Smoothness Prior Approach to Explore Mean Structure in Large-scale Time Series

Genshiro Kitagawa, Tomoyuki Higuchi and Fumiyo N. Kondo

The Institute of Statistical Mathematics,
4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569 Japan

Abstract

This article is addressed to the problem of modeling and exploring mean value structure of large-scale time series data and time-space data. A smoothness prior modeling approach [13] is taken. In this approach, the observed series are decomposed into several components each of which are expressed by smoothness priors models. In the analysis of POS and GPS data, various useful information were extracted by this decomposition, and result in discoveries in these areas.

Key words: Smoothness priors, State space model, Time series, Space-time data, Data mining, Seasonal adjustment, POS, GPS

1 Introduction

In statistical information processing, introduction of the information criterion AIC [1,19] facilitated to compare various types of statistical models freely and changed the conventional paradigm of statistical research which consisted of estimation and statistical test. It revealed the importance of proper statistical modeling, and the use of parametric models become very popular since then [4,14]. AIC criterion suggests that if the available data set is short, we have to use simpler model to obtain reliable information from that data. However, by the progress of various measuring devices, it has become possible to use huge amount of data in various fields of sciences and societies. In this situation, a more important problem is to extract useful information from huge amount

Email address: {kitagawa, higuchi, kondo}@ism.ac.jp (Genshiro Kitagawa, Tomoyuki Higuchi and Fumiyo N. Kondo).

URL: http://www.ism.ac.jp/{~kitagawa/, ~higuchi} (Genshiro Kitagawa, Tomoyuki Higuchi and Fumiyo N. Kondo).

of data, which is difficult to achieve by a simple parametric model. Namely, in this situation, modeling with small number of parameters is sometimes inadequate and a more flexible tool for extracting useful information from data is necessary.

In an analysis of input-output relationship of econometric time series, Shiller [20] introduced the notion of "smoothness priors", and considered constrained least squares problem. A similar concept has already appeared in Whittaker [22] addressing a problem of the estimation of a smooth trend. The trade-off parameters were determined subjectively until Akaike [2,3] proposed the method of choosing the trade-off parameters or hyperparameters in a Bayesian framework, by maximizing the likelihood of a Bayes model [18]. The calculation of the likelihood of a Bayes model for time series requires intensive computation, of which burden Gersch and Kitagawa [6] eased by employing a state space representation of the model and recursive algorithm of Kalman filtering [11].

In this paper, we will present applications of this smoothness priors approach for exploring large-scale time series data or space-time data. Specifically, we consider the POS (Point of Sales scanner) data and GPS (Global Positioning System) data, because automatic transaction of these data is one of the most attractive and potential targets in statistical science. By the analyses of these data, it will be shown that by removing trend and seasonal components by a proper smoothness prior modeling, useful information such as the trading day effect (for economic data), competitive relation (for POS data) and local fluctuation associated with an atmospheric condition (for GPS) are discovered.

2 Smoothness Prior Modeling

2.1 Flexible Semi-Parametric Modeling

A smoothing approach attributed to [22], is as follows: Let

$$y_n = f_n + \varepsilon_n, \quad n = 1, ..., N \tag{1}$$

denote observations, where f_n is an unknown smooth function of n, and ε_n is an independently identically distributed (i.i.d.) normal random variable with zero mean and unknown variance σ^2. The problem is to estimate $f_n, n = 1, ..., N$ from the observations, $y_n, n = 1, ..., N$, in a statistically sensible way. Here the number of parameters to be estimated is equal to the number of observations. Therefore, the ordinary least squares method or the maximum likelihood method yield meaningless results. Whittaker [22] suggested that

the solution $f_n, n = 1, ..., N$ balances a tradeoff between infidelity to the data and infidelity to a k-th order difference equation constraint. Namely, for fixed values of λ^2 and k, the solution satisfies

$$\min_f \left[\sum_{n=1}^{N} (y_n - f_n)^2 + \lambda^2 \sum_{n=1}^{N} (\Delta^k f_n)^2 \right]. \tag{2}$$

The first term in the brackets in (2) is the infidelity-to-the-data measure, the second is the infidelity-to-the-constraint measure, and λ^2 is the smoothness tradeoff parameter. Whittaker left the choice of λ^2 to the investigator.

2.2 Automatic Parameter Determination via Bayesian Interpretation

A smoothness priors solution [2] explicitly solves the problem posed by Whittaker [22]. A version of the solution is as follows: Multiply (2) by $-1/(2\sigma^2)$ and exponentiate it. Then the solution that minimizes (2) achieves the maximization of

$$\exp\left\{ -\frac{1}{2\sigma^2} \sum_{n=1}^{N} (y_n - f_n)^2 \right\} \exp\left\{ -\frac{\lambda^2}{2\sigma^2} \sum_{n=1}^{N} (\Delta^k f_n)^2 \right\}. \tag{3}$$

Under the assumption of normality, (3) yields a Bayesian interpretation

$$\pi(f|y, \lambda^2, \sigma^2, k) \propto p(y|\sigma^2, f)\pi(f|\lambda^2, \sigma^2, k), \tag{4}$$

where $\pi(f|\lambda^2, \sigma^2, k)$ is the prior distribution of f and $p(y|\sigma^2, f)$ the data distribution, conditional on σ^2 and f, and $\pi(f|y, \lambda^2, \sigma^2, k)$ the posterior of f. Akaike [2] obtained the marginal likelihood for λ^2 and k by integrating (3) with respect to f. This facilitates an automatic determination of the tradeoff parameters in constrained least squares which has been treated subjectively for many years and eventually led to the frequent use of Bayesian method in statistical and information science communities. Several interesting applications of this method can be seen in [4].

2.3 Time Series Interpretation and State Space Modeling

Consider a problem of fitting polynomial of order $k - 1$ defined by

$$y_n = t_n + \varepsilon_n, \quad t_n = a_0 + a_1 n + \cdots + a_{k-1} n^{k-1}, \tag{5}$$

where $\varepsilon_n \sim N(0, \sigma^2)$. It is easy to see that this polynomial is the solution to the difference equation

$$\Delta^k t_n = 0, \tag{6}$$

with appropriately defined initial conditions. This suggests that by modifying the above difference equation so that it allows for a small deviation from the equation, namely by letting $\Delta^k t_n \approx 0$, it might be possible to obtain a more flexible regression curve than the usual polynomials. A possible formal expression is the stochastic difference equation model

$$\Delta^k t_n = v_n, \tag{7}$$

where $v_n \sim N(0, \tau^2)$ is an i.i.d. Gaussian white noise sequence. For small noise variance τ^2, it reasonably expresses our expectation that the noise is mostly very "small" and with a small probability it may take a relatively "large" value. Actually, the solution to the model is, at least locally, very close to a $(k-1)$-th order polynomial. However, globally a significant difference arises and (7) can express a very flexible function. For $k = 1$, it is locally constant and becomes a well-known random walk model, $t_n = t_{n-1} + v_n$. For $k = 2$, the model becomes $t_n = 2t_{n-1} - t_{n-2} + v_n$ and the solution is a locally linear function.

The models (5) together with (7) can be expressed in a special form of the state space model

$$\begin{aligned} x_n &= Fx_{n-1} + Gv_n \qquad \text{(system model)}, \\ y_n &= Hx_n + w_n \qquad \text{(observation model)}, \end{aligned} \tag{8}$$

where $v_n \sim N(0, \tau^2)$, $w_n \sim N(0, \sigma^2)$ and $x_n = (t_n, ..., t_{n-k+1})'$ is a k-dimensional state vector, F, G and H are $k \times k$, $k \times 1$ and $1 \times k$ matrices, respectively. For example, for $k = 2$, they are given by

$$x_n = \begin{bmatrix} t_n \\ t_{n-1} \end{bmatrix}, \quad F = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad H = [1, 0]. \tag{9}$$

One of the merits of using this state space representation is that we can use computationally efficient Kalman filter for state estimation. Since the state vector contains unknown trend component, by estimating the state vector x_n, the trend is automatically estimated. Also unknown parameters of the model, such as the variances σ^2 and τ^2 can be estimated by the maximum likelihood method. In general, the likelihood of the time series model is given by

$$L(\theta) = p(y_1, \ldots, y_N|\theta) = \prod_{n=1}^{N} p(y_n|Y_{n-1}, \theta), \tag{10}$$

where $Y_{n-1} = \{y_1, \ldots, y_{n-1}\}$ and each component $p(y_n|Y_{n-1}, \theta)$ can be obtained as byproduct of the Kalman filter [11]. It is interesting to note that the tradeoff parameter λ^2 in the penalized least squares method (2) can be interpreted as the ratio of the system noise variance to the observation noise variance, or the signal-to-noise ratio.

The individual terms in (10) are given by, in general p-dimensional observation case,

$$p(y_n|Y_{n-1}, \theta) = \frac{1}{(\sqrt{2\pi})^p} \left|W_{n|n-1}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\varepsilon'_{n|n-1} W^{-1}_{n|n-1} \varepsilon_{n|n-1}\right\}, \tag{11}$$

where $\varepsilon_{n|n-1} = y_n - y_{n|n-1}$ is one-step-ahead prediction error of time series and $y_{n|n-1}$ and $V_{n|n-1}$ are the mean and the variance covariance matrix of the observation y_n, respectively, and are defined by

$$y_{n|n-1} = H x_{n|n-1}, \tag{12}$$

$$W_{n|n-1} = H V_{n|n-1} H' + \sigma^2. \tag{13}$$

Here $x_{n|n-1}$ and $V_{n|n-1}$ are the mean and the variance covariance matrix of the state vector given the observations Y_{n-1} and can be obtained by the Kalman filter [11].

If there are several candidate models, the goodness of the fit of the models can be evaluated by the AIC criterion defined by

$$\text{AIC} = -2 \log L(\hat{\theta}) + 2(\text{number of parameters}). \tag{14}$$

AIC is derived from an asymptotically unbiased estimate of the expected log-likelihood, or equivalently the Kullback-Leibler information of the model, and the model with the smallest AIC is considered to be the best one [1], [19].

2.4 Modeling of Space-Time Data

Let Z_n^i, $(n = 1, \ldots, N; i = 1, \ldots, I)$ be scalar observation at a discrete time of n for a station (site) i. Along the line mentioned above, we consider the following model to decompose Z_n^i into trend, T_n^i, and irregular component, D_n^i, namely,

$$Z_n^i = T_n^i + D_n^i, \quad D_n^i \sim N(0, \sigma^{2,i}). \tag{15}$$

A direct approach to realize the Bayesian space-time (space-temporal) model is given by considering the following system model for each n

$$T_n^i = 2\,T_{n-1}^i - T_{n-2}^i + E_n^i, \quad E_n^i \sim N(0, \tau^{2,i}) \qquad \text{for } \forall \;\; i, \tag{16}$$

$$T_n^i - T_n^j = V_n^i, \qquad V_n^i \sim N\left(0, (\phi(\Delta^{ij}))^2 s^2\right) \quad \text{for } \forall \; (i,j), \tag{17}$$

where Δ^{ij} is some measure of a distance between station i and j, and ϕ is usually assumed as a linear function truncated at Δ_{th} which is set to be the mean of distance between the neighboring points. Although this approach is desirable from the statistical viewpoint, its numerical realization on computer is impractical due to large memory required for a large number of $I \approx 1,000$ that we usually deal with. For a case with lower dimensional model like $I \leq 100$, a simple approach to deal with $\mathbf{T}_n = [T_n^1, T_n^2, \ldots, T_n^I \mid T_{n-1}^1, T_{n-1}^2, \ldots, T_{n-1}^I]'$ as a state vector can be implemented on a computer with large memory [12].

A simple way to mitigate this computational difficulty in the direct Bayesian approach for a case with $I \approx 1,000$ is to assume that each time series $\mathbf{Z}^i = [Z_1^i, Z_2^i, \ldots, Z_N^i]'$ is mutually independent vector. This assumption allows the smoothness priors approach mentioned earlier to be employed. Then we use the system model given by (16) only. The maximum likelihood estimates for $\sigma^{2,i}$ and $\tau^{2,i}$ are denoted by $\widehat{\sigma}^{2,i}$ and $\widehat{\tau}^{2,i}$, respectively. The Kalman filter and smoother with $\widehat{\sigma}^{2,i}$ and $\widehat{\tau}^{2,i}$ yield the estimates for the trend component, $\widehat{T}_n^i$. The estimated irregular components $\widehat{D}_n^i = Z_n^i - \widehat{T}_n^i$ is called the residual hereafter. A vector of the residual components for a station i is denoted by $\mathbf{D}^i$ and the median of $\widehat{T}_n^i$, for each n, by $\overline{T}_n$. Similarly, their percentile points corresponding to $\pm\sigma$ and $\pm 2\sigma$ intervals of $\widehat{T}_n^i$ versus n are denoted by $T_n^{\pm 1}$ and $T_n^{\pm 2}$, respectively.

The next step for exploring the mean structure of the space-time data is to examine the spatial correlation of the residual components in terms of a correlation coefficient C^{ij} between $\mathbf{D}^i$ and $\mathbf{D}^j$. For a fixed station i, a spatial distribution of C^{ij} as a function of a distance measure Δ^{ij} has to be examined visually. In fact, a large number of C^{ij} hampers such kind of visual examination. Therefore, a plot of C^{ij} versus Δ^{ij} guides us to further improvements on the mean structure of the space-time data. Obviously, when there appear many points with high correlation in the small value of Δ, taking the spatial correlation into account would improve an initial estimate on the mean structure of the space-time data, $\widehat{T}_n^i$. Such kind of improvements can be realized by considering the following smoothness priors model for a spatial data

$$\begin{aligned} &\widehat{T}_n^i = \mu_n^i + U_n^i, \quad U_n^i \sim N(0, r^2) && \text{for } \forall \;\; i, \\ &\mu_n^i - \mu_n^j = V_n^i, \quad V_n^i \sim N(0, (\phi(\Delta^{ij}))^2 s^2) && \text{for } \forall \; (i,j), \end{aligned} \tag{18}$$

where μ_n^i is an improved trend component. The iterative procedure mentioned above is practical for improving the estimates of the mean structure of the space-time data set [9].

3 Applications

3.1 An Illustrative Example: Seasonal Adjustment

The smoothness priors method has been applied to many real world problems [4,13]. Most of the economic time series contain trend and almost periodic components which make it difficult to capture the essential change of economic activities. Therefore in economic data analysis, removal of these effects is important. In our modeling it is realized by the decomposition

$$y_n = t_n + s_n + w_n, \tag{19}$$

where t_n, s_n and w_n are trend, seasonal and irregular components. A reasonable solution to this decomposition was given by the use of smoothness priors for both t_n and s_n [6]. The trend component t_n and the seasonal component s_n are assumed to follow

$$\begin{aligned} t_n &= 2t_{n-1} - t_{n-2} + v_n, \\ s_n &= -(s_{n-1} + \cdots + s_{n-11}) + u_n, \end{aligned} \tag{20}$$

where v_n, u_n and w_n are Gaussian white noise with $v_n \sim N(0, \tau_t^2)$, $u_n \sim N(0, \tau_s^2)$ and $w_n \sim N(0, \sigma^2)$.

We fit this model to BDHWWS (Wholesale Hardware Sales, U.S. Bureau of the Census, January 1967 – February 1989) data. The variance of the irregular component, the log-likelihood and AIC of the model are 0.001193, 454.6 and 3095.2, respectively. Fig. 1A, B and C show the log-transformed original data and the estimated trend, seasonal component and the irregular component, respectively. The estimated seasonal component is very stable over the whole period and the trend clearly captures the depression of the sales in 1975 and 1982. The irregular component is small compared with the seasonal variation. Although, by this seasonal adjustment, it is possible to extract or remove seasonal component, we can extract more information by a smoothness prior modeling. Many of the economic time series related to sales or production are affected by the number of days of the week. For example, the sales of a department store will be strongly affected by the number of Sundays and Saturdays in each month. Such kind of effect is called the trading day effect.

To extract the trading day effect, we consider the decomposition

$$y_n = t_n + s_n + td_n + w_n, \tag{21}$$

where t_n, s_n and w_n are as above and the trading day effect component, td_n, is assumed to be expressed as

$$td_n = \sum_{j=1}^{7} \beta_j d_{jn}, \tag{22}$$

where d_{jn} is the number of j-th day of the week (e.g., j=1 for Sunday and j=2 for Monday, etc.) and β_j is the unknown trading day effect coefficient. To assure the identifiability, it is necessary to put constraint that $\beta_1 + \cdots + \beta_7 = 0$. The variance of the irregular component, the log-likelihood and AIC of the model are 0.000602, 515.4 and 2985.7, respectively. The reduction of the variance and the AIC value clearly indicates the existence of the trading day effect. Fig. 2A shows the estimated trading day effect coefficients $\beta_j, j = 1, \ldots, 7$. It reveals that Sunday (j=1) and Saturday (j=7) have negative effect. This suggests that many wholesale stores are closed on Sunday and Saturday. The coefficients for the weekend are positive. However, those of Monday, Wednesday and Friday are close to zero.

To check the reliability of these coefficients, we considered a constrained model that assumes

$$\beta_1 = \beta_7, \quad \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6. \tag{23}$$

The variance, log-likelihood and AIC of this model are 0.000636, 512.8, 2980.9, respectively. Since AIC of the model is smaller than the former model, it indicates that this constrained model is better than the former one. Namely, the difference of the trading day coefficients within weekdays and also that of Sunday and Saturday are not significant. Fig. 2B shows the trading day coefficients obtained by this model.

Fig. 1D and E show the trading day effect and the irregular component obtained by this model. The trend and seasonal components are visually indistinguishable from the ones shown in Fig. 1A and B. The trading day effect is very small compared with the seasonal variation. However, the irregular component becomes considerably small. Actually the variance of the residual becomes a half. Fig. 1F shows the plot of the seasonal component plus trading day effect. Comparing with Fig. 1A, it can be seen that the seasonal component plus trading day effect reproduces the detailed behavior of the series.

Since the numbers of day of the week are completely determined by the cal-

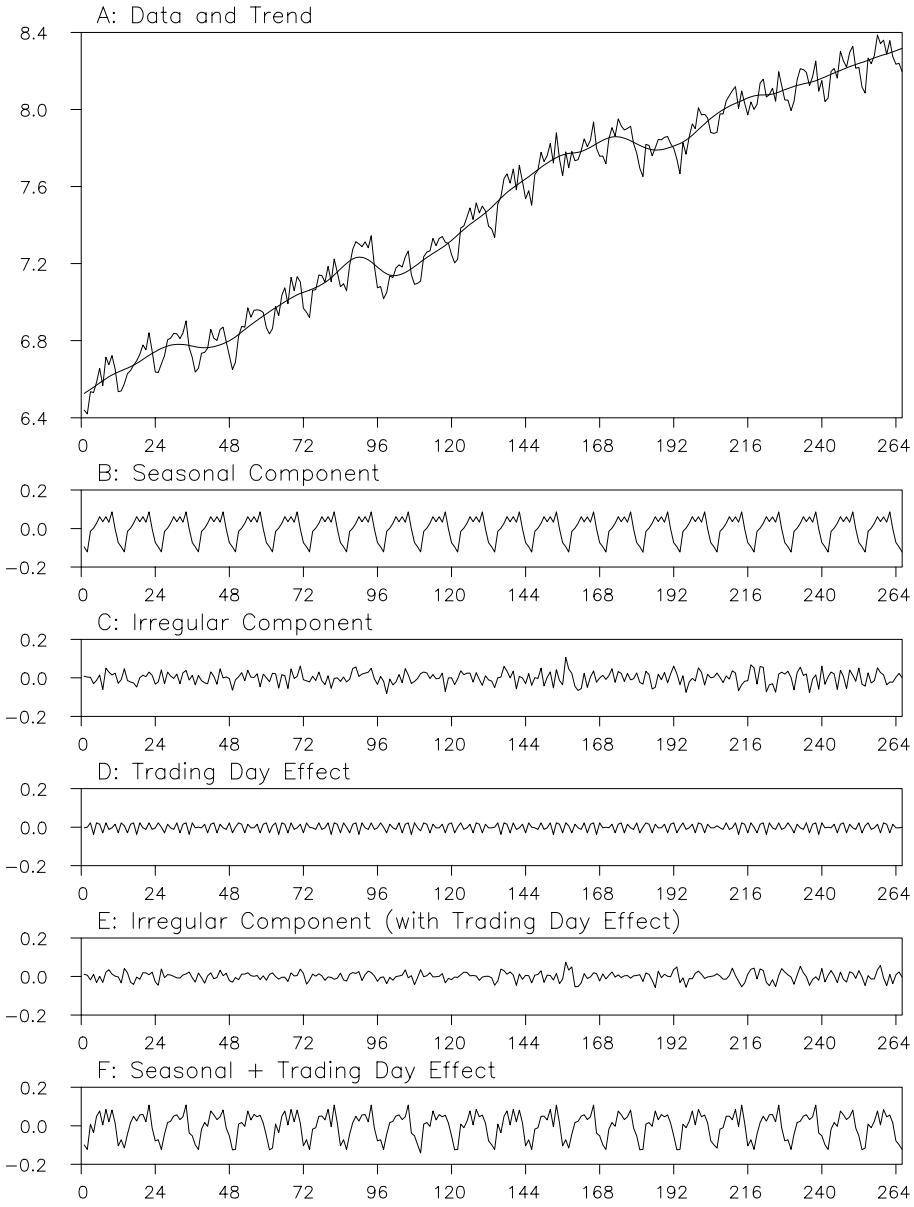

Fig. 1. Seasonal adjustment of BDHWWS.

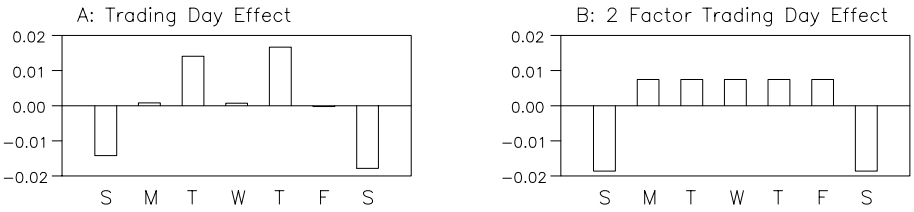

Fig. 2. Trading day effect coefficients. A: 7-factor model, B: 2-factor model.

endar, if we obtain good estimates of the trading day effect coefficients, then it will greatly contribute to the increase of prediction ability.

Similar decomposition methods are developed for the analysis of earth tide

data and groundwater data. In these applications, the time series is decomposed as

$$y_n = t_n + p_n + e_n + r_n + w_n, \tag{24}$$

where p_n, e_n and r_n are the barometric air pressure effect, the earth tide effect and the precipitation effect, respectively [4]. By the decomposition of 10 years groundwater data with this model, the effects of earthquakes are clearly detected, and various knowledge on the relation between occurrence of earthquakes and the groundwater level are obtained [14,15].

3.2 Analysis of POS Data

Analysis of Point-of-Sales (POS) scanner data is an important research area of "data mining" and discovery science, which may provide store managers with useful information to control price or stock levels of goods. The effect measurements responding to price changes and semi-automatic sales forecasts of each brand may be useful in order to pursue price promotions efficiently and reduce the risk of "dead-stock" or "out-of-stock".

POS data set consists of a huge number of items and the analyses so far are mostly concentrated on the detection of mutual relation between items. In this subsection, we will show that, by the smoothness prior modeling of multivariate time series which takes into account of various components such as long term baseline sales trend, weekly pattern and competitive effects, it is possible to discover the effect of temporary price-cut and competitive relation between several items.

Assume that $y_n = [y_n^{(1)}, \ldots, y_n^{(\ell)}]'$ denotes ℓ dimensional time series of sales of a certain product category, and $p_n = [p_n^{(1)}, \ldots, p_n^{(\ell)}]'$ the covariate expressing the price of each brand. The generic model we consider here for the analysis of POS data is given by

$$y_n = t_n + d_n + x_n + w_n, \tag{25}$$

where t_n, d_n, x_n and w_n are the baseline sales trend, weekly pattern, sales promotion effect, and observation noise, respectively. Each component of the baseline sales trend, $t_n^{(j)}$, is assumed to follow the first order trend model

$$t_n^{(j)} = t_{n-1}^{(j)} + u_n^{(j)}. \tag{26}$$

The weekly pattern, $d_n^{(j)}$, can be considered as a special form of seasonal component with period length 7 and is assumed to follow

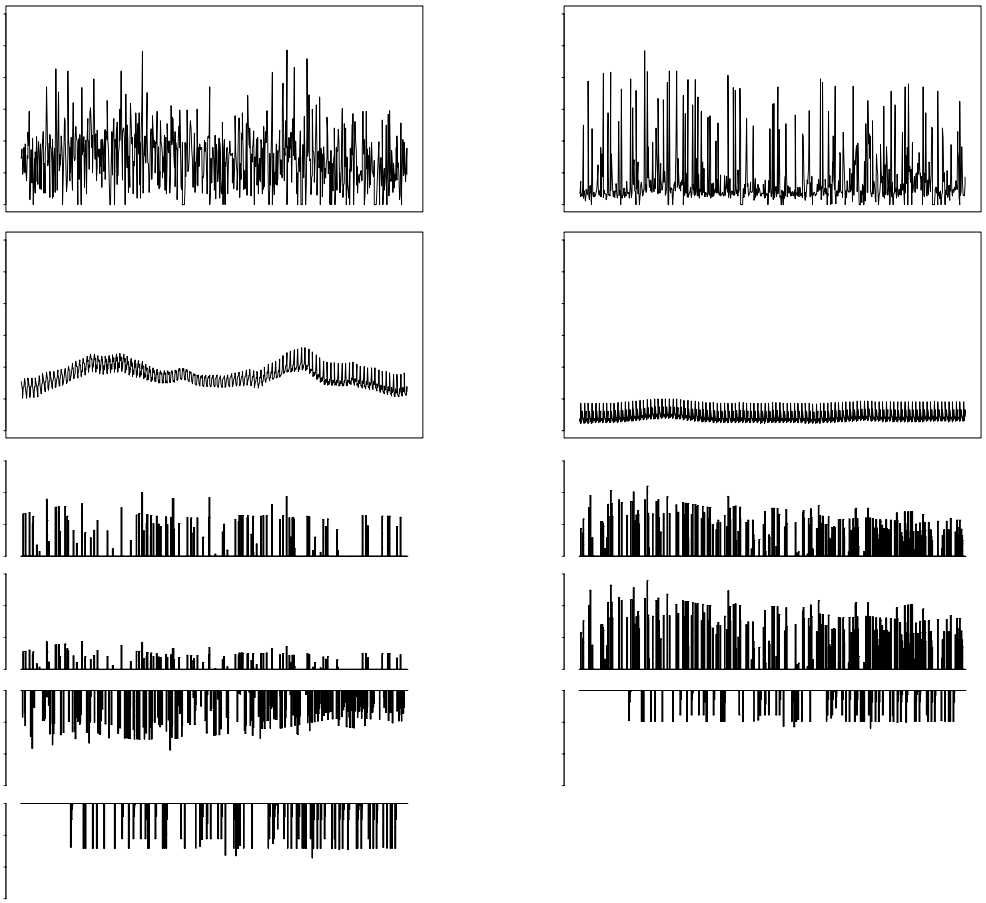

Fig. 3. Decomposition of the brand B1 (left) and B2(right). Top plots: observed series, second plots: baseline trend plus weekly pattern, third plots: category expansion, fourth to sixth plots: brand substitution, positive due to own price-cuts and negative due to competitors' price-cuts.

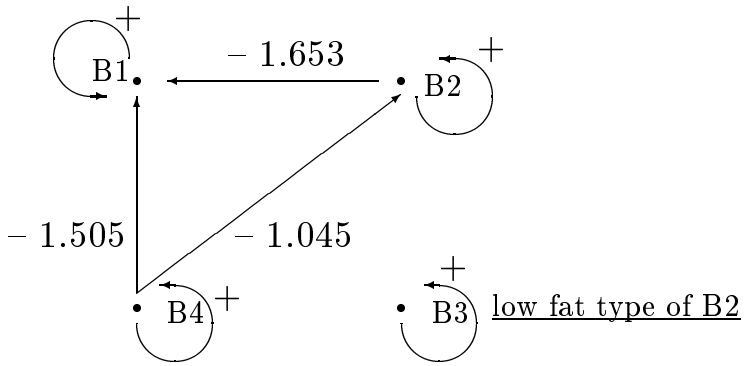

Fig. 4. Competitive relationships between four brands.

$$d_n^{(j)} = -(d_{n-1}^{(j)} + \cdots + d_{n-6}^{(j)}) + v_n^{(j)}. \tag{27}$$

The price promotion effect is assumed to be expressed by a linear function of nonlinear transformation of the price (price function)

$$x_n = B_n f(p_n). \tag{28}$$

In the analysis that follows, we assume that the price function is given by

$$f(p_n)^{(j)} = \exp\{-\gamma(n - n_0)\}\, I_A\left(\Delta p_n^{(j)} - c_n^{(j)}\right) \tag{29}$$

where $\Delta p_n^{(j)}$ denotes the temporary price-cut from its regular (precisely the maximum) price, γ, a parameter, n_0 a starting point of price-cut, $c_n^{(j)}$ a condition that a price-cut is effective to cause sales increases, and $I_A(\ )$ an indicator function. In actual modeling, this price promotion effect is further decomposed into $x_n = g_n + z_n$, where g_n is the category expansion effect and corresponds to the contribution to the increase of total sales. On the other hand, z_n is the brand switch effect which is the increase of the sales of a brand obtained at the expense of the decrease of other brands and does not contribute to the increase of category total.

This model can be conveniently expressed in linear state space model and thus the numerically efficient Kalman filter can be used for state estimation, namely for the decomposition into components, and parameter estimation. Within various possible candidate models, the best model was found by the AIC criterion.

The presented model was applied to scanner data sets of daily milk category, for the period of 1994/2/28 – 1996/3/3 (N=735). Five-variate series consisted of top four brands and the others total were analyzed. Only two brands B1 and B2 are shown on top of Fig. 3. The second plots show the estimated baseline trend components plus the estimated weekly pattern. Only about 20% of the variation of the original series is explained by this day of the week effect. However, for other stores where the prices of brands did not change so significantly, the weekly pattern contribute much more than this present case.

Fig. 4 shows the detected competitive relationship among four major brands discovered via identified model. The competitive coefficients are shown as well. The brand, B3, is a low fat type of B2 and is identified to be independent of other brand's price promotion due to the type difference. Price-cut of B4 (B2) increases sales of B4(B2), but reduces those of B1 and B2 (B1). Price-cut of B1 increases sales of B1 and does not affect the sales of the competitive three brands.

The decomposition of price promotion effect into brand switch effect and category expansion effect is achieved by using the category total sales instead of the others total sales with a zero constraint on the brand switch effect of the category total for each price function. The third plots of Fig. 3 show the estimated category expansion effect. The price-cuts of B1 and B2 contribute to the expansion of category total. The fourth plots show the estimated brand switch components, being positive by own price-cuts, and the fifth or sixth plots, being negative by the competitors' price-cuts.

The brand switch components of B1 and B2 are quite different. The plot for B1 indicates that the price-cut of B1 slightly contributes to the expansion of its own sales. However, B1 is vulnerable to the price-cut of B2 (see the fifth plots) and B4 (see the sixth plots). On the other hand, the price-cut of B2 considerably contributes to the increase of the own sales (see the fourth plots) and B2 is slightly affected by the price-cut of B4 (see the fifth plots) .

3.3 Analysis of GPS Data

The GPS (Global Positioning System) is one of the most interesting and important data set which allows us to investigate a global change in environment precisely. Its high precision information on positions of permanent stations can be supplied by signal processing of microwave signal from GPS satellite. Several physical quantities of media existing between the GPS satellite and ground stations, such as electron, water vapor, and so on, affect phase information of microwave signals and result in propagation delays [5,8,10]. Therefore, a careful treatment of propagation delays is required to extract reliable information as to measurements of the positions.

Dominant sources to bring about propagation delays are (1) ionosphere origin and (2) troposphere origin, such as atmospheric pressure and atmospheric water vapor [21]. The propagation delay generated by the atmospheric water vapor, called the wet delay, is most difficult to evaluate among these factors. A good estimation on the propagation delay can be given to the ionosphere origin and atmospheric pressure origin sources, by utilizing other physical quantities measured simultaneously. As a result, the wet delay turns out to appear as "noise source" in the processing of the GPS data and has to be subtracted prior to diagnosing the GPS data in terms of information on positions.

In Japan, considerable efforts to establish a nationwide GPS array has been kept making by the Geographical Survey Institute of Japan (GSI) [7]. The Japanese GPS array is characterized by its high spatial resolution; the array is composed of nearly one thousand stations separated typically by 15-30 km from one another [21]. Then, a proper processing of the GPS data set taking

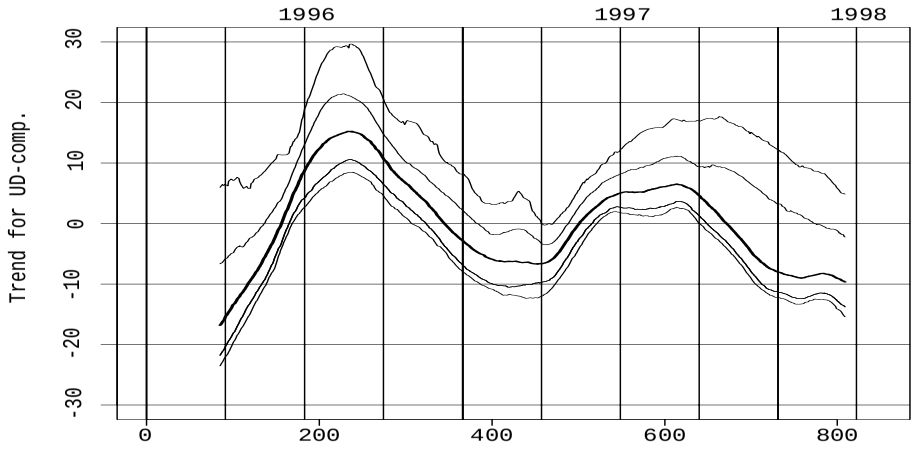

Fig. 5. The median, $\pm 1\sigma$, and $\pm 2\sigma$ percentile points of the estimated trend of the up-down component versus n, $\overline{T}_n, T_n^{\pm 1}$, and $T_n^{\pm 2}$.

the wet delay effect into account allows us to estimate a high-frequent spatial pattern of the atmospheric water vapor, in particular, precipitable water vapor (PWV) which plays an important role in forecasting a weather map. Actually, an approach to extract information concerning the PWV from the GPS data draws much attention in a field of the meteorology and now is referred to as the GPS meteorology [5,8,21].

Many previous works to infer a quantitative relationship between the PWV and GPS data used the hourly GPS data sequences for some special events in limited local areas, because an association of space-time variation of the GPS data with other information obtained by radar echo and radiosonde measurements would be useful and direct approach [8,10,21]. Our objective in this study is aimed at finding empirical rules to give a quantitative description for the relationship between the fluctuations observed in the GPS data and PWV. We begin with a statistical analysis of the daily GPS array data provided by the GSI. Let $\mathbf{U}_n^i$ be the nth day starting from January 1st, 1996 at the station (site) i:

$$\mathbf{U}_n^i = [X_n^i, Y_n^i, Z_n^i],' \quad i = 1, \ldots, I; n = N_s^i, \ldots, N_e^i \tag{30}$$

where X, Y, and Z correspond to the north-south, east-west, and up-down components, respectively. N_s^i and N_e^i represent the starting and last date of the GPS data available to us now. I is the number of stations.

Our preparatory analysis shows that the fluctuations associated with the PWV are most clearly seen in the up-down component, Z_n^i, among the three components. Then, in this study, we focus on the up-down component Z_n^i. Unfortunately, the original GPS array data contains the outliers as well as the missing observations. These unsatisfactory cases can be easily treated by smoothness

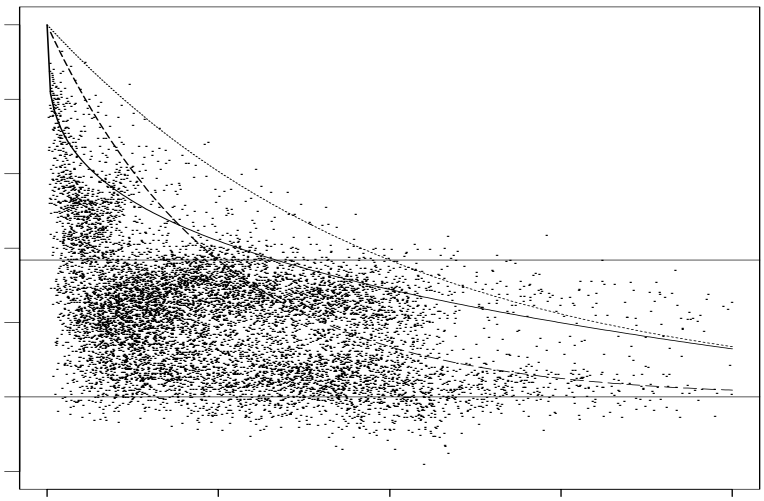

Fig. 6. Plots of Δ^{ij} versus C^{ij}.

priors approach with the state space model, presented in Sect. 2.3, which provides us with the reasonable interpolated data (see [4] for details). Denoising procedure based on another modeling approach has been proposed to deal with an identification of outliers and discontinuities and has produced the similar estimates on T_n^i [16]. The interpolation allows us to determine $\overline{T}_n$, $T_n^{\pm 1}$, and $T_n^{\pm 2}$ systematically. In Fig. 5 we show $\overline{T}_n$, $T_n^{\pm 1}$, and $T_n^{\pm 2}$ obtained by applying the smoothness priors approach to Z_n^i. A seasonal pattern, which is expected to be associated with the PWV, is clearly seen in this figure. A spatial distribution of $\overline{T}_n$ can be illustrated by a GIF animation (`http://www.ism.ac.jp/~higuchi/GPS/SpaAll.gif`). In addition, a relatively significant amplitude of the seasonal variation is found to be larger than the typical amplitude of the residuals, which can be approximated by a mean of the standard deviation of $\mathbf{D}^i$. Therefore, it is apparent that an extraction of precise information on the position from the GPS array requires an elimination of an effect of the PWV from the GPS data. A power spectrum analysis is performed on the $\overline{T}_n$ component and find no eminent peak except for a yearly cycle in a frequency domain. A detail investigation is being made on this figure to discover with what factors is associated from the viewpoint of a climatology.

Fig. 6 shows a plot of Δ^{ij} versus C^{ij}, where a unit of Δ is degree; roughly speaking, a distance of a degree corresponds to 111 km. In this figure, only 10,000 points that are randomly drawn from about $180,000$ C^{ij} are shown for the sake of reducing a file size for this figure. An appearance of many points with high correlation in a small value of Δ clearly suggests that a residual sometimes shows a similar fluctuation with that in the neighboring stations.

Three lines superposed on this figure are:

$$\begin{aligned} C(\Delta) &= \exp\left(-\tfrac{\Delta}{7}\right) && \textbf{[Exp. dacay type]} && \text{(Thin line)}, \\ C(\Delta) &= (0.82)^{\Delta} && \textbf{[AR type]} && \text{(Broken line)}, \\ C(\Delta) &= 1 - 0.36 \cdot (\Delta)^{0.29} && \textbf{[Long Memory type]} && \text{(Thick line)}. \end{aligned} \tag{31}$$

The horizontal line indicates a value of $1/e$. Each curve represents a correlation function induced from a model denoted in bold face. The thin and thick lines are drawn so as to have them resemble an envelope of the upper bound and $+2\sigma$ percentile as a function of Δ in a range of $\Delta \leq 10$. A good agreement of the thick line to $+2\sigma$ envelope would imply that a long-memory-type spatial correlation ($H \sim 0.15$) [17] happens to be observed for an atmospheric spatial pattern. An examination of the weather map for these cases is interesting, but the detailed discussion will be left to other places.

4 Conclusion

The key to the success of a statistical procedure is the appropriateness of the model used in the analysis. The smoothness prior approach facilitates to develop various types of models based on prior information on the subject and the data. In this paper, we applied the smoothness priors method for the modeling of large scale time series and space-time data with mean value structure and competitive relations between variables. In the analysis of POS data, the time series is decomposed into several components and various knowledge to make a strategy concerning price promotion and risk control of dead-stock are obtained. By the analysis of GPS data, a useful information for making a conjecture on the relationship between the propagation delay and PWV is successfully extracted based on the detailed investigation on the trend and residual components.

Softwares based on the smoothness prior approach are available at the following Web sites. The seasonal adjustment method discussed in section 3.1 can be directly performed on Web-Decomp (http://www.ism.ac.jp/~sato) without installing any software. Some other models based on the smoothness prior approach or state space modeling can be run on the Web site of Institute of Geoscience, National Institute of Advanced Industrial Science and Technology (http://150.29.8.26/GSJ_E/analysis/index.html).

Acknowledgments. The GPS array data were provided by the Geographical Survey Institute of Japan (GSI). One of the authors (T.H.) thanks to Dr. S. Miyazaki (GSI) for his help to understand the original data

structure. One of the authors (F.N.K) wishes to thank Mr. Ono and Mr. Nishiyama, The Distribution Systems Research Institute, for providing us with valuable daily scanner data.

References

[1] Akaike, H.: A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **AC-19**, 716–723 (1974)

[2] Akaike, H.: Likelihood and the Bayes procedure (with discussion), in *Bayesian Statistics*, edited by J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F.M. Smith, University press, Valencia, Spain, 143–166 (1980)

[3] Akaike, H.: Seasonal adjustment by a Bayesian modeling, *J. Time Series Analysis*, **1**, 1–13 (1980)

[4] Akaike, H., and Kitagawa, G. eds.: *The Practice of Time Series Analysis*, Springer-Verlag New York (1999)

[5] Bevis, M., Businger, S., Herring, T. A., Rocken, C., Anthes, R. A., and Ware, R. H.: Remote sensing of atmospheric water vapor using the Global Positioning System, *J. Geophysical Research*, **97**, 15,787–15,801 (1992)

[6] Gersch, W., and Kitagawa, G.: The prediction of time series with trends and seasonalities, *Journal of Business & Economic Statistics*, **1**, No. 3, 253–264 (1983)

[7] Heki, K., Miyazaki, S., and Tsuji, H.: Silent fault slip following an interplate thrust earthquake at the Japan trench, *Nature*, **386**, 595–598 (1997)

[8] Heki, K., Kato, T., Rizos, C., and Xu, P. eds.: Application of GPS and other space geodesic techniques to Earth Sciences (1), *Earth, Planets and Space*, **52**, No.10, (2000)

[9] Higuchi, T.: A method to separate the spin synchronized signals using a Bayesian approach (in Japanese with English Abstract), *Proceedings of the Institute of Statistical Mathematics*, **41**, 115–130 (1993)

[10] Hirahara, K., Ivins, E.R., Saito, A., and Tsuda, T. eds.: Application of GPS and other space geodedic techniques to Earth Sciences (2), *Earth, Planets and Space*, **52**, No.11, (2000)

[11] Kalman, R.E.: A new approach to linear filtering and prediction problems, *Transactions on American Society for Mechanical Engineers, Journal of Basic Engineering*, **82**, 35–45 (1960)

[12] Kashiwagi, N.: On the use of the Kalman filter for spatial smoothing, *Annals of the Institute of Statistical Mathematics*, **45**, 21–34 (1993)

[13] Kitagawa, G. and Gersch, W.: *Smoothness Priors Analysis of Time Series*, Lecture Notes in Statistics, No. 116, Springer-Verlag, New York (1996)

[14] Kitagawa, G. and Higuchi, T.: Automatic transaction of signal via statistical modeling, *The Proceedings of The First International Conference on Discovery Science*, Springer-Verlag Lecture Notes in Artificial Intelligence Series, 375–386 (1998)

[15] Kitagawa, G. and Matsumoto, N.: Detection of coseismic changes of underground water level, *Journal of the American Statistical Association*, **91**, No. 434, 521–528 (1996)

[16] Li, J., Miyashita, K., Kato, T., and Miyazaki, S.: GPS time series modeling by autoregressive moving average method: Application to the crustal deformation in central Japan, *Earth, Planets and Space*, **52**, No.3, 155–162 (2000)

[17] Mandelbrot, B.: *Fractals: Form, Chance and Dimension*, Freeman, San Francisco (1977)

[18] Rissanen, J.: Modeling by shortest data description, *Automatica*, Vol. 14, 465–471 (1978)

[19] Sakamoto, Y., Ishiguro, M. and Kitagawa, G.: *Akaike Information Criterion Statistics*, D-Reidel, Dordlecht, (1986)

[20] Shiller, R.: A distributed lag estimator derived from smoothness priors, *Econometrica*, **41**, 775–778 (1973)

[21] Tsuda, T., Heki, K., Miyazaki, S., Aonashi, K., Hirahara, K., Nakamura, H., Tobita, M., Kimata, F., Tabei, T., Matsushima, T., Kimura, F., Satomura, M., Kato, T., and Naito, I.: GPS meteorology project of Japan — Exploring frontiers of geodesy—, *Earth, Planets and Space*, **50**, No.10, i–v (1998)

[22] Whittaker, E.T: On a new method of graduation, *Proc. Edinborough Math. Assoc.*, **78**, 81–89 (1923)