Biometrika (2024), **111**, 4, *pp.* 1257–1275
https://doi.org/10.1093/biomet/asae015
Advance Access publication 8 April 2024

Individualized dynamic latent factor model for multi-resolutional data with application to mobile health

BY J. ZHANG

Department of Statistics, University of California, Irvine, Donald Bren Hall 2219, Irvine, California, 92697, U.S.A.
jiuchez@uci.edu

F. XUE

Department of Statistics, Purdue University, 150 N. University St, West Lafayette, Indiana, 47907, U.S.A
feixue@purdue.edu

Q. XU

Department of Statistics, University of California, Irvine, Donald Bren Hall 2219, Irvine, California, 92697, U.S.A.
qxu6@uci.edu

J. LEE

Sue & Bill Gross School of Nursing, University of California, Irvine, Nursing & Health Sciences Hall, Office 4305, Irvine, California, 92697, U.S.A.
jungahl@uci.edu

AND A. QU

Department of Statistics, University of California, Irvine, Donald Bren Hall 2212, Irvine, California, 92697, U.S.A.
aqu2@uci.edu

SUMMARY

Mobile health has emerged as a major success for tracking individual health status, due to the popularity and power of smartphones and wearable devices. This has also brought great challenges in handling heterogeneous, multi-resolution data that arise ubiquitously in mobile health due to irregular multivariate measurements collected from individuals. In this paper, we propose an individualized dynamic latent factor model for irregular multi-resolution time series data to interpolate unsampled measurements of time series with low resolution. One major advantage of the proposed method is the capability to integrate multiple irregular time series and multiple subjects by mapping the multi-resolution data to the latent space. In addition, the proposed individualized dynamic latent factor model is applicable to capturing heterogeneous longitudinal information through individualized dynamic latent factors. Our theory provides a bound on the integrated interpolation error and the convergence rate for *B*-spline approximation methods. Both the simulation studies and the application to smartwatch data demonstrate the superior performance of the proposed method compared to existing methods.

Some key words: Data integration; Interpolation; Mobile health; Nonparametric approximation; Wearable device data.

1. Introduction

With recent developments in technology, mobile health has begun to play an important role in personalized treatment and intervention due to the wide usage of smartphones and wearable devices. A wealth of longitudinal data from wearable devices track people's physical activities and health status, which enables us to deliver noninvasive interventions in real time. To address the unique challenges presented by these data, including their high heterogeneity, multi-resolution and nonlinearity, we need to develop statistical methods, theories and computational tools. The data often include both dense observations over a long period of time and also sparse observations due to a large proportion of missing data. Figure 1 illustrates an example of mobile health data for monitoring the heart rate, stress and daily wellness, where the heart rate is measured much more frequently than stress and daily wellness.

In this paper, we are particularly interested in irregular multi-resolution time series as the data present three aspects of irregularity (Sun et al., 2020): irregular intra-series due to irregular time intervals within each time series, irregular inter-series due to varying sampling rates among multivariate time series from the same subject and irregular inter-subject measurement variations due to different time stamps across different subjects. In addition to irregular and multi-resolution data features, subjects can be highly heterogeneous in terms of demographics, genetic characteristics, medical history, lifestyle and many unobserved attributes (Conway et al., 2011). Thus, each subject is expected to have a unique trajectory on measurements of interest. Traditional homogeneous models are no longer suitable for this type of data (Petris et al., 2009; Wang et al., 2016; Hamilton, 2020), so individualized modelling and learning for heterogeneous data are in great demand.

In particular, we are motivated by a stress management study for caregivers of dementia patients that uses mobile health data to intervene with subjects experiencing high stress (Lee & Gibbs, 2021). To administer intervention for subjects experiencing high stress, we need to capture the trajectories of the subject's physiological information such as the heart rate, heart rate variability, physical activities and daily wellness. Some measurements are low-resolution time series, and therefore interpolating unobserved data from irregular multi-resolution time series could play an important role in performing downstream analyses in prediction, classification or clustering (Jensen et al., 2012).

Traditional polynomial and spline methods can provide interpolation (De Boor, 1978) for a single time series. However, they are not effective for incorporating correlations among time series, which may lead to information loss shared by multivariate time series from the same subject. Multivariate time series (Hamilton, 2020) and the dynamic linear model (Petris et al., 2009) are capable of handling multiple time series with missing values (Gómez et al., 1999). However, the high missing rate of low-resolution time series makes it difficult to infer and predict trajectories of longitudinal data (Jones, 1980), especially when the occurrence of ultra-sparse time series could lead to degenerated interpolation of missing

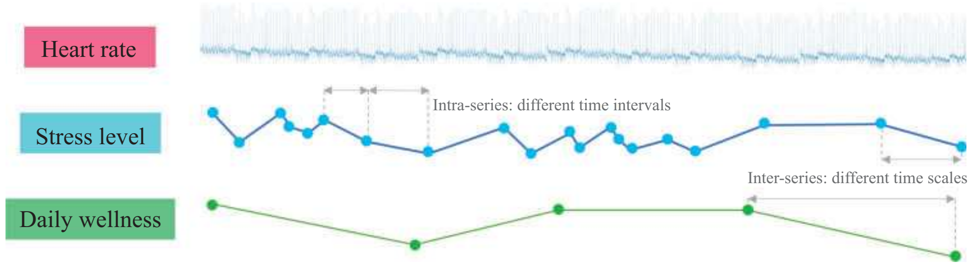

Fig. 1. Intra-series and inter-series irregularities of irregular multi-resolution time series in wearable device data.

values due to a large gap from prior observations. In addition, existing approaches require the stationarity assumption that might be difficult to satisfy or verify.

Functional data methods such as functional principal component analysis (Hall & Hosseini-Nasab, 2006; Yao & Lee, 2006; Wang et al., 2016) and functional regression (James & Hastie, 2001; Yao et al., 2005) are useful when analysing longitudinal data. Among these works in functional data analysis, to address the classification of new curves and account for subject heterogeneity, James & Hastie (2001) extended linear discriminant analysis to functional data. They modelled each predictor as a smooth curve and transformed the curve to a vector of coefficients through basis functions. James (2002) further applied a similar idea to generalized linear models with functional predictors. Additionally, James et al. (2000) proposed a principal component method for irregular and sparse data. These existing methods are designed for a single outcome. For instance, James (2002) modelled each time series separately when dealing with multiple time series in prediction models, which ignored the potential correlations among multivariate time series.

To the best of our knowledge, only a limited literature addresses multivariate functional data analysis (Berrendero et al., 2011; Chiou & Müller, 2014, 2016; Jacques & Preda, 2014), and even fewer consider the case of irregular observed data (Happ & Greven, 2018). Specifically, Happ & Greven (2018) introduced a functional principal component analysis tailored for multivariate functional data with varying dimensions. Although the covariance between time series are considered in multivariate functional data analyses, they generally do not address heterogeneity from different subjects. Volkmann et al. (2023) proposed an additive mixed-effect model for multivariate functional data, where random effects incorporate subject heterogeneity for each subgroup. However, when subject trajectories do not exhibit grouping structures, the mixed-effect model has a limitation.

The interactive fixed-effect models also incorporate subject heterogeneity in longitudinal data/time series data through interactions between individual effect factors and time-effect factors (Bai, 2009; Bonhomme & Manresa, 2015; Athey et al., 2021). Interactive fixed-effect models are mainly applied as an extension to linear functional regression models, whereas our goal is to interpolate missing values utilizing multiple time series. Furthermore, their approaches account for the interaction between individual effects and time effects for each time series. In contrast, we map multivariate time series onto a shared latent space and directly estimate individualized dynamic latent factors without additional decomposition steps.

State-of-the-art deep learning methods are widely used for supervised and unsupervised learning for performing both interpolation and prediction tasks (Sun et al., 2020). For example, recurrent neural networks (Hochreiter & Schmidhuber, 1997) are powerful for sequential data. However, due to their complex architecture and the large number of parameters involved, recurrent neural networks require massive training data to guarantee good performance. Furthermore, recurrent neural networks require a homogeneous assumption for subjects, making them unsuitable for individualized predictions on trajectories, especially when the sample size of the training data is limited.

We propose an individualized dynamic latent factor model to integrate multivariate data from heterogeneous subjects. The proposed method incorporates irregular multi-resolution time series from each subject utilizing individual-wise dynamic latent factors, in addition to integrating population-wise information via shared latent factors across different subjects. Specifically, we estimate the dynamic latent factors through a nonparametric model such as B-spline approximation and establish the corresponding algorithm based on the alternating gradient descent (Tseng & Yun, 2009). In addition, we extend the dynamic latent factor model to a more general nonparametric framework beyond the B-spline approximation and establish consistency of the proposed interpolation model.

The proposed method has the following advantages. First, through mapping observed irregular time series to the unobserved latent space, the dynamic latent factor model allows us to effectively utilize the multi-resolution time series since the trajectories of correlated multiple time series information can be borrowed from each other through shared latent space. Consequently, the proposed interpolation for the missing data is more precise compared to interpolation from a single time series.

Second, our method integrates data, not only from multivariate time series, but also across multiple subjects. Through characterizing a population-wise association between dynamic latent factors and observed time series, the latent factors shared across subjects allow us to capture homogeneous features in addition to heterogeneous features. Thus, the proposed individualized dynamic latent factor model aggregates time series from all subjects to interpolate missing data, which can make a significant improvement in interpolation, especially when the resolution of a time series of interest is sparse.

Third, the proposed individualized dynamic latent factor model is applicable for time series with a complex trajectory. In particular, in contrast to stationary or Markov chain assumptions required by multivariate time series (Hamilton, 2020) and the dynamic linear model (Petris et al., 2009), we only require a smoothness assumption (Claeskens et al., 2009) if the B-spline approximation is implemented in the dynamic latent factor modelling. Therefore, the proposed method can model nonstationary processes or time series with abrupt changes, which is particularly useful in practice as abrupt changes in time series data can often occur.

2. The proposed method

2.1. *General methodology*

In this subsection, we propose an individualized dynamic latent factor model to capture the trajectory of multi-resolution time series while preserving time-invariant shared information across subjects for each time series.

We consider a J-dimensional multivariate time series for subject i, $i = 1, \ldots, I$:

$$Y_i(t) = \{Y_{i1}(t), Y_{i2}(t), \ldots, Y_{iJ}(t)\}^{\mathrm{T}}$$

with $Y_{ij}(t)$ the jth time series for $t \in [0, T]$, a finite interval. For each time series $Y_{ij}(t)$, there are K_{ij} observations at time-points in $\mathbb{T}_{ij} = \{t_{ijk} \mid k = 1, \ldots, K_{ij},\ t_{ijk} \in [0, T]\}$. We illustrate

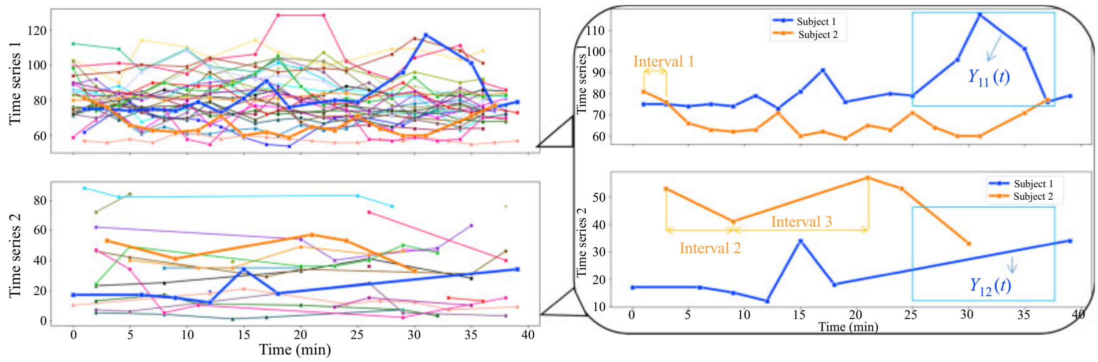

Fig. 2. Irregular multi-resolution time series with multi-resolution and irregular time intervals. The left plot provides two time series obtained from the smartwatch data where different colours represent different caregivers. The right plot shows two subjects from these two time series.

observations of the $J = 2$ time series in Fig. 2, where different colours represent different subjects.

The irregularity of multivariate time series imposes great challenges in that the number of observations K_{ij} could be different for different subjects and time series due to the multi-resolution nature. Specifically, for any pair of time series $Y_{ij}(t)$ and $Y_{ij'}(t)$, the time interval $t_{ijk} - t_{ij(k-1)}$ could be different from $t_{ij'k} - t_{ij'(k-1)}$, as illustrated in interval 1 and interval 2 of Fig. 2. Similarly, within a single time series $Y_{ij}(t)$, time intervals $t_{ijk} - t_{ij(k-1)}$ and $t_{ijk'} - t_{ij(k'-1)}$ could also be different, as shown in interval 2 and interval 3 of Fig. 2. Furthermore, the sets of time-points $\mathbb{T}_{ij}$ may also differ between subjects. That is, even if the resolution or sample rate for each time series j is the same for all subjects, the time-points t_{ijk} and $t_{i'jk}$ may not be the same.

One of our goals is to interpolate unsampled points for a low-resolution time series. Without loss of generality, let the Jth time series $Y_{iJ}(t)$ be the time series of interest. Because of the low resolution of the time series, values of the time series $Y_{iJ}(t)$ at some time-points $\{t \in \mathbb{T}_{ij} \mid j = 1, \ldots, J-1\}$ might not be observed. For illustration, on the right-hand side of Fig. 2, there are observations in the blue box for time series $Y_{11}(t)$, while time series $Y_{12}(t)$ is not observed. In addition, on the left-hand side of Fig. 2, we observe that time series with one particular lower resolution $Y_{i2}(t)$ have far fewer observed time-points than other time series, yet contain observed time-points the other series $Y_{i1}(t)$ do not have.

In the following, we propose to model each time series $Y_{ij}(t)$ for $i = 1, \ldots, I$, $j = 1, \ldots, J$, $t \in [0, T]$ by

$$Y_{ij}(t) = f_j^{\mathrm{T}} \theta_i(t) + \epsilon_{ij}(t), \tag{1}$$

where $f_j \in \mathbb{R}^R$ is a vector of population-wise latent factors corresponding to the jth time series, R is the dimension of the latent space, the dynamic latent factor $\theta_i(t) = \{\theta_{i1}(t), \ldots, \theta_{iR}(t)\}^{\mathrm{T}}$ is a vector of continuous functions of t for subject i capturing individual-specific features and the random noises $\epsilon_{ij}(t)$ are independent and identically distributed. We let $F = (f_1, \ldots, f_J)^{\mathrm{T}} \in \mathbb{R}^{J \times R}$ denote the latent factor matrix.

By modelling each time series through the inner product of latent factors f_j and $\theta_i(t)$ in (1), we are able to integrate data from multi-resolution time series and different subjects. In addition, mapping multivariate time series to a latent space via the dynamic latent factor $\theta_i(t)$ allows us to utilize information from multi-resolution time series. On the other hand, we require the latent factor F to be shared among subjects so information across subjects can be borrowed. The data integration can help us to improve the estimation accuracy of F and $\theta_i(t)$, and thus accurate interpolation of the target time series $Y_{iJ}(t)$ can be achieved. Specifically, we utilize the latent factor f_j to capture time-invariant features of the jth time series, while using the dynamic latent factor $\theta_i(t)$ for time-varying features.

In contrast to F, the dynamic latent factor $\theta_i(t)$ represents an individualized time trajectory that can be heterogeneous for different individuals. In most longitudinal data, such as our mobile health study, different time series could be correlated for the same subject; therefore, we can use a few latent factors to represent time-varying individual features. Thus, using the existing homogeneous models (Petris et al., 2009; Goodfellow et al., 2016), including the recurrent neural network model, may lead to high interpolation error, as these models are not suitable for learning individualized trajectories.

Additionally, the dynamic latent factor $\theta_i(t)$ for the ith subject allows for incorporation of the multi-resolution time series Y_{ij} for $j = 1, \ldots, J$, whereas traditional interpolation methods, such as polynomial interpolation and spline interpolation (De Boor, 1978), only use data from a single time series $Y_{iJ}(t)$. As we mentioned earlier, one of the interests in the time series study is to provide interpolation of low-resolution time series, which may be necessary due to budget or technical limitations in obtaining high-resolution data. Borrowing information from other time series of the same subject is robust for interpolation, especially in the case of low-resolution time series in multi-resolution data. This is because certain variations in the time series may not be captured by low-resolution observations. For example, in Fig. 2, the time series $Y_{i2}(t)$ is of interest. As highlighted in blue boxes, the observations of time series $Y_{11}(t)$ suggest that time series $Y_1(t)$ changes abruptly and manages to return back to the previous trend. However, the trajectory of time series $Y_{12}(t)$ might miss the abrupt change due to its low resolution, leading to high interpolation errors based only on observed data of $Y_{12}(t)$. In contrast to traditional interpolation methods, the proposed method is able to preserve the abrupt change through estimating $\theta_1(t)$ using additional time series $Y_{11}(t)$ and therefore provides more precise interpolation for $Y_{12}(t)$. Our method can be broadly applicable to time series with more complex trajectories, such as nonstationary processes or sparse time series. This is particularly useful when prior knowledge of time series is unknown, or when the data patterns indicate that stationary or non-stationary model assumptions (Petris et al., 2009; Hamilton, 2020) are not satisfied with the data. In practice, the stationary assumption could be too restrictive (Hamilton, 2020). The nonstationary random process assumption (Petris et al., 2009) could also be restrictive, as it typically requires that the state process be a Markov chain.

2.2. *Latent factor estimation*

In this subsection, we propose to estimate the dynamic latent factors $\theta_i(t)$ using B-spline functions to capture nonlinear function patterns. Specifically, we estimate the latent factors F and parameters associated with the B-spline by minimizing a regularized square loss on $Y_{ij}(t)$.

We assume that each dynamic latent factor element $\theta_{ir}(t)$ is a function in the Sobolev space $W_q^{\alpha}[0, T]$ equipped with a finite L_q norm, where α is a smooth parameter such that $\theta_{ir}(t)$ and its weak derivatives up to order α have a finite L_q norm. We approximate $\theta_{ir}(t)$ by a linear combination of B-spline basis functions of order $\alpha + 1$, that is,

$$\theta_{ir}(t) \approx \sum_{m=1}^{M} w_{irm} B_m(t) \qquad (r = 1, \ldots, R),$$

where B_m $(m = 1, 2, \ldots, M)$ are basis functions of smoothing degree α and $W = \{w_{irm}\} \in \mathbb{R}^{I\times R\times M}$ consists of weights for each basis function B_m. Specifically, on the time interval $[0, T]$, we use a sequence of a interior knots $0 < \kappa_1 < \kappa_2 < \cdots < \kappa_a < T$, and therefore the number of basis functions $M = a + \alpha + 1$.

In the context of time series or longitudinal data, the spline method is effective in modelling nonlinear trends over time (Welham, 2009) and is also flexible for modelling correlated longitudinal data (De Boor, 1978). To model irregular multi-resolution time series data with correlations among time-points and multiple time series within the same subject, we estimate the parameters F and W by minimizing the following square loss on $Y_{ij}(t)$ with an L_2 penalty (Salakhutdinov et al., 2007; Agarwal & Chen, 2009; Bi et al., 2017):

$$L(F, W) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t\in\mathbb{T}_{ij}}\{Y_{ij}(t) - f_j^{\mathrm{T}} W_i B(t)\}^2 + \lambda(\|F\|_{\mathrm{F}}^2 + \|W\|_{\mathrm{F}}^2). \tag{2}$$

Here, $W_i = \{w_{irm}\} \in \mathbb{R}^{R\times M}$ and $B(t) = \{B_m(t)\} \in \mathbb{R}^M$, λ is the tuning parameter and $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. We estimate F and W through

$$(\hat{F}, \hat{W}) = \arg\min_{F,W} L(F, W).$$

We use the Frobenius-norm penalty to control the nonsmoothness of the fitted curve by filtering out spurious coefficients of the latent factor matrix F and spline tensor W. The regularization of W enables us to use a relatively large number of interior knots without knowing the number of knots, while shrinking some spline coefficients towards zero. Allowing more interior knots leads to more flexibility in modelling the nonlinear trajectory (Claeskens et al., 2009).

We can also choose other approximation methods for dynamic latent factors $\theta_i(t)$, such as the kernel approach (Wenzel et al., 2021) or deep learning methods (Goodfellow et al., 2016). However, the interpolation accuracy is influenced by the approximation error of the dynamic latent factors, which is determined by the choice of approximation methods.

Once the estimators $\hat{F}$ and $\hat{W}$ are obtained, the proposed interpolation at any time $t \in [0, T]$ is calculated by

$$\hat{Y}_{ij}(t) = \hat{f}_j^{\mathrm{T}} \hat{W}_i B(t). \tag{3}$$

Equation (3) provides a general formula for all time series at any time-point in the range $[0, T]$. However, in practice, we might only be interested in interpolating a single time series $\hat{Y}_{iJ}(t)$ for $\{t \in \mathbb{T}_{ij} \mid j = 1, \ldots, J - 1\}$.

3. Theory

In this section, we develop the theoretical properties of the proposed method based on a sample estimator from a Sobolev space in addition to providing the theoretical properties of the estimation with the B-spline approximation. Specifically, we establish the asymptotic property of the integrated interpolation error and provide the rate of convergence for the proposed estimator when the parameter space is a Sobolev space. Finally, we provide a concrete convergence analysis for B-spline approximation to demonstrate the theoretical properties of our implemented model in § 2.

We first consider a general result for $\theta_{ir}(t) \in W_q^\alpha[0,T]$ for $\alpha > 1$ and $q \geqslant 2$, where $W_q^\alpha[0,T]$ is a Sobolev space with a finite L_q norm. The parameter α is a smooth parameter such that $\theta_{ir}(t)$ and its weak derivatives up to order α have a finite L_q norm. Additionally, we assume that $K_{ij} \sim K$ for some K, where $a \sim b$ when a and b have the same order.

Since our primary goal is the interpolation of $Y(t)$, we focus on the convergence property of the interpolation values instead of the latent factor recovery. Consider the time series

$$Y_{ij}(t) = \psi_{ij}(t) + \epsilon_{ij}(t),$$

where $\psi_{ij}(t) = \sum_{r=1}^{R} f_{jr}\theta_{ir}(t)$, $\Psi(t) = \{\psi_{ij}(t)\}$ is a vector of $\psi_{ij}(t)$ for $i = 1,\ldots,I$ and $j = 1,\ldots,J$, and the random noises $\epsilon_{ij}(t)$ are independently and identically distributed with mean 0 and variance σ^2. For simplicity, we write $\Psi(t)$ as Ψ in this paper. As $\theta_{ir}(t) \in W_q^\alpha[0,T]$, we have $\psi_{ij}(t) \in W_q^\alpha[0,T]$ by construction. For time series $Y_{ij}(t)$ of $Y(t)$, we define the L_2-loss function as $l\{\Psi, Y_{ij}(t)\} = \{Y_{ij}(t) - \psi_{ij}(t)\}^2$.

Let Ω be the set of observations, $|\Omega| = \sum_{ij} K_{ij}$ be the number of observations and $J(\Psi)$ be a nonnegative penalty function. For example, we have $J(\Psi) = \sum_{ij}\{\int_0^T |\psi_{ij}^{(\alpha)}(t)|^q \,\mathrm{d}t\}^{1/q}$ since $\psi_{ij}(t) \in W_q^\alpha[0,T]$. Then the overall object function is

$$L(\Psi \mid Y) = \frac{1}{|\Omega|} \sum_{(i,j,t)\in\Omega} l\{\Psi, y_{ij}(t)\} + \lambda_{|\Omega|} J(\Psi),$$

where $\lambda_{|\Omega|}$ is a tuning parameter for the penalization. To establish the convergence rate, we introduce the following assumption.

Assumption 1. For the empirical distribution $Q_{ij,n}$ of $t_{ij1},\ldots,t_{ijK_{ij}}$,

$$Q_{ij,n}(t) = \frac{1}{K_{ij}} \sum_{k=1}^{K_{ij}} \mathbf{1}_{t_{ijk}<t},$$

where $\mathbf{1}_A$ is the indicator function of event A, there exists a distribution function $Q_{ij}(t)$ with positive continuous density such that

$$\sup_{t\in[0,T]} |Q_{ij,n}(t) - Q_{ij}(t)| = o(K_{ij}^{-1}).$$

Assumption 1 assumes that the empirical distributions of observed time-points converge to a distribution of t with positive continuous density, which is typically the uniform distribution for random samples. Such an assumption is common in spline approximation. Similarly, in kernel approximation, the observed points are assumed to be asymptotically uniformly distributed (Wenzel et al., 2021). When the observed time-points t_{ijk} are random and sampled from the true distribution $Q_{ij}(t)$, Assumption 1 is satisfied naturally by the Glivenko–Cantelli theorem (Sharipov, 2011).

Let Ψ_0 be the true parameters and $\mathcal{S} = \{\Psi\colon \psi_{ij}(t) = \sum_r f_{jr}\theta_{ir}(t),\ \|F\|_\infty \leqslant c_0,\ \theta_{ir}(t) \in W_q^\alpha[0,T]$ for $i = 1,\ldots,I, j = 1,\ldots,J\}$ be the parameter space that depends on a positive constant c_0. We denote by $\hat{\Psi}_{|\Omega|}$ the sample estimator of Ψ_0, satisfying

$$L(\hat{\Psi}_{|\Omega|} \mid Y) \leqslant \inf_{\Psi\in\mathcal{S}} L(\Psi \mid Y) + \tau_{|\Omega|}, \tag{4}$$

where $\lim_{|\Omega|\to\infty} \tau_{|\Omega|} = 0$. This condition implies that $\hat{\Psi}_{|\Omega|}$ is close to the global minimizer of $L(\Psi \mid Y)$ when $|\Omega| \to \infty$. This necessity arises because, in practice, obtaining the exact global minimum is often impractical due to the nonconvex nature of function L.

In addition to the above assumptions, Assumptions S1 and S2 in the Supplementary Material are standard assumptions for nonparametric approximations on the basis functions, observation distribution and random noises.

We establish the convergence of $\hat{\Psi}_{|\Omega|}$ in the following theorem.

THEOREM 1. *Suppose that $\hat{\Psi}_{|\Omega|}$ is a sample estimator satisfying* (4). *Then, for $K_{ij} \sim K$ for some order K, and under Assumption* 1 *and Assumptions S1 and S2 in the Supplementary Material, we have*

$$\left(\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\int_0^T \{\hat{\psi}_{ij}(t) - \psi_{0,ij}(t)\}^2 \, \mathrm{d}Q_{ij}(t)\right)^{1/2} = O_p\{(IJK)^{-\alpha/(2\alpha+1)}\} + o_p(K^{-1/2}), \quad (5)$$

when $\tau_{|\Omega|} = o\{(IJK)^{-2\alpha/(2\alpha+1)}\}$ and $\lambda_{|\Omega|} \sim (IJK)^{-2\alpha/(2\alpha+1)}$.

Theorem 1 indicates that the average interpolation error in terms of integrated squared loss with respect to time converges to zero when K and IJ go to infinity. The first term of the error bound in (5) is due to the approximation bias of the dynamic latent factor that is determined by the smoothness of $\theta_{ir}(t)$. This approximation bias becomes negligible when the total number of observations goes to infinity. The second term of the error bound is determined by the difference between the empirical distribution and the reference distribution $Q_{ij}(t)$, which converges to zero when the number of observations of each time series K_{ij} goes to infinity, according to Assumption 1.

Theorem 1 demonstrates the benefits of integrating information across subjects and time series. The interpolation $\hat{\Psi}_{|\Omega|}$ converges faster to the true value Ψ_0 if the number of observed subjects I or the number of observed time series J is larger. That is, when we integrate more time series and more subjects, we obtain better interpolation.

Next, we extend Theorem 1 to the case when the B-spline approximation is applied, that is, the parameter space becomes $\mathcal{S}_M = \{\Phi\colon \phi_{ij}(t) = \sum_{r,m} f_{jr} w_{irm} B_m(t), \|F\|_\infty, \|W\|_\infty \leqslant c_0\}$, where $F = (f_{jr})$, $W = (w_{irm})$ and M is the number of basis functions. Similar to the previous notation, we define the L_2-loss function as $l\{\Phi, Y_{ij}(t)\} = \{Y_{ij}(t) - \phi_{ij}(t)\}^2$ and the overall object function as $L(\Psi \mid Y) = \sum_{(i,j,t)\in\Omega} l\{\Psi, y_{ij}(t)\}/|\Omega| + \lambda_{|\Omega|} J(\Psi)$. Additionally, we let the penalty be the L_2 penalty defined in § 2, that is, $J(\Phi) = \|F\|_2^2 + \|W\|_2^2$.

We also denote by $\hat{\Phi}_{|\Omega|} = \{\hat{\phi}_{ij}(t)\}$ the sample estimator of Ψ_0, satisfying

$$L(\hat{\Phi}_{|\Omega|} \mid Y) \leqslant \inf_{\Phi\in\mathcal{S}_M} L(\Phi \mid Y) + \tau_{|\Omega|}, \quad (6)$$

where $\lim_{|\Omega|\to\infty} \tau_{|\Omega|} = 0$.

We establish the asymptotic property of $\hat{\Phi}_{|\Omega|}$ in the following theorem.

THEOREM 2. *Let $\hat{\Phi}_{|\Omega|}$ be a sample estimator satisfying* (6). *Then, for $K_{ij} \sim K$ for some order K, and under Assumption* 1 *and Assumptions S1 and S2 in the Supplementary Material,* *we have*

$$\left(\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\int_0^T \{\hat{\phi}_{ij}(t) - \psi_{0,ij}(t)\}^2 \,\mathrm{d}Q_{ij}(t)\right)^{1/2} = O_p\{(IJK)^{-\alpha/(2\alpha+1)}\} + o_p(K^{-1/2}),$$

when $\tau_{|\Omega|} = o\{(IJK)^{-2\alpha/(2\alpha+1)}\}$ *and* $\lambda_{|\Omega|} \sim (IJK)^{-2\alpha/(2\alpha+1)}$.

Theorem 2 shows that the convergence rate of the proposed estimator in §2 is $O_p\{(IJK)^{-\alpha/(2\alpha+1)}\} + o_p(K^{-1/2})$. To obtain the convergence rate, we require that the penalty parameter λ shrink to zero at a rate of $(IJK)^{-2\alpha/(2\alpha+1)}$.

4. Computation

In this section, we introduce an algorithm and implementation details of the proposed method. Specifically, we utilize the alternating gradient descent algorithm (Tseng & Yun, 2009) to estimate latent factors F and W.

The alternating gradient descent algorithm provided in Algorithm 1 below is a generalization of the block coordinate gradient descent method (Tseng & Yun, 2009), which is especially useful in matrix decomposition and tensor decomposition (Zhao et al., 2015; Bi et al., 2018; Zhang et al., 2022). The main idea of the algorithm is to iteratively update each F and W_i for $i = 1, \ldots, I$, while keeping the others fixed. The advantage of this algorithm is that the latent factor matrices naturally provide a block structure of the parameters, and updating F and W_i enables us to transform the nonconvex optimization to a convex optimization. In addition, it can further decrease the number of iterations and lead to faster convergence compared with the gradient descent algorithm. This is because we can use a larger step size when updating blocks of parameters instead of entire parameters (Jain et al., 2013).

Specifically, let $F^{(s)}$ and $W_i^{(s)}$ denote the estimated F and W_i at the sth iteration, and let $L^{(s)} = L(F^{(s)}, W^{(s)})$ denote the corresponding loss. We update each $F^{(s-1)}$ and $W_i^{(s-1)}$ along the direction of the partial derivatives $\partial L(F, W^{(s-1)})/\partial F$ and $\partial L(F^{(s-1)}, W_i)/\partial W_i$ at each iteration.

Algorithm 1. Alternating gradient descent.

1. Initialization. Set the stopping error ϵ, rank R, tuning parameter λ, step size α, basis functions B_m ($m = 1, \ldots, M$), and initial values $F^{(0)}$ and $W^{(0)}$.
2. Latent factor update. At the sth iteration ($s \geqslant 1$)
   (i) update $F^{(s)}$: $F^{(s)} \leftarrow F^{(s-1)} - \alpha\partial L(F, W^{(s-1)})/\partial F$,
   (ii) update each $W_i^{(s)}$: $W_i^{(s)} \leftarrow W_i^{(s-1)} - \alpha\partial L(F^{(s-1)}, W_i)/\partial W_i$.
3. Stop if $|L^{(s+1)} - L^{(s)}|/L^{(s)} < \epsilon$.

To select rank R, tuning parameter λ and step size α, we conduct a grid search by minimizing the mean square error on the validation set. Our empirical study shows that the tuning parameter λ is quite robust and would not change the numerical performance much compared to rank R and step size α. Thus, to save computational cost, we tune the λ first, and conduct a grid search on pairs of rank R and step size α after. The results from cross-validation simulations, as well as the performance of the proposed method across different rank values, can be found in the Supplementary Material.

In addition, selection of basis functions and determining the number of knots are important here. Based on the practice of B-spline approximation, we require the number of basis functions M to be large enough so that there is at least one observation in each interval. However, in practice, even if this assumption is mildly violated, we can still obtain a reasonable interpolation accuracy due to the penalty term. In our numerical study, we utilize the evenly spaced a knots that are smaller or equal to the smallest K_{ij} for $i = 1, \ldots, I$ and $j = 1, \ldots, J$. This allows us to model the trajectory sufficiently well by utilizing a relatively large number of knots.

The L_2 penalty in (2) is selected to avoid overfitting and scale indeterminacy, which balances computational complexity and model complexity (Acar et al., 2011). We can also consider penalty functions used in penalized spline functions, e.g., the integrated squared qth-order derivative used in the spline function (Claeskens et al., 2009) or the total variation penalty used by Jhong et al. (2017). However, based on our simulation studies, utilizing these penalty terms results in similar interpolation accuracy as the L_2 penalty after proper hyperparameter tuning.

5. Simulations

5.1. *General setting*

In this section, we conduct simulations to investigate the empirical performance of the proposed individualized dynamic latent factor model and compare it with existing methods under six different settings. Specifically, we compare the proposed method with six competing methods, namely, the smoothing spline (De Boor, 1978), multivariate time series (Hamilton, 2020), the dynamic linear model (Petris et al., 2009), functional principal component analysis (Wang et al., 2016), the recurrent neural network and the deep recurrent neural network (Hochreiter & Schmidhuber, 1997).

Here, the smoothing spline applies to the Jth time series for each subject separately with smoothing degree $k = 3$, which is implemented in the Python package `scipy.interpolate` (De Boor, 1978), where knots are selected by the function `UnivariateSpline` automatically. The multivariate time series and dynamic linear model implement multivariate time series for each subject. To deal with irregular time intervals, unobserved points are treated as missing values. The multivariate time series and dynamic linear model are implemented in the Python packages `statsmodels.tsa` (Hamilton, 2020) and `pyro` (Petris et al., 2009). The functional principal component analysis integrates the Jth time series for all subjects together. To handle irregular time intervals, we apply the functional principal component analysis through the B-spline functional basis using the Python package `scikit-fda` (Wang et al., 2016). We implemented two recurrent neural network models using the Python package `tensorflow` (Hochreiter & Schmidhuber, 1997) using masking layers to handle irregularity, where the recurrent neural network model contains one recurrent neural network layer with 32 units and one dense layer, and the deep recurrent neural network model contains the same three recurrent neural network layers with 32 units and one dense layer. Both recurrent neural network models are trained by the Adam optimizer (Kingma & Ba, 2017) with 20 epochs. For the individualized dynamic latent factor model, we utilize evenly spaced internal knots with the number of basis functions $M = 300$ and the smoothing degree $p = 3$.

We generate simulated data according to (1) with $f_j \sim N(0, I_R)$ and $\epsilon_{ij}(t) \sim N(0, 0.5^2)$. In each setting, we let the latent space of dimension, the number of subjects, the number of time series and the time range be $R = 3$, $I = 30$, $J = 5$ and $T = 1000$, respectively. In each setting, we only generate time-points $t = 1, \ldots, T$ as most of the competing methods cannot handle continuous time-points $t \in [0, T]$, while the proposed method is able to handle them.

We assess interpolation performance by examining the mean square error over the training and testing sets based on 50 replications. Specifically, we are interested in the interpolation of the Jth time series. Thus, we calculate the mean square error over the training and testing sets of the Jth time series, that is,

$$\frac{\sum_{i=1}^{I}\sum_{t\in\mathbb{T}_{iJ}}\{Y_{iJ}(t)-\hat{Y}_{iJ}(t)\}^2}{\sum_{i=1}^{I}|\mathbb{T}_{iJ}|} \quad \text{and} \quad \frac{\sum_{i=1}^{I}\sum_{t\notin\mathbb{T}_{iJ}}\{Y_{iJ}(t)-\hat{Y}_{iJ}(t)\}^2}{IT-\sum_{i=1}^{I}|\mathbb{T}_{iJ}|},$$

where $|\mathbb{T}_{iJ}|$ denotes the number of observations for the time series Y_{iJ}.

5.2. *Multi-resolution time series*

In this subsection, we investigate interpolation performance for multi-resolution time series under three settings with observed time-points at $\mathbb{T}_{ij}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$.

In each setting, we let the dynamic latent factors be

$$\theta_{i1}(t) = 2\exp\left\{-\frac{(t-60-10i)^2}{50}\right\} + 4\exp\left\{-\frac{(t-70-10i)^2}{20}\right\},$$
$$\theta_{i2}(t) = i \times 0.02\log(t+1),$$
$$\theta_{i3}(t) = \cos(0.12\pi t + 1),$$

where θ_{i1} represents two pulses at time-points $60 + 10i$ and $70 + 10i$ for $i = 1, \ldots, I$; θ_{i2} represents a time trend that varies among subjects and θ_{i3} represents a seasonal trend.

To evaluate the performance of the proposed method under different observation processes, we consider three settings, where the numbers of observations are similar. Specifically, the three settings of the observation points of training sets are as follows.

Setting 1.1. We let

$$Y_{i1}, \ldots, Y_{i4}: \mathbb{P}(t \in \mathbb{T}_{ij}) = 0.8 \text{ for } t = 1, \ldots, 1000; j = 1, \ldots, 4,$$
$$Y_{i5}: \mathbb{P}(t \in \mathbb{T}_{i5}) = 0.2 \text{ for } t = 1, \ldots, 1000.$$

Setting 1.2. We let

$$Y_{i1}, \ldots, Y_{i3}: \mathbb{T}_{ij} = \{1, 2, 3, \ldots, 1000\} \text{ for } j = 1, 2, 3,$$
$$Y_{i4}: \mathbb{T}_{i4} = \{1, 3, 5, 7, \ldots, 999\},$$
$$Y_{i5}: \mathbb{T}_{i5} = \{1, 5, 9, 13, \ldots, 997\}.$$

Setting 1.3. We let

$$Y_{i1}, \ldots, Y_{i5}: \mathbb{P}(t \in \mathbb{T}_{ij}) = 0.7 \text{ for } t = 1, \ldots, 1000; j = 1, \ldots, 5.$$

Setting 1.1 mimics the most complicated situation of multi-resolution data, where the $\mathbb{T}_{ij}$ are different for different subjects and each time series has an irregular time interval. Setting

Table 1. *The mean square error of the proposed method and competing methods under settings 1.1–1.3. The standard errors are reported in parentheses*

MSE	Setting 1.1		Setting 1.2		Setting 1.3	
	Training	Testing	Training	Testing	Training	Testing
IDLFM (proposed)	0.209 (0.041)	0.415 (0.161)	0.212 (0.044)	0.341 (0.067)	0.215 (0.023)	0.327 (0.041)
SS	0.602 (0.257)	2.983 (10.453)	0.556 (0.282)	0.598 (0.346)	0.632 (0.280)	0.749 (0.460)
RNN	4.186 (6.180)	4.399 (6.226)	0.350 (0.565)	6.786 (6.840)	0.894 (0.894)	1.259 (1.078)
DRNN	1.712 (2.504)	5.793 (7.878)	0.099 (0.050)	6.610 (6.843)	0.604 (0.552)	3.915 (4.865)

IDLFM, the proposed individualized dynamic latent factor model; SS, smoothing spline; RNN, recurrent neural network; DRNN, deep recurrent neural network; MSE, mean square error.

1.2 considers the situation where multi-resolution time series have evenly spaced and fixed time-points. Setting 1.3 considers the same resolution time series with varying time-points $\mathbb{T}_{ij}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$. For testing sets of three settings, we use unobserved time-points.

Table 1 provides the mean square error (MSE) results under settings 1.1–1.3. The proposed method has the best performance on the testing set under all three settings, with more than 40% improvement in the mean square error compared to other methods. The competing methods such as the multivariate time series, dynamic linear model and functional principal component analysis cannot be applied to time series with multi-resolution or different observation time-points for each subject. Thus, only the methods of the smoothing spline, recurrent neural network and deep recurrent neural network are compared. As one of the most popular interpolation methods, the smoothing spline performs the second best in most settings, except it performs the worst under setting 1.1 with randomly selected time-points for low-resolution time series. In contrast, the proposed method is able to borrow information from other time series from the same subject, especially that with high resolution, and therefore attains better interpolation accuracy. For the recurrent neural network and deep recurrent neural network, the performance varies under different settings. In general, the deep recurrent neural network performs better than the recurrent neural network on the training set. However, the deep recurrent neural network performs poorly in interpolation under multi-resolution situations. Overall, the two recurrent neural network models perform the worst as they do not incorporate heterogeneity among subjects. The proposed individualized dynamic latent factor model performs similarly under the two multi-resolution settings, settings 1.1 and 1.2. The difference is that, for randomly generated observation time-points, the standard deviation of the mean square error is higher on the testing set.

We also investigate the setting with more time series, where $J = 101$, for each subject, where there are 100 time series as covariates and one time series of interest. Our numerical study shows that the proposed method attains a lower standard deviation when J is higher as more time series are integrated. The mean square error does not improve much because the convergence rate of the proposed method is related to the number of observation points K_{ij}. Additional simulations also illustrate the robustness of the proposed method under various missing mechanisms, such as missing completely at random, missing at random and missing not at random. The detailed simulation results are provided in Tables S1 and S2 in the Supplementary Material.

5.3. *Multiple time series with heterogeneity and nonstationarity*

In this subsection, we focus on time series with the same resolution where all subjects and all time series are observed at the same time-points $\mathbb{T}_{ij} = \mathbb{T}_{11}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$. We investigate how the heterogeneity of subjects and nonstationarity of time series affect the performance of interpolation in the following three settings. They are the same as the settings in § 5.2, except for $\mathbb{T}_{ij}$ and the dynamic latent factors. The sets $\mathbb{T}_{ij}$ are the same within each setting, which we refer to as $\mathbb{T}$. For set $\mathbb{T}$, the observation index $t \in \{1, \ldots, T\}$ is selected according to the Bernoulli distribution with a probability 0.7, and the unobserved points are treated as the testing set. This mimics the time series setting where all subjects and time series are observed at the same time-points, but at unevenly spaced time intervals. The dynamic latent factors for each setting are generated as follows, for $i = 1, 2, \ldots, 30$ and $t = 1, 2, \ldots, 1000$.

Setting 2.1. We let

$$\theta_{i1}(t) = 2\exp\left\{-\frac{(t-60)^2}{50}\right\} + 4\exp\left\{-\frac{(t-70)^2}{20}\right\},$$
$$\theta_{i2}(t) = 0.2\log(t+1),$$
$$\theta_{i3}(t) = \cos(0.12\pi t + 1).$$

Setting 2.2. We let

$$\theta_{i1}(t) = 2\exp\left\{-\frac{(t-60-10i)^2}{50}\right\} + 4\exp\left\{-\frac{(t-70-10i)^2}{20}\right\},$$
$$\theta_{i2}(t) = 0.2\log(t+1),$$
$$\theta_{i3}(t) = \cos(0.12\pi t + 1).$$

Setting 2.3. We let

$$\theta_{i1}(t) = 2\exp\left\{-\frac{(t-60-10i)^2}{50}\right\} + 4\exp\left\{-\frac{(t-70-10i)^2}{20}\right\},$$
$$\theta_{i2}(t) = i \times 0.02\log(t+1),$$
$$\theta_{i3}(t) = \cos(0.12\pi t + 1).$$

In setting 2.1, the dynamic latent factors are the same for all subjects. In setting 2.2, we change the locations of two pulses in θ_{i1} to be different for different subjects and keep θ_{i2} and θ_{i3} the same as in setting 2.1. Additionally, in setting 2.3, we keep θ_{i1} and θ_{i3} the same as in setting 2.2 and change θ_{i2} to a function where the time trend varies across subjects, where the stationary assumption is highly violated for subjects with large i.

Table 2 provides the results of all methods under settings 2.1–2.3. We observe that the proposed method has the best performance under all three settings, with more than 50% improvement in the mean square error compared to other methods. All methods except the smoothing spline perform best in setting 2.1 and perform worst in setting 2.3. This is because the subjects are homogeneous in setting 2.1, and are highly heterogeneous in setting 2.3. The proposed method is the most robust compared to other competing methods due to the advantage of integrating multiple time series and information across subjects. In setting

Table 2. *The mean square error of the proposed method and competing methods under settings 2.1–2.3. The standard errors are reported in parentheses*

MSE	Setting 2.1		Setting 2.2		Setting 2.3	
	Training	Testing	Training	Testing	Training	Testing
IDLFM (proposed)	0.201 (0.020)	0.359 (0.046)	0.208 (0.026)	0.334 (0.049)	0.213 (0.024)	0.350 (0.046)
SS	0.726 (0.275)	0.876 (0.501)	0.564 (0.278)	0.668 (0.466)	0.608 (0.253)	0.720 (0.457)
MTS	0.652 (0.384)	0.761 (0.488)	0.607 (0.414)	0.722 (0.523)	0.773 (0.919)	0.919 (1.142)
DLM	3.410 (3.575)	3.405 (3.629)	3.421 (3.821)	3.536 (4.036)	4.840 (8.506)	4.838 (8.340)
FPCA	4.429 (4.415)	4.501 (4.453)	4.516 (4.885)	4.546 (4.901)	6.665 (6.128)	6.762 (6.127)
RNN	0.869 (0.550)	0.865 (0.569)	0.805 (0.665)	0.818 (0.704)	1.584 (3.416)	1.568 (3.324)
DRNN	0.783 (0.515)	0.784 (0.531)	0.744 (0.581)	0.756 (0.604)	1.350 (2.646)	1.333 (2.589)

IDLFM, the proposed individualized dynamic latent factor model; SS, smoothing spline; MTS, multivariate time series; DLM, dynamic linear model; FPCA, functional principal component analysis; RNN, recurrent neural network; DRNN, deep recurrent neural network; MSE, mean square error.

2.3, the multivariate time series and dynamic linear model perform worse than the other two settings, settings 2.1 and 2.2, since the stationary or Markov chain assumption is violated when the nonlinear trend dominates the time series. In contrast, the proposed method can model nonlinear trends with dynamic latent factors and provides better interpolation. In addition, the functional principal component analysis and two neural network models fail to deliver accurate interpolation due to heterogeneity for different individuals. The individualized dynamic latent factor model is capable of integrating time series information from heterogeneous subjects.

6. Real data application

In this section, we apply the proposed method to smartwatch data for caregivers of dementia patients from the Dementia Family Caregiver Study, which was carried out by Lee & Gibbs (2021) in California.

Studies show that the under-served caregivers of dementia patients frequently experience physical and emotional distress (Anthony-Bergstone et al., 1988). To better monitor and manage caregivers' stress, Lee & Gibbs (2021) conducted home-based, culturally and language-specific interventions for dementia family caregivers, which included stress self-management and caregiving education. Community health workers offered stress reduction techniques, including mindful deep breathing and compassionate listening during up to four home visits with the caregiver. Additionally, community health workers also provided caregiving education to improve caregivers' health, well-being and positive interactions with dementia patients during a one-month study period. The intervention employed is not in an online form, as it is preplanned and remains unaffected by the collected data. To monitor the caregiver's physical activities and physiology, caregivers are equipped with wearable internet-of-things devices, such as the smartwatch, for a month. These wearable devices are used to measure the intervention's impact, offering noticeable changes due to the intervention. There are five time series of measurements from the smartwatch data, which are steps, heart rate, activity level, movement and stress. Among these five measurements, we are interested in evaluating and predicting the stress level of caregivers, which are highly associated with the other four time series variables (Saykrs, 1973). Our goal is to interpolate unobserved

Table 3. *The mean square error of the proposed method and competing methods for smartwatch data. Standard errors are reported in parentheses*

MSE	IDLFM (proposed)	SS	RNN	DRNN
Training	0.149 (0.011)	0.149 (0.013)	0.701 (0.296)	0.313 (0.158)
Testing	0.275 (0.009)	0.367 (0.198)	0.713 (0.144)	0.992 (0.666)

IDLFM, the proposed individualized dynamic latent factor model; SS, smoothing spline; RNN, recurrent neural network; DRNN, deep recurrent neural network; MSE, mean square error.

stress levels, so we can assess the effects of the intervention more accurately through the investigation. These measurements are observed with multi-resolution. Specifically, activity level and movement are observed every minute; heart rate is collected every two minutes; stress is measured every three minutes and steps are counted every 15 minutes. Additionally, due to the heterogeneous nature of individuals, the observed time also varies for different caregivers, leading to irregular time series.

In our study, we remove the first four days and last 10 days of data to ensure data quality, as subjects might not be familiar with the devices and have trouble setting up, or stop wearing the devices at the end of the study. There are 33 caregivers and more than 23 000 time-points from day 5 to day 20. The average missing rate for the stress variable is 80%, while missing rates for steps, heart rate, activity level and movement are 97%, 64%, 37% and 18%, respectively. The varying missing rates among different time series primarily result from differences in the collection frequency of each variable. Additionally, the measurement techniques for each time series and caregivers' charging habits contribute to the missingness. Since the scales of different time series measurements are different, we standardize each observed value $Y_{ij}(t)$ by subtracting the mean and dividing by the standard deviation of the corresponding time series Y_{ij}.

To train the interpolation models and evaluate the performance of the proposed method, we randomly select 30% of the observations for stress levels as a testing set and assign the remaining stress observations and observations of the other four time series measurements as a training set. The interpolation model predictions are evaluated by the mean square error of stress predictions on the training and testing sets based on 50 replications.

We compare the proposed individualized dynamic latent factor model method with the three competing methods, including the smoothing spline, recurrent neural network and deep recurrent neural network, as described in § 5. Table 3 provides the average mean square error based on both training and testing sets. The proposed individualized dynamic latent factor model method significantly outperforms the other methods in terms of achieving the lowest mean square error. Specifically, the proposed method reduces 24%, 61% and 72% of the mean square errors of the smoothing spline, recurrent neural network and deep recurrent neural network methods on testing sets, respectively, and achieves the smallest standard error among all competing methods. This is because the smoothing spline does not fully utilize the other four time series observations. The recurrent neural network using one layer neural network cannot accommodate complex time series data. Although the three-layer deep recurrent neural network model performs better than the recurrent neural network model, it still produces much larger mean square errors than the proposed individualized dynamic latent factor model, as the deep recurrent neural network is not capable of handling low sample-size heterogeneous data.

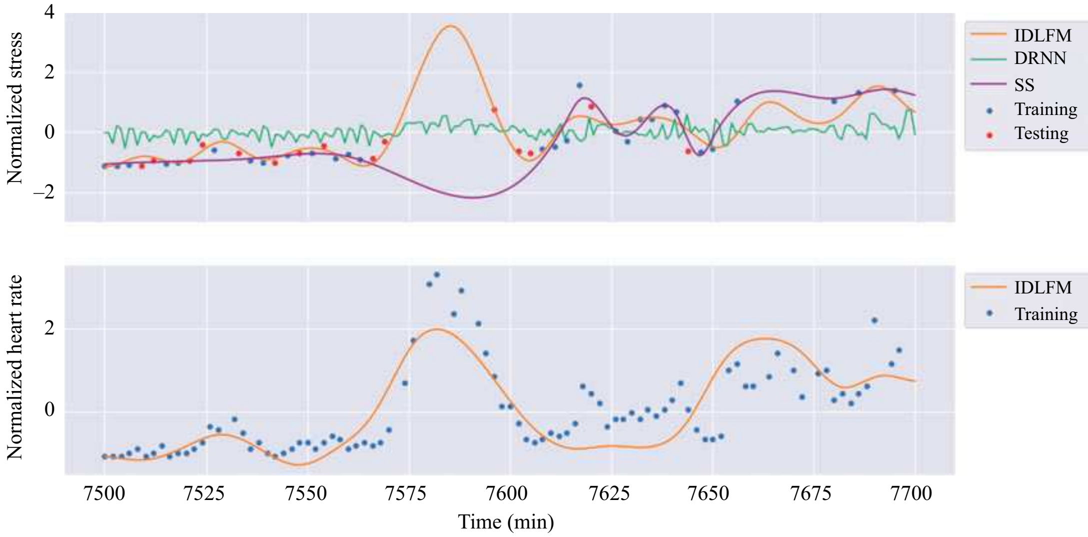

Fig. 3. The top plot provides estimated normalized stress for one caregiver by the proposed individualized dynamic latent factor model (orange line) and four competing methods for both training and testing data. The bottom plot provides a fitted normalized heart rate for the same caregiver by the proposed method. Observations of the caregiver in the training set are marked as blue dots, while those in the testing set are marked as red dots.

Figure 3 illustrates that the proposed individualized dynamic latent factor model interpolates the stress of caregivers better than competing methods. In particular, estimated normalized stress for one caregiver by the proposed individualized dynamic latent factor model and competing methods along with the fitted normalized heart rate for the same caregiver by the proposed method are provided using the training and testing data. The outstanding performance of interpolation power by the proposed method is due to the integration of information across multiple time series measurements with multi-resolution. For example, the proposed method can identify the high stress in time interval [7560, 7610] based on borrowing information from the increased heart rate during the same time period, while the competing methods, such as the deep recurrent neural network, fail to predict such a trend. In practice, monitoring the stress level and identifying high-stress moments is the first step in managing the caregiver's stress level. The proposed method can provide more precise interpolation for the unobserved stress levels, and is thus able to provide effective interventions for caregivers and lessen stress levels during dementia caregiving.

7. Discussion

We point out two future research directions. The first one is to establish a general algorithm for different approximation methods, since the proposed algorithm is based on the *B*-spline approximation that may not be applicable for kernel methods. Secondly, in clinical applications, the ultimate goal is to carry out medical tasks such as outcome prediction and patient subtyping. Interpolating unsampled values and then processing downstream tasks may lead to suboptimal analyses and predictions (Wells et al., 2013). Thus, a potential direction is to extend the proposed interpolation model in prediction, clustering and classification problems and process the downstream tasks directly based on modelling the time series with missing data.

Acknowledgement

This study was supported by the Arthur N. Rupe Foundation and the University of California, Irvine Sue & Bill Gross School of Nursing, and partially supported by the National Institute of Health (NIH)/National Institute on Aging (NIA) and National Science Foundation (NSF). The information or content and conclusions contained herein are those of the authors and should not be construed as the official position or policy of, nor should any endorsements be inferred by, the NIH, NIA, NSF or the U.S. Government. The authors would like to thank the editor, the associate editor and the referees for providing constructive comments and suggestions that improved the quality of the paper.

Supplementary material

The Supplementary Material contains further simulations, additional assumptions and proofs of the theorems.

References

Acar, E., Dunlavy, D. M., Kolda, T. G. & Mørup, M. (2011). Scalable tensor factorizations for incomplete data. *Chemom. Intell. Lab. Syst.* **106**, 41–56.

Agarwal, D. & Chen, B.-C. (2009). Regression-based latent factor models. In *Proc. 15th ACM SIGKDD Int. Conf. Know. Disc. Data Mining*, pp. 19–28. New York: Association for Computing Machinery.

Anthony-Bergstone, C. R., Zarit, S. H. & Gatz, M. (1988). Symptoms of psychological distress among caregivers of dementia patients. *Psychol. Aging* **3**, 245–8.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G. & Khosravi, K. (2021). Matrix completion methods for causal panel data models. *J. Am. Statist. Assoc.* **116**, 1716–30.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* **77**, 1229–79.

Berrendero, J. R., Justel, A. & Svarc, M. (2011). Principal components for multivariate functional data. *Comp. Statist. Data Anal.* **55**, 2619–34.

Bi, X., Qu, A. & Shen, X. (2018). Multilayer tensor factorization with applications to recommender systems. *Ann. Statist.* **46**, 3308–33.

Bi, X., Qu, A., Wang, J. & Shen, X. (2017). A group-specific recommender system. *J. Am. Statist. Assoc.* **112**, 1344–53.

Bonhomme, S. & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica* **83**, 1147–84.

Chiou, J.-M. & Müller, H.-G. (2014). Linear manifold modelling of multivariate functional data. *J. R. Statist. Soc.* B **76**, 605–26.

Chiou, J.-M. & Müller, H.-G. (2016). A pairwise interaction model for multivariate functional and longitudinal data. *Biometrika* **103**, 377–96.

Claeskens, G., Krivobokova, T. & Opsomer, J. D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika* **96**, 529–44.

Conway, M., Berg, R. L., Carrell, D., Denny, J. C., Kho, A. N., Kullo, I. J., Linneman, J. G., Pacheco, J. A., Peissig, P., Rasmussen, L. et al. (2011). Analyzing the heterogeneity and complexity of electronic health record oriented phenotyping algorithms. In *AMIA Ann. Symp. Proc. 2011*, pp. 274–83. Washington, DC: American Medical Informatics Association.

De Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer.

Gómez, V., Maravall, A. & Pena, D. (1999). Missing observations in ARIMA models: skipping approach versus additive outlier approach. *J. Economet.* **88**, 341–63.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press.

Hall, P. & Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *J. R. Statist. Soc.* B **68**, 109–26.

Hamilton, J. D. (2020). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Happ, C. & Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J. Am. Statist. Assoc.* **113**, 649–59.

Hochreiter, S. & Schmidhuber, J. (1997). Long short-term memory. *Neural Comp.* **9**, 1735–80.

Jacques, J. & Preda, C. (2014). Model-based clustering for multivariate functional data. *Comp. Statist. Data Anal.* **71**, 92–106.

Jain, P., Netrapalli, P. & Sanghavi, S. (2013). Low-rank matrix completion using alternating minimization. In *Proc. 45th ACM Symp. Theory of Computing*, pp. 665–74. New York: Association for Computing Machinery.
James, G. M. (2002). Generalized linear models with functional predictors. *J. R. Statist. Soc.* B **64**, 411–32.
James, G. M. & Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *J. R. Statist. Soc.* B **63**, 533–50.
James, G. M., Hastie, T. J. & Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87**, 587–602.
Jensen, P. B., Jensen, L. J. & Brunak, S. (2012). Mining electronic health records: towards better research applications and clinical care. *Nature Rev. Genet.* **13**, 395–405.
Jhong, J.-H., Koo, J.-Y. & Lee, S.-W. (2017). Penalized B-spline estimator for regression functions using total variation penalty. *J. Statist. Plan. Infer.* **184**, 77–93.
Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics* **22**, 389–95.
Kingma, D. P. & Ba, J. (2017). Adam: a method for stochastic optimization. *arXiv:* 1412.6980v9.
Lee, J.-A. & Gibbs, L. (2021). Effect of a home-visit intervention on emotional well-being for ethnic minority dementia family caregivers: use of mindful breathing and compassionate listening. *Alzheimer's Dement.* **17**, e053221.
Petris, G., Petrone, S. & Campagnoli, P. (2009). Dynamic linear models. In *Dynamic Linear Models with R*, pp. 31–84. New York: Springer.
Salakhutdinov, R., Mnih, A. & Hinton, G. (2007). Restricted Boltzmann machines for collaborative filtering. In *Proc. 24th Int. Conf. Mach. Learn.*, pp. 791–8. New York: Association for Computing Machinery.
Saykrs, B. M. (1973). Analysis of heart rate variability. *Ergonomics* **16**, 17–32.
Sharipov, O. S. (2011). Glivenko-Cantelli theorems. In *International Encyclopedia of Statistical Science*, Ed. M. Lovric, pp. 612–4. Berlin: Springer.
Sun, C., Hong, S., Song, M. & Li, H. (2020). A review of deep learning methods for irregularly sampled medical time series data. *arXiv:* 2010.12493v2.
Tseng, P. & Yun, S. (2009). A coordinate gradient descent method for nonsmooth separable minimization. *Math. Program* **117**, 387–423.
Volkmann, A., Stöcker, A., Scheipl, F. & Greven, S. (2023). Multivariate functional additive mixed models. *Statist. Mod.* **23**, 303–26.
Wang, J.-L., Chiou, J.-M. & Müller, H.-G. (2016). Functional data analysis. *Ann. Rev. Statist. Appl.* **3**, 257–95.
Welham, S. (2009). Smoothing spline models for longitudinal data. In *Longitudinal Data Analysis*, Ed. G. Fitzmaurice, M. Davidian, G. Verbeke and G. Molenberghs, pp. 253–89. New York: Chapman and Hall/CRC.
Wells, B. J., Chagin, K. M., Nowacki, A. S. & Kattan, M. W. (2013). Strategies for handling missing data in electronic health record derived data. *EGEMS* **1**, 1035.
Wenzel, T., Santin, G. & Haasdonk, B. (2021). A novel class of stabilized greedy kernel approximation algorithms: convergence, stability and uniform point distribution. *J. Approx. Theory* **262**, 105508.
Yao, F. & Lee, T. C. (2006). Penalized spline models for functional principal component analysis. *J. R. Statist. Soc.* B **68**, 3–25.
Yao, F., Müller, H.-G. & Wang, J.-L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33**, 2873–903.
Zhang, J., Yuan, Y. & Qu, A. (2022). Tensor factorization recommender systems with dependency. *Electron. J. Statist.* **16**, 2175–205.
Zhao, T., Wang, Z. & Liu, H. (2015). A nonconvex optimization framework for low rank matrix estimation. In *Proc. 28th Int. Conf. Neural Info. Proces. Syst.*, pp. 559–67. Cambridge, MA: MIT Press.

[*Received on* 24 *January* 2023. *Editorial decision on* 27 *February* 2024]