

PAPER

Traffic data reconstruction based on Markov random field modeling

To cite this article: Shun Kataoka *et al* 2014 *Inverse Problems* **30** 025003

View the [article online](#) for updates and enhancements.

You may also like

- [Inatorial forecasting method considering macro and micro characteristics of chaotic traffic flow](#)
Yue Hou, , Di Zhang et al.
- [Future Traffic Prediction from Short Period Traffic Data](#)
I Srinivasula Reddy and Vikas Mendi
- [A Fast Deep Learning Method for Network Intrusion Detection Without Manual Feature Extraction](#)
Wang Yue, Jiang Yiming and Lan Julong

Traffic data reconstruction based on Markov random field modeling

Shun Kataoka¹, Muneki Yasuda², Cyril Furtlehner³
and Kazuyuki Tanaka¹

¹ Graduate School of Information Science, Tohoku University, 6-3-09
Aramaki-aza-aoba, Aobaku, Sendai 980-8579, Japan

² Graduate School of Science and Engineering, Yamagata University, 4-3-16 Jonan,
Yonezawa, Yamagata 992-8510, Japan

³ INRIA Saclay, LRI, Bât. 660, Université Paris Sud, 91405, Orsay, Cedex, France

E-mail: xkataoka@smapi.is.tohoku.ac.jp

Received 28 June 2013, revised 18 November 2013

Accepted for publication 29 November 2013

Published 28 January 2014

Abstract

We consider the traffic data reconstruction problem. Suppose we have the traffic data of an entire city that are incomplete because some road data are unobserved. The problem is to reconstruct the unobserved parts of the data. In this paper, we propose a new method to reconstruct incomplete traffic data collected from various sensors. Our approach is based on Markov random field modeling of road traffic. The reconstruction is achieved by using a mean-field method and a machine learning method. We numerically verify the performance of our method using realistic simulated traffic data for the real road network of Sendai, Japan.

Keywords: Markov random fields, traffic data reconstruction, machine learning, Gaussian graphical model, probabilistic information processing

(Some figures may appear in colour only in the online journal)

1. Introduction

An intelligent transportation system (ITS) is a large-scale information system whose objective is to provide guidance information to drivers and optimize transportation traffic by analyzing vehicle traffic over an entire city. In order to provide accurate information, an ITS needs to collect accurate and comprehensive road traffic data. Due to the development of information and sensing technologies, various types of road traffic data, density, flow, speed and so on, can be collected from different sensing devices such as optical beacons and probe vehicles. These

sensors each have different features. For example, a beacon, which is a fixed type of traffic sensor, can steadily collect the traffic data of the road where it is located in a short time period; however, the detection area is narrow. A probe vehicle, which is a GPS-equipped vehicle, can collect the traffic data of a comprehensive area, but cannot collect the data steadily and needs a long time period to acquire comprehensive traffic data. Therefore, the fusion of various data collected from different sensors for the purpose of traffic prediction has recently attracted much attention [1].

Traffic prediction is a major research topic in the machine learning field. In fact, the analysis of freeway traffic has been researched since the 1970s [2]. Travel time prediction [3], density prediction [4] and route planning [5] are other active topics. In the machine learning approach, the existence of two databases, real time (RDB) and historical (HDB), is assumed. An RDB consists of road traffic data collected from sensors at the present time and represents a situation for which traffic prediction is required. An HDB contains road traffic data collected from sensors and traffic surveys in the past and is used to facilitate the prediction. That is, we use an HDB for learning and make a traffic prediction based on an RDB.

There still remains an important problem related to traffic prediction based on an RDB. The quality of the prediction depends on the quality of the RDB. However, the complete traffic data of an entire road network cannot be acquired since sensors are not installed on all roads. In fact, only 22% of the total length of trunk roads in Nagoya, Japan is covered by beacons [6]. Furthermore, at the present time, there are not enough probe vehicles to allow sufficient data to be acquired. If the number of probe vehicles in Japan were a hundred thousand, on average, it would take an hour to acquire one or two traffic data of an entire road network [7]. Therefore, in practice, it is difficult to collect sufficiently comprehensive road traffic data in a short time period to make a traffic prediction. Therefore, a method to reconstruct the unobserved parts in an RDB is required to solve a realistic traffic prediction problem. Recently, some researchers have tackled this problem. Kumagai *et al* proposed a method to reconstruct the traffic data of unobserved parts in an RDB based on feature space projection [8], which Kumagai and co-workers then applied to the dynamical traffic prediction problem [9]. In the field of statistical mechanics, Furtlehner *et al* modeled road traffic as an Ising model, where the state is determined by whether a road is congested or not, and addressed the traffic reconstruction and prediction problem that arises when the observed data are incomplete using belief propagation [10].

In this paper, we propose a new algorithm to reconstruct the traffic density data of the unobserved parts in an RDB. We use a Bayesian approach to express a posterior probability density function of unobserved roads. Our method is based on Markov random field (MRF) modeling of road traffic and the reconstruction of the traffic data of the unobserved parts in an RDB is achieved by solving simple simultaneous equations derived by a mean-field method after learning our MRF model on an HDB.

The remainder of this paper is organized as follows. In section 2, we introduce a graph representation of a road network and MRF modeling of road traffic. In section 3, we propose a traffic density reconstruction algorithm based on the MRF modeling of road traffic described in section 2. In section 4, we give a framework for determining the hyperparameters in the posterior probability density function derived in section 3 using the machine learning method. In section 5, we describe our numerical verification of the performance of our MRF model, which was achieved by conducting leave-one-out cross-validation and using large-scale simulation data for the road network of Sendai, Japan (the number of roads is 9582). Finally in section 6, we present our concluding remarks.

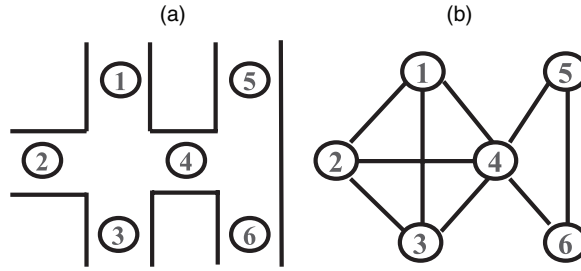


Figure 1. A simple case to explain our graph representation of a road network. (a) A toy road network with six roads. (b) A graph representation of the toy road network consists of six vertices and nine edges.

2. MRF modeling of road traffics

In this section, we explain how road traffic is expressed by MRF modeling. First, we define the undirected graph representation (V, E) of a real road network. Let us consider a road network consisting of N roads or road segments. A vertex $i \in V := \{1, \dots, N\}$ corresponds to the i th road in the road network. A set of all edges E includes edge ij if a vehicle on road i can move to road j without passing along the other roads.

We assign a random continuous variable $x_i \in (-\infty, \infty)$ associated with the traffic density of road $i \in V$. For each vertex and edge, we assign a potential function $\psi_i(x_i)$ and $\psi_{ij}(x_i, x_j)$, respectively. Then, the joint probability density function of $\mathbf{x} := \{x_i \mid i \in V\}$ is written as a product of a potential function

$$P(\mathbf{x}) := \frac{1}{Z} \prod_{i \in V} \psi_i(x_i) \prod_{ij \in E} \psi_{ij}(x_i, x_j). \quad (1)$$

The quantity Z is a partition function defined as

$$Z := \int d\mathbf{x} \prod_{i \in V} \psi_i(x_i) \prod_{ij \in E} \psi_{ij}(x_i, x_j), \quad (2)$$

where $\int d\mathbf{x}$ is taken over all the configurations of random variables \mathbf{x} . If we want to use a discrete random variable, the integration over continuous variables in equation (2) becomes a summation over discrete variables.

To explain our model, a simple case is shown in figure 1. There are six roads, represented as encircled numbers, and two intersections in figure 1(a). In this toy road network, vehicles on road 1 can directly move to roads 2, 3, or 4, but cannot move to roads 5 and 6 without passing along road 4. Then, this road network is translated to its graph representation, shown in figure 1(b). In this case, the joint probability density function is expressed as

$$P_{\text{ex}}(\mathbf{x}) := \frac{1}{Z_{\text{ex}}} \prod_{i=1}^6 \psi_i(x_i) \prod_{ij \in E_{\text{ex}}} \psi_{ij}(x_i, x_j), \quad (3)$$

$$Z_{\text{ex}} := \int d\mathbf{x} \prod_{i=1}^6 \psi_i(x_i) \prod_{ij \in E_{\text{ex}}} \psi_{ij}(x_i, x_j), \quad (4)$$

where $E_{\text{ex}} := \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}, \{4, 5\}, \{4, 6\}, \{5, 6\}\}$. It should be noted that we ignore road direction relationships throughout this paper for simplicity, as shown in figure 1; however, extending our model to one that includes road direction relationships is straightforward.

3. Traffic data reconstruction algorithm based on MRF

As mentioned in the introduction, a problem that affects traffic prediction is that complete traffic data of all roads cannot be collected due to a lack of sensors. In this section, we propose a method to reconstruct the traffic densities of unobserved roads from observed traffic densities based on MRF modeling and the Bayesian point of view. Suppose that $\{y_i \mid i \in V_o \subset V\}$ is a set of traffic densities of observed roads collected by sensors at a certain time and that the information about the traffic densities is incomplete. Our goal is to reconstruct the traffic densities of unobserved roads $i \in V_u := V \setminus V_o$.

In the Bayesian point of view, a reconstruction of unobserved roads is inferred by using the posterior probability density function $P(\mathbf{x} \mid \mathbf{y})$ expressed as

$$P(\mathbf{x} \mid \mathbf{y}) := \frac{P(\mathbf{y} \mid \mathbf{x}_o)P(\mathbf{x})}{\int d\mathbf{x} P(\mathbf{y} \mid \mathbf{x}_o)P(\mathbf{x})}, \quad (5)$$

where $\mathbf{y} := \{y_i \mid i \in V_o\}$ is a set of the observed results collected by sensors and $P(\mathbf{y} \mid \mathbf{x}_o)$ is a conditional density function expressing how \mathbf{y} is obtained from the set of true traffic densities $\mathbf{x}_o := \{x_i \mid i \in V_o\}$. It should be noted that, since y_i is a specific value, a denominator in equation (5) gives a constant value.

To define a concrete joint probability density function of \mathbf{x} , we assume that the potential functions in equation (1) are expressed as

$$\psi_i(x_i) := \exp(\beta_i x_i) \quad (6)$$

and

$$\psi_{ij}(x_i, x_j) := \exp \left\{ -\frac{\eta}{2}(x_i - x_j)^2 - \frac{\epsilon\eta}{2} \left(\frac{x_i^2}{|\partial i|} + \frac{x_j^2}{|\partial j|} \right) \right\}, \quad (7)$$

respectively, where $\partial i := \{j \in V \mid ij \in E\}$ and $|S|$ is the number of elements in set S . $\boldsymbol{\beta} = \{\beta_i \mid i \in V\}$ and η are hyperparameters that determine the features of our MRF model. $\beta_i \in (-\infty, \infty)$ is an external field that controls the largeness of the traffic density of road $i \in V$, while $\eta \in (0, \infty)$ is an interaction strength that controls the closeness of the traffic density values of neighboring roads. That is, if β_i is set at a large value, the density of road i tends to take a large value. Similarly, the densities of neighboring roads take close values if η is set at a large value. Here, we assume that the closeness of the traffic density values of neighboring roads is governed by a single parameter η for applying a large-scale problem and reducing computation cost. In appendix A, we discuss the more realistic setting that the densities of neighboring roads are governed by an edge-dependent parameter, η_{ij} . It should be noted that the second term of the exponent in equation (7) meets the mathematical demand to guarantee the normalization of the resulting prior density function. Therefore, we set ϵ at a small positive value to reduce the effect of this term.

Then, the joint probability density function, which is regarded as the prior density function in the Bayesian point of view, of \mathbf{x} is written as

$$\begin{aligned} P(\mathbf{x}; \boldsymbol{\beta}, \eta) &= \frac{1}{Z} \exp \left[\sum_{i \in V} \beta_i x_i - \frac{\epsilon\eta}{2} \sum_{i \in V} x_i^2 - \frac{\eta}{2} \sum_{ij \in E} (x_i - x_j)^2 \right] \\ &= \sqrt{\frac{\det C}{(2\pi)^N}} \exp \left[-\frac{1}{2} (\mathbf{x} - C^{-1} \boldsymbol{\beta})^T C (\mathbf{x} - C^{-1} \boldsymbol{\beta}) \right]. \end{aligned} \quad (8)$$

The $N \times N$ matrix C is defined by

$$C_{ij} := \begin{cases} (\epsilon + |\partial i|)\eta, & i = j \\ -\eta, & ij \in E \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

and called the precision matrix, which is the inverse of the covariance matrix. This form of probability density function is known as a Gaussian MRF and has been widely used in various applications [11].

We define a conditional density function $P(\mathbf{y} \mid \mathbf{x}_o)$ as

$$P(\mathbf{y} \mid \mathbf{x}_o) := \prod_{i \in V_o} \delta(y_i - x_i), \quad (10)$$

where $\delta(p - q)$ is the Dirac delta function. Here, it is assumed that the densities of the observed roads are not corrupted by measurement noises. That is, the densities of the observed roads are equivalent to the true densities of corresponding roads.

From equation (5), the posterior probability density function $P(\mathbf{x} \mid \mathbf{y})$ is written as $P(\mathbf{x} \mid \mathbf{y}) \propto P(\mathbf{y} \mid \mathbf{x}_o)P(\mathbf{x}; \boldsymbol{\beta}, \eta)$. Thus, the marginal posterior probability density function over the traffic densities of observed roads is expressed as

$$\begin{aligned} P(\mathbf{x}_u \mid \mathbf{y}; \boldsymbol{\beta}, \eta) &\propto \int d\mathbf{x}_o P(\mathbf{y} \mid \mathbf{x}_o) P(\mathbf{x}; \boldsymbol{\beta}, \eta) \\ &\propto \exp \left[\sum_{i \in V_u} \beta_i x_i - \frac{\epsilon \eta}{2} \sum_{i \in V_u} x_i^2 - \frac{\eta}{2} \sum_{ij \in E_1} (x_i - x_j)^2 - \frac{\eta}{2} \sum_{ij \in E_2} (x_i - y_j)^2 \right] \\ &\propto \exp \left[-\frac{1}{2} (\mathbf{x}_u - A^{-1} \mathbf{b})^T A (\mathbf{x}_u - A^{-1} \mathbf{b}) \right], \end{aligned} \quad (11)$$

where $\mathbf{x}_u := \{x_i \mid i \in V_u\}$, $E_1 := \{ij \in E \mid i, j \in V_u\}$ and $E_2 := \{ij \in E \mid i \in V_u, j \in V_o\}$. The $|V_u| \times |V_u|$ matrix A and vector $\mathbf{b} := \{b_i \mid i \in V_u\}$ are defined as

$$A_{ij} := \begin{cases} (\epsilon + |\partial i|)\eta, & i = j \\ -\eta, & ij \in E_1 \\ 0, & \text{otherwise,} \end{cases} \quad (12)$$

$$b_i := \beta_i + \eta \sum_{j \in \partial i_2} y_j, \quad (13)$$

where $\partial i_2 := \{j \in \partial i \mid ij \in E_2\}$. The reconstruction of unobserved traffic densities in the RDB can be achieved to find values \mathbf{x}_u^* , such that

$$\mathbf{x}_i^* := \begin{cases} x'_i, & x'_i \geq 0 \\ 0, & x'_i < 0, \end{cases} \quad (14)$$

$$\mathbf{x}'_u := \arg \max_{\mathbf{x}_u} P(\mathbf{x}_u \mid \mathbf{y}; \boldsymbol{\beta}, \eta) \quad (15)$$

for $i \in V_u$. Because the marginal posterior probability density function in equation (11) is a multivariate Gaussian distribution, values \mathbf{x}'_u are given by the mean vector of

$$\mathbf{x}'_u = A^{-1} \mathbf{b}, \quad (16)$$

and \mathbf{x}'_u can be calculated exactly by applying the naive mean-field approximation [12]. The problem of estimating road traffic densities is reduced to solving the following simultaneous equations by an iteration method:

$$x'_i = \frac{1}{A_{ii}} \left(\beta_i + \eta \sum_{j \in \partial i} z_j \right) \quad (17)$$

for $i \in V_u$, where

$$z_i = \begin{cases} x'_i, & i \in V_u \\ y_i, & i \in V_o. \end{cases} \quad (18)$$

The derivation of equation (17) is given in appendix B. Since matrix A is a sparse matrix, solving equation (17) is more efficient than calculating equation (16) directly.

The proposed algorithm for reconstructing the traffic densities of unobserved roads in an RDB is summarized as following.

- Step 1.* Determine the sets V_o and V_u from a graph representation of a road network. Input the values of observed traffic densities \mathbf{y} .
- Step 2.* Calculate matrix A according to equation (12).
- Step 3.* Solve simultaneous equations (17) by an iteration method and then use equations (14) and (15) to obtain reconstructed traffic densities \mathbf{x}_u^* .

4. Determining hyperparameters from HDB

We derived a reconstruction algorithm for traffic densities of unobserved roads based on a naive mean-field method described in the previous section. However, we have not yet specified the values of the hyperparameters. The purpose of this section is to show how these parameters are determined from the HDB using a machine learning method. In this section, it is assumed that a large number of complete traffic data are available. An explanation that justifies this assumption is that real complete data are not needed to determine hyperparameters and artificial data will suffice if they express the situations of road traffic well. When we have permitted the assumption that daytime road traffic situations are similar on different days, we can create these pseudo complete traffic data at a certain time by merging the data collected on multiple days because, in contrast to the RDB, the HDB consists of many traffic data for long time periods and a comprehensive area. This assumption seems reasonable, in particular, at rush hour in an urban area where traffic predictions are necessary. The extension to the area where this assumption is violated is mentioned in section 6 as well as the difficulty associated with it.

Let us suppose that a set of K complete road data of traffic densities, $\mathcal{D} = \{\mathbf{d}^k \mid k = 1, \dots, K\}$, $\mathbf{d}^k := \{d_i^k \in (-\infty, \infty) \mid i \in V\}$, created from the HDB. The empirical distribution of the complete road data is given by

$$Q(\mathbf{x}) := \frac{1}{K} \sum_{k=1}^K \prod_{i \in V} \delta(x_i - d_i^k). \quad (19)$$

A standard approach to determining the hyperparameters is finding that which maximizes the likelihood function defined as

$$L(\boldsymbol{\beta}, \eta) := \int d\mathbf{x} Q(\mathbf{x}) \log P(\mathbf{x}; \boldsymbol{\beta}, \eta). \quad (20)$$

However, this approach often gives rise to the over-fitting problem, which occurs when the number of hyperparameters is larger than the number of data. Typically, this problem occurred when the magnitude of estimated hyperparameters gets too large to fit the small set of data. Therefore, the regularized likelihood function written as

$$L_\lambda(\boldsymbol{\beta}, \eta; \lambda_\beta, \lambda_\eta) := L(\boldsymbol{\beta}, \eta) - \frac{\lambda_\beta}{2} \sum_{i \in V} \beta_i^2 - \frac{\lambda_\eta}{2} \eta^2 \quad (21)$$

is sometimes maximized to avoid the over-fitting problem. This regularization method is called ridge regression [13] and regarded as a MAP estimation to maximize $P(\boldsymbol{\beta}, \eta \mid \mathcal{D}) \propto P(\mathcal{D} \mid \boldsymbol{\beta}, \eta) \prod_{i \in V} \mathcal{N}(\beta_i; \lambda_\beta^{-1}) \mathcal{N}_+(\eta; \lambda_\eta^{-1})$ from the Bayesian point of view, where $P(\mathcal{D} \mid \boldsymbol{\beta}, \eta) := \prod_{k=1}^K P(\mathbf{d}^k; \boldsymbol{\beta}, \eta)$. $\mathcal{N}(v; u)$ is a Gaussian distribution of a random variable v with zero mean and variance u , and $\mathcal{N}_+(v; u)$ is a rectified Gaussian distribution,

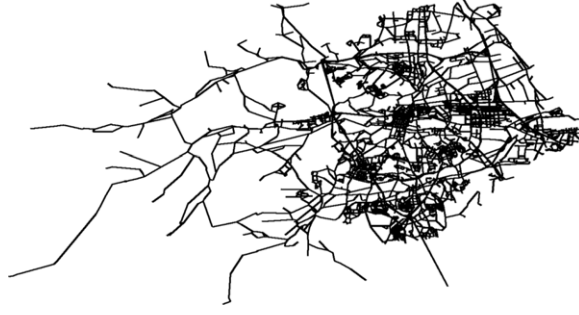


Figure 2. Road network of Sendai, Japan that we used in numerical experiments. There are 9582 vertices and 20 482 edges in the graph representation of this road network.

which is a modification of $\mathcal{N}(v; u)$ that resets its negative element to zero and renormalized. The parameters λ_β and λ_η are called the regularization parameter; it controls the magnitude of the estimates of the hyperparameters and is often determined by hand in advance. That is, the larger values at which λ_β and λ_η are set, the smaller are the values of the estimates of the hyperparameters. It should be noted that because our MRF model belongs to the exponential family without latent variables, the regularized likelihood function $L_\lambda(\beta, \eta; \lambda_\beta, \lambda_\eta)$ is a convex function [12]. Therefore, there are unique values of β and η that maximize $L_\lambda(\beta, \eta; \lambda_\beta, \lambda_\eta)$.

From equations (8) and (20), we can write equation (21) as

$$L_\lambda(\beta, \eta; \lambda_\beta, \lambda_\eta) = \sum_{i \in V} \beta_i \langle x_i \rangle_{\mathcal{D}} - \frac{\eta}{2} \sum_{i \in V} (\epsilon + |\partial i|) \langle x_i^2 \rangle_{\mathcal{D}} + \eta \sum_{ij \in E} \langle x_i x_j \rangle_{\mathcal{D}} - \frac{1}{2} \beta^T C^{-1} \beta + \frac{1}{2} \log \det C - \frac{\lambda_\beta}{2} \sum_{i \in V} \beta_i^2 - \frac{\lambda_\eta}{2} \eta^2 + \text{const.}, \quad (22)$$

where the notation $\langle \cdots \rangle_{\mathcal{D}}$ denotes the expectation with respect to $Q(\mathbf{x})$, i.e., the sample average of the complete traffic density data set. Using the gradient ascent method, the values of β and η that maximize $L_\lambda(\beta, \eta; \lambda_\beta, \lambda_\eta)$ can be obtained. The gradients of $L_\lambda(\beta, \eta; \lambda_\beta, \lambda_\eta)$ with respect to β and η are calculated as

$$\frac{\partial L_\lambda(\beta, \eta; \lambda_\beta, \lambda_\eta)}{\partial \beta_i} = \langle x_i \rangle_{\mathcal{D}} - \frac{1}{\eta} \sum_{j \in V} D_{ij}^{-1} \beta_j - \lambda_\beta \beta_i, \quad (23)$$

$$\frac{\partial L_\lambda(\beta, \eta; \lambda_\beta, \lambda_\eta)}{\partial \eta} = -\frac{1}{2} \sum_{i \in V} (\epsilon + |\partial i|) \langle x_i^2 \rangle_{\mathcal{D}} + \sum_{ij \in E} \langle x_i x_j \rangle_{\mathcal{D}} + \frac{1}{2\eta^2} \beta^T D^{-1} \beta + \frac{N}{2\eta} - \lambda_\eta \eta, \quad (24)$$

where the $N \times N$ matrix D is given by $C = \eta D$. It should be noted that, although the inverse of matrix D is needed in equations (23) and (24), it is enough to calculate the inverse matrix once in pre-processing, because, from equation (9), it depends on only the structure of a given road network.

5. Numerical experiments

In this section, we describe the numerical verification of the performance of our MRF model. We used the real road network of Sendai, Japan, described in figure 2, and 360 vehicle traffic data, which constitute a snapshot of its simulated vehicle traffic, to represent a realistic situation. In the graph representation of the Sendai road network, there are 9582 vehicles and 20 482 edges.

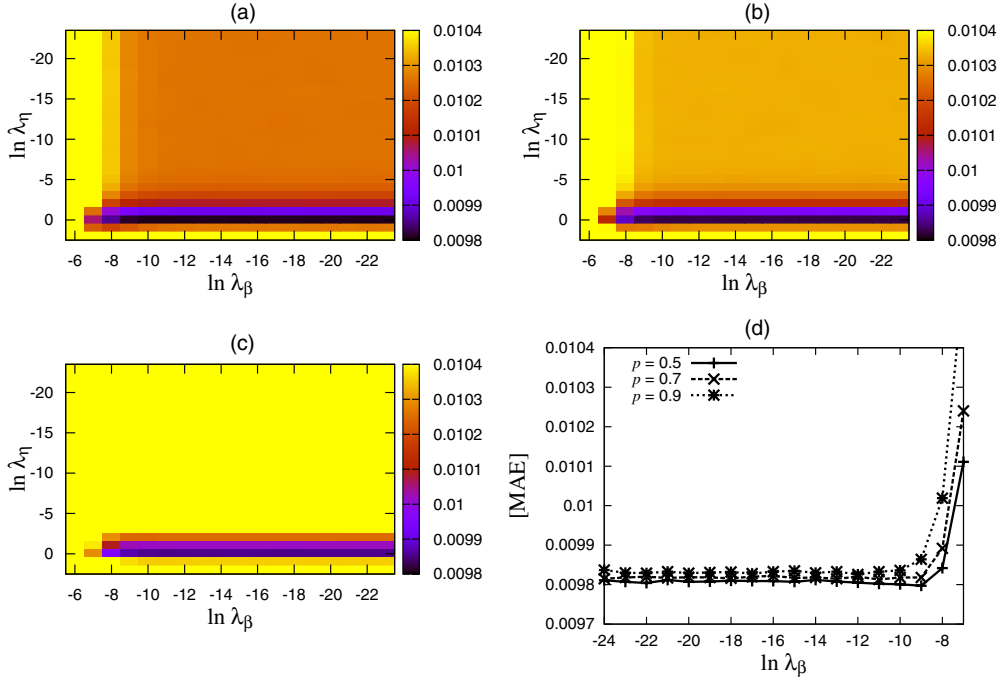


Figure 3. [MAE] versus $\ln \lambda_\beta$ and $\ln \lambda_\eta$ when $p = 0.5$, $p = 0.7$ and $p = 0.9$. Each point is obtained by averaging over 360 test data and 100 trials for each test data. (a) [MAE] versus $\ln \lambda_\beta$ and $\ln \lambda_\eta$ when $p = 0.5$. (b) [MAE] versus $\ln \lambda_\beta$ and $\ln \lambda_\eta$ when $p = 0.7$. (c) [MAE] versus $\ln \lambda_\beta$ and $\ln \lambda_\eta$ when $p = 0.9$. (d) [MAE] versus $\ln \lambda_\beta$ when $\ln \lambda_\eta = 0$ for $p = 0.5$, $p = 0.7$ and $p = 0.9$.

To evaluate the performance of our model, we conducted leave-one-out cross-validation [14] in which only one data item was used to check the performance and the remaining ones were used to determine the hyperparameters by the method described in section 4. The total performance of the model is then given by the performance averaged over all the chosen test data. That is, for each choice of test data, we regarded the remaining data as the complete data created from the HDB, and the test data were used to create the data in the RDB. In the test phase, we randomly selected unobserved roads with equal probability p from all roads and then reconstructed the traffic densities of the unobserved roads using our algorithm. In each test data, we evaluated the performance of our model by the average of mean absolute errors (MAE) between the true and reconstructed traffic density over 100 trials defined by

$$[\text{MAE}]_m := \frac{1}{100} \sum_{l=1}^{100} \left(\frac{1}{|V_u^{(l)}|} \sum_{i \in V_u^{(l)}} |x_i^* - x_i^{(m)}| \right), \quad (25)$$

where $V_u^{(l)}$ is the set of unobserved roads at the l th trial and $x_i^{(m)}$ is the true traffic density of road i in the m th data. Hence, the results of leave-one-out cross-validation are given by

$$[\text{MAE}] := \frac{1}{360} \sum_{m=1}^{360} [\text{MAE}]_m \quad (26)$$

for each λ_β and λ_η .

Figure 3(a)–(c) show the plots of [MAE] versus $\ln \lambda_\beta$ and $\ln \lambda_\eta$ when $p = 0.5$, $p = 0.7$ and $p = 0.9$, and figure 3(d) show the plot of [MAE] versus $\ln \lambda_\beta$ when $\ln \lambda_\eta = 0$ for $p = 0.5$,

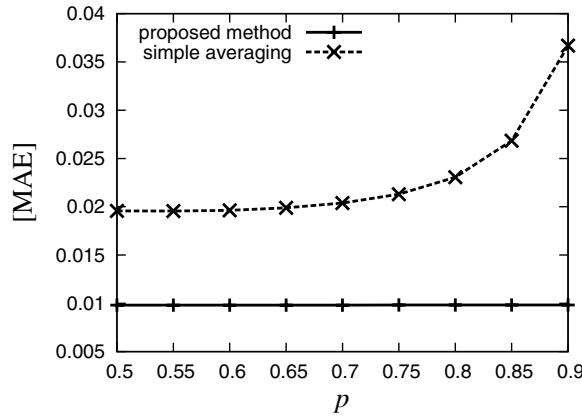


Figure 4. Plot of [MAE] versus p . Plus marks represent the results obtained by the proposed method when $(\lambda_\beta, \lambda_\eta) = (0, 1)$, while cross marks represent the results of a simple reconstruction method.

$p = 0.7$ and $p = 0.9$. Here, we set $\epsilon = 10^{-4}$ in equation (7). In the region where $\ln \lambda_\eta = 0$ and $\ln \lambda_\beta$ is sufficiently small, our reconstruction algorithm yields a good performance for all values of p , and in this region, the [MAE]s are close to the value obtained when $(\lambda_\beta, \lambda_\eta) = (0, 1)$. When $(\lambda_\beta, \lambda_\eta) = (0, 1)$, [MAE] was 0.009 808, 0.009 817 and 0.009 829 for $p = 0.5$, $p = 0.7$ and $p = 0.9$, respectively.

We show the plot of [MAE] versus p in figure 4. The plus marks represent the result obtained by our proposed method when $(\lambda_\beta, \lambda_\eta) = (0, 1)$. On the other hand, the cross marks represent the result obtained by a simple sequential interpolation method. In the simple method, the density of unobserved road i is reconstructed using the average of densities of observed roads neighboring i . When all the roads neighboring i are unobserved, nothing is done for road i at that step. All the reconstructed roads at that step are regarded as the observed roads at the next step and the same procedure is repeated until all unobserved roads become observed ones. Our proposed method yields better reconstruction results than does the simple interpolation method.

We show an example of our numerical experiments when $p = 0.7$ and $(\lambda_\beta, \lambda_\eta) = (0, 1)$ in figures 5 and 6. Figure 5(a) shows the original traffic densities and figure 5(e) shows the reconstructed traffic densities using our model. In figures 5(a) and (e), the road colors are changed from black to blue, green, yellow and red in order of increasing traffic density by 0.03 intervals, where a black road is one where the density takes a value between 0 and 0.03. Figure 5(c) shows the positions of unobserved roads; we colored the roads red when they were selected as unobserved roads with probability $p = 0.7$. That is, about 70% of roads are unobserved. The black roads in figure 5(c) denote the positions of traffic sensors that collect the traffic densities of the observed parts in the RDB. The MAE between figure 5(a) and (e) is 0.009 312. Figures 5(b), (d) and (f) are enlarged images of the downtown area of Sendai, Japan shown in figures 5(a), (c) and (e), respectively. Figure 6(a) shows the magnitudes of true and estimated traffic densities of unobserved roads in figures 5(a) and (e) with an auxiliary line drawn where estimated densities and true densities are equivalent. If the estimated traffic density is equivalent to the true one, the corresponding point is on this line. Figure 6(b) is an enlarged plot of figure 6(a) in the region where traffic density is low. The solid line in figure 6 is the auxiliary line. The correlation coefficient between true and estimated traffic densities of unobserved roads is 0.9168. This result suggests that there exists a strong linear correlation dependence between true and estimated traffic densities.

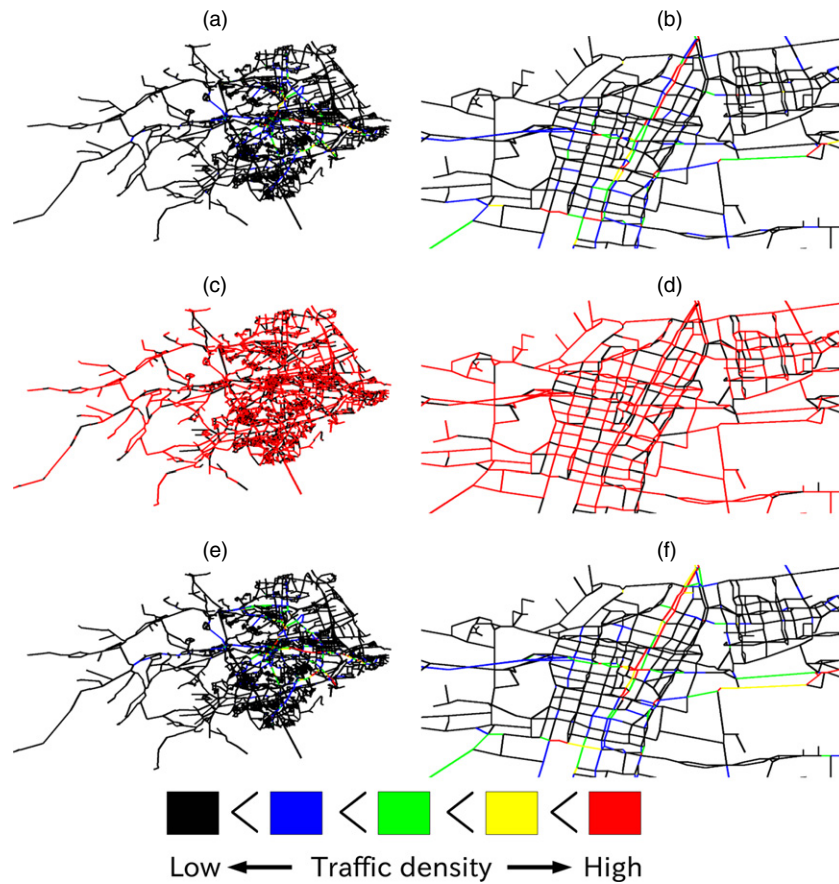


Figure 5. An example of our numerical experiments using simulated data for the road network of Sendai, Japan. (a) True traffic density data where each road is colored according to its traffic density. (b) Enlarged image of a part of (a). (c) Positions of unobserved roads where unobserved roads are colored red. About 70% of roads in this road network are unobserved. (d) Enlarged image of part of (c). (e) Result of reconstruction using our model. The MAE between (a) and these results is 0.009 312. (f) Enlarged image of part of (e).

6. Concluding remarks

In this paper, we proposed a traffic density reconstruction method based on MRF modeling. The reconstruction of unobserved parts of an road network in an RDB is reduced to a simple simultaneous equation of a mean-field method. The hyperparameters in our model are determined utilizing past traffic data in an HDB. We evaluated the performance of our model by conducting leave-one-out cross validation, as described in section 5. In the numerical experiments, we used large-scale simulated data of traffic in Sendai, Japan. We think it difficult to apply the previous reconstruction method to such a large-scale road network. It should be noted that, in this study, we reconstructed only the traffic density data; however, the extension of our MRF model to other data types, such as speed or flow, and furthermore, to combinations of these data types, is straightforward.

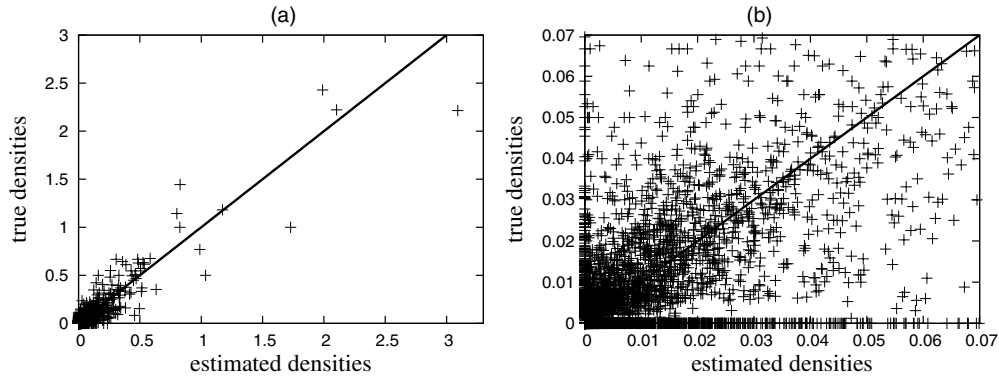


Figure 6. Plots of the magnitudes of true and estimated traffic densities of unobserved roads in figure 5(a) and (e). The solid line is an auxiliary line drawn where estimated densities and true densities are equivalent. (a) Plot of the magnitudes of true and estimated traffic densities of all unobserved roads. The correlation coefficient of these points is 0.9168. (b) Enlarged plot of (a).

In our scheme, we made two assumptions about the HDB and traffic densities for analytical convenience. The first assumption was that a number of complete traffic data can be created from an HDB because it can contain many traffic data for a long time period and comprehensive area and the daily conditions of road traffic seem similar, especially in an urban area. This assumption might be unrealistic in an area where the amount of traffic is small, as in a rural area. Our learning framework can be modified by using an expectation maximization algorithm [15] for determining hyperparameters from an incomplete data set in an HDB of such an area. However, the inverse of K different matrices A_k ($k = 1, \dots, K$) has to be calculated in this framework. The definition of the matrix A_k is similar to equation (12), but the dimensions corresponding to the number of unobserved roads in the k th data may be different. Although the calculation amount of inverting a matrix is polynomial time, calculating the inverse of such large matrices is still a computationally hard task. Therefore, a straightforward implementation of this framework will take a long time to be conducted and an approximate method to calculate A_k^{-1} needs to be found. It should be noted that the reconstruction scheme described in section 3 does not change after this modification. The second assumption was that the traffic density can take any real value and its potential functions have quadratic form, as equations (6) and (7). This assumption allows the Gaussian MRF modeling of traffic densities, which is a single-mode density function. In our definition of MRF modeling of traffic in section 2, we did not need to restrict the form of the potential functions and their arguments. One extension that would result in a more complex MRF is using a non-negative Boltzmann machine, which is a multi-modal density function for the joint density function of \mathbf{x} [16]; however, an approximation method [17] is required because its analytical treatment is difficult.

The other direction in which our MRF model can be developed is to innovate the traffic properties. In traffic flow theory [18], there are many properties pertaining to vehicle traffic, such as the flow conservation law, fundamental diagram and so on. Because we considered only the reconstruction problem of present traffic densities in this study, we utilized only snapshot data of vehicle traffic as shown in figure 5(a) and did not take account of any traffic properties statically or dynamically in our model. However, the extension of our MRF model to take account of the dynamical properties of traffic flow is a very important issue in the analysis of vehicle traffic. If such a model is designed, we can apply it to the prediction problem of future

traffic, to which the present model cannot be applied. We aim to develop our MRF model in these directions.

Acknowledgments

The authors thank Professor Masao Kuwahara and Jinyoung Kim of the Graduate School of Information Science, Tohoku University, for providing road network data and traffic simulation data. This work was partly supported by grants-in-aid (nos 25280089, 24700220 and 25.7259) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. SK was partially supported by a Research Fellowships of Japan Society for the Promotion of Science for Young Scientists.

Appendix A. The extension to edge-depending parameters model

In this appendix, we discuss an extension of our model to the edge-dependent parameter η_{ij} . In typical vehicle traffic pattern, more vehicles tend to go straight than turn left or right. Therefore, this setting is more realistic than the model represented by equation (8). However, determining the hyperparameters in this model is computationally hard because the number of the determining hyperparameters is $O(E)$. Therefore, we confine ourselves to showing the framework of this extension.

In this extension, the potential function in equation (7) is modified as

$$\psi_{ij}(x_i, x_j) := \exp \left\{ -\frac{\eta_{ij}}{2} (x_i - x_j)^2 - \frac{\epsilon \eta_{ij}}{2} \left(\frac{x_i^2}{|\partial i|} + \frac{x_j^2}{|\partial j|} \right) \right\}. \quad (\text{A.1})$$

Then, the precision matrices of the prior density function in equation (8) and of the marginal posterior density function are modified as

$$C_{ij} := \begin{cases} \left(1 + \frac{\epsilon}{|\partial i|}\right) \sum_{j \in \partial i} \eta_{ij}, & i = j \\ -\eta_{ij}, & ij \in E \\ 0, & \text{otherwise} \end{cases} \quad (\text{A.2})$$

and

$$A_{ij} := \begin{cases} \left(1 + \frac{\epsilon}{|\partial i|}\right) \sum_{j \in \partial i} \eta_{ij}, & i = j \\ -\eta_{ij}, & ij \in E_1 \\ 0, & \text{otherwise,} \end{cases} \quad (\text{A.3})$$

respectively. Therefore, the traffic density reconstruction algorithm in equation (17) is modified to solve the following simultaneous equations by an iteration method:

$$x'_i = \frac{1}{A_{ii}} \left(\beta_i + \sum_{j \in \partial i} \eta_{ij} z_j \right), \quad (\text{A.4})$$

for $i \in V_u$. The definition of z_i is the same as equation (18) in this modification. It should be noted that the calculation amount in each iteration of this algorithm is the same as the previous one.

However, the difficulty pertaining to this extension arises in the learning phase. In the present model, the gradients of the regularized likelihood function in equation (23) and (24) are modified as

$$\frac{\partial L_\lambda(\boldsymbol{\beta}, \boldsymbol{\eta}; \lambda_\beta, \lambda_\eta)}{\partial \beta_i} = \langle x_i \rangle_{\mathcal{D}} - \sum_{j \in V} C_{ij}^{-1} \beta_j - \lambda_\beta \beta_i \quad (\text{A.5})$$

and

$$\begin{aligned} \frac{\partial L_\lambda(\boldsymbol{\beta}, \boldsymbol{\eta}; \lambda_\beta, \lambda_\eta)}{\partial \eta_{ij}} = & -\frac{1}{2} \left(1 + \frac{\epsilon}{|\partial i|} \right) \{ \langle x_i^2 \rangle_{\mathcal{D}} - (C^{-1} \boldsymbol{\beta})_i^2 - C_{ii}^{-1} \} \\ & - \frac{1}{2} \left(1 + \frac{\epsilon}{|\partial j|} \right) \{ \langle x_j^2 \rangle_{\mathcal{D}} - (C^{-1} \boldsymbol{\beta})_j^2 - C_{jj}^{-1} \} \\ & + \langle x_i x_j \rangle_{\mathcal{D}} - (C^{-1} \boldsymbol{\beta})_i (C^{-1} \boldsymbol{\beta})_j - C_{ij}^{-1} - \lambda_\eta \eta_{ij}, \end{aligned} \quad (\text{A.6})$$

respectively, where $\boldsymbol{\eta} := \{\eta_{ij} \mid ij \in E\}$ and the notation $(\boldsymbol{a})_i$ is the i th element of vector \boldsymbol{a} . Obviously, the inverse of precision matrix C , which depends on the set of hyperparameters $\boldsymbol{\eta}$, is needed to calculate these gradients. Therefore, the large inverse matrix C^{-1} should be computed repeatedly when the regularized likelihood function is maximized by a gradient ascent method. This task is computationally hard. Therefore, a fast approximation method to compute the large inverse matrix is required in order to apply this extension model to the road reconstruction problem.

Appendix B. The derivation of naive mean-field equations

We derive the equations (17) by using a naive mean-field method in this appendix. In the naive mean-field method, instead of computing the exact probability density function $P(\mathbf{x}_u \mid \mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\eta})$ in equation (11), we compute the probability density function $P^{\text{MF}}(\mathbf{x}_u)$ that minimizes the Kullback–Leibler (KL) divergence

$$\text{KL}(P^{\text{MF}} \parallel P) := \int d\mathbf{x}_u P^{\text{MF}}(\mathbf{x}_u) \log \frac{P^{\text{MF}}(\mathbf{x}_u)}{P(\mathbf{x}_u \mid \mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\eta})} \quad (\text{B.1})$$

between $P^{\text{MF}}(\mathbf{x}_u)$ and $P(\mathbf{x}_u \mid \mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\eta})$ under the constraint $P^{\text{MF}}(\mathbf{x}_u) = \prod_{i \in V_u} P_i^{\text{MF}}(x_{ui})$. Because $P(\mathbf{x}_u \mid \mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\eta})$ is a multivariate Gaussian distribution, $P^{\text{MF}}(\mathbf{x}_u)$ can be chosen as a product of single variate Gaussian distributions expressed as

$$P^{\text{MF}}(\mathbf{x}_u) = \prod_{i \in V_u} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left[-\frac{1}{2\sigma_i^2} (x_i - \mu_i)^2 \right], \quad (\text{B.2})$$

where $\boldsymbol{\mu} := \{\mu_i \mid i \in V_u\}$ and $\boldsymbol{\sigma} := \{\sigma_i \mid i \in V_u\}$ are parameters that determine the concrete form of probability density function $P^{\text{MF}}(\mathbf{x}_u)$. Our goal is to determine the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ that minimize $\text{KL}(P^{\text{MF}} \parallel P)$. By substituting equations (B.2) and (11) into equation (B.1), $\text{KL}(P^{\text{MF}} \parallel P)$ is expressed as

$$\begin{aligned} \text{KL}(P^{\text{MF}} \parallel P) = & -\frac{1}{2} \sum_{i \in V_u} \log \sigma_i^2 - \sum_{i \in V_u} b_i \mu_i + \frac{1}{2} \sum_{i \in V_u} A_{ii} (\sigma_i^2 + \mu_i^2) \\ & + \sum_{ij \in E_1} A_{ij} \mu_i \mu_j + \text{const.} \end{aligned} \quad (\text{B.3})$$

From the minimum condition of equation (B.3) with respect to μ_i and σ_i^2 , equations that determine the value of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are given by

$$\mu_i = \frac{1}{A_{ii}} \left(b_i - \sum_{j \in \partial i_1} A_{ij} \mu_j \right) \quad (\text{B.4})$$

and

$$\sigma_i^2 = \frac{1}{A_{ii}}, \quad (\text{B.5})$$

respectively, where $\partial i_1 := \{j \in \partial i \mid ij \in E_1\}$. Substituting equations (12) and (13) into equations (B.4), we reach equations (17).

References

- [1] Faouzi N E E, Leung H and Kurian A 2011 *Inform. Fusion* **12** 4–10
- [2] Ahmed M S and Cook A R 1979 *Transp. Res. Rec.* **722** 1–9
- [3] Ide T and Sugiyama M 2011 Trajectory regression on road networks AAAI'11: *Proc. 24th Association for the Advancement of Artificial Intelligence (AAAI) Conf. on Artificial Intelligence* pp 203–8
- [4] Kriegel H P, Renz M, Schubert M and Zuefle A 2008 Statistical density prediction in traffic networks *Proc. Society for Industrial and Applied Mathematics (SIAM) Int. Conf. on Data Mining (SDM_08)* pp 692–703
- [5] Nikolova E and Karger D R 2008 Route planning under uncertainty: the canadian traveler problem AAAI'08: *Proc. 23rd Association for the Advancement of Artificial Intelligence (AAAI) Conf. on Artificial Intelligence* pp 969–74
- [6] Morikawa T, Yamamoto T, Miwa T and Wan L 2007 *Koutsuukougaku* **42** 65–75 (in Japanese)
- [7] Fushiki T, Yokota T, Kimita K and Kumagai M 2004 Study on density of probe cars sufficient for both level of area coverage and traffic information update cycle *Proc. 11th World Congress on ITS*
- [8] Kumagai M, Fushiki T, Yokota T and Kimita K 2006 *Inform. Process. Soc. Japan J.* **47** 2133–40 (in Japanese)
- [9] Kumagai M, Hiruta T, Okude M and Yokota T 2012 *Inform. Process. Soc. Japan J.* **53** 243–50 (in Japanese)
- [10] Furtlehner C, Lasgouttes J M and Fortelle A D L 2007 A belief-propagation approach to traffic prediction using probe vehicles *Proc. 10th IEEE Conf. Intelligent Transportation Systems* pp 1022–7
- [11] Rue H and Held L 2005 *Gaussian Markov Random Fields: Theory and Application* (London: Chapman and Hall)
- [12] Wainwright M J and Jordan M I 2008 *Found. Trends[®] Mach. Learn.* **1** 1–305
- [13] Hoerl A E and Kennard R W 1970 *Technometrics* **12** 55–67
- [14] Bishop C M 2006 *Pattern Recognition and Machine Learning* (Berlin: Springer)
- [15] Dempster A P, Laird N M and Rubin D B 1977 *J. R. Stat. Soc.* **39** 1–38
- [16] Downs O B, Mackay D J C and Lee D D 2000 *Adv. Neural Inform. Process. Syst.* **12** 428–34
- [17] Yasuda M and Tanaka K 2012 *Phil. Mag.* **92** 192–209
- [18] Daganzo C F 1997 *Fundamentals of Transportation and Traffic Operations* (Oxford: Pergamon)