2016 IEEE 24th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems

Data Modelling with Gaussian Process in Sensor Networks for Urban Environmental Monitoring

Xiuming Liu*, Teng Xi†, Edith Ngai*
*Department of Information Technology
Uppsala University, Uppsala, Swden
Email: {xiuming.liu, edith.ngai}@it.uu.se
†State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications, Beijing, China
Email: xiteng@bupt.edu.cn

***Abstract*—In this paper, the multidimensional output Gaussian process (GP) is applied to model urban environmental data collected by sensor networks. Measurements from sensors at different locations are correlated. Moreover, we observe that the pollution level in urban area is highly coupled with human activities and shows periodic patterns accordingly. Based on these observations, we discuss the design of mean and kernel functions with two approaches: (1) composed kernel and maximum likelihood estimation of hyper-parameters; (2) Wiener-Khinchin theorem based approximation of sample covariances. To validate the models, the accuracy of interpolations given by different approaches are compared. The experimental results show that, for the application of interpolation, the dependent GP with the approximated sample covariances as kernels can provide better performance than the independent GP model with composed kernels.**

***Keywords*-data modelling, Gaussian process, kernel design**

I. Introduction

The global urbanization in recent decades has brought more than half of the global population into cities, and therefore produced many mega-cities worldwide. Urban environment pollution (especially air pollution) is one of the major challenges for sustainable growth of our society. The main air pollutants include carbon monoxide (CO), nitrogen dioxide (NO_2), and particulate matters (PM). According to reports from World Health Organization (WHO) [1], long-term exposure in highly polluted air environment leads to several health risks, and consequently produces significant economic costs for both individual citizen and the entire society. Therefore, air quality information is important for people who live in urban areas.

Many research projects have been conducted on developing urban environment monitoring systems [2, 3]. From the quality of service (QoS) perspective, challenges of designing sensor networks for urban environment monitoring exist in two major aspects: data accuracy and service coverage. The data accuracy is mostly determined by the quality of sensing devices. For instance, the performance of low-cost electrochemical sensors can be unstable (e.g., drift of sensitivity) when the operating condition varies. The challenge of service coverage appears when the energy and density of sensor network are constrained, and therefore available observations are often sparse in both time and space. Interpolation [4] is a common approach to improve the sensor network's coverage. Most of data processing methods require probabilistic models of the interested dynamics, which motivates our study in the this paper.

Gaussian process (GP) is a frequently used non-parametric model in analysis of time series data. A GP is uniquely defined by its mean and kernel functions. By carefully designing the mean and kernel functions, we are able to encode properties of the interested dynamics into the model. Moreover, the multidimensional output GP (or dependent GPs [5]) is capable of modelling the correlations between measurements from different locations. The generality and flexibility of GP makes it a powerful tool in time series analysis [6]. In this work, we apply GP to model air pollution data collected by sensor networks in urban area. First, the multidimensional output GP model is defined for sensor network data. Second, we present a kernel function design approach based on the Wiener-Khinchin theorem. Comparing to the likelihood based method, the Wiener-Khinchin theorem based method can be totally automatic and more computationally efficient.

The rest of this paper is organized as the following. In section II, we review the related work for sensor data modelling. A general GP model for sensor network data is presented in section III. Next, in section IV, we discuss the hidden periodic pattern in interested dynamics and different approaches of designing mean and kernel functions. Experimental results are shown in section V. Finally we conclude this paper in section VI.

II. Related Work

The data collected by different nodes in a sensor network are often correlated in both time and space domains. These correlations can be used as the side information when processing sensor network data. Such idea has been explored in the following previous works.

2375-0227/16 $31.00 © 2016 IEEE
DOI 10.1109/MASCOTS.2016.45

Deshpande *et al*. [7] proposed a Bayesian probabilistic model for data acquisition in sensor network, aiming to extend the lifetime of sensors by reducing the communication and data acquisition costs. Since then the model-driven data acquisition has become a popular approach in studies of sensor network. Raza *et al*. [8] examines this methodology and shows that the model driven approach does save the energy, conditioning on coordinated operations of sensor network. Thus, the critical task is to find a good probabilistic data model which characterizes the interested dynamics.

Gaussian process (GP) has been used widely in different applications [9]. There are two key aspects of modelling sensor network data with GP. Firstly, the GP model is required to encode both temporal and spatial correlations of sensor network data. One way of achieving this is to view the measurements from different nodes as dependent (or coupled) GPs [5]. In such cases, not only the auto-covariance but also the cross-covariance functions are calculated in the multidimensional output GP model. Secondly, appropriate mean and kernel functions must be designed for measurements from each sensor node (or pair of sensors in case of cross-covariance), because measurements from different sensors might have various patterns. Wilson summaries commonly used basic kernels as well as the method of constructing new kernels in his doctoral dissertation [10]. The evidence for designing kernel in the GP is usually the sample covariance and estimated power spectrum. In [11], the spectral mixture (SM) kernels is designed based on Gaussian mixture approximation of power spectrum density.

This paper distinguishes itself from the previous works in the following aspects. First, we introduce a general framework for sensor network data modelling based on multidimensional output GP. The temporal and spatial correlations of data are encoded in the multidimensional GP model simultaneously. Second, we focus on cases where the data collected by sensor network shows strong periodic patterns. We compare the traditional likelihood based kernel function design approach with the Wiener-Khinchin theorem based approach, and show that the latter case is more efficient to capture the hidden patterns and adapt to the training data.

III. System Model

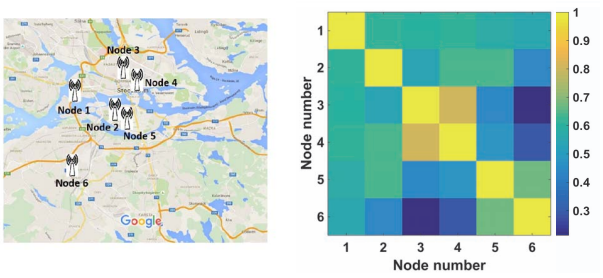
A. The Multidimensional Output GP

Let $X_t = \begin{bmatrix} x_t^1 & \dots & x_t^N \end{bmatrix}^T$ be the N-dimensional output of a GP at the input of time stamp t:

$$X_t \sim GP(\mathbf{m}(t), \mathbf{k}(t,t')), \tag{1}$$

where $\mathbf{m}(t) = \begin{bmatrix} m_{x^1}(t) & \dots & m_{x^N}(t) \end{bmatrix}^T$ is the vector of mean functions, and the matrix $\mathbf{k}(t,t')$ is given by

$$\begin{bmatrix} k_{x^1,x^1}(t,t') & k_{x^1,x^2}(t,t') & \dots & k_{x^1,x^N}(t,t') \\ k_{x^2,x^1}(t,t') & k_{x^2,x^2}(t,t') & \dots & k_{x^2,x^N}(t,t') \\ \vdots & \vdots & \ddots & \vdots \\ k_{x^N,x^1}(t,t') & k_{x^N,x^2}(t,t') & \dots & k_{x^N,x^N}(t,t') \end{bmatrix}. \tag{2}$$



(a) Locations of sensors. (b) Correlation matrix.

Figure 1: An example of sensor network for urban air quality monitoring and the scaled correlation matrix at $\Delta t = 0$.

The diagonal elements $k_{x^i,x^i}(t,t')$ $(i \in \{1,\dots,N\})$ are the auto-covariance functions; the non-diagonal elements $k_{x^i,x^j}(t,t')$ $(i,j \in \{1,\dots,N\}$ and $i \neq j)$ are the cross-covariances.

The model in Eq. (1) is also referred as N dependent (or coupled) GPs [5]. In real-life scenarios, an output x_t^i might be coupled with a subset of the remaining outputs, resulting in several elements in $\mathbf{k}(t,t')$ close to matrices of zeros. In the extreme case, an arbitrary dimension of outputs x_t^i is only dependent on itself. The matrix $\mathbf{k}(t,t')$ will therefore become diagonal and the coupled GP models becomes N independent one-dimensional output GPs.

B. Urban Sensor Network

Considering the planning of urban infrastructures (e.g., road systems) and different functions of districts of cities, such as residential and industrial, the correlations between sensors' outputs are more complicate than the a function of spatial distance. For example, two air pollution sensors deployed at interchanges of city highway system might provide highly correlated measurements, although their geographical distance can be relatively large; while measurements from other nodes, such as the ones deployed in the recreation area, can be less correlated with the roadside sensors. An example of deployment of air quality monitoring stations is showed in Fig. 1a.

Denoting the measurements from N sensors as $X_t = \begin{bmatrix} x_t^1 & \dots & x_t^N \end{bmatrix}^T$, the sensor data can be modelled as a N dimensional output GP in Eq. (1). The auto-covariances of measurements from a node i is characterized by the kernel $k_{x_t^i,x_t^i}(t,t')$. The cross-covariance of measurements from node i and node j is presented by the kernel $k_{x^i,x^j}(t,t')$. Fig. 1b illustrates the covariance matrix in Eq. (2) evaluated at zero time difference and scaled to $[0,1]$. The diagonal elements are auto-covariance functions evaluated with zero lag, resulting in maximum values. The non-diagonal elements present various level of correlation. For example, nodes 3 and 4 have highest correlation among all pairs of sensors.

The mean function $\mathbf{m}(t)$ and kernel matrix $\mathbf{k}(t,t')$ are obtained from the historical data of the sensor network (will be discussed in Section IV). Based on today's wireless communication and cloud computing technologies, it is feasible to collect data from the sensor network and update the GP model in a real time manner.

C. Joint Data Reconstruction

The urban sensor network for environmental monitoring usually has long sampling periods, especially in the case of battery powered sensor network where energy efficiency must be considered. Therefore, to reconstruct the sensory data from measurements with low sampling rate is an important task.

In the following example, assume that measurements $\mathbf{x}_\mathbf{t}^i$ and $\mathbf{x}_\mathbf{t}^j$ from node i and j are available with a vector of inputs $\mathbf{t}$, we are interested in inferring values of outputs $\mathbf{x}_{\mathbf{t}'}^i$ at unobserved inputs $\mathbf{t}'$ from node i. The joint probability of observations and unknown values, $p(\mathbf{x}_\mathbf{t}^i, \mathbf{x}_\mathbf{t}^j, \mathbf{x}_{\mathbf{t}'}^i)$, is a Gaussian distribution

$$\mathcal{N}(\begin{bmatrix}\mathbf{m}_i(\mathbf{t})\\ \mathbf{m}_j(\mathbf{t})\\ \mathbf{m}_i(\mathbf{t}')\end{bmatrix}, \begin{bmatrix}\mathbf{K}_{ii}(\mathbf{t},\mathbf{t}) & \mathbf{K}_{ij}(\mathbf{t},\mathbf{t}) & \mathbf{K}_{ii}(\mathbf{t},\mathbf{t}')\\ \mathbf{K}_{ji}(\mathbf{t},\mathbf{t}) & \mathbf{K}_{jj}(\mathbf{t},\mathbf{t}) & \mathbf{K}_{ji}(\mathbf{t},\mathbf{t}')\\ \mathbf{K}_{ii}(\mathbf{t}',\mathbf{t}) & \mathbf{K}_{ij}(\mathbf{t}',\mathbf{t}) & \mathbf{K}_{ii}(\mathbf{t}',\mathbf{t}')\end{bmatrix}), \quad (3)$$

and the conditional probability of unknown value is therefore given by

$$p(\mathbf{x}_{\mathbf{t}'}^i | \mathbf{x}_\mathbf{t}^i, \mathbf{x}_\mathbf{t}^j) \sim \mathcal{N}(\mu_{\mathbf{t}'|\mathbf{t}}, \mathbf{\Sigma}_{\mathbf{t}'|\mathbf{t}}), \quad (4)$$

where $\mu_{\mathbf{t}'|\mathbf{t}}$ in Eq. (5) is the mean of estimation and $\mathbf{\Sigma}_{\mathbf{t}'|\mathbf{t}}$ in Eq. (6) is the covariance matrix. And $\hat{\mathbf{x}}_{\mathbf{t}'}^i = \mu_{\mathbf{t}'|\mathbf{t}}$ is the estimated value of unknown output $\mathbf{x}_{\mathbf{t}'}^i$.

IV. Periodic Patterns and Kernel Design

A. Discover the Periodic Patterns

Before starting the discussion of designing mean and kernel functions for the GP, we looked into a sequence of the measurements[1], which are illustrated in Fig. 2. In this example, node 1 (solid blue) is a station deployed on top of a building's roof, measuring the background air quality; node 2 (dot red) is a station deployed on roadside in city centre, measuring the local air quality.

The time series illustrated in Fig. 2 has two properties. First, the value of this time series at any time stamp is a random number with a distribution (mean and variance). Hence the time series is stochastic process. Second, the time series shows periodic patterns. At least 3 periodic patterns can be discovered from the data showed in Fig. 2:

- Weekly pattern: the existence of weekly periodicity cane be observed in the figure. The NO_2 concentration is lower during weekends and higher during weekdays.
- Daily pattern: consider 24 hours as a period, the NO_2 concentration is lower at night and higher during daytime.
- Rush-hour pattern: hourly NO_2 measurements within a working day have two peaks, one is around 8:00 and the other is around 17:00. This is due to the large traffic volume and therefore high emission from vehicles.

In light of those three observations, next we discuss the design of mean and kernel functions for the GP model.

B. Design of Mean Function

Let's consider a time series with the periodic pattern of period T_p. Assume that a sequence of training data which contains N_p periods $\{x_{n*T_p+t} | n = 1, \ldots, N_p,\ t = 1, \ldots, T_p\}$ are available, the mean function of GP is designed to be

$$m_x(kT_p + t) = \frac{1}{N_p} \sum_{n=0}^{N_p-1} x_{n*T_p+t},\ \forall\ k \in \mathbb{Z}, \quad (7)$$

where $t \in \{1, \ldots, T_p\}$ are the time indexes within one period and k is the number of period. Equation (7) is usually referred as the sample average estimator for the mean function. The sample average is the minimum variance unbiased (MVU) estimator for mean values of Gaussian distribution [12]. In the example of NO_2 measurements, we separate the data set into weekdays and weekends, and then apply Eq. (7) to calculate the averaged hourly NO_2 measurements with $T_p = 24$ hours, which are illustrated in Fig. 3.

Compared to the hourly average NO_2 concentration during weekends, the process during weekdays shows the "double-peak" pattern, which is a reflection of rush-hours during a working day.

C. Design of Kernel Function

The kernel function $k_{x^i,x^j}(t,t')$ characterizes the correlation of time series samples with distance $\Delta t = t - t'$. In case of $i = j$, the kernel is also referred to the auto-covariance; in case $i \neq j$, the kernel represents the cross-covariance. We discuss two approaches in the following: (1) likelihood based approach, and (2) Wiener-Khinchin theorem based approach.

1) Likelihood Based Approach: The commonly used approach of kernel design is to first select a type of kernel, and then determine the value of hyper-parameters in the selected format via maximum likelihood estimation (MLE). This approach is briefly reported in the following. We refer readers to [9] for the details.

Assume that the stochastic process shows M periodic patterns with periods $\{\omega_1, \ldots, \omega_M\}$, a composed kernel [11] can be designed as

$$\sum_{m=1}^{M} \alpha_m^2 \exp[\frac{-\sin^2(\pi\omega_m(t-t'))}{2\theta_m^2}] + \alpha_0^2 \exp[\frac{-(t-t')^2}{2\theta_0^2}], \quad (8)$$

which is a superposition of M basic periodic kernels of different periodicities and a squared-exponential kernel.

[1]The data set is available at this website: http://slb.nu/elvf/

$$\mu_{\mathbf{t}'|\mathbf{t}} = \mathbf{m}_i(\mathbf{t}') + \begin{bmatrix}\mathbf{K}_{ii}(\mathbf{t}',\mathbf{t}) \\ \mathbf{K}_{ij}(\mathbf{t}',\mathbf{t})\end{bmatrix}^T \begin{bmatrix}\mathbf{K}_{ii}(\mathbf{t},\mathbf{t}) & \mathbf{K}_{ij}(\mathbf{t},\mathbf{t}) \\ \mathbf{K}_{ji}(\mathbf{t},\mathbf{t}) & \mathbf{K}_{jj}(\mathbf{t},\mathbf{t})\end{bmatrix}^{-1} \begin{bmatrix}\mathbf{x}_\mathbf{t}^i - \mathbf{m}_i(\mathbf{t}) \\ \mathbf{x}_\mathbf{t}^j - \mathbf{m}_j(\mathbf{t})\end{bmatrix}, \tag{5}$$

$$\mathbf{\Sigma}_{\mathbf{t}'|\mathbf{t}} = \mathbf{K}_{ii}(\mathbf{t}',\mathbf{t}') - \begin{bmatrix}\mathbf{K}_{ii}(\mathbf{t}',\mathbf{t}) \\ \mathbf{K}_{ij}(\mathbf{t}',\mathbf{t})\end{bmatrix}^T \begin{bmatrix}\mathbf{K}_{ii}(\mathbf{t},\mathbf{t}) & \mathbf{K}_{ij}(\mathbf{t},\mathbf{t}) \\ \mathbf{K}_{ji}(\mathbf{t},\mathbf{t}) & \mathbf{K}_{jj}(\mathbf{t},\mathbf{t})\end{bmatrix}^{-1} \begin{bmatrix}\mathbf{K}_{ii}(\mathbf{t},\mathbf{t}') \\ \mathbf{K}_{ji}(\mathbf{t},\mathbf{t}')\end{bmatrix}. \tag{6}$$

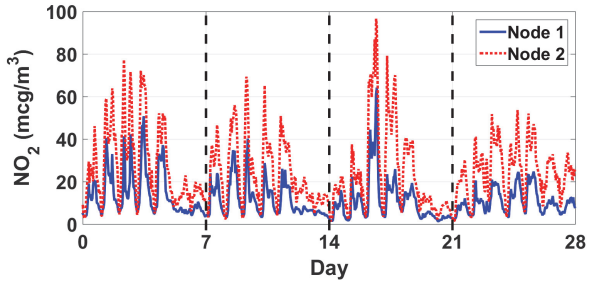


Figure 2: NO_2 measurements from 2 stations in 4 weeks.

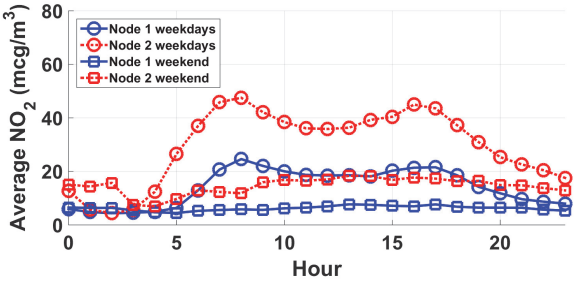


Figure 3: Sample average of hourly NO_2 measurements from node 1 and 2 during weekdays and weekends.

Next, to determine the value of parameters $\mathcal{H} = \{\alpha_0 \dots \alpha_M, \theta_0 \dots \theta_M\}$, we write the likelihood of hyper-parameters $\mathcal{H}$ given training data $\mathbf{X}$ (a sequence of measurements from N sensors):

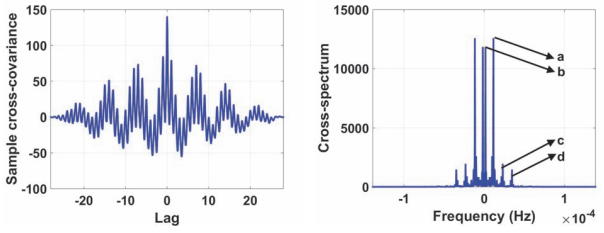
$$L(\mathcal{H}|\mathbf{X}) = -\frac{1}{2}[(\mathbf{X}-\mathbf{M})^T\mathbf{K}_\mathcal{H}^{-1}(\mathbf{X}-\mathbf{M}) + N\ln 2\pi + \ln|\mathbf{K}_\mathcal{H}|], \tag{9}$$

where $\mathbf{M}$ is the matrix of mean values for N sensors over the same time period as training data, and $\mathbf{K}_\mathcal{H}$ is the covariance matrix which is a function of hyper-parameters $\mathcal{H}$.

The maximum likelihood estimation of the hyper-parameters, $\hat{\mathcal{H}}_{\text{ML}}$, is therefore obtained via solving the following optimization problem:

$$\hat{\mathcal{H}}_{\text{ML}} = \arg\max_{\hat{\mathcal{H}}} L(\hat{\mathcal{H}}|\mathbf{X}), \tag{10}$$

which can be solved by gradient based algorithms. However, it is non-trivial and time-consuming to obtain the global optimum due to the non-convexity of the likelihood function and the large number of variables.



(a) Cross-covariance (b) Cross-spectrum

Figure 4: Frequency components a, b, c, and d are 1.157×10^-5, 1.653×10^-6, 2.315×10^-5, and 3.472×10^-5 (all in Hz), corresponding to 1 day, 1 week, 12 and 8 hours periodicities.

2) Wiener-Khinchin Theorem Based Approach: The Wiener-Khinchin theorem states that the auto-correlation and the power spectrum density (PSD) (or the cross-correlation and the cross-spectrum) of stochastic processes are a pair of Fourier transform. Based on this theorem, we introduce the Wiener-Khinchin theorem based approach in the following.

Assume that a set of historical data $\mathbf{x} = \{x_t^i, x_t^j | t = 1, \dots, T\}$ are available and the length of data set T is longer than the maximum period, we first calculate the sample covariance

$$\text{cov}(\Delta t) = \frac{1}{T}\sum_{t=1}^{T}(x_{t+\Delta t}^i - \text{E}[x_t^i])(x_t^j - \text{E}[x_t^j]), \tag{11}$$

where $\Delta t = -T+1 \dots T-1$ are the lags. In case of $i = j$, Eq. (11) is the auto-covariance for node i; otherwise it is the cross-covariance between nodes i and j.

Next, we apply discrete time Fourier transform (DTFT) to the sample covariances, resulting to the estimated PSD of the stochastic process:

$$\hat{R}(f) = \int_{-\infty}^{\infty} \text{cov}(\Delta t) e^{-j2\pi f \Delta t} df. \tag{12}$$

Note that since the covariance is time-discrete, therefore the estimated PSD function $\hat{R}_{\mathbf{xx}}(f)$ is periodic with period of f_s (the sampling frequency). We sample the first period of estimated PSD ($f \in [-\frac{f_s}{2}, \frac{f_s}{2}]$) to $2T-1$ frequency-discrete numbers

$$\hat{R}(k\frac{f_s}{2T}), \ \forall k = \{-T+1, \dots, T-1\}. \tag{13}$$

Finally the analytical expression of the estimated covariance function can be obtained via inverse discrete Fourier transform (IDFT) on the sampled estimated PSD

$$\begin{aligned} \hat{\text{cov}}(\Delta t) &= \frac{1}{2T-1} \sum_{k=-T+1}^{T-1} \hat{R}(k\frac{f_s}{2T}) e^{j2\pi k \frac{f_s}{2T}\Delta t} \\ &= A_0 + \sum_{k=1}^{T-1} 2A_k \cos(2\pi k \frac{f_s}{2T}\Delta t), \end{aligned} \tag{14}$$

where $A_k = \hat{R}(k\frac{f_s}{2T})/(2T-1)$, $\forall k \in \{0, \ldots, T-1\}$. The estimated covariance function is then used as the kernel function for GP

$$k(\Delta t) = \hat{\text{cov}}(\Delta t). \tag{15}$$

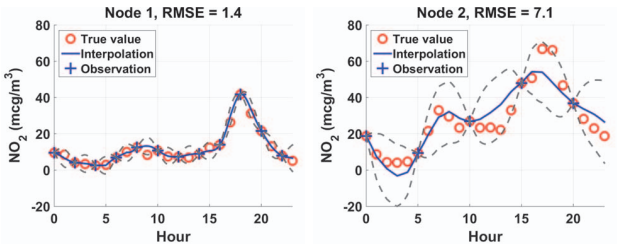
The proposed Wiener-Khinchin theorem based approach of designing kernel function has advantages from both theoretical and computational perspective. First, comparing to the manually selected type of kernel function, the spectrum based kernel is a better approximation of the true kernel. In fact, Eq. (14) is a projection of the sample covariance onto a space spanned by orthonormal basis (trigonometric functions). Secondly, the spectrum based approach can be implemented with fast Fourier transform (FFT), which is computationally efficient.

The cross-covariance and cross-spectrum of measurements from node 1 and 2 are shown in Fig. 4. From the power spectrum (Fig. 4b) we recognize 4 major periodic components of 1 week, 24 hours, 12 hours, and 8 hours periodicities, respectively. To apply the likelihood based approach, one might design a composed kernel by summing up a SE kernel and four periodic kernels, such as Eq. (8), where $\omega_1 = 1.653 \times 10^6$, $\omega_2 = 1.157 \times 10^5$, $\omega_3 = 2.315 \times 10^5$, and $\omega_4 = 3.472 \times 10^5$ (all in seconds). Hence there will be 10 hyper-parameters in one kernel function. Consider a multi-dimensional output GP, the number of variables in solving the MLE is very large, which makes the optimization problem (10) non-trivial and time-consuming to solve.

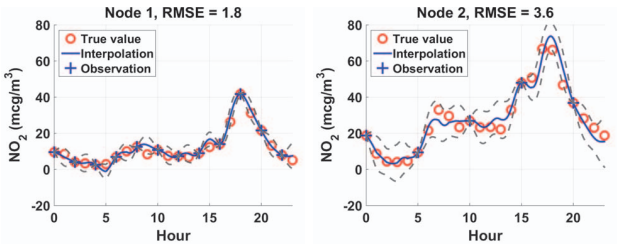
V. Experimental Results

Given a set of observations from the sensor networks, one can infer the values at unobserved time stamps. This task is referred as temporal interpolation. In this section, we use interpolation as an application to examine our model. The data set is downloaded from the hourly database of *the Stockholm - Uppsala County Air Quality Management Association*. 10 air quality monitoring stations are operated in the association, measuring the concentration of air pollutants such as NO_2, O_3, CO, and PM10. The hourly aggregated NO_2 data in 2014 is selected as an example.

We take the measurements from 2 nodes with most significant correlations as an illustration. Each node provides hourly aggregated NO_2 measurements (24 samples per day). To test our model, we first down-sample the measurements,



(a) Independent GPs.



(b) Dependent GPs.

Figure 5: A comparison between the model of independent GPs (5a), and the model of coupled GPs (5b). The true value (red circle), observation (blue cross), interpolation (blue solid), and 95% confidence interval (grey dash) are shown in each figure.

then interpolate missing values, and finally compare the interpolated values with the true values and calculate the root mean square error (RMSE): $\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{x}_t - x_t)^2}$, where T is the length of data to be interpolated, x_t is the true value, and $\hat{x}_t$ is the interpolated value at time stamp t.

A. Comparison Between Independent and Dependent GP

We begin with a comparison between the independent GP model and the dependent GP model. In the independent GP model, the cross-covariance matrices ($\mathbf{K}_{ij}(\mathbf{t}, \mathbf{t})$ and $\mathbf{K}_{ji}(\mathbf{t}, \mathbf{t})$ in (5) and (6)) are replaced by matrices zeros. In the coupled GP model, the cross-covariance matrices are calculated with kernel functions $k_{x^i,x^j}(t, t')$ and $k_{x^j,x^i}(t, t')$.

In this comparison, we down-sample the measurements from node 1 to $T_s = 2$ hour (12 samples per day), and measurements from node 2 to $T_s = 5$ hour (5 samples per day), then use both independent GP and dependent GP model to reconstruct the missing values. The results are shown in Fig. 5. We see that by coupling measurements from nodes 1 and 2, the RMSE of interpolation on node 2's measurements is significantly reduced from 7.1 to 3.6, while the RMSE for node 1's interpolation is slightly increased from 1.4 to 1.8.

B. Experiment on One-Year Data

Next we conduct an experiment on the performance of different GP models on the one-year hourly NO_2 measurements from nodes 3 and 4 showed in Fig. 1a. The covariance matrix

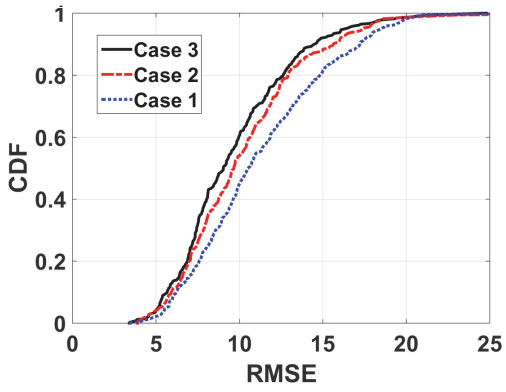


Figure 6: A comparison of cumulative distribution functions of RMSEs given by 3 experiment cases.

illustrated in Fig. 1b implies strong correlation ($0.8 \sim 0.9$) between the measurements of those two nodes. Therefore, we coupled the measurements from those two nodes as a dependent GP and perform joint data reconstruction after down-sampling the original hourly data to sampling period of 6 hours. In this experiment, we use the most recent 4-week historical data as train data to learn the mean and kernel functions, then perform the interpolation on one day's measurements and calculate the RMSE. This validation is repeated on a year's NO_2 measurements with three test cases:

- Case 1: independent GP model with composed kernel (SE + PE) in Eq. (8).
- Case 2: independent GP model with spectrum based kernel in Eq. (14).
- Case 3: dependent GP model with spectrum based kernel in Eq. (14).

The cumulative distribution function of RMSEs of three cases are illustrated in Fig. 6. The statistical results validate that the dependent GP model with spectrum based kernel achieves better performance comparing to other 2 cases.

VI. Conclusions

In this paper, we present a general framework for sensor network data modelling with multidimensional Gaussian process. Measurements from different sensors are correlated with various significances, which are characterized by the matrix of kernels (auto-covariances and cross-covariances). On top of commonly applied MLE method, we introduce a kernel design approach based on Wiener-Khinchin theorem. Experiments are conducted on the proposed GP model for sensor networks with the data set of NO_2 measurements in urban area. Results show that, comparing to the independent GP model, dependent GP model has better performance in terms of the accuracy for data reconstruction.

Acknowledgment

This work has been supported by the Vinnova GreenIoT project (2015-00347) in Sweden, the National Natural Science Foundation of China (No.61370197 and 61271041), and the China Scholarship Council. These supports are gratefully acknowledged.

References

[1] "Economic cost of the health impact of air pollution in europe: Clean air, health and wealth," World Health Organization, Reginal Office for Europe, 2015.

[2] K. Hu, V. Sivaraman, B. G. Luxan, and A. Rahman, "Design and evaluation of a metropolitan air pollution sensing system," *IEEE Sensors Journal*, vol. 16, no. 5, pp. 1448–1459, March 2016.

[3] X. Liu, Z. Song, E. Ngai, J. Ma, and W. Wang, "Pm2:5 monitoring using images from smartphones in participatory sensing," in *2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2015, pp. 630–635.

[4] J. Cortes, "Distributed kriged kalman filter for spatial estimation," *IEEE Transactions on Automatic Control*, vol. 54, no. 12, pp. 2816–2827, Dec 2009.

[5] P. Boyle and M. Frean, "Dependent gaussian processes," in *Advances in neural information processing systems*, 2004, pp. 217–224.

[6] R. H. Shumway and D. S. Stoffer, *Time series analysis and its applications: with R examples*, 3rd ed. New York: Springer, 2011.

[7] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, and W. Hong, "Model-driven data acquisition in sensor networks," in *Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30*, ser. VLDB '04. VLDB Endowment, 2004, pp. 588–599.

[8] U. Raza, A. Camerra, A. L. Murphy, T. Palpanas, and G. P. Picco, "What does model-driven data acquisition really achieve in wireless sensor networks?" in *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*, March 2012, pp. 85–94.

[9] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. Cambridge, Mass: MIT Press, 2006;2005;.

[10] A. G. Wilson, "Covariance kernels for fast automatic pattern discovery and extrapolation with gaussian processes," Ph.D. dissertation, PhD thesis, University of Cambridge, 2014.

[11] A. G. Wilson and R. Prescott Adams, "Gaussian Process Kernels for Pattern Discovery and Extrapolation," *ArXiv e-prints*, Feb. 2013.

[12] S. M. Kay, *Fundamentals of statistical signal processing: Vol. 1, Estimation theory*. Upper Saddle River, N.J: Prentice Hall PTR, 1993.