector space based on "*Takens' theorem of embedding and local reconstruction*" [20]. Hence m should be a sufficiently large value.

Then the correlation dimension can be estimated by calculating correlation integral on reconstructed vector space. The correlation integral is represented by the following formulae;

$$D^m_h(r) = \frac{2}{T(T-1)} \sum_{i=1}^{T} \sum_{j=1}^{i-1} H(r - |Y^h_i - Y^h_j|),$$

where $H(d)$ is an unit step function which is defined as follows.

$$H(d) = \begin{cases} 1 & (d \geq 0), \\ 0 & (d < 0), \end{cases}$$

$D^m_h(r)$ represents the average number of the reconstructed vectors for which the distance from a vector is less than r. Thus, $D^m_h(r)$ represents density of the state in the space which corresponds to the complexity of the state change. For example, when the states are uniformly distributed in a line which is dimension of 1, the number of states in the range of r is in proportion to r as shown in **Fig. 2** (a). When the states are uniformly distributed in a plain whose dimension is 2, the states in the circle with the radius of r is in proportion to r^2 as shown in Fig. 2 (b). Similarly, when the states is uniformly distributed in the n–dimensional state space, the number of points in a n–dimensional hyper–sphere with the radius of r is in proportion to r^n. Hence $D^m_h(r)$ has the following relation with r, where the range of r covers the state distribution. $\nu_h(m)$ is called

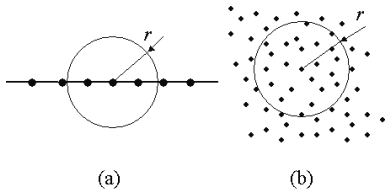

Fig. 2 State variables in a plain.

"*correlation exponent*".

$$D_h^m(r) \propto r^{\nu_h(m)}.$$

$D_h^m(r)$ is calculated for various values of r. Then $\nu_h(m)$ is calculated by the average gradient of $\log(D_h^m(r))$ with respect to $\log r$. Moreover, $\nu_h(m)$ is calculated for various values of m which is the dimension of the phase space. If the value of m becomes large enough, $\nu_h(m)$ asymptotically approaches to the dimension of the state space. Hence the dimension of state space of $\boldsymbol{x}(t)$ is approximately estimated from $\nu_h(m)$ under sufficiently large m. $\nu_h(m)$ is computed for each $y^h(h = 1, ..., dim(\boldsymbol{y}))$, and the nearest integer of its maximum, $\nu_{max}(m)$, is used for $dim(\boldsymbol{x})$, since some measurement variables do not have any relations to some state variables.

3.2 Identification of scale-types of $\boldsymbol{x}$

After $dim(\boldsymbol{x})$ is calculated, the candidate scale-types of $\boldsymbol{x}(t)$ are derived based on "*scale-type constraint*" which limits admissible scale-types of variables based on their mathematical relations. scale-type is defined by the rules of measurement. Physical variables are mainly categorized into two scale-types that are Ratio scale and Interval scale. Ratio scale variables have the absolute origin, and the ratio between two variables is identical in any unit because the admissible unit conversion is similarity group $(x' = cx)$. Mass, velocity and frequency are the examples of Ratio scale variables. On the other hand, Interval scale variables have an arbitrary origin, and the variables represent distance from the origin. The admissible unit conversion is generic linear group $(x' = cx + z)$. Celsius temperature, energy and entropy are the examples of Interval scale variables.

scale-types of some state variables can be estimated from scale-types of measurement variables based on scale-type constraint. **Table 1** shows the examples of scale-type constraints among two state variables x_1, x_2 and a measurement variable y^h. c_{h1} and c_{h2} are constants corresponding to the measurement facility. The first, second, fourth and fifth rows show the cases that a measurement variable y^h depends on only one of the state variables x_i. The origin of Ratio scale variable is defined absolutely while the origin of Interval scale variable is defined arbitrarily. When the scale-type of x_i is Interval, the origin of $c_{hi}x_i$ is not absolute since the origin of x_i is arbitrary defined. Hence, the scale-type of $y^h(= c_{hi}x_i)$ is Interval. In contrast, when the scale-type of x_i is Ratio, the origin of $c_{hi}x_i$ is absolute since the origin of x_i is defined absolutely. Thus the scale-type of $y^h(= c_{hi}x_i)$ is Ratio. Accordingly, the scale-type of x_i is identical to scale-type of y^h in these cases. On the other hand, the scale-type of the other state variable having no relation to y^h are unknown. Its scale-types may be estimated from scale-types of the other measurement variables.

Table 1 scale-types of state variables.

y^h	relation	(x_1, x_2)
R	$y^h = c_{h1}x_1(c_{h1} \neq 0)$	(R,X)
R	$y^h = c_{h2}x_2(c_{h2} \neq 0)$	(X,R)
R	$y^h = c_{h1}x_1 + c_{h2}x_2(c_{h1}c_{h2} \neq 0)$	(I,I)(R,R)
I	$y^h = c_{h1}x_1(c_{h1} \neq 0)$	(I,X)
I	$y^h = c_{h2}x_2(c_{h2} \neq 0)$	(X,I)
I	$y^h = c_{h1}x_1 + c_{h2}x_2(c_{h1}c_{h2} \neq 0)$	(I,X)(X,I)

R:Ratio scale, I:Interval scale, X:unknown (any)

In the third row and sixth raw, the measurement variable is represented by linear summation of the state variables. If the state variables are both Ratio scale, their linear combination is Ratio scale since the relation does not introduce any arbitrary origins to the measurement variables. If the state variables are both Interval scale, the measurement variable can be Ratio scale, since the arbitrary origin of the state variables may mutually cancel out. On the other hand, when the scale-type of a state variable is Ratio and another Interval, the scale-type of y^h is Interval, because the unique arbitrary origin of the Interval state variable remains in the measurement variable through the linear combination. In summary, when the scale-type of measurement variable is Ratio, the admissible scale-types of state variables are both Ratio or both Interval. When the scale-type of measurement variable is Interval, at least the scale-type of one state variable is Interval.

The principle of the scale-type constraints is further generalized for multiple state variables as follows. Let y^h be an h–th measurement variable and $\boldsymbol{x}^h$ a set of state variables where each variable x_i in it has a nonzero coefficient c_{hi}

in the h–th row of the measurement matrix C. Given $y^h = \sum_{x_i \in \boldsymbol{x}^h} c_{hi} x_i$ and the scale-type of y^h,

1. When y^h is Ratio scale,
   all state variables in $\boldsymbol{x}^h$ are Ratio scale. Or two or more state variables in $\boldsymbol{x}^h$ are Interval scale and the others Ratio scale.
2. When y^h is Interval scale,
   at least one state variable in $\boldsymbol{x}^h$ is Interval scale and the others Ratio scale.

scale-types of state variables are estimated by using these constraints, the measurement equation and the scale-types of measurement variables. In case that the scale-types of some state variables are left to be unknown, both possibilities of Ratio scale and Interval scale of the variables are considered. Based on these principles, "*Candidate Combinations of Scale-types* (CCS)", which is the set of all admissible combinations of scale-types of the state variables, is derived.

3.3 Candidate Generation of $\boldsymbol{f}$

Once CCS is given, the candidates of state equation $\boldsymbol{f}$ are generated for each combination of scale-types in CCS based on "*Extended Product Theorem*" [23] which is the extension of the scale-type constraint for admissible multivariate relations. Then a set of all Candidate State Equations (CSE) is obtained.

The following is the Extended Product Theorem where some notions are adapted to our descriptions.

Extended Product Theorem: *Let R be a set of state variables which are Ratio scale, and another I be a set of state variables which are Interval scale, the state variables have the following relation*

$$\dot{x}_i = \alpha \left(\prod_{x_j \in R} |x_j|^{\beta_j} \right) \times \left(\prod_{I_k \subseteq (I - I_g)} \left(\sum_{x_k \in I_k} |x_k| + \delta_k \right)^{\beta_k} \right) \times \prod_{x_l \in I_g} \exp\left(\gamma_{gl} |x_l|\right),$$

where $x_i \in R \cup I$, all coefficients are constants, I_g a subset of I and $\{I_k\}$ a covering of $(I - I_g)$.

For example, given a two dimensional state vector $\boldsymbol{x} = (x_1, x_2)^T$ where $R = \{x_1\}$ and $I = \{x_2\}$, admissible candidate formulae for an element $\dot{x}_i$ of $\dot{\boldsymbol{x}}$ in a state equation are the followings.

$$\dot{x}_i = \begin{cases} \alpha x_1^{\beta_1} (x_2 + \delta_1)^{\beta_2}, \\ \alpha x_1^{\beta_1} \exp(\gamma x_2), \end{cases} \qquad (i = 1, 2)$$

where αs, βs, γ and δ are constants, According to this expression, four candidate state equations of $\boldsymbol{x}$ are generated by taking one of these two formulae for each x_i.

3.4 Identification of Coefficients of $\boldsymbol{f}$

Once CSE is generated, every candidate state equation $eq (\in CSE)$ is tested in this step, and its constant coefficients providing the least error are identified. Moreover, the n–best candidates are selected in terms of their accuracy for solutions. In SCALETRACK, the power coefficients and the other coefficients are processed in different manners, because the locus of a state variable is very sensitive to the change of power coefficients. For example, given the following state equation, a vector of its power coefficients, $\boldsymbol{B} = (\beta_1, \beta_2, \beta_3, \beta_4)$, and another vector of the other coefficients, $\boldsymbol{A} = (\alpha_1, \alpha_2)$, are defined. The former is called "*beta vector*", and the latter is called "*alpha vector*".

$$\begin{cases} \dfrac{\mathrm{d}x_1}{\mathrm{d}t} = \alpha_1 x_1^{\beta_1} x_2^{\beta_2}, \quad \text{and} \\ \dfrac{\mathrm{d}x_2}{\mathrm{d}t} = \alpha_2 x_1^{\beta_3} x_2^{\beta_4}, \end{cases}$$

The mutual distance between two n–dimensional beta vectors is defined as follows.

$$distance(\boldsymbol{B}_1, \boldsymbol{B}_2) = \sum_{i=1}^{n} |\beta_{1i} - \beta_{2i}|,$$

SCALETRACK limits the power coefficients into integers because the power coefficients in most of scientific equations are integers.

3.4.1 Identification Algorithm

The algorithm to identify the plausible state equation representing the objective process is shown in **Fig. 3**. The function **Identify–f** takes some arguments. CSE is the Candidate State Equations generated through the "Candidate Generation" phase, $\boldsymbol{y}_{1:t}$ the measured time series data from time step 1 to t, T the number of trials to obtain the n–best beta vectors, and $\boldsymbol{BV}_0$ the starting point to search beta vectors. T and $\boldsymbol{BV}_0$ are given by users. Currently, T is set to 5, and $\boldsymbol{BV}_0$ consists of three vectors where the first vector consists of 0 valued elements, the second 1 valued elements, and the third -1 valued elements.

In line 2, a candidate eq is selected from CSE, and then its coefficients are identified. SCALETRACK fixes the values of beta vec-

```
1.  Identify–f (CSE, y_{1:t}, T, BV_0, C)
2.    forall eq ∈ CSE do
3.      BV_1 = BV_0
4.      SOL = {φ}
5.      for i=1 to T do
6.        TV = BV_i
7.        forall B_s ∈ BV_i do
8.          TV = TV∪Generate–neighbors(B_s)
9.        forall B_n ∈ TV do
10.         (A_op, MSE)
            = Optimize–alpha(eq, B_n, y_{1:t}, C)
11.         If In–top–n–solutions(SOL, MSE)
            then
12.           sol = (eq, B_n, A_op, MSE)
13.           SOL = SOL ∪ {sol}
14.       BV_{i+1} = {B_n|(eq, B_n, A_op, MSE)
                ∈ Top–n–solutions(SOL)}
15.     TSOL
              = TSOL∪ Top–n–solutions(SOL)
16.   Output(Top–n–solutions(TSOL))
```

Fig. 3 The algorithm of the identification.

tor first, and then every element of alpha vector is optimized under the beta vector. The function **Generate–neighbors**($\boldsymbol{B}_s$) in line 8 generates a set of beta vector $\boldsymbol{B}_n$s satisfying $distance(\boldsymbol{B}_s, \boldsymbol{B}_n) = 1$. This step generates beta vectors which are neighbor of the vectors in $\boldsymbol{BV}_i$. The function **Optimize–alpha** in line 10 provides an optimized alpha vector of the given candidate equation *eq* together with Mean Square Error (MSE) of the optimization based on "*Golden Ratio Division Method*" and "*Sequential Importance Sampling/Resampling Monte Carlo filter* (SIS/RMC filter) [3]".

After the identification of alpha vector is finished, the combination of the identified state equation and its MSE is selected if the MSE is the n–smallest in $\boldsymbol{SOL}$. The function **In–top–n–solutions** (MSE) in line 11 returns true when MSE is in the n–smallest values in $\boldsymbol{SOL}$, otherwise false. Once the solutions $(eq, \boldsymbol{B}_n, \boldsymbol{A}_{op}, MSE)$ are obtained for all neighbor beta vectors, the beta vectors of the n–best solutions among the obtained solutions are stored to $\boldsymbol{BV}_{i+1}$. The generation of $\boldsymbol{BV}_{i+1}$ and the optimization of alpha vector are repeated until the counter i becomes T. The function **Top–n–solutions** in line 14, 15 and 16 returns the list of solutions having the values of MSE which are n–smallest in $\boldsymbol{SOL}$ or $\boldsymbol{TSOL}$. Subsequently, the solutions having the values of MSE which are within the n–smallest in $\boldsymbol{SOL}$ are added to the temporary result $\boldsymbol{TSOL}$. This process is applied to all candidates in CSE. Finally, the n–best solutions of the identified state equations in terms of MSE are provided as the final result.

```
1.  Optimize–alpha(eq, B_n, y_{1:t}, C)
2.    set A =Non–power–coefficients–of(eq)
3.    forall α_i ∈ A α_i = 1
3.    forall α_i ∈ A do
4.      (α_i, E_op)
          =GoldenRatioDM(eq,A,i,B_n,y_{1:t},C)
5.    return (A, E_op)
```

Fig. 4 The algorithm of alpha vectors' identification.

3.4.2 Optimize–alpha

For the estimation of alpha vector, we adopt Golden Ratio Division Method instead of well known Gradient Descent Method. The SIS/RMC filter estimates the state variables so that they fit to the measurement data even if the shapes of state equation and the values of the coefficients are different from those of the original. Accordingly, the gradient of the MSE with respect to a coefficient in the alpha vector becomes to be gentle. In such a situation, Gradient Descent Method is not practical, since the method requires much computation to achieve the convergence of the values of coefficients in the alpha vector. In contrast, the values of coefficients in the alpha vector are converged within almost a constant computation time under Golden Ratio Division Method, since the method narrows down the ranges to search the values in a constant speed.

The algorithm of **Optimize–alpha** is shown in **Fig. 4**. First, all elements of an alpha vector are initialized to 1. Then, the values of all coefficients of $\boldsymbol{A}$ are optimized one by one in terms of MSE of the candidate equation *eq* under fitting to the time series $\boldsymbol{y}_{1:t}$, and the value of MSE, E_{op}, is obtained under the alpha vector, $\boldsymbol{A}$.

3.4.3 SIS/RMC Filter

In SCALETRACK, we adopt the SIS/RMC filter to track the state variables instead of the well-known Kalman Filter (KF) [9] and Extended Kalman Filter (EKF) [7] approaches. The applicability of the KF is limited to linear state equations, and that of the EKF which applies linear approximations of the state equations is also limited to weak non-linearity. In contrast, the SIS/RMC filter which uses a simulations/tests based approach to track the states can be applied to the state equations having strong non-linearity.

SIS/RMC filter is a tracking method based on Bayesian Monte Carlo sampling. Given a time series of measurement vector $\boldsymbol{y}_{1:t}$ and candidate state space model eq which is the state equation in Eq. (2), the probability of the state vector $\boldsymbol{x}(t)$ under given $\boldsymbol{x}(t-1)$ and $\boldsymbol{y}(t)$ is approximated as follows.

$p\left(\boldsymbol{x}(t)|\boldsymbol{x}(t-1),\boldsymbol{y}(t)\right) \propto N(\boldsymbol{m}_t,\Sigma)$, where $\Sigma^{-1}=\Sigma_v^{-1}+\boldsymbol{C}^T\Sigma_w^{-1}\boldsymbol{C}$, and $\boldsymbol{m}_t=\Sigma\left(\Sigma_v^{-1}\bar{\boldsymbol{x}}(t)+\boldsymbol{C}^T\Sigma_w^{-1}\boldsymbol{y}_k\right)$. Here, $\bar{\boldsymbol{x}}(t)$ is the state vector predicted from $\boldsymbol{x}(t-1)$ and the given candidate state equation $\boldsymbol{eq}$. The probability of $\boldsymbol{y}(t)$ under $\boldsymbol{x}(t-1)$ is proportional to the following Gaussian distribution.

$$p(\boldsymbol{y}(t)|\boldsymbol{x}(t-1)) \propto \exp\left(-\frac{1}{2}(\boldsymbol{y}(t)-\boldsymbol{C}\bar{\boldsymbol{x}}(t))^T \times(\Sigma_v+\boldsymbol{C}\Sigma_w\boldsymbol{C}^T)^{-1} \times(\boldsymbol{y}(t)-\boldsymbol{C}\bar{\boldsymbol{x}}(t))\right).$$

$\pi(\boldsymbol{x}(1)|\boldsymbol{y}(1))$ is the probability distribution of an initial state. This is estimated from the measurement vector $\boldsymbol{y}(1)$ and measurement matrix C. As the initial distributions of hidden state variables cannot be estimated, the distributions are given by the users. Under these formulation, the procedure of SIS/RMC filter is designed as follows. The operator "$\sim$" means that a value is sampled under the probability distribution allocated to the right side of the operator.

0. Initialize
   - 0–1 For $i=1,2,\ldots,N$, sample $\boldsymbol{x}^{(i)}(1)\sim\pi(\boldsymbol{x}(1)|\boldsymbol{y}(1))$ and $\tilde{w}^{*(i)}(1)=1/N$.
1. Importance sampling
   - 1–1 For $i=1,2,\ldots,N$, sample $\tilde{\boldsymbol{x}}^{(i)}(t)\sim p(\boldsymbol{x}(t)|\boldsymbol{x}^{(i)}(t-1),\boldsymbol{y}(t))$.
   - 1–2 For $i=1,2,\ldots,N$, evaluate the importance weights up to a normalizing constant: $\tilde{w}^{*(i)}(t)=\tilde{w}^{*(i)}(t-1)p(\boldsymbol{y}(t)|\boldsymbol{x}^{(i)}(t-1))$.
   - 1–3 For $i=1,2,\ldots,N$, normalize the importance weights: $\tilde{w}^{(i)}(t)=\tilde{w}^{*(i)}(t)/(\sum_{i=1}^{N}\tilde{w}^{*(i)}(t))$.
   - 1–4 Let MAP estimation, $\tilde{\boldsymbol{x}}(t)$, be $\tilde{\boldsymbol{x}}^{(i)}(t)$ having the maximum $\tilde{w}^{(i)}(t)$.
   - 1–5 $N_{eff}=(\sum_{i=1}^{N}(\tilde{w}^{(i)}(t))^2)^{-1}$.
   - 1–6 If $N_{eff}\geq N_{thres}$ then goto 1, otherwise goto 2.
2. Resampling
   - 2–1 For $i=1,2,\ldots,N$, sample an index $j(i)$ distributed according to the discrete distribution with N elements satisfying $\Pr\{j(i)=l\}=\tilde{w}^{(l)}(t)$ for $l=1,2,\ldots,N$.
   - 2–2 For $i=1,2,\ldots,N$, $\boldsymbol{x}_{1:t}^{(i)}=\tilde{\boldsymbol{x}}_{1:t}^{j(i)}$ and $w^{(i)}(t)=1/N$, then goto 1.

The filter generates N candidates of time series of $\boldsymbol{x}(t)$ to explain $\boldsymbol{y}_{1:t}$ under a given state space expression. First, the probability distributions of the state vector are calculated based on the input candidate state equation, eq, the estimated state variables, $\tilde{\boldsymbol{x}}(t-1)$, and the measurement vector, $\boldsymbol{y}(t)$. Then the estimated state vector, $\tilde{\boldsymbol{x}}^{(i)}(t)$, is sampled based on the probability distributions. Subsequently, the importance weights are assigned to each state vector, and a sequence consisting of state vectors having the largest value of the weights at each time index is selected.

However, when the importance sampling process is repeated many times, the weight distribution on the N candidates becomes strongly uneven. Accordingly some estimated state vectors in the filter become to have no contribution to the estimation process. This situation is called "*degeneracy*". To avoid this degeneracy, a procedure named "*resampling*" is performed. In resampling, the sequences are sampled N times so that the probability to sample the sequence j is equal to the normalized weight of $\tilde{w}_t^{(j)}$. Then all sequences are replaced with resampled sequences while resetting the weights to escape from the degeneracy. In short, sequences having large weights are copied, and the other sequences are replaced by the copies. The likely sequences survive, while the unlikely sequences die.

Once tracking is done, a sequence of state vectors having the largest weights at each time step is obtained as a tracked history of $\boldsymbol{x}(t)$, $\hat{\boldsymbol{x}}_{1:t}$. Then the time series of measurement variables is estimated from the measurement equation,

$$\hat{\boldsymbol{y}}_{1:t}=C\hat{\boldsymbol{x}}_{1:t},$$

and Mean Square Error (MSE) between the measurement and the estimation is calculated as follows.

$$MSE(\hat{\boldsymbol{y}}_{1:t})=\frac{1}{t}\sum_{i=1}^{t}|\hat{\boldsymbol{y}}(i)-\boldsymbol{y}(i)|^2$$

4. Result

4.1 Implementation

The evaluation of candidate state equations by the SIS/RMC filter is the most time consum-

ing step. Any search cannot be skipped, since the search space is discrete and nonmonotonic. We experienced that one run of stand alone SCALETRACK took a month even if we used an efficient algorithm. Accordingly, the current SCALETRACK introduced a simple grid computing framework using a PC cluster consisting of a server and 10 clients. The server has an AthlonXP 1900+(1.6 GHz) CPU and 2 GB RAM, and each client has an AthlonXP 3000+(2.17 GHz) CPU and 512 MB RAM. The server computes the first three steps, *i.e.*, Identification of dimension of $\boldsymbol{x}$, Identification of scale-types of $\boldsymbol{x}$ and Candidate generation of $\boldsymbol{f}$. Then it allocates a task to evaluate 10% of candidate state equations to each client. Because the task is mutually independent, and occupies the most of computation of SCALETRACK, this implementation accelerates the run speed almost 10 times.

4.2 Basic Performance Evaluation

Basic performance of SCALETRACK in terms of scale-types of state variables, hidden state variables and measurement noise levels is evaluated by using the following two artificial formulae of two dimensions.

1. Model RR:
$$\begin{aligned} \dot{x}_1(t) &= x_1(t)x_2(t), \\ \dot{x}_2(t) &= -0.5x_1(t), \\ \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \boldsymbol{w}_t, \end{aligned}$$
where $y_1(t) = x_1(t)$ and $y_2(t) = x_2(t)$ are Ratio scale. The measurement data were generated by the simulations under one time step $\Delta t = 0.005$ and total steps $n = 600$.
2. Model RI:
$$\begin{aligned} \dot{x}_1(t) &= 0.4x_1(t)(x_2(t) + 0.2), \\ \dot{x}_2(t) &= -0.1(x_2(t) + 0.6), \\ \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \boldsymbol{w}_t, \end{aligned}$$
where $y_1(t) = x_1(t)$ is Ratio scale and $y_2(t) = x_2(t)$ Interval scale. The measurement data were generated by the simulations under one time step $\Delta t = 0.05$ and total steps $n = 600$.

The elements of measurement noise $\boldsymbol{w}_t$ are determined as follows.

$$w_t^h \sim N\left(0, \sigma_w x^h(t)\right),$$

where w_t^h is the h–th element of $\boldsymbol{w}_t$, $x^h(t)$ the h–th element of $\boldsymbol{x}(t)$ and σ_w a relative amplitude of measurement noise. Empirically, m in the correlation dimension analysis and N in the state tracking were chosen to be 8 and 500 respectively. The amplitude of the process noise is set to be 0 to check the pure effect of the measurement noise in the rest of this paper.

Table 2 $\nu_{max}(8)$ for each noise level.

case	σ_w(%)					
	0.1	0.5	1.0	2.0	5.0	10.0
RR	1.11	1.74	2.21	2.59	3.23	3.44
RRH	1.05	1.74	2.21	2.59	3.23	3.44
RI	1.25	1.90	2.19	2.40	2.41	2.58
RIH	1.25	1.90	2.19	2.40	2.41	2.58

Table 3 Basic performance.

case	*ct*	σ_w(%)				
	(hrs.)	0.1	0.5	1.0	2.0	5.0∼
RR	1.5	++	+	+	+	-
RRH	5.5	+	+	-	-	-
RI	4.0	++	+	+	+	-
RIH	5.5	++	+	-	-	-

Table 2 shows the result of correlation dimension analysis. The cases of RR and RI in the table correspond to the above two state space models, and RRH and RIH are the cases where the second measurement variable y_2 is not available, and hence x_2 is hidden. The correlation dimension analysis estimated the dimension of state vectors as around 2 in many cases, and thus the existence of two state variables was assumed in the subsequent step. **Table 3** shows the result of the evaluation. The column *ct* is the computation time to complete the search. The computation times required for RRH, RI and RIH were longer than that of RR, because the variety of admissible formulae containing Interval scale variables is far larger than that of Ratio scale variables. The result is marked by '++' in the table where the formula having the correct shape is top ranked in the accuracy. If the formula having the correct shape is derived within the top five solutions, it is marked by '+', and otherwise it is marked by '-'. **Table 4** shows the top 5 results in case of RR, where the relative noise level is 0.1%. The first, the third and the fifth results in terms of MSE have the correct formulation, and the coefficients are very close to those of the original. The tables show that at least 2.0% measurement noise in relative amplitude is acceptable to discover the correct formulae, if all state variables can be measured. Even if a hidden state variable exists, SCALETRACK can discover correct equation in case that the mea-

Table 4 The top 5 results in case of RR.

	state equation	MSE
1	$dx_1/dt = 0.943x_1x_2$ $dx_2/dt = -0.552x_2$	7.50×10^{-7}
2	$dx_1/dt = 1.00x_1x_2$ $dx_2/dt = 0.617x_1$	7.79×10^{-7}
3	$dx_1/dt = 1.07x_1x_2$ $dx_2/dt = -0.424x_2$	7.80×10^{-7}
4	$dx_1/dt = 1.06x_1x_2$ $dx_2/dt = 0.647x_1$	7.95×10^{-7}
5	$dx_1/dt = 1.02x_1x_2$ $dx_2/dt = -0.475x_2$	7.96×10^{-7}

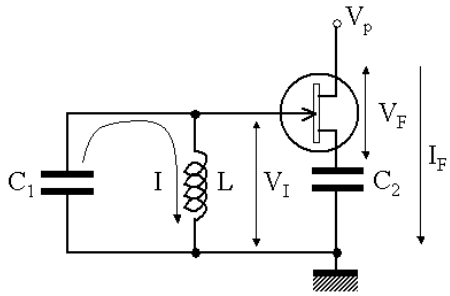

Fig. 5 An LC and FET circuit.

surement noise is 0.1–0.5%.

4.3 Discovery of Circuit Dynamics

SCALETRACK has been applied to synthetic data of an electric circuit consisting of LCs and a Field Effect Transistor (FET) as shown in **Fig. 5**. Its state equation is represented as follows.

$$\begin{aligned} \dot{V}_I(t) &= -\frac{I(t)}{C_1} &&= -100I(t), \\ \dot{I}(t) &= \frac{V_I(t)}{L} &&= 50V_I(t), \\ \dot{V}_F(t) &= \frac{V_I(t)V_F(t)}{rC_2} &&= 250.0V_I(t)V_F(t), \end{aligned}$$

where the definitions of V_I, I, V_F, $L = 20$mH, $C_1 = 10$mF and $C_2 = 1$mF are clear in the figure and $r = 4.0\Omega$V a voltage–resistance coefficient of FET. All state variables are Ratio scale, and can be measured via corresponding Ratio scale measurement variables respectively. The measurement data were sampled under one time step $\Delta t = 0.001$, total time steps $n = 800$ and the relative measurement noise $\sigma_w = 0.1\%$. Because $\nu_{max}(8) = 2.94$ was obtained in the correlation dimension analysis, the state equation consisting of three state variables was searched.

In case that every state variables are directly measured, the following state equation having the best accuracy was derived.

$$\begin{aligned} \dot{V}_I(t) &= -133.3I(t), \\ \dot{I}(t) &= 6.94V_I(t)V_F(t), \\ \dot{V}_F(t) &= 249.0V_I(t)V_F(t). \end{aligned}$$

The shapes of the first and the third expressions of the equation are identical with those in the original equation though the values of coefficients are moderately different from the original.

Subsequently, the measurement of I was omitted to make I a hidden state variable. The following correct formula except for the discrepancy of coefficient values showed up within the solutions having top five accuracies.

$$\begin{aligned} \dot{V}_I(t) &= -26.9I(t), \\ \dot{I}(t) &= 298.0V_I(t), \\ \dot{V}_F(t) &= 250.0V_I(t)V_F(t). \end{aligned}$$

These results indicate that SCALETRACK has ability to discover state equations of engineering objects having three-dimensional dynamics at least.

5. Discussion

In this paper, we proposed a method named SCALETRACK which discovers the first principle based dynamic models of an objective process represented by simultaneous time differential equations. According to the experiments, SCALETRACK has an ability to discover state equations even if hidden states exist. SCALETRACK accepts at least 2.0% measurement noise in relative amplitude when hidden states do not exist. This is comparable with the noise level in practical cases where 1.0–2.0% measurement noise is the most widely seen in scientific and engineering applications. Even when a hidden state exists, 0.5% measurement noise in relative amplitude can be accepted by SCALETRACK. This noise level can be also achieved by using proper measurement facilities in many applications. The performance of SCALETRACK shows robustness against measurement noise to some extent.

We have seeked any other approaches which have comparable functions of the law equation discovery with SCALETRACK. However, as discussed earlier, all the conventional approaches including LAGRAMGE and Process Modeling can not address the class of problem to discover the dynamic and simultaneous time differential law equations without using any domain specific knowledge base. Accordingly, the experimental comparison of SCALE-

TRACK with the other approaches has not been conducted in this paper.

Computational complexity of SCALETRACK is NP–hard in terms of the number of state variables, because the number of possible combinations of scale-types, the number of candidate state equations and the number of the possible values of the coefficients to be searched shows combinational explosions when the number of state variables increases. This fact is reflected to the computation time required by SCALETRACK, where it takes over 3 days to complete the search of the solutions having three state variables of Ratio scales. Although the computational time can be reduced by setting less trials to search beta vectors' neighbors, the correctness of the solutions is also reduced. More efficient search algorithm should be studied in future work.

Another issue remained in this work is the noise robustness. This problem is also very important to establish wilder applicability of SCALETRACK to noisy situations. There are two major issues on the noise robustness. First is the noise characteristic of MSE resulted in SIS/RMC filter tracking. Because the SIS/RMC filter tries good estimation of state variables even under erroneous models and noise, the MSE between the real and the estimated measurements is not very sensitive to the noise. Though this is a desirable characteristic for state estimation, many spurious models which provide good MSE appears in the model search, and the spurious models remove the correct model from the search space. Second issue is the accuracy of the correlation dimension analysis. Some past studies reported that the analysis does not provide a correct dimension under noisy and high state dimensions [6),10)]. Because only the cases having low state dimensions are accessed in this paper, this problem does not appear significantly. However, a more robust approach to estimate accurate state dimensions for large scale problems should be explored in future.

The advantage of SCALETRACK is that the discovered equations are guaranteed that they are the first principle based equations because the candidates generated in SCALETRACK are constrained by the scale-types of variables and Extended Product Theorem. Scientists can easily avoid the solutions not reflecting the underlying the first principles by using this method.

6. Conclusion

We showed a novel method to discover simultaneous time differential law equations representing the first principles governing dynamic behavior of an objective process from passively observed data. Significant advantages of this approach are the reduction of the strong bias introduced by the domain knowledge specific to the objective process and the wide applicability to the cases including hidden state variables in the objective process. The remained major issues are to overcome the computational complexity of the search and the problem of noise robustness. The study to overcome these issues is currently underway.

Acknowledgments This research has conducted under the support of JSPS Kiban Kenkyuu (B) (2) 16300045.

References

1) Akaike, H.: Autoregressive model fitting for control, *Ann. Inst. Statist. Math.*, Vol.23, pp.163–180 (1971).
2) Berge, P., Vidal, C. and Pomeau, Y.: *Order within chaos*, Sangyo-tosho (1992). (In Japanese) Translated by Aizawa, Y.
3) Doucet, A., Godsill, S. and Andrieu, C.: On Sequential Monte Carlo sampling methods for Bayesian filtering, *Statistics and Computing*, Vol.10, pp.197–208 (2000).
4) Dzeroski, S. and Todorovski, L.: Discovering Dynamics: From Inductive Logic Programming to Machine Discovery, *Journal of Intelligent Information Systems*, Vol.4, pp.89–108 (1995).
5) Falkenhainer, B.C. and Michalski, R.S.: Integrating Quantitative and Qualitative Discovery: The ABACUS System, *Machine Learning*, Vol.1, pp.367–401 (1986).
6) Fueda, K. and Yanagawa, T.: Estimating the Embedding Dimension and Delay Time from Chaotic Time Series with Dynamic Noise, *Journal of the Japan Statistical Society*, Vol.31, No.1, pp.27–38 (2001).
7) Haykin, S.: *Kalman Filtering and Neural Networks*, John Wiley & Sons, Inc., Hoboken, New Jersey (2001).
8) Iwasaki, Y. and Simon, H.A.: Causality in Device Behavior, *Artificial Intelligence*, Vol.29, pp.3–32 (1986).
9) Kalman, R.E.: A New Approach to Linear Filtering and Prediction Problems, *Transaction of ASME—Journal of Basic Engineering*, pp.35–45 (1960).
10) Kawaguchi, A., Yonemoto, K. and Yanagawa, T.: Estimating the Correct Dimension from

a Chaotic System with Dynamic Noise, *MHF Preprint Series* (2004). http://www.math.kyushu-u.ac.jp/coe/report/pdf/2004-4.pdf
11) Koehn, B. and Zytkow, J.M.: Experimenting and theorizing in theory formation, *Proc. International Symposium on Methodologies for Intelligent Systems*, pp.296–307 (1986).
12) Kokar, M.M.: Determining Arguments of Invariant Functional Descriptions, *Machine Learning*, Vol.1, No.4, pp.403–422 (1986).
13) Langley, P., Sanchez, J. and S., D.: Inductive Process Models from Continuous Data, *Proc. 19th International Conference on Machine Learning* (*ICML*), pp.347–354 (2002).
14) Langley, P., Simon, H.A., Bradshaw, G.L. and Zytkow, J.M.: *Scientific discovery; Computational explorations of the creative process*, Canbridge, MA, MIT Press (1987).
15) Ljung, L.: *System Identification*, PTR Prentice-Hall (1986).
16) Murota, K.: *Systems Analysis by Graphs and Matroids — Structural Solvability and Controllability*, Algorithms and Combinatorics, Vol.3, Springer Verlag (1987).
17) Nordhausen, B. and Langley, P.: *An Integrated Approach to Empirical Discovery, Computational Models of Scientific Discovery and Theory Formulation*, Morgan Kaufman Publishers Inc., San Mateo, California (1990).
18) Roweis, S. and Ghahramani, Z.: Learning Nonlinear Dynamical Systems using the EM Algorithm, *KalmanFiltering and Neural Networks*, pp.175–220 (2001).
19) Shumway, R. and Stoffer, D.: An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm, *Journal of Time Series Analysis*, Vol.3, No.4, pp.253–264 (1982).
20) Takens, F.: Detecting strange attractors in turbulence, *Dynamical Systems and turbulence*, Vol.898, pp.366–382 (1981).
21) Todorovski, L. and Dzeroski, S.: Declarative bias in equation discovery, *Proc. Fourteenth International Conference on Machine Learning*, San Mateo, California, Morgan Kaufmann, pp.376–384 (1997).
22) Washio, T. and Motoda, H.: Discovering Admissible Models of Complex Systems Based on scale-types and Identity Constraints, *Proc. Fifteenth International Jouint Conference on Artificial Intelligence*, Nagoya, Japan, pp.810–817 (1997).
23) Washio, T. and Motoda, H.: Discovery of First Principle Based on Data-Driven Reasoning, *Proc. First Pacific-Asia Conference on Knowldge Discovery and Data Mining*, pp.169–182 (1997).
24) Washio, T. and Motoda, H.: Discovering Admissible Simultaneous Equations of Large Scale Systems, *Proc Fifteenth National Conference on Artificial Intelligence*, Madison, Wisconsin, pp.189–196 (1998).
25) Washio, T., Motoda, H. and Niwa, Y.: Discovering admissible model equations from observed data based on scale-types and identity constraints, *Proc. Sixteenth International Joint Conference on Artificial Intelligence*, Stockholm, Sweden (1999).
26) Washio, T., Motoda, H. and Niwa, Y.: Discovering admissible simultaneous equation models from observed data, *Proc. twelfth European conference on machine learning*, Freiburg, Germany (2001).
27) Wasserman, P.D.: *Newral Computing: Theory and Practice*, New York: Van Nostrand Reinhold (1989).

(Received April 5, 2005)
(Revised May 30, 2005)
(Accepted July 9, 2005)
(Released November 1, 2006)
(Paper version of this article can be found in the IPSJ Transactions on Mathematical Modeling and Its Applications, Vol.47 No.SIG14(TOM15), pp.31–42.)

Fuminori Adachi received his Master of engineering from Osaka University in 2003. He is enrolled in the doctoral course of Osaka University from 2003. His current research interest includes scientific discovery.

Takashi Washio received his Ph.D. from Tohoku University in 1988. In 1988, he became a visiting reseacher in Massachusetts Institute of Technology. In 1990, he joined Mitsubishi Research Institute Inc., and is working for Osaka University from 1996. His current research interests include scientific discovery, data mining and machine learning techniques.

Hiroshi Motoda received his Ph.D. from the University of Tokyo in 1972. In 1967, he joined Hitachi Ltd, and has been working for Osaka University since 1996. His current research interest includes scientific discovery, data mining and machine learning.