Structural Degradation Modeling Framework for Sparse Data Sets With an Application on Alzheimer's Disease

Abdallah Chehade and Kaibo Liu, *Member, IEEE*

***Abstract*— The rapid development of information technologies provided unprecedented big data environments for condition monitoring and degradation analyses. However, the available big data sets are often sparse with a limited number of observations per recorded unit. For example, in many healthcare systems, data are collected from a large number of patients, but the available observations from each patient are quite limited. Unfortunately, most of the existing approaches for data-driven degradation modeling may not work well in this scenario as they either pool the information from the population or require rich historical observations in each unit. To address the challenges in "sparse data environments," this paper proposes a structural degradation modeling framework (SDM). The SDM is inspired by the recommender system, which provides recommendations about specific items for the user. In addition, it is also tailored to the needs of degradation modeling. In particular, the framework takes into consideration: 1) the available data from the unit of interest; 2) the population characteristics; 3) the relationship between the available units; and 4) the precision of the available units. Simulation studies and a case study that involves the Alzheimer's disease (AD) neuroimaging initiative data set are conducted, which shows satisfactory performance of the proposed method.**

***Note to Practitioners*—This paper proposes a framework for modeling and predicting the degradation level and/or condition of units with time. Our framework is particularly useful where many units have missing and/or limited degradation observations. Essentially, our proposed method integrates two important ideas: 1) leveraging the available data from the unit of interest to improve the modeling fitting of the individual unit over the observed time domain and 2) considering the relationship between the available units to extract proper and accurate population characteristics to address the challenge of limited observations. The proposed approach is validated via simulation studies as well as a healthcare case study based on AD. In the future research, we will further explore the extension of the proposed method such as considering more generic degradation models and optimal parameter tunings.**

Manuscript received February 9, 2018; revised April 19, 2018; accepted April 20, 2018. Date of publication May 18, 2018; date of current version January 4, 2019. This paper was recommended for publication by Associate Editor X. Xie and Editor J. Li upon evaluation of the reviewers' comments. This work was supported in part by the National Science Foundation under Grant CMMI-1435809 and in part by the Office of Naval Research under Grant N00014-17-1-2261. *(Corresponding author: Abdallah Chehade.)*

A. Chehade is with the Department of Industrial and Manufacturing Systems Engineering, University of Michigan–Dearborn, Dearborn, MI 48128 USA (e-mail: achehade@umich.edu).

K. Liu is with the Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI 53706 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2018.2829770

***Index Terms*— Collaborative model, condition monitoring, dissimilarity, healthcare, predictive analytics, prognostics, recommender system.**

NOMENCLATURE

$s_{j,t}$	Degradation status for unit/recommender j at time t.
w_{ij}	Weight of recommender i in construction of the model of unit j.
b_j	Bias term for unit j.
$\epsilon_{j,t}$	Random noise for unit j at time t.
m	Number of available recommenders.
$t_i(r)$	Time at which the rth observation is obtained for unit/recommender i.
$\eta(\boldsymbol{\theta}, t)$	Functional form of the degradation model.
$\boldsymbol{\theta}_i$	Parameters of the degradation model for recommender i.
$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$	Covariance matrix of the recommenders' parameters.
d_{ik}	Dissimilarity between recommenders i and k.
d_{ii}^*	Sum of dissimilarities between recommender i and the remaining recommenders.
z_{ik}	Indicator if recommenders i and k are identical, with 1 if they are identical and 0 otherwise.
$h(.)$	Tuning function.
λ	Tuning parameter.
n_j	Number of available observations for unit/recommender j.

I. INTRODUCTION

ACCURATE modeling and prediction of the future degradation evolution has been a critically important task in many applications. For example, in manufacturing equipment, an unexpected failure may lead to significant economic losses, production downtime, customer dissatisfaction, and safety issues. Also in healthcare systems, an unexpected disease onset may lead to severe medical complexities, ineffective treatment planning, and long-term side effects. Therefore, it is crucial to accurately monitor the health status of a unit (e.g., a system, equipment, and a patient) and understand its degradation process. To achieve this goal, condition-based

techniques have been rapidly developed, which aim to fully understand the degradation mechanism of each individual unit so that optimal intervention decisions can be made [1]–[5]. For example, in healthcare systems, appropriate implementation of condition-based techniques can significantly improve early disease diagnosis, treatment effect monitoring and evaluation, and reduce medical and economic costs [6]. Similarly, in manufacturing applications, effectively employing a condition-based strategy can extend the lifespan of a unit, lower the maintenance cost, enhance equipment safety, and improve operator experience [7]–[9].

Fortunately, the accessibility and development of information technologies (e.g., clouds) have facilitated the collection and storage of information from a massive number of units nowadays, which provide a great opportunity to better understand the degradation processes. For example, the Alzheimer's disease neuroimaging initiative (ADNI) has been collecting longitudinal measurements of biomarkers from hundreds of participating patients. While the data environment is *rich* in the number of studied patients, the available observations from each patient are often quite limited. The "sparse data environments" result from several reasons such as: 1) data loss during transmission; 2) high cost or limited feasibility of acquiring the biomarkers; and 3) the patient being a new participant in the study. Unfortunately, most of the existing literature on degradation modeling is not specifically designed for such sparse data sets [10]–[19].

Generally speaking, the existing literature on data-driven degradation modeling can be classified into two main categories. The first category directly constructs an individual-level degradation model for a unit using only the available data from that unit such as the least squares model (LSM) [12] and the maximum likelihood estimation model [14]. These methods often require a large amount of historical data to maintain a certain level of model accuracy. Another category focuses on the population characteristics across a set of units, and then, leverages the available data from the unit of interest to construct an individual-level degradation model [10], [11], [15]. Direct implementation of the above-mentioned techniques to sparse data sets may lead to several issues, including: 1) the high sensitivity of the estimated parameters of the degradation model to the given data; 2) the lack of interpretability of the degradation model; and 3) the poor prediction accuracy of the degradation model. In Section II, we review in depth these conventional approaches for degradation modeling as well as some recent studies that are related to the sparse data sets. Furthermore, we review some existing methods for the multivariate analysis [e.g., recommender systems, multioutput Gaussian process (MOGP)] that model multiple signals simultaneously. These models may be effective for sparse data sets because they capture the relationship between the multiple signals (e.g., biomarkers for patients); however, they are not mainly designed for degradation modeling.

To the best of our knowledge, the existing literature still lacks a reliable degradation modeling approach that is suitable for the units with a limited number of observations. This paper seeks to fill this gap and tackle the unique challenges in sparse data environments by developing a structural degradation modeling (SDM) framework. Our approach is inspired by the recommender system, which recommends an item to a user based on: 1) other recommenders who rated that item before and 2) the historical ratings of the user. We first define each available unit with historical observations as a recommender and the unit of interest as the user; and then, we model the degradation status of each unit of interest as a combination of the recommenders by considering: 1) the available data from the unit of interest; 2) the population characteristics; 3) the relationship between the recommenders; and 4) the precision of the recommenders. Essentially, our proposed method integrates two important ideas: 1) leveraging the available data from the unit of interest to improve the modeling fitting of the individual unit over the observed time domain and 2) considering the relationship between the available units to extract proper and accurate population characteristics to address the challenge of limited observations.

II. Literature Review

As mentioned earlier, our proposed method is inspired by the recommender system framework, which provides recommendations about specific items for the user of interest. Therefore, we first introduce some background information of the existing recommender systems in Section II-A. We then discuss potential and existing degradation modeling approaches in Section II-B, which includes: 1) the MOGP and its potential application for degradation modeling; 2) the Bayesian mixed-effects model (MEM) that is commonly used for degradation modeling; and 3) some degradation-based approaches that attempt to address the sparse data sets.

A. Recommender Systems

There exist different frameworks for the recommender system including matrix completion [20], collaborative filtering [21], user-oriented neighborhood [22], and so on. However, most of those frameworks are oriented to estimate the attitude of the user toward a finite set of discrete items, which cannot be leveraged for the continuous time analysis. Also, the existing frameworks are often application specific and assume prior correlated characteristics between different items, such as movies that belong to the same genre, and advertisements that belong to the same category.

Fig. 1 illustrates one conventional recommender system framework, which consists of learning the attitude of user j for item y from other recommenders who previously rated item y. For example, Koren *et al.* [22] implemented the recommender system to provide better movie recommendations to the users. The dashed box contains the following unknown values that are calculated upon request: 1) the estimated rating for item y by user j and 2) the probability that recommender i shares the same attitude as user j which is denoted by $P(Z_j = R_i)$. Specifically, the attitude of user j for item y, $Z_{j,y}$, can be estimated as the following:

$$Z_{j,y} = \sum_{i=1}^{m} P(Z_j = R_i) * R_{i,y} + \epsilon_{j,y} \tag{1}$$

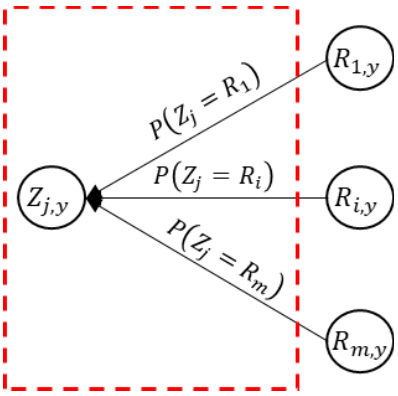

Fig. 1. Conventional recommender system framework.

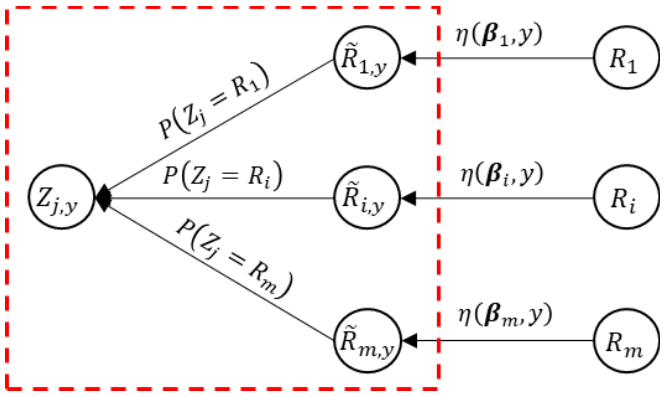

Fig. 2. Extended recommender system framework.

where $R_{i,y}$ is the rating of recommender i for item y; $\epsilon_{j,y}$ represents the error; and m is the number of available recommenders for item y.

Different approaches have been proposed in the literature to estimate $P(Z_j = R_i)$ such as the constrained least squares estimator that aims to minimize $\sum_{y \in \boldsymbol{Y}^j_{\text{hist}}} (\sum_{i=1}^{m} P(Z_j = R_i) * R_{i,y} - Z_{j,y})^2$ given that $P(Z_j = R_i) \geq 0$, $\sum_{i=1}^{m} P(Z_j = R_i) = 1$, and $\boldsymbol{Y}^j_{\text{hist}}$ is the list of items that were previously rated by user j. However, these approaches often require that all the suggested recommenders for item y have already rated enough common items with user j to achieve an accurate estimation of $P(Z_j = R_i)$. As mentioned in [23], many approaches alternatively approximate $P(Z_j = R_i)$ based on the similarity $\text{sim}(i, j)$ between user j and recommender i

$$P(Z_j = R_i) \propto \text{sim}(i, j). \tag{2}$$

For more details about the similarity metric, refer to [23].

The conventional recommender system framework requires that all the utilized recommenders should have rated item y before; otherwise, these recommenders will not be informative for learning the attitude of user j for item y. However, it is common in practice that only a few recommenders have previously rated an item y (e.g., a new movie that has been recently featured and not yet rated by many recommenders). In such a case, the estimated attitude for user j for item y based on (1) is dependent on a limited number of recommenders, which may not be reliable. To address this issue, further extensions have been proposed in the literature as shown in Fig. 2 [21]. Here, the dashed box contains the following unknown values that are calculated upon request: 1) the probability measure $P(Z_j = R_i)$; 2) estimated ratings from the recommenders; and 3) the estimated rating for the user. Unlike the conventional approach that only focuses on the actual ratings, the extended framework utilizes the estimated ratings from the recommenders based on their historical preference. In other words, the attitude of user j for item y, $Z_{j,y}$, is estimated as the following:

$$Z_{j,y} = \sum_{i=1}^{m} P(Z_j = R_i) * \tilde{R}_{i,y} + \epsilon_{j,y} \tag{3}$$

where $\tilde{R}_{i,y}$ is the estimated rating for recommender i and it is assumed to follow the parametric form $\eta(\boldsymbol{\beta}_i, y)$, where $\boldsymbol{\beta}_i$ is estimated based on the recommenders' historical preference.

Although this approach tackles the challenge of missing values from each recommender (i.e., items that a recommender has never rated), it is very critical to accurately calculate $\boldsymbol{\beta}_i$ to obtain a reliable estimation of $\tilde{R}_{i,y}$. In addition to that, $P(Z_j = R_i)$ may not be accurately estimated, if there is only a limited rating history for user j.

While the recommender system aims to address a different problem from degradation modeling, it provides a good framework to address sparse data sets by leveraging the relationship between the user of interest and the recommenders, which inspires our proposed method in Section III.

B. Degradation Modeling

1) Multioutput Gaussian Process: Gaussian process is a statistical model where observations occur in a continuous domain, e.g., time or space. In a Gaussian process, every point in some continuous input space is associated with a normally distributed random variable and the main task of the Gaussian process is to model the covariance between any two input points. Gaussian processes are mostly designed to analyze single outputs and their main advantages include: 1) they are often used as a nonparametric technique to flexibly capture complex signal forms and 2) they can be utilized for the continuous time analysis. The MOGP, as an extension of the conventional Gaussian process further collaboratively captures the correlation between the different outputs to better predict a set of multiple outputs.

For degradation modeling, we may consider the degradation signal from each unit as an output. This allows the MOGP to simultaneously model the signals from multiple units. Here, we focus on a recent method, the collaborative MOGP, introduced in [24] that showed good performance, especially when some of the outputs have missing data. This is highly desired for modeling the sparse data set of our interest. Specifically, Ngyuen and Bonilla [24] assumed all the outputs (e.g., the degradation signals from the units) share Q latent functions, where each latent function follows an independent Gaussian process. They also assumed that each output function has a Gaussian process prior. Finally, they mixed the stochastic

processes from the latent functions and the output functions to construct the MOGP model. They also provided the MATLAB code with the necessary libraries. In this paper, we will use the provided code as a comparison study.

Despite the promise, there are also several limitations for using the MOGP method. As discussed earlier, MOGP is commonly used as a nonparametric technique to capture complex signal forms. However, such characteristics limit its extrapolation performance [25]. While some studies considered introducing a parametric form for characterizing the mean function as an extension, it removes the nonparametric nature of the MOGP and diminishes one of its major contributions. Finally, another major drawback for the MOGP is the high computational cost which limits it from real-time analysis in comparison to other parametric methods such as the MEM.

2) Mixed-Effects Degradation Modeling: For degradation modeling, the MEM that considers both the population and individual characteristics, is one of the most commonly used techniques in the literature. The MEM model was first introduced in [10] and can be written as

$$s_t = \eta(\boldsymbol{\varphi}, \boldsymbol{\nu}, t) + \epsilon_t \tag{4}$$

where $\eta(\cdot)$ is the parametric form of the degradation model; s_t is the measurement for describing the underlying degradation status at time$t$; $\boldsymbol{\varphi}$ is a vector of fixed-effect parameters that represents common characteristics of the population; $\boldsymbol{\nu}$ is a vector of random-effect parameters that characterizes the unit-to-unit variability; and ϵ_t is an error term that represents the measurement noises. Depending on the parametric form of $\eta(\cdot)$, this degradation model can be used to describe a variety of functional forms according to the evolution of the degradation signal. Based on the MEM, extensive studies have been developed in the literature. For example, Gebraeel *et al.* [11] proposed a Bayesian version of the MEM [(Bayesian MEM (BMEM)] that first fitted a MEM given a set of historical units, and then, utilized the Bayesian approach for updating the random-effect parameters given the degradation history of the unit of interest. This approach accounts for: 1) the population characteristics via the MEM and 2) the individual characteristics via the Bayesian update. Without loss of generality, in the following, we focus on the pth order polynomial degradation model considered in [15] for a demonstration:

$$s_t = \sum_{\alpha=0}^{p} (\nu_\alpha) t^\alpha + \epsilon_t = \boldsymbol{\Gamma}_t \boldsymbol{\nu} + \epsilon_t \tag{5}$$

where p is the order of the polynomial model; $\boldsymbol{\nu} = [\nu_0, \ldots, \nu_p]^T$ is the random-effect parameters and often assumed to follow a multivariate normal distribution, $\boldsymbol{\nu} \sim N_{p+1}(\boldsymbol{u}^0, \boldsymbol{\Sigma}^0)$; ϵ_t is the random noise and follows $N(0, \sigma^2)$; and $\boldsymbol{\Gamma}_t = [1, t, \ldots, t^p]$.

Then, the estimated degradation status at time t can be calculated as

$$\hat{s}_t | \boldsymbol{s}_\cdot = \boldsymbol{\Gamma}_t * \boldsymbol{\nu}^1 \tag{6}$$

where $\boldsymbol{s}_\cdot = [s_{t(1)}, \ldots, s_{t(n)}]^T$ is the vector of the observed measurements for the unit of interest up to current time $t(n)$; and $\boldsymbol{\nu}^1$ is the posterior distribution of the random-effect parameters and follows $N_{p+1}(\boldsymbol{u}^1, \boldsymbol{\Sigma}^1)$, where

$$\boldsymbol{u}^1 = \left(\frac{\boldsymbol{\Psi}^T \boldsymbol{\Psi}}{\sigma^2} + (\boldsymbol{\Sigma}^0)^{-1} \right)^{-1} \left(\frac{\boldsymbol{\Psi}^T \boldsymbol{s}_\cdot}{\sigma^2} + (\boldsymbol{\Sigma}^0)^{-1} \boldsymbol{u}^0 \right)$$

$$\boldsymbol{\Sigma}^1 = \left(\frac{\boldsymbol{\Psi}^T \boldsymbol{\Psi}}{\sigma^2} + (\boldsymbol{\Sigma}^0)^{-1} \right)^{-1},$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Gamma}_{t(1)}^T & \cdots & \boldsymbol{\Gamma}_{t(n)}^T \end{bmatrix}^T \in R^{n \times (p+1)}$$

and n is the number of available observations.

Accordingly, the prediction of the degradation status at time t is just a realization of the distribution $\hat{s}_t | \boldsymbol{s}_\cdot \sim N(\boldsymbol{\Gamma}_t \boldsymbol{u}^1, \boldsymbol{\Gamma}_t \boldsymbol{\Sigma}^1 \boldsymbol{\Gamma}_t^T)$. While the BMEM shows promising results in rich data sets, the Bayesian update procedure may not be effective for units with a small number of available observations. This is because, for units with a limited number of observations, the updating procedure tends to focus more on the population characteristics which may not be effective for future predictions of the degradation status for the unit of interests.

3) Degradation Modeling Approaches for Sparse Data Sets: Currently, there are few approaches for degradation modeling, which attempt to address the challenges of sparse data environments. For example, Lin *et al.* [26] assumed that each individual degradation model can be written as a weighted combination of K canonical models, where the weights and the K canonical models are learned simultaneously by an iterative algorithm. However, there are several limitations of the proposed method: 1) this iterative algorithm only leads to a stationary point, which may not be globally optimal; 2) how to choose the initial starting point to reach a solution near the optimal parameters is not discussed, which makes the entire proposed approach difficult to be implemented in practice; and 3) the approach cannot be easily employed for real-time monitoring of the unit of interest due to the high computational cost of the iterative algorithm involved in learning the weights as well as the K canonical models.

There also exist other approaches that utilize stochastic processes to analyze the degradation processes with sparse data (e.g., large interarrival times between the degradation observations). Specifically, Peng *et al.* [27] proposed a Bayesian framework with inverse Gaussian process models to analyze the degradation of heavy duty machine tool's spindle systems where the position accuracy is measured at intermittent discrete time points. Although such methods showed good performance to continuously characterize and interpolate the degradation status of a unit over the observed time domain, they are ineffective for extrapolation and predictive analytics as mentioned in [25] because they: 1) mainly focus on modeling the relationship between the available observations within each individual unit and 2) fail to capture the functional form of the model, which is critical for extrapolation.

III. Methodology Development

In this paper, we propose an SDM framework that utilizes the recommender system framework for degradation modeling

by considering: 1) the available data from the unit of interest; 2) the population characteristics; 3) the relationship between the recommenders; and 4) the precision of the recommenders. Specifically, we consider the unit of interest as the user and the remaining available units with historical observations as recommenders. Then, we adopt and modify the extended recommender system framework in Fig. 2 to satisfy the specific needs for degradation modeling. Section III-A first provides the problem formulation, and then, a motivation example. Next, Section III-B discusses the details of the proposed SDM framework.

A. Problem Formulation

Our main idea is to construct the degradation model of unit j of interest as a combination of the expected values from the recommenders

$$s_{j,t} = \sum_{i=1}^{m} (w_{ij} E[s_{i,t}|\boldsymbol{s}_{i,\cdot}]) + b_j + \epsilon_{j,t} \tag{7}$$

where $\sum_{i=1}^{m} (w_{ij} E[s_{i,t}|\boldsymbol{s}_{i,\cdot}]) + b_j$ is the predicted degradation status for unit j at time t; $\boldsymbol{s}_{i,\cdot}$ is the vector of available observations for recommender i; m is the number of available recommenders; w_{ij} is the weight of recommender i in construction of the model of unit j; b_j is a bias term; and $\epsilon_{j,t}$ quantifies the measurement errors. Note that index i in the summation represents recommender i and it is different from the unit of interest j.

In (7), w_{ij} is a key parameter because it quantifies the similarity between the degradation profiles for unit j and recommender i. Once w_{ij} is accurately estimated, we will then be able to leverage the degradation profiles of the recommenders to accurately predict the future degradation status for the unit of interest. It may be intuitive to calculate w_{ij} via the existing methods such as: 1) maximizing the likelihood or 2) minimizing the sum of squared errors over the observed time domain (i.e., least squares approach). However, such approaches do not explain that w_{ij} quantifies the similarity between the degradation profiles for unit j and recommender i; and therefore, the interpretability is limited. In addition, for a unit j with limited observations, there is high uncertainty involved in the estimation of w_{ij} if it is based only on the limited observations from unit j.

Next, we provide an example to illustrate the challenges for modeling the AD with limited observations. Specifically, Fig. 3 shows in red circles, the mini mental score examination (MMSE) measurements for an AD patient of interest; in blue squares, a recommender for the MMSE of a typical mild AD patient; in green stars, a recommender for the MMSE of a typical moderate AD patient; and in black diamonds, a recommender for the MMSE of a typical severe AD patient.

Our goal is to estimate the degradation model of the patient of interest using (7). With the maximum likelihood estimator, the model parameters learned for this example are

$$\begin{bmatrix} b \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 1.2 \\ 0.86 \\ -1.0 \\ 0.89 \end{bmatrix}$$

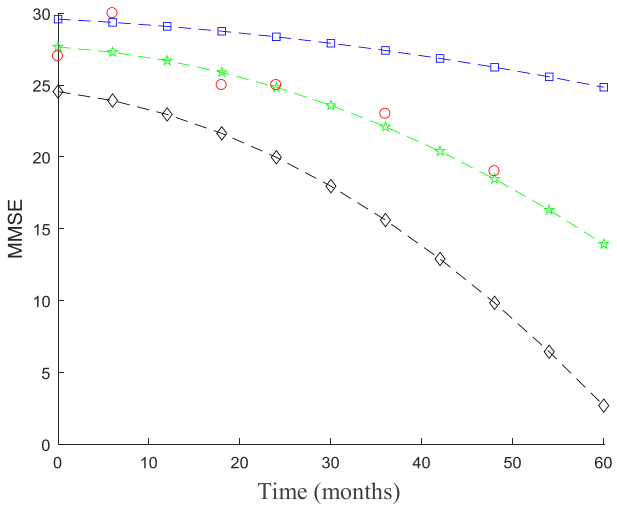

Fig. 3. Blue squares: mild AD. Green stars: moderate AD. Black diamonds: severe AD. Red circles: MMSE measurements of a patient of interest.

which also minimizes the sum of squared errors. However, this model is limited in interpretation and does not provide any insights on the level of sickness. To be specific, the results indicate that the degenerative process of the patient of interest is almost equally close to the degradation processes of the severe patient as well as the mild patient ($w_1 = 0.86$ and $w_3 = 0.89$). In other words, this model indicates that the patient should follow the path of a moderate patient. Contradictorily, the results also show that the patient's degradation model is far from that of the moderate patient ($w_2 = -1$). Furthermore, to understand the sensitivity of the model to newly observed data, we intentionally hide the last observation and derive the model parameters again:

$$\begin{bmatrix} b \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} -73 \\ 2.8 \\ 1.9 \\ 0.95 \end{bmatrix}.$$

It can be seen that there is a large difference in the estimated parameters after hiding the last observation, which indicates that the estimated parameters are highly sensitive to newly observed data and the derived degradation model is not reliable for extrapolation.

B. Structural Degradation Modeling Framework

To address the challenges in sparse data sets, we propose to model the degradation status for the unit of interest based on the recommender system framework by considering the precision of the recommenders and the relationship between the recommenders as shown in the solid box in Fig. 4.

However, unlike the conventional recommender system, the proposed approach estimates the expected value and variance of the degradation status for recommender i at time t as $E[s_{i,t}] = \eta(\boldsymbol{\theta}_i, t)$ and $\text{Var}[s_{i,t}] = \sigma_i^2$, where $\eta(\boldsymbol{\theta}_i, t)$ is assumed to be monotonic. The monotonicity assumption on degradation is natural and common in practice [9], [28], [29]. For example, the National Institutes of Health (NIH) described AD as an irreversible, progressive brain disorder that slowly

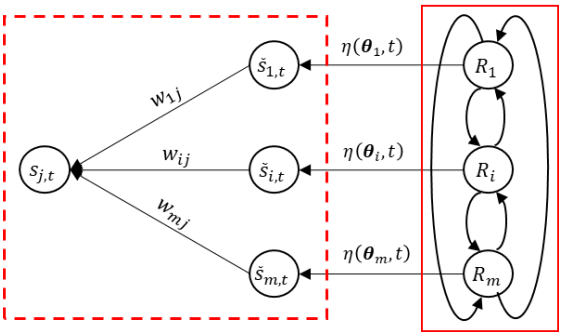

Fig. 4. SDM framework.

destroys memory and thinking skills. The NIH also mentioned that "current Alzheimer's treatments cannot stop AD from progressing," and "they can temporarily slow the worsening of dementia symptoms."

To highlight our main ideas, we focus on the polynomial degradation process because: 1) it is commonly used to model the MMSE in AD [26] and 2) various functional forms such as the exponential and logarithmic models can be transformed to polynomial functional forms (check Appendix C for more details on estimating the parameters of the recommenders).

The following SDM formulation is proposed to estimate the parameters w_{ij} and b_j in (7):

$$\begin{aligned}
&\min_{w_{ij},b_j} \sum_{r=1}^{n_j} \{E[s_{j,t_j(r)}] - s_{j,t_j(r)}\}^2 \\
&\qquad + h(n_j) \left\{ \sum_{i=1}^{m} \sum_{k=1}^{m} (w_{ij} d_{ik} w_{kj}) + \sum_{i=1}^{m} w_{ij} \sigma_i \sigma_i w_{ij} \right\} \\
&\text{s.t.} \sum_{i=1}^{m} w_{ij} = 1, \\
&\qquad w_{ij} \geq 0, \quad \text{for all pairs } (i,k) \\
&\qquad z_{ik}(w_{ij} - w_{kj}) = 0, \quad \text{for all pairs } (i,k) \\
&z_{ik} = \begin{cases} 1 & \text{if recommenders } i \text{ and } k \text{ are identical} \\ 0 & \text{otherwise} \end{cases}
\end{aligned} \tag{8}$$

where $E[s_{j,t}] = \sum_{i=1}^{m} (w_{ij} E[S_{i,t} | \boldsymbol{S}_{i,\cdot}]) + b_j$ is the predicted degradation status for unit j at time t, where $s_{i,t} | \boldsymbol{s}_{i,\cdot}$ is a random variable that represents the degradation status of recommender i given its historical degradation information; z_{ik} is an indicator variable to check if recommenders i and k are identical, i.e., share the same degradation model; n_j is the number of available observations for unit j; $h(n_j)$ is a tuning parameter that depends on n_j; $t_j(r)$ is the time at which the rth observation is obtained for unit j; and $d_{ki} = d_{ik} = \sqrt{d_{ik}} * \sqrt{d_{ki}} \geq 0$ is the dissimilarity between recommenders i and k. Thus, in the objective function, $w_{ij} d_{ik} w_{kj} = w_{ij} \sqrt{d_{ik}} * \sqrt{d_{ki}} w_{kj}$ is the weighted dissimilarity between recommenders i and k, which depends on the estimated weights w_{ij} and w_{kj}.

Unlike the existing methods in degradation modeling, the SDM framework takes into consideration of all the followings.

1) The precision of the recommenders via the term $\sum_{i=1}^{m} w_{ij} \sigma_i \sigma_i w_{ij}$.
2) The relationship between the recommenders via the term $\sum_{i=1}^{m} \sum_{k=1}^{m} (w_{ij} d_{ik} w_{kj})$.
3) The population characteristics via modeling the unit of interest j as a combination of the recommenders and by the definition of the weight w_{ij} that it quantifies the similarity between the degradation profiles for unit j and recommender i.
4) The individual characteristics via minimizing the sum of squared errors $\sum_{r=1}^{n_j} \{E[s_{j,t_j(r)}] - s_{j,t_j(r)}\}^2$.

In the following, we provide detailed discussions for each term considered earlier. For 1), when predicting the degradation status of unit j, it is preferred to focus more on precise recommenders. In other words, if σ_i^2 is relatively high with respect to other recommenders, then we prefer to assign a low value for w_{ij}. Therefore, to achieve a reliable estimation, we add the term $\sum_{i=1}^{m} w_{ij} \sigma_i \sigma_i w_{ij}$ to diminish the influence from recommenders with low levels of precision.

For 2), we focus on minimizing the overall weighted dissimilarity for controlling the stability in the constructed degradation model. Considering two nonsimilar recommenders i and k (i.e., d_{ik} is large), a large value in $w_{ij} d_{ik} w_{kj}$ means that large weights w_{ij} and w_{kj} are assigned to construct the degradation model for unit j, which thus, leads to an uninterpretable and unstable model similar to the case in the motivation example in Fig. 3 (w_1 and w_3 are very large and close to each other). To avoid this issue, we propose to minimize the overall pairwise weighted dissimilarity between the recommenders, $\sum_{i=1}^{m} \sum_{k=1}^{m} (w_{ij} d_{ik} w_{kj})$, to enhance the stability and interpretation of the constructed degradation model. Ideally, we prefer $w_{ij} d_{ik} w_{kj} \to 0$ when recommenders i and k do not share similar degradation characteristics (i.e., d_{ik} is large). This is equivalent to mitigating the contribution of recommender i (i.e., $w_{ij} \to 0$), or recommender k (i.e., $w_{kj} \to 0$), or both the recommenders (i.e., $w_{ij} \to 0$ and $w_{kj} \to 0$).

The dissimilarity metric is a commonly used measure in the literature of clustering and classification such as density-based spatial clustering of applications with noise [30], k-nearest neighbors [31], k-means [32], [33], and ratio cut [34]. In particular, most of the existing methods set the dissimilarity metric d_{ii} to 0, because there is no dissimilarity between the recommender and itself. Furthermore, extensive studies have been done in the literature of the recommender systems as well [35]. Without loss of generality, in this paper, we consider the square of the Mahalanobis distance, $d_{ki} = (\boldsymbol{\theta}_i - \boldsymbol{\theta}_k)^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\theta}_k)$ for $i \neq k$ to measure the dissimilarity between recommenders i and k. Here, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is the covariance matrix of $\boldsymbol{\theta}$ and it can be estimated by the sample covariance matrix from the parameters of the recommenders.

For 4), $\sum_{r=1}^{n_j} (E[s_{j,t_j(r)}] - s_{j,t_j(r)})^2$ is the sum of squared errors over the observed time domain, which is used to ensure the constructed model accurately characterizes the degradation evolution of the unit of interest j. In the objective function, the weighted dissimilarity $\sum_{i=1}^{m} \sum_{k=1}^{m} (w_{ij} d_{ik} w_{kj}) + \sum_{i=1}^{m} w_{ij} \sigma_i \sigma_i w_{ij}$ is more focused on the population characteristics, whereas the sum of squared errors $\sum_{r=1}^{n_j} (E[s_{j,t_j(r)}] - s_{j,t_j(r)})^2$ is more focused on the individual characteristics

of the unit of interest. $h(n_j)$ balances this tradeoff. In this paper, we propose to consider $h(n_j)$ as a decreasing positive function with respect to n_j. This is because when there are more observations available from unit j, we will be more confident to rely on the individual observations to characterize the unit. As an illustration, we set $h(n_j) = \lambda/n_j$, where λ is a tuning parameter and it is calculated via cross validation.

Finally, by the definition of w_{ij} that it quantifies the similarity between the degradation profiles for unit j and recommender i, the following constraints should be satisfied: 1) $w_{ij} \geq 0$ for $i = 1, \ldots, m$ and 2) weights from equivalent recommenders should be equal, which mathematically can be written as $z_{ik}(w_{ij} - w_{kj}) = 0$ such that $z_{ik} = 1$ if recommenders i and k are identical; otherwise, $z_{ik} = 0$. Since we are given a finite set of recommenders, we focus on the normalized similarity measure (i.e., $\sum_{i=1}^{m} w_{ij} = 1$).

With all the above-mentioned efforts, we expect the proposed model to be more interpretable and robust to overfitting. As an illustration, we apply the proposed framework to the example shown in Fig. 3. The results show that the optimal solution is

$$\begin{bmatrix} b \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 0.97 \\ 0.09 \\ 0.77 \\ 0.14 \end{bmatrix}$$

before hiding the last observation; and

$$\begin{bmatrix} b \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 0.88 \\ 0.10 \\ 0.78 \\ 0.12 \end{bmatrix}$$

after hiding the last observation. From these results, we can see that the proposed method has a nice interpretation that the patient is more likely to have a moderate AD status. In addition, the weights do not vary much compared to the maximum likelihood and the least squares approaches before and after hiding the last observation. In Sections IV and V, we will further thoroughly study our proposed method under different scenarios.

For simplicity, we rewrite the SDM formulation as the following:

$$\begin{aligned} \min_{\boldsymbol{w}_j} \ & \boldsymbol{w}_j^T \boldsymbol{\Psi}^T \boldsymbol{\Psi} \boldsymbol{w}_j - 2\boldsymbol{s}_{j,\cdot}^T \boldsymbol{\Psi} \boldsymbol{w}_j + \boldsymbol{s}_{j,\cdot}^T \boldsymbol{s}_{j,\cdot} + h(n_j) \boldsymbol{w}_j^T \boldsymbol{D}^* \boldsymbol{w}_j \\ \text{s.t.} \ & \boldsymbol{o}^T \boldsymbol{w}_j = 1 \\ & \boldsymbol{A} \boldsymbol{w}_j \geq \boldsymbol{0} \end{aligned} \tag{9}$$

where $\boldsymbol{s}_{j,\cdot} = [s_{j,t_j(1)}, s_{j,t_j(2)}, \ldots, s_{j,t_j(n_j)}]^T$ is the vector of available observations for the unit of interest j; $\boldsymbol{w}_j = [b_j, w_{1j}, \ldots, w_{mj}]^T \in R^{(m1)\times 1}$ is a vector that contains both the bias term and the weights

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & \eta(\boldsymbol{\theta}_1, t_j(1)) & \ldots & \eta(\boldsymbol{\theta}_m, t_j(1)) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \eta(\boldsymbol{\theta}_1, t_j(n_j)) & \ldots & \eta(\boldsymbol{\theta}_m, t_j(n_j)) \end{bmatrix} \in R^{n_j \times (m+1)}$$

is the design matrix; $\boldsymbol{o} = [0, 1, \ldots, 1]^T \in R^{(m1)\times 1}$ is a vector containing all ones except for the first entry to be 0; and $\boldsymbol{A} = \mathbf{diag}(\boldsymbol{o})$ is a diagonal matrix with a diagonal vector $\boldsymbol{o}$; and

$$\boldsymbol{D}^* = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & d_{11}^* + \sigma_1^2 & \ldots & d_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & d_{m1} & \ldots & d_{mm}^* + \sigma_m^2 \end{bmatrix} \in R^{(m+1)\times(m+1)}$$

is the modified dissimilarity matrix, in which $d_{ii}^* = \sum_{k=1}^{m} d_{ik}$, $d_{ki} = d_{ik} \geq 0$, and $d_{ii} = 0$. The rationale of the definition of d_{ii}^* is that if recommender i is mistakenly chosen, then it adds up to the uncertainty of the constructed model. This uncertainty would be quantitatively high if recommender i is totally off from the true degradation model, which is estimated as a combination of the remaining recommenders' models. Accordingly, if recommender i is highly nonsimilar from the remaining recommenders, then it is risky to include recommender i in the constructed model. The third and fourth constraints of (8) are removed by the results of Lemma 1.

In Lemma 1, we prove that any two equivalent recommenders will have equal weights. This Lemma allows removing the last two constraints in formulation (8) and maintaining the interpretation of w_{ij}. In Lemma 2, we further prove that the proposed SDM formulation (9) is convex. Therefore, the formulation can be efficiently solved at a low computational cost by many existing solvers [36].

Lemma 1: If recommenders i^* and k^* are identical (i.e., $\boldsymbol{\theta}_{i^*} = \boldsymbol{\theta}_{k^*}$ and $\sigma_{i^*}^2 = \sigma_{k^*}^2$), then the resulting weights w_{i^*j} and w_{k^*j} from the SDM formulation (9) are also identical (see Appendix A for details).

Lemma 2: The SDM formulation (9) is convex (see Appendix B for details).

IV. Simulation Studies

In this section, we investigate the performance of the proposed SDM via simulation studies and compare it with the following benchmark methods: 1) recommender system model (RM); 2) BMEM; 3) LSM; 4) collaborative MOGP with a constant mean function (MOGP); and 5) collaborative MOGP with a quadratic mean function (MOGP-Q). For the RM, we estimate w_{ij} by its restricted maximum likelihood estimator subject to $\sum_{i=1, i\neq j}^{m} w_{ij} = 1$ and $w_{ij} \geq 0$ for $i = 1, 2, \ldots, m$. In our simulation studies, we focus on the AD application and simulate the MMSE measurements for AD patients. The MMSE ranges between 0 (worst condition) and 30 (best condition) and it is expected to decrease with time for all patients. In total, 100 patients were simulated over a period of 36 months with measurements taken monthly.

For each simulated patient i, the remaining 99 simulated patients are considered to be the recommenders. The MMSE measurement is assumed to follow a second-order polynomial model [26], [37]:

$$M_{i,t} = \theta_{i,0} + \theta_{i.1} t + \theta_{i,2} t^2 + \epsilon_{i,t} \tag{10}$$

where $M_{i,t}$ is the simulated MMSE measurement for patient i at time t (t is the time in months); $\epsilon_{i,t}$ represents the random noise in the MMSE measurements and it is assumed to follow a normal distribution $N(0, \sigma_i^2)$; and $\boldsymbol{\theta}_i = [\theta_{i,0}, \theta_{i,1}, \theta_{i,2}]^T$ are

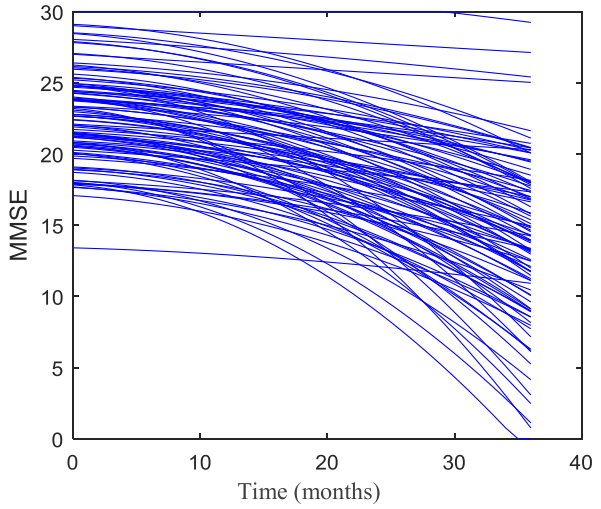

Fig. 5. Underlying degradation models for AD patients over 36 months.

the parameters for patient i and assumed to follow the normal distribution $N_3(\boldsymbol{u}, \boldsymbol{\Sigma})$ with $\boldsymbol{u} = [23; -0.05; -0.005]^T$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 10 & -10^{-4} & -10^{-5} \\ -10^{-4} & 10^{-4} & 10^{-5} \\ -10^{-5} & 10^{-5} & 10^{-5} \end{bmatrix}.$$

Fig. 5 shows the underlying degradation models for AD patients over 36 months.

In our simulation studies, nine different scenarios are considered with two varying factors: 1) the variance of the random noise σ_i^2 with values 1, 9, and 16 and 2) the percentage of missing data (sparsity level) by randomly hiding 20%, 50%, and 80% of the observations from the full data set, i.e., some units may have more hidden observations than other units. For each scenario, we 1) calculate λ via holdout cross validation and 2) evaluate the metrics in (11) and (12).

Specifically, for each simulated patient j, we define the mean squared error (MSE) based on the hidden observations as the following:

$$\text{MSE}(j) = \frac{1}{n_j^h} \sum_{t \in H} (M_{j,t} - \hat{M}_{j,t})^2 \tag{11}$$

where H is the set of time points for the hidden measurements; n_j^h is the number of hidden measurements for patient j; and $\hat{M}_{j,t}$ is the estimated measurement for patient j at time t (i.e., $\hat{M}_{j,t} = \hat{\theta}_{j,0} + \hat{\theta}_{j,1} t + \hat{\theta}_{j,2} t^2$ via the LSM and the BMEM approaches, and $\hat{M}_{j,t} = \sum_{i=1}^{m} w_{ij} \hat{M}_{i,t} + b_j$ via the SDM and the RM frameworks).

Since the true parametric simulation model is known, we can also calculate the MSE of the parameters (pMSE) for patient j as the following:

$$\text{pMSE}(j) = \frac{(\boldsymbol{\theta}_j - \hat{\boldsymbol{\theta}}_j)^T (\boldsymbol{\theta}_j - \hat{\boldsymbol{\theta}}_j)}{p+1} \tag{12}$$

where p is the order of the polynomial degradation model and it is equal to 2 in our simulation study; and $\hat{\boldsymbol{\theta}}_j = [\hat{\theta}_{j,0}, \hat{\theta}_{j,1}, \hat{\theta}_{j,2}]^T$ is the estimated set of degradation parameters for patient j based on the available observations (i.e., excluding the hidden observations). Note that if a constructed model fits the available observations of a unit accurately but fails to estimate the true degradation model, it is expected that the future predictions from the constructed model will be inaccurate. Therefore, we utilize pMSE here to measure the long-term prediction performance of the constructed model, such that a lower error in the model parameters is expected to produce better predictions in the future. On the contrary, the MSE metric in (11) characterizes the prediction performance within the observation window.

Table I summarizes the performance comparisons regarding the MSE metric. From Table I, we can see that the SDM shows a relatively better performance compared to the benchmark methods, especially at high percentages of missing data. This is an expected consequence because: 1) the constructed models from the LSM and BMEM are unstable and highly sensitive to the variance of the available data; 2) the MOGP and MOGP-Q may falsely characterize the correlation between some of the units due to the limited data availability or the high level of noise involved in the data; and 3) the MOGP is suitable for interpolation but not extrapolation [25]. On the contrary, the SDM framework focuses on precise recommenders and considers the relationship between the recommenders, which stabilize the constructed degradation model. As a conclusion from the MSE metric results, the proposed SDM framework is expected to have a better performance than the benchmark methods for sparse data sets and/or in the presence of high variance of noise. Note that data sets with high noise variance share some challenges with sparse data sets such as the risk of overfitting and lack of interpretability. This is because noisy data sets typically require more observations than a cleaned data set to achieve a comparable model fitting performance.

To have a better understanding of the computational cost, we further conduct numerical experiments by using MATLAB V9.2 in a 64-bit Windows 7 Enterprise operating system with Intel Core i7-6600U 2.60-GHz CPU and 16-GB RAM. The overall computational times (seconds) for the SDM, BMEM, LSM, MOGP, and MOGP-Q are 3.36, 0.046, 0.023, 2330, and 3223, respectively. This time includes model construction and model prediction for all the patients. The results are expected because the computational time increases with the increase of model complexity. For example, 1) the LSM model focuses only on the data from the patient of interest, which thus is expected to be the fastest and 2) the MOGP requires solving a challenging and complex optimization problem, which thus leads to an extremely slow performance.

Fig. 6 provides a better illustration with the noise variance equal to 9. It clearly shows that the SDM outperforms the benchmark methods at high percentages of missing data. Note that the MSE metric is based on randomly selected hidden observations; therefore, it measures the interpolation performance because those hidden observations are scattered and not concentrated in a specific time window.

To better understand the predictive performance for the long run, we further evaluate the results of the pMSE metric, which are present in Table II. Here, we do not show the results from

TABLE I

SAMPLE MEAN (μ_{MSE}) AND SAMPLE STANDARD DEVIATION (σ_{MSE}) OF THE MSE UNDER DIFFERENT SIMULATED SCENARIOS WITH THE BEST PERFORMING MODEL IN BOLD

Noise Variance	Missing Data (%)	SDM $[\mu_{MSE}, \sigma_{MSE}]$	RM $[\mu_{MSE}, \sigma_{MSE}]$	BMEM $[\mu_{MSE}, \sigma_{MSE}]$	LSM $[\mu_{MSE}, \sigma_{MSE}]$	MOGP $[\mu_{MSE}, \sigma_{MSE}]$	MOGP-Q $[\mu_{MSE}, \sigma_{MSE}]$
1	20	[0.25, 0.15]	[0.25, 0.15]	[0.25, 0.15]	[0.26, 0.16]	[0.73, 0.81]	**[0.25, 0.15]**
1	50	[1.13, 0.60]	[1.13, 0.58]	**[1.12, 0.56]**	[1.19, 0.63]	[3.09, 3.52]	[1.13, 0.58]
1	80	**[7.38, 10.0]**	[9.67, 22.1]	[7.96, 10.9]	[22.5, 83.7]	[21.3, 25.0]	[9.74, 22.4]
9	20	**[2.24, 1.51]**	[2.26, 1.51]	[2.27, 1.52]	[2.35, 1.53]	[2.73, 2.08]	[2.28, 1.51]
9	50	**[9.83, 4.16]**	[10.1, 4.42]	[10.1, 4.22]	[11.0, 5.02]	[11.3, 5.68]	[10.2, 4.44]
9	80	**[48.5, 30.0]**	[54.9, 45.6]	[55.7, 32.4]	[107, 201]	[57.8, 34.3]	[55.0, 46.6]
16	20	**[3.96, 2.67]**	[4.01, 2.68]	[4.03, 2.71]	[4.16, 2.72]	[4.44, 3.19]	[4.04, 2.68]
16	50	**[17.3, 7.17]**	[17.9, 7.80]	[17.8, 7.37]	[19.5, 8.96]	[18.6, 8.21]	[18.0, 7.83]
16	80	**[84.1, 48.8]**	[95.1, 73.7]	[96.1, 53.6]	[166, 245]	[91.5, 48.8]	[95.4, 75.8]

TABLE II

SAMPLE MEAN (μ_{pMSE}) AND SAMPLE STANDARD DEVIATION (σ_{pMSE}) OF THE PMSE UNDER DIFFERENT SIMULATED SCENARIOS WITH THE BEST PERFORMING MODEL IN BOLD

Noise Variance	Percentage Missing Data	SDM $[\mu_{pMSE}, \sigma_{pMSE}]$	RM $[\mu_{pMSE}, \sigma_{pMSE}]$	BMEM $[\mu_{pMSE}, \sigma_{pMSE}]$	LSM $[\mu_{pMSE}, \sigma_{pMSE}]$
1	20	**[0.030, 0.038]**	[0.061, 0.070]	[0.043, 0.050]	[0.097, 0.11]
1	50	**[0.055, 0.078]**	[0.10, 0.14]	[0.089, 0.11]	[0.21, 0.31]
1	80	**[0.17, 0.24]**	[0.28, 0.49]	[0.56, 2.86]	[5.59, 35.4]
9	20	**[0.28, 0.37]**	[0.45, 0.64]	[0.35, 0.42]	[0.83, 1.02]
9	50	**[0.38, 0.47]**	[0.72, 1.03]	[0.63, 0.76]	[1.53, 2.29]
9	80	**[0.81, 1.06]**	[1.59, 2.33]	[3.49, 5.35]	[20.9, 119]
16	20	**[0.44, 0.57]**	[0.74, 1.08]	[0.59, 0.71]	[1.48, 1.82]
16	50	**[0.56, 0.70]**	[1.19, 1.64]	[1.07, 1.27]	[2.73, 4.07]
16	80	**[1.31, 1.69]**	[2.59, 3.66]	[6.13, 9.33]	[37.2, 212]

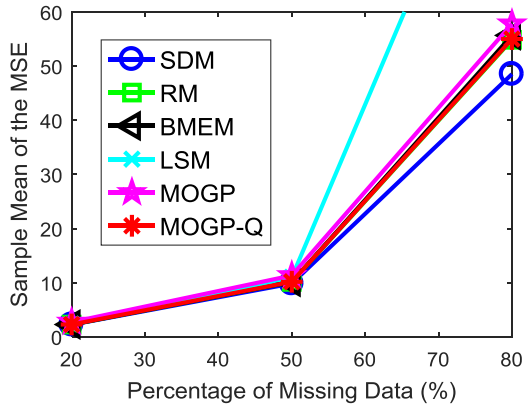

Fig. 6. Sample mean of the MSE for noise variance equals to 9.

the MOGP and MOGP-Q because they are based on data and do not give estimates for the parametric model in (10).

From Table II, we can see that: 1) the proposed SDM framework outperforms the benchmark methods in all the scenarios and 2) the benchmark methods perform poorly at high percentages of missing data. The first observation is expected because the SDM framework considers the relationship between the recommenders, which ensures the stableness of the constructed degradation models even at high levels of missing data. The second observation stems from the facts that: 1) the RM and LSM tend to overfit the limited observations, which thus leads to a good performance over the observed time domain but not for the future predictions and 2) in presence of limited observations, the BMEM tends to focus more on the population characteristics that may not accurately capture the unique characteristics of the unit of interest. As a conclusion, Table II gives the predictive power of the proposed method. For a better visual illustration, Fig. 7 shows the sample mean value of the pMSE for the scenarios with a noise variance equal to 9.

Note that: 1) the LSM considers only the sum of squared errors; 2) the BMEM considers the population characteristics, but it does not consider the relationship between the recommenders and the unit of interest; and 3) the RM considers the relationship between the recommenders and the unit of interest via the constraints $w_{ij} \geq 0$ and $\sum_{i=1}^{m} w_{ij} = 1$, and the component $\sum_{r=1}^{n_j} (E[s_{j,t_j(r)}] - s_{j,t_j(r)})^2$. However, the RM does not consider the relationship between the recommenders themselves, which is critical for the stability of the model. As a result, this comparison study further shows the importance of the considered components in the proposed approach.

Additional simulations are conducted to better understand the performance of the proposed approach versus the number of available observations. Specifically, 500 patients are sim-

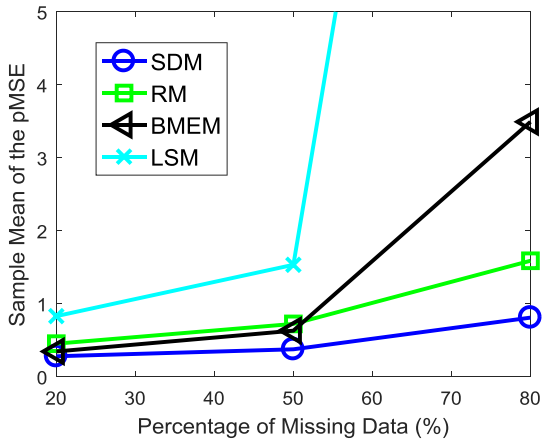

Fig. 7. Sample mean of the pMSE for noise variance equals to 9.

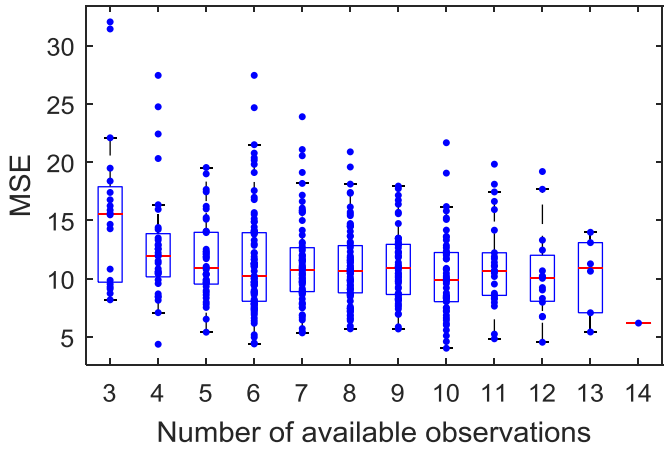

Fig. 8. MSE when the noise variance equals to 9 and the level of missing data equals to 80%.

ulated when the variance of the noise is set to 9 and the percentage of missing data is set at 80%. We follow (11) to calculate the MSE of the hidden observations for each patient. The results are summarized in Fig. 8, which shows the MSE for each of the 500 patients versus the corresponding number of available observations and the boxplots of the MSE at different levels of available observations.

From Fig. 8, 1) there is no clear trend showing that the performance of the proposed approach improves with more available observations and 2) there are few extreme patients that perform poorly, which is mostly due to the limited available data, and/or the lack of having similar recommenders to those patients. As a conclusion, this simulation study further shows the advantage of our proposed method for sparse data environments.

V. Case Study

This section further investigates the performance of the proposed SDM framework based on the ADNI data set [38], which involves participants between the age of 55 and 90 from USA and Canada. We only consider the RM, BMEM, and LSM as the three benchmark methods and exclude the MOGP. This is because there are a little shared observational time points between the patients, which result in an ill-conditioned MOGP. The data set contains personal information and longitudinal measurements of examinations and biomarkers for the participating patients. Here, we focus on the MMSE because it has been widely used to predict the AD status of a patient in practice. Furthermore, in this paper, we consider all of the 583 participating patients that have four or more MMSE measurements. This requirement is needed to ensure the construction of valid models via the LSM and BMEM approaches. For a better visualization, Fig. 9 shows the MMSE degradation curves for a subset of the patients. We can observe that different patients may have a different number of available observations and that the observations are collected at different time points. In addition, the data set of MMSE measurements is sparse in nature because: 1) MMSE measurement is recorded semiannually for each patient typically; 2) patients may join the program at different AD stages; and 3) some patients may skip some of their semiannual visits to the clinic. Thus, the given MMSE data set presents a real practical challenge and also provides a good example to test the efficacy of the proposed structural model for sparse data sets.

A. Problem Setup

In this paper, all the participating patients excluding the patient of interest are considered to be the recommenders, which totals to 582 recommenders for every patient of interest. Specifically, we first model the MMSE measurements of the recommenders as a quadratic model [26], [37]

$$M_{i,t} = \theta_{i,0} + \theta_{i.1}t + \theta_{i,2}t^2 + \epsilon_{i,t} \tag{13}$$

where $M_{i,t}$ is the MMSE measurement for recommender i at time t; t is the time after the first visit for each recommender; $\theta_{i,0}$ is the MMSE score of recommender i for the first visit to the clinic; and $\boldsymbol{\theta}_i = [\theta_{i,0}, \theta_{i,1}, \theta_{i,2}]^T$ are the parameters of the degradation model for recommender i.

Second, to set up the dissimilarity matrix, we focus on $d_{ik} = d_{ki} = (\boldsymbol{\theta}_i - \boldsymbol{\theta}_k)^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\theta}_k)$ to measure the dissimilarity between recommenders i and k, and $d_{ii}^* = \sum_{k=1}^{582} [(\boldsymbol{\theta}_i - \boldsymbol{\theta}_k)^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\theta}_k)]$ to measure the dissimilarity between recommender i and all the remaining recommenders.

Third, with our proposed SDM framework, we model the patient of interest j given the available MMSE measurements for that patient

$$\boldsymbol{M}_{j,\cdot} = [M_{j,t_j(1)}, \ldots, M_{j,t_j(n_j)}]^T$$

as

$$M_{j,t} = \left\{ \sum_{i=1}^{582} [w_{ij}(\theta_{i,0} + \theta_{i.1}t + \theta_{i,2}t^2)] \right\} + b_j + \epsilon_{j,t} \tag{14}$$

where $M_{j,t}$ is the true MMSE score for patient j at time t, and $\boldsymbol{w}_j = [b_j, w_{1,j}, \ldots, w_{582,j}]^T$ is the solution of the optimization problem in (9) by simply replacing $\boldsymbol{s}_{j,\cdot}$ by $\boldsymbol{M}_{j,\cdot}$. Note that for every patient of interest j, there is a different set of 582 recommenders. To calculate the tuning parameter λ, we conduct a leave-one-out cross validation (i.e., consider each patient as a validating patient once). Specifically, we first initialize λ with a random positive value, estimate $\boldsymbol{w}$ for every validating patient by considering the remaining patients as recommenders, and then, we calculate the prediction errors

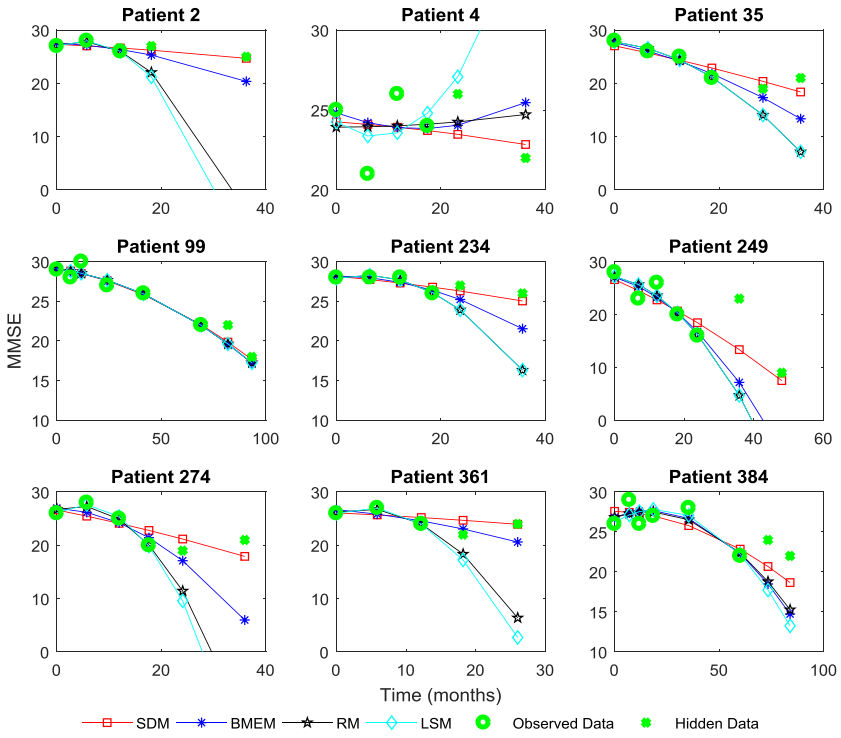

Fig. 9. True and predicted MMSE measurements from a subset of patients.

for the validating patients. Next, we update λ iteratively to minimize the MSE of all the predictions for the validating patients via a gradient descent algorithm [36].

B. Prognostics Performance

To measure the prognostics performance of the proposed and benchmark degradation models, we conduct the leave-one-out cross validation. In other words, for each validating patient j, we calculate $\boldsymbol{w}_j$ such that all the remaining 582 patients are considered as recommenders. Furthermore, we hide the last two observations for each validating patient j, and we calculate the root of the squared difference (rSD) between the predicted and true measurements for each of the hidden observations. Mathematically, rSD is defined as

$$\text{rSD}(j,t) = \sqrt{(M_{j,t} - \hat{M}_{j,t})^2} \tag{15}$$

where

$$\hat{M}_{j,t} = \left\{ \sum_{i=1}^{582} [w_{ij}(\theta_{i,0} + \theta_{i.1}t + \theta_{i,2}t^2)] \right\} + b_j$$

is the expected MMSE estimate for patient j at time t.

To better visualize and compare the prognostic performance of the proposed SDM framework and the benchmark methods, Fig. 9 shows the prediction results of the MMSE for some patients, and Figs. 10 and 11 show the boxplots of rSD of the two hidden measurements for all the participating patients.

From Figs. 9–11, we can see that: 1) the SDM framework results in lower means and lower variances for the rSD of both hidden measurements compared to the benchmark methods and 2) the SDM method shows more stable results and smaller increases in the errors from the first hidden measurement to the second hidden measurement than the benchmark methods. The first observation stems from the fact that the SDM framework considers the relationship between the recommenders and the precision of the recommenders. The second observation shows the importance of considering: 1) the relationship between the user and the recommenders and 2) the pairwise relationship between the recommenders to construct a stable and accurate degradation model that shows consistently satisfactory prognostic performance. Note that unlike the simulation study, here we do not know the underlying true model, and thus, we cannot calculate the pMSE metric defined in (12). In summary, this case study further validates our conclusion that for sparse data sets, the proposed SDM framework outperforms the benchmark methods.

VI. Discussion and Conclusion

Predicting and modeling the progression of degradation is important and critical to a wide set of applications. Fortunately, the development of information technologies (e.g., clouds) has facilitated the collection and storage of information from a massive number of operating units and provided a great environment for condition monitoring. While the number of recorded units can be large, many units often have limited available observations due to different reasons such as data loss during transmission, and high cost or limited feasibility of data acquisition. Thus, the direct implementation of the existing degradation modeling approaches in such sparse data environments may lead to: 1) a low model accuracy for

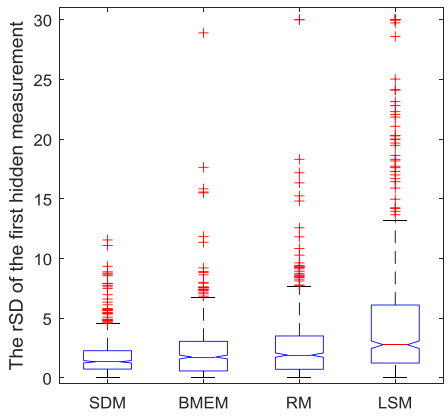

Fig. 10. Boxplot for the rSD of the first hidden measurement.

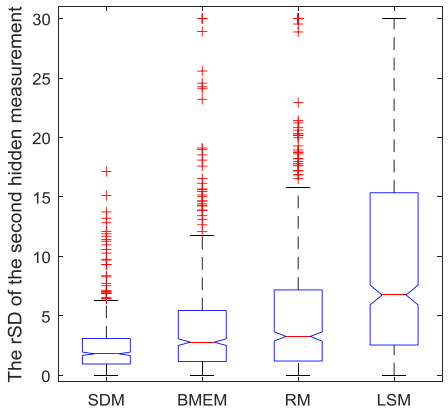

Fig. 11. Boxplot for the rSD of the second hidden measurement.

prediction; 2) a high uncertainty of the model parameters; and 3) a lack of interpretability of the model.

This paper aims to fill the literature gap by developing an SDM framework that addresses the unique challenges of the sparse data environments. Specifically, the degradation model of each individual unit is structured as a combination of a set of recommenders by considering: 1) the available data from the unit of interest; 2) the relationship between the recommenders; 3) the precision of the recommenders based on past performance; and 4) the population characteristics. The developed framework is tested and validated by simulation studies as well as the ADNI data set and the results show that the SDM framework outperforms the benchmark methods for degradation modeling.

There are several important topics for the future research. First, it is important to further explore functional data analysis techniques and nonparametric approaches to model the degradation status of the recommenders because for some applications, the true degradation model may be complicated and cannot be well characterized by a parametric form. Second, in this paper, we propose to consider $h(n_j) = \lambda/n_j$. In the future studies, it would be interesting to investigate more on the choice of $h(n_j)$. Third, the proposed method is expected to perform well in the presence of a large number of recommenders, i.e., the degradation profiles from the available recommenders are likely to span the entire set of possible degradation profiles. How to extend the proposed approach to more general scenarios is worth studying. Finally, it would be of high interest to integrate the system performance analysis [39], [40] with the proposed approach for better decision-making and preventive control.

APPENDIX A

This Appendix proves Lemma 1 that if $\exists i^*, k^*$ such that $\boldsymbol{\theta}_{i^*} = \boldsymbol{\theta}_{k^*}$ and $\sigma^2_{i^*} = \sigma^2_{k^*}$, then the solution of formulation (9) guarantees $w_{i^*j} = w_{k^*j}$. First, we split the objective function of the SDM formulation (9) into two pieces

$$q(w_{ij}) = \boldsymbol{w}_j^T \boldsymbol{\Psi}^T \boldsymbol{\Psi} \boldsymbol{w}_j - 2\boldsymbol{s}_{j,\cdot}^T \boldsymbol{\Psi} \boldsymbol{w}_j + \boldsymbol{s}_{j,\cdot}^T \boldsymbol{s}_{j,\cdot} = \sum_{r=1}^{n_j} \left(\sum_{i=1}^{m} (w_{ij}\eta(\boldsymbol{\theta}_i, t_j(r))) + b_j - s_{j,t_j(r)} \right)^2$$

and $f(w_{ij}) = \boldsymbol{w}_j^T \boldsymbol{D}^* \boldsymbol{w}_j$. Next, we isolate the indexes i^* and k^* in $f(w_{ij})$. This is equivalent to writing

$$\begin{aligned} f(w_{ij}) &= \sum_{i=1}^{m} w_{ij}^2 \sigma_i^2 + \sum_{i=1}^{m} \sum_{k=i+1}^{m} (w_{ij} + w_{kj})^2 d_{ik} \\ &= \sum_{\substack{i=1 \\ i \neq i^*, k^*}}^{m} w_{ij}^2 \sigma_i^2 + w_{i^*j}^2 \sigma_{i^*}^2 + w_{k^*j}^2 \sigma_{k^*}^2 \\ &\quad + \sum_{\substack{i=1 \\ i \neq i^*, k^*}}^{m} \sum_{\substack{k=i+1 \\ k \neq i^*, k^*}}^{m} (w_{ij} + w_{kj})^2 d_{ik} + \sum_{\substack{i=1 \\ i \neq k^*}}^{m} (w_{ij} + w_{i^*j})^2 d_{i^*i} \\ &\quad + \sum_{\substack{i=1 \\ i \neq i^*}}^{m} (w_{ij} + w_{k^*j})^2 d_{k^*i} + (w_{i^*j} + w_{k^*j})^2 d_{i^*k^*} = g(w_{ij}) \\ &\quad + w_{i^*j}^2 \sigma_{i^*}^2 + w_{k^*j}^2 \sigma_{k^*}^2 + \sum_{i=1}^{m} (w_{ij} + w_{i^*j})^2 d_{i^*i} \\ &\quad + \sum_{i=1}^{m} (w_{ij} + w_{k^*j})^2 d_{k^*i} - (w_{i^*j} + w_{k^*j})^2 d_{i^*k^*} \end{aligned}$$

where

$$g(w_{ij}) = \sum_{\substack{i=1 \\ i \neq i^*, k^*}}^{m} w_{ij}^2 \sigma_i^2 + \sum_{\substack{i=1 \\ i \neq i^*, k^*}}^{m} \sum_{\substack{k=i+1 \\ k \neq i^*, k^*}}^{m} (w_{ij} + w_{kj})^2 d_{ik}$$

and it is independent of w_{i^*j} and w_{k^*j}.

Given that $\boldsymbol{\theta}_{i^*} = \boldsymbol{\theta}_{k^*}$ then $d_{ii^*} = d_{ik^*}$ and $d_{i^*k^*} = 0$, then

$$\begin{aligned} f(w_{ij}) &= g(w_{ij}) + w_{i^*j}^2 \sigma_{i^*}^2 + w_{k^*j}^2 \sigma_{k^*}^2 + \sum_{i=1}^{m} \\ &\quad [(w_{ij} + w_{i^*j})^2 + (w_{ij} + w_{k^*j})^2] d_{ik^*} = g(w_{ij}) \\ &\quad + w_{i^*j}^2 \sigma_{i^*}^2 + w_{k^*j}^2 \sigma_{k^*}^2 + \sum_{i=1}^{m} \\ &\quad (w_{ij}^2 + 2w_{ij}w_{i^*j} + w_{i^*j}^2 + w_{ij}^2 + 2w_{ij}w_{k^*j} + w_{k^*j}^2) d_{ik^*}. \end{aligned}$$

Also, given that $\sigma_{i^*}^2 = \sigma_{k^*}^2$, then

$$f(w_{ij}) = g(w_{ij}) + (w_{i^*j}^2 + w_{k^*j}^2)\sigma_{k^*}^2 + \sum_{i=1}^{m}(w_{ij}^2 + 2w_{ij}w_{i^*j} + w_{i^*j}^2 + w_{ij}^2 + 2w_{ij}w_{k^*j} + w_{k^*j}^2)d_{ik^*}$$

Now, we prove that for any feasible solution $\tilde{w}_{ij}$, there exists another feasible solution

$$\hat{w}_{ij} = \begin{cases} \tilde{w}_{ij} & i \neq i^*, k^* \\ \dfrac{\tilde{w}_{i^*j} + \tilde{w}_{k^*j}}{2} & i = i^*, k^* \end{cases}$$

such that $h(n_j)f(\hat{w}_{ij}) + q(\hat{w}_{ij}) \leq h(n_j)f(\tilde{w}_{ij}) + q(\tilde{w}_{ij})$. Note that $g(\hat{w}_{ij}) = g(\tilde{w}_{ij})$ because $g(w_{ij})$ is independent of w_{i^*j} and w_{k^*j}. Also, $q(\hat{w}_{ij}) = q(\tilde{w}_{ij})$ because $\boldsymbol{\theta}_{i^*} = \boldsymbol{\theta}_{k^*}$ and $\tilde{w}_{i^*j}\eta(\boldsymbol{\theta}_{i^*}, \cdot) + \tilde{w}_{k^*j}\eta(\boldsymbol{\theta}_{k^*}, \cdot) = \hat{w}_{i^*j}\eta(\boldsymbol{\theta}_{i^*}, \cdot) + \hat{w}_{k^*j}\eta(\boldsymbol{\theta}_{k^*}, \cdot)$. Therefore, $h(n_j)\{f(\tilde{w}_{ij}) + q(\tilde{w}_{ij}) - h(n_j)f(\hat{w}_{ij}) - q(\hat{w}_{ij}) = h(n_j)(\tilde{w}_{i^*j}^2 + \tilde{w}_{k^*j}^2 - \hat{w}_{i^*j}^2 - \hat{w}_{k^*j}^2)\sigma_{k^*}^2 + \sum_{i=1}^{m}(2\tilde{w}_{ij}\tilde{w}_{i^*j} + \tilde{w}_{i^*j}^2 + 2\tilde{w}_{ij}\tilde{w}_{k^*j} + \tilde{w}_{k^*j}^2 - 2\hat{w}_{ij}\hat{w}_{i^*j} - \hat{w}_{i^*j}^2 - 2\hat{w}_{ij}\hat{w}_{k^*j} - \hat{w}_{k^*j}^2)d_{ik^*}\}$. Here, $2\tilde{w}_{ij}\tilde{w}_{i^*j} + 2\tilde{w}_{ij}\tilde{w}_{k^*j} = 2\tilde{w}_{ij}(\tilde{w}_{i^*j} + \tilde{w}_{k^*j}) = 2\hat{w}_{ij}(\hat{w}_{i^*j} + \hat{w}_{k^*j})$ because $\hat{w}_{i^*j} = \hat{w}_{k^*j} = (\tilde{w}_{i^*j} + \tilde{w}_{k^*j})/2$; and $\tilde{w}_{i^*j}^2 + \tilde{w}_{k^*j}^2 - \hat{w}_{i^*j}^2 - \hat{w}_{k^*j}^2 = \tilde{w}_{i^*j}^2 + \tilde{w}_{k^*j}^2 - ((\tilde{w}_{i^*j} + \tilde{w}_{k^*j})/2)^2 - ((\tilde{w}_{i^*j} + \tilde{w}_{k^*j})/2)^2 = \tilde{w}_{i^*j}^2 + \tilde{w}_{k^*j}^2 - (\tilde{w}_{i^*j}^2/2) - (\tilde{w}_{k^*j}^2/2) - \tilde{w}_{i^*j}\tilde{w}_{k^*j} = ((\tilde{w}_{i^*j} - \tilde{w}_{k^*j})/\text{sqrt}(2))^2$. Then, $f(\tilde{w}_{ij}) - f(\hat{w}_{ij}) = (\tilde{w}_{i^*j}^2 + \tilde{w}_{k^*j}^2 - \hat{w}_{i^*j}^2 - \hat{w}_{k^*j}^2)\sigma_{k^*}^2 + \sum_{i=1}^{m}(\tilde{w}_{i^*j}^2 + \tilde{w}_{k^*j}^2 - \hat{w}_{i^*j}^2 - \hat{w}_{k^*j}^2)d_{ik^*} = ((\tilde{w}_{i^*j} - \tilde{w}_{k^*j})/\text{sqrt}(2))^2\sigma_{k^*}^2 + \sum_{i=1}^{m}((\tilde{w}_{i^*j} - \tilde{w}_{k^*j})/2)^2 d_{ik^*} = ((\tilde{w}_{i^*j} - \tilde{w}_{k^*j})/\text{sqrt}(2))^2 * (\sigma_{k^*}^2 + \sum_{i=1}^{m} d_{ik^*}) > 0$ if $\tilde{w}_{i^*j} \neq \tilde{w}_{k^*j}$ because $d_{ik^*} > 0$ for $\boldsymbol{\theta}_i \neq \boldsymbol{\theta}_{k^*}$. Accordingly, if $\boldsymbol{\theta}_{i^*} = \boldsymbol{\theta}_{k^*}$ and $\sigma_{i^*}^2 = \sigma_{k^*}^2$ then to minimize $h(n_j)f(w_{ij}) + q(w_{ij})$, it is necessary to have $w_{i^*j} = w_{k^*j}$. To finalize the proof, we show that $\hat{w}_{ij}$ is feasible by showing that it satisfies $\sum_{i=1}^{m}\hat{w}_{ij} = 1$ and $\hat{w}_{ij} \geq 0$, $i = 1, 2, \ldots, m$. Since $\tilde{w}_{ij}$ is a feasible solution, then $\sum_{i=1}^{m}\tilde{w}_{ij} = 1$ and $\tilde{w}_{ij} \geq 0, i = 1, 2, \ldots, m$. Therefore

$$\hat{w}_{ij} = \begin{cases} \tilde{w}_{ij} \geq 0 & i \neq i^*, k^* \\ \dfrac{\tilde{w}_{i^*j} + \tilde{w}_{k^*j}}{2} \geq 0 & i = i^*, k^* \end{cases}$$

and $\sum_{i=1}^{m}\hat{w}_{ij} = \sum_{i=1}^{m}\tilde{w}_{ij} = 1$. Thus, if $\boldsymbol{\theta}_{i^*} = \boldsymbol{\theta}_{k^*}$ and $\sigma_{i^*}^2 = \sigma_{k^*}^2$ then the solution of the SDM formulation will satisfy $w_{i^*j} = w_{k^*j}$.

Appendix B

This Appendix proves Lemma 2 that formulation (9), $\min_{\boldsymbol{w}_j} \boldsymbol{w}_j{}^T\boldsymbol{\Psi}^T\boldsymbol{\Psi}\boldsymbol{w}_j - 2\boldsymbol{s}_{j,\cdot}^T\boldsymbol{\Psi}\boldsymbol{w}_j + \boldsymbol{s}_{j,\cdot}^T\boldsymbol{s}_{j,\cdot} + h(n_j)\boldsymbol{w}_j^T\boldsymbol{D}^*\boldsymbol{w}_j$ subject to $\boldsymbol{o}^T\boldsymbol{w}_j = 1$ and $\boldsymbol{A}\boldsymbol{w}_j \geq \boldsymbol{0}$, is convex.

Since $h(n_j) \geq 0$ by definition, then it is sufficient to prove that $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ and $\boldsymbol{D}^*$ are positive semidefinite (PSD) matrixes. First, $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ is PSD because for any nonzero vector $\boldsymbol{w}_j$, we have $\boldsymbol{w}_j^T\boldsymbol{\Psi}^T\boldsymbol{\Psi}\boldsymbol{w}_j = \|\boldsymbol{\Psi}\boldsymbol{w}_j\|_2^2 \geq 0$. Second, we proof that $\boldsymbol{D}^*$ is PSD by showing that $\boldsymbol{w}_j^T\boldsymbol{D}^*\boldsymbol{w}_j \geq 0$ for any $\boldsymbol{w}_j$.

The dissimilarity matrix can be written as $\boldsymbol{w}_j^T\boldsymbol{D}^*\boldsymbol{w}_j = \sum_{i=1}^{m} w_{ij}^2\sigma_i^2 + \sum_{i=1}^{m} w_{ij}^2 d_{ii}^* + \sum_{i=1}^{m}\sum_{\substack{k=1 \\ k\neq i}}^{m} w_{ij}d_{ik}w_{kj}$. Recall that $d_{ii}^* = \sum_{k=1}^{m} d_{ik}$ and $d_{ii} = 0$, then $\boldsymbol{w}_j^T\boldsymbol{D}^*\boldsymbol{w}_j = \sum_{i=1}^{m} w_{ij}^2\sigma_i^2 + \sum_{i=1}^{m} w_{ij}^2\sum_{k=1}^{m} d_{ik} + \sum_{i=1}^{m}\sum_{k=1}^{m} w_{ij}d_{ik}w_{kj}$. The second part of $\boldsymbol{w}_j^T\boldsymbol{D}^*\boldsymbol{w}_j$ can be written as: $\sum_{i=1}^{m} w_{ij}^2\sum_{k=1}^{m} d_{ik} = w_{1j}^2(d_{12} + d_{13} + \ldots + d_{1m}) + w_{2j}^2(d_{21} + d_{23} + \ldots + d_{2m}) + w_{3j}^2(d_{31} + d_{32} + \ldots + d_{3m}) + \ldots + w_{mj}^2(d_{m1} + d_{m2} + \ldots + d_{m(m1)}) = (w_{1j}^2 + w_{2j}^2)d_{12} + (w_{1j}^2 + w_{3j}^2)d_{13} + (w_{2j}^2 + w_{3j}^2)d_{23} + \ldots + (w_{(m1)j}^2 + w_{mj}^2)d_{(m1)m} = \sum_{i=1}^{m}\sum_{k=i+1}^{m}(w_{ij}^2 + w_{kj}^2)d_{ik}$. Similarly, the third part of $\boldsymbol{w}_j^T\boldsymbol{D}^*\boldsymbol{w}_j$ can be written as: $\sum_{i=1}^{m}\sum_{k=1}^{m} w_{ij}d_{ik}w_{kj} = \sum_{i=1}^{m}\sum_{k=i+1}^{m} 2d_{ik}w_{ij}w_{kj}$ because $d_{ik} = d_{ki}$ and $d_{ii} = 0$. Then, for any $\boldsymbol{w}_j$, $\boldsymbol{w}_j^T\boldsymbol{D}^*\boldsymbol{w}_j = \sum_{i=1}^{m} w_{ij}^2\sigma_i^2 + \sum_{i=1}^{m}\sum_{k=i+1}^{m}(w_{ij}^2 + w_{kj}^2 + 2w_{ij}w_{kj})d_{ik} = \sum_{i=1}^{m} w_{ij}^2\sigma_i^2 + \sum_{i=1}^{m}\sum_{k=i+1}^{m}(w_{ij} + w_{kj})^2 d_{ik} \geq 0$ because $d_{ik} \geq 0$. This concludes our proof that the SDM formulation is convex.

Appendix C

As an illustration for practitioners to calculate $\boldsymbol{\theta}_i$ and σ_i^2 for each recommender, we focus on the constrained least-squares approach because it is computationally efficient and it satisfies the monotonicity requirement for degradation. Specifically, $\boldsymbol{\theta}_i$ can be estimated by the solution of $\sum_{r=1}^{n_i}\{\eta(\boldsymbol{\theta}_i, t_i(r)) - s_{i,t_i(r)}\}^2$ such that

$$\begin{cases} d\eta(\boldsymbol{\theta}_i, t)/dt \geq 0, & \text{if monotonically increasing} \\ d\eta(\boldsymbol{\theta}_i, t)/dt \leq 0, & \text{if monotonically decreasing} \end{cases}$$

and σ_i^2 can be estimated by $(\sum_{r=1}^{n_i}\{\eta(\boldsymbol{\theta}_i, t_i(r)) - s_{i,t_i(r)}\}^2/\text{degrees of freedom})$. Here, n_i is the number of available observations from recommender i; $t_i(r)$ is the time at which the rth observation is obtained for recommender i; and $s_{i,t_i(r)}$ is the degradation status for recommender i at time $t_i(r)$. For example, if we consider the quadratic model for each recommender that can be written as $s_{i,t} = \theta_{i,0} + \theta_{i.1}t + \theta_{i,2}t^2 + \epsilon_t$ and given that the degradation trends are monotonically decreasing, then $\boldsymbol{\theta}_i$ is the solution of $\sum_{r=1}^{n_i}\{\eta(\boldsymbol{\theta}_i, t_i(r)) - s_{i,t_i(r)}\}^2$ such that $\theta_{i,1} \leq 0$ and $\theta_{i,2} \leq 0$; and $\sigma_i^2 \cong \sum_{r=1}^{n_i}\{\eta(\boldsymbol{\theta}_i, t_i(r)) - s_{i,t_i(r)}\}^2/(n_i - 3)$. Note that the least-squares approach is often used under the normality assumption for $s_{i,t}$ because it provides the best linear unbiased estimator.

References

[1] S. Abbate, M. Avvenuti, and J. Light, "Usability study of a wireless monitoring system among Alzheimer's disease elderly population," *Int. J. Telemed. Appl.*, vol. 2014, pp. 1–8, 2014, Art. no. 617495, doi: 10.1155/2014/617495.

[2] K.-Y. Lam, N. W.-H. Tsang, S. Han, J. K.-Y. Ng, S.-W. Tam, and A. Nath, "SmartMind: Activity tracking and monitoring for patients with Alzheimer's disease," in *Proc. IEEE 29th Int. Conf. Adv. Inf. Netw. Appl.*, Mar. 2015, pp. 453–460.

[3] Y.-S. Shin, J.-K. Wee, I. Song, and S. Lee, "Small-area low-power heart condition monitoring system using dual-mode SAR-ADC for low-cost wearable healthcare systems," *Technol. Health Care*, vol. 23, no. s2, pp. S277–S284, Jun. 2015.

[4] P. Moore, F. Xhafa, L. Barolli, and A. Thomas, "Monitoring and detection of agitation in dementia: Towards real-time and big-data solutions," in *Proc. 8th Int. Conf. P2P, Parallel, Grid, Cloud Internet Comput.*, Oct. 2013, pp. 128–135.

[5] K. F. Martin, "A review by discussion of condition monitoring and fault diagnosis in machine tools," *Int. J. Mach. Tools Manuf.*, vol. 34, no. 4, pp. 527–551, May 1994.

[6] M. Landolina *et al.*, "Remote monitoring reduces healthcare use and improves quality of care in heart failure patients with implantable defibrillators: The evolution of management strategies of heart failure patients with implantable defibrillators (EVOLVO) study," *Circulation*, vol. 125, no. 24, pp. 2985–2992, Jun. 2012.
[7] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.
[8] Y. M. Ko and E. Byon, "Condition-based joint maintenance optimization for a large-scale system with homogeneous units," *IISE Trans.*, vol. 49, no. 5, pp. 493–504, May 2017.
[9] A. Chehade, S. Bonk, and K. Liu, "Sensory-based failure threshold estimation for remaining useful life prediction," *IEEE Trans. Rel.*, vol. 66, no. 3, pp. 939–949, Sep. 2017.
[10] C. J. Lu and W. O. Meeker, "Using degradation measures to estimate a time-to-failure distribution," *Technometrics*, vol. 35, no. 2, pp. 161–174, May 1993.
[11] N. Z. Gebraeel, M. A. Lawley, R. Li, and J. K. Ryan, "Residual-life distributions from component degradation signals: A Bayesian approach," *IIE Trans.*, vol. 37, no. 6, pp. 543–557, Jun. 2005.
[12] H. Peng, D. W. Coit, and Q. Feng, "Component reliability criticality or importance measures for systems with degrading components," *IEEE Trans. Rel.*, vol. 61, no. 1, pp. 4–12, Mar. 2012.
[13] R. Zhou, N. Gebraeel, and N. Serban, "Degradation modeling and monitoring of truncated degradation signals," *IIE Trans.*, vol. 44, no. 9, pp. 793–803, Sep. 2012.
[14] N. Chen and K. L. Tsui, "Condition monitoring and remaining useful life prediction using degradation signals: Revisited," *IIE Trans.*, vol. 45, no. 9, pp. 939–952, Sep. 2013.
[15] K. Liu, N. Z. Gebraeel, and J. Shi, "A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 652–664, Jul. 2013.
[16] B. Kosasih, W. Caesarendra, K. Tieu, A. Widodo, C. A. S. Moodie, and A. K. Tieu, "Degradation trend estimation and prognosis of large low speed slewing bearing lifetime," *Appl. Mech. Mater.*, vol. 493, pp. 343–348, Jan. 2014.
[17] R. Moghaddass and M. J. Zuo, "Multistate degradation and supervised estimation methods for a condition-monitored device," *IIE Trans.*, vol. 46, no. 2, pp. 131–148, Feb. 2014.
[18] K. Liu, A. Chehade, and C. Song, "Optimize the signal quality of the composite health index via data fusion for degradation modeling and prognostic analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 3, pp. 1504–1514, Jul. 2017.
[19] A. Chehade, C. Song, K. Liu, A. Saxena, and X. Zhang, "A data-level fusion approach for degradation modeling and prognostic analysis under multiple failure modes," *J. Quality Technol.*, vol. 50, no. 2, pp. 150–165, Apr. 2018.
[20] B. Recht and C. Ré, "Parallel stochastic gradient algorithms for large-scale matrix completion," *Math. Program. Comput.*, vol. 5, no. 2, pp. 201–226, 2013.
[21] B. M. Marlin, "Modeling user rating profiles for collaborative filtering," in *Proc. NIPS*, 2003, pp. 627–634.
[22] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *IEEE Comput.*, vol. 42, no. 8, pp. 30–37, Aug. 2009.
[23] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005.
[24] T. V. Nguyen and E. V. Bonilla, "Collaborative multi-output Gaussian processes," in *Proc. 13th Conf. Uncertainty Artif. Intell.*, 2014, pp. 643–652.
[25] S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain, "Gaussian processes for time-series modelling," *Philos. Trans. Roy. Soc. London A, Math. Phys. Sci.*, vol. 371, no. 1984, pp. 1–25, 2013.
[26] Y. Lin, K. Liu, E. Byon, X. Qian, S. Liu, and S. Huang, "A collaborative learning framework for estimating many individualized regression models in a heterogeneous population," *IEEE Trans. Rel.*, vol. 67, no. 1, pp. 328–341, Mar. 2017.
[27] W. Peng, Y.-F. Li, Y.-J. Yang, S.-P. Zhu, and H.-Z. Huang, "Bivariate analysis of incomplete degradation observations based on inverse Gaussian processes and copulas," *IEEE Trans. Rel.*, vol. 65, no. 2, pp. 624–639, Jun. 2016.
[28] R. A. Sperling *et al.*, "Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimers. Dementia*, vol. 7, no. 3, pp. 280–292, May 2011.
[29] C. R. Jack *et al.*, "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade," *Lancet Neurol.*, vol. 9, no. 1, pp. 119–128, Jan. 2010.
[30] A. Smiti and Z. Eloudi, "Soft DBSCAN: Improving DBSCAN clustering method using fuzzy set theory," in *Proc. 6th Int. Conf. Hum. Syst. Interact. (HSI)*, Jun. 2013, pp. 380–385.
[31] P. Mitra, C. A. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 301–312, Mar. 2002.
[32] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Appl. Statist.*, vol. 28, no. 1, pp. 100–108, 1979.
[33] N. M. Nasrabadi, "Pattern recognition and machine learning," *J. Electron. Imag.*, vol. 16, no. 4, p. 49901, Oct. 2007.
[34] L. Hagen and A. B. Kahng, "New spectral methods for ratio cut partitioning and clustering," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 11, no. 9, pp. 1074–1085, Sep. 1992.
[35] H. Liu, D. Song, S. Rüger, R. Hu, and V. Uren, "Comparing dissimilarity measures for content-based image retrieval," in *Information Retrieval Technology*. Berlin, Germany: Springer, 2008, pp. 44–50.
[36] J. Nocedal and S. Wright, *Numerical Optimization*. New York, NY, USA: Springer, 2006.
[37] J. C. Biesanz, N. Deeb-Sossa, A. A. Papadakis, K. A. Bollen, and P. J. Curran, "The role of coding time in estimating and interpreting growth curve models," *Psychol. Methods*, vol. 9, no. 1, pp. 30–52, 2004.
[38] S. G. Mueller *et al.*, "The Alzheimer's disease neuroimaging initiative," *Neuroimag. Clin.*, vol. 15, no. 4, pp. 869–877, Nov. 2005.
[39] B. Huang, J. Xie, and Q.-M. He, "Cyclic change of server's performance: Impacts and applications," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 703–713, Mar. 2014.
[40] J. Xie, P. Cao, B. Huang, and M. E. H. Ong, "Determining the conditions for reverse triage in emergency medical services using queuing theory," *Int. J. Prod. Res.*, vol. 54, no. 11, pp. 3347–3364, Jun. 2016.

Abdallah Chehade received the B.S. degree in mechanical engineering from the American University of Beirut, Beirut, Lebanon, in 2011, and the M.S. degree in mechanical engineering and the M.S. and Ph.D. degrees in industrial engineering from the University of Wisconsin–Madison, Madison, WI, USA, in 2014, 2014, and 2017, respectively.

He is currently an Assistant Professor with the Department of Industrial and Manufacturing Systems Engineering, University of Michigan–Dearborn, Dearborn, MI, USA. His research interests are data fusion for process modeling, and optimization of data-analysis.

Dr. Chehade is a member of INFORMS and IISE.

Kaibo Liu (M'14) received the B.S. degree in industrial engineering and engineering management from the Hong Kong University of Science and Technology, Hong Kong, in 2009, and the M.S. degree in statistics and the Ph.D. degree in industrial engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2011 and 2013, respectively.

He is currently an Assistant Professor with the Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI, USA. His research interests are data fusion for process modeling, monitoring, diagnosis, and prognostics.

Dr. Liu is a member of INFORMS, IISE, and ASQ.