Evolutionary Time Series Model with Parallel Computing

Tomoyuki Higuchi *

Abstract: In this study, we consider a time series model which combines the partial non-Gaussian state space model and self-organizing state space model (SOSSM), where the SOSSM has been proposed to an extension of the generalized state space model. The competing different system/observation models for the state vector can be simultaneously dealt in this model as introducing a switching structure, and appropriate system/observation models among them is automatically determined as a function of time. As a result, we are free from a procedure of selecting models among competing models. Of course, this model allows us to consider an inclusion of two system/observation models which conflicts each other in some sense. Namely, we need not have well-organized knowledge about modeling of the time series. We therefore call this model the evolutionary time series model. We regard the approach based on the evolutionary time series model as one of the machine learning approaches. The evolutionary time series model can be formulated within a framework of an extension of SOSSM which takes a convenient form from a computational point of view. The Monte Carlo filter (MCF) is fully employed to handle computationally extensive and difficult tasks. The approach we propose demands a huge computational power and then its realization in practice relies on the massive parallel computer systems. To illustrate, we demonstrate an application of the proposed method to simulated data as well as a seasonal adjustment of the Japanese Gross Domestic Product (GDP) data.

Keywords: *latent variable, machine learning, Monte Carlo filter, parallel computing, partial non-Gaussian state space model, self-organizing state space model, switching model.*

1 Introduction

A **partial non-Gaussian state-space models** [2, 18] is a linear model whose parameter evolve with time according to an unobserved stochastic process $\mathbf{s}_t$ and defined by:

$$\begin{cases} \mathbf{x}_t = F_t(\mathbf{s}_t)\mathbf{x}_{t-1} + G_t(\mathbf{s}_t)\mathbf{v}_t \\ \mathbf{y}_t = H_t(\mathbf{s}_t)\mathbf{x}_t + E_t(\mathbf{s}_t)\mathbf{w}_t, \end{cases} \tag{1}$$

where $\mathbf{x}_t$ is an $n_x \times 1$ vector of unobserved sate variables, and $\mathbf{y}_t$ is an n_y dimensional vector observation. Given $\mathbf{s}_t$, all matrices, F_t, G_t, H_t, and E_t are known matrices of appropriate dimension. The noise sequences $\mathbf{v}_t$ and $\mathbf{w}_t$ are assumed to follow $\mathbf{v}_t \sim N(0, Q_t)$ and $\mathbf{w}_t \sim N(0, R_t)$, respectively. Conditional upon $\mathbf{s}_t$, this model is a standard linear Gaussian state space model [1]. The process $\mathbf{s}_t$ is usually assumed to be a first order Markov process of Markov transition kernel $p(\mathbf{s}_t|\mathbf{s}_{t-1})$. The **switching Gaussian state space model** is considered as a variant of the partial non-Gaussian state space model [5].

The generalized state space models (SOSSM) are also the useful framework for describing a wide variety of the time series models [11, 14] and are described by the following two equations:

$$\begin{cases} \mathbf{x}_t = f(\mathbf{x}_{t-1}, \mathbf{v}_t) \\ \mathbf{y}_t \sim r(\ \cdot\ |\mathbf{x}_t, \boldsymbol{\theta}_{obs}). \end{cases} \tag{2}$$

$f : R^{n_x} \times R^{n_v} \to R^{n_x}$ is given function. $\{\mathbf{v}_t\}$ is independent and identically distributed (i.i.d.) random process with $\mathbf{v}_t \sim q(\mathbf{v}|\boldsymbol{\theta}_{sys})$. r is conditional distribution of $\mathbf{y}_t$ given $\mathbf{x}_t$. $q(\cdot|\cdot)$ and $r(\cdot|\cdot)$ are, in general, non-Gaussian densities specified by the unknown parameter vectors, $\boldsymbol{\theta}_{sys}$ and $\boldsymbol{\theta}_{obs}$, respectively.

To realize a dependency of $\boldsymbol{\theta}_{sys}$ and $\boldsymbol{\theta}_{obs}$ on time in the generalized state space model, the **self-organizing state space model** has been introduced [14]. Namely, we can consider a model in which $\boldsymbol{\theta}_{sys}$ and $\boldsymbol{\theta}_{obs}$ also evolve with time. For a simple notation, we set $\boldsymbol{\theta}_t = [\boldsymbol{\theta}'_{sys,t}, \boldsymbol{\theta}'_{obs,t}]'$. A system model for time-varying parameter $\boldsymbol{\theta}_t$ is described by $\boldsymbol{\theta}_t = g(\boldsymbol{\theta}_{t-1}, \boldsymbol{\varepsilon}_t)$, where $\boldsymbol{\varepsilon}_t$ is a white noise sequence with density function $\phi(\boldsymbol{\varepsilon}|\boldsymbol{\xi})$. We usually adopt the random walk model for g: $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\varepsilon}_t$ [11, 14]. The parameter to be estimated by the maximum likelihood method, $\boldsymbol{\xi}$, is anticipated to be up to 5 dimensional vector.

In this study we propose a general framework which combines the partial non-Gaussian state space model

*The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan; higuchi@ism.ac.jp

and self-organizing state space model conceptually. Section 2 describes this framework which we call an evolutionary time series model. Section 3 shows three examples of the evolutionary time series models. Section 4 explains a calculation of how to estimate the state vector and to determine the posterior distribution. An implementation of the procedure on the parallel computing is also described. Section 5 illustrates an application of the evolutionary time series model to simulated data set. Section 6 demonstrates an application of the evolutionary time series model to a seasonal adjustment of the Japanese Gross Domestic Product (GDP) data.

2 Evolutionary time series model

2.1 Latent switching vector variable

In this study we assume that in addition to the state vector process $\mathbf{x}_t$ and $\boldsymbol{\theta}_t$, a discrete latent switching vector variable

$$\mathbf{I}_t = [I_t^1, \ldots, I_t^K]' \tag{3}$$

influences the distribution of $\mathbf{y}_t$, where the k-th element in (3) takes values in $\{1, \ldots, m^k\}$. The number of all combinations of possible switching variable is $M = m^1 \times m^2 \times \cdots \times m^K$. Therefore we are required to predetermine M generalized state space models. We denote the dependency of f, g, and r on the latent switching vector as $f_{\mathbf{I}}$, $g_{\mathbf{I}}$, and $r_{\mathbf{I}}$, respectively.

The general form of the **evolutionary time series model** that we propose in this study is obtained by augmenting the state vector $\mathbf{x}_t$ with both the parameter vector $\boldsymbol{\theta}_t$ and latent switching vector $\mathbf{I}_t$ as $\mathbf{z}_t = [\mathbf{x}_t', \boldsymbol{\theta}_t', \mathbf{I}_t']'$. Then the model for the augmented state vector $\mathbf{z}_t$ is given by

$$\left\{ \begin{array}{l} \mathbf{z}_t = F^*(\mathbf{z}_{t-1}, \mathbf{u}_t) \\ \mathbf{y}_t \sim r_{\mathbf{I}=\mathbf{I}_t}(\cdot|\mathbf{z}_t), \end{array} \right. \tag{4}$$

where the nonlinear function F^* is defined by

$$F^*(\mathbf{z}_{t-1}, \mathbf{u}_t) = \left[\begin{array}{c} f_{\mathbf{I}=\mathbf{I}_t}(\mathbf{x}_{t-1}, \mathbf{v}_t) \\ g_{\mathbf{I}=\mathbf{I}_t}(\boldsymbol{\theta}_{t-1}, \boldsymbol{\varepsilon}_t) \\ h(\mathbf{I}_{t-1}) \end{array} \right]. \tag{5}$$

The system noise $\mathbf{u}_t$ is defined by $\mathbf{u}_t = [\mathbf{v}_t', \boldsymbol{\varepsilon}_t']'$.

There are various ways of giving a time-dependent structure for $\mathbf{I}_t$, $\mathbf{I}_t = h(\mathbf{I}_{t-1})$. The simplest way is the *independently exchangeable prior*, where an element of $\mathbf{I}_t$, I_t^k, is assumed to an iid process with $\Pr\{I_t^k = j\} = \eta_j$ (e.g., [19]). An evolution of $\mathbf{I}_t$, $h(\cdot)$, is in this study realized by considering *Markovian switching prior* where I_t^k, is assumed to be a stationary Markov process with $\Pr\{I_t^k = j | I_{t-1}^k = i\} = \eta_{ij}^k$ (see e.g., [5, 16]). η_{ij}^k is called the transition matrix or Markov transition kernel. The reason we adopt the Markovian switching prior is that it can be regarded as a discrete version of the random walk which describes a time dependent structure of the parameter vector $\boldsymbol{\theta}_t$ in this study.

No correlation among the latent switching variables is here considered (namely, independently Markov switching model) so that η_{ij}^k is common to all k and defined by

$$\eta_{ij}^k = \eta_{ij} = \left\{ \begin{array}{ll} 0.95 & i = j \\ 0.05/(m^k - 1) & i \neq j. \end{array} \right. \tag{6}$$

3 Examples of model

3.1 local level model with switching system/observation variance

The local level model with switching system/observation variances is given by

$$\begin{aligned} \mu_t &= \mu_{t-1} + v_t, \quad v_t \sim N(0, Q_t^{[k]}) \\ y_t &= \mu_t + w_t, \quad w_t \sim N(0, R_t^{[k]}) \\ \log_{10} Q_t^{[k]} &= \log_{10} Q_{t-1}^{[k]} + \varepsilon_{Q,t}^{[k]}, \\ & \quad \varepsilon_t^{Q,[k]} \sim N(0, \xi_1^{[k]}) \quad \text{for } k = 1, 2 \\ \log_{10} R_t^{[k]} &= \log_{10} R_{t-1}^{[k]} + \varepsilon_{R,t}^{[k]}, \\ & \quad \varepsilon_{R,t}^{[k]} \sim N(0, \xi_2^{[k]}) \quad \text{for } k = 1, 2. \end{aligned} \tag{7}$$

The two type of the system noise variance sequences, $Q_t^{[1]}$ and $Q_t^{[2]}$, is discriminated by the latent switching variable, I_t^1; $Q_t^{[1]}$ for $I_t^1 = 1$ and $Q_t^{[2]}$ for $I_t^1 = 2$, respectively. Similarly, $R_t^{[1]}$ for $I_t^2 = 1$ and $R_t^{[2]}$ for $I_t^2 = 2$. Then the number of all combinations of I_t^1 and I_t^2 values are four; $M = 2 \times 2$. For this model, the function forms appearing in the evolutionary time series models, F^* and r are common to all cases. The parameter vector optimized in parallel is $\boldsymbol{\xi} = [\xi_1^{[1]}, \xi_1^{[2]}, \xi_2^{[1]}, \xi_2^{[2]}]'$. This model can be formulated to the evolutionary state space model with

$$\begin{aligned} \mathbf{z}_t &= [\mu_t, \log_{10} Q_t^{[1]}, \log_{10} Q_t^{[2]}, \log_{10} R_t^{[1]}, \\ & \quad \log_{10} R_t^{[2]}, I_t^1, I_t^2]. \end{aligned} \tag{8}$$

When we focus on a simpler one based on this model, for example, the model with constant $Q^{[k]}$ and $R^{[k]}$, it can be reduced to be the simplest model among the models called *switching Gaussian state space model* [5, 12] or modification of the *Jump Markov linear system* [2].

3.2 Switching trend model

The function form appearing the system model mentioned above, $f_{\mathbf{I}=\mathbf{I}_t}(\cdot)$, is independent of the latent

switching variable. Here we consider three different system model each of which are labeled with I_t^1:

for $I_t^1 = 1$: first order difference (**D1**) model

$$\mu_t = \mu_{t-1} + v_t, \quad v_t^{[1]} \sim N(0, Q_t^{[1]}) \tag{9}$$

for $I_t^1 = 2$: second order difference (**D2**) model

$$\mu_t = 2\mu_{t-1} - \mu_{t-2} + v_t^{[2]}, \quad v_t \sim N(0, Q_t^{[2]}) \tag{10}$$

for $I_t^1 = 3$: heavy tailed non-Gaussian noise model

$$\mu_t = \mu_{t-1} + v_t^{[3]}, \quad v_t^{[3]} \sim C(0, Q_t^{[3]}), \tag{11}$$

where $C(0, Q_t^{[3]})$ is the Cauchy distribution with the centered at 0 and dispersion parameter $Q_t^{[3]}$. We call the third model **C1** model hereafter. An employment of the heavy tailed Non-Gaussian distribution such as a Cauchy distribution for a system noise distribution is aimed at detecting the very large jumps of the trend component which shows a smooth behavior except for the jump points [13]. As in the previous model, we assume that all parameters appearing in a description of system noise distribution, $Q_t^{[1]}$, $Q_t^{[2]}$, and $Q_t^{[3]}$, evolves with time:

$$\begin{aligned} \log_{10} Q_t^{[k]} &= \log_{10} Q_{t-1}^{[k]} + \varepsilon_{Q,t}^{[k]}, \\ \varepsilon_{Q,t}^{[k]} &\sim N(0, \xi_1^{[k]}) \quad \text{for } k = 1, 2, 3 \end{aligned} \tag{12}$$

We assume that in addition to a simple description for the observation model of $y_t = \mu_t + w_t$, $w_t \sim N(0, R_t)$, the time dependence of R_t can be simply described by

$$\begin{aligned} \log_{10} R_t &= \log_{10} R_{t-1} + \varepsilon_{R,t}, \\ \varepsilon_{R,t} &\sim N(0, \xi_2). \end{aligned} \tag{13}$$

Then $K = 1$ and $M = m^1 = 3$. This model can be formulated to the evolutionary state space model with

$$\begin{aligned} \mathbf{z}_t = [&\mu_t, \mu_{t-1}, \log_{10} Q_t^{[1]}, \log_{10} Q_t^{[2]}, \\ &\log_{10} Q_t^{[3]}, \log_{10} R_t, I_t^1]. \end{aligned} \tag{14}$$

The function forms F^* are properly defined according to the system models Eq.(9–13). The parameter vector optimized in parallel is $\boldsymbol{\xi} = [\xi_1^{[1]}, \xi_1^{[2]}, \xi_1^{[3]}, \xi_2]'$.

3.3 Switching seasonal variation model

Usually, an analysis of the seasonal time series is carried out in terms of the procedure called seasonal adjustment which is designed to decompose a time series y_t into multi-factors: a trend component μ_t, seasonal component s_t, observation noise component w_t, and so on. We consider a model by which y_t is decomposed into three factors: $y_t = \mu_t + s_t + w_t$. The decomposition can be achieved by assuming the stochastically perturbed linear difference equation on each component [9, 13, 20]. We consider two type of the trend model used in Section 3.1. Namely we use the D1 and D2 models:

for $I_t^1 = 1$: **D1** model

$$\mu_t = \mu_{t-1} + v_{1,t}^{[1]}, \quad v_{1,t}^{[1]} \sim N(0, Q_{1,t}^{[1]})$$

for $I_t^1 = 2$: **D2** model

$$\mu_t = 2\mu_{t-1} - \mu_{t-2} + v_{1,t}^{[2]}, \quad v_{1,t}^{[2]} \sim N(0, Q_{1,t}^{[2]})$$

$$\begin{aligned} \log_{10} Q_{1,t}^{[k]} &= \log_{10} Q_{1,t-1}^{[k]} + \varepsilon_{Q1,t}^{[k]}, \\ \varepsilon_{Q1,t}^{[k]} &\sim N(0, \xi_1^{[k]}) \quad \text{for } k = 1, 2. \end{aligned} \tag{15}$$

When we analyze a quarterly data y_t, there are several ways to define a structure of the seasonal variation of s_t as a system model. The first way, which is labeled by $I_t^2 = 1$, is a simple representation such as $s_t = s_{t-4} + v_{2,t}^{[1]}$, $v_{2,t}^{[1]} \sim N(0, Q_{2,t}^{[1]})$. An alternative choice is the model which is frequently adopted in the standard basic structural model (BSM) [4, 9, 10]: $s_t = -(s_{t-1} + s_{t-2} + s_{t-3}) + v_{2,t}^{[2]}$, where $v_{2,t}^{[2]} \sim N(0, Q_{2,t}^{[2]})$. Of course, a time dependency of both $Q_{2,t}^{[1]}$ and $Q_{2,t}^{[2]}$ is again introduced as follows:

for $I_t^2 = 1$: **Diff** model

$$s_t = s_{t-4} + v_{2,t}^{[1]}, \quad v_{2,t}^{[1]} \sim N(0, Q_{2,t}^{[1]})$$

for $I_t^1 = 1$ and $I_t^2 = 2$: **Sum** model

$$\begin{aligned} s_t &= -(s_{t-1} + s_{t-2} + s_{t-3}) + v_{2,t}^{[2]}, \\ & \quad v_{2,t} \sim N(0, Q_{2,t}^{[2]}) \\ \log_{10} Q_{2,t}^{[k]} &= \log_{10} Q_{2,t-1}^{[k]} + \varepsilon_{Q2,t}^{[k]}, \\ \varepsilon_{Q2,t}^{[k]} &\sim N(0, \xi_2^{[k]}) \quad \text{for } k = 1, 2. \end{aligned} \tag{16}$$

This model has $K = 2$ and $m^1 = m^2 = 2$ and can be formulated to the evolutionary state space model with

$$\begin{aligned} \mathbf{z}_t = [&\mu_t, \mu_{t-1}, s_t, s_{t-1}, s_{t-2}, s_{t-3}, \log_{10} Q_{1,t}^{[1]}, \\ &\log_{10} Q_{1,t}^{[2]}, \log_{10} Q_{2,t}^{[1]}, \log_{10} Q_{2,t}^{[2]}, I_t^1, I_t^2]'. \end{aligned} \tag{17}$$

The function forms F^* are properly defined according to the system models Eq.(15,16). The parameter vector optimized in parallel is $\boldsymbol{\xi} = [\xi_1^{[1]}, \xi_1^{[2]}, \xi_2^{[1]}, \xi_2^{[2]}]'$.

4 Calculation

4.1 Recursive calculation [9, 13]

The SOSSM exploits the useful recursive formulas for estimating conditional probability distribution of the state vector $\mathbf{z}_i$ given data $\mathbf{y}_{1:j} \equiv [y_1, y_2, \ldots, y_j]$, $p(\mathbf{z}_i|\mathbf{y}_{1:j})$ [9, 13]. The recursive formulas are given by a set of the following two steps at each time t: *prediction* and *filtering* .

(1) *prediction*: Assuming knowledge of the posterior distribution for the state vector at time $t-1$,

$p(\mathbf{z}_{t-1}|\mathbf{y}_{1:t-1})$, compute the one–step–ahead predictive distribution at time t, $p(\mathbf{z}_t|\mathbf{y}_{1:t-1})$, by

$$p(\mathbf{z}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{z}_{t-1}|\mathbf{y}_{1:t-1}) \cdot p(\mathbf{z}_t|\mathbf{z}_{t-1}) d\mathbf{z}_{t-1}. \quad (18)$$

(2) *filtering*: Based on the obtained distribution, $p(\mathbf{z}_t|\mathbf{y}_{1:t-1})$, compute the posterior distribution at time t, $p(\mathbf{z}_t|\mathbf{y}_{1:t})$, by

$$\begin{aligned} p(\mathbf{z}_t|\mathbf{y}_{1:t}) &= \frac{r(\mathbf{y}_t|\mathbf{z}_t) \cdot p(\mathbf{z}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \\ &= \frac{r(\mathbf{y}_t|\mathbf{z}_t) \cdot p(\mathbf{z}_t|\mathbf{y}_{1:t-1})}{\int r(\mathbf{y}_t|\mathbf{z}_t) \cdot p(\mathbf{z}_t|\mathbf{y}_{1:t-1}) d\mathbf{z}_t}. \end{aligned} \quad (19)$$

These recursive formulas still demand the repeated necessary integrations over a state space which increases enormously with respect to the state dimension n_z. A *sequential Monte Carlo method* [3] for filtering and smoothing, called "Monte Carlo filter" (MCF) [14] or "bootstrap filter" [7] has been proposed to overcome this computational problem.

4.2 Monte Carlo filter (MCF)[7, 11, 14]

To review MCF, suppose that $p(\mathbf{z}_t|\mathbf{y}_{1:t-1})$ and $p(\mathbf{z}_t|\mathbf{y}_{1:t})$ are approximated by the N realizations,

$$Z_{t|t-1} \equiv \{\mathbf{z}^{(i)}_{t|t-1} | i = 1, \ldots, N\} \quad \text{and} \quad (20)$$

$$Z_{t|t} \equiv \{\mathbf{z}^{(i)}_{t|t} | i = 1, \ldots, N\}, \quad (21)$$

respectively. It can be shown that these particles can be generated recursively by the following algorithm:

1. For $i = 1, \ldots, N$ generate n_z-dimensional random number $\mathbf{z}^{(i)}_{0|0} \sim p(\mathbf{z}_0)$.
2. Repeat the following steps for $t = 1, \ldots, T$. For (a)–(d), repeat N times independently for $i = 1, \ldots, N$.
   (a) Get $\mathbf{I}^{(i)}_t$ by $\mathbf{I}^{(i)}_t = h(\mathbf{I}^{(i)}_{t-1})$.
   (b) Generate n_ε-dimensional system random number $\boldsymbol{\varepsilon}^{(i)}_t \sim \phi(\boldsymbol{\varepsilon}|\boldsymbol{\xi})$ and n_v-dimensional system random number $\mathbf{v}^{(i)}_t \sim q(\mathbf{v}|\boldsymbol{\theta}^{(i)}_t)$.
   (c) Compute $\mathbf{z}^{(i)}_{t|t-1} = F^*(\mathbf{z}^{(i)}_{t-1|t-1}, \mathbf{u}^{(i)}_t)$, where $\mathbf{u}^{(i)}_t = [\mathbf{v}'^{(i)}_t, \boldsymbol{\varepsilon}'^{(i)}_t]'$.
   (d) Compute $\widetilde{w}^{(i)}_t = r_{\mathbf{I}=\mathbf{I}_t}(\mathbf{y}_t|\mathbf{z}_t = \mathbf{z}^{(i)}_{t|t-1})$.
   (e) Obtain $Z_{t|t}$ by the sampling with replacement from $Z_{t|t-1}$ with sampling probabilities proportional to $\widetilde{w}^{(1)}_t, \cdots, \widetilde{w}^{(N)}_t$.

This filtering algorithm can be extended to the smoothing by storing the past particles and resampling the vector of particles $(\mathbf{z}'^{(i)}_{t|t-1}, \mathbf{z}'^{(i)}_{t-1|t-1}, \ldots, \mathbf{z}'^{(i)}_{t-L|t-1})$ rather than the single particle $\mathbf{z}^{(i)}_{t|t-1}$.

Once the Monte Carlo smoothing is performed to the evolutionary time series model, the log likelihood in MCF, $l^{\mathrm{MCF}}(\boldsymbol{\xi})$, is given by

$$\begin{aligned} \log p(\mathbf{y}_{1:T}|\boldsymbol{\xi}) &= \sum_{t=1}^{T} \log p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \boldsymbol{\xi}) \\ &\cong \sum_{t=1}^{T} \log \left(\sum_{i=1}^{N} \widetilde{w}^{(i)}_t \right) - T \log N \\ &= l^{\mathrm{MCF}}(\boldsymbol{\xi}), \end{aligned} \quad (22)$$

where

$$\begin{aligned} p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \boldsymbol{\xi}) &= \int r_{\mathbf{I}=\mathbf{I}_t}(\mathbf{y}_t|\mathbf{z}_t) p(\mathbf{z}_t|\mathbf{y}_{1:t-1}, \boldsymbol{\xi}) d\mathbf{z}_t \\ &\cong \frac{1}{N} \sum_{i=1}^{N} \widetilde{w}^{(i)}_t. \end{aligned} \quad (23)$$

The posterior distribution of the state vector $\mathbf{x}_t$ is defined by a marginal posterior distribution given by

$$p(\mathbf{x}_t|\mathbf{y}_{1:T}) = \int p(\mathbf{z}_t|\mathbf{y}_{1:T}) d\boldsymbol{\theta}_t d\mathbf{I}_t. \quad (24)$$

Based on $p(\mathbf{z}_t|\mathbf{y}_{1:T})$, we can calculate the probabilities $\Pr(I^k_t = \ell|\mathbf{y}_{1:T})$, $\ell = 1, \ldots, m^k$ which are estimated from a marginal posterior distribution by:

$$\begin{aligned} &\widehat{\Pr}(I^k_t = \ell|\mathbf{y}_{1:T}) = \\ &\frac{1}{N} \#\{\mathbf{z}^{(i)}_{t|1:T} \text{ with } I^k_t = \ell | i = 1, \ldots, N\}, \end{aligned} \quad (25)$$

where $\mathbf{z}^{(i)}_{t|1:T}$ is a particle of which distribution approximates to $p(\mathbf{z}_t|\mathbf{y}_{1:T})$. Actually, $p(\mathbf{z}_t|\mathbf{y}_{1:T})$ is determined by the fixed interval smoother explained in this section.

A successful usage of the parameter estimation procedure relies on how to set its initial distribution, $p(\boldsymbol{\theta}_0)$. Usually, it is recommended to adopt a uniform distribution of which range covers a possible range of $\boldsymbol{\theta}$. On the contrary, the initial distribution of $\mathbf{I}_0$ does not affect an estimation seriously, and then it is enough to give an equal probability on all discrete values for any k:

$$\Pr(I^k_0 = \ell) = \frac{1}{m^k}. \quad (26)$$

4.3 Grid search for an optimal $\boldsymbol{\xi}$ with parallel computing

As mentioned in Section 2, a parameter vector to be optimized is only $\boldsymbol{\xi}$ of which dimension is expected to take less than 5. Many numerical experiences suggests that $\boldsymbol{\xi}$ need not require a fine tuning, because it produces insignificant change in a behavior of the estimated $\boldsymbol{\theta}_t$ in time domain. Hence we conduct a grid search for an optimization of $\boldsymbol{\xi}$. For example, when $\boldsymbol{\xi}$ is a scaler variable, we choose an optimal ξ among three candidates: $\xi = 10^{-2}$, 10^{-3}, and 10^{-4}.

Table 1: Parallel Computer Systems at ISM

Product	CPU		CPU No.
SGI2800	MIPS R12000	400MHz	$\times$ 80 CPUs
SGI1200	PentiumIII	800MHz	$\times$ 100 CPUs

When $\boldsymbol{\xi}$ is a four dimensional vector as appearing in the previous section, only 81 (3^4) values of the log-likelihood have to be determined. Such rough optimization based on the grid search can be easily implemented on the parallel computer systems with MPI (Message Passing Interface) library [17] as follows. If we assign the ith value of $\boldsymbol{\xi}$ among 81 cases to the ith processor (node) on the parallel computer system, all information required for performing the Monte Carlo filter to calculate $l^{\mathrm{MCF}}(\boldsymbol{\xi})$ is common to a processor except for a value of $\boldsymbol{\xi}$. Then we can take an advantage of the parallel computer as much as possible, because an amount of communication between the processors is quite small; we have only to gather a value of $l^{\mathrm{MCF}}(\boldsymbol{\xi})$ from the processors, and determine an optimal $\boldsymbol{\xi}$ with the maximum value of $l^{\mathrm{MCF}}(\boldsymbol{\xi})$, which is specified by $\boldsymbol{\xi}^*$. As a result, we can enjoy a benefit of exploiting the parallel computing.

The Institute of Statistical Mathematics (ISM) has two massive parallel computer systems listed in Table 1 that we used in this study. Here we would like to defend an expected criticism against a prospect of an approach of using the evolutionary time series model which relies on a massive parallel computer system. Since an organization of possessing the massive parallel computer systems is in general limited, most of users are obliged to perform their analysis on the PC with a single CPU. However, there are new projects which frees a user from his/her computational resource. For example, the **Ninf** is an ongoing global computing infrastructure project which allows users to access computational resources including hardware, software and scientific data distributed across a wide area network with an easy-to-use interface [8]. In future, we will realize the approach proposed here in a kind of Ninf system.

5 Application to simulated data

5.1 Data simulated from a local level model with switching observation variance

First, we apply the method of Section 3.1 to a time series of length $T = 400$ simulated from a local level model with switching observation variance [5]:

$$\begin{aligned} \mu_t &= \mu_{t-1} + v_t, \quad v_t \sim N(0, 0.001) \\ y_t &= \mu_t + w_t, \quad w_t \sim N(0, R_t) \end{aligned} \tag{27}$$

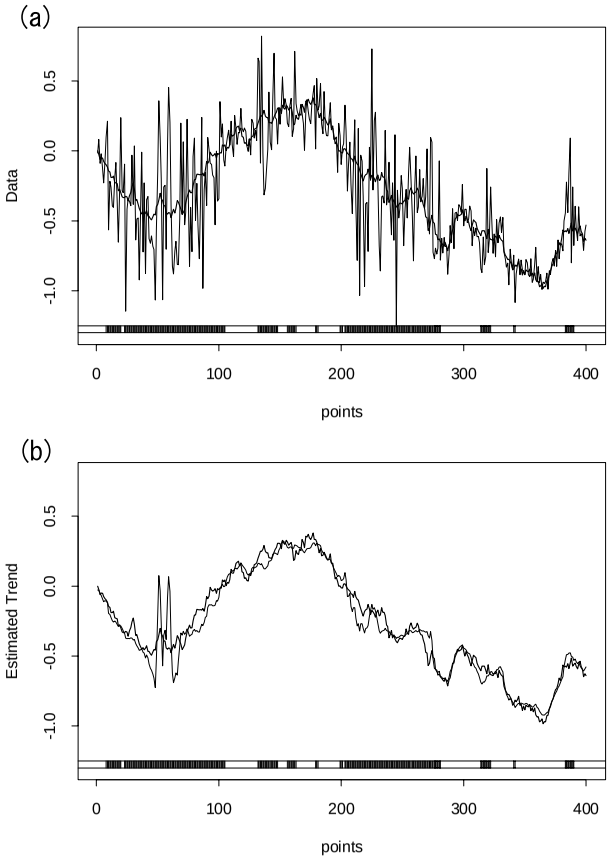

Figure 1: Analysis of a local level model with switching observation variance. (a) Given data (thin line) and simulated trend component, μ_t^{true} (thick line). The data points with larger observation noise are denoted by a black portion in the bar. (b) Estimated μ_t, $\widehat{\mu}_t$. A thick line is μ_t^{true}.

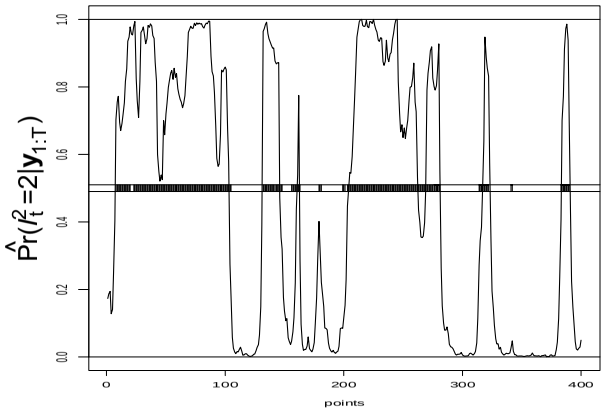

Figure 2: The probabilities of having a larger observation noise, $\widehat{\Pr}(I_t^2 = 2|\mathbf{y}_{1:T})$. A black/white bar is identical to one shown in Figure 1.

where

$$R_t = \begin{cases} 0.01, & I_t = 1, \\ 0.1, & I_t = 2, \end{cases} \tag{28}$$

and I_t is simulated as a 2-state Markov chain with transition matrix η given by Eq.(6). The simulated data y_t is denoted by a thin line in Figure 1. The simulated μ_t, μ_t^{true}, is shown by a thick line in this figure. The black portion of a bar exhibited in the bottom corresponds to the points with $I_t^1 = 2$; namely, μ_t^{true} is contaminated at these points with larger observation noise. Figure 1(b) demonstrates the estimated trend component, $\widehat{\mu}_t$. For this problem, we give the following initial distribution: $p(\mu_0) = N(0, 0.01)$, $p(\log_{10} Q_0^{[1]}) = U([-3.5, -2.5])$, $p(\log_{10} Q_0^{[2]}) = U([-1.5, -0.5])$, $p(\log_{10} R_0^{[1]}) = U([-2.5, -1.5))$, and $p(\log_{10} R_0^{[2]}) = U((-1.5, -0.5])$. We use $N = 50,000$ particles for this analysis. μ_t^{true} is again superposed in this figure by a thick line. A bar exhibited in this figure is identical to one in Figure 1(a). It is clearly shown that $\widehat{\mu}_t$ has a good agreement with μ_t^{true} except for around $t = 50$. To illustrate this good agreement, we calculate $\Pr(I_t^2 = 2|\mathbf{y}_{1:T})$. Figure 2 shows $\Pr(I_t^2 = 2|\mathbf{y}_{1:T})$ against the simulated I_t (a black/white bar). It is seen in this figure that the data points with $I_t = 2$ is clearly detected by an evolutionary time series model.

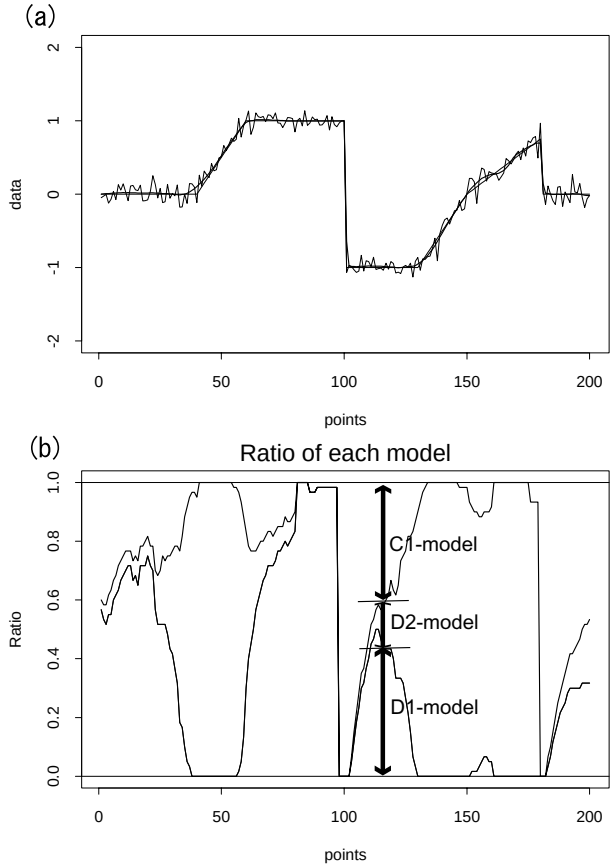

Figure 3: (a) Simulated data (thin line) and estimated trend component (thick line). (b) Ratio of the probabilities of each model among three trend models.

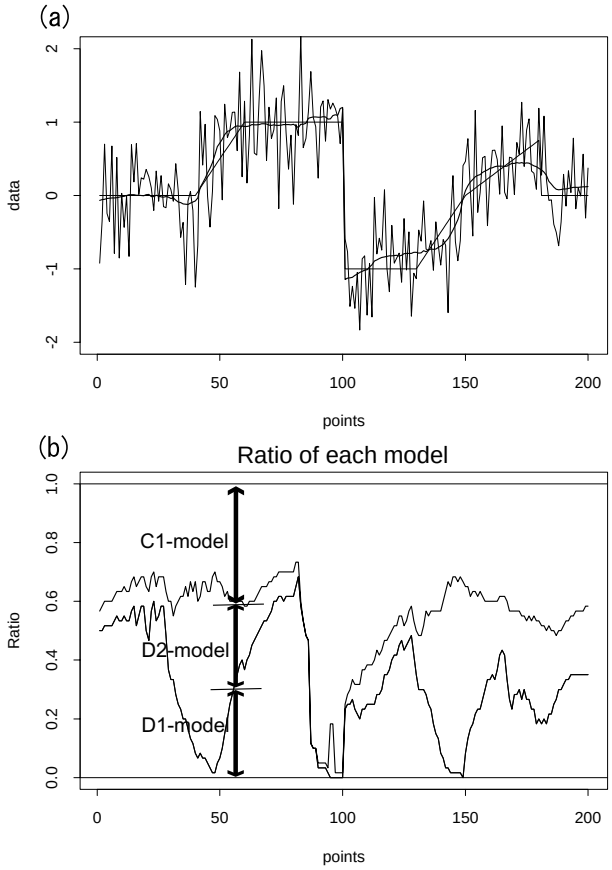

Figure 4: (a) Simulated data (thin line) and estimated trend component (thick line). (b) Ratio of the probabilities of each trend model.

5.2 Estimation of the trend component with discontinuities in both its level and slope

We consider the data of length $T = 200$ which is generated from the following model: $y_t = \mu_t^{true} + w_t$, $w_t \sim N(0, 0.01)$, where μ_t^{true} is a piece-wise liner function denoted by a thin line in Figure 3(a). There are the discontinuities in its level at $t = 101$ and $t = 181$. The intervals of $t = 41 \sim 60$, and $131 \sim 180$ have a finite value of the slope. Except for these intervals, μ_t^{true} has a constant level. We apply the switching trend model explained in Section 3.2 to this time series. $N = 10,000$ particles are used for this problem. The estimated trend component, $\widehat{\mu}_t$, is superposed on this figure by the thick line. Since we have a good agreement between $\widehat{\mu}_t$ and μ_t^{true}, a discrepancy between them is invisible.

The satisfactory results that we have obtained are most easily communicated graphically. Figure 3(b) illustrates a ratio of each model among three models; D1, D2, and C1 models. The two curves drawn are defined by:

$$\text{Upper line:} \quad \widehat{\Pr}(I_t^1 = 1|\mathbf{y}_{1:T}) + \widehat{\Pr}(I_t^1 = 2|\mathbf{y}_{1:T})$$
$$\text{Lower line:} \quad \widehat{\Pr}(I_t^1 = 1|\mathbf{y}_{1:T}).$$

Accordingly, a ratio of the length of three portions (three arrows in Figure 3(b)) separated by these two lines for each time represents to what extent each model among three trend models is employed as a function of time. As mentioned previously, only the C1 model can account for an abrupt change in the level. Figure 3 clearly shows that this expectation is realized by the evolutionary time series model. It is obvious that the D2 model is appropriate when there is a finite value of the slope. Actually, for the intervals of the slope being a finite, the D2 model has a majority among three trend models. In short, the evolutionary time series model can accommodate itself to the observations.

We apply the same evolutionary time series model to other simulated time series with a different observation noise sequences: $w_t \sim N(0, 0.25)$. A standard deviation of this noise process is 5 times larger than that for the time series analyzed previously. Figure 4(a) demonstrates the simulated data y_t and estimated trend component that are obtained. The estimated trend component is denoted by the thick line. A presence of larger observation noise in comparison with that in previous figure produces larger discrepancy of $\widehat{\mu}_t$ from μ_t^{true}. However, the time series model can also detect a sharp drop in the level at $t = 101$ as well as reasonable smoothing elsewhere. Figure 4(b) demonstrates a ratio of each model. In contrast to Figure 3(b), the C1 model keeps to account for at least one third of the probabilities throughout a time. A replacement of the trend model in response to μ_t^{true} are made slowly owing to larger observation noise.

6 Application to seasonal adjustment of GDP data

For further illustration we analyze the Japanese Gross Domestic Product (GDP) data from second quarter, 1955 to first quarter, 1995. The GDP data is a quarterly data. We apply the evolutionary time series model explained in Section 3.3 to the logarithmically transformed GDP data which is shown in Figure 5(a). The reason of applying the log transformation is that a linear decomposition suits for the logarithmically transformed GDP rather than the original one. Of course, an alternative transformation through such as the Box-Cox transformation is available (a square root transformation is better), but a detailed investigation regarding to a way of transformation is not essential for discussion.

$N = 100,000$ particles are used for this problem. The estimated trend component $\widehat{\mu}_t$ is indicated by a thick line. A thin line in Figure 5(b) shows the probabilities $\widehat{\Pr}(I_t^1 = 1|\mathbf{y}_{1:T})$ as a function of time. The thick line connects the first difference of $\widehat{\mu}_t$ normalized by its maximum value:

$$\Delta_{\mu,t} = \frac{\widehat{\mu}_{t+1} - \widehat{\mu}_t}{\max.\{\widehat{\mu}_{t+1} - \widehat{\mu}_t\}}. \tag{29}$$

It is seen in Figure 5(b) that $\Delta_{\mu,t}$ takes zero and smaller values for the intervals of the probability being nearly to one. This tendency is visually demonstrated by making a scatter plot of $\widehat{\Pr}(I_t^1 = 1)$ versus $\Delta_{\mu,t}$ shown in Figure 6. An extremely smaller percentage of the D1 model for the initial several data points seen in Figure 5(b) is owing to poor adjustments of the initial distribution for $\mathbf{x}_0$. Figure 7(a) shows the estimated seasonal component, $\widehat{s}_t$. Figure 7(b) connects the probabilities of the Diff model being appropriate as a function of time, $\widehat{\Pr}(I_t^2 = 1|\mathbf{y}_{1:T})$.

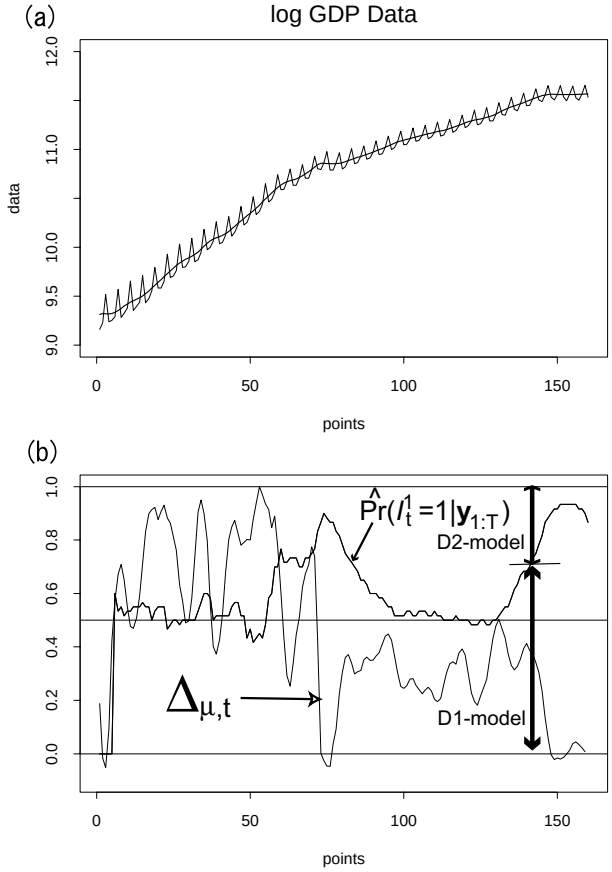

Figure 5: Analysis of the log GDP data (quarterly data). (a) Given data and estimated trend component, $\widehat{\mu}_t$ (thick line). (b) The probabilities of the D1 model being appropriate as a function of time, $\widehat{\Pr}(I_t^1 = 1|\mathbf{y}_{1:T})$. A thin line connects the first difference of the estimated trend component normalized by its maximum value.

References

[1] B.D. Anderson, and J.B. Moore, *Optimal Filtering*, Prentice-Hall, New Jersey, 1979.

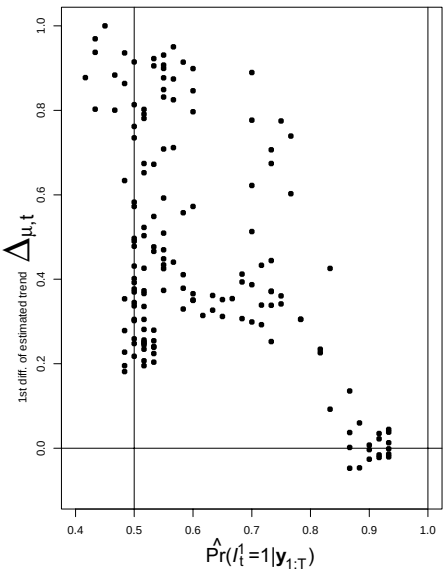

Figure 6: Plots of $\widehat{\Pr}(I_t^1 = 1|\mathbf{y}_{1:T})$ versus $\Delta_{\mu,t}$.

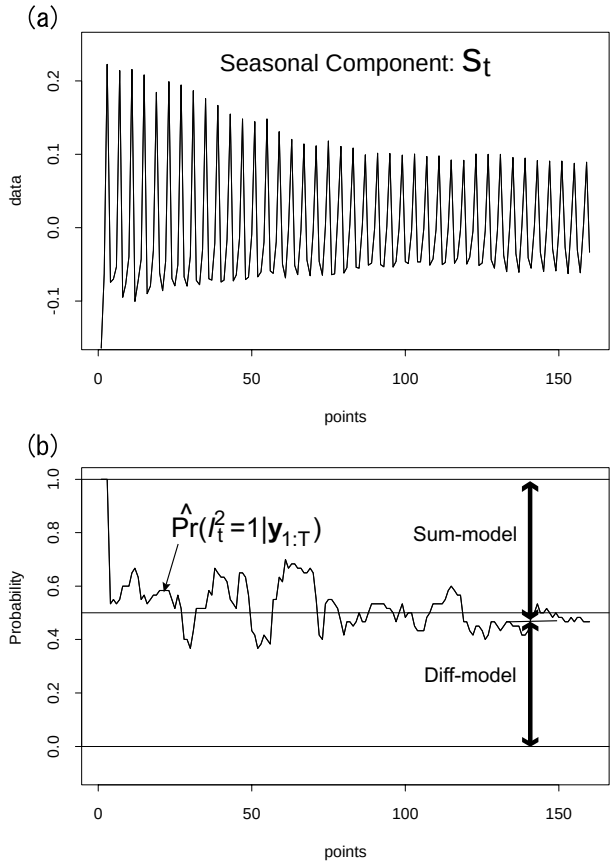

Figure 7: (a) Estimated seasonal component, $\widehat{s}_t$. (b) The probabilities of the Diff model being appropriate as a function of time, $\widehat{\Pr}(I_t^2 = 1|\mathbf{y}_{1:T})$.

[2] N. Bergman, A. Doucet, and N. Gordon, Optimal estimation and Cramér-Rao bounds for partial Non-Gaussian state space models, *Annals of Institute of Statistical Mathematics*, 53, No. 1, pp. 97–122, 2001.

[3] A. Doucet, J.F.G, de Freitas, and N.J. Gordon eds., *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, 2001.

[4] J. Durbin, and S.J. Koopman, Monte Carlo maximum likelihood estimation for non-gaussian state space models, *Biometrika*, 84, 1403–1412, 1997.

[5] S. Frühwirth-Schnatter, Fully Bayesian analysis of switching Gaussian state space models, *Annals of Institute of Statistical Mathematics*, 53, No. 1, pp. 31–49, 2001.

[6] K. Gordon, and A. F. M. Smith, Modeling and monitoring biomedical time series, *Journal of the American Statistical Association*, Vol. 85, No. 410, pp. 328–337, 1990.

[7] N.J. Gordon, D.J. Salmond, and A.F.M. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings-F*, 140, No. 2, 107–113, 1993.

[8] `http://ninf.etl.go.jp/`.

[9] A.C. Harvey, *Forecasting, Structural Time Series Models, and Kalman Filter*, Cambridge, Cambridge University Press, 1989.

[10] T. Higuchi, Monte Carlo filter using the genetic algorithm operators, *Journal of Statistical Computation and Simulation*, 59, No. 1 1–23, 1997.

[11] T. Higuchi, and G. Kitagawa, Knowledge Discovery and Self-organizing state space model, *IEICE Transactions on Information and Systems*, E83-D, No.1, pp. 36–43, 2000.

[12] C.-J. Kim, and Ch. R. Nelson, *State Space Models with Regime Switching*, MIT press, Cambridge, Massachusetts, 1999.

[13] G. Kitagawa, Non-Gaussian state-space modeling of nonstationary time series (with discussion), *Journal of the American Statistical Association*, 82. No. 400, pp. 1032–1063, 1987.

[14] G. Kitagawa, Monte Carlo filter and smoother for non-Gaussian nonlinear state space model, *Journal of Computational and Graphical Statistics*, 5, No. 1, pp. 1–25, 1996.

[15] G. Kitagawa, Self-organizing state space model, *Journal of the American Statistical Association*, 93, No. 443, pp. 1203–1215, 1998.

[16] F. Komaki, State-space modeling of time series sampled from continuous process with pulses, *Biometrika*, 80, 417–429, 1993.

[17] `http://www.mpi-forum.org/`.

[18] N. Shephard, Partial non-Gaussian time series models, *Biometrika*, 81, 115–131, 1994.

[19] R. H. Shumway, and D. S. Stoffer, Dynamic liner models with switching, *Journal of the American Statistical Association*, Vol. 86, pp. 763–769, 1991.

[20] M. West, P.J. Harrison, and H.S. Migon, Dynamic generalized linear models and Bayesian forecasting (with discussion), *Journal of the American Statistical Association*, 80, pp. 73–97, 1985.