

A spatiotemporal prediction approach for a 3D thermal field from sensor networks

Di Wang , Kaibo Liu & Xi Zhang

To cite this article: Di Wang , Kaibo Liu & Xi Zhang (2021): A spatiotemporal prediction approach for a 3D thermal field from sensor networks, Journal of Quality Technology

To link to this article: <https://doi.org/10.1080/00224065.2020.1851618>



Published online: 22 Jan 2021.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



A spatiotemporal prediction approach for a 3D thermal field from sensor networks

Di Wang^a , Kaibo Liu^b , and Xi Zhang^c 

^aDepartment of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China; ^bDepartment of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, Wisconsin; ^cDepartment of Industrial Engineering and Management, Peking University, Beijing, China

ABSTRACT

Thermal fields exist widely in engineering systems and are critical for engineering operation, product quality and system safety in many industries. An accurate prediction of thermal field distribution, that is, acquiring any location of interest in a thermal field at the present and future time, is essential to provide useful information for the surveillance, maintenance, and improvement of a system. However, thermal field prediction using data acquired from sensor networks is challenging due to data sparsity and missing data problems. To address this issue, we propose a field spatiotemporal prediction approach based on transfer learning techniques by studying the dynamics of a 3D thermal field from multiple homogeneous fields. Our model characterizes the spatiotemporal dynamics of the local thermal field variations by considering the spatiotemporal correlation of the fields and harnessing the information from homogeneous fields to acquire an accurate thermal field distribution in the future. A real case study of thermal fields during grain storage is conducted to validate our proposed approach. Grain thermal field prediction results provide a deep insight of grain quality during storage, which is helpful for the manager of grain storage to make further decisions about grain quality control and maintenance.

KEYWORDS

3D thermal field;
spatiotemporal prediction;
transfer learning

1. Introduction

Thermal fields exist widely in engineering systems, and an accurate thermal field distribution, that is, acquiring any location of interest in a thermal field at the present and future time, is essential to provide useful information for the surveillance, maintenance, and improvement of systems in various industries, including advanced manufacturing (Li, Jin, and Yu 2018), cyber infrastructure (Jiang et al. 2013), meteorology (Liu et al. 2017), and microenvironment (Yan, Li, and Tu 2017). For example, grain temperature is the most commonly used index for grain quality monitoring during storage. Obtaining an accurate grain thermal field distribution can provide thorough information for timely and effective monitoring of grain quality, which is critical for preventing unnecessary grain losses. Such thermal field distributions are generally determined by a series of extrinsic factors (e.g., ambient temperature or radiation power) and intrinsic factors (e.g., self-interaction within the system) as well as the structure and operational modes of the system. Hence, understanding extrinsic and intrinsic factors

and the associated operation mechanism that may influence a thermal field is important for acquiring the thermal field distribution.

Using partial differential equations (PDEs) to study the thermodynamics of a system enables the prediction of the thermal field distribution by considering extrinsic factors (Jia, Sun, and Cao 2000). Such PDEs are established on the basis of complete understanding of the operational mechanism of a system, and the structures and associated parameters in PDEs should be verified through experimental tests or deductive reasoning. However, in practice, solving PDEs for a complex engineering system is not feasible at all times because the explicit structures cannot be easily obtained and the parameters in PDEs are not consistently treated as constants. Finite element analysis (FEA) method is used extensively to obtain numerical solutions for PDEs and considers time-varying parameters in thermal systems (Wang et al. 2020). However, FEA models usually acquire the thermal field distribution under ideal conditions. These models are adept for obtaining the global trend of the thermal field

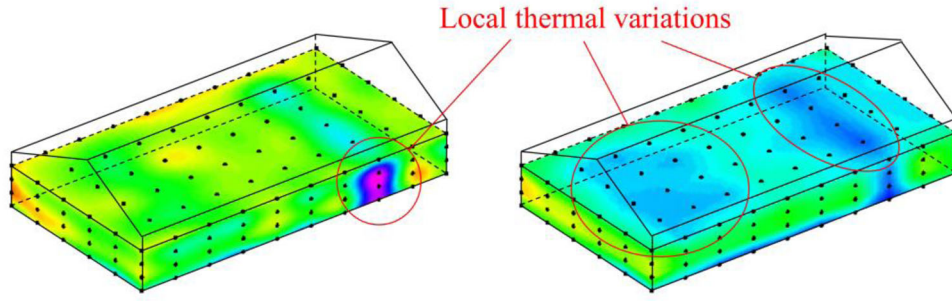


Figure 1. Thermal field distribution of a 3D granary with local thermal variations. (The cloud represents the grain thermal field, and the black points represent sensor locations.).

caused by extrinsic factors but usually fail to capture local thermal variations caused by intrinsic factors, which is of importance for providing useful information regarding the severity and risks of the actual thermal system.

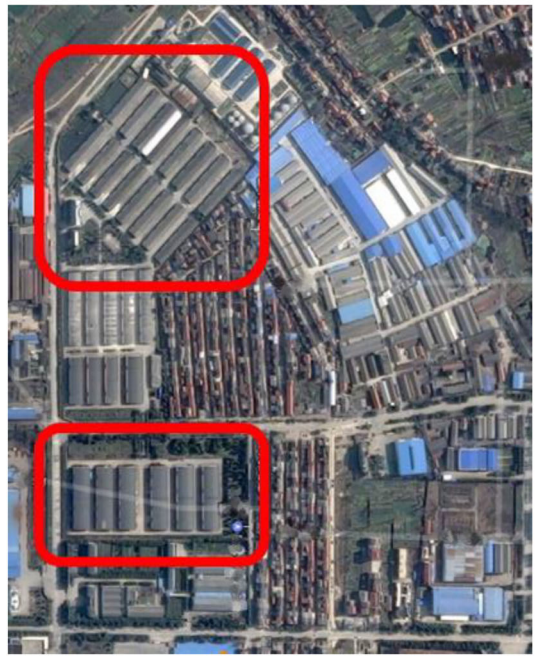
Throughout the thermal field distribution of a system, practitioners have focused on local thermal variation because such changes usually trigger systematic changes or even system failure. For example, local thermal variations exist in a 3D granary, as shown in Figure 1. A small local temperature increase due to the overheat by the grain self-breath or mildewing may induce an extensive temperature increase spreading across the granary, thereby leading to the unexpected grain deterioration before releasing to the grain processing plants. In view of the severity of local thermal variations, accurately predicting the thermal field distribution at the present and future time points is of considerable necessity.

Wireless sensor networks and information technologies have emerged as key support for linking sensor observations; moreover, studying the thermodynamics of a 3D thermal field caused by intrinsic factors (e.g., local thermal variations) using the sensor observations shows a remarkably promising research orientation. However, thermal field distribution prediction using sensor observations has encountered two major problems. First, the ‘data sparsity’ problem exists. ‘Data sparsity’ means that only sparse sensors are engaged in sensor networks for data collection due to the limited budget for sensor deployment. Therefore, only sparse sensor data are available for thermal field distribution prediction. Second, missing data invariably exist among sensor observations in the sensor network for unexpected reasons, such as sensor aging and wireless communication failure.

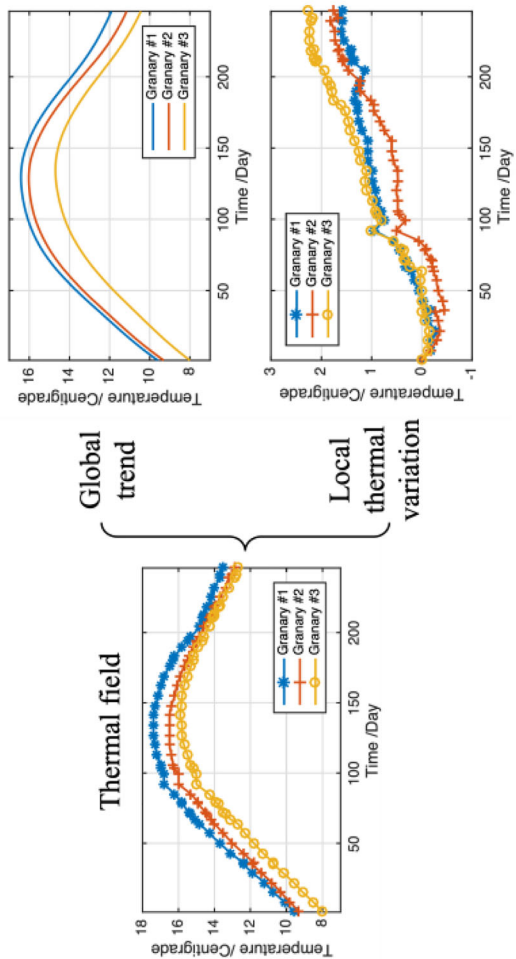
To address the issue of data sparsity and missing data, we find thermal fields with similar external conditions have homogeneous properties in engineering practices, so that temperature in this fields has similar

profiles. As shown in Figure 2(a), we use grain storage as an example. Several columns of granaries with similar storage conditions (within the red boxes), including locations, volumes of granaries, stored grains, and environmental surroundings, are found in this grain depot. Granaries with similar storage conditions have homogeneous properties, which makes temperature in these granaries have similar profiles. When we predict a grain thermal field of a target granary, we can try to capture available sensor data in its neighboring homogeneous granaries to support the thermal field prediction of the target granary. We also present a demonstrative example of thermal sensor observations at the same location in three adjacent granaries with similar storage conditions in Figure 2(b). The curves of the global trend are obtained using the thermodynamic PDE model, and the local thermal variations marked with dots are the values of the deviance of the sensor observations from the global trend. We can see the local thermal variation in each granary has similar profiles, and exhibits considerable spatial and temporal correlations.

Transfer learning, which shares knowledge or information from data sources in homogeneous subjects, provides an unprecedented opportunity for thermal field modeling with limited sensor observations and addresses data sparsity and missing data problems. In recent years, transfer learning has been studied by numerous researchers and applied to various engineering domains, including speech emotion recognition (Song, Jin, and Zhao 2014), manufacturing shape deviation (Shao et al. 2017), engine degradation processes (Lin et al. 2018), and disease telemonitoring (Yoon and Li 2018). However, existing transfer learning approaches mainly focus on the prediction of a subject at the current time, that is, estimating the current values of points of interest on the sensor locations. Thermal field prediction in the upcoming time (that is, forecasting the future values of points of interest on and off the sensor locations) cannot be directly solved by existing approaches.



(a)



(b)

Figure 2. Homogeneous grain thermal fields with similar thermal profiles: (a) Aerial view of a grain depot with multiple homogeneous granaries in the central part of China; (b) Anatomy of a thermal field. (A thermal field is decomposed of the global trend and the local thermal variation.).

Thermal field prediction generally involves twofold objectives: one is to estimate the entire thermal field distribution at the current time by using the collected data from the sparse sensors, and the other is to forecast the future distribution in the upcoming time. In recent years, studies have been done for field prediction by modeling spatiotemporal correlations, including the kriging method (Stein 2012; Inoue, Sasaki, and Washio 2012), sparse matrix algorithms (Furrer, Genton, and Nychka 2006), reduced rank techniques (Cressie and Johannesson 2008), and Gaussian Markov random field (GMRF; Xu and Huang 2012). These studies mainly focus on field distribution estimation at the current time by considering the spatiotemporal correlation structures within the field. However, they require dense observations for an accurate field distribution. When sensor sparsity or missing data problems exist, these studies fail to achieve accurate field distribution modeling and estimation results. Currently, few studies have concentrated on the field forecast in the upcoming time. Therefore, developing an effective method for thermal field prediction given limited sensor observations with data sparsity and missing data problems is essential.

Three challenges are involved in predicting a 3D thermal field distribution in our study: 1) estimating the values of points of interest off the sensor locations at the current time points; 2) forecasting the thermal field distribution (i.e., values of points of interest on and off the sensor locations) at the upcoming time points; and 3) handling sensor data sparsity and missing data problems to acquire an accurate thermal field distribution. To address these challenges, this study presents a spatiotemporal transfer learning-based prediction (TLP) approach to learn a 3D thermal field distribution when only sparse sensor observations are accessible and missing data problems exist. For the first two challenges, in view of the complex spatiotemporal structures of the thermal field, our approach combines the ideas of field estimation and forecast to acquire a distribution of the entire thermal field at the current and upcoming time points. For the third challenge, we develop the TLP approach to handle the data sparsity and missing data problems by sharing the information from data sources in homogeneous thermal fields.

The remainder of this paper is organized as follows. Section 2 presents a literature review in terms of thermal field prediction, including field estimation and forecast. Section 3 introduces the methodology of transfer learning-based thermal field prediction. Section 4 shows a real case study of grain thermal

fields for evaluating the proposed model performance. Section 5 provides a summary.

2. Literature review

The goal of this study is to predict a 3D thermal field distribution using sensor observations with data sparsity and missing data problems. In this section, we review relevant physical models and statistical spatiotemporal approaches for the modeling of thermodynamic systems and the methods for solving data sparsity and missing data problems.

2.1. Physical models and mixed models

Conventional methods for thermal field distribution modeling capture thermal field dynamics by using physical models, which consider extrinsic factors and rely on physical principles (e.g., heat exchange mechanism) to describe the field. For example, Khatchatourian and De Oliveira (2006) applied a hydromechanical model to characterize the airflow and dynamics of a thermal field. Wang and Zhang (2015) simulated a thermodynamic process in a cubic granary by considering the environmental factors, solar radiation, and heat transmission among the grains. The physical models mainly describe thermodynamics on the basis of PDEs and rely on FEA method to obtain numerical solutions (Thijssen 2017). However, these methods model thermal field distribution under ideal conditions by only considering extrinsic factors. Uncertainties caused by intrinsic factors (e.g., local thermal variations) cannot be accurately captured, thereby leading to a large discrepancy between the predicted and actual thermal fields.

Studies that combine global and local models have been conducted to address this issue. In such mixed models, the global model captures the global trend caused by extrinsic factors, whereas the local model captures the local variations caused by intrinsic factors. For example, Ba and Joseph (2012) proposed a novel nonstationary Gaussian process model that is a composite of two Gaussian processes in which the first one captures the global trend and the second one captures local variations. Yan, Paynabar, and Shi (2018) developed a spatiotemporal smooth sparse decomposition method that decomposes the original data stream into the functional mean, sparse anomalies, and random noises. These models consider the mean and local models as data-driven models and can fully capture the spatiotemporal correlation from sensor data. However, when only limited sensor data are

available, these models may not work well in capturing the complex spatiotemporal structures of fields. In addition, these methods do not consider engineering knowledge, such as thermal principles, which is critical in most engineering cases.

2.2. Statistical spatiotemporal models

Statistical spatiotemporal approaches have been vastly developed to address the uncertainties of dynamic fields using sensor observations. Previous studies have modeled spatiotemporal correlations of a dynamic field by assuming that spatial and temporal parts are independent of each other (Huang 2010; Katzfuss and Cressie 2011; Zheng, Liuand, and Hsieh 2013). One limitation in these studies is that the independent form of the spatial and temporal parts cannot characterize the spatiotemporal correlations of fields. To overcome this limitation, some studies have attempted to model spatiotemporal correlations of a dynamic field. In particular, the kriging method has been widely used in modeling spatiotemporal correlation by considering a spatiotemporal covariance function for field distribution modeling (Stein 2012; Inoue, Sasaki, and Washio 2012). However, the computational issue generally arises in the kriging method for large datasets. Many approaches, including sparse matrix algorithms (Furrer, Genton, and Nychka 2006), reduced rank techniques (Cressie and Johannesson 2008), and full-scale approximation method (Zhang, Sang, and Huang 2015), have been developed to reduce computational burdens for large datasets. However, these methods, which describe the spatiotemporal correlations of a field by a covariance function, are applicable to simple systems only. For complex systems, such as 3D thermal fields, the modeling accuracy of these methods is unsatisfactory because complex spatiotemporal dynamics in the systems cannot be adequately captured by the covariance function.

To address this issue, studies that combine spatial and temporal modeling methods have been conducted to characterize spatiotemporal dynamics among complex systems. For spatial modeling, the GMRF has received considerable attention in recent years. For example, spatial modeling by GMRF, which considers grid-based neighborhood structures to model spatial correlation of a complex system, has been well validated in nanowire growth (Xu and Huang 2012). However, given that the GMRF can only be adopted when the sampled data in the entire structured field are complete, temperature off the sensor locations cannot be directly predicted by the conventional

GMRF method. To solve the problem, some researchers have adopted methods that combine GMRF and kriging models (Hartman and Hössjer, 2008; Perdikaris et al. 2015). Xu and Choi (2012) proposed a Gaussian process built on a GMRF to model mobile distribution given resource-constrained mobile sensor networks. Wang et al. (2019) integrated a kriging model into a GMRF model by fully using grid-based sensor data. However, these methods mainly focus on field estimation for points of interest at the current time points and still fail to forecast field distribution in a short time period. In other words, these methods cannot characterize the temporal evolution of the field in the near future.

To address the aforementioned issue, Mariella and Tarantino (2010) proposed a spatiotemporal conditional autoregressive (STCAR) model, which is an autoregressive model for a temporal sequence of GMRFs, to forecast field distribution in the near future. Liu, Gopal, and Kalagnanam (2018) improved the STCAR model to forecast weather radar reflectivity fields by considering the motion of the weather system and a spatiotemporal process that governs the growth or decay of the strength of radar reflectivity. The two methods integrate a time series model into a spatial model to forecast field distribution in the near future by characterizing the spatiotemporal correlation of the field. Given that the STCAR model can only be adopted when the sampled data in the entire structured field are accessible, temperature off the sensor locations still cannot be directly predicted by the STCAR method. To predict the field distribution, the thermal quantities should not only forecast on the deployed sensor locations but also those off the sensor locations. In addition, the aforementioned methods generally require sufficient sensor observations for acquiring an accurate field distribution. However, in practice, only limited sensor observations may be available due to data sparsity and missing data problems. Therefore, given limited sensor observations, developing an effective method to predict the field distribution is necessary.

2.3. Models for data sparsity and missing data problems

Some researchers have applied interpolation methods to fill missing data using existing sensor observations of the thermal field, including linear, spline (Aliaga 2017), and Lagrange (Cheng and Mark 2010) interpolations. However, this strategy may result in inaccurate interpolated values when the number of missing data is large,

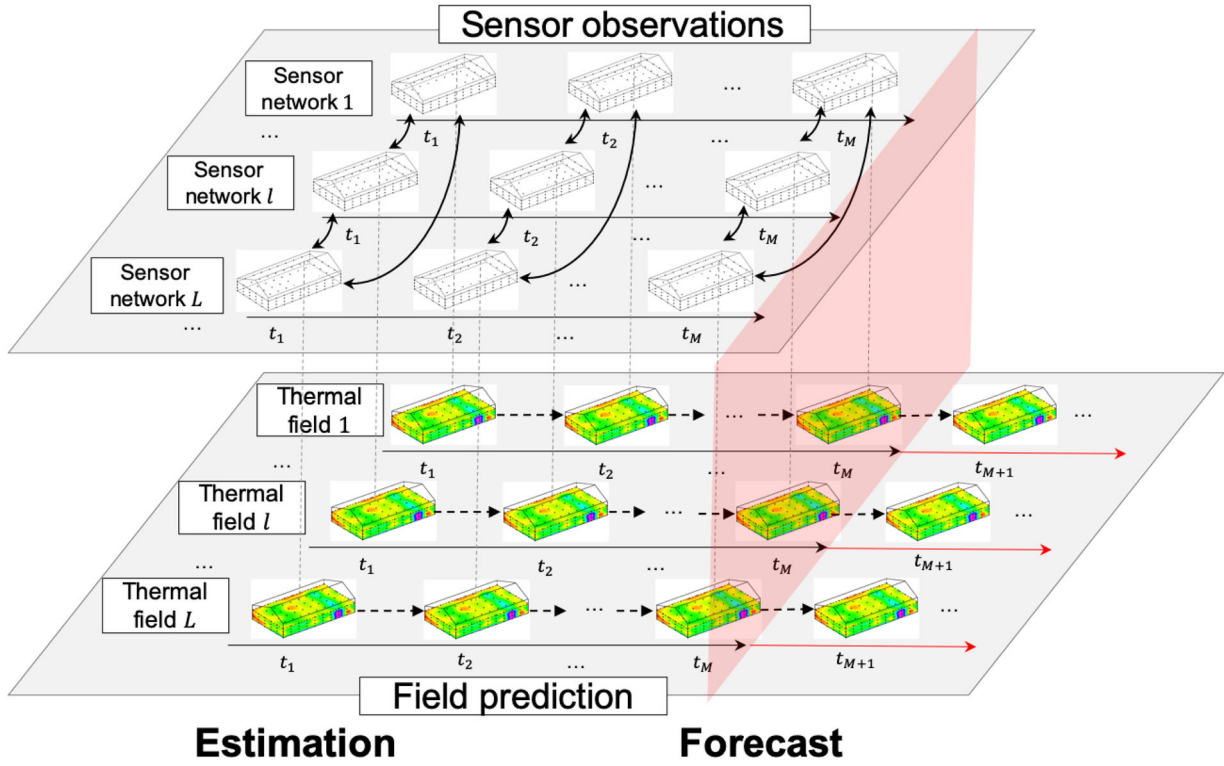


Figure 3. Main idea of the transfer learning-based prediction approach.

thereby demonstrating its tendency to lose the chance to capture local thermal variations. Transfer learning provides an opportunity for thermal field distribution modeling with limited sensor observations by capturing local thermal variations and handling data sparsity and missing data problems. This approach aims to improve the accuracy of the target thermal field by sharing knowledge or information from its homogeneous thermal fields. Multitask learning, which improves model performance of a target task by learning multiple similar-but-not-identical tasks and sharing the information of each task, has emerged as one of the popular focuses in transfer learning. Shao et al. (2017) developed a multitask learning model to predict a 2D machined-surface shape by characterizing the spatial correlation on the basis of the sensor data of related surface shapes. This model mainly focuses on static spatial surfaces but cannot describe spatiotemporal dynamic systems that vary across space and time. For spatiotemporal data modeling, Goncalves, Banerjee, and Von Zuben (2017) proposed a hierarchical multitask learning method to predict climate systems. However, this method can be only applied to 1D spatiotemporal data. Bonilla, Chai, and Williams (2008) proposed a multitask Gaussian process prediction model, which characterizes the spatiotemporal correlation of simple systems by a covariance function in the Gaussian process. However, in complex spatiotemporal systems, spatiotemporal

correlation may not be accurately characterized by only a simple covariance function. Moreover, existing transfer learning approaches model 1D profiles and 2D surfaces, and few studies model 3D dynamic fields. In addition, existing transfer learning approaches mainly focus on field distribution estimation, that is, estimating the values of points of interest off the sensor locations at current time points. Few studies regarding transfer learning have addressed the issue of forecasting a thermal field at future time points.

3. Transfer learning-based field prediction model

In this study, we propose a spatiotemporal TLP approach for a 3D thermal field using limited sensor observations with data sparsity and missing data problems. We develop the TLP model by combining the ideas of field estimation and forecast and integrating transfer learning approach to address the data sparsity and missing data problems to acquire an accurate local thermal variation distribution at current and future time points. The TLP model is substantially an autoregressive model for a temporal sequence of spatial transfer learning processes, which characterizes spatiotemporal correlation of local thermal variations using sensor observations from homogeneous data sources. Specifically, we model the spatial transfer learning

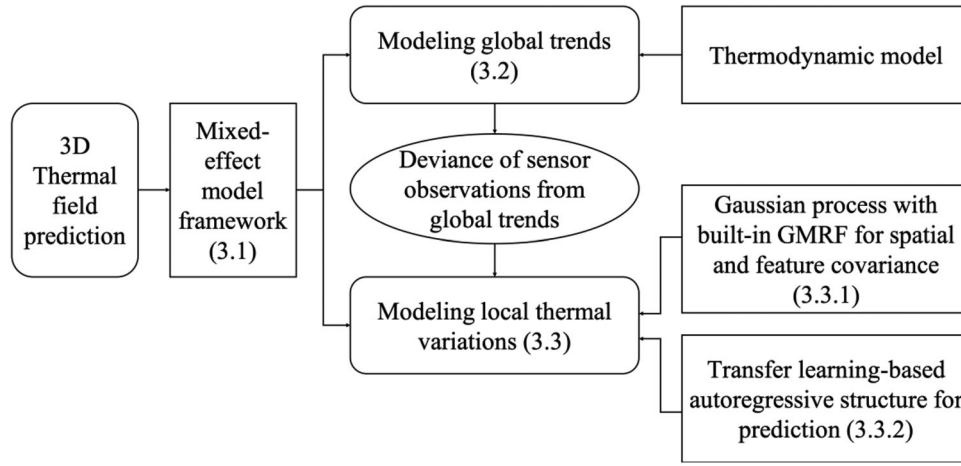


Figure 4. Flowchart of the proposed methodology.

processes by establishing a multitask Gaussian process framework, which is characterized by a covariance matrix that can describe the spatial correlation structure of a field and the feature correlation among fields.

The contributions of the proposed model are presented in the following aspects. First, our model can predict a 3D thermal field distribution at any point of interest and at any current or future time point. Second, our model can address the sensor data sparsity and missing data problems by establishing a multitask Gaussian process framework that fully utilizes limited sensor observations from homogeneous data sources. Third, our model focuses on the modeling of local thermal variations and effectively captures spatiotemporal correlations of the local thermal variations by integrating time series models and spatial transfer learning processes.

As shown in Figure 3, we assume that L homogeneous thermal fields exist, and these fields demonstrate different missing data patterns. We take a thermal field l as an example, and the idea of transfer learning is that we predict the thermal field l by using sensor data in all L thermal fields. The common feature and spatial correlation among thermal fields are captured at each time point from the sensor data in all L thermal fields, and the thermal fields also evolve over time because of temporal correlation. The thermal field prediction includes estimating the entire field distribution at the current time points (e.g., $t_1, \dots, t_M$) and forecasting the future field distribution at the upcoming time points (e.g., t_{M+1}). Figure 4 presents the flowchart of the proposed research methodology. The proposed approach will be introduced in detail as follows.

3.1. Mixed-effect model framework

A thermal field, including the global and local thermal variations, should follow the heat transfer principle.

However, this can only be achieved under the assumption that all the parameters of the heat transfer function can be determined over time. In reality, the parameters are not only affected by environmental factors but also by some intrinsic factors in most cases. Furthermore, they vary over time and could not be directly measured or determined. For example, local thermal variations tend to be triggered by intrinsic factors, such as grain self-breath and local over-heat, and the property of the heat-transfer medium varies over time. This fact could not be directly captured by using the conventional PDE model, which might mainly consider extrinsic factors, such as the thermal field boundary or environmental factors with static parameters. The sensor data collected in the thermal field provide us the opportunity to model local thermal variations using the data-driven method. On this basis, we decompose the modeling of a thermal field into two parts: the global trends and the local thermal variations.

We define a complete set of the thermal field distribution for L fields at time t , with $t = 1, \dots, M$, as $\mathbf{G}_t = (g_t^1(\mathbf{s}_1), \dots, g_t^1(\mathbf{s}_n), \dots, g_t^L(\mathbf{s}_1), \dots, g_t^L(\mathbf{s}_n))^T$, where $g_t^l(\mathbf{s}_i)$ denotes the thermal value at location $\mathbf{s}_i$ and time t of field l ; moreover, there are a total of n locations at each time point. We adopt the following mixed-effect model framework to represent the thermal field distribution for L fields and jointly characterize the global trends and the local thermal variations:

$$\mathbf{G}_t = \boldsymbol{\mu}_t + \mathbf{V}_t, \quad (1)$$

where $\boldsymbol{\mu}_t$ is the set of the global trend distribution for L fields at time t . The residual of $\mathbf{G}_t$ after the global trends $\boldsymbol{\mu}_t$ represents the set of the local thermal distribution for L fields at time t , which is denoted as $\mathbf{V}_t$. We adopt a thermodynamic function to capture the

global trends, which models the main profile of the thermal field, and develop a data-driven method to capture the local thermal variations from the sensor data that can reflect the local status of the thermal field.

3.2. Modeling global trends

In view of the extrinsic factors and thermal mechanism, we adopt a thermodynamic function to characterize the global trends of thermal fields. For the thermal field l ($l = 1, 2, \dots, L$), the thermodynamic function is modeled as the following unsteady heat transmission equation:

$$\frac{\partial}{\partial t} \mu_t^l(\mathbf{s}) - \sum_{d=1}^D \alpha_d \frac{\partial^2}{\partial \mathbf{s}_d^2} \mu_t^l(\mathbf{s}) = 0, \quad (2)$$

where d denotes the index of the space dimension, D is the number of the space dimension, and $\mathbf{s}_d$ represents the space location of dimension d . The parameter α_d of dimension d is determined on the basis of thermal properties. For stored grains, α_d denotes the thermal diffusivity of the grains. We use the FEA method to solve the unsteady heat transmission equation:

$$\mu_{t+1}^l(\mathbf{s}) = \Gamma(\mu_t^l(\mathbf{s})), \quad (3)$$

where we take a 3 D space as an example, and location $\mathbf{s}$ in Eq. (2) is denoted as (s_x, s_y, s_z) in a Cartesian coordinate system:

$$\begin{aligned} \Gamma(\mu_t^l(\mathbf{s})) &= \frac{\Delta t \alpha_x}{\Delta s_x^2} \mu_t^l(s_{x+1}, s_y, s_z) \\ &+ \frac{\Delta t \alpha_x}{\Delta s_x^2} \mu_t^l(s_{x-1}, s_y, s_z) + \frac{\Delta t \alpha_y}{\Delta s_y^2} \mu_t^l(s_x, s_{y+1}, s_z) \\ &+ \frac{\Delta t \alpha_y}{\Delta s_y^2} \mu_t^l(s_x, s_{y-1}, s_z) + \frac{\Delta t \alpha_z}{\Delta s_z^2} \mu_t^l(s_x, s_y, s_{z+1}) \\ &+ \frac{\Delta t \alpha_z}{\Delta s_z^2} \mu_t^l(s_x, s_y, s_{z-1}) \\ &+ \left(1 - \frac{2\Delta t \alpha_x}{\Delta s_x^2} - \frac{2\Delta t \alpha_y}{\Delta s_y^2} - \frac{2\Delta t \alpha_z}{\Delta s_z^2}\right) \mu_t^l(s_x, s_y, s_z). \end{aligned} \quad (4)$$

Here, Δs_x , Δs_y , Δs_z , and Δt are the grid spacings in the x -, y -, and z -directions of the spatial and temporal domains. Given the grain temperature at time t_0 as the initial value and grain temperature on the granary walls as the boundary condition, Eq. (4) can be solved.

3.3. Modeling local thermal variations

Local thermal variations throughout the thermal field distribution of a system usually trigger systematic changes or even system failures, which provides useful information regarding the severity and risks of the thermal system. In comparison with global trends, local thermal variations play a crucial role in engineering systems to provide useful information for the surveillance, maintenance, and improvement of a system. In this section, we develop a TLP approach for acquiring local thermal variation values at any point of interest and at any current or upcoming time point.

Spatiotemporal correlation should be considered in the modeling of local thermal variations to predict a 3 D thermal field distribution accurately. At each time point, local thermal variations at adjacent locations tend to exhibit strong spatial correlation when the distance between their locations is small in a thermal field. Temporal evolution also exists among thermal fields as time goes. In practice, sparse sensor observations are accessible and missing data problem exists, thereby complicating the modeling of a thermal field distribution when only using such sensor observations. Homogeneous thermal fields under similar conditions are known to share a common spatiotemporal structure. This motivates us to consider multisource sensor observations from homogeneous thermal fields for the thermal field prediction to handle the data sparsity and missing data problems effectively.

The local thermal distribution $\mathbf{V}_t$ for L fields at time t is decomposed into two independent parts, namely, the latent function $\mathbf{Z}_t$ of the local thermal distribution for L fields and the noise term δ_t . Thus, we have $\mathbf{V}_t = \mathbf{Z}_t + \delta_t$, where $\mathbf{V}_t$, $\mathbf{Z}_t$, and δ_t are the sets in terms of $v_t^l(\mathbf{s}_i)$, $z_t^l(\mathbf{s}_i)$, and $\delta_t^l(\mathbf{s}_i)$, which denote the local thermal distribution, its corresponding latent function, and the noise at location $\mathbf{s}_i$ and time t of field l , respectively. $\delta_t^l(\mathbf{s}_i)$ is assumed to be an independent and identically distributed Gaussian white noise with the variance $e_{l,t}^2$, that is, $\delta_t^l(\mathbf{s}_i) \sim N(0, e_{l,t}^2)$. Thus, we have $v_t^l(\mathbf{s}_i) \sim N(z_t^l(\mathbf{s}_i), e_{l,t}^2)$.

3.3.1. Gaussian process with built-in GMRF for spatial and feature covariance

Given the essential role of local thermal variations in engineering applications, we remove the global trends from the sensor observations of the thermal field l ($l = 1, \dots, L$) and use the residuals, which are denoted as the deviance of sensor observations $h_t^l(\mathbf{s}_i)$ from global trends $\mu_t^l(\mathbf{s}_i)$ (i.e., $y_t^l(\mathbf{s}_i) = h_t^l(\mathbf{s}_i) + \mu_t^l(\mathbf{s}_i)$) to model local thermal variations. At time t , with

$t = 1, \dots, M$, we assume that the local thermal latent functions of L fields follow a Gaussian process with zero mean. The covariance of the Gaussian process describes the spatial correlation of fields and the feature correlation among fields as follows:

$$\langle z_t^l(\mathbf{s}), z_t^{l'}(\mathbf{s}') \rangle = k_t^z(l, l') k_t^s(\mathbf{s}, \mathbf{s}'). \quad (5)$$

Eq. (5) denotes the covariance of two local thermal latent functions $z_t^l(\mathbf{s})$ at location $\mathbf{s}$ of field l and $z_t^{l'}(\mathbf{s}')$ at location $\mathbf{s}'$ of field l' , where $k_t^z(l, l')$ is a feature covariance that specifies the similarities between fields l and l' , and $k_t^s(\mathbf{s}, \mathbf{s}')$ is a spatial covariance between $\mathbf{s}$ and $\mathbf{s}'$ at time t that characterizes the spatial correlation of the fields. Particularly, for the local thermal latent function in field l at time t , the covariance between locations $\mathbf{s}$ and $\mathbf{s}'$ is proportional to the spatial covariance, that is, $\langle z_t^l(\mathbf{s}), z_t^l(\mathbf{s}') \rangle = c_t^l k_t^s(\mathbf{s}, \mathbf{s}') \propto k_t^s(\mathbf{s}, \mathbf{s}')$, where c_t^l is a non-negative constant in terms of the feature covariance for field l at time t .

3.3.1.1. Modeling spatial covariance. We now introduce the modeling of the spatial covariance for each thermal field. The GMRF has received considerable attention for modeling the spatial dynamic fields in recent years. In a 3D thermal field, sensor observations are collected from grid-based sensor networks, wherein sensors are uniformly distributed in the system; thus, we can fully use the grid-based sensor network structures that correspond to the neighborhood structures in GMRF. However, GMRF can only be used when the sampled data in the entire structured field are complete. Hence, the conventional GMRF method cannot be directly applied to predict a thermal field distribution, wherein field values off the sensor locations are unknown. To address this issue, we propose to employ a Gaussian process model with a built-in GMRF to predict thermal values off the sensor locations by characterizing the spatial covariance at each time point and fully combining the ideas of Gaussian process and GMRF under the grid-based sensor network structure.

We assume $\gamma_t = (\gamma_t(p_1), \gamma_t(p_2), \dots, \gamma_t(p_I))^T$ to be a zero-mean GMRF model in the spatial domain at time t , where p_i denotes the location of point i on the lattice, and I is the number of points on the lattice. We refer to such locations of points on the lattice as generating points, and select the grid-based sensor locations as the generating points in our model. At time t , the local thermal latent function $z_t^l(\mathbf{s})$ of field l with $l = 1, \dots, L$ is modeled by a Gaussian process with a built-in GMRF, which is defined as follows:

$$z_t^l(\mathbf{s}) = \sum_{i=1}^I \sqrt{c_t^l} \lambda(\mathbf{s}, p_i) \gamma_t(p_i) = \sqrt{c_t^l} \lambda_s \gamma_t, \quad (6)$$

where $\lambda(\mathbf{s}, p_i)$ is a weighting function for location $\mathbf{s}$ and the generating points p_i , and $\lambda_s = (\lambda(\mathbf{s}, p_1), \lambda(\mathbf{s}, p_2), \dots, \lambda(\mathbf{s}, p_I))$.

For the variable $\gamma_t(p_i)$, with $i = 1, \dots, I$, we consider a spatial domain of the generating points $S = \{1, \dots, I\}$ in the GMRF model, and the neighborhood of p_i is defined as $\mathcal{N}_i = \{j \in S : j \text{ is a neighbor of } i, j \neq i\}$. Given $\{\gamma_t(p_j), j \in \mathcal{N}_i\}$, the conditional distribution of $\gamma_t(p_i)$ is a Gaussian, which can be expressed as follows:

$$\gamma_t(p_i) | \{\gamma_t(p_j), j \neq i, j \in \mathcal{N}_i\} \sim N\left(\rho_t \sum_{j \neq i, j \in \mathcal{N}_i} \beta_{i,j} \gamma_t(p_j), \sigma_{i,t}^2\right), \quad (7)$$

where ρ_t is the overall effect of spatial dependence at time t , $\sigma_{i,t}^2$ is the variance of p_i at time t , $\beta_{i,j}$ is a normalized adjacency parameter of p_i and p_j , $\beta_{i,i} = 0$, and $\beta_{i,j} = \beta_{j,i}$. We denote $\beta_{i,j} = w_{i,j}/w_{i+}$, where

$$w_{i,j} = \begin{cases} \phi(p_i, p_j), & \text{for } j \in \mathcal{N}_i \text{ and } j \neq i, \\ 0, & \text{otherwise.} \end{cases}$$

Here, $w_{i+} = \sum_{j=1}^{n_i} w_{i,j}$, where n_i is the number of i 's neighbors, and ϕ is a non-increasing function of the distance between p_i and p_j . We denote a systematic adjacency matrix $\mathbf{W} = \{w_{i,j}\}_{i,j=1}^I$, $\mathbf{W}_D = \text{diag}\{w_{1+}, w_{2+}, \dots, w_{I+}\}$, and $\sigma_{i,t}^2 = \sigma_t^2 w_{i+}^{-1}$. Then, we obtain the joint probability density function of γ_t as follows:

$$\gamma_t \sim N\left(0, \left[\frac{1}{\sigma_t^2} (\mathbf{W}_D - \rho_t \mathbf{W})\right]^{-1}\right), \quad (8)$$

where σ_t^2 is the overall variance at time t . The inverse covariance matrix (precision matrix) of γ_t is $\mathbf{Q}_t = \frac{1}{\sigma_t^2} (\mathbf{W}_D - \rho_t \mathbf{W})$, and thus $\mathbf{Q}_t$ is a symmetric positive definite matrix.

Therefore, by inserting Eq. (8) into Eq. (6), we obtain the distribution of $z_t^l(\mathbf{s})$, which is also a Gaussian, as follows:

$$z_t^l(\mathbf{s}) \sim N(0, c_t^l \lambda_s \mathbf{Q}_t^{-1} \lambda_s^T). \quad (9)$$

3.3.1.2. Local thermal distribution with spatial and feature. We consider the set of local thermal distribution $\mathbf{V}_t$ for L fields at time t , with $t = 1, \dots, M$. We denote $\lambda \in \mathbb{R}^{n \times I}$ to be a matrix obtained by $(\lambda)_{ij} = \lambda(\mathbf{s}_i, p_j)$, where $\mathbf{s}_i$ denotes the location of the point of interest i . Then, we acquire the distribution of $\mathbf{V}_t$ as a Gaussian, as shown as follows:

$$\mathbf{V}_t \sim N(0, \mathbf{K}_t^z \otimes \mathbf{K}_t^s + \mathbf{E}_t \otimes \mathbf{I}_n), \quad (10)$$

where $\mathbf{K}_t^z = \{k^z(l_1, l_2)\}_{l_1, l_2=1}^L$ is an $L \times L$ matrix of the feature covariance among L fields; $\mathbf{K}_t^s = \lambda \mathbf{Q}_t^{-1} \lambda^T$ is an $n \times n$ matrix of the spatial covariance; $\mathbf{E}_t$ denotes an $L \times L$ diagonal matrix in terms of the noise variance in which the l th diagonal element is $\varepsilon_{l,t}^2$, that is, $\mathbf{E}_t = \text{diag}\{\varepsilon_{1,t}^2, \varepsilon_{2,t}^2, \dots, \varepsilon_{L,t}^2\}$; $\mathbf{I}_n$ is an $n \times n$ identity matrix; and $\otimes$ denotes the Kronecker product.

3.3.2. Transfer learning-based autoregressive structure for prediction

On the basis of Eq. (10), the distribution of $\mathbf{V}_t$ is a Gaussian at time t with $t = 1, \dots, M$. We denote $\mathbf{V}_t \sim N(0, \mathbf{D}_t)$, where $\mathbf{D}_t = \mathbf{K}_t^z \otimes \mathbf{K}_t^s + \mathbf{E}_t \otimes \mathbf{I}_n$, and $t = 1, \dots, M$. For ease of exposition, we illustrate our proposed model by considering $\mathbf{V}_{t-1}$ and $\mathbf{V}_t$ as the temporal sequence of two Gaussian models, namely, $\mathbf{V}_{t-1} \sim N(0, \mathbf{D}_{t-1})$ and $\mathbf{V}_t \sim N(0, \mathbf{D}_t)$, respectively. Then, the vector of $(\mathbf{V}_{t-1}, \mathbf{V}_t)^T$ follows a multivariate Gaussian distribution:

$$\begin{pmatrix} \mathbf{V}_{t-1} \\ \mathbf{V}_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{t-1,t-1} & \Sigma_{t-1,t} \\ \Sigma_{t,t-1} & \Sigma_{t,t} \end{pmatrix}\right), \quad (11)$$

where $\Sigma_{t,t-1} = \Sigma_{t-1,t}^T$, $\Sigma_{t-1,t-1} = \mathbf{D}_{t-1}$, and $\Sigma_{t,t} = \mathbf{D}_t$. The joint probability density function can be obtained as the product of the conditional and marginal probability density functions, $f(\mathbf{V}_{t-1}, \mathbf{V}_t) = f(\mathbf{V}_t | \mathbf{V}_{t-1})f(\mathbf{V}_{t-1})$. In particular, we have the following equations based on the properties of the multivariate Gaussian distribution:

$$\mathbf{E}(\mathbf{V}_t | \mathbf{V}_{t-1}) = \Sigma_{t,t-1} \Sigma_{t-1,t-1}^{-1} \mathbf{V}_{t-1},$$

$$\text{Var}(\mathbf{V}_t | \mathbf{V}_{t-1}) = \Sigma_{t,t} - \Sigma_{t,t-1} \Sigma_{t-1,t-1}^{-1} \Sigma_{t-1,t}.$$

Let $\mathbf{C}_{t-1} = \Sigma_{t,t-1} \Sigma_{t-1,t-1}^{-1}$ and $\Sigma_{t,t} - \Sigma_{t,t-1} \Sigma_{t-1,t-1}^{-1} \Sigma_{t-1,t} = \Sigma_{t,t|t-1}$. Eq. (11) can be written as

$$\begin{pmatrix} \mathbf{V}_{t-1} \\ \mathbf{V}_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{t-1,t-1} & \Sigma_{t-1,t} \\ \Sigma_{t,t-1} & \Sigma_{t,t|t-1} \end{pmatrix}\right),$$

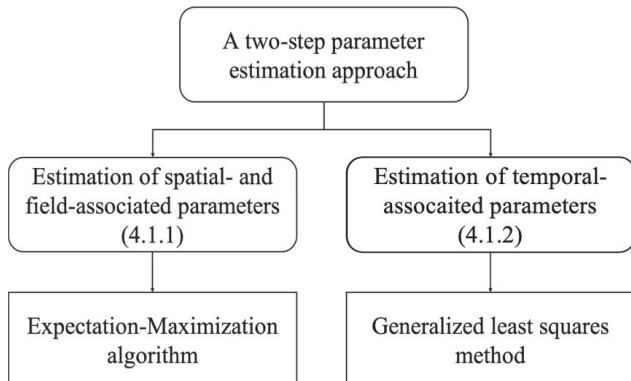


Figure 5. Two-step parameter estimation procedure.

$$\begin{pmatrix} \Sigma_{t-1,t-1} & (\mathbf{C}_{t-1} \Sigma_{t-1,t-1})^T \\ \mathbf{C}_{t-1} \Sigma_{t-1,t-1} & \Sigma_{t,t|t-1} + \mathbf{C}_{t-1} \Sigma_{t-1,t-1} \mathbf{C}_{t-1}^T \end{pmatrix}. \quad (12)$$

In this way, we propose the spatiotemporal TLP model, where $\mathbf{V}_t$ depends linearly on itself $\mathbf{V}_{t-1}$ lagged by one time unit as follows:

$$\mathbf{V}_t = \mathbf{C}_{t-1} \mathbf{V}_{t-1} + \boldsymbol{\eta}_t. \quad (13)$$

In Eq. (13), we define $\mathbf{C}_{t-1}$ as a spatiotemporal autoregressive matrix as follows:

$$\mathbf{C}_{t-1} = r_1 \mathbf{D}_t \mathbf{D}_{t-1}^{-1},$$

where r_1 is the temporal-associated parameter at time $t - 1$. Let $\boldsymbol{\eta}_t \sim (0, \mathbf{K}_t^z \otimes \mathbf{K}_t^s + \mathbf{E}_t \otimes \mathbf{I}_n)$ be a pseudo error process with spatiotemporal and feature covariance coefficients. At this point, the generalization of the model in Eq. (13) is straightforward. Given a temporal sequence of local thermal variations $\{\mathbf{V}_{t-1}, \dots, \mathbf{V}_{t-Q}\}$, we propose the general expression of the spatiotemporal TLP model for $\mathbf{V}_t$ as follows:

$$\mathbf{V}_t = \sum_{q=1}^Q r_q \mathbf{D}_t \mathbf{D}_{t-q}^{-1} \mathbf{V}_{t-q} + \boldsymbol{\eta}_t, \quad (14)$$

where r_q is the temporal-associated parameter at time $t - q$, and Q is the order of the TLP model. With $q = 1, \dots, Q$, we use a temporal-associated parameter r_q , spatial matrices $\mathbf{D}_t \mathbf{D}_{t-q}^{-1}$, and the product of a temporal-associated parameter and spatial matrices $r_q \mathbf{D}_t \mathbf{D}_{t-q}^{-1}$ to characterize the local thermal variations in the same location at different time, those in nearby locations at the same time point, and those in nearby locations at different time points, respectively.

4. Parameter estimation and field prediction

4.1. Parameter estimation

The proposed model contains a set of parameters to be estimated, including the feature covariance matrix $\mathbf{K}_t^z$ at time t , the overall variance σ_t^2 at time t , the overall effect of spatial dependence ρ_t at time t , the noise variance $\varepsilon_{l,t}^2$ of field l at time t , and the temporal-associated parameter $\mathbf{r} = \{r_1, r_2, \dots, r_Q\}$. As shown in Figure 5, we develop a two-step approach to estimate these unknown parameters. We initially estimate the spatial-associated parameters σ_t^2 and ρ_t , and field-associated parameters $\mathbf{K}_t^z$ and $\varepsilon_{l,t}^2$ at each time $t (t = 1, \dots, M)$ using an expectation-maximization (EM) algorithm by excluding the temporal dependence. Afterward, we estimate the temporal-associated parameter $\mathbf{r}$ using the generalized least squares method by excluding the spatial dependence.

We denote the set of the deviance of sensor observations from the global trends of L fields as $\mathbf{Y} = (y_{1,1}^1, \dots, y_{1,N_{1,1}}^1, \dots, y_{M,1}^1, \dots, y_{M,N_{M,1}}^1, \dots, y_{t,i}^l, \dots, y_{1,1}^L, \dots, y_{1,N_{1,L}}^L, \dots, y_{M,1}^L, \dots, y_{M,N_{M,L}}^L)$, where $y_{t,i}^l$ denotes the deviance of field l at location $\mathbf{s}_i$ and time t , and M time points and $N_{t,l}$ locations exist at time t ($t = 1, \dots, M$) in field l ($l = 1, \dots, L$). We initially interpolate the missing data of each thermal field l using the Gaussian process model in Eq. (6). Then, we obtain a processed dataset of L fields $\tilde{\mathbf{Y}}$, where M time points and N locations are found at each time point in each field.

4.1.1. Estimation of spatial- and field-associated parameters

Notably, our proposed model is reduced to the Gaussian process model in Eq. (10) by excluding the temporal dependence, that is, $r_1 = r_2 = \dots = r_Q = 0$. Let $\Lambda \in \mathbb{R}^{N \times I}$ be a matrix obtained by $(\Lambda)_{ij} = \lambda(\tilde{\mathbf{s}}_i, p_j)$, where $\tilde{\mathbf{s}}_i$ is the location of the i th processed data of each field. Afterward, the spatial covariance matrix in terms of the processed data of each field is denoted as $\tilde{\mathbf{K}}_t^s = \Lambda \mathbf{Q}_t^{-1} \Lambda^T$. We apply an EM algorithm to estimate the spatial- and field-associated parameters $\boldsymbol{\theta}_t$, $\mathbf{K}_t^z$ and $\varepsilon_{t,t}^2$. The log-likelihood function in terms of these parameters is

$$\begin{aligned} l(\mathbf{K}_t^z, \boldsymbol{\theta}_t, \varepsilon_{t,t}^2) = & -\frac{N}{2} \log|\mathbf{K}_t^z| - \frac{L}{2} \log|\tilde{\mathbf{K}}_t^s| \\ & - \frac{1}{2} \text{tr}[(\mathbf{K}_t^z)^{-1} \mathbf{Z}_t^T (\tilde{\mathbf{K}}_t^s)^{-1} \mathbf{Z}_t] \\ & - \frac{N}{2} \sum_{l=1}^L \log \varepsilon_{t,t}^2 \\ & - \frac{1}{2} \text{tr}[(\tilde{\mathbf{Y}}_t - \mathbf{Z}_t) \mathbf{E}_t^{-1} (\tilde{\mathbf{Y}}_t - \mathbf{Z}_t)^T] \\ & - \frac{NL}{2} \log 2\pi, \end{aligned} \quad (15)$$

where $\boldsymbol{\theta}_t = \{\sigma_t^2, \rho_t\}$, $\tilde{\mathbf{Y}}_t$ denotes an $L \times N$ matrix in terms of the processed data of L fields at time t , and $\mathbf{Z}_t$ indicates the expectation of $\tilde{\mathbf{Y}}_t$. $\mathbf{Z}_t$ is obtained from $p(\mathbf{z}_t | \tilde{\mathbf{y}}_t, \mathbf{K}_t^z, \boldsymbol{\theta}_t) = N((\mathbf{K}_t^z \otimes \tilde{\mathbf{K}}_t^s) \mathbf{K}_t^{-1} \tilde{\mathbf{y}}_t, (\mathbf{K}_t^z \otimes \tilde{\mathbf{K}}_t^s) - (\mathbf{K}_t^z \otimes \tilde{\mathbf{K}}_t^s) \mathbf{K}_t^{-1} (\mathbf{K}_t^z \otimes \tilde{\mathbf{K}}_t^s))$, where $\tilde{\mathbf{y}}_t = (\tilde{\mathbf{y}}_t^1, \dots, \tilde{\mathbf{y}}_t^L, \dots, \tilde{\mathbf{y}}_t^L)^T$ and $\tilde{\mathbf{y}}_t^l = (\tilde{y}_{t,1}^l, \dots, \tilde{y}_{t,N}^l)^T$ is the processed data of field l at time t ; $\mathbf{K}_t = \mathbf{K}_t^z \otimes \tilde{\mathbf{K}}_t^s + \mathbf{E}_t \otimes \mathbf{I}_N$; and $\mathbf{I}_N$ is an $N \times N$ identity matrix. The EM algorithm has the following updates on the parameters $\boldsymbol{\theta}_t$, $\mathbf{K}_t^z$ and $\varepsilon_{t,t}^2$:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_t = & \text{argmin}_{\boldsymbol{\theta}_t} \\ & \left(N \log \left| \mathbf{Z}_t^T (\tilde{\mathbf{K}}_t^s(\boldsymbol{\theta}_t))^{-1} \mathbf{Z}_t \right| + L \log |\tilde{\mathbf{K}}_t^s(\boldsymbol{\theta}_t)| \right), \end{aligned} \quad (16)$$

$$\hat{\mathbf{K}}_t^z = \frac{1}{N} \mathbf{Z}_t^T (\tilde{\mathbf{K}}_t^s(\hat{\boldsymbol{\theta}}_t))^{-1} \mathbf{Z}_t, \quad (17)$$

$$\hat{\varepsilon}_{t,t}^2 = \frac{1}{N} (\tilde{\mathbf{y}}_t^l - \mathbf{z}_t^l)^T (\tilde{\mathbf{y}}_t^l - \mathbf{z}_t^l), \quad (18)$$

where $\hat{\boldsymbol{\theta}}_t$, $\hat{\mathbf{K}}_t^z$ and $\hat{\varepsilon}_{t,t}^2$ denote the estimated $\boldsymbol{\theta}_t$, $\mathbf{K}_t^z$ and $\varepsilon_{t,t}^2$, and $\mathbf{z}_t^l$ is the expectation of $\tilde{\mathbf{y}}_t^l$. In Eq. (16), $\hat{\boldsymbol{\theta}}_t$ is obtained by minimizing the log-likelihood function $l(\boldsymbol{\theta}_t) = N \log |\mathbf{Z}_t^T (\tilde{\mathbf{K}}_t^s(\boldsymbol{\theta}_t))^{-1} \mathbf{Z}_t| + L \log |\tilde{\mathbf{K}}_t^s(\boldsymbol{\theta}_t)|$. The detailed parameter estimation algorithm of $\boldsymbol{\theta}_t$ is introduced in Appendix A.

4.1.2. Estimation of temporal-associated parameters

Estimating the temporal-associated parameters $\mathbf{r} = \{r_1, r_2, \dots, r_Q\}$ is feasible after substituting the estimated spatial-associated parameter $\hat{\sigma}_t^2$ and field-associated parameters $\hat{\mathbf{K}}_t^z$ and $\hat{\varepsilon}_{t,t}^2$ into Eq. (14). Our proposed model is reduced to a linear model by excluding the spatial dependence, that is, $\rho_t = 0$ for all t :

$$\mathbf{V}_t = \sum_{q=1}^Q r_q \mathbf{U}_{t-q} + \boldsymbol{\eta}_t, \quad (19)$$

where $\mathbf{U}_{t-q} = \mathbf{C}_{t,q} \mathbf{V}_{t-q}$, $\mathbf{C}_{t,q} = \hat{\mathbf{D}}_t \hat{\mathbf{D}}_{t-q}^{-1}$ is a constant matrix, and $\boldsymbol{\eta}_t \sim N(0, \hat{\sigma}_t^2 \hat{\mathbf{K}}_t^z \otimes (\lambda \mathbf{W}_D^{-1} \lambda^T))$. $\mathbf{r}$ can be estimated using the generalized least squares method, which is introduced in Appendix B in details.

4.2. Field prediction

Given the sensor observations of L fields at time t ($t = 1, \dots, M$), we obtain the processed dataset of deviance $\tilde{\mathbf{Y}}$. At the current time points, we predict the points of interest off the sensor locations by $\hat{\mathbf{V}}_t = (\mathbf{K}_t^z \otimes \tilde{\mathbf{K}}_t^s)^T \mathbf{K}_t^{-1} \tilde{\mathbf{y}}_t^l$ to acquire the local thermal distribution at each time point t ($t = 1, \dots, M$). The thermal field distribution of L fields is then obtained by combining the global trend and the predicted local thermal distribution using Eq. (1).

At future time points, we forecast $\hat{\mathbf{V}}_{t+1}$ by using the conditional distribution of $\mathbf{V}_{M+1} | (\hat{\mathbf{V}}_M, \dots, \hat{\mathbf{V}}_{M-Q+1})$ at time $M+1$, which can be expressed as follows:

$$\begin{aligned} & \mathbf{V}_{M+1} | (\hat{\mathbf{V}}_M, \dots, \hat{\mathbf{V}}_{M-Q+1}) \\ & \sim N \left(\sum_{q=1}^Q r_q \mathbf{D}_{M+1} \mathbf{D}_{M-q+1}^{-1} \hat{\mathbf{V}}_{M-q+1}, \right. \\ & \quad \left. \mathbf{K}_{M+1}^z \otimes \mathbf{K}_{M+1}^s + \mathbf{E}_{M+1} \otimes \mathbf{I}_n \right), \end{aligned} \quad (20)$$

where we initially denote $\hat{\sigma}_{M+1}^2 = \hat{\sigma}_M^2$, $\hat{\rho}_{M+1} = \hat{\rho}_M$, $\hat{\mathbf{K}}_{M+1}^z = \hat{\mathbf{K}}_M^z$, and $\hat{\varepsilon}_{l,M+1}^2 = \hat{\varepsilon}_{l,M}^2$ because the spatial-

and field-associated parameters are similar at two consecutive time points. After obtaining $\hat{\mathbf{V}}_{M+1}$, we estimate the spatial- and field-associated parameters at time $M+1$ to update the current values. Then, the thermal field distribution of L fields at time $M+1$ can be acquired by

$$\hat{\mathbf{G}}_{M+1} = \hat{\boldsymbol{\mu}}_{M+1} + \hat{\mathbf{V}}_{M+1},$$

where the global trend $\hat{\boldsymbol{\mu}}_{M+1}$ is obtained by Eq. (3). The thermal field distribution at time $M+2$, $M+3$, and other future times can be forecasted similarly.

In summary, we acquire the thermal field distribution of L fields at current and future time points. If we regard thermal field l^* , which is one of the L homogeneous fields, as the target task, then we can directly obtain its distribution from the thermal field distribution of L fields using our proposed approach. In addition, we provide the computational cost of our proposed approach in Big-O notation and add the illustration to Appendix C.

5. Real case study

5.1. Dataset description

We conducted a real case study of grain thermal fields during storage to validate the performance of our proposed model. We selected two adjacent conventional cuboidal granaries comprising the same volume and stored the same type of grains in a national grain depot located in Central China. The thermal sensor observations of each granary were collected from grid-based sensor networks (Figure 1). In Figure 1, 240 evenly spaced thermal sensors were distributed in the granary with the length, width, and height of 46, 26, and 6 m, respectively. The sensors were deployed at the initial positions of 0.5, 0.5, and 0.3 m in the x -, y -, and z -directions, respectively, and located at 5 m intervals in the x - and y -directions and 1.8 m intervals in the z -direction.

Table 1. Estimated values of ρ_t and σ_t^2 .

Date	1	8	15	22	29	36	43	50	57	64
ρ_t	5.546	5.504	5.505	5.519	5.484	5.491	5.506	5.527	5.546	5.519
σ_t^2	17.143	17.155	15.657	17.542	16.296	15.677	16.254	15.193	17.001	17.540

Table 2. Estimated values of $\mathbf{K}_t^z$.

Date	1	8	15	22	29
$\mathbf{K}_t^z$	$\begin{bmatrix} 1.466 & 1.457 \\ 1.457 & 1.584 \end{bmatrix}$	$\begin{bmatrix} 1.585 & 1.498 \\ 1.498 & 1.455 \end{bmatrix}$	$\begin{bmatrix} 2.391 & 2.264 \\ 2.264 & 2.217 \end{bmatrix}$	$\begin{bmatrix} 2.841 & 2.861 \\ 2.861 & 2.913 \end{bmatrix}$	$\begin{bmatrix} 3.298 & 3.456 \\ 3.456 & 3.702 \end{bmatrix}$
Date	36	43	50	57	64
$\mathbf{K}_t^z$	$\begin{bmatrix} 3.398 & 3.356 \\ 3.356 & 3.465 \end{bmatrix}$	$\begin{bmatrix} 4.194 & 3.952 \\ 3.952 & 3.772 \end{bmatrix}$	$\begin{bmatrix} 4.957 & 4.912 \\ 4.912 & 4.918 \end{bmatrix}$	$\begin{bmatrix} 4.328 & 4.408 \\ 4.408 & 4.593 \end{bmatrix}$	$\begin{bmatrix} 4.606 & 4.658 \\ 4.658 & 4.767 \end{bmatrix}$

We selected a sensor dataset that was synchronously collected every 7 days, and a total of 13 time points (91 days) were set. At the first 10 time points, we selected sensor data once every other point in the x -, y -, and z -directions at each time as the training dataset and set the remaining sensor data to the test dataset to be estimated. At the last three time points, all the sensor data were set to the test dataset to be forecasted.

We adopted an estimation and a τ -step-ahead forecast ($\tau = 1, 2, 3$) to obtain the grain thermal field distribution of the target granary. We considered the root mean square errors (RMSEs) between the test data and the estimated values and the forecast values of the target granary, as shown as follows:

$$\begin{aligned} \text{RMSE - Estimation} &= \sqrt{\frac{1}{NM} \sum_{t=1}^M \sum_{i=1}^{N_E} \left(g_t^{l^*}(\mathbf{s}_i) - \hat{g}_t^{l^*}(\mathbf{s}_i) \right)^2}, \\ \text{RMSE - Forecast} &= \sqrt{\frac{1}{N\tau} \sum_{t=M+1}^{M+\tau} \sum_{i=1}^{N_F} \left(g_t^{l^*}(\mathbf{s}_i) - \hat{g}_t^{l^*}(\mathbf{s}_i) \right)^2}, \end{aligned} \quad (21)$$

where $g_t^{l^*}(\mathbf{s}_i)$ is the observation in the test dataset of the target granary at location $\mathbf{s}_i$ and time t , and $\hat{g}_t^{l^*}(\mathbf{s}_i)$ is the corresponding thermal value acquired by the proposed model; N_E and N_F denote the number of test data at each time point for estimation and forecast, respectively. Here, RMSE - Estimation is used to evaluate the model estimation performance for the points of interest off the sensor locations at the current time points. Meanwhile, RMSE - Forecast is used to evaluate the model forecast performance for the entire thermal field distribution (including points on and off the sensor locations) at the near future time points.

5.2. Parameter estimation

For the mean function, the static parameters in the thermodynamic model were determined by physical

Table 3. Estimated temporal-associated parameter and its corresponding RMSE – Forecast with different orders.

Order	Estimated temporal associated parameter $\hat{\mathbf{r}}$	RMSE – Forecast
1	$\hat{\mathbf{r}} = (1.0128)$	0.0708
2	$\hat{\mathbf{r}} = (0.7649, 0.2567)$	0.0704
3	$\hat{\mathbf{r}} = (0.7121, 0.3243, 0.0132)$	0.0719

properties of the stored grains. Mixed wheat was stored in the granaries. The thermal diffusivity of the grains in the x -, y -, and z -directions were 5.56, 5.55, and $1.50 (\times 10^{-8} \text{m}^2/\text{s})$, respectively. We acquired the sensor data of the granary walls collected by the sensors on the walls and adopted a Gaussian process model (Rasmussen 2004) for interpolation using the sensor data of the granary walls. In this manner, we obtained the boundary conditions of the thermodynamic model. Then, we acquired the global trends of the two homogeneous granaries using Eqs. (2)–(4).

We calculated the deviance of the training data from the global trends for each granary to model the local thermal variations. The spatial-associated parameters σ_t^2 and ρ_t and field-associated parameters $\mathbf{K}_t^z$ and $\varepsilon_{t,t}^2$ at each time $t (t = 1, \dots, M)$ were estimated using the EM algorithm described in Eqs. (15)–(19). A unique feature in grain thermal fields was the various spatial correlation patterns of local thermal distribution in different directions of the granary. Thus, we considered an anisotropic spatial-correlated weighting function to characterize the spatial correlation between a location in the granary and the generating points; that is,

$$\lambda(\mathbf{s}, \mathbf{p}) = \exp \left\{ - \left[\left(\frac{s_x - p_x}{\delta_x} \right)^2 + \left(\frac{s_y - p_y}{\delta_y} \right)^2 + \left(\frac{s_z - p_z}{\delta_z} \right)^2 \right] \right\},$$

where $\mathbf{s}$ denotes a location in the granary; $\mathbf{p}$ indicates the location of a generating point; s_x , s_y , and s_z correspond to the coordinates of $\mathbf{s}$ in the x -, y -, and z -directions, respectively; p_x , p_y , and p_z are the coordinates of $\mathbf{p}$ in the x -, y -, and z -directions, respectively; and δ_x , δ_y , and δ_z are the range parameters in the x -, y -, and z -directions, respectively; these range parameters correspond to the intervals between two adjacent sensor locations. We defined $\phi(p_i, p_j)$ similarly to λ to characterize the weights in the neighborhood structures of the GMRF.

Table 4. RMSE – Estimation values in the six groups.

Group	S1-A	S1-B	S1-C	S2-A	S2-B	S2-C
Mean of RMSE – Estimation	0.2416	0.2615	0.2818	0.2411	0.2613	0.2812
Standard deviation of RMSE – Estimation	0.0060	0.0066	0.0069	0.0156	0.0164	0.0169

Tables 1 and 2 list the estimated spatial-associated parameters ρ_t and σ_t^2 and feature covariance matrix of fields $\mathbf{K}_t^z$ using the training dataset. In Table 1, the ρ_t values, which characterize the overall effect of spatial dependence at time t , remain steady during the first 10 time points. The results are consistent with the actual conditions of grain storage, that is, the spatial dependence structure slowly changes over time. The σ_t^2 values, which characterize the overall variance at time t , do not dramatically change over the first 10 time points, thereby indicating that the grain storage is under favorable condition. In Table 2, the feature covariance matrix $\mathbf{K}_t^z$ captures the correlation between the target thermal field and its homogeneous thermal field. The target thermal field and its homogeneous thermal field are highly correlated at each time point. Then, we estimated the temporal-associated parameter $\mathbf{r}$ using the generalized least squares method. As shown in Table 3, we estimated the temporal-associated parameter $\hat{\mathbf{r}}$ with different orders $q = 1, 2, 3$ and obtained the corresponding RMSE – Forecast using the local thermal variation data in the training dataset. That is, we forecasted the local thermal distribution at time $t = Q + 1$ using the training data at $t = 1, \dots, Q$. Similarly, we acquired the forecast values of the local thermal distribution at time $t = Q + 2, \dots, M$. Then, we calculated the RMSE – Forecast using the forecast values and the actual values at time $t = Q + 1, \dots, M$. We selected the order $q = 2$, which had the minimal RMSE – Forecast value. The order q can also be selected by some model selection criteria such as AIC or BIC.

5.3. Model performance

5.3.1. Missing data generation

To validate the model performance of field prediction when data are missing, we generated two missing data situations S1 and S2 because the missing data scenario usually emerges due to two aspects, namely, sensor failure and data reading errors.

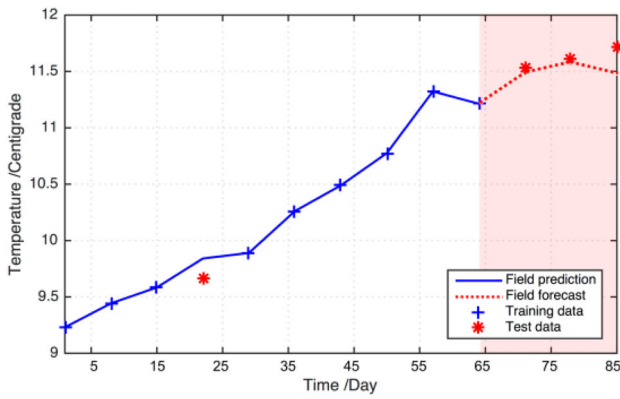
1. Situation S1: The missing data problem was caused by the failure of some sensors. Consequently, sensor observations were not gathered at fault sensor locations continuously until the fault sensors were repaired. To generate the missing data scenario caused by sensor failure, we

Table 5. RMSE — Forecast values in the six groups.

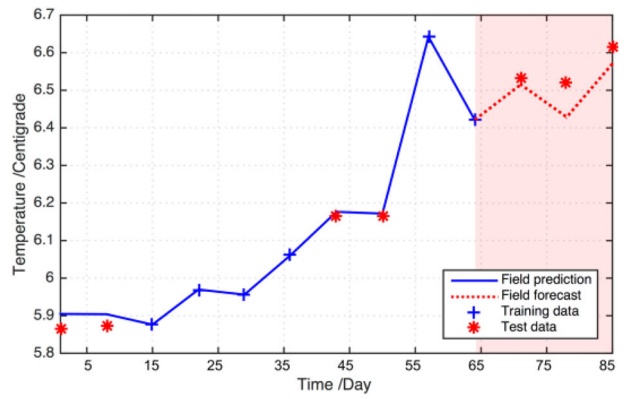
Group	Date	Mean of RMSE — Forecast	Standard deviation of RMSE — Forecast	Group	Date	Mean of RMSE — Forecast	Standard deviation of RMSE — Forecast
S1-A	Day 71	0.2428	0.0087	S2-A	Day 71	0.2482	0.0114
	Day 78	0.2515	0.0116		Day 78	0.2510	0.0116
	Day 85	0.2680	0.0137		Day 85	0.2669	0.0119
S1-B	Day 71	0.2514	0.0102	S2-B	Day 71	0.2482	0.0126
	Day 78	0.2592	0.0139		Day 78	0.2587	0.0137
	Day 85	0.2766	0.0183		Day 85	0.2761	0.0157
S1-C	Day 71	0.2610	0.0115	S2-C	Day 71	0.2645	0.0142
	Day 78	0.2655	0.0153		Day 78	0.2680	0.0147
	Day 85	0.2884	0.0242		Day 85	0.2880	0.0215

Table 6. Computational time of the proposed model.

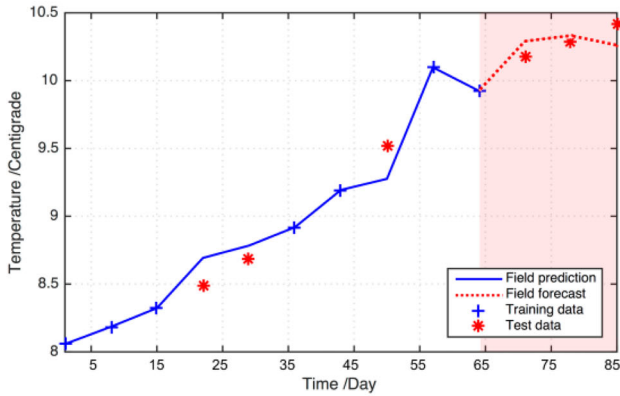
Group	S1-A	S1-B	S1-C	S2-A	S2-B	S2-C
Computational time /s	51.27	51.41	51.44	51.24	51.31	51.47



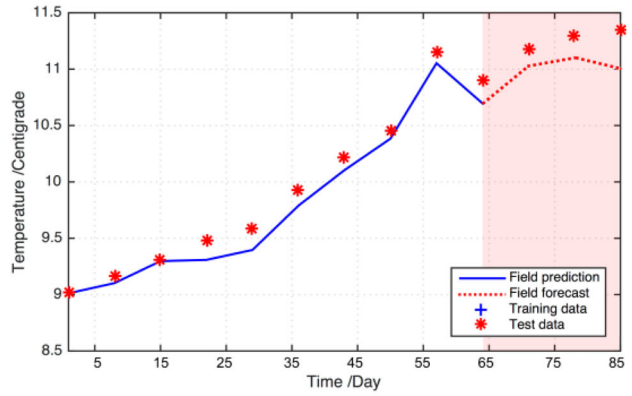
(a)



(b)



(c)



(d)

Figure 6. Examples of the thermal profiles at different locations acquired by our proposed model using the training data of Group S1-A. (Example coordinates: (a) (5.5, 15.5, 3.9)m; (b) (15.5, 40.5, 2.1)m; (c) (15.5, 25.5, 3.9)m; (d) (10.5, 15.5, 3.9)m).

randomly selected sensor locations from the training dataset and assumed that the sensor data at the selected sensor locations were missing during the entire period. We considered three groups of data with different levels of missing data to evaluate the proposed model performance, that is, the proportions of missing data to the entire training

data were set as 30% (S1-A), 40% (S1-B), and 50% (S1-C).

2. Situation S2: The missing data problem was caused by data reading errors that could occur at any location and time point. To generate the missing data scenario caused by data reading errors, we randomly selected data at any location

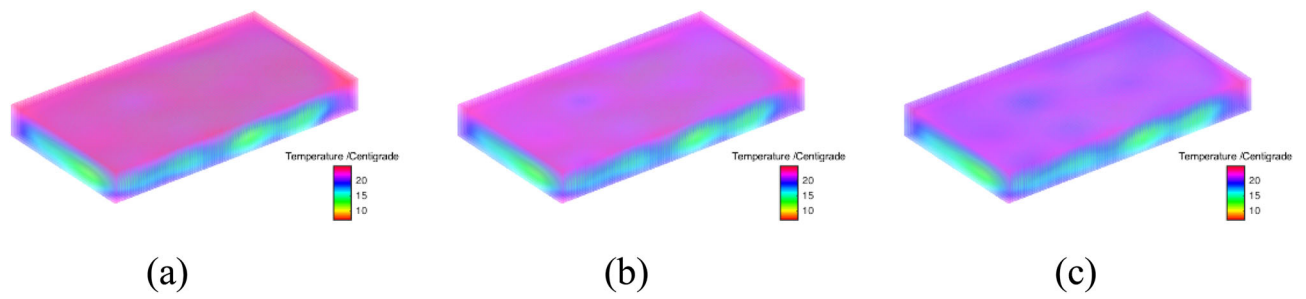


Figure 7. Forecast thermal field distributions on days 71, 78, and 85.

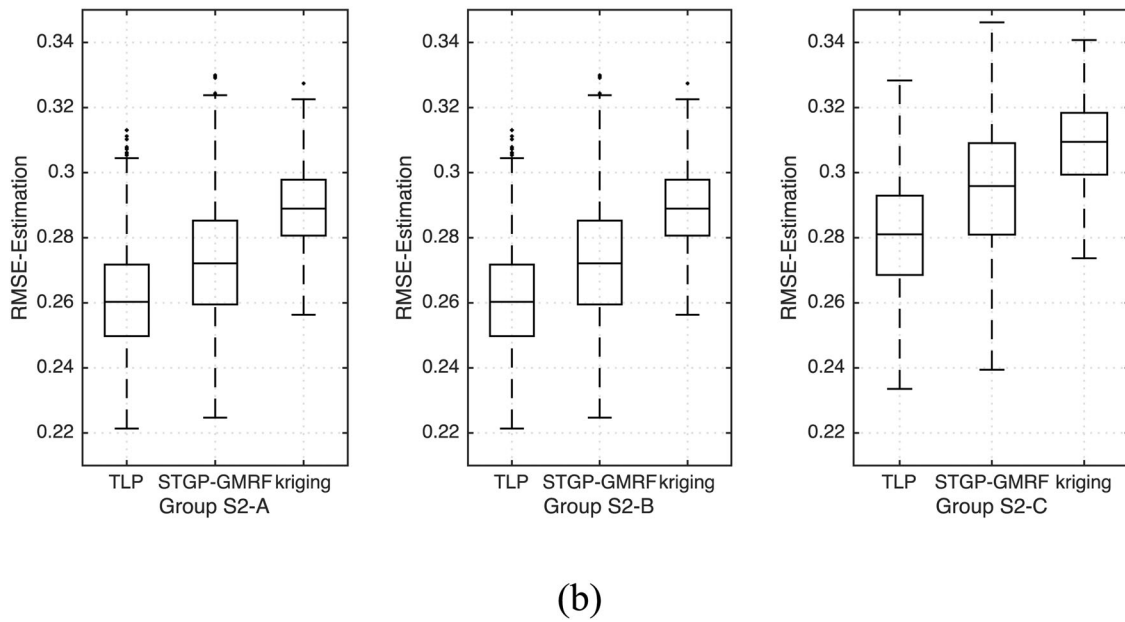
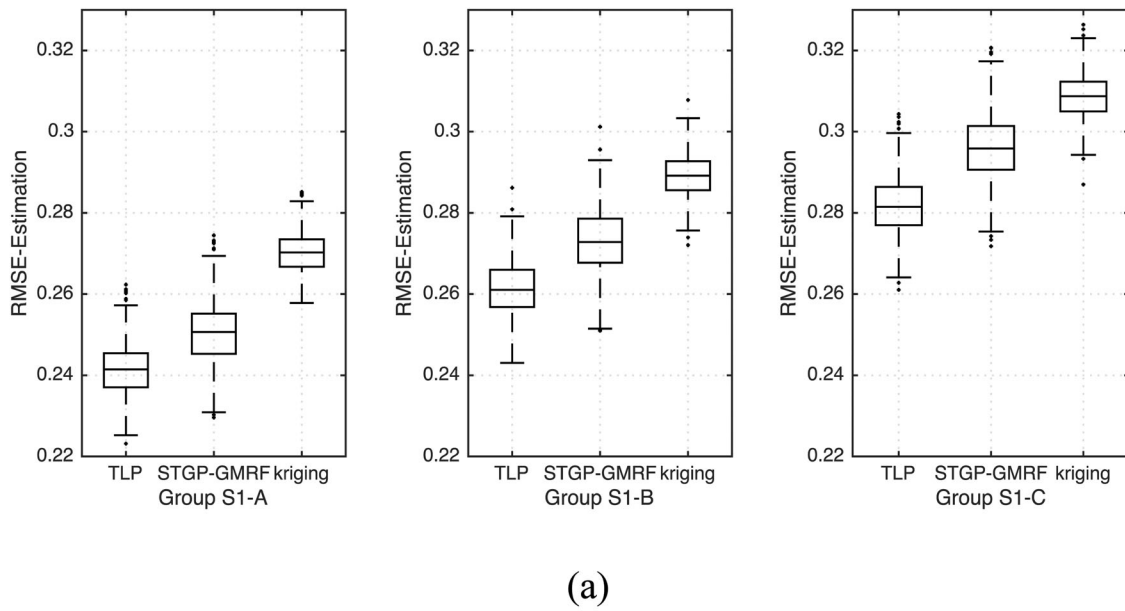
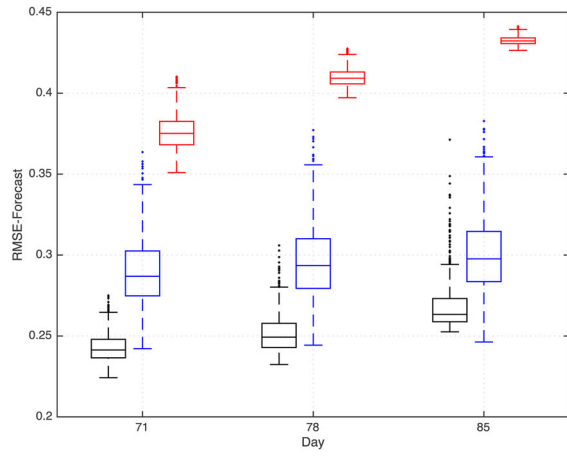
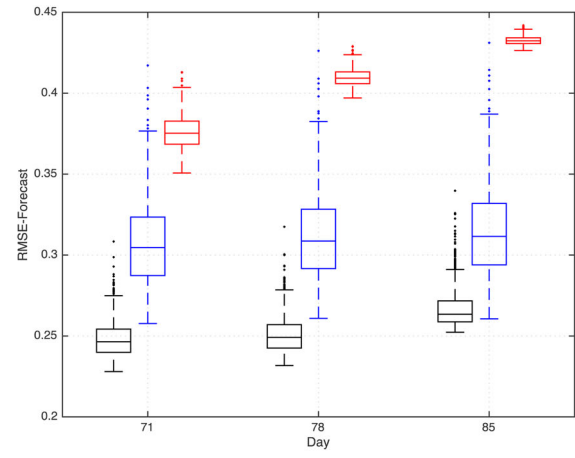


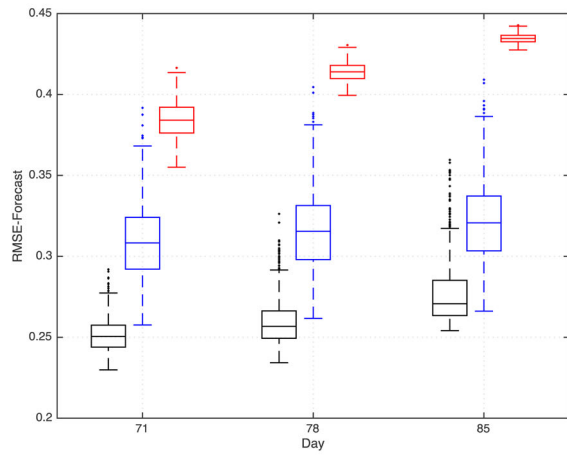
Figure 8. Comparison of RMSEs-Estimation using three models under 1000 replications.



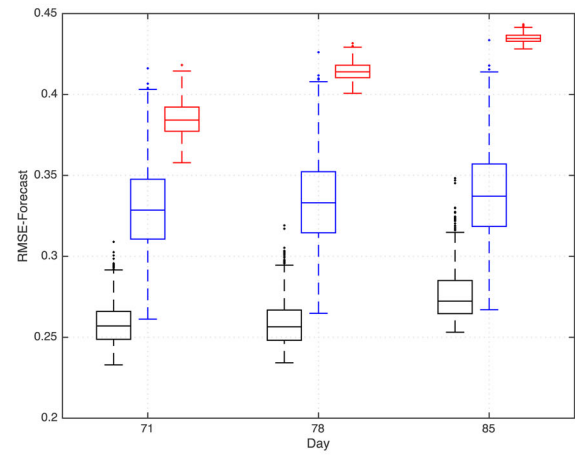
(a)



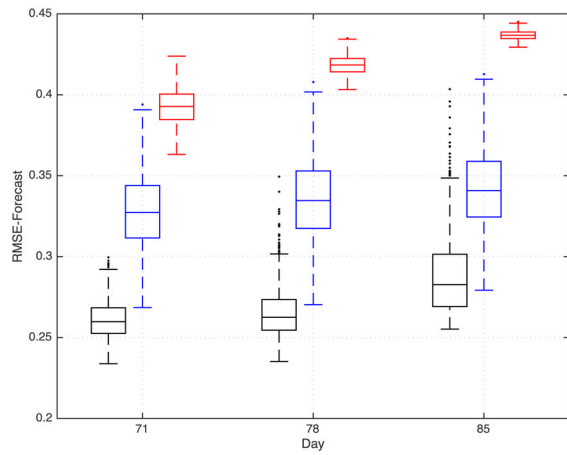
(d)



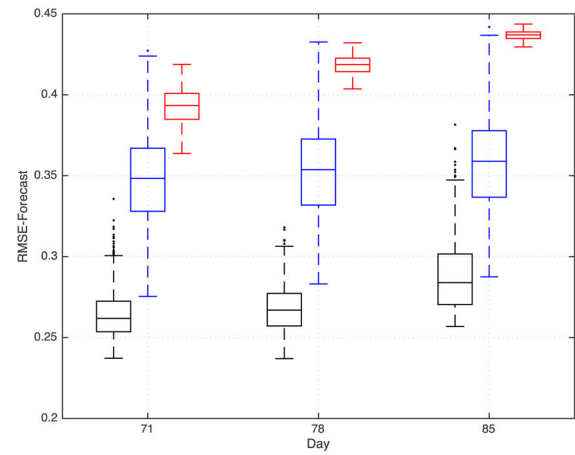
(b)



(e)



(c)



(f)

Black box(left)—TLP; Blue box(middle) — STGP-GMRF; Red box(right)—kriging

Figure 9. Comparison of RMSEs — Forecast using three models under 1000 replications: (a) Group S1-A; (b) Group S1-B; (c) Group S1-C; (d) Group S2-A; (e) Group S2-B; (f) Group S2-C.

Table 7. Model comparison results for field estimation.

Group	RMSE-Estimation				Improvement		
	Thermodynamic model	Kriging	STPF-GMRF	TLP	TLP vs Thermo-dynamic model	TLP vs Kriging	TLP vs STPF-GMRF
S1-A	0.3504	0.2702	0.2506	0.2416	31.05%	10.58%	3.59%
S1-B		0.2892	0.2732	0.2615	25.37%	9.58%	4.28%
S1-C		0.2960	0.3087	0.2818	19.58%	4.80%	8.71%
S2-A		0.2701	0.2500	0.2411	31.19%	10.74%	3.56%
S2-B		0.2892	0.2730	0.2613	25.43%	9.65%	4.29%
S2-C		0.3088	0.2957	0.2815	19.66%	8.84%	4.80%

Note: the improvement of model 1 vs model 2 is equal to (RMSE-Estimation of model 1 – RMSE-Estimation of model 2)/ RMSE-Estimation of model 2.

Table 8. Model comparison results for field forecast.

Date	Group	RMSE				Improvement		
		Thermodynamic model	Kriging	STPF-GMRF	TLP	TLP vs Thermo-dynamic model	TLP vs Kriging	TLP vs STPF-GMRF
Day 71	S1-A	0.4320	0.3759	0.2892	0.2428	43.80%	35.41%	16.04%
	S1-B		0.3843	0.3089	0.2514	41.81%	34.58%	18.61%
	S1-C		0.3926	0.3284	0.2610	39.58%	33.52%	20.52%
	S2-A		0.3759	0.3073	0.2482	42.55%	33.97%	19.23%
	S2-B		0.3847	0.3294	0.2582	40.23%	32.88%	21.62%
	S2-C		0.3929	0.3487	0.2645	38.77%	32.68%	24.15%
Day 78	S1-A	0.4393	0.4097	0.2957	0.2515	42.75%	38.61%	14.95%
	S1-B		0.4140	0.3159	0.2592	41.00%	37.39%	17.95%
	S1-C		0.4183	0.3358	0.2655	39.56%	36.53%	20.94%
	S2-A		0.4097	0.3115	0.2510	42.86%	38.74%	19.42%
	S2-B		0.4142	0.3338	0.2587	41.11%	37.54%	22.50%
	S2-C		0.4185	0.3535	0.2680	38.99%	35.96%	24.19%
Day 85	S1-A	0.4473	0.4325	0.3000	0.2680	40.08%	38.03%	10.67%
	S1-B		0.4346	0.3214	0.2786	37.72%	35.90%	13.32%
	S1-C		0.4368	0.3423	0.2884	35.52%	33.97%	15.75%
	S2-A		0.4325	0.3144	0.2669	40.33%	38.29%	15.11%
	S2-B		0.4347	0.3380	0.2761	38.27%	36.48%	18.31%
	S2-C		0.4368	0.3588	0.2880	35.61%	34.07%	19.73%

Note: the improvement of model 1 vs model 2 is equal to (RMSE-Forecast of model 1 – RMSE-Forecast of model 2)/ RMSE-Forecast of model 2.

and time point from the training dataset and assumed that these data were missing. Similar to S1, we also considered three groups of data with different levels of missing data to evaluate our model performance, that is, the proportions of the missing data to the entire training data were set as 30% (S2-A), 40% (S2-B), and 50% (S2-C).

Six groups of data, namely, S1-A, S1-B, S1-C, S2-A, S2-B, and S2-C, were used. Furthermore, we repeated the procedure in each group 1000 times by randomly selecting missing data to evaluate the proposed model performance.

5.3.2. Field estimation and forecast

For each group, the parameter estimation procedure was similar to the aforementioned ones without missing data in Section 5.2 and are thus not detailed here. We estimated the points of interest (test data + missing data) using the remaining training data at the first 10 given time points from days 1 to 64 and calculated the corresponding RMSE – Estimation. We forecasted the thermal field distribution (test data) on day 71 (one-step forecast) using our proposed model and calculated the

corresponding RMSE – Forecast. Similarly, the thermal field distributions on days 78 (two-step forecast) and 85 (three-step forecast) were iteratively forecasted.

Tables 4 and 5 summarize the mean and standard deviation of RMSE-Estimation and RMSE-Forecast under 1000 replications using the six groups of training data. Our proposed model achieves a favorable estimation result and excellent one- and two-step forecast results in both situations. Our model still performs well when the proportion of missing data was moderate ($\leq 40\%$, i.e., S1-A, S1-B, S2-A, and S2-B). The forecast accuracy decreases but remained within the acceptable range when the proportion of missing data increases ($> 40\%$, i.e., S1-C and S2-C). We calculated the computational time for the six groups S1-A, S1-B, S1-C, S2-A, S2-B, and S2-C using the proposed model. Table 6 lists the average computational time for field prediction, and the average computational time for each group is about 51 s. In the case study of grain storage, the grain temperature changes slowly during storage, and thus the sampling frequency of the grain temperature sensor data is low to reduce the data storage and processing costs. The computational time of the proposed model is comparably acceptable under this scenario.

Figure 6 shows four examples of thermal profiles at different locations acquired by our proposed model using the training data of Group S1-A. Figures 6(a)–(c) present the thermal profiles at the sensor locations. Our model can evidently estimate the thermal value on sensor locations with missing data accurately before day 64 and provide a good forecast performance for days 71 and 78. Figure 6(d) presents the thermal profiles off the sensor locations. Our model can accurately capture local thermal variations for field estimation and one- and two-step field forecasts. Figure 7 depicts the forecasted thermal field distributions of the entire granary on days 71, 78, and 85 using the training data of Group S1-A. Our proposed model characterizes the spatiotemporal correlation of the thermal field. The forecasted thermal field distribution is practically important for monitoring the quality of grains and reducing grain loss during storage.

We compared our proposed model with two alternative models, namely, the kriging model and the spatiotemporal Gaussian process with a built-in GMRF (STGP-GMRF) model. Similar to our proposed model, the two models use the mixed-effect model framework (Eq. (1)) and the thermodynamic model (Eq. (2)) to characterize the global trends; however, these models characterize local thermal variations using only the sensor data collected from the target granary. In the kriging model, the spatiotemporal correlation of local thermal variations is characterized by a covariance function in terms of space and time. The detailed information about kriging can be found in the paper of Inoue, Sasaki, and Washio (2012). Similar to our proposed model, STGP-GMRF uses a Gaussian process with a built-in GMRF to characterize the spatial correlation at each time point and applies a conditional autoregressive temporal sequence model to characterize the temporal correlation. Compared with our proposed model, the STGP-GMRF model disregards transfer learning. We implemented the three models 1000 times and calculated their RMSEs – Estimation and RMSEs – Forecast.

Figures 8 and 9 show the boxplots of the RMSE-Estimation and RMSE-Forecast values using the three models under 1000 replications. Our proposed model clearly performs better in terms of estimation and forecasting compared with the kriging model, especially in field forecasting, because our model accurately captures the spatiotemporal effects of the thermal field, that is, the spatial-correlated neighborhood structure and the conditional autoregressive temporal sequence. Additionally, our proposed model obtains better estimation and forecast results compared with the STGP-GMRF model by adopting the

transfer learning idea, given other sensor observations from multiple homogeneous fields. Specifically, we listed the model comparison results about the proposed model, the STPF-GMRF model, the kriging model, and the thermodynamic model in Tables 7 and 8 for field estimation and forecast. In Table 7 for field estimation, the comparison of our proposed model and the STGP-GMRF model aims to show the superiority by using transfer learning. The RMSE of our proposed model has about 4% improvement compared with the STPF-GMRF model, which shows our proposed model performs better by using transfer learning. The comparison of our proposed model and the kriging model aims to show the superiority of the spatiotemporal modeling part and the transfer learning part in our proposed model. The RMSE of our proposed model has about 9% improvement compared with the Kriging model, which shows our proposed model performs better for spatiotemporal modeling and transfer learning. The thermodynamic model is used to capture the global trends. The comparison of our proposed model and the thermodynamic model aims to show the superiority of the proposed model by developing the transfer learning-based spatiotemporal statistical method for the modeling of local thermal variations. The RMSE of our proposed model has about 25% improvement compared with the thermodynamic model, which shows our proposed model performs much better for developing the novel method for the modeling of local thermal variations. Similarly in Table 8 for field forecast, we can see that our proposed model performs much better by developing the transfer learning-based spatiotemporal statistical method for local thermal variations.

6. Conclusion

Thermal field prediction plays an essential part in engineering domains. Obtaining an accurate thermal field distribution can provide useful and thorough information for quality control, system improvement and maintenance. Estimating and forecasting a thermal field distribution using data acquired from sensor networks is typically challenging due to the complex spatiotemporal structures of the thermal field and data sparsity and missing data problems. This study proposes an innovative TLP approach to predict a 3D thermal field distribution using sensor observations from homogeneous data sources. The TLP approach is characterized by an autoregressive model for a temporal sequence of spatial transfer learning processes to capture spatiotemporal dynamics of the thermal field and address the issue of data sparsity and missing data. Particularly, we model the spatial transfer learning processes by establishing a

multitask Gaussian process framework, in which we model spatial correlation at each time point by a Gaussian process with a built-in GMRF on the basis of a grid-based sensor network.

Our proposed method offers deep insights into the spatiotemporal dynamics of a thermal field and has been validated to have a good prediction performance through a real case study of grain storage. The proposed model framework can be generalized and applied to the thermal field prediction of other engineering cases, such as the air temperature prediction for research on regional weather change, ocean field prediction to identify climate variability and global ocean circulation, and land temperature prediction to capture interactions between the Earth's surface and the atmosphere. In addition, the idea of the proposed method can be applied to other PDE-based systems in engineering cases, which satisfy the following conditions: (1) Sensor data of the systems are collected; (2) Physical parameters of the PDE are known, which can be either obtained by engineering knowledge or experimental studies; (3) The boundary conditions and initial values of the PDE are known, so that the PDE can be solved analytically or has the numerical solution by approximation methods, e.g., the finite different method and the variation method. In engineering cases, there are a great number of PDE-based systems that satisfy the aforementioned conditions, such as the thermal stratification of sodium fast reactors based on the mass, momentum, and energy conservation equations, the acoustic field based on the wave equation, and the convection–diffusion of pollution sources based on the convection–diffusion equation. To apply this approach successfully on these engineering cases, the target field and its homogeneous fields should have similar structures, and data acquisition should be operated under similar external conditions.

In the future work, given the predicted thermal field distribution by the proposed model, we will focus on establishing an effective strategy for simultaneously monitoring multiple homogeneous thermal fields.

Funding

This work was supported in part by the National Science Foundation of China under Grants 71771004, 71932006, 71690232, 71571003, and in part by the National Science Foundation under Grant 1637772.

About the authors

Di Wang received the B.S. degree in industrial engineering from Nankai University, Tianjin, China, in 2015, and the Ph.D. degree in management science and engineering from

Peking University, Beijing, China, in 2020. She is currently an Assistant Professor with the Department of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China. Her research interests include statistical modeling of spatiotemporal data and artificial intelligence of complex engineering systems. Dr. Wang is a member of INFORMS and IISE.

Kaibo Liu received the B.S. degree in industrial engineering and engineering management from 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 Associate Professor with the Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI, USA, where he is also the Associate Director of the UW–Madison IoT Systems Research Center. His research interests include system informatics, big data analytics, and data fusion for process modeling, monitoring, diagnosis, prognostics, and decision making. Dr. Liu is a member of ASQ, INFORMS, SME, and IISE.

Xi Zhang received the B.S. degrees in mechanical engineering and business administration from Shanghai Jiao Tong University, Shanghai, China, in 2006, and the Ph.D. degree in industrial and management systems engineering from the University of South Florida, Tampa, FL, USA, in 2010. He is currently an Associate Professor with the Department of Industrial Engineering and Management, Peking University, Beijing, China. His research interests include physics-based engineering data integration and analytics for process monitoring, diagnosis, control and optimization in complex engineering, and service systems. Dr. Zhang is a member of ASQ, INFORMS, and IISE.

ORCID

Di Wang  <http://orcid.org/0000-0001-7030-6521>

Kaibo Liu  <http://orcid.org/0000-0003-2863-5748>

Xi Zhang  <http://orcid.org/0000-0003-3415-5345>

References

- Aliaga, R. J. 2017. Real-time estimation of zero crossings of sampled signals for timing using cubic spline interpolation. *IEEE Transactions on Nuclear Science* 64 (8): 2414–22. doi: [10.1109/TNS.2017.2721103](https://doi.org/10.1109/TNS.2017.2721103).
- Ba, S., and V. R. Joseph. 2012. Composite Gaussian process models for emulating expensive functions. *The Annals of Applied Statistics* 6 (4):1838–60. doi: [10.1214/12-AOAS570](https://doi.org/10.1214/12-AOAS570).
- Bonilla, E. V., K. M. A. Chai, and C. K. I. Williams. 2008. Multi-task Gaussian process prediction. *Advances in Neural Information Processing Systems* 20:153–60.
- Cheng, Z., and J. W. Mark. 2010. Channel estimation by modulated Lagrange interpolation. *Signal Processing* 90 (9):2749–59. doi: [10.1016/j.sigpro.2010.03.026](https://doi.org/10.1016/j.sigpro.2010.03.026).
- Cressie, N., and G. Johannesson. 2008. Fixed rank kriging for very large spatial data sets. *Journal of the Royal*

- Statistical Society: Series B (Statistical Methodology)* 70 (1):209–26. doi: [10.1111/j.1467-9868.2007.00633.x](https://doi.org/10.1111/j.1467-9868.2007.00633.x).
- Furrer, R., M. G. Genton, and D. Nychka. 2006. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics* 15 (3):502–23. doi: [10.1198/106186006X132178](https://doi.org/10.1198/106186006X132178).
- Goncalves, A. R., A. Banerjee, and F. J. Von Zuben. 2017. Spatial projection of multiple climate variables using hierarchical multitask learning. In *The 31st AAAI Conference on Artificial Intelligence*, 4509–15.
- Hartman, L., and O. Hössjer. 2008. Fast kriging of large data sets with Gaussian Markov random fields. *Computational Statistics and Data Analysis* 52 (5): 2331–49. doi: [10.1016/j.csda.2007.09.018](https://doi.org/10.1016/j.csda.2007.09.018).
- Huang, Q. 2010. Physics-driven Bayesian hierarchical modeling of the nanowire growth process at each scale. *IIE Transactions* 43 (1):1–11. doi: [10.1080/07408171003795335](https://doi.org/10.1080/07408171003795335).
- Inoue, T., T. Sasaki, and T. Washio. 2012. Spatio-temporal kriging of solar radiation incorporating direction and speed of cloud movement. In *The 26th Annual Conference of the Japanese Society for Artificial Intelligence*, 1–4.
- Jia, C., D. Sun, and C. Cao. 2000. Finite element prediction of transient temperature distribution in a grain storage bin. *Journal of Agricultural Engineering Research* 76 (4): 323–30. doi: [10.1006/jaer.2000.0533](https://doi.org/10.1006/jaer.2000.0533).
- Jiang, H. J., X. Deng, V. Lopez, and H. Hamann. 2013. Online updating and scheduling of computer model with application to data center thermal management. In *Proceedings of ASME IPACK2013*, 73042.
- Katzfuss, M., and N. Cressie. 2011. Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *Journal of Time Series Analysis* 32 (4):430–46. doi: [10.1111/j.1467-9892.2011.00732.x](https://doi.org/10.1111/j.1467-9892.2011.00732.x).
- Khachatourian, O. A., and F. A. De Oliveira. 2006. Mathematical modeling of airflow and thermal state in large aerated grain storage. *Biosystems Engineering* 95 (2): 159–69. doi: [10.1016/j.biosystemseng.2006.05.009](https://doi.org/10.1016/j.biosystemseng.2006.05.009).
- Li, J., R. Jin, and H. Z. Yu. 2018. Integration of physically-based and data-driven approaches for thermal field prediction in additive manufacturing. *Materials and Design* 139:473–85. doi: [10.1016/j.matdes.2017.11.028](https://doi.org/10.1016/j.matdes.2017.11.028).
- Lin, Y., K. Liu, E. Byon, X. Qian, S. Liu, and S. Huang. 2018. A collaborative learning framework for estimating many individualized regression models in a heterogeneous population. *IEEE Transactions on Reliability* 67 (1): 328–41. doi: [10.1109/TR.2017.2767941](https://doi.org/10.1109/TR.2017.2767941).
- Liu, K., S. Huang, C. Jin, H. Xie, and F. Wang. 2017. Prediction models of the thermal field on ice-snow melting pavement with electric heating pipes. *Applied Thermal Engineering* 120:269–76. doi: [10.1016/j.applthermaleng.2017.04.008](https://doi.org/10.1016/j.applthermaleng.2017.04.008).
- Liu, X., V. Gopal, and J. Kalagnanam. 2018. A spatio-temporal modeling approach for weather radar reflectivity data and its applications in tropical southeast Asia. *The Annals of Applied Statistics* 12 (1):378–407. doi: [10.1214/17-AOAS1064](https://doi.org/10.1214/17-AOAS1064).
- Mariella, L., and M. Tarantino. 2010. Spatial temporal conditional auto-regressive model: A new autoregressive matrix. *Austrian Journal of Statistics* 39 (3):223–44. doi: [10.17713/ajs.v39i3.246](https://doi.org/10.17713/ajs.v39i3.246).
- Perdikaris, P., D. Venturi, J. O. Royset, and G. E. Karniadakis. 2015. Multi-fidelity modelling via recursive co-kriging and Gaussian-Markov random fields. *Proceedings Mathematical, Physical, and Engineering Sciences* 471 (2179):20150018. doi: [10.1098/rspa.2015.0018](https://doi.org/10.1098/rspa.2015.0018).
- Rasmussen, C. E. 2004. Gaussian process in machine learning. In *Advanced lectures on machine learning*, eds. O. Bousquet, U. von Luxburg, and G. Rätsch, 63–71. Tbingen, Germany: Springer.
- Shao, C., J. Ren, H. Wang, J. Jin, and S. Hu. 2017. Improving machined surface shape prediction by integrating multi-task learning with cutting force variation modeling. *Journal of Manufacturing Science and Engineering* 139 (1):1–11. doi: [10.1115/1.4034592](https://doi.org/10.1115/1.4034592).
- Song, P., Y. Jin, and L. Zhao. 2014. Speech emotion recognition using transfer learning. *IEICE Transactions on Information and Systems* E97 (9):2530–2. doi: [10.1587/transinf.2014EDL8038](https://doi.org/10.1587/transinf.2014EDL8038).
- Stein, M. L. 2012. *Interpolation of spatial data: Some theory for kriging*. New York: Springer Science and Business Media.
- Thijssen, J. 2017. *Computational physics*. Cambridge: Cambridge University Press.
- Wang, D., and X. Zhang. 2015. A prediction method for interior temperature of grain storage via dynamics model: A simulation study. In *Proceedings of IEEE International Conference of Automation Science and Engineering*, 1477–83.
- Wang, D., K. Liu, and X. Zhang. 2019. Modeling of a three-dimensional dynamic thermal field under grid-based sensor networks in grain storage. *IIE Transactions* 51 (5): 531–46. doi: [10.1080/24725854.2018.1504356](https://doi.org/10.1080/24725854.2018.1504356).
- Wang, D., K. Liu, and X. Zhang. 2020. Spatiotemporal thermal field modeling using partial differential equations with time-varying parameters. *IEEE Transactions on Automation Science and Engineering* 17 (2):646–57. doi: [10.1109/TASE.2019.2940269](https://doi.org/10.1109/TASE.2019.2940269).
- Xu, L., and Q. Huang. 2012. Modeling the interactions among neighboring nanostructures for local feature characterization and defect detection. *IEEE Transactions on Automation Science and Engineering* 9 (4):745–54. doi: [10.1109/TASE.2012.2209417](https://doi.org/10.1109/TASE.2012.2209417).
- Xu, Y., and J. Choi. 2012. Spatial prediction with mobile sensor networks using Gaussian processes with built-in Gaussian Markov random fields. *Automatica* 48 (8): 1735–40. doi: [10.1016/j.automatica.2012.05.029](https://doi.org/10.1016/j.automatica.2012.05.029).
- Yan, H., K. Paynabar, and J. Shi. 2018. Real-time monitoring of high-dimensional functional data streams via spatio-temporal smooth sparse decomposition. *Technometrics* 60 (2):181–97. doi: [10.1080/00401706.2017.1346522](https://doi.org/10.1080/00401706.2017.1346522).
- Yan, Y., X. Li, and J. Tu. 2017. Effects of manikin model simplification on CFD predictions of thermal flow field around human bodies. *Indoor and Built Environment* 26 (9):1185–97. doi: [10.1177/1420326X16653500](https://doi.org/10.1177/1420326X16653500).
- Yoon, H., and J. Li. 2018. A novel positive transfer learning approach for telemonitoring of Parkinson’s disease. *IEEE Transactions on Automation Science and Engineering* 16 (1):180–191. doi: [10.1109/TASE.2018.2874233](https://doi.org/10.1109/TASE.2018.2874233).
- Zhang, B., H. Sang, and J. Z. Huang. 2015. Full-scale approximations of spatio-temporal covariance models for large datasets. *Statistica Sinica* 25 (1):99–114. doi: [10.5705/ss.2013.260w](https://doi.org/10.5705/ss.2013.260w).
- Zhao, P., S. C. H. Hoi, J. Wang, and B. Li. 2014. Online transfer learning. *Artificial Intelligence* 216:76–102. doi: [10.1016/j.artint.2014.06.003](https://doi.org/10.1016/j.artint.2014.06.003).

Zheng, Y., F. Liuand, and H. Hsieh. 2013. U-Air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1436–44.

Appendix A. Parameter estimation of θ_t

We denote $\theta_t = \{\sigma_t^2, \rho_t\}$. The log-likelihood function of θ_t is $l(\theta_t) = N \log |Z_t^T (\tilde{\mathbf{K}}_t^s(\theta_t))^{-1} Z_t| + L \log |\tilde{\mathbf{K}}_t^s(\theta_t)|$. θ_t is estimated by minimizing the log-likelihood function $l(\theta_t)$ on the basis of its partial derivatives as follows:

$$\begin{aligned} \frac{\partial l(\theta_t)}{\partial \theta_t} = & -N \text{trace} \\ & \left\{ \left(Z_t^T (\tilde{\mathbf{K}}_t^s)^{-1} Z_t \right)^{-1} Z_t^T (\tilde{\mathbf{K}}_t^s)^{-1} \frac{\partial \tilde{\mathbf{K}}_t^s}{\partial \theta_t} (\tilde{\mathbf{K}}_t^s)^{-1} Z_t \right\} \\ & + L \text{trace} \left\{ (\tilde{\mathbf{K}}_t^s)^{-1} \frac{\partial \tilde{\mathbf{K}}_t^s}{\partial \theta_t} \right\}, \end{aligned} \quad (\text{A1})$$

where

$$\begin{aligned} \frac{\partial \tilde{\mathbf{K}}_t^s}{\partial \sigma_t^2} &= \Lambda (\mathbf{W}_D - \rho_t \mathbf{W}) \Lambda^T, \\ \frac{\partial \tilde{\mathbf{K}}_t^s}{\partial \rho_t} &= \sigma_t^2 \Lambda (\mathbf{W}_D - \rho_t \mathbf{W})^{-1} \mathbf{W} (\mathbf{W}_D - \rho_t \mathbf{W})^{-1} \Lambda^T. \end{aligned}$$

Eq. (A1) can be conveniently solved by using a numerical optimization approach, such as steepest gradient method.

Appendix B. Generalized least squares method for estimating temporal-associated parameters

Following Eq. (19), the temporal-associated parameters $\mathbf{r} = \{r_1, r_2, \dots, r_Q\}$ can be estimated using the generalized least squares method as follows:

$$\hat{\mathbf{r}} = (\Gamma_0^T \Omega \Gamma_0)^{-1} \Gamma_0^T \Omega \Gamma_1, \quad (\text{A2})$$

where $\Omega = \text{diag} \left\{ \left(\hat{\sigma}_{Q+1}^2 \hat{\mathbf{K}}_{Q+1}^z \otimes (\Lambda \mathbf{W}_D^{-1} \Lambda^T) \right)^{-1}, \dots, \left(\hat{\sigma}_M^2 \hat{\mathbf{K}}_M^z \otimes (\Lambda \mathbf{W}_D^{-1} \Lambda^T) \right)^{-1} \right\}$ is an $LN(M-Q) \times LN(M-Q)$ matrix, $\Gamma_0 = \left\{ \mathbf{C}_{Q+1,1} \tilde{\mathbf{y}}_1, \dots, \mathbf{C}_{Q+1,Q} \tilde{\mathbf{y}}_Q; \dots; \mathbf{C}_{M,M-Q} \tilde{\mathbf{y}}_{M-Q}, \dots, \mathbf{C}_{M,M-1} \tilde{\mathbf{y}}_{M-1} \right\}$ is an $LN(M-Q) \times Q$ matrix, and $\Gamma_1 = \left\{ \tilde{\mathbf{y}}_{Q+1}, \dots, \tilde{\mathbf{y}}_M \right\}^T$ is an $LN(M-Q) \times 1$ vector.

Appendix C. Illustration of the computational complexity of using TLP

At each time point t_m , the local thermal variation $\mathbf{V}_t$ follows a Gaussian distribution with the covariance matrix $\mathbf{D}_t$. We consider L homogeneous fields and have N points in the spatial domain, and thus $\mathbf{D}_t$ is a $LN \times LN$ matrix. The covariance matrix $\mathbf{D}_t$ contains $(LN)^2$ elements, and each element should be considered in the operation for each spatial point $\mathbf{s}_i$ ($i = 1, \dots, N$) in each field l ($l = 1, \dots, L$). In conclusion, the computational cost of TLP is $O((LN)^3)$ at each time point.