SCALETRACK: A System to Discover Dynamic Law Equations Containing Hidden States and Chaos

Takashi Washio, Fuminori Adachi, and Hiroshi Motoda

I.S.I.R., Osaka University, 8-1, Mihogaoka, Ibaraki City, Osaka, 567-0047, Japan
washio@ar.sanken.osaka-u.ac.jp

Abstract. This paper proposes a novel system to discover simultaneous time differential law equations reflecting first principles underlying objective processes. The system has the power to discover equations containing hidden state variables and/or representing chaotic dynamics without using any detailed domain knowledge. These tasks have not been addressed in any mathematical and engineering domains in spite of their essential importance. Its promising performance is demonstrated through applications to both mathematical and engineering examples.

1 Introduction

A set of well known pioneering approaches of scientific law equation discovery is called BACON family [1]. They try to figure out a static equation on multiple quantities over a wide state range under a given laboratory experiment where quantities are actively controlled. Their drawback is the low likelihood to discover the law equations, since they do not use certain essential criteria to capture relations induced by the first principles. A law equation reflecting the first principle here is an observable, reproducible and concise relation satisfying generality, soundness and mathematical admissibility. The generality is to be widely observed in the objective domain of the equation, the soundness not to conflict with any observations and the mathematical admissibility to follow some constraints deduced from the invariance of the relation under various times, places and measurement expressions. Especially, the mathematical admissibility can be used to narrow down the equation formulae for the search. Some systems introduced unit dimension constraints and "*scale-type constraints*" to limit the search space to mathematically admissible equations [2,3,4]. Especially, the scale-type constraints have wide applicability since they do not need unit information of quantities. LAGRANGE addressed the discovery of "*simultaneous time differential law equations*" reflecting the dynamics of objective processes under "*passive observations*" where none of quantities are experimentally controllable [5]. Its extended version called LAGRAMGE introduced domain knowledge of the objective process to limit the search space within plausible law equations [6]. IPM having similar functions with LAGRAMGE further identified plausible law equations containing "*hidden state variables*" when the variables are known in

A. Hoffmann, H. Motoda, and T. Scheffer (Eds.): DS 2005, LNAI 3735, pp. 253–266, 2005.

the detailed domain knowledge [7]. PRET identified "*chaotic dynamics*" under similar conditions where very rich domain knowledge is available [8].

However, scientists and engineers can develop good models of the objective dynamics without using the discovery systems in many practical cases when detailed domain knowledge is available. Accordingly, the main applications of the discovery systems are to identify simultaneous time differential equations reflecting the first principles under passive observation and "*little domain knowledge*." One of such important applications is the discovery of "*hidden state variables*." In many problems, some state variables are not directly observed, and even the number of unobserved state variables is not known. Another important issue is the analysis of the observed data representing "*chaotic dynamics*." If the detailed domain knowledge on the dynamics underlying the chaos is given, some of the aforementioned systems can construct the dynamic equations appropriately representing the chaos. However, scientists can hardly grasp the dynamic laws on many chaotic behaviors based on their domain knowledge, since the background mechanisms of the chaos are usually very complex [9].

In this paper, we propose a novel scientific equation discovery system called SCALETRACK (SCALE-types and state TRACKing based discovery system) to discover a model of an objective process under the following requirements.

(1) The model is simultaneous time differential equations representing the dynamics of an objective process.
(2) The model is not an approximation but a plausible candidate to represent the underlying first principles.
(3) The model is discovered from passively observed data without using domain knowledge specific to the objective process.
(4) The model can include hidden state variables.
(5) The model can represent chaotic dynamics.

2 Outline

2.1 Basic Problem Setting

We adopt the following "*state space model*" of objective dynamics and measurement without loss of generality.

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t)) + \boldsymbol{v}(t) \quad (\boldsymbol{v}(t) \sim N(0, \boldsymbol{\Sigma}_v)), \text{ and} \tag{1}$$
$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{w}(t) \quad (\boldsymbol{w}(t) \sim N(0, \boldsymbol{\Sigma}_w)), \tag{2}$$

where the first equation is called a "*state equation*" and the second a "*measurement equation*." $\boldsymbol{x}$ is called a state vector, $\boldsymbol{f(x)}$ a system 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. $\boldsymbol{f(x)}$, a model of the objective dynamics over its wide state range, is not limited to linear formulae in general, and any state transition of $\boldsymbol{x}$ can be represented by this formulation. $\boldsymbol{C}$, the model of measurement, is represented by a linear transformation matrix, because the measurement facilities are artificial and linear in most cases, and some state variables in $\boldsymbol{x}$ are often observed indirectly as their linear combinations through measurement variables in $\boldsymbol{y}$. If $\boldsymbol{C}$ is column full rank, the values of all state variables with the measurement noise are estimated by solving the measurement equation with $\boldsymbol{x}$. Otherwise, some state variables are not estimated within the measurement equation, and these variables are called "*hidden state variables*."

In the scientific law equation discovery, $\boldsymbol{f}(\boldsymbol{x})$ is initially unknown, and even $\boldsymbol{x}$ is not known correctly. Only a state subvector $\boldsymbol{x}'(\subseteq \boldsymbol{x})$ and a submatrix $\boldsymbol{C}'(\subseteq \boldsymbol{C})$ representing an artificial measurement facility are initially known to relate $\boldsymbol{x}'$ with $\boldsymbol{y}$ as $\boldsymbol{y} = \boldsymbol{C}'\boldsymbol{x}'$. To derive $\boldsymbol{C}$ from $\boldsymbol{C}'$, the number of missing state variables, *i.e.*, the difference between the dimensions of $\boldsymbol{x}$ and $\boldsymbol{x}'$, must be estimated. Thus, SCALETRACK identifies the number of elements in $\boldsymbol{x}$ including hidden state variables based on passively observed data at first. Then, it searches plausible candidates of $\boldsymbol{f}(\boldsymbol{x})$ reflecting the first principles from the data.

2.2 Entire Approach

The entire approach of SCALETRACK is outlined in Figure 1. Given a set of measurement data, the dimension of $\boldsymbol{x}$ is identified through a statistical analysis called "*correlation dimension analysis*." Once the dimension is known, all possible combinations of scale-types of the elements in $\boldsymbol{x}$ are enumerated based on scale-type constraints, the known measurement submatrix $\boldsymbol{C}'$ and the known scale-types of the elements in $\boldsymbol{y}$. Then, for every combination, the candidate formulae of a state equation admissible to the scale-type constraints are generated. Subsequently, through a set of state tracking simulations called "SIS/RMC filter" combined with parameter search on the given measurement data, the parameter values in every candidate formula are estimated. Finally, some candidates providing highly accurate tracking in terms of "*Mean Square Error* (MSE)" are selected as the discovered dynamic models of the objective process. The details of each step in Figure 1 are described in the following section.

3 Methods

3.1 Estimating Dimension of x

"*Correlation dimension analysis*" estimates the dimension of $\boldsymbol{x}$, $dim(\boldsymbol{x})$, from given measurement data $\boldsymbol{y}$ over n sampling time steps [9]. Given an element y_h $(h = 1, .., dim(\boldsymbol{y}))$ of $\boldsymbol{y}$, let τ_h be the minimum time step lag that the time lagged autocorrelation of $y_h(t)$ becomes 0 as follows.

$$\tau_h = argmin_{\tau \in [1,n]} \{ \frac{1}{n} \sum_{t=1}^{n-\tau} (y_h(t) - \bar{y}_h)(y_h(t+\tau) - \bar{y}_h) \simeq 0 \}, \quad (3)$$

where $\bar{y}_h$ is the time average of $y_h(t)$ over $[1, n]$. τ_h is the time steps within that the local dependency among the observed states is vanished. Then the following time lagged vectors of length m are constructed from y_h.

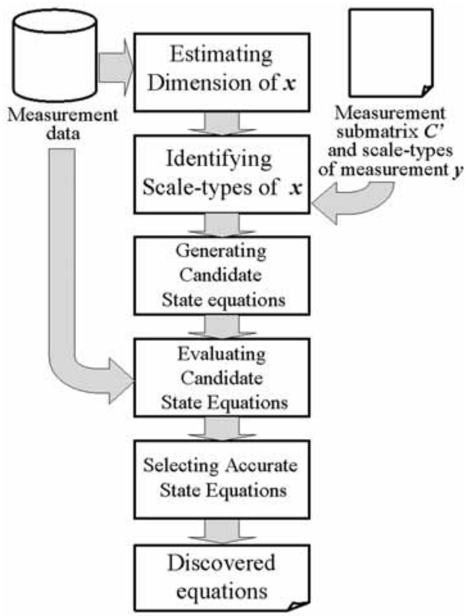

Fig. 1. Outline of SCALETRACK

$$Y_h^m(1) = [y_h(1), y_h(1+\tau_h), \ldots, y_h(1+(m-1)\tau_h)]$$
$$\cdots\cdots\cdots$$
$$Y_h^m(n-(m-1)\tau_h) = [y_h(n-(m-1)\tau_h), y_h(n-(m-2)\tau_h), \ldots, y_h(n)]$$

If m is sufficiently large, each of these vectors reflects a global relation among the states, since the time intervals among the elements in a vector are equal to or longer than τ_h. Then the following correlation integral in the time lagged phase space is calculated.

$$R_h^m(r) = \frac{2}{n'(n'-1)}[\text{number of } (i,j)\text{s; } \Delta Y_h^m(i,j) < r], \tag{4}$$

where $n' = n-(m-1)$, $1 \le i,j \le n'$ and $\Delta Y_h^m(i,j) = |Y_h^m(i) - Y_h^m(j)|$. $R_h^m(r)$ represents the density of states in the space, and shows the following power law relation in general over the range of r covering the state distribution.

$$R_h^m(r) \propto r^{\nu_h(m)}, \tag{5}$$

where $\nu_h(m)$ is called a "*correlation exponent.*" Theoretically it is an approximation of the fractal dimension of the global state distribution which is equivalent to $dim(\boldsymbol{x})$ under the condition of $m \ge 2dim(\boldsymbol{x})+1$. $dim(\boldsymbol{x})$ is estimated through the least square fitting of Eq.(5) to $R_h^m(r)$s derived by Eq.(4) under a sufficiently large m. $\nu_h(m)$ is computed for each y_h $(h = 1, .., dim(\boldsymbol{y}))$, and the nearest integer of its maximum, $\nu_{max}(m)$, among them is used for $dim(\boldsymbol{x})$, since some measurement variables may miss the behaviors of some state variables.

3.2 Identifying Scale-Types of $\boldsymbol{x}$

Once $dim(\boldsymbol{x})$ is known, the "*scale-type*" of each element of $\boldsymbol{x}$ is identified for the candidate $f(\boldsymbol{x})$ generation in the next step. This is done based on "*scale-type constraints*" [10] and the scale-types of elements of $\boldsymbol{y}$. Representatives of quantitative scale-types are ratio scale and interval scale. Examples of the ratio scale quantities are physical mass and absolute temperature where each has an absolute origin. The admissible unit conversion of the ratio scale follows $x' = \alpha x$. Examples of the interval scale quantities are temperature in Celsius and sound pitch where the origins of their scales are not absolute and arbitrary changed by human's definitions. The admissible unit conversion of the interval scale follows $x' = \alpha x + \beta$. Though the scale-type is strongly related with the unit dimension, they are different each other.

As noted in the previous section, only a state subvector $\boldsymbol{x'}(\subseteq \boldsymbol{x})$ is measured by $\boldsymbol{y}$ through a measurement facility $\boldsymbol{C'}(\subseteq \boldsymbol{C})$ as $\boldsymbol{y} = \boldsymbol{C'x'}$. Because the structure of the facility is independent of the units of the elements of $\boldsymbol{x'}$ and $\boldsymbol{y}$, $\boldsymbol{C'}$ is invariant against the change of their units. Then the following theorem holds.

Linear Formula Theorem. *Let $\boldsymbol{x'}$ be a known state subvector of $\boldsymbol{x}$, y_h an element of a measurement vector $\boldsymbol{y}$ and $\boldsymbol{x'_h}$ a state subvector of $\boldsymbol{x'}$ where each $x_i \in \boldsymbol{x'_h}$ has a nonzero (h,i)-element, c_{hi}, in the known measurement submatrix $\boldsymbol{C'}$. The scale-types of x_is in $\boldsymbol{x'_h}$ are constrained by the scale-type of y_h and the following rules.*

(1) If y_h is a ratio scale, all x_is are ratio scales, or more than one x_i are interval scales and the rest ratio scales.

(2) If y_h is an interval scale, one x_i at least is an interval scale and the rest ratio scales.

Proof. Because of the relation $\boldsymbol{y} = \boldsymbol{C'x'}$, $y_h = \sum_{x_i \in \boldsymbol{x'_h}} c_{hi}x_i$ holds. Let the set of interval scale quantities in $\boldsymbol{x'_h}$ be I_h. Every $x_i \in I_h$ follows the admissible unit conversion $x'_i = \alpha_i x_i + \beta_i$, and every x_i in the rest, *i.e.*, ratio scales, follows $x'_i = \alpha_i x_i$. When y_h is a ratio scale, it follows $y'_h = \alpha y_h$. Because of the invariance of $\boldsymbol{C'}$, $y'_h = \sum_{x'_i \in \boldsymbol{x'_h}} c_{hi}x'_i$ holds. By substituting the admissible unit conversions and $y_h = \sum_{x_i \in \boldsymbol{x'_h}} c_{hi}x_i$ to this linear relation, the following is obtained.

$$\sum_{x_i \in \boldsymbol{x'_h}} \alpha c_{hi} x_i = \sum_{x_i \in \boldsymbol{x'_h}} c_{hi}\alpha_i x_i + \sum_{x_i \in I_h} c_{hi}\beta_i$$

Because this is an identity equation for every $x_i \in \boldsymbol{x'_h}$, $\alpha_i = \alpha$ for every x_i and $\sum_{x_i \in I_h} c_{hi}\beta_i = 0$ hold. If I_h is empty, the last relation trivially holds. If I_h has only a unique x_i, $\beta_i = 0$ must hold, and this is contradictory to the interval scale $x_i \in I_h$. If I_h has more than one x_i, the last relation can hold for non-zero β_is while β_is are mutually dependent in the relation. This concludes the rule (1). When y_h is an interval scale, it follows $y'_h = \alpha y_h + \beta$. Through the similar discussion with the rule (1), $\sum_{x_i \in I_h} c_{hi}\beta_i = \beta$ hold. If I_h is empty, $\beta = 0$ must hold, and this is contradictory to the interval scale y_h. If I_h is not empty, this relation can hold for non-zero β_is and β while they are mutually dependent in the relation. This concludes the rule (2). ∎

Based on this theorem and the scale-types of all $y_h \in \boldsymbol{y}$, a set of constraints on the scale-types of all $x_i \in \boldsymbol{x'}$ is obtained. Because the scale-types of all $x_i \in \boldsymbol{x}$ which are not in $\boldsymbol{x'}$ are unknown, they can be either ratio or interval scale. Then, every admissible combination (R_x, I_x) where R_x is a set of ratio scale state variables and I_x a set of interval scale state variables in $\boldsymbol{x}$ satisfying these constraints are enumerated by using a simple search. Though this search is combinatorial, it is tractable in practice as far as the dimension of $\boldsymbol{x}$ is not very large.

3.3 Generating Candidate State Equations

"*Extended Product Theorem*" [4] provides a basis of the candidate generation of state equations. This theorem comes from the invariance of the formula shape against the unit conversions and the scale-type constraints similarly to the aforementioned Linear Formula Theorem, and it has been used in several law equation discovery systems in the past. The following is the theorem where some notions are adapted to our descriptions.

Extended Product Theorem. *Given a combination* (R_x, I_x) *for* $\boldsymbol{x}$*, the state variables have the following relation.*

$$\dot{x}_i = \prod_{x_j \in R_x} |x_j|^{\alpha_j} \prod_{I_k \subseteq (I_x - I_g)} (\sum_{x_j \in I_k} \beta_{kj}|x_j| + \beta_k)^{\alpha_k} \prod_{x_j \in I_g} \exp(\beta_{gj}|x_j|)$$

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

In a state equation $\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t))$, all elements in $\dot{\boldsymbol{x}}$ are ratio scales, since the time derivative of an element of $\boldsymbol{x}$ is the difference of two states divided by a time interval in essence. The formulae following this theorem are called "*regimes*" having the invariance against the unit conversions. Since this is a required character of the formulae to represent the first principles, the candidates have high plausibility to be law equations. Under ratio scale time derivatives $\dot{\boldsymbol{x}}$ and a given combination (R_x, I_x), the multiple candidates of a state equation are enumerated based on this theorem. The set of combinations of (R_x, I_x) derived in the previous step provides a set of many candidate state equations, CSE.

3.4 Evaluating Candidate State Equations

Once CSE for an objective process is provided in the previous step, a fitting error $E(c)$ of every candidate $c \in CSE$ under given measurement data is evaluated through adjustment of its coefficients and state tracking.

Searching for Power Coefficients. As shown in Extended Product Theorem, the formulae of the state equations have two types of constants, *i.e.*, power coefficients αs and proportional coefficients βs. The search space of a power

Table 1. Nonlinearity of T^α.

case of α		range of T	nonlinearlity
range	parity		
$\alpha > 1$	α is even.	$T \geq 0$	monotonic increase
		$T < 0$	monotonic decrease
$0 < \alpha < 1$	$1/\alpha$ is even.	$T \geq 0$	monotonic increase
		$T < 0$	not admissible
$\alpha > 0$	α or $1/\alpha$ is odd.	$T \geq 0$	monotonic increase
		$T < 0$	monotonic increase
$\alpha = 0$			1
$\alpha < 0$	α or $1/\alpha$ is odd.	$T \geq 0$	monotonic decrease
		$T < 0$	monotonic decrease
$-1 < \alpha < 0$	$1/\alpha$ is even.	$T \geq 0$	monotonic decrease
		$T < 0$	not admissible
$\alpha < -1$	α is even.	$T \geq 0$	monotonic decrease
		$T < 0$	monotonic increase

where $T \to 0 \Rightarrow T^\alpha \to 0(\alpha > 0)$ and $T^\alpha \to \pm\infty(\alpha < 0)$.

coefficient α is limited to small integers, within $[-5, 5]$ for instance, and their inverses. This is because the power coefficients reflect the dimensions of space and units where the objective process operates, and their complexities are limited. Moreover, given a term T having a power coefficient α in the formulae of Extended Product Theorem, the range and the parity of α strongly affect the nonlinearity of T^α as shown in Table 1. Because of these discrete characteristics of α, the standard approaches for continuous and nonlinear optimization such as gradient descent method are not applicable. Instead, for every combination of the cases over all αs appearing in the candidate c, a monotonic and discrete search on integer αs is applied to reduce the fitting error $E(c)$. Because the number of αs in c is not very large, this part does not cause severe combinatorial explosion.

Searching for Proportional Coefficients. The search of the proportional coefficients βs minimizing $E(c)$ under every combination of αs provided in abovementioned scheme is performed. We experienced in our preliminary study that the standard nonlinear optimization of βs such as gradient descent method again does not converge to their right values within tractable time, because the influence of some βs to $\dot{x}_i$ can be very small under some αs. Accordingly, the following Golden Ratio Search [11] which is a well-known opportunistic line search without using the quantitative gradient information has been applied to βs. Under a combination of values of all αs and a combination of default values of all βs appearing in c, given initial upper bound β_j^u and lower bound β_j^l of β_j in c, $E(c)$s are evaluated on the following β_j^1 and β_j^2 by the state tracking which will be described shortly.

$$\beta_j^1 = \beta_j^l + r(\beta_j^u - \beta_j^l), \text{ and} \tag{6}$$

$$\beta_j^2 = \beta_j^u - r(\beta_j^u - \beta_j^l), \tag{7}$$

where $r = (3 - \sqrt{5})/2$ is the golden ratio. Let $E(c)$s evaluated on β_j^1 and β_j^2 be $E(c|\beta_j^1)$ and $E(c|\beta_j^2)$ respectively. If $E(c|\beta_j^1) \geq E(c|\beta_j^2)$ then $\beta_j^1 \rightarrow \beta_j^l$, $\beta_j^2 \rightarrow \beta_j^1$ and calculate new β_j^2 by Eq.(7), else $\beta_j^2 \rightarrow \beta_j^u$, $\beta_j^1 \rightarrow \beta_j^2$ and calculate new β_j^1 by Eq.(6). This rule is applied iteratively until $|\beta_j^2 - \beta_j^1|$ becomes less than a threshold ϵ. After this convergence, the converged value becomes a new default value of β_j. Subsequently, another β in c is selected in place of β_j, and this Golden Ratio Search is repeated until the default values of all βs in c becomes stable. Finally, the estimated βs are rounded off to integers when the values are close enough to the integers within the statistically expected estimation errors, since the parameters tend to be integers in many physical processes. After obtaining values of all βs for each combination of values of all αs, the unique combination of values of all αs, A_c, and that of all βs, B_c, providing the minimum $E(c)$ is chosen to be the coefficients of c.

State Tracking. Given a time series of measurement vector $\boldsymbol{y}(t)$s, a candidate state equation c and its $A_c \cup B_c$, the fitting error $E(c)$ is evaluated through state tracking. The recent massive increase in computational power became to allow the introduction of direct and sequential Monte Carlo integration of the state probability distributions within Bayesian framework. This approach is called "*Sequential Importance Sampling/Resampling Monte Carlo filter* (SIS/RMC filter)" [12], and can track the states generated in c without introducing any essential approximation. This state tracking has many advantages comparing with the other nonlinear state tracking approaches such as the conventional Extended Kalman Filter [13] and the qualitative reasoning based PRET [8]. The former using the linearization of the state equations does not work well when the equations include some singular points and/or some state regions having strong sensitivity to the tracking error. The latter faces a combinatorial explosion of qualitative states when the dimension and/or the complexity of the state space structure are high. In contrast, SIS/RMC filter does not require any approximation to be spoiled by the singularity and the strong nonlinearity, and does not face the combinatorial explosion of the states to be considered.

Because of the space limit, readers should refer the literature [12] to learn the background theory of SIS/RMC filter. In this paper, only the procedure of the state tracking adapted to our basic problem setting is indicated. The SIS/RMC filter is represented by the following procedures where the probabilities $p(\boldsymbol{x}(t)|\boldsymbol{x}(t-1), \boldsymbol{y}(t))$ and $p(\boldsymbol{y}(t)|\boldsymbol{x}(t-1))$ are defined by $\boldsymbol{y}(t)$, c and its $A_c \cup B_c$.

1 Importance sampling

(1-1) For $i = 1, ..., N$, sample $\tilde{\boldsymbol{x}}^{(i)}(t) \sim p(\boldsymbol{x}(t)|\boldsymbol{x}^{(i)}(t-1), \boldsymbol{y}(t))$.

(1-2) For $i = 1, ..., N$, evaluate the importance weights:
$w^{*(i)}(t) = w^{*(i)}(t-1)p(\boldsymbol{y}(t)|\boldsymbol{x}^{(i)}(t-1))$.

(1-3) For $i = 1, ..., N$, normalize the importance weights: $\tilde{w}^{(i)}(t) = \frac{w^{*(i)}(t)}{\sum_{j=1}^{N} w^{*(j)}(t)}$.

(1-4) Let MAP estimation, $\tilde{\boldsymbol{x}}(t)$, be $\tilde{\boldsymbol{x}}^{(i)}(t)$ having the maximum $\tilde{w}^{(i)}(t)$.

(1-5) $N_{eff} = \frac{1}{\sum_{i=1}^{N} (\tilde{w}_k^{(i)})^2}$.

(1-6) If $N_{eff} \geq N_{thres}$ then $\boldsymbol{x}^{(i)}(t) = \tilde{\boldsymbol{x}}^{(i)}(t)$ for $i = 1, ..., N$, $t = t + 1$ and go to 1. Otherwise go to 2.

2 Resampling

(2-1) Generate random integers $j(i)$ $(i = 1, ..., N)$ in proportion to the probabilities $\tilde{w}^{(l)}(t)$ $(l = 1, ..., N)$ so that l having larger $\tilde{w}^{(l)}(t)$ appears more as $j(i)$.

(2-2) $x^{(i)}(t) = \tilde{x}^{j(i)}(t)$, $w^{(i)}(t) = 1/N$ for $i = 1, ..., N$, $t = t + 1$ and go to 1.

In the importance sampling, many $\tilde{\boldsymbol{x}}^{(i)}(t)$s called "*particles*" derive "*Maximum A Posteriori* (MAP)" estimation of the state vector in concert with the normalized weight $\tilde{w}^{(i)}(t)$. An index N_{eff} monitors the ratio of probable particles having high weights. When the ratio becomes lower than a predefined threshold N_{thres}, resampling is applied to increase the probable particles.

Once the MAP estimation $\tilde{\boldsymbol{x}}(t)$s are obtained over $t = 1, ..., n$ time steps, the time series of $\tilde{\boldsymbol{y}}(t)$s $(t = 1, ..., n)$ are estimated via Eq.(2). Then the fitting error $E(c)$ can be evaluated by the following "*Mean Square Error* (MSE)."

$$E(c) = \frac{1}{n} \sum_{t=1}^{n} |\boldsymbol{y}(t) - \tilde{\boldsymbol{y}}(t)|^2$$

3.5 Selecting Accurate State Equations

The previous steps provide CSE and $< c, A_c \cup B_c, E(c) >$ for all $c \in CSE$. The solutions $< c, A_c \cup B_c, E(c) >$ having the top K accuracy, *i.e.* the K least $E(c)$s, are selected as discovered dynamic state equations in large CSE. The value of K is empirically chosen according to the complexity of the objective process and the quality of measurement data. $K = 5$ is used throughout this paper to check the variation of the search space.

4 Result

4.1 Implementation

The evaluation of candidate state equations by the SIS/RMC filter is the most time consuming step. Any search can not be skipped, since the search space is discrete and nonmonotonic. We experienced that one run of a stand alone SCALETRACK to discover a simple state equation took more than weeks even if we used an efficient algorithm. Accordingly, the current SCALETRACK introduced a simple grid computing framework using a PC cluster consisting of a control server and 10 clients, where the server has an AthlonXP1900+ 1.6GHz CPU and 2GB RAM, and each client has an AthlonXP3000+ 2.7GHz and 512MB RAM. The server computes the first three steps and then allocates the task to evaluate 10% of candidate state equations to each computer. Because this 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.

$$\left.\begin{array}{l}\dot{x}_1(t) = x_1(t)x_2(t) \\ \dot{x}_2(t) = -0.5x_1(t)\end{array}\right\} RR,$$

where $y_1 = x_1$ and $y_2 = x_2$ are ratio scale.

$$\left.\begin{array}{l}\dot{x}_1(t) = 0.4x_1(t)(x_2(t) + 0.2) \\ \dot{x}_2(t) = -0.1(x_2(t) + 0.6)\end{array}\right\} RI,$$

where $y_1 = x_1$ is ratio scale and $y_2 = x_2$ interval scale. The measurement data were generated by the simulations under one time step $\Delta t = 0.005$ and total steps $n = 600$. Empirically, m in the correlation dimension analysis and N in the state tracking were chosen to be 7 and 500 respectively. The process noise is set to be 0 to check the pure effect of the measurement noise. These settings were used in every demonstration in the rest of this paper.

Table 2. Basic Performance

case	$\nu_{max}(7)$	ct (h)	$\sigma_w(\%)$				
			0.1	0.5	1.0	2.0	5.0
RR	2.21	1.5	$+$	$\pm$	$\pm$	$\pm$	$-$
RRH	2.21	5.5	$\pm$	$\pm$	$-$	$-$	$-$
RI	2.19	4.0	$+$	$\pm$	$\pm$	$\pm$	$-$
RIH	2.19	5.5	$+$	$\pm$	$-$	$-$	$-$

ct is a required comp. time and σ_w a measurement noise level.

Table 2 shows the result of the evaluation. The case names, RR and RI, in the table correspond to the above two state equations, 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 properly estimated the dimension of state vectors as nearly 2 in each case, and thus two state variables were assumed in the subsequent steps. 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 larger than that of ratio scale variables. The result in that the formula having the correct shape is top ranked by the accuracy is marked by $+$ in the table. If the formula having the correct shape is derived within the top five solutions, it is marked by $\pm$, otherwise it is marked by $-$. The table shows that almost 2.0% relative noise is acceptable for the discovery of the correct formulae, if all state variables are measured. On the other hand, noise less than 1.0% is required to discover the correct formulae, if a hidden state variable exists. Similar results were obtained under the other n samplings more than a hundred. Since $0.5 - 2.0\%$ noise is widely seen in many scientific and engineering process measurements, the basic performance of SCALETRACK is considered to be acceptable for practical use, though further improvements on the noise robustness is needed in future study.

4.3 Discovery of Circuit Dynamics

SCALETRACK has been applied to synthetic data of an electric circuit consisting of LCs and a Field Effect Transister (FET) as shown in Figure 2 . Its state equation is represented as follows.

$$\dot{V}_I(t) = -\frac{I(t)}{C_1} = -100I(t),\ \dot{I}(t) = \frac{V_I(t)}{L} = 50V_I(t),\ \text{and}$$
$$\dot{V}_F(t) = \frac{V_I(t)V_F(t)}{rC_2} = 250V_I(t)V_F(t),$$

where the definitions of $V_I, I, V_F, L = 20mH, C_1 = 10mF$ and $C_2 = 1mF$ are clear in the figure, and $r = 4.0\Omega V$ is a voltage-resistance coefficient of the 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 relative noise level $\sigma_w = 1.0\%$. Because $\nu_{max}(7) = 2.94$ was obtained in the correlation dimension analysis, the state equations consisting of three state variables were searched.

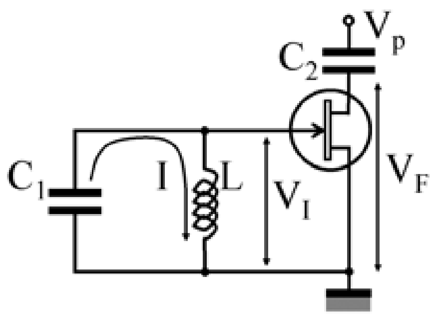

Fig. 2. An LC and FET Circuit

When every state variables are directly measured, the computation time was 18.5 hours, and the following equation having the best accuracy was derived.

$$\dot{V}_I(t) = -133.3I(t),\ \dot{I}(t) = 60.2V_I(t),\ \text{and}$$
$$\dot{V}_F(t) = 249.0V_I(t)V_F(t).$$

Though the values of coefficients are moderately different from the the originals, the entire shape of the formulae is identical. Next, the measurement of I was omitted to make I a hidden state variable. The computation time was 24 hours.

In this case, the following correct formula except the discrepancy of coefficient values showed up within the solutions having top five accuracies.

$$\dot{V}_I(t) = -26.9I(t),\ \dot{I}(t) = 298.0V_I(t),\ \text{and}$$
$$\dot{V}_F(t) = 250.0V_I(t)V_F(t).$$

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

4.4 Discovery of Chaos

The future state of a chaotic process will never be identical with its past state, and thus the state changes as if it is partially at random. Due to this nature, the state of the process gradually loses the dependency on its past state in a long term, and this makes harder to identify the dynamic equations governing the process. Nevertheless the trajectory of the state evolution is determined by the current state in the chaotic dynamics. Accordingly, the dynamic equation of the process can be discovered, if the state of the process is observed in sufficiently short sampling intervals comparing with the term length in which the state dependency dies out. Because the maximum term length of the state dependency is known by τ_h of Eq.(3) introduced in the aforementioned correlation analysis, the appropriate sampling interval can be easily known.

Under this consideration, the identification of chaotic dynamics was attempted. The state equation to be discovered is the following Altered Rossler Chaos.

$$\dot{x}_1 = -x_2 - x_3, \dot{x}_2 = x_1 + 0.36 * x_2,\ \text{and}$$
$$\dot{x}_3 = 0.01 * (x_1 - 4.5) * (x_1 + 1000 * x_3 - 4.5).$$

This has an attractor in a (x_1, x_2, x_3)-phase space as depicted in Figure 3. All state variables are interval scale, and can be measured through the corresponding interval scale measurement variables respectively. The measurement data

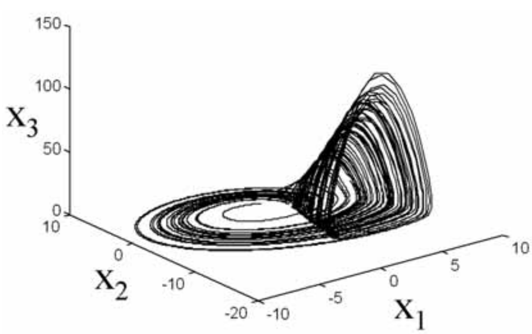

Fig. 3. An Attractor of Altered Rossler Chaos

were simulated under one time step $\Delta t = 0.001$, total time steps $n = 1500$ and relative noise level $\sigma_w = 1.0\%$. Because $\nu_{max}(7) = 3.33$ was obtained in the correlation dimension analysis, SCALETRACK searched for state equations consisting of three state variables. The computation time was 15.0 hours, and SCALETRACK resulted the following most accurate state equation. This formula has an identical shape with the original except some discrepancies of coefficients. This result indicates the high ability of SCALETRACK to discover the dynamic models of chaotic behaviors reflecting the underlying first principles.

$$\dot{x}_1 = -x_2 - x_3, \dot{x}_2 = x_1 + 0.33 * x_2, \text{ and}$$
$$\dot{x}_3 = 0.064 * (x_1 - 6.34) * (x_1 + 1002 * x_3 - 4.75).$$

5 Discussion

SCALETRACK is the first discovery system which introduced a state tracking approach in the search mechanism. In the demonstrations, the discrepancies of coefficient values are frequently observed. This may be because the particles and the weights are updated to follow the time series of $\boldsymbol{y}(t)$s at each time step in the SIS/RMC filter. This correction derives the robustness of the state tracking, but reduces the precision of the coefficient values. This may be a reason why the standard approaches for continuous and nonlinear optimization of the coefficients such as gradient descent do not perform well in the search. Another observation is that the candidate state equations top ranked by the accuracy often have formulae shapes different from the originals. This also may be due to the robustness of the state tacking against the modeling error in addition to the existence of many local minima of the accuracy in the nonlinear search space. Accordingly, the performance improvement of SCALETRACK is expected by introducing less robust state tracking, and this is a future research topic.

Another remaining issue is the current limitation of the search space. The search space of SCALETRACK is currently limited to a class of equation formulae called "*regime*"s specified by Extended Product Theorem. Although this class captures ample law equation formulae, another class of dynamic equations called "*ensemble*"s which are coupled with dimensionless variables are known not to be covered by this class. Further extension of criteria and algorithm for the search must be introduced in future while maintaining the tractability of the computation.

Introducing further valid constraints to narrow down the formulae within the law equations may enhance the plausibility of the discovered equations while reducing the search space. One of the candidate constraints is the relational templates representing conservation and flow of entities and interactions similar to Bond-Graph approach [14]. Though this type of constraints significantly contributes to the plausibility and the search space reduction in some domains including physics, they may not be applied to the wider domains such as economy and psychology where these templates do not hold, and thus the discovery systems become domain dependent. Introduction of new search constraints must be explored by carefully considering both the domain dependency and the efficiency.

6 Conclusion

SCALETRACK achieved three advantages which has not been addressed in any past work of mathematics, physics and engineering not limited to scientific discovery. The first is the discovery of first principle based simultaneous time differential equations without using detailed domain knowledge. The second is the discovery of hidden state variables. The third is the discovery of chaotic dynamics. These advantages are essentially important in many scientific and engineering fields due to the wide existence of such dynamics in nature.

References

1. Langley, P.W., Simon, H.A., Bradshaw, G.L., Zytkow, J.M.: Scientific Discovery; Computational Explorations of the Creative Process. MIT Press, Cambridge, Massachusetts (1987)
2. Koehn, B.W., Zytkow., J.M.: Experimeting and theorizing in theory formation. In: Proceedings of the International Symposium on Methodologies for Intelligent Systems, Knoxville, Tennessee, ACM SIGART Press. (1986) 296–307
3. Falkenhainer, B.C., Michalski, R.S.: Integrating quantitative and qualitative discovery: The abacus system. Machine Learning **1** (1986) 367–401
4. Washio, T., Motoda, H.: Discovering admissible models of complex systems based on scale-types and identity constraints. In: Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, Nagoya, Japan (1997) 810–817
5. Dzeroski, S., Todorovski, L.: Discovering dynamics: from inductive logic programing to machine discovery. Journal of Intelligent Information Systems **4** (1995) 89–108
6. Todorovski, L., Dzeroski, S.: Declarative bias in equation discovery. In: Proceedings of the Fourteenth International Conference on Machine Learning, San Mateo, California, Morgan Kaufmann (1997) 376–384
7. Langley, P., George, D., Bay, S., Saito, K.: Robust induction of process models from time-series data. In: Proceedings of the Twentieth International Conference on Machine Learning, Menlo Park, California, The AAAI Press (2003) 432–439
8. Bradley, E.A., O'Gallagher, A.A., Rogers, J.E.: Global solutions for nonlinear systems using qualitative reasoning. Annals of Mathematics and Artificial Intelligence **23** (1998) 211–228
9. Berge, P., Pomeau, Y., Vidal, C.: Order in Chaos - For understanding turbulent flow. Hermann, Paris, France (1984)
10. Luce, D.R.: On the possible psychological laws. Psychological Review **66** (1959) 81–95
11. Luenberger, D.G.: Linear and Nonlinear Programing. Ed. Adison-Wesley, Cambridge, Massachusetts (1989)
12. Doucet, A., Godsill, S., Andrieu, C.: On sequential monte carlo sampling methods for bayesian filtering. Statistics and Computing **10** (2000) 197–208
13. Haykin, S.S.: Kalman Filtering and Neural Networks. John Wiley & Sons, Inc., Hoboken, New Jersey (2001)
14. Gawthrop, P.J., Smith, L.S.: Metamodelling: Bond Graphs and Dynamic Systems. Prentice-Hall, Englewood Cliffs, New Jersey (1996)