Novel Geospatial Interpolation Analytics for General Meteorological Measurements

Bingsheng Wang
Department of Computer Science, Virginia Tech
7054 Haycock Road
Falls Church, VA 22043 USA
claren89@vt.edu

Jinjun Xiong
Smarter Energy Research Institute
IBM Thomas J. Watson Research Center
Yorktown Heights, NY 10598 USA
jinjun@us.ibm.com

ABSTRACT

This paper addresses geospatial interpolation for meteorological measurements in which we estimate the values of climatic metrics at unsampled sites with existing observations. Providing climatological and meteorological conditions covering a large region is potentially useful in many applications, such as smart grid. However, existing research works on interpolation either cause a large number of complex calculations or are lack of high accuracy. We propose a Bayesian compressed sensing based non-parametric statistical model to efficiently perform the spatial interpolation task. Student-t priors are employed to model the sparsity of unknown signals' coefficients, and the Approximated Variational Inference (AVI) method is provided for effective and fast learning. The presented model has been deployed at IBM, targeting for aiding the intelligent management of smart grid. The evaluations on two real world datasets demonstrate that our algorithm achieves state-of-the-art performance in both effectiveness and efficiency.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data Mining—*Machine Learning*; I.5 [**PATTERN RECOGNITION**]: I.5.4 Applications—*Miscellaneous*

Keywords

Geospatial Interpolation; Analytics; Meteorological Measurements; Bayesian Inference; Smart Grid

1. INTRODUCTION

Complete and accurate source of meteorological data are prerequisites for the efficient modelling of a wide variety of environmental processes [14]. For example, the productivity for wheat largely depends on the climate patterns, such as precipitation [24]. Crime rates are primarily driven by pleasant weather [13], while cherry blossom is mainly influenced by temperature, sunshine, etc. [20]. However, the most common source of climatic data is meteorological stations, which only supply measurements in some limited locations. Geospatial interpolation has thus attracted many researchers for the estimation of weather or climate variables at unsampled locations.

This paper looks specifically at the application of spatial interpolation to smart grid applications. Smart grid refers to a modernized utility electricity delivery systems, where the power distribution and management is enhanced by employing computer-based remote control and automation along with advanced two-way communications and pervasive computing capabilities, targeting for improved control, efficiency, reliability, safety and sustainability [5, 30]. A smart grid intelligently and bidirectionally delivers electricity between suppliers and consumers, where meteorological variables play an important role [12]. For example, temperature and precipitation are two climate measurements that have great impact on the electricity grid. From the perspective of temperature, we anticipate that when the thermometer rises in summer, so does the demand for power to achieve comfortable temperatures. The increased degree of hotness would also decrease the efficiency of power plants no matter the power is generated from stream, water, nuclear, natural resources, solar, or wind. At the same time, the increased temperatures might reduce transmission lines' efficiency and lead to line sags, which would potentially disrupt the service [8]. Considering precipitation's impacts, we expect that less precipitation might cause the drop of water level, further reducing the generation of hydro-power plants. Drought situation can pose a serious threat to the transmission lines, especially in the states where there is a fire season, such as California. Due to the lack of precipitation, fires are more like to occur and destroy the brittle infrastructure that transports power. It's of critical importance to know about the climate and weather conditions covering the power grid for intelligent peak-load management and effective measures taken to avoid lines' impairment. However, it's impossible to install weather sensors all around the transmission lines and power plants. An effective and efficient interpolation method is expected to provide maps of climatological and meteorological conditions for the whole grid.

Motivated by the needs of smart grid on weather data products, we focus on geospatial interpolation for general meteorological measurements, the procedure of estimating unsampled sites' values with a few observations. Much work has been done on estimating weather or climate variables by spatial interpolation. One branch is to apply physical-deterministic model to simulate the spatial variations in climate metrics by solving a large scale of partial differential equations. Although such models are most accurate, however, it's of high complexity and requires high performance computing, such as the national weather service [21] and the Deep Thunder system deployed by IBM [27]. To balance the efficiency and accuracy, many researchers proposed a variety of learning based interpolation techniques, ranging from inverse distance weighting, splines,

KDD '14 August 24–27 New York City USA

http://dx.doi.org/10.1145/2623330.2623367 .

regression and Kriging to neural networks and machine learning techniques [23, 17, 1]. These learning based models are easy to set up and can complete interpolation in a more efficient manner. However, they are lack of high accuracy.

The approach we present in this paper builds upon compressed sensing methods and recent work in reconstruction problems [18, 7, 31]. Specifically, the proposed model is learned with few measurements in some weather stations and then the meteorological values at all other locations are estimated and evaluated. While the interpolation procedure can be naturally formulated as such a reconstruction problem, we know of no previous application of these methods to the geospatial interpolation task. Indeed, the most common application of such algorithm is image reconstruction, which typically has small scale in spatial dimension. The large-scale geospatial interpolation task poses a new set of challenges for such methods, and existing approaches alone cannot perform well enough. Instead of applying Laplace priors to model coefficients' sparsity [26, 15, 3], we propose to use Student-t priors, and show that this improves the performance in the meteorological domain. Then, an efficient inference algorithm named Approximated Variational Inference (AVI) is developed based on Taylor series to deal with the intractability caused by the non-conjugate priors and the approximation algorithm is theoretically justified. Besides, we present a sampling based strategy as an alternative inference method to experimentally show the correctness of AVI. With respect to the task of geospatial interpolation, our model significantly outperforms the traditional methods: Thin Plate Spline (TPS) [23] and Kriging [17]. In summary, the primary contributions of this paper are as follows:

- **Provision of smarter interpolation service:** An in-depth study on the meteorological metrics is performed and our system can be potentially used to estimate the ambient weather conditions of power plants, transformers and transmission lines to support intelligent management of electrical grid.
- **Design of a novel geospatial interpolation model:** We incorporate Student-t prior into Bayesian compressed sensing to accurately model the distributions of sparse coefficients, and an efficient and effective learning algorithm is provided.
- **Comprehensive experiments to validate the effectiveness and efficiency:** We demonstrated that the proposed model outperformed the baselines, and the interpolated results are visualized to show climate patterns.

The remainder of this paper is organized as follows. In Section 2, we introduce the application deployment for the proposed approach, and the background knowledge. The Bayesian Compressed Sensing using student-T priors model (BCST) is presented in Section 3, including the introduction of the generative model and the inference algorithms. The effectiveness and efficiency of the model are illustrated with extensive experiments in Section 4. Section 5 presents our conclusions.

2. BACKGROUND & PROBLEM

2.1 Application Deployment

The intelligent management of electrical grid largely depends on the weather and climate conditions around the power plants, transformers and transmission lines. We focus on the application of smart grid asset management. For example, transformers are essential assets for utility companies, and it is impossible to construct a weather station for every transformer due to the large number of transformers. To estimate the electrical age of transformers (which is different from calendar age) [2, 28], it is of critical importance to consider the weather conditions as input variables for modeling and computing. Acting as the smarter interpolation service, our system is capable of producing a map of meteorological measurements to potentially support the smart management of electricity grid. As shown in Figure 1, the system is consists of three key modules:

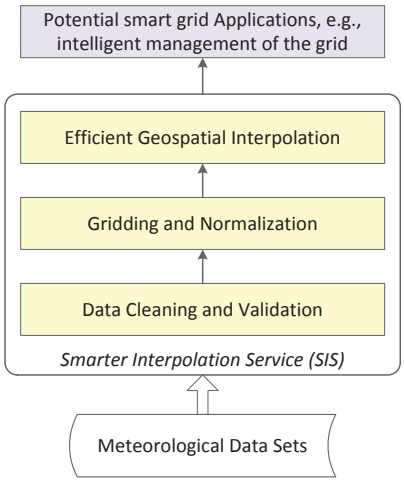

Figure 1: Overview of the deployed system: smarter interpolation service provides support for the intelligent management of power grid.

- *Data Cleaning and Validation:* This module is responsible to handle the input meteorological data. Clean the error data items, such as the data out of spatial range or containing wrong time stamp. Then we check the validity of the data, for example, the measurement of humidity (i.e., relative humidity) which is bigger than 100% or less than 0% is invalid.
- *Gridding and Normalization:* Gridding is to split the spatial region into cells, where each cell refers to the location of interest for interpolation. A large region can be divided in a recursive way and then each sub-region could be interpolated concurrently, while the edge effects caused by recursive division could be eliminated by overlapping adjacent sub-regions. Normalization is executed to provide a limited range values, which could potentially advance the process of learning spatial variations.
- *Efficient Geospatial Interpolation:* This is the core part of the system. A Bayesian compressed sensing based model is developed for fast and accurate spatial interpolation of meteorological measurements. The interpolation model is wrapped into a service provider to support smart grid. We can provide the service to detect the bottleneck of transmission lines using their ambient weather conditions. Other services can also be supplied, such as maximal workload estimation of power plants under current temperature, humidity, etc.

The preliminary evaluation results show that the deployed system can provide on-line interpolation service with very short latency. Additionally, by analysing the large amount of historical weather data, our system can discover the seasonal patterns and identify the extreme climatic conditions. An example application to smart grid is to help for wind turbine placement, since the performance and useful life of wind turbines is sensitive to wind direction/velocity, precipitation, humidity, etc. [25]. A visualization module is also

developed to visualize the interpolation results. This can help to quickly recognize the places with extreme weathers, which limit the normal functionality of power grid.

2.2 Preliminaries

A spatial performance of interest, such as temperature, can be expressed as a 2-D function $g(i, j)$, where i, j denote the coordinate of a spatial location in the two-dimensional plane. We can discretize the spatial region into equal-sized grids with P rows and Q columns, and then the coordinates i and j are discrete numbers accordingly: $i \in \{1, 2, \cdots, P\}, j \in \{1, 2, \cdots, Q\}$.

Formally, the function $g(i, j)$ can be projected to the frequency domain by numerous 2-D transforms, such as Discrete Fourier Transform (DFT) [19], Discrete Cosine Transform (DCT) [11]. Note that DCT is used throughout the paper to illustrate the idea of our proposed model, which can be easily generalized to other transforms.

Mathematically, the DCT transform can be represented as [11],

$$G(u, v) = \sum_{i=1}^{P} \sum_{j=1}^{Q} \Gamma(u, v, i, j) g(i, j) \tag{1}$$

where $G(u, v)$ for $u = 1, 2, \cdots, P$, $v = 1, 2, \cdots, Q$ is the set of DCT coefficients, and

$$\begin{aligned} \Gamma(u, v, i, j) &= \alpha'_u \cdot \beta'_v \cdot \cos \frac{\pi(2i-1)(u-1)}{2P} \cdot \\ &\qquad \cos \frac{\pi(2j-1)(v-1)}{2Q} \\ \alpha'_u &= \begin{cases} \sqrt{1/P} & , u = 1 \\ \sqrt{2/P} & , 1 < u \leq P \end{cases} \\ \beta'_v &= \begin{cases} \sqrt{1/Q} & , v = 1 \\ \sqrt{2/Q} & , 1 < v \leq Q \end{cases} \end{aligned} \tag{2}$$

While the Inverse Discrete Cosine Transform (IDCT) [11] can be applied to project the DCT coefficients ($G(u, v)$) to be the performance measurements in spatial domain,

$$g(i, j) = \sum_{u=1}^{P} \sum_{v=1}^{Q} \Gamma(u, v, i, j) G(u, v) \tag{3}$$

where $\Gamma(u, v, i, j)$ is defined in Eq. (2).

Compressed Sensing (CS) is an effective approach to recover original signal from few observations [6], which also has been validated to be powerful in spatial interpolation [31]:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{z} + \mathbf{u} \tag{4}$$

where $\mathbf{y} \in \mathbb{R}^{M \times 1}$ denotes the few observations, $\mathbf{z} \in \mathbb{R}^{K \times 1}$ denotes the sparse coefficients, $\mathbf{\Phi} \in \mathbb{R}^{M \times K}$ is the matrix to transform from $\mathbf{z}$ to $\mathbf{y}$, and $\mathbf{u} \in \mathbb{R}^{M \times 1}$ represents the noise with zero mean and precision τ. Let $\mathbf{f} \in \mathbb{R}^{K \times 1}$ denote the unknown signal. Then we know that

$$\begin{aligned} \mathbf{f} &= [g(1,1), g(1,2), \cdots, g(P,Q)]^T \in \mathbb{R}^{K \times 1} \\ \mathbf{y} &= [g(i_1, j_1), g(i_2, j_2), \cdots, g(i_M, j_M)]^T \in \mathbb{R}^{M \times 1} \\ \mathbf{\Phi} &= \begin{bmatrix} \phi_{1,1,1} & \phi_{1,1,2} & \cdots & \phi_{1,P,Q} \\ \phi_{2,1,1} & \phi_{2,1,2} & \cdots & \phi_{2,P,Q} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{M,1,1} & \phi_{M,1,2} & \cdots & \phi_{M,P,Q} \end{bmatrix} \in \mathbb{R}^{M \times K} \\ \phi_{m,u,v} &= \Gamma(u, v, i_m, j_m) \\ \mathbf{z} &= [G(1,1), G(1,2), \cdots, G(P,Q)]^T \in \mathbb{R}^{K \times 1} \end{aligned} \tag{5}$$

Under the conditions that $\mathbf{z}$ is sparse, even when the number of measurements in $\mathbf{y}$ is far smaller than the number of coefficients in $\mathbf{z}$, i.e., $M \ll K$, the original signal $\mathbf{f}$ can be accurately reconstructed using Eq. (3), where $\mathbf{z}$ can be estimated by solving the optimization problem:

$$\mathbf{z} = \underset{\mathbf{z}}{\operatorname{argmin}} \ \left\{ \|\mathbf{z}\|_1 + \zeta \|\mathbf{y} - \mathbf{\Phi}\mathbf{z}\|_2^2 \right\} \tag{6}$$

The l_1-regularization formulation in Eq. (6) is equivalent to using a Laplace prior[1] on the coefficients $\mathbf{z}$ [3, 15],

$$P(\mathbf{z}) = \frac{b}{2} \exp(-\frac{b}{2} \|\mathbf{z}\|_1) \tag{7}$$

2.3 Limitations of Laplace Prior

With respect to the large-scale geospatial interpolation task, we suspect the effectiveness of Lapalce prior in Eq. (7) for modelling the distributions of the sparse coefficients. From the visualization of the coefficients' histograms in Figure 2, we observe that the Lapalce prior over-penalizes the true coefficients. Student-t is thus selected to compare with Laplace, since Student-t's heavy tail property enables it to better model large coefficients, while the peak near zero helps to keep $\mathbf{z}$ to be sparse. To explore the suitable regularizations on $\mathbf{z}$, Log-Likelihood Ratio (LLR) test is performed with two real-world meteorological datasets: KNMI dataset [16] and a dataset provided by the vendor of IBM (Vendor dataset). Let $\mathbf{z} = [z_1, z_2, \cdots, z_K]^T$, LLR is formulated as

$$LLR = \sum_{k=1}^{K} \ln \left[\frac{P(z_k \mid \mu, \sigma, \nu)}{P(z_k \mid \lambda, b)} \right] \tag{8}$$

where Student-t distribution's parameters (μ: location parameter, σ: scale parameter, ν: degree of freedoms) and Laplace distribution's parameters (λ: location parameter, b: scale parameter) are respectively estimated by maximizing corresponding likelihood functions. Table 1 shows the parameters of these two distributions and the LLR values of Student-t over Laplace. The LLR values are far greater than 0, indicating that Student-t is a better fit than Laplace. Figure 2 also illustrates similar conclusions, where we respectively use Laplace and Student-t distributions to fit the histograms of 1000 samples' Discrete Cosine Transform (DCT) coefficients. It figures out that Student-t can better fit the histograms for all of the meteorological measurements. This motivates us to extend compressed sensing by applying Student-t priors to model the sparsity of the coefficients.

[1] Using Maximum a Posterior (MAP) with Eq. (10) and (7), Eq. (6) can be achieved with $\zeta = \tau / b$.

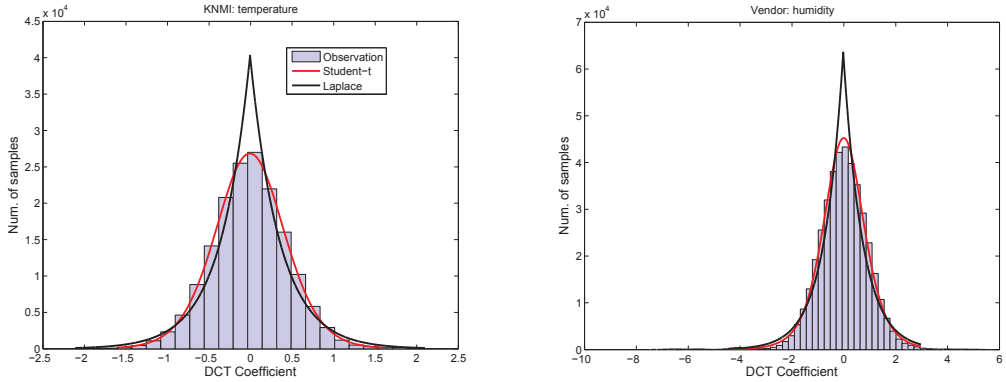

Figure 2: Histograms and distribution fitting of the normalized DCT coefficient for the two real-world meteorological datasets calculated from 100 samples for both dataset.

Table 1: Estimated parameters for Laplace distribution (λ: location parameter, b: scale parameter), and Student-t distribution (μ: location parameter, σ: scale parameter, ν: degree of freedoms) , and the log-likelihood ratio (LLR) values.

Meteorological Measurements	λ	b	μ	σ	ν	LLR
KNMI: temperature	-7.3×10^{-3}	0.3559	-8.6×10^{-3}	0.4200	11.5277	4499.0
Vendor: humidity	-5.6×10^{-3}	0.7392	-4.9×10^{-3}	0.7937	5.8467	10418

3. A NOVEL GEOSPATIAL INTERPOLATION METHOD

The geospatial interpolation of large-scale meteorological measurements has become a critical challenge since the prior of coefficients can not be captured with Laplace distribution. This section introduces a novel geospatial interpolation method by applying Student-t to model coefficient's distribution, and corresponding inference and learning algorithms are provided.

3.1 Model Definition

We begin our treatment of the generative model by expressing it as a directed graph, as illustrated in Figure 3. The measurement noise $\mathbf{u}$ is Gaussian with zero mean and unknown precision τ. With Eq. (4), we know that for each sample y_i, it depends on the latent variables $\mathbf{z}$ and τ

$$P(y_i \mid \mathbf{z}, \tau) = \mathcal{N}\left(y_i \mid \boldsymbol{\phi}_i \mathbf{z}, \tau^{-1}\right) \tag{9}$$

The M measurements are independent with each other given all the latent variables

$$P(\mathbf{y} \mid \mathbf{z}, \tau) = \prod_{i=1}^{M} \mathcal{N}\left(y_i \mid \boldsymbol{\phi}_i \mathbf{z}, \tau^{-1}\right) \tag{10}$$

The **prior on the precision** τ of noise follows a Gamma distribution governed by hyper-parameters α_0 (shape parameter) and β_0 (rate parameter)

$$\begin{aligned} P(\tau \mid \alpha_0, \beta_0) &= \text{Gamma}(\tau \mid \alpha_0, \beta_0) \\ &= \beta_0^{\alpha_0} \frac{1}{\Gamma(\alpha_0)} \tau^{\alpha_0 - 1} e^{-\beta_0 \tau} \end{aligned} \tag{11}$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$.

The **prior on the coefficients** $\mathbf{z}$ is given by a Student-t distribution governed by hyper-parameters μ_0 (mean value), $\nu_0 > 0$ (degree of freedom) and $\sigma_0^2 > 0$ (scale parameter)

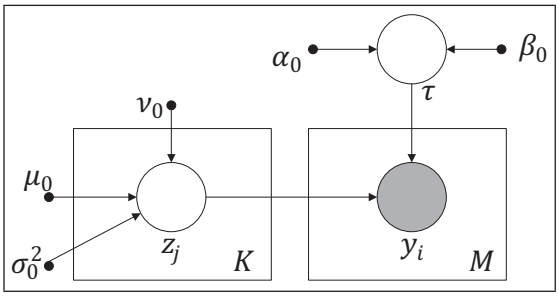

Figure 3: Representation of the generative model as a directed acyclic graph. The observed variable y_i is shown by the shaded node, while the latent variables z_j and τ are shown by circles. The right box represents the M independent measurements, while the left box represents the K independent DCT coefficients. $\alpha_0, \beta_0, \mu_0, \sigma_0$, and ν_0 are hyper-parameters.

$$\begin{aligned} P(z_j \mid \mu_0, \nu_0, \sigma_0^2) &= t_{\nu_0}(z_j \mid \mu_0, \sigma_0^2) \\ &= c\left[1 + \frac{1}{\nu_0}\left(\frac{z_j - \mu_0}{\sigma_0}\right)^2\right]^{-\frac{\nu_0+1}{2}} \\ \text{where } c &= \frac{\Gamma(\nu_0/2 + 1/2)}{\Gamma(\nu_0/2)} \frac{1}{\sqrt{\nu_0 \pi}\sigma_0} \end{aligned} \tag{12}$$

The K sparse coefficients are independent with each other

$$P(\mathbf{z} \mid \mu_0, \nu_0, \sigma_0^2) = \prod_{j=1}^{K} P(z_j \mid \mu_0, \nu_0, \sigma_0^2) \tag{13}$$

Armed with the probabilistic formulation of BCST, the predictive density can be obtained by marginalizing over the latent variables (parameters become latent variables after applying prior dis-

tributions on them), so that

$$P(\hat{y}_i \mid \mathbf{y}) = \int P(\hat{y}_i \mid \boldsymbol{\theta}) P(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta} = \int \mathcal{N}(\hat{y}_i \mid \boldsymbol{\phi}_i \mathbf{z}, \tau^{-1}) P(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta} \tag{14}$$

where $\boldsymbol{\theta} = \{\mathbf{z}, \tau\}$ denotes the latent variables. To implement this framework, we must address the formulation of a tractable procedure for marginalization over the posterior distribution. Since the exact computation will inevitably cause high computational cost, approximation is required.

3.2 Approximated Variational Inference

Starting from Eq. (10), the evidence (marginal likelihood) of BCST is obtained by averaging the likelihood under priors for the latent variables (parameters which have their own hyper-parameters)

$$P(\mathbf{y}) = \int P(\boldsymbol{\theta}) P(\mathbf{y} \mid \boldsymbol{\theta}) d\boldsymbol{\theta} \tag{15}$$

Since the exact inference is intractable, we need to resort to some form of approximation. Let q be any functional form of $\boldsymbol{\theta}$ satisfying $\int q(\boldsymbol{\theta}) d\boldsymbol{\theta} = 1$ and $q(\boldsymbol{\theta}) > 0$. By Jensen's inequality, we have

$$\ln P(\mathbf{y}) = \ln \int P(\mathbf{y}, \boldsymbol{\theta}) d\boldsymbol{\theta} \geq \int q(\boldsymbol{\theta}) \ln \frac{P(\mathbf{y}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \equiv \mathcal{F}(q) \tag{16}$$

where $\mathcal{F}(q)$ forms a lower bound of $\ln P(\mathbf{y})$. The significance of this approximation is that the lower bound might be tractable by selecting suitable $q(\boldsymbol{\theta})$ although the original marginal log likelihood is not. Note that maximizing $\mathcal{F}(q)$ is equivalent to minimizing the KL-divergence between $q(\boldsymbol{\theta})$ and $P(\boldsymbol{\theta} \mid \mathbf{y})$ [4].

Without the loss of generality, let $q_j = q(z_j)$ for $j = 1, 2, \cdots, K$ and let $q_{K+1} = q(\tau)$. Similar to [4], we consider to restrict the family of distributions $q(\boldsymbol{\theta})$ by factorizing the functional form

$$q(\boldsymbol{\theta}) = \prod_{j=1}^{K+1} q_j \tag{17}$$

The form to maximize $\mathcal{F}(q)$ with respect to q_k is given by

$$\mathcal{F}(q) = \int \prod_j q_j \left\{ \ln P(\mathbf{y}, \boldsymbol{\theta}) - \ln q(\boldsymbol{\theta}) \right\} d\boldsymbol{\theta} = -\mathrm{KL}\left(q_k \| \mathbb{E}_{j \neq k}[\ln P(\mathbf{y}, \boldsymbol{\theta})]\right) + \text{const} \tag{18}$$

where $\mathbb{E}_{j \neq k}[P(\mathbf{y}, \boldsymbol{\theta})]$ is the expectation of $P(\mathbf{y}, \boldsymbol{\theta})$ with respect to the q distributions over all variables θ_j for $j \neq k$. Then the optimal q_k is

$$q_k^\star = \frac{\exp(\mathbb{E}_{j \neq k}[\ln P(\mathbf{y}, \boldsymbol{\theta})])}{\int \exp(\mathbb{E}_{j \neq k}[\ln P(\mathbf{y}, \boldsymbol{\theta})]) d\theta_k} \tag{19}$$

We can easily observe that $q_k^\star$ will have the same functional forms with its priors for conjugate models. However, Student-t priors are not conjugate, which leads to an analytical intractable model and infeasible for learning. Note that the Laplace priors on $\mathbf{z}$ (see Eq. (7)) are also not conjugate, two and three stage hierarchical priors are respectively considered for approximating Laplace in [15] and [3]. Instead, we consider a much faster and more effective way for the approximation of Student-t:

$$\ln q^\star(z_j) = -\frac{\nu_0 + 1}{2} \ln\left[1 + \frac{1}{\nu_0}\left(\frac{z_j - \mu_0}{\sigma_0}\right)^2\right] - \frac{\mathbb{E}_\tau[\tau]}{2} \times \sum_{i=1}^{M} [(\phi_{ij} z_j)^2 + 2 \sum_{k \neq j} \phi_{ik} z_k \phi_{ij} z_j - 2 y_i \phi_{ij} z_j] + \text{const} \tag{20}$$

To deal with the intractability, under the condition that $\frac{1}{\nu_0}\left(\frac{z_j - \mu_0}{\sigma_0}\right)^2 = \epsilon_1 \to 0$, based on Taylor series expansion, we have

$$\ln(1 + \epsilon_1) = \frac{1}{\nu_0}\left(\frac{z_j - \mu_0}{\sigma_0}\right)^2 + O(\epsilon_1^2) \approx \frac{1}{\nu_0}\left(\frac{z_j - \mu_0}{\sigma_0}\right)^2$$

Then Eq. (20) is simplified to be

$$\ln q^\star(z_j) = -\frac{\nu_N + 1}{2}\left[\frac{1}{\nu_N}\left(\frac{z_j - \mu_N}{\sigma_N}\right)^2\right] \tag{21}$$

where

$$\mu_N = \frac{\mu_0(\nu_0 + 1) + \varphi_2 \nu_0 \sigma_0^2}{\nu_N + 1}, \qquad \sigma_N^2 = \frac{\nu_0 \sigma_0^2}{\nu_N}, \qquad \nu_N = \varphi_1 \nu_0 \sigma_0^2 + \nu_0 \tag{22}$$

and the parameters in Eq. (22) are given:

$$\varphi_1 = M \overline{\phi_{\bar{i}\bar{j}}^2} \mathbb{E}_\tau[\tau],$$

$$\varphi_2 = -\mathbb{E}_\tau[\tau] M \overline{(K-1)\phi_{i\bar{k}} z_{\bar{k}} \phi_{ij}} + \mathbb{E}_\tau[\tau] M \overline{y_{\bar{i}} \phi_{\bar{i}j}}$$

$$\overline{\phi_{\bar{i}\bar{j}}^2} = \frac{1}{M}\sum_{i=1}^{M} \phi_{ij}^2, \qquad \overline{\phi_{\bar{i}j}} = \frac{1}{M}\sum_{i=1}^{M} \phi_{ij},$$

$$\overline{\phi_{i\bar{k}} z_{\bar{k}}} = \frac{1}{K-1}\sum_{k \neq j} \phi_{ik} z_k = \frac{1}{K-1}(\boldsymbol{\phi}_i \mathbf{z} - \phi_{ij} z_j),$$

$$\overline{(K-1)\phi_{i\bar{k}} z_{\bar{k}} \phi_{\bar{i}j}} = \frac{1}{M}\sum_{i=1}^{M} (K-1)\overline{\phi_{i\bar{k}} z_{\bar{k}}} \phi_{ij},$$

$$\overline{y_{\bar{i}} \phi_{\bar{i}j}} = \frac{1}{M}\sum_{i=1}^{M} y_i \phi_{ij}$$

Then with $\frac{1}{\nu_N}\left(\frac{z_j - \mu_N}{\sigma_N}\right)^2 = \epsilon_2 \to 0$, use Taylor series again

$$\frac{1}{\nu_N}\left(\frac{z_j - \mu_N}{\sigma_N}\right)^2 \approx \epsilon_2 + O(\epsilon_2^2) = \ln(1 + \epsilon_2)$$

Eq. (21) thus can be converted to

$$\ln q^\star(z_j) = -\frac{\nu_N + 1}{2} \ln\left[1 + \frac{1}{\nu_N}\left(\frac{z_j - \mu_N}{\sigma_N}\right)^2\right] + \text{const} \tag{23}$$

Following Eq. (19), the optimal $q(\tau)$ can be easily achieved since it's in conjugate form

$$\ln q^\star(\tau) = \alpha_N \ln \tau - \beta_N \tau + \text{const} \tag{24}$$

Algorithm 1 Geospatial Interpolation with BCST-AVI

Input: M geospatial locations $\{(i_1, j_1), \cdots, (i_M, j_M)\}$, and the observations $\mathbf{y} \in \mathbb{R}^{M \times 1}$ at these locations.
Output: The interpolated measurements $\mathbf{f} \in \mathbb{R}^{K \times 1}$ ($K \gg M$).

1: Construct the matrix $\boldsymbol{\Phi}$ based on Eq. (5).
2: Randomly initialize vector $\mathbf{z} \in \mathbb{R}^{K \times 1}$
3: **repeat**
4: Update μ_N, σ_N, ν_N using Eq. (22).
5: Update α_N, β_N using Eq. (25).
6: Evaluate the lower bound of model evidence, i.e., $\mathcal{F}(q)$, using Eq. (16).
7: **until** Convergence
8: Estimate the meteorological measurements at the other $(K - M)$ locations based on Eq. (27).

where

$$\alpha_N = \alpha_0 + \frac{M}{2}, \qquad \beta_N = \beta_0 + \frac{1}{2}\sum_{i=1}^{M} \mathbb{E}_{\mathbf{z}}\left[(y_i - \boldsymbol{\phi}_i \mathbf{z})^2\right] \quad (25)$$

The expected values in the above formulas are given by

$$\begin{aligned} \mathbb{E}_\tau[\tau] &= \frac{\alpha_N}{\beta_N} \ \ (\alpha_N > 0, \beta_N > 0) \\ \mathbb{E}_{z_j}[z_j] &= \mu_N \ \ (\nu_N > 1) \\ \mathbb{E}_{z_j}[z_j^2] &= \frac{\nu_N \sigma_N^2}{\nu_N - 2} + \mu_N^2 \ \ (\nu_N > 2) \end{aligned} \quad (26)$$

Then the optimal $q_k^\star$ can be achieved by iteratively updating the variational parameters μ_N, σ_N, ν_N (see Eq. (22)) and α_N, β_N (see Eq. (25)) until convergence. The theoretical analysis of AVI is shown in Section A.1.

Evaluation of Predictive Density: Based on Jensen's inequality, we consider to evaluate the lower bound of the predictive density

$$\begin{aligned} &\ln P(\hat{y}_i \mid \mathbf{y}) \approx \ln \int \mathcal{N}(\hat{y}_i \mid \boldsymbol{\phi}_i \mathbf{z}, \tau^{-1}) q(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &\geq \int q(\boldsymbol{\theta}) \ln \mathcal{N}(\hat{y}_i \mid \boldsymbol{\phi}_i \mathbf{z}, \tau^{-1}) d\boldsymbol{\theta} \\ &= \mathbb{E}_\tau[\ln \tau] - \frac{\mathbb{E}_\tau[\tau]}{2} \mathbb{E}_{\mathbf{z}}[(\hat{y}_i - \boldsymbol{\phi}_i \mathbf{z})^2] + \text{const} \end{aligned} \quad (27)$$

where $\mathbb{E}_\tau[\ln \tau]$ is the negative entropy of $q^\star(\tau)$, and $\mathbb{E}_\tau[\tau]$ is defined in Eq. (26). The approximation in Line 1 of Eq. (27) is due to using variational distributions to replace the exact posterior distribution. The inequality in Line 2 is a simple Jensen bound.

The procedure to use BCST-AVI (Bayesian Compressed Sensing using the Approximated Variational Inference) for geospatial interpolation is summarized in Algorithm 1.

3.3 Sampling based Inference

Alternatively, we use sampling method to approximate the posterior distributions $P(\boldsymbol{\theta} \mid \mathbf{y})$ as a benchmark for testing the correctness of AVI. Gibbs sampling on $P(\boldsymbol{\theta} \mid \mathbf{y})$ can be performed by flipping the latent variables:

$$\begin{aligned} &P(\tau \mid \mathbf{y}, \mathbf{z}) \propto \text{Gamma}(\tau \mid \alpha_s, \beta_s) \\ &P(z_j \mid \mathbf{y}, \mathbf{z} \setminus z_j, \tau) \propto \prod_{i=1}^{M} \mathcal{N}(y_i \mid \phi_i \mathbf{z}, \tau^{-1}) t_{\nu_0}(z_j \mid \mu_0, \sigma_0^2) \end{aligned} \quad (28)$$

where $\alpha_s = \alpha_0 + \frac{M}{2}, \beta_s = \beta_0 + \frac{1}{2}\sum_{i=1}^{M}(y_i - \phi_i \mathbf{z})^2$. Since $P(z_j \mid \mathbf{y}, \mathbf{z} \setminus z_j, \tau)$ might be non log-concave, we apply Adaptive Rejection Metropolis Sampling (ARMS) for the generation of z_j [10]. Then the predictive density is evaluated by

$$P(\hat{y}_i \mid \mathbf{y}) \simeq \frac{1}{T-b} \sum_{t=s+1}^{T} \mathcal{N}(\hat{y}_i \mid \boldsymbol{\phi}_i \mathbf{z}^{(t)}, (\tau^{(t)})^{-1}) \quad (29)$$

where T is the number of total generated samples, and b is the number of abandoned starting samples, targeting for reducing the effects of unreasonable initializations.

4. EMPIRICAL EVALUATIONS

In this section, we perform the geospatial interpolation task to demonstrate the advantages of the proposed model compared to the baselines.

Table 2: Data Sets

Dataset	Metrics	Sampling Rate
KNMI	Temperature	Daily Reporting
Vendor	Temperature, Humidity, Barometric Pressure	5 Minutes

4.1 Experimental Setup

We use 2 general meteorological data sets which are summarized in Table 2. KNMI is available online [16], which contains about 12 years (2001/05 – 2013/05) climate data for 33 weather stations spreading throughout Netherlands. We use 11×17 grid to split Netherlands into equal-sized cells (edge length of one cell is about 25KM and the 33 stations can be fully separated), and consider one daily snapshot as a sample. The other dataset is provided by the vendor of IBM (Vendor dataset), reporting the climate records of 14,290 weather stations in North United States for almost 3 years (2009/04 – 2012/03). Since the entire North United States is too large for one sample, we consider a 20×20 equal-sized grid as a sample for this dataset, where edge length of each cell is about 10KM.

We compare the proposed models BCST-AVI (Bayesian Compressed Sensing using student-T priors with Approximated Variational Inference), BCST-GS (Bayesian Compressed Sensing using student-T priors with Sampling based inference) with BCSL (Bayesian Compressed Sensing using Laplace prior), CS (Compressed Sensing), TPS (Thin Plate Spline), and UK (Universal Kriging). The BCSL is implemented based on the package VB-BCS, where the Laplace prior is modelled in a hierarchical way [15]. The CS is extended to perform geospatial interpolation and implemented based on Virtual Probe (VP) solver [31]. TPS is implemented with R package "fields" [9], while UK is implemented with R package "DiceKriging" [22]. All tests run on an Intel i7-2760QM processor using up to eight cores (2.40 GHz).

Average Normalized Error (ANE) is used to measure the interpolation performance, i.e.,

$$\text{ANE} = \sqrt{\frac{\sum_i \sum_j \left[g(i,j) - \hat{g}(i,j)\right]^2}{\sum_i \sum_j g(i,j)^2}} \quad (30)$$

where $g(i,j)$ is the observed values at the evaluation weather stations, and $\hat{g}(i,j)$ is the estimated value.

4.2 Quantitative Evaluation of the Geospatial Interpolation Methods

Geospatial Interpolation Results on KNMI: With respect to KNMI, 33 weather stations almost uniformly locate in a spatial region split with a 17×11 equal-sized grid. To evaluate the performance of models with different M, we vary the number of measurements from 6 to 16 (about from 20% to 50% of 33 weather stations), and the residual sites are used for computing the ANE values. For a certain M, we repeat for 100 runs. Table 3 shows the mean±std (standard derivation) of ANE values for all models.

Geospatial Interpolation Results on Vendor Dataset: In terms of Vendor dataset, there are about 170 weather stations in a grid of size 20×20 equal-sized cells. We vary the number of measurements from 20 to 120 (about from 10% to 70% of 170 weather stations) to show how performance of models changes with M, while the remaining stations are used for evaluations. Similarly, 100 runs are repeated for any certain M. As shown in Table 4, three rows of mean±std of ANE values are reported for any M: the first row is

for metric Temperature, the second is for Humidity, while the last row is for Barometric Pressure.

As shown in Tables 3 and 4, regardless of the datasets and metrics, the proposed two models (BCST-AVI and BCST-GS) significantly outperform TPS and UK, and are competitive with BCSL and CS. Specifically, under most cases, both BCST-AVI and BCST-GS perform better than BCSL and CS, indicating that the Student-t priors can more accurately model the sparsity of climate data's coefficients. We also observe that the performance of BCST-AVI, BCST-GS, BCSL and CS slightly increases with the growth of M while that of TPS and UK significantly increase, and the best performance of TPS and UK is chasing the worst of other four methods for many cases; thus conclude that the compressed sensing based methods can achieve good performance with much fewer measurements than TPS and UK. The similar performance of BCST-AVI and BCST-GS validates that the Approximated Variational Inference (AVI) can achieve as good approximation as Gibbs sampling based inference.

4.3 Efficiency Evaluation

We study the behaviour of BCST-AVI, BCST-GS, CS, TPS and UK in terms of running time. Considering the worst case for time consumption, M is set to be 16 for KNMI, and 120 for Vendor dataset. Since the varying metrics (i.e., temperature, humidity and barometric pressure) lead to very small change in the time consumption, we do not distinct them here. 10 runs are repeated to report the mean±std of the time consumption in millisecond (ms). As shown in Table 5, BCST-AVI consumes much less time than all other methods for both datasets, CS, TPS and UK can achieve acceptable performance and perform better than BCSL. As expected, BCST-GS costs quite more time than the others. Generally, increasing M (from KNMI's 16 to Vendor's 120) will inevitably cause more time consumption.

Table 5: Time consumption (mean±std, and the unit of time is ms) comparison between models: the first column is for KNMI, while the second is for Vendor dataset. The bold entries denote the best average performance.

Model \ Dataset	KNMI	Vendor
BCST-AVI	**14.162**±1.7789	**31.169**±2.0685
BCST-GS	14205±269.91	63375±1792.7
BCSL	116.20±3.1875	489.65±12.750
CS	69.073±30.499	276.31±30.312
TPS	74.000±6.6500	106.00±8.4327
UK	50.000±5.9470	160.00±10.824

4.4 Visualization of Geospatial Interpolation

Our system provides a function to visualize the geospatial interpolation results with a high resolution in a large spatial scale: using Vendor data, we are capable of estimating the meteorological measurements in a 10KM × 10KM resolution throughout USA. This allows decision makers to quickly identify the bottlenecks which limit the functionality of power grid. The detailed visualization process is implemented in a recursive way. Suppose the task is to interpolate and visualize the climate variables for USA. Since it's neither efficient nor accurate to perform interpolation with respect to the entire region, the first step we employed is to recursively grid the region until the edge-length of each cell is approximately 10KM. Then we select a size of $P \times Q$ adjacent cells as a sub-region and the size of sub-regions can vary to adapt to real scenarios. The sub-regions may also overlap with each other to reduce the edge effects. Moreover, for each sub-region, BSC-AVI is applied to perform interpolation, and this process can be concurrently executed. Finally, sub-regions' interpolation results are combined to produce the result for the whole region. Using the above stated procedure, the interpolated results would cover the sea or large water area, where the values are estimated based on the observations of sensors on land. Due to the inherent different properties of land and sea in determining climate conditions, this would inevitably lead to the inaccurate estimations for the sea area. We thus can ignore the results on the sea area in the following figures[2]. We respectively

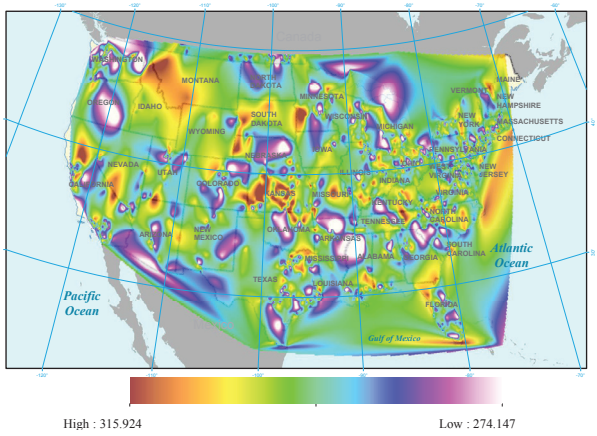

Figure 4: Example geospatial interpolation results: temperature map in USA.

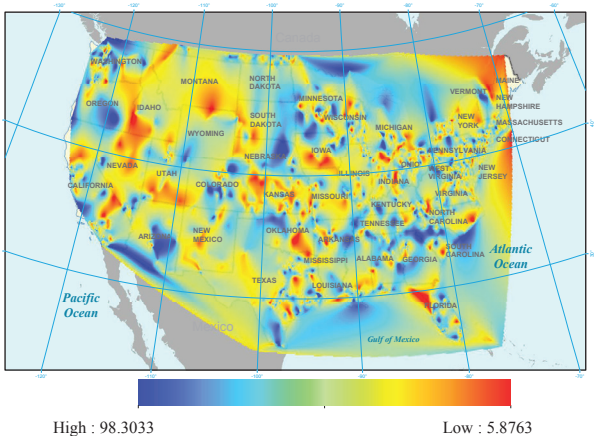

Figure 5: Example geospatial interpolation results: humidity map in USA.

show example interpolation results of temperature (in Kelvin), humidity (in %, i.e., relative humidity), barometric pressure (in 1/100 inch of mercury) in Figure 4, 5, and 6. The observations collected from nation-wide weather stations are used for interpolations. It is easy to identify the hot areas in USA (i.e., the areas with high temperature) from Figure 4. As illustrated in 5, the dry areas (i.e.,

[2]Since we don't have land boundary data, it's impossible to remove the estimations for sea area in the visualization process.

Table 3: Geospatial interpolation performance on KNMI dataset, reporting the ANE (mean±std) of Temperature (Kelvin) with varying M (the number of measurements). The bold entries denote the best average performance values.

Model / M	BCST-AVI	BCST-GS	BCSL	CS	TPS	UK
6	(**4.4655**±0.2090)%	(4.5271±0.2043)%	(6.0500±0.2360)%	(6.5341±0.2765)%	(24.075±1.0912)%	(21.531±1.2026)%
8	(4.1963±0.2071)%	(**4.0133**±0.2058)%	(5.5939±0.2126)%	(6.0756±0.2391)%	(15.071±1.1238)%	(14.295±1.2815)%
10	(3.9797±0.2077)%	(**3.9246**±0.2001)%	(5.0283±0.2247)%	(5.7360±0.2337)%	(11.676±1.0394)%	(10.578±1.2292)%
12	(**3.8415**±0.2070)%	(3.8518±0.2085)%	(4.9300±0.2586)%	(5.5593±0.2274)%	(10.209±0.9548)%	(8.0038±0.9005)%
14	(3.6969±0.2041)%	(**3.6728**±0.2097)%	(4.5629±0.2256)%	(5.3472±0.2199)%	(8.3562±0.4875)%	(6.4566±0.3853)%
16	(3.4824±0.2059)%	(**3.4815**±0.2056)%	(4.4169±0.2698)%	(5.1859±0.4248)%	(6.9671±0.4248)%	(5.2924±0.4530)%

Table 4: Geospatial interpolation performance on Vendor dataset, reporting the ANE (mean±std) of corresponding metrics with varying M (the number of measurements): the first row is for Temperature (Kelvin), the second is for Humidity (%), and the third is for Barometric Pressure (1/100 inch of mercury). The bold entries denote the best average performance values.

Model / M	BCST-AVI	BCST-GS	BCSL	CS	TPS	UK
20	(2.5943±0.2285)%	(**2.4585**±0.2056)%	(2.9598±0.2749)%	(3.0167±0.3075)%	(19.022±1.5085)%	(21.263±1.9554)%
	(28.172±2.7023)%	(**27.414**±1.6046)%	(28.112±2.6669)%	(28.685±2.5680)%	(37.721±3.2162)%	(34.141±2.3860)%
	(**0.6073**±0.0274)%	(0.6603±0.0244)%	(0.7334±0.0331)%	(0.7759±0.0363)%	(4.5072±0.2701)%	(5.2732±0.1194)%
40	(2.2124±0.2291)%	(**2.1119**±0.2051)%	(2.7849±0.3012)%	(2.9782±0.2752)%	(12.067±1.0984)%	(13.708±1.4975)%
	(27.518±2.3710)%	(**26.513**±1.5725)%	(28.047±2.3536)%	(28.404±2.2523)%	(32.695±2.6955)%	(30.075±2.2480)%
	(**0.6105**±0.0282)%	(0.6205±0.0201)%	(0.7095±0.0318)%	(0.7498±0.0326)%	(2.7117±0.1739)%	(3.4868±0.0989)%
60	(2.0481±0.2043)%	(**2.0115**±0.2061)%	(2.4077±0.2091)%	(2.5880±0.2055)%	(9.2703±1.0487)%	(10.455±1.0390)%
	(26.649±2.3825)%	(**26.092**±1.5882)%	(27.542±2.7628)%	(28.050±2.1569)%	(31.086±2.5170)%	(29.257±2.2803)%
	(0.6062±0.0199)%	(**0.5866**±0.0127)%	(0.6785±0.0294)%	(0.7009±0.0302)%	(1.9887±0.1311)%	(2.7652±0.0841)%
80	(**1.5401**±0.1692)%	(1.6952±0.1347)%	(2.1335±0.1047)%	(2.2983±0.1307)%	(7.5462±1.0250)%	(8.3303±1.0369)%
	(**25.582**±1.6172)%	(25.867±1.5653)%	(26.796±1.7270)%	(27.068±1.9557)%	(30.229±2.4530)%	(28.942±2.2088)%
	(0.6031±0.0132)%	(**0.5933**±0.0129)%	(0.6284±0.0275)%	(0.6729±0.0283)%	(1.5779±0.1094)%	(2.2954±0.0718)%
100	(1.5341±0.1056)%	(**1.4843**±0.1712)%	(1.8688±0.1323)%	(2.0106±0.1344)%	(6.4482±0.9551)%	(6.6483±0.9954)%
	(25.853±1.4301)%	(**25.828**±1.4718)%	(26.541±1.3728)%	(26.891±1.9638)%	(29.776±2.4438)%	(28.800±2.2599)%
	(0.5998±0.0135)%	(**0.5630**±0.0105)%	(0.6199±0.0294)%	(0.6706±0.0274)%	(1.2842±0.0910)%	(1.9843±0.0594)%
120	(1.5437±0.1257)%	(**1.4672**±0.1041)%	(1.7912±0.1507)%	(2.0146±0.1134)%	(5.4632±0.8838)%	(5.4822±0.7508)%
	(**25.469**±1.7490)%	(25.618±1.4718)%	(26.339±1.8365)%	(26.933±1.9116)%	(29.534±2.4075)%	(28.661±2.1423)%
	(0.5944±0.0119)%	(**0.5604**±0.0103)%	(0.6132±0.0265)%	(0.6573±0.0254)%	(1.0383±0.0767)%	(1.6784±0.0490)%

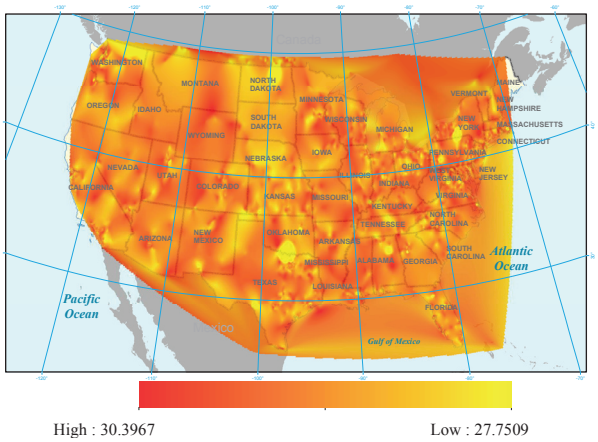

Figure 6: Example geospatial interpolation results: barometric pressure map in USA.

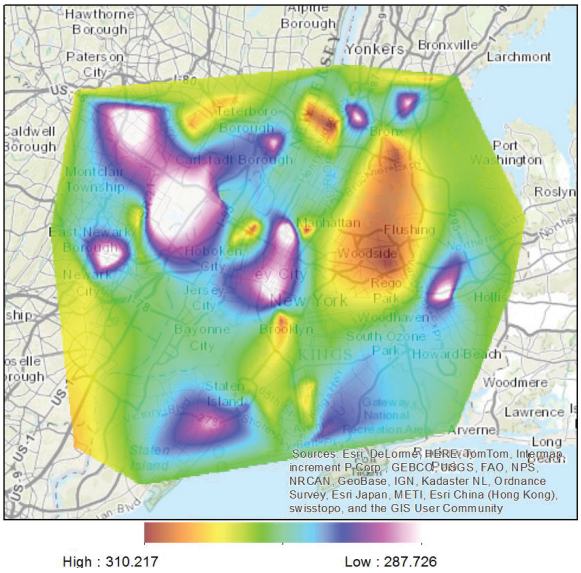

Figure 7: Example geospatial interpolation results: temperature map in New York City.

the areas with low humidity) have covered most of the deserts. It is expected that the barometric pressure changes very slowly with spatial locations as shown in Figure 6. The example interpolation results in New York City for temperature, humidity and barometric

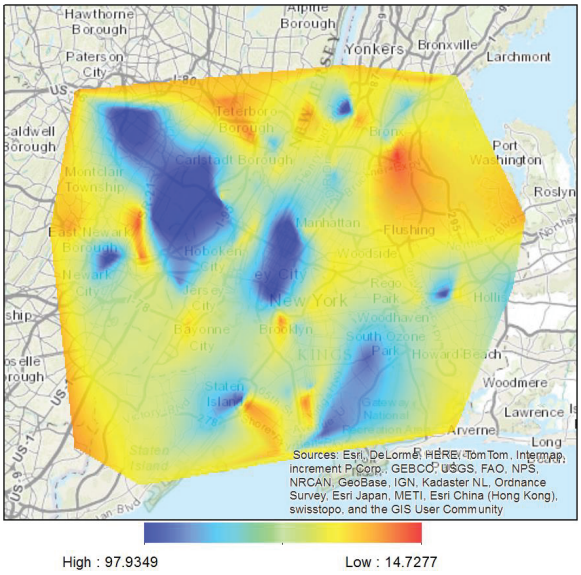

Figure 8: Example geospatial interpolation results: humidity map in New York City.

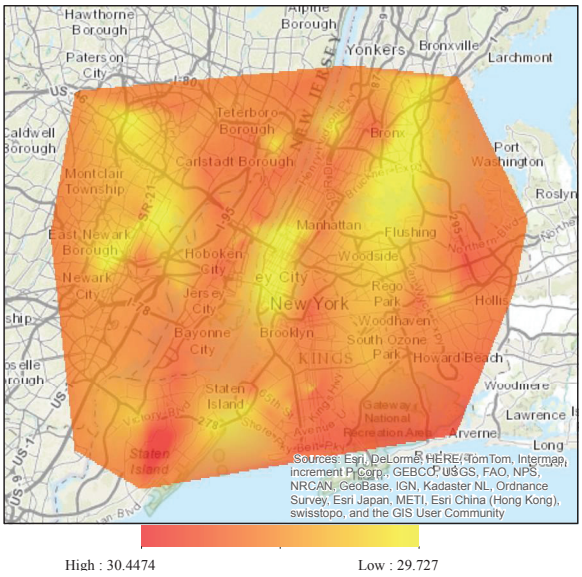

Figure 9: Example geospatial interpolation results: barometric pressure map in New York City.

pressure are respectively shown in Figure 7, 8 and 9. The city level maps of meteorological measurements convey more detailed information, which can be potentially used for city planning and guiding users to find places with pleasant weathers.

5. CONCLUSIONS

In this paper, we propose a new method to perform geospatial interpolation for large-scale meteorological measurements. Specifically, Student-t prior is incorporated into a Bayesian compressed sensing model to capture the distributions of coefficients, and the corresponding fast inference algorithm is provided. Using two real world datasets, we show that the proposed models are capable of effectively and efficiently estimating the values at unknown places with only a few measurements, while significantly outperformed the state-of-the-art methods. At the same time, our deployed system is of critical importance in supporting the intelligent management of electrical grid by providing the ambient climate conditions of power plants, transformers and transmission lines in high spatiotemporal resolutions.

References

[1] O. AntonicÂť, J. KrizËĞan, A. Marki, and D. Bukovec. Spatio-temporal interpolation of climatic variables over large region of complex terrain using neural networks. *Ecological Modelling*, 138:255 – 263, 2001.

[2] M. Arshad, S. M. Islam, and A. Khaliq. Power transformer aging and life extension. In *IEEE Conference on Probabilistic Methods Applied to Power Systems*, pages 498–501, 2004.

[3] S. Babacan, R. Molina, and A. Katsaggelos. Bayesian compressive sensing using laplace priors. *IEEE Transactions on Image Processing*, 19(1):53–63, 2010.

[4] A. Corduneanu and C. M. Bishop. Variational bayesian model selection for mixture distributions. In *Artificial Intelligence and Statistics*, pages 27–34, 2001.

[5] U. S. Department of Energy. Smart grid, 2012.

[6] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[7] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25:83–91, 2008.

[8] N. Farouk, L. Sheng, and Q. Hayat. Effect of ambient temperature on the performance of gas turbines power plant. *International Journal of Computer Science Issues (IJCSI)*, 10(1), 2013.

[9] R. Furrer, D. Nychka, and S. Sain. fields, 2013.

[10] W. R. Gilks, N. G. Best, and K. K. C. Tan. Adaptive rejection metropolis sampling within gibbs sampling. *Journal of the Royal Statistical Society*, 44(4):455–472, 1995.

[11] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (2nd ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2002.

[12] L. Hernández, C. Baladrón, J. M. Aguiar, L. Calavia, B. Carro, A. Sánchez-Esguevillas, D. J. Cook, D. Chinarro, and J. Gómez. A study of the relationship between weather variables and electric power demand inside a smart grid /smart world framework. *Sensors*, 12(9):11571–11591, 2012.

[13] J. R. Hipp, D. J. Bauer, P. Curran, and K. A. Bollen. Crimes of opportunity or crimes of emotion? testing two explanations of seasonal change in crime. *Social Forces*, 82:1333–1372, 2004.

[14] S. J. Jeffrey, J. O. Carter, K. B. Moodie, and A. R. Beswick. Using spatial interpolation to construct a comprehensive archive of australian climate data. *Environmental Modelling & Software*, 16(4):309–330, 2001.

[15] S. Ji, Y. Xue, and L. Carin. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356, 2008.

[16] KNMI. www.knmi.nlclimatologydaily_data.

[17] M. Knotters, D. Brus, and J. O. Voshaar. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma*, 67(3âĂŞ4):227–246, 1995.

[18] M. Lustig, D. Donoho, and J. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Society of Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[19] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab. *Signals and systems (2nd ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.

[20] R. B. Primacka, H. Higuchib, and A. J. Miller-Rushing. The impact of climate change on cherry trees and other species in japan. *Biological Conservation*, 142:1943–1949, 2009.

[21] P. F. D. Roosevelt. National weather service. *Disasters, Accidents, and Crises in American History: A Reference Guide to the Nation's Most Catastrophic Events*, page 1969, 2008.

[22] O. Roustant, D. Ginsbourger, and Y. Deville. Dicekriging, 2013.

[23] A. Tait, R. Henderson, R. Turner, and X. Zheng. Thin plate smoothing spline interpolation of daily rainfall for new zealand using a climatological rainfall surface. *International Journal of Climatology*, 26(14):2097–2115, 2006.

[24] G. Tan, R. Shibasaki, and K. S. Rajan. The study of global land suitability evaluation: A case of potential productivity estimation for wheat. *International Archives of Photogrammetry and Remote Sensing*, 33:1045–1050, 2000.

[25] P. J. Tavner, D. M. Greenwood, M. W. G. Whittle, R. Gindele, S. Faulstich, and B. Hahn. Study of weather and location effects on wind turbine failure rates. *Wind Energy*, 16(2):175–187, 2013.

[26] M. E. Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.

[27] L. A. Treinish, A. P. Praino, and Z. D. Christidis. Implementation of mesoscale numerical weather prediction for weather-sensitive business operations. In *19th International Conference on IIPS*, 2003.

[28] R. Walling and G. B. Shattuck. Distribution transformer thermal behavior and aging in local-delivery distribution systems. In *Proceedings of the 19th Conference on Electricity Distribution, Vienna*, 2007.

[29] B. Wang and J. Xiong. Appendix: Novel geospatial interpolation analytics for general meteorological measurements. *http://filebox.vt.edu/users/claren89/Publications/2014/kdd2014_more.pdf*, 2014.

[30] Y. Yan, Y. Qian, H. Sharif, and D. Tipper. A survey on smart grid communication infrastructures: Motivations, requirements and challenges. *Communications Surveys & Tutorials, IEEE*, 15(1):5–20, 2013.

[31] W. Zhang, X. Li, and R. Rutenbar. Virtual probe solver, 2010.

APPENDIX

A. DERIVATIONS AND MORE RESULTS

The detailed derivations and additional experimental evaluations can be found at [29].

A.1 Proposition and Proof

PROPOSITION 1. *For any $j = 1, 2, \cdots, M$, following the schema defined in Eq.* (19)*, if $\epsilon_1 \to 0$ and $\epsilon_2 \to 0$, then $q^\star(z_j)$ follows a Student-t distribution defined in Eq.* (23)*.*

Proof: Without the loss of generality, let's take any $j = 1, 2, \cdots, M$, and we know that

$$\ln q^\star(z_j) = -\frac{\nu_0+1}{2}\ln[1+\frac{1}{\nu_0}(\frac{z_j-\mu_0}{\sigma_0})^2] - \frac{\mathbb{E}_\tau[\tau]}{2}\times \sum_{i=1}^{M}[(\phi_{ij}z_j)^2 + 2\sum_{k\neq j}\phi_{ik}z_k\phi_{ij}z_j - 2y_i\phi_{ij}z_j] + \text{const} \tag{31}$$

Note that in the following derivations, we will neglect the const factor in Eq (31) since it has nothing to do with z_j. Since $\frac{1}{\nu_0}\left(\frac{z_j-\mu_0}{\sigma_0}\right)^2 = \epsilon_1 \to 0$, applying Taylor series, we have

$$\ln(1+\epsilon_1) = \frac{1}{\nu_0}(\frac{z_j-\mu_0}{\sigma_0})^2 + O(\epsilon_1^2) \tag{32}$$

Substituting Eq. (32) into Eq. (31), we get

$$\begin{aligned}
&\ln q^\star(z_j)\\
&= -\frac{1}{2}\left(\varphi_1 + \frac{\nu_0+1}{\nu_0\sigma_0^2}\right)z_j^2\\
&\quad + \left[\frac{\mu_0(\nu_0+1)}{\nu_0\sigma_0^2} + \varphi_2\right]z_j + O(\epsilon_1^2)\\
&= -\frac{1}{2}\left(\frac{\nu_N+1}{\nu_N\sigma_N^2}\right)z_j^2 + \left[\frac{\mu_N(\nu_N+1)}{\nu_N\sigma_N^2}\right]z_j + O(\epsilon_1^2)\\
&= -\frac{\nu_N+1}{2}\left[\frac{1}{\nu_N}\left(\frac{z_j-\mu_N}{\sigma_N}\right)^2\right] + O(\epsilon_1^2)
\end{aligned} \tag{33}$$

where μ_N, σ_N, ν_N are provided in Eq. (22) in the paper. Further, with the condition that $\frac{1}{\nu_N}\left(\frac{z_j-\mu_N}{\sigma_N}\right)^2 = \epsilon_2 \to 0$, applying Taylor series again, we know

$$\frac{1}{\nu_N}(\frac{z_j-\mu_N}{\sigma_N})^2 + O(\epsilon_2^2) = \epsilon_2 + O(\epsilon_2^2) = \ln(1+\epsilon_2) \tag{34}$$

Substituting Eq. (34) into Eq. (33) we get

$$\begin{aligned}
&\ln q^\star(z_j)\\
&= -\frac{\nu_N+1}{2}\ln\left[1+\frac{1}{\nu_N}\left(\frac{z_j-\mu_N}{\sigma_N}\right)^2\right]\\
&\qquad + \left[O(\epsilon_1^2) + O(\epsilon_2^2)\right]
\end{aligned} \tag{35}$$

As $\epsilon_1 \to 0$ and $\epsilon_2 \to 0$, the higher-order terms are negligible. We thus conclude

$$\begin{aligned}
&\ln q^\star(z_j)\\
&= -\frac{\nu_N+1}{2}\ln\left[1+\frac{1}{\nu_N}\left(\frac{z_j-\mu_N}{\sigma_N}\right)^2\right]\\
&= t_{\nu_N}(z_j \mid \mu_N, \sigma_N^2)
\end{aligned} \tag{36}$$