Gaussian process dynamical models (GPDMs) are used for nonlinear dimensionality reduction in time series by means of Gaussian process priors. An extension of GPDMs is proposed for visualizing the states of time series. The conventional GPDM approach associates a state with an observation value. Therefore, observations changing over time cannot be represented by a single state. Consequently, the resulting visualization of state transition is difficult to understand, as states change when the observation values change. To overcome this issue, autoregressive GPDMs, called ARGPDMs, are proposed. They associate a state with a vector autoregressive (VAR) model. Therefore, observations changing over time can be represented by a single state. The resulting visualization is easier to understand, as states change only when the VAR model changes. We demonstrate experimentally that the ARGPDM approach provides better visualization compared with conventional GPDMs.

Keywords: gaussian process dynamical models, vector autoregressive models, time series, visualization

1. Introduction

In multivariate time-series analysis [1, 2], data whose state changes over time are occasionally handled. For example, the non-scheduled worked hours index and operating profit are time-series data that cyclically fluctuate depending on the domestic business condition. Moreover, stock prices and exchange rate are time-series data that could drastically change owing to a financial crisis or policy change by the government. In this study, the data of time series whose states change over time are subjected to dimensional compression to visualize state transition. Using this method, the state transition is drawn as a trajectory in a low-dimensional characteristics space. This facilitates the visual understanding of the circumstances of the creation of the time-series data. For example, the method allows one to visually analyze whether the state changes cyclically, or whether the state changes smoothly. In addition, the method enables visual analysis of whether a regression or classification error in a regression or classification problem of time-series data is caused by insufficient learning pattern or by insufficient explanatory variables.

Several data visualization techniques have been proposed, such as principal component analysis (where data are visualized by linear functions), self-organization maps (where data are visualized by nonlinear functions [3]), and Gaussian Process Latent Variable Models (GPLVMs) [4]. However, these methods assume independent identically distributed samples and could yield erroneous results if they are applied to inter-correlated time-series data. To resolve this, Wang et al. extend the GPLVM to Gaussian process dynamical models (GPDMs) [5] by including dynamics. GPDMs are stochastic models consisting of a state transition function $f$ in a state space and a function $g$ that maps from the state space to an observation data space. They suppress over-learning by treating the parameters that determine $f$ and $g$ as stochastic variables. In GPDMs, unknown parameters are hyper-parameters that determine prior distributions of the functions $f$ and $g$, and time-series data are visualized by estimating the unknown parameters with maximum a posteriori probability estimation. GPDMs achieve smooth transition in the state space and, in particular, have succeeded as models for generating human body motion data or motion picture series [6–9].

However, in conventional GPDMs, the function $g$ associates a single state with a single regular variable. With a single state, one cannot express time series whose expectation value changes over time. Therefore, if GPDMs are used for time-series data where data values are constantly changing, it is difficult to visually understand state transition. To resolve this, we propose ARGPDMs where each single state is associated with a vector auto-regressive model [2]. ARGPDMs can represent time-series data whose expectation value changes over time by a single state. Therefore, if ARGPDMs are used to visualize time-series data, state transition occurs when the vector auto-regressive model switches; thus, smoother and more comprehensible visualization can be achieved in comparison with GPDMs.

Notably, auto-regressive hidden Markov models [10] were proposed for associating each single state with a single vector auto-regressive model, as in the proposed method. However, auto-regressive hidden Markov models
use a Markov process to discretely switch between a finite number of states (vector auto-regressive model). By contrast, ARGPDMS use a linear base function model to continuously switch between states (vector auto-regressive model). Therefore, the proposed method, which changes states continuously, achieves more comprehensible visualization with smoother state changes than the auto-regressive hidden Markov model, which changes states discretely.

In this study, we derive a visualization algorithm for ARGPDMS using maximum a posteriori probability estimation and perform visualization experiments using artificial time-series data and diffusion indexes to show that ARGPDMS yield a lower average misclassification rate than GPDMs.

The paper is organized as follows. In Section 2, an overview of the GPDMs proposed in [5] is presented. In Section 3, ARGPDMS where a single state is associated with a single vector autoregressive model are proposed, and subsequently, a visualization method for time-series data using ARGPDMS is proposed. Computer simulations are presented in Section 4. Section 5 concludes the paper.

2. GPDMs

In this section, we overview the GPDMs proposed in [5]. GPDMs are used for dimensional compression of the observation data set \( y_t = (y_t,1, \ldots, y_t,D_y)\) at time \( t \) to a data set on a lower dimensional state space, namely, \( z_t = (z_t,1, \ldots, z_t,D_z)\). GPDMs extend GPLVMs, which are used for nonlinear dimensional compression, to handle the dynamics of the state \( z_t \). GPDMs are stochastic models consisting of a state transition function \( f \) in a state space and a function \( g \) that maps from the state space to an observation data space. They suppress over-learning by treating the parameters that determine \( f \) and \( g \) as stochastic variables. GPDMs visualize the state \( z_t \) by setting the number of dimensions of the state space to \( D_z = 2 \) or 3. In the following sections, we will present such a stochastic model and the estimation algorithm.

2.1. Stochastic Model

In GPDMs, the state \( z_t \) and observation data set \( y_t \) are obtained according to the following statespace model:

\[
\begin{align*}
\mathbf{z}_t &= f(\mathbf{z}_{t-1}) \tag{1} \\
\mathbf{y}_t &= g(\mathbf{z}_t) \tag{2}
\end{align*}
\]

Here,

\[
\begin{align*}
f(\mathbf{z}_t) &= (f_1(\mathbf{z}_t), \ldots, f_{M_f}(\mathbf{z}_t))^T \tag{3} \\
g(\mathbf{z}_t) &= (g_1(\mathbf{z}_t), \ldots, g_{M_g}(\mathbf{z}_t))^T \tag{4}
\end{align*}
\]

where \( f \) is a function that determines the state change, and \( g \) is a function that maps the state \( z_t \) onto an observation data space. In GPDMs, the functions \( f \) and \( g \) are defined as a linear combination of \( M_f \) base functions, \( \phi_{f,1}, \ldots, \phi_{f,M_f} \), and a linear combination of \( M_g \) base functions, \( \phi_{g,1}, \ldots, \phi_{g,M_g} \), respectively, as follows:

\[
f(\mathbf{z}_t) = \sum_{m=1}^{M_f} a_{f,m} \phi_{f,m}(\mathbf{z}_t) + \mathbf{n}_{f,t} \tag{5}
\]

\[
g(\mathbf{z}_t) = \sum_{m=1}^{M_g} a_{g,m} \phi_{g,m}(\mathbf{z}_t) + \mathbf{n}_{g,t} \tag{6}
\]

where \( \mathbf{A}_f = [a_{f,1}, \ldots, a_{f,M_f}]^T \) and \( \mathbf{A}_g = [a_{g,1}, \ldots, a_{g,M_g}]^T \) are the parameters that determine the functions \( f \) and \( g \), respectively. \( \mathbf{n}_{f,t} \) and \( \mathbf{n}_{g,t} \) are noise terms independent of each other. The noise terms have zero mean value and follow normal distributions with variance-covariance matrices \( \alpha_f^{-1} \) and \( \alpha_g^{-1} \):

\[
p(\mathbf{n}_{f,t}) = \mathcal{N}(\mathbf{n}_{f,t}; \mathbf{0}, \alpha_f^{-1} I) \tag{7}
\]

\[
p(\mathbf{n}_{g,t}) = \mathcal{N}(\mathbf{n}_{g,t}; \mathbf{0}, \alpha_g^{-1} I) \tag{8}
\]

where \( \mathcal{N}(\mathbf{0}, \Sigma) \) is a normal distribution with mean \( \mathbf{0} \) and variance-covariance matrix \( \Sigma \).

In GPDMs, a stochastic model is defined by assuming prior distributions of the parameters \( \mathbf{A}_f \) and \( \mathbf{A}_g \) without directly handling the integrals of \( \mathbf{A}_f \) and \( \mathbf{A}_g \). The prior distributions of the parameters \( \mathbf{A}_f \) and \( \mathbf{A}_g \) are assumed to be standard normal distributions to make \( f \) and \( g \) smoothly-changing continuous functions.

\[
p(\mathbf{A}_f) = \prod_{m=1}^{M_f} \mathcal{N}(a_{f,m}; 0, I) \tag{9}
\]

\[
p(\mathbf{A}_g) = \prod_{m=1}^{M_g} \mathcal{N}(a_{g,m}; 0, I) \tag{10}
\]

The distributions of the observation data matrix \( \mathbf{Y} = [y_1, \ldots, y_N]^T \) and the state matrix \( \mathbf{Z} = [z_1, \ldots, z_N]^T \) are given in Eqs. (11) and (12), respectively.

\[
p(\mathbf{Y} | \mathbf{Z}, \mathbf{A}_g) = \prod_{d=1}^{D_y} \mathcal{N}(y_{d,t}; 0, \mathbf{K}_g) \tag{11}
\]

\[
= (2\pi)^{-\frac{1}{2}ND_y} |\mathbf{K}_g|^{-\frac{1}{2}} \times \exp \left( -\frac{1}{2} |\mathbf{Y} - \mathbf{K}_g^{-1}\mathbf{Z}|^2 \right) \tag{12}
\]

where \( y_{d,t} \) is the \( d_t \)-th column vector of the matrix \( \mathbf{Y} \) and \( \mathbf{K}_g \) is an \( N \times N \) kernel matrix whose elements are the kernel functions \( k_g(z_t, z_{t'}) \). In Eq. (11), the prior distribution of the vector \( y_{d,t} \) is assumed to be a normal distribution with mutually-independent mean 0 and variance-covariance matrix \( \mathbf{K}_g \). Therefore, GPDMs are stochastic models where the prior distribution of the \( d_t \)-th element \( g_{d,t} \) of the function \( g \) is assumed to be a Gaussian process with mutually-independent mean 0 and variance-covariance function \( k_g(z_t, z_{t'}) \) [11].

\[
p(\mathbf{Z} | \mathbf{A}_f) = p(z_1)(2\pi)^{-\frac{1}{2}(N-1)D_z} |\mathbf{K}_f|^{-\frac{1}{2}} \times \exp \left( -\frac{1}{2} |\mathbf{Z} - \mathbf{K}_f^{-1}\mathbf{Z}_{2,N}|^2 \right) \tag{12}
\]

where \( \mathbf{Z}_{2,N} = [z_{2,1}, \ldots, z_{2,N}]^T \) and \( \mathbf{K}_f \) is an \( (N-1) \times (N-1) \) kernel matrix whose elements are the kernel functions
where $1 ≤ i, j ≤ N − 1$. The prior distribution of the state $z_i$ is assumed to be a standard normal distribution:

$$p(z_i) = \mathcal{N}(z_i; 0, I) \quad . . . . . . . . . . . . . (13)$$

We note that Eq. (12), unlike Eq. (11), is not a normal distribution, as the kernel matrix $K_f$ depends on the state matrix $Z$.

In Section 4 (Computer Simulations), the kernel functions $k_f(z_i, z_j)$ and $k_g(z_i, z_j)$ are set to RBF + linear kernel and defined as follows:

$$k_f(z_i, z_j) = \alpha_{f,1} \exp \left( -\frac{\alpha_{t,2}}{2} ||z_i - z_j||^2 \right) + \alpha_{f,3}^T z_i z_j + \alpha_{f,4} \delta_i z_j \quad . . . . . . (14)$$

$$k_g(z_i, z_j) = \alpha_{g,1} \exp \left( -\frac{\alpha_{g,2}}{2} ||z_i - z_j||^2 \right) + \alpha_{g,3}^T z_i z_j + \alpha_{g,4} \delta_i z_j \quad . . . . . . (15)$$

Here, $\alpha_f = (\alpha_{f,1}, \ldots, \alpha_{f,4})^T$ and $\alpha_g = (\alpha_{g,1}, \ldots, \alpha_{g,4})^T$ are hyper-parameters that determine the kernel functions. The prior distributions of the hyper-parameters $\alpha_f$ and $\alpha_g$ are assumed to be uninformative prior distributions, as follows:

$$p(\alpha_f) \propto \prod_{i=1}^{4} |\alpha_{f,i}| \quad . . . . . . . . . . . . . (16)$$

$$p(\alpha_g) \propto \prod_{i=1}^{4} |\alpha_{g,i}| \quad . . . . . . . . . . . . . (17)$$

In GPDMs, the joint distribution of the observation data matrix $Y$ and the unknown parameter $\boldsymbol{\theta} = \{Z, \alpha_f, \alpha_g\}$ are defined using the distributions of Eqs. (11)–(17), as follows:

$$p(Y, \theta) = p(Y \mid Z, \alpha_g)p(Z \mid \alpha_f)p(\alpha_f)p(\alpha_g) \quad . . . . . . . . . . . . . (18)$$

2.2. Parameter Estimation

In GPDMs, the unknown parameter $\theta$ is estimated by maximizing a maximum a posteriori probability estimation. This is accomplished by maximizing the negative logarithm of a posteriori probability $L$ of the unknown parameter $\theta$ for given observation data $Y$. The negative logarithm a posteriori probability $L$ is calculated as in Eq. (19) from the stochastic model in Eq. (18).

$$L = -\ln p(\theta \mid Y) = l_y + l_z + \sum_{i=1}^{4} \ln |\alpha_{f,i}| + \sum_{i=1}^{4} \ln |\alpha_{g,i}| + \text{const} \quad . . . . . . . (19)$$

$$l_y = \frac{D_y}{2} \ln |K_f| + \frac{1}{2} \text{tr}(K_f^{-1} YY^T) \quad . . . . . . . . . . . . (20)$$

$$l_z = \frac{D_z}{2} \ln |K_g| + \frac{1}{2} \text{tr}(K_f^{-1} Z_{2:N} Z_{2:N}^T) + \frac{1}{2} z_i^T z_i \quad . . . . . . . . . . . . (21)$$

Here, $\text{const}$ is a constant term independent of the unknown parameter $\theta$. The unknown parameter $\theta$ is estimated by maximizing the negative logarithmic a posteriori probability $L$ given in Eq. (19), based on the SCG (scaled conjugate gradient) method [12].

In addition, $l_z$ in Eq. (19) can be considered a penalty term for a non-smooth state transition. Therefore, B-GPDMs (Balanced GPDMs), which estimate parameters by replacing $l_z$ in Eq. (19) with $l_z D_y / D_z$, are proposed in [13] for smoother state transition. In Section 4 (Computer Simulations), we estimate the unknown parameter $\theta$ using B-GPDMs.

3. ARGPDMs

3.1. Stochastic Model

In this section, we propose ARGPDMs for visualizing time-series data. ARGPDMs are stochastic models consisting of a state transition function $f$ in a state space and a function $\pi$ that maps from the state space to the parameter space of a vector autoregressive model. They suppress over-learning by treating the parameters that determine $f$ and $\pi$ as stochastic variables. In ARGPDMs, each single state is associated with a vector auto-regressive model and can represent time-series data whose expectation value changes over time by a single state. Therefore, if ARGPDMs are used to visualize time-series data, state transition occurs when the vector auto-regressive model switches. Thus, smoother and more comprehensible visualization can be achieved in comparison with GPDMs.

In ARGPDMs, the distributions of the state matrix $Z$ and the parameter $\theta$ are assumed to take the forms of Eqs. (12)–(14), and (16), as in GPDMs. Moreover, the observation data set $y_i$ is assumed to be generated according to the vector auto-regressive model that uses the past values $y_{i-1}, \ldots, y_{i-p}$ in the latest period $p$ and the base function $\phi_{y_i}$.

$$y_i = \sum_{q=1}^{p} \Pi_{q,i} \phi_{y_q} (y_{i-q}) \quad . . . . . . . . . . . . . . . . . (22)$$

Here, $\Pi_{1,i}, \ldots, \Pi_{p,i}$ are parameter matrices that determine the vector auto-regressive model. In addition, for simplicity, the vector $\phi_{y_q}$ consisting of the past values of $\phi_{y_q}(y_{i-p})$ in the period $p$ is denoted by $x_i$, and the matrix $D_y \times M_p$ consisting of $\Pi_{q,i}$ is denoted by $\Pi_i$. The observation data set $y_i$ is generated according to the following vector auto-regressive model:

$$y_i = \Pi_i x_i \quad . . . . . . . . . . . . . . . . . . . . . . . (23)$$

The matrix $\Pi_i$ is generated according to the following stochastic model:

$$\text{vec}(\Pi_i^T) = \pi(z_i) \quad . . . . . . . . . . . . . . . . . . . . . . . (26)$$

Here, $\text{vec}(\mathbf{B})$ is an operator that stacks the columns of an $N \times M$ matrix $\mathbf{B}$ into an $MN \times 1$ column vector.

$$\text{vec}([b_1 \ldots b_M]) = [b_1^T \ldots b_M^T]^T \quad . . . . . . . . . . . . . . . . . . . . . . . (27)$$
where $\pi$ is a function that maps the state $\mathbf{z}_t$ to the parameter space of the vector auto-regressive model. The function $\pi$ is defined as in Eq. (28) using a linear combination of $M_\pi$ base functions, $\phi_{\pi,1}, \ldots, \phi_{\pi,M_\pi}$.

$$\pi(\mathbf{z}_t) = \sum_{m=1}^{M_\pi} \mathbf{a}_{\pi,m}\phi_{\pi,m}(\mathbf{z}_t) \quad \ldots \ldots \ldots \quad (28)$$

Here, $\mathbf{A}_\pi = [\mathbf{a}_{\pi,1}, \ldots, \mathbf{a}_{\pi,M_\pi}]^T$ is the parameter that determines the function $\pi$.

A stochastic model is defined by assuming a prior distribution of the parameter $\mathbf{A}_\pi$ without directly handling the parameter. This is assumed to be a standard normal distribution to make $\pi$ a smoothly-changing continuous function.

$$p(\mathbf{A}_\pi) = \prod_m N(\mathbf{a}_{\pi,m}|0, I) \quad \ldots \ldots \ldots \quad (29)$$

The distribution of the observation data matrix $\mathbf{Y}$ can be written as follows:

$$p(\mathbf{Y}|\mathbf{Z}, \mathbf{A}_\pi, \pi) = p(\mathbf{Y}_{1:p}|2\pi)^{-\frac{1}{2}(N-p)D_\pi}|\mathbf{K}|^{-\frac{1}{2}D_\pi} \times \exp \left( -\frac{1}{2}(\mathbf{K}^{-1}\mathbf{Y}_{p+1:N}\mathbf{Y}_{p+1:N}^T) \right) \quad (30)$$

$k_{\pi}(\mathbf{y}_i, \mathbf{y}_j)$ and $k_{\pi}(\mathbf{z}_i, \mathbf{z}_j)$ are kernel functions. $\mathbf{Y}_{i:t'} = [\mathbf{y}_i, \ldots, \mathbf{y}_{t'}]^T$ and $\circ$ is Hadamard product. $\mathbf{K}_\pi$ is an $(N-p) \times (N-p)$ matrix with elements $\mathbf{K}_{i,j} = k_{\pi}(\mathbf{y}_{i', \ldots, \mathbf{y}_{j'}}, \mathbf{y}_j)$, where $p + 1 \leq i, j \leq N$, and $k_{\pi}(\mathbf{y}_i, \mathbf{y}_j)$ is a kernel function. $\mathbf{K}_\pi$ is an $(N-p) \times (N-p)$ kernel matrix whose elements are the kernel function $k_{\pi}(\mathbf{z}_i, \mathbf{z}_j)$, where $p + 1 \leq i, j \leq N$.

The kernel functions $k_{\pi}(\mathbf{y}_i, \mathbf{y}_j)$ and $k_{\pi}(\mathbf{z}_i, \mathbf{z}_j)$ are set to be RBF + linear kernel, as in GPDMs, and defined as follows:

$$k_{\pi}(\mathbf{y}_i, \mathbf{y}_j) = \alpha_{y_{\pi}} \exp \left( -\frac{\alpha_{y_{\pi}2}}{2}||\mathbf{y}_i - \mathbf{y}_j||^2 \right) + \alpha_{y_{\pi}4} \delta_{\mathbf{y}_i, \mathbf{y}_j} + \ldots \ldots \ldots \quad (32)$$

$$k_{\pi}(\mathbf{z}_i, \mathbf{z}_j) = \alpha_{z_{\pi}} \exp \left( -\frac{\alpha_{z_{\pi}2}}{2}||\mathbf{z}_i - \mathbf{z}_j||^2 \right) + \alpha_{z_{\pi}4} \delta_{\mathbf{z}_i, \mathbf{z}_j} + \ldots \ldots \ldots \quad (33)$$

Here, $\mathbf{a}_y = (\alpha_{y_{\pi}1}, \ldots, \alpha_{y_{\pi}4})^T$ and $\mathbf{a}_z = (\alpha_{z_{\pi}1}, \ldots, \alpha_{z_{\pi}4})^T$ are hyper-parameters that determine the kernel functions $k_{\pi}(\mathbf{y}_i, \mathbf{y}_j)$ and $k_{\pi}(\mathbf{z}_i, \mathbf{z}_j)$. The prior distributions of the hyper-parameters $\mathbf{a}_y$ and $\mathbf{a}_z$ are assumed to be uninformative prior distributions as follows:

$$p(\mathbf{a}_y) = \prod_{j=1}^{4} \alpha_{y_{\pi}j}^{-1} \quad \ldots \ldots \ldots \quad (34)$$

$$p(\mathbf{a}_z) = \prod_{j=1}^{4} \alpha_{z_{\pi}j}^{-1} \quad \ldots \ldots \ldots \quad (35)$$

The joint distribution of the observation data matrix $\mathbf{Y}$ and the unknown parameter $\theta_{ar} = [\mathbf{Z}, \mathbf{A}_\pi, \mathbf{a}_y, \mathbf{a}_z]$ are defined using Eqs. (12)–(14), (16), and (30)–(35), as follows:

$$p(\mathbf{Y}, \mathbf{A}_\pi) = p(\mathbf{Y}|\mathbf{Z}, \mathbf{A}_\pi, \pi)p(\mathbf{Z}|\mathbf{A}_\pi)p(\mathbf{A}_\pi)p(\mathbf{a}_y) \times p(\mathbf{a}_z) \quad \ldots \ldots \ldots \quad (36)$$

### 3.2. Parameter Estimation

In ARGPDMS, the unknown parameter $\theta_{ar}$ is estimated by maximum a posteriori probability estimation. That is, the negative logarithmic posterior probability $L_{ar}$ of $\theta_{ar}$ with given observation data matrix $\mathbf{Y}$ is maximized to estimate the unknown parameter $\theta_{ar}$. The stochastic model of Eq. (36) yields the negative logarithmic posterior probability $L_{ar}$ as follows:

$$L_{ar} = -\ln p(\theta_{ar}|\mathbf{Y}) = l_y + l_z$$

$$+ \sum_{i=1}^{4} \ln \alpha_{y_{\pi}j} + \sum_{i=1}^{4} \ln \alpha_{z_{\pi}j} + \sum_{i=1}^{4} \ln \alpha_{y_{\pi}j} + \ldots \ldots \ldots \quad (37)$$

$$l_y = \frac{D_y}{2} \ln |\mathbf{K}| + \frac{1}{2} tr(\mathbf{K}^{-1}\mathbf{Y}_{p+1:N}\mathbf{Y}_{p+1:N}^T) \quad \ldots \ldots \ldots \quad (38)$$

$$l_z = \frac{D_z}{2} \ln |\mathbf{K}_z| + \frac{1}{2} tr(\mathbf{K}_z^{-1}\mathbf{Z}_{2:N}\mathbf{Z}_{2:N}^T) + \frac{1}{2} \mathbf{z}_1^T \mathbf{z}_1 \quad \ldots \ldots \ldots \quad (39)$$

Here, const is a constant term independent of the unknown parameter $\theta_{ar}$. The unknown parameter $\theta_{ar}$ is estimated by maximizing the negative logarithmic posterior probability $L_{ar}$ given in Eq. (37) based on the SCG method.

In addition, as in Eq. (19), $l_z$ in Eq. (37) can be considered a penalty term for a non-smooth state transition. Therefore, the parameters are estimated by replacing $l_z$ in Eq. (37) with $l_z, D_y/D_z$ for smoother state transition.

### 4. Computer Simulations

We use two types of time-series data to perform visualization experiments with an ARGPDG: artificial time-series data and indexes adopted to calculate the composite index. To compare the obtained results, we also use a GPDM and perform similar simulations. In the experiments, the order of the state space is set to $D_z = 2$.

#### 4.1. Artificial Time-Series Data

An auto-regressive model with a single variable of the second order given by Eqs. (40)–(42) is used to generate 400 time-series data elements for $t = 1, \ldots, 400$ and perform simulations.

$$y_t = \Pi_{t,1}y_{t-1} + \Pi_{t,2}y_{t-2} + n_{y,t} \quad \ldots \ldots \ldots \quad (40)$$

$$y_{t,1}, y_{t,2} \sim \text{i.i.d. } N(0, 10^{-1}) \quad \ldots \ldots \ldots \quad (41)$$

$$n_{y,t} \sim \text{i.i.d. } N(0, 10^{-3}) \quad \ldots \ldots \ldots \quad (42)$$

Here, the parameters $\Pi_{t,1}$ and $\Pi_{t,2}$, which circulate with a period of 100 data elements, are given by

$$\Pi_{t,1} = \cos \left( \frac{2\pi t}{100} \right) \quad \ldots \ldots \ldots \quad (43)$$
Visualizing States of Time-Series Data by ARGPDM

**Table 1.** Average misclassification rate of artificial time-series data.

<table>
<thead>
<tr>
<th></th>
<th>ARGPDM</th>
<th>GPDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCR</td>
<td>3.81%</td>
<td>41.94%</td>
</tr>
</tbody>
</table>

Hence, it is difficult to visually recognize the cyclicity of the auto-regressive model that generates the data. Moreover, it fails to group $\mathbf{z}_t$ for each class. Therefore, it is difficult to visually recognize that the artificial time-series data explain the cyclicity of the auto-regressive model that generates the data only with the given explanatory variables.

**Table 1** shows the average misclassification rates of the two models. The average misclassification rate is calculated by randomly choosing 70% of the subscripts 1, . . . , $N$ of the states $\mathbf{z}_1, \ldots, \mathbf{z}_N$ as $T_{\text{learn}}(s)$, the remaining 30% as $T_{\text{test}}(s)$, and taking the average over 100 misclassification rates $\text{MCR}(s)$ of $s = 1, \ldots, 100$. The misclassification rate $\text{MCR}(s)$ is calculated for the test data $\{\mathbf{z}_t | t \in T_{\text{test}}(s)\}$ for which 1-nearest neighbor is developed using $\{\mathbf{z}_t | t \in T_{\text{learn}}(s)\}$ as learning data. That is, the misclassification rate $\text{MCR}(s)$ is calculated using Eq. (46) and is used to evaluate ease of classification of the visualization results.

$$\text{MCR}(s) = \frac{1}{\#T_{\text{test}}(s)} \sum_{t \in T_{\text{test}}(s)} \left[ c_t \neq c_{\text{NN}_t(s)} \right] \quad \ldots \quad (46)$$

$$\text{NN}_t(s) = \arg \min_{t' \in T_{\text{learn}}(s)} \| \mathbf{z}_t - \mathbf{z}_{t'} \|^2 \quad \ldots \quad (47)$$

Here, $\left[ \cdot \right]$ is an indicator function that is equal to 1 when the argument is true and 0 otherwise, and $\#T$ is the number of elements of the group $T$. **Table 1** shows that the average misclassification rate is lower in the ARGPDM than in the GPDM.

### 4.2. Indexes Adopted for the Composite Index

The composite index was released by the Cabinet Office and created to study current business conditions by integrating business activity indexes that are important and sensitive to these conditions. For the visualization experiments, we generated time-series data with five variables by choosing five indexes out of those used in the calculation of the composite index (CI): Index of Industrial Production (Mining and Manufacturing), Index of Producer’s Shipments (Producer Goods for Mining and Manufacturing), Index of Non-Scheduled Worked Hours (Industries Covered), Operating Profits (All Industries), and Index of Shipment in Small and Medium Sized Enterprises. For these indexes, data of 273 months from October 1993 to June 2016 were used. **Fig. 3** shows these indexes and colored markers for different class labels $c_t \in \{1, 2\}$. The class label $c_t$ is 1 in the business expansion period determined by the Cabinet Office and 2 in the recession period. In the experiments, the order of the auto-regression of the
ARGPDM was set to $p = 1$.

Figure 4 shows the visualization of the indexes using $\mathbf{z}_t$. The visualization is performed with different colors, depending on the class labels $c_t$. It is seen that in the ARGPDM, the states $\mathbf{z}_t$ are grouped for each business expansion and regression period. Therefore, it can be visually recognized that the indexes explain a change of the business condition only with the given explanatory variables. By contrast, in the GPDM, $\mathbf{z}_t$ of the business expansion periods and the regression periods are scattered and cannot be grouped. Consequently, it is difficult to visually recognize that the indexes explain a change of the business condition only with the given explanatory variables.

Table 2 shows the average misclassification rates of the two models. It is seen that the average misclassification rate is lower in the ARGPDM than in the GPDM.

Moreover, it can be seen in Fig. 3 that the adopted indexes decreased in the regression period A (February 2008 to February 2009, Global Financial Crisis) more drastically than in the other regression periods. The result of the ARGPDM in Fig. 4 shows that $\mathbf{z}_t$ changes along a trajectory with a larger radius in period A than in the other regression periods, visually indicating a drastic change in that period. Likewise, the result of the GPDM shows that $\mathbf{z}_t$ changes along a trajectory with a larger radius in period A than in the other regression periods.

### 5. Conclusions

An ARGPDM was proposed that associated a state with a vector auto-regressive model. Thus, ARGPDMs could represent observation changes over time by a single state and provide a more comprehensible visualization with smoother state transition than GPDMs. In computer simulations, artificial time-series data and indexes adopted for the composite index were used for visualization experiments. It was confirmed that visualization using ARGPDMs allowed for easier classification with lower average misclassification rate compared with the conventional GPDMs.

We conclude by summarizing certain issues in regard to the proposed method. It should be noted that the focus was on a method for visualizing time-series data used for learning rather than a method for visualizing test data, which are not used for learning, on the basis of parameters after learning. Visualization of test data should be considered in future work. In addition, the focus was on visualization of time-series data rather than forecast, smoothing, or compensation of deficits of the data. These issues should be considered as well.

### References:


Appendix A. Derivation of Eq. (30)

Herein, the $d_i$-th column vector of the matrix $Y_{p+1:N}$ is written as $y_{(d_i)}$. Moreover, we define $\Phi_{\pi}(z_i) = (\Phi_{\pi,1}(z_i), \ldots, \Phi_{\pi,M_t}(z_i))^T$ and divide the matrix $A_{\pi}$ into $D_p$ submatrices, i.e., $D_p \times p$ column matrices), denoted by $A_{\pi,1}, \ldots, A_{\pi,D_p}$. Namely, we have

$$A_{\pi} = \begin{bmatrix} A_{\pi,1} & \cdots & A_{\pi,D_p} \end{bmatrix}$$

where $y_{(d_i)}$ is calculated from the stochastic model in Section 3.1 as follows:

$$y_{(d_i)} = Wm_{d_i} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

Here we have

$$W = \begin{bmatrix} \text{vec}(\Phi_{\pi}(z_{p+1}))^T \ldots \text{vec}(\Phi_{\pi}(z_N))^T \end{bmatrix}^T$$

$$m_{d_i} = \text{vec}(A_{\pi,d_i}) \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

Using Eqs. (29) and (49), the left-hand side of Eq. (30) becomes

$$p(Y | Z, \alpha_y, \alpha_{\pi})$$

$$= p(Y_{1:p}) \prod_{t=p+1}^N p(y_t | x_t, z_t, \alpha_y, \alpha_{\pi})$$

$$= p(Y_{1:p}) \prod_{t=p+1}^N \prod_{d_i=1}^{D_p} p(y_t | x_t, z_t, \alpha_y, \alpha_{\pi})$$

$$= p(Y_{1:p}) (2\pi)^{-\frac{1}{2}(N-p)D_y} |WW^T|^{-\frac{1}{2}}$$

$$\times \exp \left( -\frac{1}{2} \text{tr}((WW^T)^{-1}Y_{p+1:N}Y_{p+1:N}^T) \right)$$

$WW^T$ in Eq. (52) can be written as follows:

$$WW^T = \left\{ \sum_{q=1}^{p} \Phi_{\pi,1-p+q,N-q}^T \Phi_{\pi,1-p+q,N-q} \right\} = \Phi_{\pi}^T \Phi_{\pi}$$

where

$$\Phi_{\pi} = [ \Phi_{\pi}(z_{p+1}) \cdots \Phi_{\pi}(z_N)]^T$$

Equation (30) can be obtained from Eqs. (52) and (53) by setting $\Phi_{\pi}(y_t)^T \phi_{\pi}(y_t)$ to $k_\pi(y_t, y_t)$ and $\Phi_{\pi}(z_t)^T \phi_{\pi}(z_t)$ to $k_\pi(z_t, z_t)$.