

Sparse VARX model with Kalman-Smoother for metabolomics

DESHUAI SU^{1,a)} SHIGEYUKI OBA^{1,b)} MASANORI KOYAMA¹ MASASHI FUJII² SHINYA KURODA²
SHIN ISHII¹

Abstract: Dynamics of blood chemicals often reflects metabolic states of patients and has hence been broadly used as disease markers, e.g. oral glucose tolerance test is a standard method to diagnose diabetes that is typically caused by pancreatic insufficiency. In order to reveal metabolic system involving many chemical species, and its variability among patients, we need statistical analysis of multi-subject and multi-condition observations. In this study, we propose a sparse VARX model with standard Kalman-Smoother method. The sparse Vector Auto-Regressive with exogenous input (VARX) model resolved the multi-subject, multi-condition structure and the Kalman-Smoother resolved the problem of irregular-intervals in observations. We showed that the proposed method could identify the simulated system that involved sparse variability among subjects.

1. Introduction

Recently, with the rising needs of treatment options of pancreas diseases, further understanding of the pancreas functions has become more and more important. Diabetes is a group of metabolic diseases in which the blood sugar is of high level for a prolonged period that can cause many acute complications and serious long-term complications. Diabetes is considered to occur, because the pancreas does not produce enough insulin or the cells of the body do not respond properly to the produced insulin. The dynamics of blood chemicals turn to be very important, since all the metabolic compounds (e.g. glucose) from the cells and hormone productions (mainly insulin) from pancreas are released into our circulatory system, and hence involved in the homeostasis of the whole body. As an example, the oral glucose tolerance test (OGTT) has been commonly performed to diagnose diabetes, in which a standard dose of glucose is ingested by mouth, and blood levels are measured several hours later. The test often needs to draw blood several times and requires two hours or up to six hours, in order to allow only the glucose (sometimes with insulin) levels to be measured. In many cases, rich information from various chemicals in the metabolic system has been ignored, although it can shorten the test time and also detect different types of diabetes.

The goal of our project is the identification of the metabolic system consisting of blood chemicals, which is involved in glucose induced pancreas functions, and its variability among individuals. Previously, Kuroda and colleagues have successfully

modeled intracellular insulin pathway[1] and found two different patterns of insulin activity in circulatory system[2]. This study is an attempt to analyze sequential data of blood chemicals for identifying the metabolic system controlled by pancreas as an endocrine gland. In this study, we have faced several challenges. The first challenge is due to multiple subjects and multiple conditions in the blood chemical analysis. The “common” features over all the subjects and conditions are of course very important for diagnosing diseases, while the “individual” features of each subject or dosing condition also provides important information for personalized treatment programs, reasonable eating habits advisement, and other future research directions. The second challenge is an irregular-interval observation in data sampling. Considering both the health of subjects and the experimental cost, blood sampling often cannot be frequent or equal. So, the time courses of observations become of irregular intervals. This also requires sophisticated treatment of the time-series.

Auto-Regressive (AR) model is a type of random process that specifies the output values depending on its own previous values, and has been used to describe various time-series found in nature and economics. Vector AR(VAR) model of more than one stochastic differential equations has been used in analyzing sequential data of medicine and physiology, such as RNA and DNA sequence[3][4][5]. There are many disadvantages in the VAR model: the output values can only depend on its own previous values; it is also weak against missing observations or nonlinear factors of the dynamics. VAR models can incorporate nonlinear factors inside the dynamics. There are two basic ways to make a VAR model nonlinear: One is to add nonlinear terms into the VAR model (then it becomes a Vector Auto-Regressive with exogenous variable (VARX) model)[6], and the other is to use an ARX model[7] with state-dependent radial basis functions[8].

¹ Graduate School of Informatics, Kyoto university, 36-1 Yoshidahonmachi, Sakyo-ku, Kyoto-city, 606-8501, Japan

² Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

^{a)} su-ds@sys.i.kyoto-u.ac.jp

^{b)} oba@i.kyoto-u.ac.jp

Referring to the previous study by Kuroda and others[2], we used a sparse VARX model with standard Kalman-Smoother method in this study. To solve the multi-subject, multi-condition problem, we assumed the parameter matrix of each condition of each individual as a sum of common, condition dependent, and individual dependent parameter matrices in the sparse VARX model. With this assumption and the sparse estimation of the parameters, this block structure not only increased the accuracy but also provided important information for studying the influence of individuals and conditions. Irregular-interval-observation of the time-course was found to significantly affect the estimation of regular VARX models. In order to handle this problem, we assumed that an irregular-interval time-course obtained by the actual blood sampling had been sub-sampled from a regular-interval time-course so that there were missing observations within it. Based on this assumption, we introduced standard Kalman-smoother into the state-space model and found that this modification was effective in dealing with irregular-interval time-courses. Among the blood chemicals, glucose and insulin are seen as corner stones, since they are the triggers of anabolism and catabolism; so we focused on these species. Nonlinear terms that had been introduced to the ODE model by Kuroda and others were also added into our VARX model to increase the reproducibility of important blood chemicals.

2. Methods

In this study, we developed a combined model of a sparse VARX model and standard Kalman-Smoother and applied it to the dataset of time-sequences of blood chemicals. We expected the VARX model represents quantitatively the metabolic dynamics of blood chemicals, whereas sparse regularization resolves multi-subject and multi-condition structure of the experimental data. Moreover, we expected the Kalman-smoother resolves the problem of irregular interval observations. Before explaining our proposed model, we briefly introduce the VARX model.

2.1 VARX model

A vector auto-regressive (VAR) model of order p (VAR(p)) is defined as

$$\mathbf{y}(t) = \sum_{i=1}^p \mathbf{W}_i \mathbf{y}(t-i) + \epsilon(t) \quad (1)$$

where $\mathbf{y}(t)$ is a $k \times 1$ variable vector at time $t = 1, \dots, T$, \mathbf{W}_i is a $k \times k$ matrix representing in total time-invariant coefficients, p is the order, and $\epsilon(t)$ is a $k \times 1$ vector of white noise. All the variables in the VAR above are treated individually and symmetrically over the vector; each variable has an equation explaining its evolution based on the previous values of its own and the other variables. A VAR model only requires a list of variables that hypothetically affect each other temporally as a prior knowledge, but includes different prior relationship between variables from those in structural models like simultaneous equations.

Sometimes, we additionally consider exogenous input $\mathbf{f}(t)$ that affects the VAR dynamics of eq.(1) which is formally written as follows:

$$\mathbf{y}(t) = \sum_{i=1}^p \mathbf{W}_i \mathbf{x}(t-i) + \epsilon(t) \quad (2)$$

where $\mathbf{x}(t)$ is a $k' \times 1$ variable vector that includes $\mathbf{y}(t)$ itself, the exogenous input $\mathbf{f}(t)$, and other extended terms $\mathbf{y}_{\text{ext}}(t)$, representing some non-linear effects. The time-invariant coefficient \mathbf{W}_i becomes a $k \times k'$ matrix. We call this extended model a VARX model. The extended terms $\mathbf{y}_{\text{ext}}(t)$ and the exogenous input $\mathbf{f}(t)$ should be designed to reflect background knowledge of the system.

2.2 Sparse VARX model

We propose a sparse model as an extension of the VARX(1) model to analyze metabolites data obtained by multi-subject and multi-condition measurements. Provided that we have a multi-subject and multi-condition metabolomics dataset $\mathbf{Y} = \{\mathbf{y}^{ij}(t) \in \mathbb{R}^k, i = 1, \dots, N_S, j = 1, \dots, N_C\}$ where N_S and N_C are the numbers of individual subjects and dosing conditions, respectively. To deal with this dataset, we assume a VARX model for each individual i and dosing condition j :

$$\mathbf{y}^{ij}(t) = \mathbf{W}^{ij} \mathbf{x}^{ij}(t-1) + \epsilon^{ij}(t) \quad (3)$$

where \mathbf{W}^{ij} is a $k \times k'$ matrix of time-invariant coefficient parameters for each individual i and dosing condition j . $\mathbf{x}^{ij}(t)$ is the same with $\mathbf{x}(t)$ in eq.(2), except for the dependence on individual i or dosing condition j . $\epsilon^{ij}(t)$ is a residual noise.

In our sparse VARX model, we assume either of the following four sub-model structures among subjects $i = 1, \dots, N_S$ and dosing conditions $j = 1, \dots, N_C$. The first model is

$$\mathbf{W}^{ij} = \mathbf{A}$$

saying that the parameters are common over all the individuals and dosing conditions; this is called model A. And, the second one is

$$\mathbf{W}^{ij} = \mathbf{A} + \mathbf{B}^i$$

saying that the parameters are common over different dosing conditions but different between individuals; this is called model AB. The third one is

$$\mathbf{W}^{ij} = \mathbf{A} + \mathbf{B}^i + \mathbf{D}^j$$

saying that there are individual effect \mathbf{B}^i and dosing effect \mathbf{D}^j , in total, called model ABD. Finally, the most complicated model is

$$\mathbf{W}^{ij} = \mathbf{A} + \mathbf{E}^{ij}$$

saying that there are both individual and dosing effects \mathbf{E}^{ij} , called model AE.

We assume all the matrices: \mathbf{A} , \mathbf{B}^i , \mathbf{D}^j and \mathbf{E}^{ij} , are sparse. In order to obtain sparse solutions for them, we applied the following L1 regularization[9][10]:

$$\text{Reg}(\mathbf{W}) = |\mathbf{A}| + \sum_i |\mathbf{B}^i| + \sum_j |\mathbf{D}^j| + \sum_{ij} |\mathbf{E}^{ij}| \quad (4)$$

where $|\mathbf{A}|$ denotes the sum of absolute values of all the elements of matrix \mathbf{A} .

2.3 VARX model with Kalman-Smoother

Usual AR and VARX models assume time-sequences with equal intervals, thus missing observations or observations with irregular intervals make the analysis with AR/VARX difficult. To

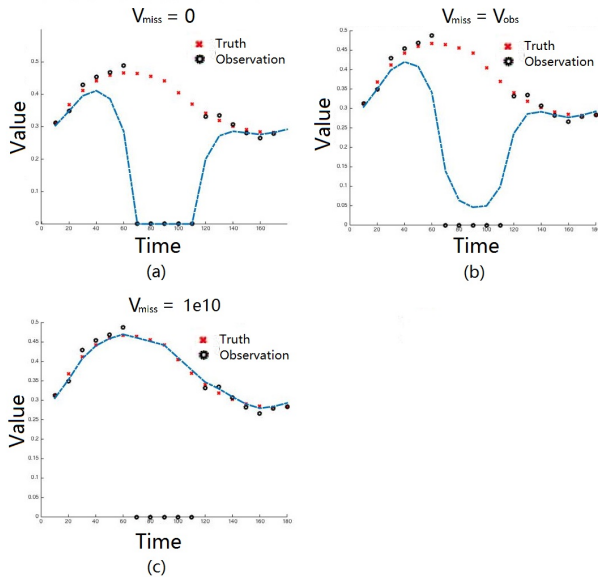


Fig. 1 Comparison of smoothed results among different variance for observation noise; three different value of V_{miss} was compared. Horizontal and vertical axes denote the time and value of time sequence, respectively. Black “o” and red “x” denote observation and truth values. Notice that we set the values of missing observations at zeros. Blue dash lines denote smoothed results. The observed values are available every ten minutes except for missing moments, at {70, 80, 90, 100, 110}. (a) smoothed result for $V_{miss} = 0$, (b) smoothed result for $V_{miss} = V_{obs}$, (c) smoothed result for V_{miss} equals a very large value

deal with irregular interval observations, we introduce Kalman-Smoother algorithm with a state-space modeling. Kalman-Smoother is an algorithm to estimate unknown system variables, given a series of measurements along time which may contain observation and system noises; the estimates are expected to be more precise than those based on a single measurement alone. We consider the following state space model.

$$\bar{\mathbf{x}}^{ij}(t) = \mathbf{W}^{ij} \bar{\mathbf{x}}^{ij}(t-1) + \epsilon^{ij}(t) \quad (5)$$

$$\mathbf{y}^{ij}(t) = \bar{\mathbf{x}}^{ij}(t) + \eta^{ij}(t) \quad (6)$$

where $\bar{\mathbf{x}}^{ij}(t)$ is a $k \times 1$ dimensional vector whose values are the same with the top k elements of the vector $\mathbf{x}^{ij}(t)$. $\eta^{ij}(t)$ is an observation noise whose variance $\mathbf{V}(t)$ may depend on time. We set the variance $\mathbf{V}(t) = V_{miss}$ for a time-point t at which the observation $\mathbf{y}^{ij}(t)$ was regarded as missing and $\mathbf{V}(t) = V_{obs}$ otherwise, where V_{miss} and V_{obs} are fixed hyper-parameters. In **Fig. 1**, if we set $V_{miss} = 0$, smoother will totally trust the observation, then state space model is exactly the same with the VARX model, eq.(2), in total; if $V_{miss} = V_{obs}$, the missing points will be treated equally as normal observations, This still greatly affects the accuracy of estimation method; if $V_{miss} = a$ very large value ($1e10$ in our simulation), smoother will have no trust to the observations (set as zero values) and choose the system prediction, make the dynamic smoothed.

This formalism, however, brings us a new problem that we have to deal with both unknown VAR model parameters and missing observations at the same time. To this end, expectation-and-maximization (EM) algorithm[11][12] was introduced into this framework. We regard $\mathbf{x}^{ij}(t)$ as a hidden variable vector if $\mathbf{V}(t) > 0$, and let \mathbf{X} be the set of all hidden variables. We can

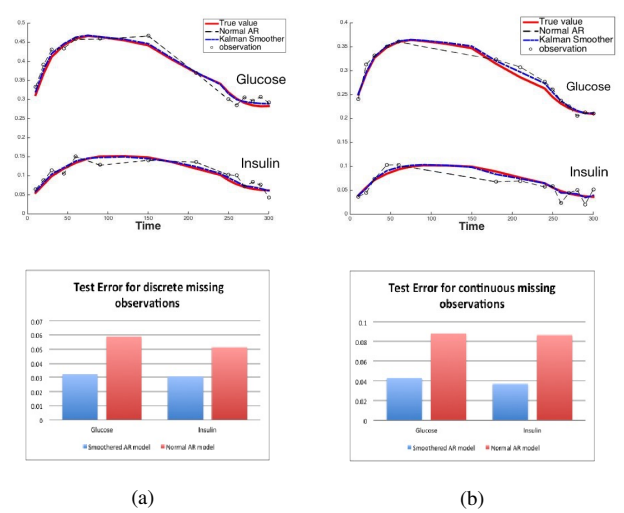


Fig. 2 A demonstration of Kalman-Smoother with comparison to simple AR model. (a) and (b) correspond to two different time-interval patterns, *sparse* and *connected*, produced by simulations, respectively. The top panels show the observed and the estimated time sequences, in which horizontal and vertical axes denote time and normalized concentrations of blood chemicals, respectively. Red, black dash, and blue lines denote ground truth series, estimation by AR, and that by Kalman-Smoother, respectively. Black “o” denotes a single observation. The observed values are available every ten minutes except for missing moments, at {60, 80, 100, 120, 140} in the *sparse* case (a) and at {90, 100, 110, 120, 130} in the *connected* case (b). The bottom panels show estimation errors of the missing observations corresponding to the two cases above. Blue bars depict the errors by Kalman-Smoother, and red bars by AR.

estimate a posterior distribution of the hidden variables \mathbf{X} by a Kalman-Smoother algorithm if the parameter value of \mathbf{W} is given. On the other hand, we can calculate maximum likelihood estimation of the parameter \mathbf{W} if the posterior distribution of the hidden variable \mathbf{X} is given. Although the latter maximum likelihood estimation is done with an L1 regularization, simultaneous estimation of \mathbf{X} and \mathbf{W} can be performed with an alternate optimization procedure, i.e., the EM algorithm, in total.

In the E-step, posterior density $q(\mathbf{X})$ is calculated for each individual and condition independently by standard Kalman-Smoother[13][14] with parameters of the VARX(1) model and partially missing observations (An irregular interval observations in real blood sampling experiment can be seen as regular interval observations with missing points),

$$q(\mathbf{X}) = p(\mathbf{X}|\mathbf{Y}, \hat{\mathbf{W}}) \quad (7)$$

where \mathbf{X} , \mathbf{Y} , and $\hat{\mathbf{W}}$ are internal (system) variables, observations, and the parameters of the VARX(1), respectively. The objective of the standard Kalman-Smoother algorithm is to obtain the marginal probability $\gamma(\mathbf{x}(t)) = p(\mathbf{x}(t)|\mathbf{y}(1), \dots, \mathbf{y}(T)) = N(\mathbf{x}_t|\hat{\mu}(t), \hat{\mathbf{V}}(t))$ for each $t = 1, \dots, T$.

In the M-step, parameter matrix \mathbf{W} is estimated by minimization of the following objective function:

$$L(\mathbf{W}, \mathbf{Y}) = Err(\mathbf{Y}, \mathbf{W}) + \lambda Reg(\mathbf{W}) \quad (8)$$

Quadratic error function *Err* is defined by

$$Err(\mathbf{Y}, \mathbf{W}) = \frac{1}{2} \sum_{t=2}^T (\mathbf{Y}(t) - \mathbf{W}\mathbf{X}(t-1))^2 \quad (9)$$

As described in section 2.2, the L1-norm regularization term is given by

$$Reg(\mathbf{W}) = |\mathbf{A}| + \sum_i |\mathbf{B}| + \sum_j |\mathbf{D}| + \sum_{ij} |\mathbf{E}| \quad (10)$$

λ denotes a regularization coefficient whose value is determined by a cross-validation procedure.

Kalman-Smoother is expected to estimate the unknown system variables more accurately than the simple AR model, especially when the observations consist of unequal-interval observations or missing observations. Fig. 2 demonstrates that the Kalman-Smoother provided a better estimation of the unknown variables than the AR model in the both cases of discrete and connected missing observations.

2.4 Hyper-parameters and cross-validation

The proposed method requires not only an appropriate setting of regularization hyper-parameter λ but also model selection from the four options, A, AB, ABD, and AE. In order to determine the best hyper-parameter setting and model, we compared cross-validation error that was defined as follows. In each sub-task in the cross-validation, we picked a time sequence of five continuous time points, regarded the observed values at these five time points as artificial missing values, estimated the parameters and all the missing values, and assessed the error of the estimation of the missing values by comparing with the ground truths (actual observations). We calculated the cross-validation error for every combination of possible values of the regularization hyper-parameter, and model structures, to determine the best combination. Note that we picked up five succession of time points in the cross-validation process, because we wanted to see fitting performance over large time gap, rather than to see one for interpolation of small gaps; the latter seems much easier than the former.

3. Experiment Data

An artificial dataset of the concentration of certain blood plasma chemicals, was generated in order to test the performance of the proposed method at a condition including sufficient amount of data. Another advantage would be availability of ground truth. We generated the artificial data by an application of the VARX(1) model with a given set of parameters and initial conditions. The dataset contained 20 time points of 10 chemical species from 20 individuals and in 2 different dosing conditions. All the values of chemical concentrations were normalized so that minimum and maximum values of each chemical became one and zero, respectively. The 10 chemical species we used in our experiment are Glucose, Insulin, GIP(Active), Alanine, Leucine, Tyrosine, Total ketone body, Pancreatic glucagon, Citrate, and Succinate, which are important hormone and typical metabolites of catabolism. The information of subjects was blinded. Dosing conditions were separated into two types: *bolus* and *ramp*. The *bolus* denotes a standard Oral Glucose Tolerance Test (OGTT), namely, the subject took a dose of 75g glucose in a time after 2 hours fasting. In the *ramp* condition, the subject kept a series of small dose portions of total 75g glucose over 2 hours after 2 hours fasting. The parameter matrix \mathbf{W}^{ij} for each individual $i = 1, \dots, 20$ and each

dosing condition $j \in \{bolus, ramp\}$ was prepared as follows. The common parameter \mathbf{A} was set consistently with the one estimated from the real data described in the previous subsection. We set the parameter matrices \mathbf{B}^i at zero matrices for 10 individuals and sparse random matrices for the other 10 individuals. This setting of $\mathbf{B}^i, i = 1, \dots, 20$ reflected an assumption that many subjects had identical characters and the others had their own characters in some coefficients being deviated from the population average. We set \mathbf{D}^j at zero matrices for both of the two dosing conditions because there was no plausible assumption for preparing different metabolic systems in the two dosing conditions.

4. Results

To evaluate our method, we performed one experiment based on the artificial dataset, we estimated the parameters of the four model structures, and optimized hyper-parameter λ of the L1-norm regularization with respect to the cross-validation error for each model structure. We also estimated the parameters of three model structures and compared the parameters with the ground-truth values of the VARX(1) models that had generated the data. Table 1 shows the cross-validation error for these settings.

Table 1 Comparison of cross-validation error among the four model structures and various settings of regularization hyper-parameter λ . Bold text denotes the smallest error in the four models.

λ	0.3	0.1	0.06	0.03	0.01	0.001
AE	8.60E-04	2.91E-04	2.06E-04	8.69E-05	2.55E-05	7.17E-05
ABD	8.59E-04	2.79E-04	1.82E-04	8.22E-05	2.08E-05	6.86E-05
AB	8.58E-04	2.81E-04	1.84E-04	8.46E-05	2.37E-05	6.67E-05
A	8.63E-04	3.25E-04	2.54E-04	2.17E-04	1.68E-04	3.84E-04

As we expected, good performances were obtained by models AB and ABD than those by models A and AE. Too simple (sub-model A) and too complicated (sub-model AE) models could not perform better. Based on the large difference in the performance between models AB and A, we speculate individual characters, which are represented by matrices $\mathbf{B}^i, i = 1, \dots, 20$, had strong influence on the results. There was improvement in the performance by model ABD to by AB, possibly because we have some false positive elements in the matrices \mathbf{D}^j .

In Fig. 3, we compared the true and estimated values of the parameter matrices \mathbf{A}, \mathbf{B} , and \mathbf{D} , with a particular interest in reproducibility of the sparse structure of the true matrices. We also show the numbers of true positives, true negatives, false positives, and false negatives, to examine the accurate reproduction of the binary structure of the true sparse matrices. Seeing Fig. 3, we found that the estimated matrices were sparser than the true ones. The precision rate of the estimation of matrix \mathbf{A} to detect non-zero parameter values was as high as 81%. There were zero and three false positives in estimation of matrices \mathbf{B}^i (Fig. 3(b)) and \mathbf{D}^j (Fig. 3(c)), respectively. In the simulation, VARX coefficients \mathbf{W}^{ij} of five out of ten subjects were set same with the average and those of two dosing conditions were also the same. Because of this sparse settings of individual-based and condition-based matrices, the low false positives in the result were preferable, reflecting the sparseness in the real data generation process.

We visualized the ground truth and the estimated parameters of sub-model ABD in Fig. 4. The figure shows that sparseness

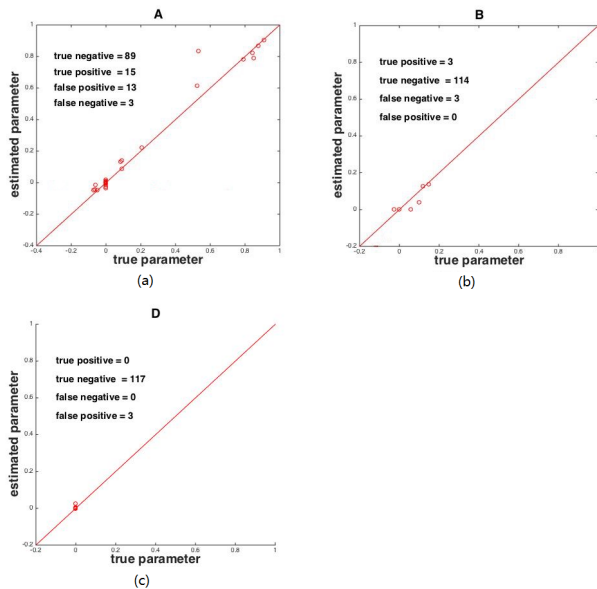


Fig. 3 Comparison between true and estimated parameter values for matrices **A**, **B**, and **D** of sub-model ABD. Horizontal and vertical axes denote the true and the estimated parameters, respectively. Notice that matrix **D** and half of the matrix **B** is set empty for our method to detect the sparseness of the matrices. Numbers of true positives, true negatives, false positives, and false negatives to detect non-zero entry in the matrices are shown in each panel. (a) comparison of matrix **A**(common), (b) comparison of matrix **B**(individuals), (c) comparison of matrix **D**(dosing conditions)

Table 2 Improvement of model fitting by introducing the extended variables. Cross validation errors in the reproduction of time-sequences of the two important chemicals, glucose and insulin.

Chemicals	Terms	Model A	Model AB	Model ABD
Glucose	Non-extended	7.7E-5	5.7E-6	3.6E-6
	With extended	5.8E-5	4.9E-6	3.2E-6
Insulin	Non-extended	5.2E-5	7.4E-6	4.5E-6
	With extended	3.5E-5	5.0E-6	3.1E-6

of the estimated matrices well reflected that of the ground truth, although the estimation was a little sparser than the truth. All the empty matrices of individual differences were correctly detected (Fig. 4(b)).

Our method detected a few false positive elements in the matrices D^j ; however, the values were very close to zero, the true value.

5. Discussion

In our VARX model, two extended terms were added to the set of cause variables y , in order to allow the model to better fit to dynamics of important molecules, glucose and insulin. Comparison shows that the extended terms contributed to improving the accuracy.

We also added the terms that represent the absorption of oral glucose, in which *bolus* condition values were taken from the Elashoffs model, and *ramp* condition values were the dose averages. These terms, however, did not lead to substantial improvement of our models, possibly because condition difference was not considered when estimating the parameters. Since condition-dependent estimation may exhibit further improvement, it remains as our future study.

According to the cross validation, model structure ABD

showed the slight advantage in the accuracy over model AB (see **Table 1**). The estimated matrix **D** was also found very sparse with few small values, being comparable with the "ground truth" which had been set empty. There is still a room to be discussed, whether we can say that the dosing condition has little influence to the dynamics of blood chemicals, based on these two experimental results.

In Fig. 4, the estimation of matrix **D** was not good enough, because it did not reproduce the emptiness of matrix **D** so good as in matrix **B**. There should be many possible reasons for this result; for example, arbitrary set up of the external input may have been inappropriate, condition differences was not under the consideration for the different external inputs, the react insulin flow in the bolus condition was too strong for the extended terms to absorb. Accordingly, there is still great room for improvement of the estimation of matrix **D**.

6. Summary

In this study, we developed a time-sequence analysis method that can deal with heterogeneous data coming from multiple individuals and multiple conditions, and also temporally complicated observations with irregular intervals. As a typical experiment, we considered a blood chemical monitoring to examine the dynamics of the blood chemicals related with pancreas as an endocrine gland. In specific, we proposed a sparse VARX model with the Kalman-Smoother method. To represent individual and condition dependent factors, the model parameters were divided into four groups; common, individual-dependent, condition-dependent and individual-condition-dependent ones. Extended terms were also added to the model to improve the fitting to the dynamics of important blood chemical species (glucose and insulin in our case). Sparse estimation based on the L1-norm regularization was effective in exacting the underlying sparse structures. Applications to the artificial dataset suggested that some of our model structures, like ABD and AB, showed the best performance, by incorporating appropriate dependence on the individuals. Through the simulation study, we found our sparse estimation could sufficiently reproduce the underlying sparse dependence on the individuals or conditions. There is still much room for improvement, especially in selecting the extended terms or construction of model structures, such as introduction of layered models incorporating clustered secondary features; they are remaining for future studies.

References

- [1] Kubota, H., Noguchi, R., Toyoshima, Y., Ozaki, Y.-i., Uda, S., Watanabe, K., Ogawa, W. and Kuroda, S.: Temporal coding of insulin action through multiplexing of the AKT pathway, *Molecular cell*, Vol. 46, No. 6, pp. 820–832 (2012).
- [2] Noguchi, R., Kubota, H., Yugi, K., Toyoshima, Y., Komori, Y., Soga, T. and Kuroda, S.: The selective control of glycolysis, gluconeogenesis and glycogenesis by temporal insulin patterns, *Molecular systems biology*, Vol. 9, No. 1, p. 664 (2013).
- [3] Oh, S., Song, S., Grabowski, G., Zhao, H. and Noonan, J. P.: Time series expression analyses using RNA-seq: a statistical approach, *BioMed research international*, Vol. 2013 (2013).
- [4] Fan, Y., Wu, R., Chen, M.-H., Kuo, L. and Lewis, P. O.: A Conditional Autoregressive Model for Detecting Natural Selection in Protein-Coding DNA Sequences, *Topics in Applied Statistics*, pp. 203–212 (2013).

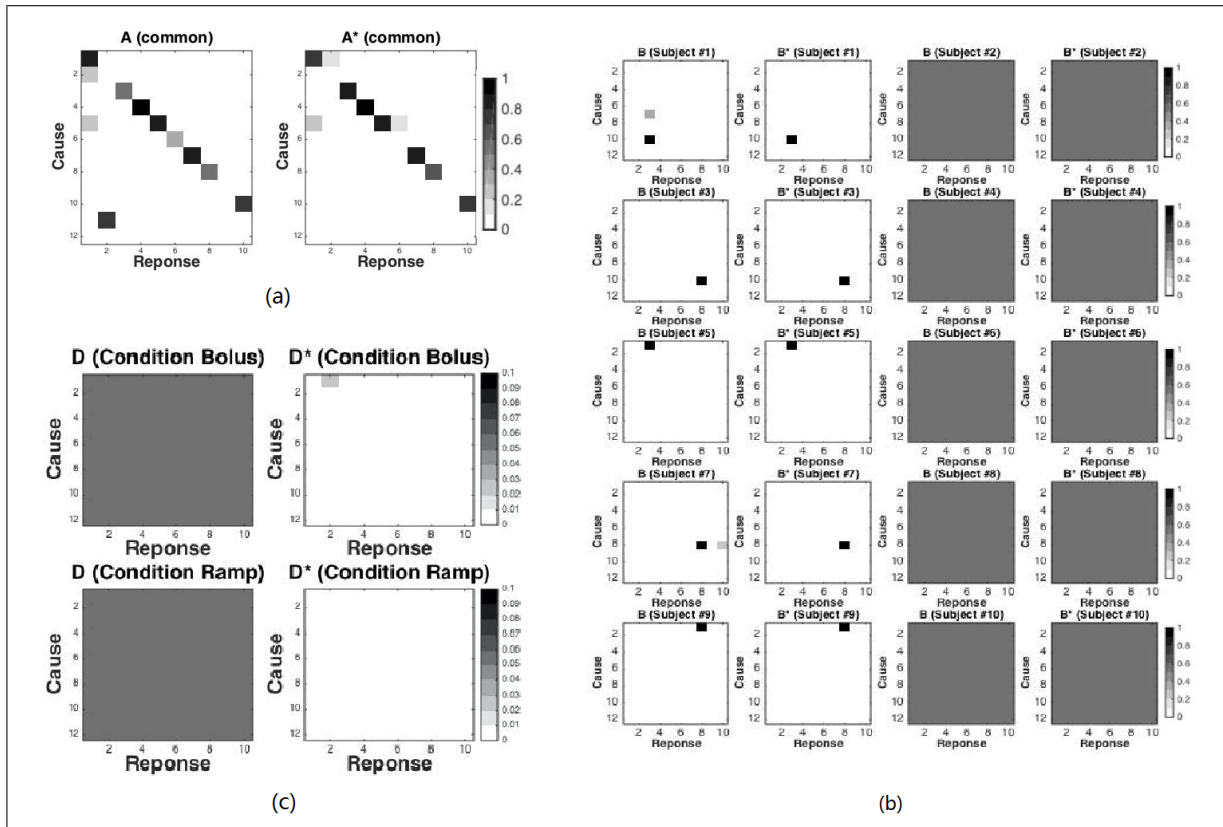


Fig. 4 Comparison of the estimated and ground-truth values of the 22 parameter matrices **A**, **B**, and **D** of sub-model ABD. The true matrices (e.g. **B**) and the estimated (e.g. **B***) matrices of ten subjects, #1, ..., #10, are shown (see titles of the panels). Rows and columns of the matrices denote 12 causes and 10 responses of the VARX(1) model, respectively. The matrix **D** and the matrix **B** for a half of 10 subjects were set empty as the ground truth. (a) comparison of matrix **A**(common) between true and estimated parameters, (b) comparison of matrix **B**(individuals) between true and estimated parameters, (c) comparison of matrix **D**(dosing conditions) between true and estimated parameters

[5] Fujita, A., Sato, J. R., Garay-Malpartida, H. M., Yamaguchi, R., Miyano, S., Sogayar, M. C. and Ferreira, C. E.: Modeling gene expression regulatory networks with the sparse vector autoregressive model, *BMC Systems Biology*, Vol. 1, No. 1, p. 39 (2007).

[6] Chen, R. and Tsay, R. S.: Nonlinear additive ARX models, *Journal of the American Statistical Association*, Vol. 88, No. 423, pp. 955–967 (1993).

[7] Liu, Y. and Allen, R.: Analysis of dynamic cerebral autoregulation using an ARX model based on arterial blood pressure and middle cerebral artery velocity simulation, *Medical and Biological Engineering and Computing*, Vol. 40, No. 5, pp. 600–605 (2002).

[8] Peng, H., Nakano, K. and Shioya, H.: Nonlinear predictive control using neural nets-based local linearization ARX model Stability and industrial application, *Control Systems Technology, IEEE Transactions on*, Vol. 15, No. 1, pp. 130–143 (2007).

[9] Li, W., Feng, J. and Jiang, T.: IsoLasso: a LASSO regression approach to RNA-Seq based transcriptome assembly, *Journal of Computational Biology*, Vol. 18, No. 11, pp. 1693–1707 (2011).

[10] Hoerl, A. E., Kannard, R. W. and Baldwin, K. F.: Ridge regression: some simulations, *Communications in Statistics-Theory and Methods*, Vol. 4, No. 2, pp. 105–123 (1975).

[11] Moon, T. K.: The expectation-maximization algorithm, *Signal processing magazine, IEEE*, Vol. 13, No. 6, pp. 47–60 (1996).

[12] Wallin, R., Isaksson, A. and Ljung, L.: An iterative method for identification of ARX models from incomplete data (2000).

[13] Wan, E. and Van Der Merwe, R.: The unscented Kalman filter for nonlinear estimation, *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, IEEE, pp. 153–158 (2000).

[14] Rauch, H. E., Striebel, C. and Tung, F.: Maximum likelihood estimates of linear dynamic systems, *AIAA journal*, Vol. 3, No. 8, pp. 1445–1450 (1965).