IISE Transactions on Healthcare Systems Engineering

ISSN: 2472-5579 (Print) 2472-5587 (Online) Journal homepage: https://www.tandfonline.com/loi/uhse21

Nested Gaussian process modeling and imputation of high-dimensional incomplete data under uncertainty

Farhad Imani, Changqing Cheng, Ruimin Chen & Hui Yang

To cite this article: Farhad Imani, Changqing Cheng, Ruimin Chen & Hui Yang (2019): Nested Gaussian process modeling and imputation of high-dimensional incomplete data under uncertainty, IISE Transactions on Healthcare Systems Engineering, DOI: 10.1080/24725579.2019.1583704

To link to this article: https://doi.org/10.1080/24725579.2019.1583704

Accepted author version posted online: 08 Mar 2019.
Published online: 09 Apr 2019.

IISE TRANSACTIONS ON HEALTHCARE SYSTEMS ENGINEERING
https://doi.org/10.1080/24725579.2019.1583704

Nested Gaussian process modeling and imputation of high-dimensional incomplete data under uncertainty

Farhad Imani[a], Changqing Cheng[b], Ruimin Chen[a], and Hui Yang[a]

[a]Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, University Park, PA, USA; [b]Department of Systems Science and Industrial Engineering, State University of New York, Binghamton, NY, USA

ABSTRACT

Modern healthcare systems are increasingly investing in advanced sensing and information technology, leading to data-rich environments in hospitals. For example, when patients are admitted into an intensive care unit (ICU), it is a common practice to monitor a number of clinical variables, such as blood pressure, heart rate, gas exchange, pulse oximetry and metabolic panel. However, heterogeneous sensing and measurement methods often lead to data uncertainty and incompleteness. Missing values exist pervasively for ICU clinical variables pertinent to a patient's health condition. This adversely affects time-critical decision making in patient care. Hence, there is an urgent need to develop advanced analytical methods that address the challenges of ICU data uncertainty, provide a robust estimate of health conditions and derive in-depth knowledge for decision making from heterogeneous healthcare recordings. This article presents a novel nested Gaussian process (NGP) model that is tailored to represent multi-dimensional covariance structure of time, variable and patient for high-dimensional data imputation. We evaluate and validate the proposed NGP method on both simulation and real tensor-form ICU data with high-level missing information. Experimental results show that the proposed methodology effectively handles the data uncertainty in ICU settings, which helps further improve the biomarker extraction, patient monitoring and decision making. The proposed NGP model can also be generally applicable to a variety of engineering and medical domains that entail high-dimension data imputation and analytics.

ARTICLE HISTORY
Received 20 April 2018
Accepted 13 February 2019

KEYWORDS
Heterogeneous sensing; intensive care unit; missing data imputation; nested Gaussian process; tensor-form data

1. Introduction

The annual admission to intensive care units (ICUs) in the US is more than 5.7 million with an economic cost of $81.7 billion (Pastores *et al.*, 2012). Generally, the ICU provides deliberated care services; e.g., advanced life support and intensive monitoring, to seriously ill patients. ICU patients, albeit a heterogeneous population, have the compelling necessity for real-time monitoring and regular lab tests compared to those admitted to non-ICU beds. In particular, as the population ages, the prevalence of multi-morbidity and the resulting complexity of treatments spur the implementation of multimodal sensing technologies to improve the quality of ICU care.

Advanced sensing gives rise to the ubiquitous data-rich environment in ICUs. It is not uncommon that a large number of sensors are used for postoperative recovery monitoring (Chen and Yang, 2016b; Zhu *et al.*, 2019). Modern ICUs require the monitoring of important clinical variables, including laboratory tests (e.g., blood and urine), heart rate and rhythm, blood pressure, respiratory rate and blood-oxygen saturation. Traditionally, clinicians make inferences about patient conditions only with the most recent monitoring variables, while overlooking past results, patient similarity and variable correlation. Realizing the full potential of rich ICU data for postoperative care depends on the development of new analytical methods and tools to handle data uncertainty, delineate hidden patterns and provide effective clinical decision support (Yang *et al.*, 2014).

However, heterogeneous sensing poses significant challenges for information extraction and decision making. As shown in Fig. 1(b), ICU monitoring leads to a new tensor form of datasets that provide rich information on the underlying dynamics of postoperative recovery processes. ICU tensor data also present distinct characteristics, such as patient heterogeneity, time asynchronization and variable heterogeneity, as opposed to the traditional table form of predictor and response variables (see Fig. 1(a)) commonly seen in predictive modeling (Chen and Yang, 2014).

- ***Patient heterogeneity:*** The patients admitted to the ICU may undergo different treatment procedures, suffer from different kinds of diseases, or belong to various age groups. New analytical methods and tools are therefore needed to handle the heterogeneity in the patient population, and shed insights on the data-driven estimation and modeling of patient conditions for better care and resource allocation.

CONTACT Hui Yang huy25@psu.edu Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, University Park, PA, USA.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uhse.

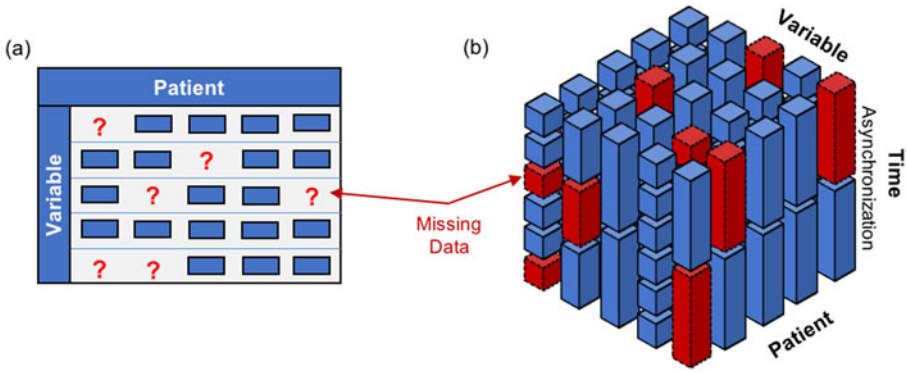

Figure 1. (a) Traditional table-form data for predictive modeling; (b) new tensor-form data generated in ICU settings.

- ***Time asynchronization*:** In current clinical practice, ICU data collection protocol is not standardized. It is common that the frequency of measurements may be subjected to the discretion of clinicians. As such, the time stamp for clinical variables is often not uniform. Some variables may be observed at a low sampling rate (e.g., Lactate, Creatinine and Glucose), whereas others may be at a high sampling rate (e.g., temperature, heart rate and a fraction of inspired oxygen (FiO_2)).
- ***Variable heterogeneity:*** As mentioned earlier, ICU sensing involves more than 44 clinical variables and they provide rich and diverse information on patient conditions from different perspectives (Chen and Yang, 2016a; Liu and Yang, 2018). For instance, FiO_2, platelets, bilirubin and hypertension as ICU variables reveal health condition for respiratory, coagulation, liver and cardiovascular systems, respectively.

Furthermore, missing data are prevalent in modern ICUs due to heterogeneous types of sensing and measurement methods, attrition in longitudinal assessments, poor record keeping or human errors, to name a few (Cismondi *et al.*, 2013). Missing data pose significant challenges for the estimate of a patient's condition. For example, the presence of missing data could negatively influence predictive modeling. The cumulative impact of missing data causes considerable loss of precision and power in clinical decision making (Little *et al.*, 2012; Vesin *et al.*, 2013). In particular, ICU data are in the tensor form and missing values are embedded in this structure, as shown in Fig. 1(b). New analytical methods are urgently needed to capture high-dimensional covariance structure for missing data imputation in ICU settings. Yet, traditional imputation methods are designed to work with the table-form data and are not tailored to handle the high-dimensional correlation structure in the tensor data.

This article presents a novel nested Gaussian process (NGP) approach to model and impute high-dimensional tensor data that is conducive to improving the quality of care in ICU settings. As a nonparametric approach, NGP is flexible enough to represent multi-dimensional covariance structure across time, patient and variable, and provides a predictive distribution (i.e., posterior mean and variance for Gaussian distribution) for missing data imputation rather than a point estimate from an explicit function. In short, NGP accounts for the correlation in time domain, similarity among patients, and inter-relationships of variables, thereby improving the performance of missing data imputation through the proposed hierarchical structure of covariance functions.

The remainder of this article is organized as follows: Section 2 introduces a research background on missing data handling. Section 3 presents the research methodology of NGP. Section 4 considers experimental design and materials for both simulation and real case studies. Sections 5 and 6 present experimental results and conclude the article, respectively.

2. Research background

As human discretion directs the collection of clinical data in the ICU environment, missing data have become a common problem facing postoperative monitoring and intervention. Generally, there are three different types of missing data mechanisms (Little and Rubin, 2014). The first type is called missing completely at random (MCAR), in which missingness does not depend on values in a dataset. Let $\boldsymbol{X} = (x_{ij})$ denote a data matrix where x_{ij} is the value of the variable j for sample i. $\boldsymbol{M} = (m_{ij})$ symbolizes the missing-data indicator matrix, where $m_{ij} = 1$ if x_{ij} is missing and $m_{ij} = 0$ if x_{ij} is observed. In the MCAR mechanism, the missing-data indicator matrix $\boldsymbol{M}$ is statistically independent of the data $\boldsymbol{X}$ as:

$$p(\boldsymbol{M} \mid \boldsymbol{X}) = p(\boldsymbol{M}) \tag{1}$$

In other words, the mechanism of missing is not because of anything other than the randomness. The second mechanism is missing at random (MAR), in which, if we divide $\boldsymbol{X}$ to two groups, i.e., observed data $\boldsymbol{X}_{obs}$ and missing data $\boldsymbol{X}_{mis}$, then the probability of missing is independent of $\boldsymbol{X}_{mis}$ but dependent on part or all of the observed data $\boldsymbol{X}_{obs}$:

$$p(\boldsymbol{M} \mid \boldsymbol{X}) = p(\boldsymbol{M} \mid \boldsymbol{X}_{obs}) \tag{2}$$

The third mechanism is missing not at random (MNAR), which dictates that missingness depends on both observed data and the missing values. Here, the probability of missing hinges on both $\boldsymbol{X}_{obs}$ and $\boldsymbol{X}_{mis}$:

$$p(\boldsymbol{M} \mid \boldsymbol{X}) = p(\boldsymbol{M} \mid \boldsymbol{X}_{obs}, \boldsymbol{X}_{mis}) \tag{3}$$

In the literature, missing data are often handled in two ways: complete case analysis and data imputation. Complete

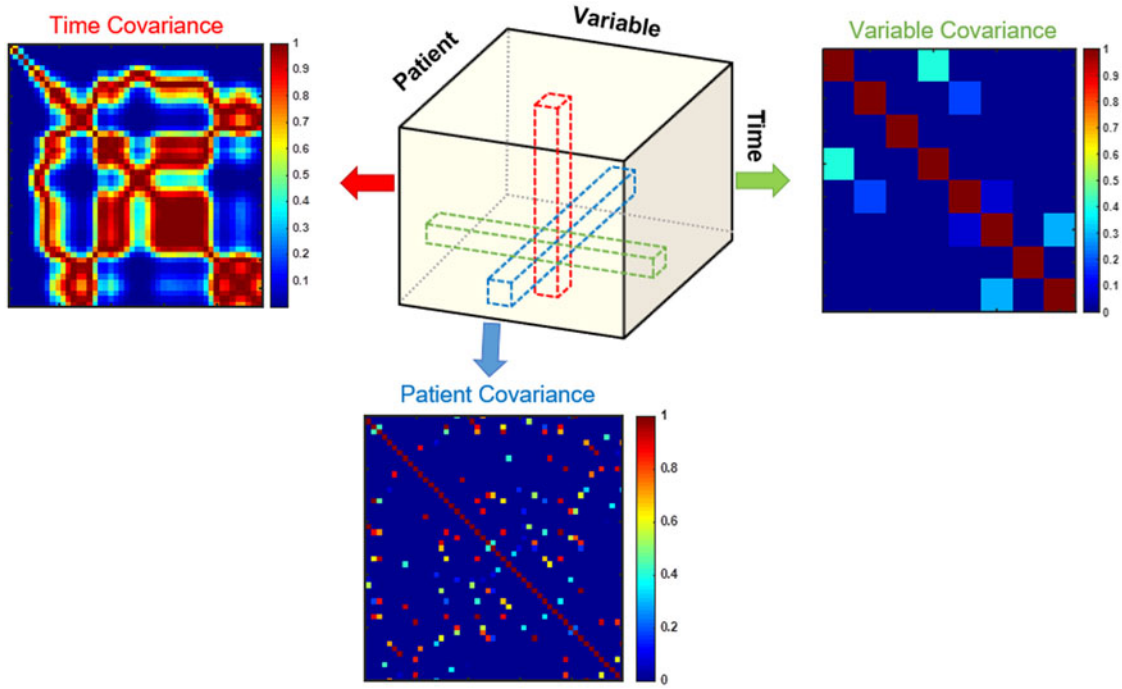

Figure 2. Heat maps of time, patient and variable covariance over the 48-hour period for Patient ID 1802 and the variable mean arterial pressure (MAP) in the MIMIC-II clinical database.

case analysis excludes the cases with missing data and only relies on the complete data (Haukoos and Newgard, 2007; Newgard and Haukoos, 2007). Listwise and pairwise deletion are the two commonly used methods. However, this approach throws away a sizable portion of data and diminishes the power of statistical analysis. It is worth mentioning that the pivotal assumption of complete case analysis is the MCAR with limited missing values, which is often violated due to the heterogeneous sensing methods in ICU settings (García-Laencina *et al.*, 2015; Little *et al.*, 2012).

In contrast, missing data imputation fills in predicted values at the locations of incomplete observations (Little and Schenker, 1995; Schafer and Graham, 2002). Single imputation replaces missing data with a single statistic derived from observations (e.g., mean or median) in the scenario of MAR. Single imputation retains a great portion of information which could otherwise be relinquished by the deletion methods. Similarly, the hot deck method was developed to replace missing values with the most frequent observations (Fuller and Kim, 2005). A major drawback of these imputation methods is the lack of uncertainty quantification associated with the prediction.

On the other hand, multiple imputation estimates the posterior distribution of the missing data given the observations (Kenward and Carpenter, 2007; Sinharay *et al.*, 2001). Maximum likelihood estimates the parameters that optimize the likelihood to predict the observed data with numerical optimization algorithms such as Newton–Raphson or the expectation–maximization algorithm (Dempster *et al.*, 1977). However, most of the maximum likelihood methods use only linear models, and are therefore sensitive to the choice of initial values and suffer from the issue of "curse of dimensionality" in the presence of high-dimensional data such as those in the ICU settings (Johnson *et al.*, 2016).

More recently, analytical methods such as support vector machines (SVM) and K nearest neighbor (KNN) are increasingly used for missing data imputation (Pan *et al.*, 2015; Pelckmans *et al.*, 2005). However, most of those methods focus on the imputation of the table-form data (see Fig. 1(a)) rather than high-dimensional tensor data (see Fig. 1(b)). Notably, table-form data, including predictor and target variables, are not time-varying. In other words, the temporal dimension is not a concern in many previous studies. Therefore, existing approaches are limited in the ability to handle high-dimensional missing data imputation in ICU settings.

As a nonparametric approach, Gaussian process (GP) is more flexible than traditional imputation methods to represent *multi-dimensional* data and provides the quantification of posterior uncertainty rather than point estimate (Cheng, 2018). Note that the Kriging and GP models have attracted considerable attention in the domain of geo-spatial analysis and computer experiments (Joseph, 2006; Joseph and Kang, 2011). Both models provide the best linear unbiased prediction of missing values in a spatial region or a design space. However, a major limitation is the assumption of stationarity (i.e., constancy of mean, variance, etc.) for the underlying stochastic process, thereby making it incongruous to deal with the tensor form of clinical variables that exhibit substantial nonstationarity (Cheng *et al.*, 2015). Furthermore, spatio-temporal structure of tensor data significantly challenges the conventional formulation of Kriging and GP models (Liu and Yang, 2013; Yang *et al.*, 2012, 2013). Hence, there is an urgent need to develop a new GP framework to delineate the inherent multi-dimensional and hierarchical covariance structure (i.e., time, patient and variable correlations), instead of one-dimensional correlation, and

tackle the challenges of incomplete and uncertain data in ICU settings.

3. Research methodology

In this article, a novel NGP approach is proposed to impute missing values in high-dimensional tensor data generated from heterogeneous sensing in an ICU. NGP is nonparametric with a new multi-dimensional covariance structure among time, patient and variable, and offers a greater level of flexibility to impute missing data in the tensor form. To portray the inherent hierarchical structure, we consider the time covariance in the first level, patient covariance in the second level and variable covariance in the third level:

- Time covariance has the highest influence on ICU missing data imputation, as the historical data tend to have a higher correlation along the time dimension for the same variable (see Fig. 2).
- Patient similarity is set forth in the second level, as similar patients may possess a comparable evolution trajectory in the same variable. Notably, the exogenous factors such as environmental, physical and psychological conditions could blur the imputation of missing value for a patient just through the similarity with other patients. Indeed, the estimate based on patient similarity provides *a priori* expectation of the missing value at a specific time stamp, which will be fed to the first level.
- Variable covariance is depicted in the third level because it does not provide as high correlations as in the other two levels. Intuitively, patient similarity in the second level hinges on the similarity of their variables. That said, the more the variables of one patient bear a resemblance to those of the other, the higher the similarity of the two patients.

To further illustrate these three correlations, we examine the covariance configuration using real data from the multiparameter intelligent monitoring in intensive care (MIMIC-II) clinical database (Goldberger *et al.*, 2000; Saeed *et al.*, 2011). The covariance function across each factor is estimated by fixing two other factors. As shown in Fig. 2, the heat map of time covariance exhibits the strongest similarity for the variable of mean arterial pressure (MAP) and Patient ID 1802. This is in stark contrast with the scatter bright spots in the heat map for patient similarity, and the highly sparse bright spots for variable similarity.

3.1. Gaussian process

In this section, we first introduce one-level GP with the time covariance, and then illustrate the proposed NGP and its posterior prediction to impute missing tensor data in ICU. Let $\boldsymbol{x} = (x_1, ..., x_n)$ be time-varying realization of a certain ICU variable recorded at time $\boldsymbol{t} = (t_1, ..., t_n)$, imputation at the first level seeks to find the following map:

$$x \sim f(t) + \epsilon \tag{4}$$

where $\epsilon \overset{\text{iid}}{\sim} N(0, \sigma_{nt}^2)$ is the error term and $f(t)$ is a GP determined by the mean function μ_t and covariance function K_t; i.e.,

$$f(t) \sim \mathcal{GP}(\mu_t, K_t) \tag{5}$$

where the GP defines the distribution of functions, μ_t is the mean function, and K_t is the covariance matrix. Note that different types of mean and covariance functions can be considered in GP modeling, such as the linear mean (i.e., $\mu_t = \alpha t + \beta$, where α and β are the slope and intercept of the linear mean function) and squared exponential covariance functions (i.e., $K_t = \sigma_t^2 \exp[-(t-t')^2/2l^2]$, where σ_t^2 is the signal variance and l is length-scale of the exponential covariance function). For posterior distribution in GP, the joint prior distribution is constrained to contain functions that comply with observations from data. Therefore, predictive distribution for the missing x^* at t^* is represented by the posterior distribution, conditional on the observations $(\boldsymbol{t}, \boldsymbol{x})$ (Williams and Rasmussen, 1996). Concretely, we have a joint prior distribution as follows:

$$\begin{bmatrix} \boldsymbol{x} \\ x^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ \mu^* \end{bmatrix}, \begin{bmatrix} K_{\boldsymbol{tt}} + \sigma_{nt}^2 \boldsymbol{I} & K_{\boldsymbol{t}t^*} \\ K_{t^*\boldsymbol{t}} & K_{t^*t^*} \end{bmatrix} \right) \tag{6}$$

where μ^* is the prior mean for x^*, and $K_{t^*\boldsymbol{t}}$ is the covariance between x^* and $\boldsymbol{x}$. The posterior distribution of x^* is given as

$$P(x^* \mid \boldsymbol{x}, \boldsymbol{t}, t^*) \sim \mathcal{N}\left(\overline{x^*}, cov(x^*)\right) \tag{7}$$

Here, the posterior mean $\overline{x^*}$ and covariance $cov(x^*)$ are obtained as

$$\overline{x^*} = \mu^* + K_{t^*\boldsymbol{t}}\left[K_{\boldsymbol{tt}} + \sigma_{nt}^2 \boldsymbol{I}\right]^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \tag{8}$$

$$cov(x^*) = K_{t^*t^*} - K_{t^*\boldsymbol{t}}\left[K_{\boldsymbol{tt}} + \sigma_{nt}^2 \boldsymbol{I}\right]^{-1} K_{\boldsymbol{t}t^*} \tag{9}$$

The predictive performance depends on hyper-parameters $\boldsymbol{\theta} = \{\boldsymbol{\theta_m}, \boldsymbol{\theta_c}\}$, which can to be learned from the training data. Here, $\boldsymbol{\theta_m}$ and $\boldsymbol{\theta_c}$ represent the set of hyper-parameters for mean and covariance function, respectively. In the case of the linear mean and squared exponential covariance function, hyper-parameters sets are: $\boldsymbol{\theta_m} = \{\alpha, \beta\}$ and $\boldsymbol{\theta_c} = \{l, \sigma_t, \sigma_{nt}\}$, which can be learned from the observed values using maximum marginal likelihood. Explicitly, marginal log-likelihood function and partial derivative regarding to $\boldsymbol{\theta_m}$ and $\boldsymbol{\theta_c}$ can be expressed as:

$$\begin{aligned} \log p(\boldsymbol{x} \mid \boldsymbol{t}, \boldsymbol{\theta}) = & -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \left[K_{\boldsymbol{tt}} + \sigma_{nt}^2 \boldsymbol{I}\right]^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \\ & -\frac{1}{2}\log \mid K_{\boldsymbol{tt}} + \sigma_{nt}^2 \boldsymbol{I} \mid -\frac{n}{2}\log(2\pi) \end{aligned} \tag{10}$$

$$\frac{\partial}{\partial \boldsymbol{\theta_m}} \log p(\boldsymbol{x} \mid \boldsymbol{t}, \boldsymbol{\theta}) = -(\boldsymbol{x} - \boldsymbol{\mu})^T \left(K_{\boldsymbol{tt}} + \sigma_{nt}^2 \boldsymbol{I}\right)^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta_m}} \tag{11}$$

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta_c}} \log p(\boldsymbol{x} \mid \boldsymbol{t}, \boldsymbol{\theta}) = & -\frac{1}{2} Tr\left[\left(K_{\boldsymbol{tt}} + \sigma_{nt}^2 \boldsymbol{I}\right)^{-1} \frac{\partial\left(K_{\boldsymbol{tt}} + \sigma_{nt}^2 \boldsymbol{I}\right)}{\partial \boldsymbol{\theta_c}}\right] \\ & + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \left[K_{\boldsymbol{tt}} + \sigma_{nt}^2 \boldsymbol{I}\right]^{-1} \frac{\partial\left(K_{\boldsymbol{tt}} + \sigma_{nt}^2 \boldsymbol{I}\right)}{\partial \boldsymbol{\theta_c}} \left[K_{\boldsymbol{tt}} + \sigma_{nt}^2 \boldsymbol{I}\right]^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \end{aligned} \tag{12}$$

Once we have this marginal likelihood function and its derivatives, gradient-based methods (e.g., conjugate gradient

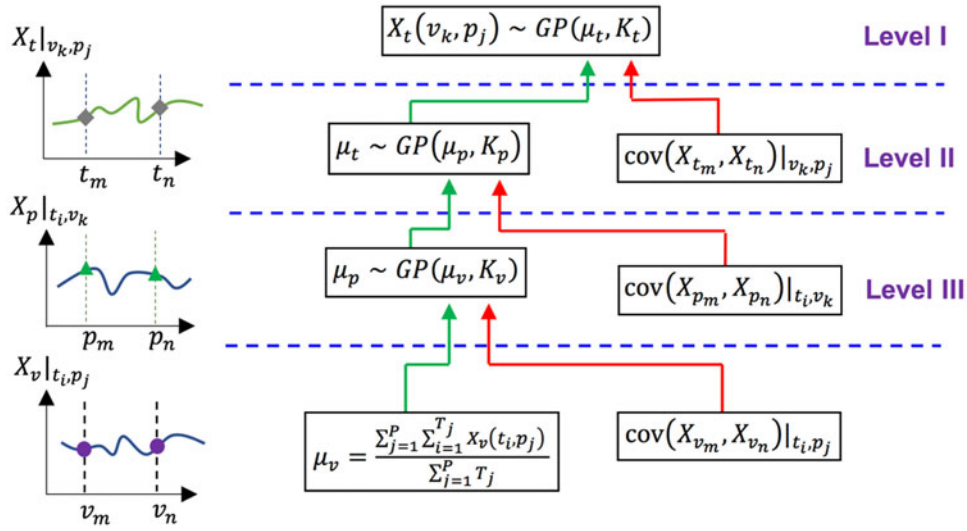

Figure 3. The framework of the multi-level NGP approach.

descent or resilient back propagation (Rprop)) can be taken into account to solve the problem (Riedmiller and Braun, 1993). We implement the fast Rprop algorithm, which only relies on the sign of the gradient:

$$\theta_i^{(q+1)} = \theta_i^{(q)} - sign\left(\frac{\partial \log p(\boldsymbol{x}\,|\,\boldsymbol{t},\boldsymbol{\theta})^{(q)}}{\partial \theta_i}\right)\Delta_i^{(q)} \tag{13}$$

where $\theta_i^{(q)}$ shows the ith component in hyper-parameters set $\boldsymbol{\theta}$ and q expresses the replication number. If the sign function in Eq. (13) remains the same, the update value Δ_i is increased by a factor $\xi^+>1$ so that the convergence of the algorithm is accelerated, and the step size is decreased by a factor $\xi^-<1$ if the direction reverses as:

$$\Delta_i^{(q)} = \begin{cases} \xi^+\Delta_i^{(q-1)} & if\left(\frac{\partial \log p(\boldsymbol{x}\,|\,\boldsymbol{t},\boldsymbol{\theta})^{(q-1)}}{\partial \theta_i}\right)\times\left(\frac{\partial \log p(\boldsymbol{x}\,|\,\boldsymbol{t},\boldsymbol{\theta})^{(q)}}{\partial \theta_i}\right)>0 \\ \xi^-\Delta_i^{(q-1)} & if\left(\frac{\partial \log p(\boldsymbol{x}\,|\,\boldsymbol{t},\boldsymbol{\theta})^{(q-1)}}{\partial \theta_i}\right)\times\left(\frac{\partial \log p(\boldsymbol{x}\,|\,\boldsymbol{t},\boldsymbol{\theta})^{(q)}}{\partial \theta_i}\right)<0 \\ \Delta_i^{(q-1)} & otherwise \end{cases} \tag{14}$$

3.2. Nested Gaussian process

Figure 3 shows the framework of the proposed NGP approach that jointly considers multi-dimensional covariance structures among time, patients and variables in a hierarchical way. Suppose there are V variables of P patients for the time duration of T in order-3 tensor data (see Fig. 1(b)). Let $X_t(v_k,p_j)$ denote the value of tensor data at time t for the variable v_k of patient p_j. First, we construct a level I GP model as:

$$X_t(v_k,p_j) \sim \mathcal{GP}(\mu_t, K_t) \tag{15}$$

where μ_t is the mean function and K_t is the covariance function (see Fig. 3). As the observations at two different time points tend to have a stronger correlation when they are closer to each other, we use the covariance function K_t as:

$$cov(X_{t_m}, X_{t_n})\,|\,v_k,p_j = \sigma_t^2 \exp\left[-\frac{(t_m-t_n)^2}{2l_t^2(v_k,p_j)}\right] \tag{16}$$

where σ_t^2 is the signal variance in the dimension of time, and $l_t(v_k,p_j)$ the length scale. Note that X_{t_m} and X_{t_n} should be similar if t_m and t_n are sufficiently close in the temporal dimension. Therefore, the length scale $l_t(v_k,p_j)$ defines the closeness in the time domain. Second, we model μ_t using a level II GP model as:

$$\mu_t \sim \mathcal{GP}(\mu_p, K_p) \tag{17}$$

where K_p represents the covariance between patients (e.g., p_m and p_n). The proposed hierarchical design is aimed at incorporating nonstationarity in the underlying stochastic process through GP modeling of mean functions. If two patients share closer values in clinical variables, they tend to have a stronger correlation. Therefore, the covariance function K_p is defined as:

$$cov(X_{p_m}, X_{p_n})\,|\,t_i,v_k = \sigma_p^2 \exp\left[-\frac{(X_{p_m}-X_{p_n})^2}{2l_p^2}\right] \tag{18}$$

Third, the mean function of μ_p is modeled as a level III GP

$$\mu_p \sim \mathcal{GP}(\mu_v, K_v) \tag{19}$$

where the covariance function K_v captures the similarity between clinical variables (e.g., v_m and v_n), and it is akin to the definition of K_p. Here, μ_v is the average of tensor data across the time and patients, for a specific clinical variable:

$$\mu_v = \frac{\sum_{j=1}^{p}\sum_{i=1}^{T_j} X_v(t_i,p_j)}{\sum_{j=1}^{p} T_j} \tag{20}$$

In this equation, T_j is the time duration for patient p_j. An attractive feature of the NGP model is the capability and flexibility to predict missing values in high-dimensional tensor data. We investigate the missing data imputation at time t^* of the variable v^* for a patient p^*. Based on the observed values $X_t(v^*,p^*)$ at time stamp t, the posterior mean of level

I GP is given as:

$$\bar{X}_{t^*}(v^*,p^*) = \bar{X}_t(v^*,p^*) + K_{t^*\boldsymbol{t}}\left[K_{\boldsymbol{tt}} + \sigma_{nt}^2\boldsymbol{I}\right]^{-1}\left[X_t(v^*,p^*) - \mu_t(v^*,p)\right] \tag{21}$$

where $K_{t^*\boldsymbol{t}}$ is the temporal covariance, σ_{nt}^2 is the noise variance in temporal domain and $\bar{X}_t(v^*,p^*)$ can be calculated as follows:

$$\bar{X}_t(v^*,p^*) = \frac{\sum_t X_t(v^*,p^*)}{T_{v^*p^*}} \tag{22}$$

$T_{v^*p^*}$ is the time duration of the variable v^* for patient p^*. The mean function in Eq. (21) can be derived from level II GP as:

$$\mu_t(v^*,p) = X_{p^*}(t,v^*) + K_{p^*\boldsymbol{p}}\left[K_{\boldsymbol{pp}} + \sigma_{np}^2\boldsymbol{I}\right]^{-1}\left[\bar{X}_t(t,v^*) - \mu_t(t,v^*)\right] \tag{23}$$

where $K_{p^*\boldsymbol{p}}$ is the covariance between patients, σ_{np}^2 is the noise variance in patient domain and $\bar{X}_p(t,v^*)$ is mean for v^* across all patients in the dataset. Consequently, we have $\mu_p(t,v^*)$ in the level III GP model as follows:

$$\mu_p(t,v^*) = X_{v^*}(t,p^*) + K_{v^*\boldsymbol{v}}\left[K_{\boldsymbol{vv}} + \sigma_{nv}^2\boldsymbol{I}\right]^{-1}\left[\bar{X}_v(t,p) - \mu_v\right] \tag{24}$$

where $K_{v^*\boldsymbol{v}}$ is the covariance between variables, σ_{nv}^2 is the noise variance in variable domain and $\bar{X}_v(t,p)$ is the prior of variable v for all patients in the datasets. In NGP, we have a set of hyper-parameters for each level. In the first level (i.e., time domain), hyper-parameters are signal standard deviation σ_t, noise standard deviation σ_{nt} and length scale $l_t(v_k,p_j)$, which belong to the squared exponential covariance function. In the second and third levels (i.e., patient and variable domains), we have a similar set of hyper-parameters. Note that the NGP hyper-parameters are updated from the third level to the first level sequentially. The updating procedure for NGP hyper-parameters can be obtained via Eq. (10) to Eq. (14).

4. Experimental design and materials

In this article, tensor data from both simulation and real-world ICU are utilized to evaluate and validate the proposed NGP approach.

4.1. Simulation study

As shown in Table 1, we simulate an order-3 tensor data with dimensions of time (T), patients (P) and variables (V). The tensor data are initialized as 100 time points, 50 patients, and 20 variables. We also choose the type of mean, covariance functions and inference method from a design of experiments. For each patient, hyper-parameter values are sampled standard normal distribution contaminated by exogenous noises (i.e., additive white Gaussian noise) representing the heterogeneous population characteristics. Then, the covariance function for a specific patient and the related variable is evaluated. Next, singular value decomposition (SVD) is utilized to decompose the positive-definite covariance matrix of GP (i.e., K_t) for an efficient and robust numerical solution. The SVD of a matrix K_t is the factorization of it into the product of three matrices $K_t = USQ^T$ where the columns of U and Q^T are orthonormal and the matrix S is diagonal with positive real entries. Finally, column-specific data in the tensor (i.e., a specific patient, specific variable over time) are obtained by multiplying generated data with the SVD results and adding evaluated mean result.

Table 1. The algorithm for tensor data generation in the simulated study.

Input:
$T \leftarrow$ the length of time period
$V \leftarrow$ total number of variables
$P \leftarrow$ total number of patients

For $j = 1 : P$ // loop for patients
// initialize hyper-parameters for each patient
  $\theta = \{\theta_m, \theta_c\} \quad \forall m = 1 : M$ and $c = 1 : C$
  $\theta_m \leftarrow rand(M)$ // generate M normally distributed random numbers
  $\theta_c \leftarrow rand(C)$ // generate C normally distributed random numbers
  For $k = 1 : V$ // loop for variables
  // generate variation for the hyper-parameters for each variable
    $\theta_m \leftarrow \theta_m + noise()$
    $\theta_c \leftarrow \theta_c + noise()$ // noise > 0
    // evaluate covariance matrix and mean value
    $\mu_t \leftarrow \mu_t$ *and* θ_m
    $K_t \leftarrow K_t$ *and* θ_c
    $\eta \leftarrow$ *Generate pseudo-random numbers*
    $[U, S, Q^T] = svd(K_t)$ // singular value decomposition
    $z(k,:,j) = U \times sqrt(S) \times \eta + \mu_t$
  End
End
Output: z

Next, we randomly select a patient, a variable and remove 25%, 50% and 75% of temporal observations (i.e., three scenarios), and then impute missing values with the proposed NGP. Figure 4 illustrates the NGP imputation of order-3 tensor data for a randomly chosen patient and variable, where blocks with the red color are missing data to be imputed.

We performed the experiments for three different missing percentages, each for 100 replications for robust estimation. The root-mean-square error (RMSE) is utilized as the performance metric in the experiments. We also evaluated the performance of missing data imputation with different prior mean, inference, and covariance functions with GP and NGP models. Figure 5 shows the cause-and-effect diagram of the experimental design.

As shown in Fig. 5, we examine two types of prior mean function, zero and linear, two types of covariance function, Squared Exponential (SEard) and Isotropic Squared Exponential (SEiso), as well as two types of inference function, Exact (see section 3) and Laplace. Note that, for NGP, zero and linear prior mean function are only considered in the third level (i.e., μ_v) as the prior mean in the first and second levels (i.e., μ_t and μ_p) are estimated from GP results in the second and third levels, respectively.

4.2. Real-world case study

Furthermore, we use the MIMIC-II Clinical Database (Goldberger *et al.*, 2000; Saeed *et al.*, 2011) that consists of

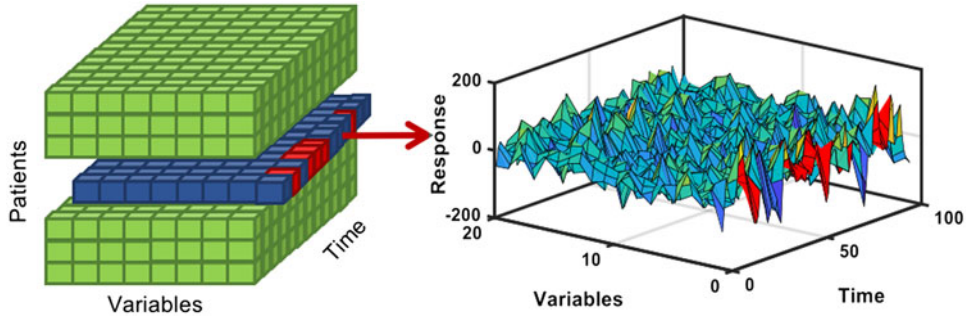

Figure 4. The schematic view of missing data imputation in tensor-form structure.

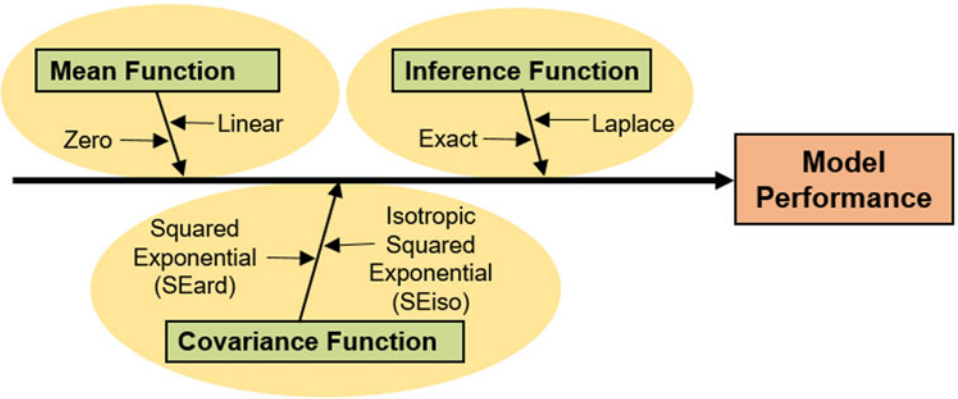

Figure 5. The design of a simulation experiment to evaluate and validate the proposed NGP approach.

clinical data collected from patients admitted to ICUs in the Beth Israel Deaconess Medical Center in Boston. The missing data imputation for MIMIC II database includes the following steps:

1. ***Categorization:*** 26 ICU variables are categorized into three groups (namely general descriptor, one-hour sampling, four-hour sampling) according to the fraction of missing data and sampling frequency per variable. General descriptors are a group of variables including general attributes of a patient that are recorded during the patient's first visit to the ICU; e.g., Record ID, Height, Mechanical Ventilation (Mech Vent), Age, ICU Type and Gender. One-hour sampling is a group of variables with an approximate sampling rate of 12 per 24 hours or less. The group of four-hour sampling includes those variables with the sampling rate less than 6 per 24 hours. Figure 6 shows the fraction of missing data for clinical variables in the MIMIC-II database. Note that general descriptors (i.e., six variables) are removed due to the fact that they are documented only at the start of ICU stay. It is worth mentioning that none of the variables are completely observed for all patients. Here, amongst variables that have missing values, Lactate has the highest missing percentage at 45.42%, and HR is the lowest with 1.57%.
2. ***Preprocessing:*** We preprocess the data based on clinical inputs and expert knowledge from physicians. The details of preprocessing steps are shown in Table 2. First, invalid height and weight values were excluded, and missing height/weight values were substituted via a regression model and with consideration of standard values of height/weight by the sex and age group. Then, TroponinT was multiplied by a constant (i.e., 100) and then pooled with another type of regulatory protein (TroponinI) as a new clinical variable called Troponin. In the third step, Creatinine is substituted by creatinine clearance, which is achieved by solving the Cockcroft Gault equation:
3. $$CreatinineClearance = \frac{(140-Age) \times Weight \times (0.85 + 0.15 \times Gender)}{(72 \times Creatinine)} \quad (25)$$
4. Fourth, we used a variable, Urine.Sum, the aggregate sum of the Urine measurements, to replace the old variable Urine, which is more informative for use the physicians. Finally, three pairs of variables (i.e., NIMAP, SysABP and NISysABP, DiasABP and NIDiasABP, MAP) were pooled respectively as three new variables.
5. ***Missing data imputation:*** We performed missing data imputation on the MIMIC-II clinical database with 4000 patient records and over 48 hours of ICU stays and for 26 variables. Here, we randomly chose a patient and a variable from the MIMIC-II dataset and randomly removed 25%, 50% and 75% of data (see Fig. 4, the red blocks in tensor). In each case, we impute the missing values via NGP, GP, Geo-kriging and KNN methods and repeat this procedure for 100 different sets of patient and variable. It may be noted that if a

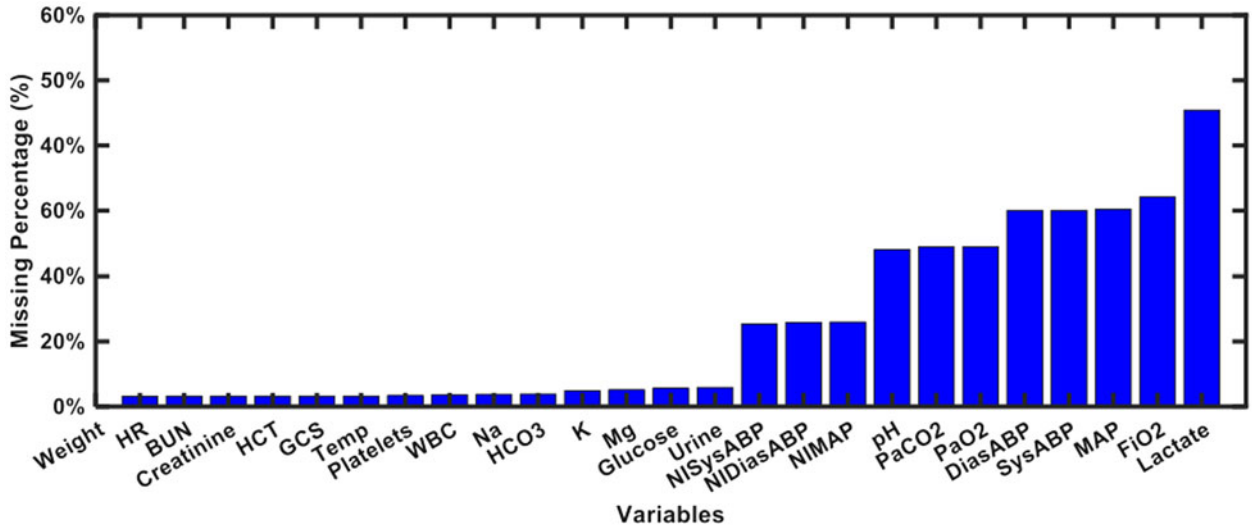

Figure 6. The percentage of missing data for ICU variables in the real-world case study.

Table 2. ICU data categorization, characteristics, and preprocessing (Chen *et al.*, 2016).

Data Category	Variables	Normal Range	Missing Percentage (%)	# Observations (median ±) std	Data Processing
General Descriptor	Record ID		All Available	Observed only at the starting point of the ICU stay	
	Height				
	Gender				
	Age				
	ICU Type				
	Mech Vent		35.87	7 ± 7.56	1: patient required mechanical ventilation; 0 otherwise
Four-hour Sampling	BUN	6–20	1.6	3 ± 1.68	
	Lactate	3.7–5.2	45.42	1 ± 3.15	
	Creatinine	F: 0.6–1.1 M: 0.7–1.3	1.6	3 ± 1.7	Change to Creatinine Clearance[c]
	HCO3	23–29	1.9	3 ± 1.7	
	Glucose	70–100	2.82	3 ± 1.8	
	Mg	1.7–2.2	2.57	3 ± 1.77	
	K	0.5–2.2	2.4	3 ± 1.92	
	WBC	4.5–10	1.82	3 ± 1.57	
	PaCO2	35–45	24.42	5 ± 5.72	
	Na	135–145	1.87	3 ± 1.86	
	HCT	F: 35–48 M: 40–53	1.6	4 ± 2.58	
	Platelets	150–450	1.7	3 ± 1.91	
	pH	7.38–7.42	24	5 ± 5.91	
	PaO2	75–100	24.42	5 ± 5.71	
One-hour Sampling	GCS	0–3	1.6	13 ± 7.88	
	FiO2	0.21–0.5	32.07	8 ± 7.34	
	Urine	1500	2.92	37 ± 12.49	Change to Urine.Sum[d]
	Temp	36–40	1.6	14 ± 17.45	
	HR	60–100	1.57	55 ± 16.05	
	Weight		All Available	37 ± 26.43	[a]
	MAP	70–100	30.2	21 ± 20.48	[a]
	NIMAP	70–100	12.97	21 ± 20.7	
	NISysABP	100–140	12.67	21 ± 20.7	[a]
	SysABP	100–140	30.02	43 ± 29.59	
	NIDiasABP	60–90	12.92	21 ± 20.7	[a]
	DiasABP	60–90	30.02	43 ± 29.57	

[a]Invalid values were omitted, and missing values were substituted by regression model and according to the standard values by sex.
[b]Pool TroponinI and constant (i.e., 100)* TroponinT as a new variable called Troponin.
[c]Creatinine Clearance is ased on 13.
[d]Urine.Sum is the aggregate sum of urine measurements.
[e]Merge two variables together.

randomly chosen patient and his/her associated variable has a high level of missing data (i.e., more than 75%), we exclude this case from analysis and randomly select another case because it is difficult to set up the ground truth for the experimental scenario with 75% missing data already in the database.

Table 3. The performance comparison of GP and NGP tensor-data imputations in the simulation study under different mean, inference and covariance functions.

			RSME		
Mean function	Inference function	Covariance function	GP	NGP	Improvement (%)
Zero	Exact	SEard	26.23 (±21.99)	12.40 (±9.74)	49.41%
		SEiso	33.28 (±17.23)	13.56 (±8.01)	56.70%
	Laplace	SEard	38.64 (±13.76)	13.68 (±5.08)	63.06%
		SEiso	34.03 (±20.13)	13.91 (±10.20)	59.12%
Linear	Exact	SEard	21.78 (±14.13)	11.15 (±7.84)	51.68%
		SEiso	41.55 (±17.69)	13.23 (±8.10)	67.02%
	Laplace	SEard	30.16 (±7.16)	9.78 (±4.54)	67.02%
		SEiso	21.90 (±16.74)	7.60 (±8.09)	60.30%

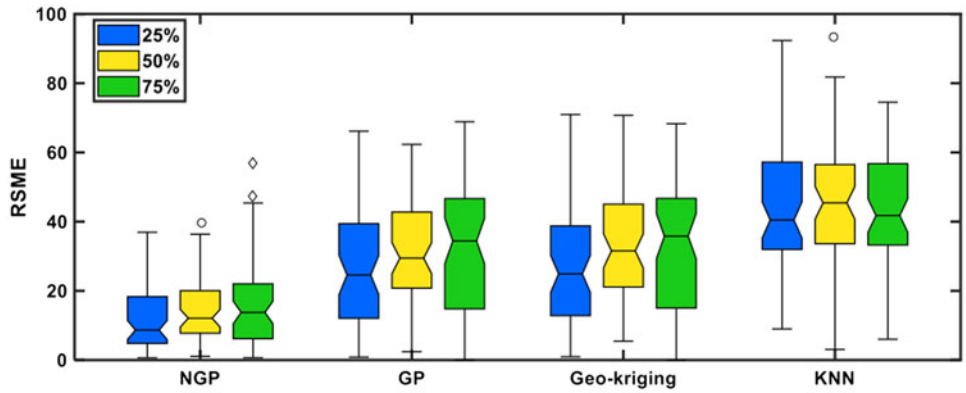

Figure 7. Performance comparison of missing data imputation with NGP, GP, Geo-kriging and KNN models in the simulation study. The imputation results are obtained from 100 replicates in each experimental scenarios for 25%, 50% and 75% missing percentage.

5. Experimental results

5.1. Simulation study

Table 3 shows the performance comparison between GP and NGP models with different mean, inference and covariance functions in the simulation experiments (see Fig. 5). The RMSE values in Table 3 are the average results of 100 replications for the experimental scenario of 25% missing percentage in the tensor data. Note that NGP leads to smaller errors in comparison with GP, regardless of the types of mean, inference and covariance functions. The best improvement of NGP occurs when the prior mean function for variables (i.e., μ_v) is considered as a linear function, which provides more flexibility than zero function. This experiment compares the performance of different mean, covariance and inference functions for the GP and NGP.

Furthermore, we have conducted experiments to benchmark the performance of NGP models with other popular imputation methods, such as Geo-kriging and KNN imputation. It is worth mentioning that GP utilizes only the temporal information for performing imputation, while Geo-kriging is commonly used for spatial inference and estimates the missing values from spatial covariance among the sample values. Also, KNN is an instance-based learning algorithm and predicts missing value according to the closest training instance in the predefined neighborhood.

As shown in Fig. 7, the proposed NGP model with hierarchical covariance functions yields the lowest RMSE values compared to other imputation methods. Specifically, NGP registered a reduction of RMSE of 51.9%, 54.01% and 67.81% when compared to GP, Geo-kriging and KNN, respectively. The NGP leverages the multi-dimensional time, patient and variable correlations for tensor missing data imputation, as opposed to the conventional onedimensional correlation, and thereby achieves better results in the experiments.

5.2. Real-world case study

In addition to simulation experiments, we have also performed a real-world case study on the MIMIC-II database. As shown in Fig. 8, the average RMSE of the NGP model is significantly lower than GP, Geo-kriging and KNN imputation methods for 12 one-hour sampling variables (i.e., high sampling rate) as well as 14 four-hour sampling variables (i.e., medium sampling rate) for all three levels of missing percentages. Notably, the RMSE for high sampling variables is relatively large owing to the huge inherent fluctuations in their values compared to the medium sampling variables. The results show that NGP imputation decreases the error rate on average by 44.76% in comparison with GP. Also, it reduces the error rate by 30.04% and 83.68% in comparison with Geo-kriging and KNN, respectively. Experimental results demonstrate that the proposed NGP is more effective in the imputation performance of missing data in the high-dimensional tensor form.

Figure 9 shows an illustration of tensor data imputation for NGP, GP, Geo-kriging and KNN models based on a real-world case study. Here, the performance comparison is between the imputed values (and confidence interval when applicable) and the ground truth for Patient ID 102, variable urine and missing percentage of 50% by NGP, GP, Geo-

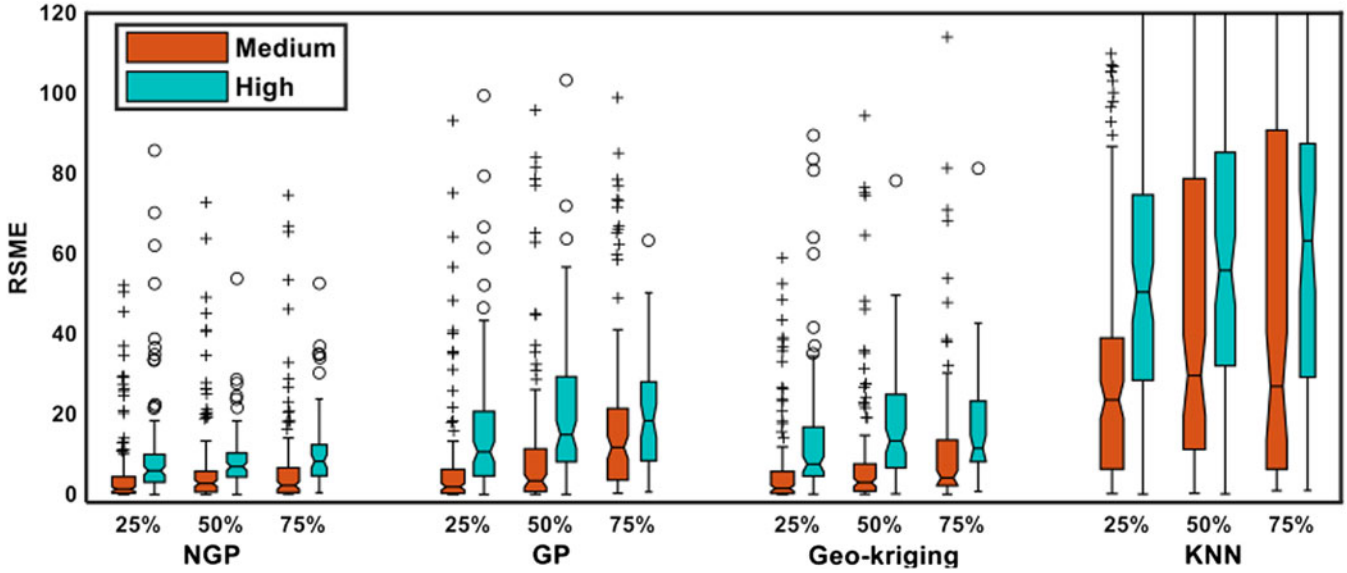

Figure 8. Performance comparison of missing data imputation with NGP, GP, Geo-kriging and KNN models in the real-world case study. The imputation results are obtained from 100 replicates in each experimental scenario for 25%, 50% and 75% missing percentage.

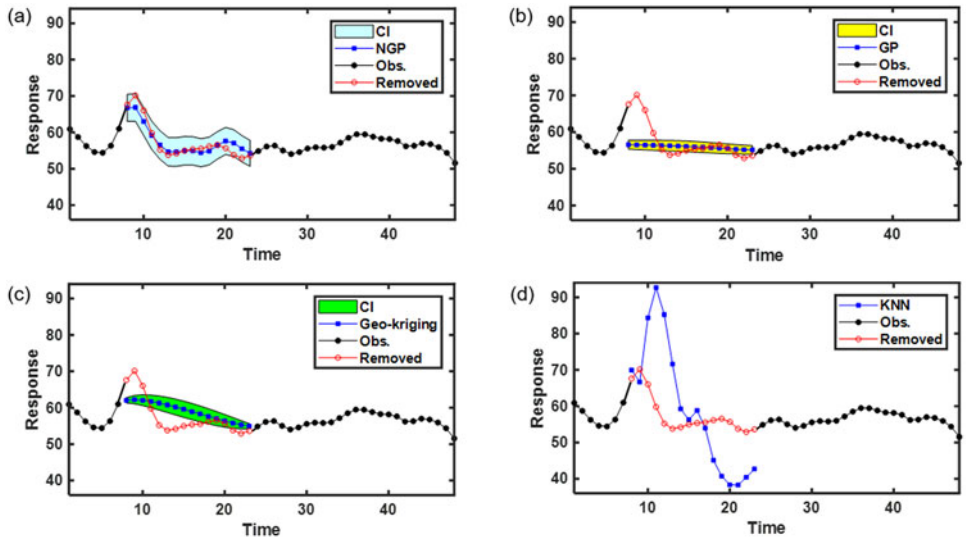

Figure 9. An illustration of tensor data imputation with (a) NGP, (b) GP, (c) Geo-kriging and (d) KNN in a real-world case study for the patient with ID 102, variable urine and the missing percentage of 50%.

kriging and KNN. As shown in Fig. 9(a), NGP effectively captures the trend of data through multi-dimensional covariance structure of time, patient and variable and imputes missing values close to the ground truth. When there is only sparse temporal information (i.e., a considerable number of consecutive missing values), the prediction error of other imputation methods increases drastically, while remaining low for the NGP model. Note that GP, Geo-kriging and KNN yield pronounced RMSE values at 4.986, 3.765 and 45.018, respectively, compared to 1.819 achieved by NGP.

6. Conclusions

Recent progress in advanced sensing and information technology provides a variety of measurement and monitoring systems to improve the quality of ICU care, leading to the data-rich environment in the hospital. ICU data present unique characteristics, such as patient heterogeneity, time asynchronization, and variable heterogeneity in the new tensor form. The prevalent issue of missing data undermines the data-driven decision making in the ICU setting, because conventional approaches focus more on the table-form data and are less concerned about high-dimensional tensor data. These limitations make it difficult for conventional imputation approaches to handle tensor-form ICU datasets. To this end, we propose a novel nested Gaussian process (NGP) method to manipulate the uncertainty and incompleteness in the high-dimensional tensor-form data. Specifically, we model the multi-dimensional correlation structure of a time-patient-variable interrelationship for statistical inference and prediction of missing values. The proposed approach is evaluated and validated with both simulation studies and real-world ICU datasets from the MIMIC-II database.

Experimental results show that, on average, the NGP model diminishes prediction error by 48.3% compared to GP, 42.0% for Geo-kriging, and 75.7% for KNN with the different percentage of missing ICU data. Remarkably, the NGP method registers strikingly better performance for the high level of missing data scenario than traditional imputation methods attributed to the innovative hierarchical structure of time-patient-variable covariance in ICU settings. The proposed NGP model can also be generally applicable in a variety of engineering and medical domains that entail high-dimensional data imputation and analytics.

Acknowledgments

The authors would like to thank editors and anonymous reviewers for their constructive comments and suggestions to improve the quality of this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors would like to thank the National Science Foundation (No. CMMI-1646660) for the support of this research. The author (HY) thanks Harold and Inge Marcus Career Professorship for additional financial support.

References

Chen, Y., Leonelli, F., and Yang, H. (2016) Heterogeneous sensing and predictive modeling of postoperative outcomes. In H. Yang and E. K. Lee (Eds.), *Healthcare Analytics: From Data to Knowledge to Healthcare Improvement* (pp. 463–501). Hoboken, NJ: Wiley.

Chen, Y., and Yang, H. (2014) Heterogeneous postsurgical data analytics for predictive modeling of mortality risks in intensive care units. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* (pp. 4310–4314).

Chen, Y., and Yang, H. (2016a) A novel information-theoretic approach for variable clustering and predictive modeling using dirichlet process mixtures. *Scientific Reports*, **6**, 38913.

Chen, Y., and Yang, H. (2016b) Sparse modeling and recursive prediction of space–time dynamics in stochastic sensor networks. *IEEE Transactions on Automation Science and Engineering*, **13**(1), 215–226.

Cheng, C. (2018) Multi-scale Gaussian process experts for dynamic evolution prediction of complex systems. *Expert Systems with Applications*, **99**, 25–31.

Cheng, C., Sa-Ngasoongsong, A., Beyca, O., Le, T., Yang, H., Kong, Z., and Bukkapatnam, S. T. (2015) Time series forecasting for nonlinear and non-stationary processes: A review and comparative study. *IIE Transactions*, **47**(10), 1053–1071.

Cismondi, F., Fialho, A. S., Vieira, S. M., Reti, S. R., Sousa, J. M., and Finkelstein, S. N. (2013) Missing data in medical databases: Impute, delete or classify? *Artificial Intelligence in Medicine*, **58**(1), 63–72.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodologi-cal)*, **39**(1), 1–38.

Fuller, W. A., and Kim, J. K. (2005) Hot deck imputation for the response model. *Survey Methodology*, **31**(2), 139–149.

García-Laencina, P. J., Abreu, P. H., Abreu, M. H., and Afonoso, N. (2015) Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values. *Computers in Biology and Medicine*, **59**, 125–133.

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., and Stanley, H. E. (2000) PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, **101**(23), 215–220.

Haukoos, J. S., and Newgard, C. D. (2007) Advanced statistics: Missing data in clinical research part 1: An introduction and conceptual framework. *Academic Emergency Medicine*, **14**(7), 662–668.

Johnson, A. E., Ghassemi, M. M., Nemati, S., Niehaus, K. E., Clifton, D. A., and Clifford, G. D. (2016) Machine learning and decision support in critical care. *Proceedings of the IEEE*, **104**(2), 444–466.

Joseph, V. R. (2006) Limit kriging. *Technometrics*, **48**(4), 458–466.

Joseph, V. R., and Kang, L. (2011) Regression-based inverse distance weighting with applications to computer experiments. *Technometrics*, **53**(3), 254–265.

Kenward, M. G., and Carpenter, J. (2007) Multiple imputation: Current perspectives. *Statistical Methods in Medical Research*, **16**(3), 199–218.

Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Frangakis, C., Hogan, J. W., Molenberghs, G., Murphy, N. J., Susan, A, Rotnitzky, A., Scharfstein, D., Shih, W. J., Siegel, J. P., and Stern, H. (2012) The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, **367**(14), 1355–1360.

Little, R. J., and Rubin, D. B. (2014) *Statistical Analysis with Missing Data* (Vol. **333**). Hoboken, NJ: John Wiley & Sons.

Little, R. J., and Schenker, N. (1995) Missing data. In *Handbook of Statistical Modeling for the Social and Behavioral Sciences* (pp. 39–75). Boston, MA: Springer.

Liu, G., and Yang, H. (2013) Multiscale adaptive basis function modeling of spatiotemporal vectorcardiogram signals. *IEEE Journal of Biomedical and Health Informatics*, **17**(2), 484–492.

Liu, G., and Yang, H. (2018) Self-organizing network for variable clustering. *Annals of Operations Research*, **263**(1-2), 119–140.

Newgard, C. D., and Haukoos, J. S. (2007) Advanced statistics: Missing data in clinical research part 2: Multiple imputation. *Academic Emergency Medicine*, **14**(7), 669–678.

Pan, R., Yang, T., Cao, J., Lu, K., and Zhang, Z. (2015) Missing data imputation by k nearest neighbours based on grey relational structure and mutual information. *Applied Intelligence*, **43**(3), 614–632.

Pastores, S. M., Dakwar, J., and Halpern, N. A. (2012) Costs of critical care medicine. *Critical Care Clinics*, **28**(1), 1–10.

Pelckmans, K., De Brabanter, J., Suykens, J. A., and De Moor, B. (2005) Handling missing values in support vector machine classifiers. *Neural Networks*, **18**(5-6), 684–692.

Riedmiller, M., and Braun, H. (1993) A direct adaptive method for faster backpropagation learning: The RProp algorithm. In IEEE International Conference on Neural Networks (pp. 586–591).

Saeed, M., Villarroel, M., Reisner, A. T., Clifford, G., Lehman, L. W., Moody, G., Heldt, T., Kyaw, T. H., Moody, B., and Mark, R. G. (2011) Multiparameter intelligent monitoring in intensive care II (MIMIC-II): A public-access intensive care unit database. *Critical Care Medicine*, **39**(5), 952.

Schafer, J. L., and Graham, J. W. (2002) Missing data: Our view of the state of the art. *Psychological Methods*, **7**(2), 147–177.

Sinharay, S., Stern, H. S., and Russell, D. (2001) The use of multiple imputation for the analysis of missing data. *Psychological Methods*, **6**(4), 317–329.

Vesin, A., Azoulay, E., Ruckly, S., Vignoud, L., Rusinovà K., Benoit, D., Soares, M., Azeivedo-Maia, P., Abroug, F., Benbenishty, J., and Timsit, J. F. (2013) Reporting and handling missing values in clinical studies in intensive care units. *Intensive Care Medicine*, **39**(8), 1396–1404.

Williams, C. K., and Rasmussen, C. E. (1996) Gaussian processes for regression. In *Advances in Neural Information Processing Systems* (pp. 514–520). Cambridge, MA: MIT Press.

Yang, H., Bukkapatnam, S. T., and Komanduri, R. (2012) Spatiotemporal representation of cardiac vectorcardiogram (vcg) signals. *Biomedical Engineering Online*, **11**(16), 1–15.

Yang, H., Kan, C., Liu, G., and Chen, Y. (2013) Spatiotemporal differentiation of myocardial infarctions. *IEEE Transactions on Automation Science and Engineering*, **10**(4), 938–947.

Yang, H., Kundakcioglu, E., Li, J., Wu, T., Mitchell, J. R., Hara, A. K., Pavlicek, W., Hu, L. S., Silva, A. C., Zwart, C. M., et al. (2014) Healthcare intelligence: turning data into knowledge. *IEEE Intelligent Systems*, **29**(3), 54–68.

Zhu, R., Yao, B., Yang, H., and Leonelli, F. (2019) Optimal sensor placement for space–time potential mapping and data fusion. *IEEE Sensors Letters*, **3**(1), 1–4.