Latent Variable Model for Weather-Aware Traffic State Analysis

Akira Kinoshita[1,2], Atsuhiro Takasu[2], and Jun Adachi[2]

[1] The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo, Japan,
[2] National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda, Tokyo, Japan,
{kinoshita,takasu,adachi}@nii.ac.jp

Abstract. Because vehicular traffic is affected by weather conditions, knowledge of the relationship between weather and traffic enables attempts to improve social services through applications such as situation-aware anomaly vehicle detection and snow-removal planning in snowy countries. We propose a *weather-aware traffic state model* for vehicular traffic analysis in consideration of weather conditions. The model is a probabilistic latent variable model that integrates weather and traffic data, whereby the characteristics of the traffic according to location, time, and weather condition are obtained automatically. After we observe both weather and travel times along road segments, we derive the expectation–maximization algorithm for model parameter estimation and the predictive distribution of travel time given the weather observation values. We evaluated the model qualitatively and quantitatively using winter traffic and weather data for the city of Sapporo, Japan, which is a large city that suffers heavy snowfalls. The empirical analysis with model visualization outcomes demonstrated the relationship between the expected vehicular speed and weather conditions, and showed the potential bottleneck segments for given weather conditions. The quantitative evaluation showed that our model fits the data better than a linear regression model, which suggests the potential for anomaly detection from vehicular observation data.

Keywords: Data integration · Data mining · Latent variable models · Probe-car data · Social cyber–physical systems · Weather-aware traffic state analysis

1 Introduction

Real-world traffic is complex and involves various factors. One important factor is the weather conditions. These change the driving environment, including visibility and road surface conditions, which affects the movement of vehicles in terms of running speed, vehicular gaps, and so on. Bad weather also affects the behavior of people: they may change their destination or visiting order, or avoid traveling at all, which affects the traffic volume and travel route. Knowledge of the relationship between weather and traffic enables attempts to improve social services.

For example, if the roads and intersections in which heavy snow often hinders the traffic are known, effective road improvement and snow-removal planning will be available. Situation-aware detection of anomalies in vehicle movements will also be possible with such knowledge and the awareness of the current condition. This approach also promises to provide finer-grained traffic information than existing approaches: for instance, that traffic is congested because the average speed has fallen below a prescribed value.

The relationship between weather and vehicular traffic has been studied over decades. Traffic engineers studied the effect of weather on freeway traffic in 1988 [5]. Keay and Simmonds analyzed the relationship between rainfall and traffic volume in Melbourne, Australia, using a linear regression method [7]. Recent studies also utilized linear regression techniques but they are application-oriented. Lee et al. developed a linear regression model to predict traffic congestion using weather data [9]. Tanimura et al. also used a linear regression model to predict the reductions in vehicle speeds in snowy conditions [11]. Xu et al. predicted traffic flows based on weather data using an artificial neural network as well as a linear regression method [13]. These studies model the relationship between weather and traffic in terms of traffic statistics or aggregated values. However, they do not model the relationship between weather and the behavior of individual vehicles.

The movement of a vehicle differs substantially among individuals. In particular, on ordinary roads, the speeds of vehicles vary greatly because they frequently slow down for intersections, traffic lights, and pedestrians. Therefore, in this paper, we model the traffic by probabilistic means. With the probabilistic distribution of the traffic observation values, such as travel time at a certain location over a selected period, the degree of anomaly of the observed behavior of an individual vehicle can be evaluated quantitatively. Statistics such as mean travel time and average speed can also be calculated based on the distribution. Earlier studies proposed several probabilistic distribution models [2,4,14]; however, they have not considered the weather conditions.

In this paper, we develop a *weather-aware traffic state model* (WATS model), a probabilistic model of observed values for traffic with consideration of weather conditions. Our probabilistic model aims at learning the "normal" patterns of traffic using data archives, which would be used for the applications described above, i.e., traffic incident detection and snow-removal planning. We have previously proposed a latent variable model for traffic state, and have shown its effectiveness for incident detection on expressways [8]. This model introduced latent traffic states such as "smooth" and "congested," and assumed the traffic observation values depend on the latent states. This paper extends the model by introducing variables related to weather conditions and relationships among the variables. The WATS model assumes not only that the traffic observation value depends on latent traffic states, but also that the traffic state depends on the weather. We introduce weather states to relate the traffic data to the weather data observed at the same time of day, thus allowing the relationship between weather and traffic to be learned. We borrow the idea of a Pachinko allocation model [10], which is a latent variable model that analyzes topics in textual information by considering correlations among latent information, to realize a feasible model for the problem.

We also aim to apply the WATS model to ordinary roads as well as expressways. In this study, we conducted an experiment in the city of Sapporo, Hokkaido, Japan. Sapporo is the metropolis of Hokkaido prefecture, and the fourth largest city in Japan in terms of population. While more than 1.9 M people live there, the city is located in a heavy snowfall region. The city spent more than 18 billion yen (about $US 150 M) on plans to counter snow in the 2015 fiscal year, with more than three quarters of the budget devoted to snow removal [3]. It is important to enhance cost effectiveness while reducing the bad influence of snow on traffic. Against this background, we attempted an empirical analysis on the weather–traffic relationship in Sapporo in winter. The visualized results of the WATS model indicate that it is possible to find bottlenecks that change according to weather and time. Our quantitative evaluation showed that the model fits the data better than an existing weather–traffic model, suggesting the potential for incident detection.

The main contributions of this paper are as follows.

- We propose a *weather-aware traffic state* (WATS) *model*. It is a new probabilistic latent variable model that integrates weather and traffic data, deriving the characteristics of the traffic according to location, time, and weather condition automatically.
- We show the effectiveness and the potential of the WATS model by our empirical qualitative and quantitative evaluations. The evaluation was conducted in the city of Sapporo, located in Japan's snow country.

The rest of this paper is organized as follows. In Section 2, we propose the WATS model. Section 3 reports the result of our empirical experiment and discusses the results, issues and future work. Finally, Section 4 concludes the paper.

2 Weather-Aware Traffic State Model

In this section, we present our WATS model. The aim of our model is to integrate observation data, traffic conditions, and weather conditions, so that the relationship between weather and vehicular traffic can be analyzed. Table 1 summarizes the notation used in this paper.

2.1 Model Design Concepts

We use two kinds of data:

- weather observation data, which are obtained periodically (e.g., hourly),
- traffic observation data, which are obtained intermittently or continuously.

Table 1: Notation

Notation	Definition
s	Road segment.
d_s	Length of road segment s.
S	Number of road segments.
t	Time.
T	Number of weather observation data.
n	Index of traffic observation data.
N_{ts}	Number of traffic observation data in segment s in time t.
$\boldsymbol{w}$	Weather observation data.
D	Dimension of $\boldsymbol{w}$.
x	Traffic observation data.
$\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$	Mean and precision, respectively, of $\boldsymbol{w}$.
λ	Mean travel speed.
l, k	Index of latent states.
L, K	Number of latent weather and traffic states, respectively.
$\boldsymbol{\pi}$	Mixing coefficient of weather states.
v	Latent weather state.
y, z	Latent weather and traffic state, respectively.
$\boldsymbol{\theta}$	Mixing coefficient of traffic states.
$\boldsymbol{\Theta}$	Set of parameters: $(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \lambda)$.
η	All hyperparameters: $(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, b, \boldsymbol{\mu}_0, \gamma, \boldsymbol{W}, \nu)$.

Here we develop a latent variable model, whereby weather-aware traffic performance is described in terms of a probability distribution.

Weather observation data indicate the weather conditions in the subject area. At time t, only one weather observation data point $\boldsymbol{w}$ is obtained. The observation value is numerical and we assume $\boldsymbol{w}$ follows a Gaussian distribution with mean $\boldsymbol{\mu}$ and precision (i.e., the inverse of the covariance matrix) $\boldsymbol{\Lambda}$. The mean and covariance can change according to weather conditions. For example, temperature tends to be low and snowfall tends to be large in "snowy" conditions, and temperature tends to be high and snowfall tends to be zero in "sunny" conditions. We introduce L weather states, each of which is characterized by the mean $\boldsymbol{\mu}_l$ and the precision $\boldsymbol{\Lambda}_l$. Then the probability distribution of the weather observation at time t is described in terms of a mixture of these components:

$$p(\boldsymbol{w}_t \mid \boldsymbol{\Theta}) = \sum_{l=1}^{L} \pi_{tl} \mathcal{N}(\boldsymbol{w}_t \mid \boldsymbol{\mu}_l, {\boldsymbol{\Lambda}_l}^{-1}), \tag{1}$$

where $\boldsymbol{\pi}_t$ is the mixing coefficient vector at time t. π_{tl} is equivalent to the probability of being in the l-th weather state at time t. The mixing coefficient varies according to the time t while each of L component Gaussians is identical over time. Therefore, $\boldsymbol{\pi}_t$ characterizes the weather condition at time t.

Intuitively, we can identify traffic states as being "smooth" or "congested," regardless of location, and the traffic state information is strongly related to geographical and time-of-day conditions [8]. Therefore, traffic observations are conducted for each *segment* s, which is the unit for traffic observation and is specified by road segment, direction, and time period, such as "morning" or "evening." At time t, N_{ts} traffic observation data points are obtained from segment s.

There are several options for traffic observation values, e.g., travel time and average speed. In this paper, we use travel time as the traffic observation value x. Travel time can be measured directly using probe cars [12]. The travel time depends on both the length of the road segment and traffic condition, and we model the relationship among them. Although several models have been proposed for travel time distribution [2,4], they are too complicated to incorporate into our model or they need additional features for estimation. Intuitively, traffic conditions can be characterized by the average speed λ: the traffic is "smooth" if λ is large, and "congested" if λ is small. We therefore introduce K traffic states with different average speeds $\{\lambda_k\}$ and the gamma distribution $\mathrm{Gamma}(d_s, \lambda_k)$, where d_s is the road length of the segment s. This distribution makes sense with the assumption of a Poisson process: considering the *event* that a vehicle goes forward a unit length, the rate λ, or average number of times the event occurs, is equivalent to the average speed, and the total time required for k occurrences of the event is equivalent to the travel time on a road of length k and follows an Erlang distribution. The gamma distribution is the generalized form of the Erlang distribution by allowing k to be a positive real number rather than a positive integer. Note that we use the notation $\mathrm{Gamma}(d_s, \lambda_k)$ in this paper for the gamma distribution with the following probability density function:

$$\mathrm{Gamma}(x \mid d_s, \lambda_k) = \frac{\lambda_k^{d_s} x^{d_s - 1} \exp(-\lambda_k x)}{\Gamma(d_s)}, \tag{2}$$

where $\Gamma(z)$ is the gamma function.

Traffic conditions vary according to the time and place and depend on the "condition" there, e.g., road shape and congestion occurrence. The condition can also change according to the weather even at a particular place. For example, traffic congestion may occur when it snows. There seems to be a hierarchical property: traffic observations depend on the traffic conditions, and the traffic conditions depend on the weather conditions. In this study, we model the relationship as a latent variable model, borrowing the idea of the Pachinko allocation model (PAM) [10]. PAM was proposed to analyze topics in textual information with consideration of correlations among latent information. It introduces a hierarchical structure among latent variables: word occurrence in a document depends on its topic and the topic depends on the supertopic. In our WATS model, the traffic observation value depends on the traffic state and the traffic state depends on the weather state. The probability distribution of the traffic observation value is described in terms of a hierarchical mixture:

$$p(x_{tsn} \mid \boldsymbol{\Theta}) = \sum_{l=1}^{L} \pi_{tl} \sum_{k=1}^{K} \theta_{slk} \mathrm{Gamma}(x_{tsn} \mid d_s, \lambda_k), \tag{3}$$

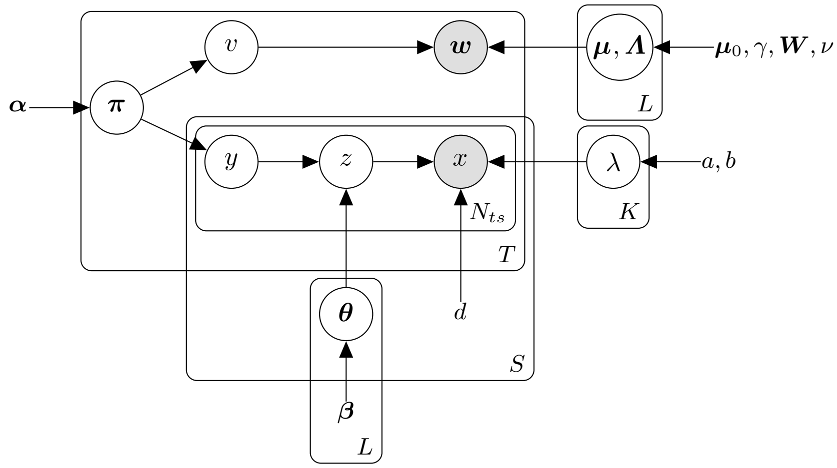


Fig. 1: Graphical model for the WATS model.

where $\boldsymbol{\theta}_{sl}$ is the mixing coefficient vector, the k-th element of which is equivalent to the probability of being in the k-th traffic state in segment s in the l-th weather state. Each of K component gamma distributions is identical regardless of the segment and the weather state. Therefore, $\boldsymbol{\theta}_{sl}$ characterizes the traffic performance for each road segment and for each weather state.

2.2 Generative Model

Figure 1 shows the graphical model for the WATS model. The upper part generates the weather observation data $\boldsymbol{w}_t$ while the lower part generates the traffic observation data x_{tsn}. The two kinds of data in the same time t are associated with each other by the parameter $\boldsymbol{\pi}_t$. The generative process is as follows:

1. Generate parameters:
   (a) Generate the mean speed $\lambda_k \sim \text{Gamma}(a, b)$ for each of K traffic states;
   (b) Generate the pair of the mean vector and the precision matrix of the weather observation values $(\boldsymbol{\mu}_l, \boldsymbol{\Lambda}_l) \sim \text{Gauss–Wishart}(\boldsymbol{\mu}_0, \gamma, \boldsymbol{W}, \nu)$ for each of L weather states;
   (c) Generate the mixing coefficient $\boldsymbol{\theta}_{sl} \sim \text{Dirichlet}(\boldsymbol{\beta}_l)$ for each segment s and for each weather state l;
2. Generate the data at time t:
   (a) Generate a mixing coefficient $\boldsymbol{\pi}_t \sim \text{Dirichlet}(\boldsymbol{\alpha})$;
   (b) Generate weather observation data:
      i. Generate the weather state $v_t \sim \text{Multinomial}(\boldsymbol{\pi}_t)$;
      ii. Generate the weather observation vector $\boldsymbol{w}_t \sim \mathcal{N}(\boldsymbol{\mu}_l, {\boldsymbol{\Lambda}_l}^{-1})$, where $v_t = l$.
   (c) Generate the nth traffic observation data for each segment s:
      i. Generate $y_{tsn} \sim \text{Multinomial}(\boldsymbol{\pi}_t)$;

ii. Generate $z_{tsn} \sim \text{Multinomial}(\boldsymbol{\theta}_{sl})$, where $y_{tsn} = l$;
iii. Generate $x_{tsn} \sim \text{Gamma}(d_s, \lambda_k)$, where $z_{tsn} = k$.

According to the generative process above, the log-likelihood is derived as follows:

$$\sum_t \ln \sum_l \pi_{tl} \mathcal{N}(\boldsymbol{w}_t \mid \boldsymbol{\mu}_l, \boldsymbol{\Lambda}_l{}^{-1}) + \sum_{t,s,n} \ln \sum_{l,k} \pi_{tl} \theta_{slk} \text{Gamma}(x_{tsn} \mid d_s, \lambda_k). \quad (4)$$

2.3 Model Estimation

Maximum-likelihood (ML) estimations or maximum a posteriori (MAP) estimations for latent variable models generally use the expectation–maximization (EM) algorithm [1]. The EM algorithm iterates E- and M-steps alternately until the log-likelihood converges. We show the update formulas for our WATS model; they can be derived by considering some conditional probabilities based on the graphical model, but the mathematical details of the derivation are are omitted here because of space limitations.

The E-step calculates the posteriors:

$$\zeta_{tl} \equiv p(v_t = l \mid \boldsymbol{w}_t, \boldsymbol{\Theta}) \propto \pi_{tl} \mathcal{N}(\boldsymbol{w}_t \mid \boldsymbol{\mu}_l, \boldsymbol{\Lambda}_l{}^{-1}), \quad (5)$$

$$\xi_{tsnlk} \equiv p(y_{tsn} = l, z_{tsn} = k \mid x_{tsn}, \boldsymbol{\Theta}) \propto \pi_{tl} \theta_{slk} \text{Gamma}(x_{tsn} \mid d_s, \lambda_k). \quad (6)$$

The M-step maximizes the following Q function for the ML estimation, or the $\tilde{Q}$ function for the MAP estimation:

$$Q = \sum_{t,l} \zeta_{tl} \left[\ln \mathcal{N}(\boldsymbol{w}_t \mid \boldsymbol{\mu}_l, \boldsymbol{\Lambda}_l{}^{-1}) + \ln \pi_{tl} \right] + \sum_{t,s,n,l,k} \xi_{tsnlk} \left[\ln \text{Gamma}(x_{tsn} \mid d_s, \lambda_k) + \ln \theta_{slk} + \ln \pi_{tl} \right], \quad (7)$$

$$\tilde{Q} = Q + \ln p(\boldsymbol{\Theta}). \quad (8)$$

This Q or $\tilde{Q}$ is maximized by introducing Lagrange multipliers and setting its partial derivatives with respect to each parameter to zero. The update formulas for MAP estimation are derived as follows:

$$\pi_{tl} \propto \zeta_{tl} + \sum_{s,n,k} \xi_{tsnlk} + \alpha_l - 1, \quad \theta_{slk} \propto \sum_{t,n} \xi_{tsnlk} + \beta_{lk} - 1, \quad (9)$$

$$\lambda_k = \frac{\sum_{t,s,n,l} \xi_{tsnlk} d_s + a - 1}{\sum_{t,s,n,l} \xi_{tsnlk} x_{tsn} + b}, \qquad \boldsymbol{\mu}_l = \frac{\sum_t \zeta_{tl} \boldsymbol{w}_t + \gamma \boldsymbol{\mu}_0}{\sum_t \zeta_{tl} + \gamma}, \quad (10)$$

$$\boldsymbol{\Lambda}_l{}^{-1} = \frac{\sum_t \zeta_{tl} (\boldsymbol{w}_t - \boldsymbol{\mu}_l)(\boldsymbol{w}_t - \boldsymbol{\mu}_l)^\mathsf{T} + \gamma (\boldsymbol{\mu}_l - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_l - \boldsymbol{\mu}_0)^\mathsf{T} + \boldsymbol{W}^{-1}}{\sum_t \zeta_{tl} + \nu - D}. \quad (11)$$

π_{tl} and θ_{slk} should be normalized so that $\sum_l \pi_{tl} = 1$ and $\sum_k \theta_{slk} = 1$ respectively. As for the ML estimation, constant terms (i.e., terms that do not include ζ_{tl} or ξ_{tsnlk}) are simply eliminated.

2.4 Prediction

Once the model parameter $\mathbf{\Theta}$ is estimated, we can calculate the predictive distribution of travel time x in the segment s as the conditional probability given a weather observation vector $\boldsymbol{w}$. Based on the graphical model, the predictive distribution is derived as follows:

$$p(x \mid \boldsymbol{w}, s, \mathbf{\Theta}) = \sum_k \omega_k \mathrm{Gamma}(x \mid d_s, \lambda_k), \tag{12}$$

where

$$\omega_k \equiv \frac{\sum_l \theta_{slk} \alpha_l \left(q(\boldsymbol{w}) + \mathcal{N}(\boldsymbol{w} \mid \boldsymbol{\mu}_l, \boldsymbol{\Lambda}_l{}^{-1}) \right)}{\left(\sum_l \alpha_l + 1\right) q(\boldsymbol{w})}, \tag{13}$$

$$q(\boldsymbol{w}) \equiv \sum_l \alpha_l \mathcal{N}(\boldsymbol{w} \mid \boldsymbol{\mu}_l, \boldsymbol{\Lambda}_l{}^{-1}). \tag{14}$$

Therefore, the predicted travel time follows a gamma mixture distribution. The expected travel time is obtained as follows:

$$\mathbb{E}[x \mid \boldsymbol{w}, s, \mathbf{\Theta}] = \sum_k \omega_k \frac{d_s}{\lambda_k}. \tag{15}$$

The predictive distribution and the expected value above include the hyperparameter $\boldsymbol{\alpha}$ in the formulas. It is given in the MAP estimation, but it is not given in the ML estimation. According to Equations (7) and (8), the ML estimation can be regarded as a MAP estimation that assumes that the prior $p(\mathbf{\Theta})$ is uniform, i.e., constant. $\boldsymbol{\alpha}$ is the parameter of a Dirichlet distribution, which is equivalent to a uniform distribution when $\boldsymbol{\alpha} = \mathbf{1}$, i.e., $\alpha_l = 1$ for all l. Therefore, we propose to use $\boldsymbol{\alpha} = \mathbf{1}$ for prediction with the estimated value of ML estimation.

3 Experiment

We have conducted an empirical winter traffic analysis in the city of Sapporo by applying the WATS model to real weather and traffic data. This section reports and discusses the experimental results.

3.1 Data set

Figure 2 shows the subject area, a part of Sapporo, Hokkaido, Japan. Snow falls in Sapporo from the end of October to April and the snow depth reaches about one meter in midwinter every year. Surface weather such as temperature and precipitation is observed at about 1300 stations in Japan using the Automated Meteorological Data Acquisition System (AMeDAS) developed and operated by the Japan Meteorological Agency [6]. In this experiment, we used weather

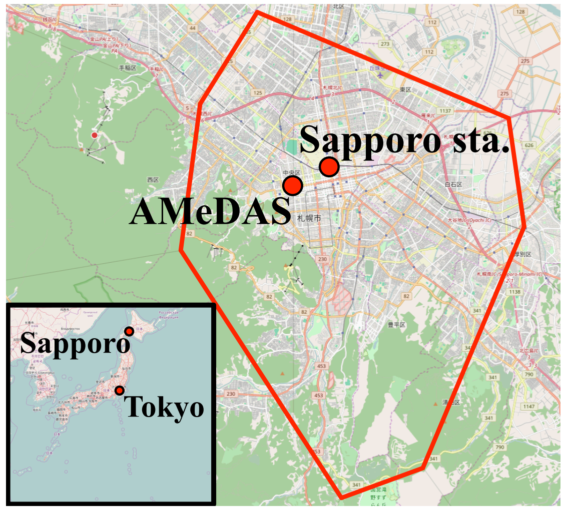


Fig. 2: Map of Sapporo, Hokkaido, Japan. The red-line polygon shows the subject area. The Sapporo AMeDAS station is also shown. Map tiles ©OpenStreetMap contributors, CC BY-SA 2.0.

observation data from Sapporo AMeDAS station and traffic data within the region shown in Figure 2; the nearest AMeDAS station is Sapporo.

We obtained probe-car data in Sapporo city, and then we preprocessed them to generate traffic data for each road segment. The original probe-car data include trajectory information, i.e., a sequence of time–location data points for each active probe car. The preprocess had five phases: we defined segments according to the road segments and time period; map matching was conducted to associate a vehicular location with a road segment; trajectory data were interpolated linearly so that the time when the vehicle entered or exited the segment was determined; and the travel time for each segment was observed. As mentioned previously, each *segment* was determined by road segment, direction, and time period. Our road segments follow the main road segment data in a commercial digital road map. They are divided by day (weekday or holiday) and by time period: morning rush hour (7–10 h), daytime (10–17 h), evening rush hour (17–20 h), and night (20–7 h on the following day). For the weather observation data, we used hourly data at Sapporo AMeDAS station, including temperature [°C], snowfall [cm/h], snow depth [cm], and precipitation [mm/h].

We obtained weather and traffic data from January 2010 to February 2015, but we used only winter data from October to April of each year. We used data in or before April 2014 for model training, and that in and after October 2014 for testing. We removed noisy segments that have less than 1000 training data points from the data set. The final training data include 31,891 segments and 74,826,640 traffic observations.

3.2 Parameter Estimation

We first trained the WATS model using the training data. In this experiment, we conducted ML estimation. We assumed eight weather states and 16 traffic states, but we did not assume any other parameter values such as mean speed or mean temperature, which were estimated from the training data.

We implemented the EM algorithm described in the previous section using OpenMP for multiprocessing. The EM algorithm required about 13 minutes using parallel processing with 36 CPU cores. Table 2 and 3 show the estimated parameter values. The eight estimated weather states include snowy conditions (l =1,2,4), warm conditions (7,8), and snow-accumulated conditions (1–6). The mean speed of the traffic states extends over a wide range from “almost stopped” to “very fast.” The estimated mean speed for the 16th traffic state was quite fast; this seems to be caused by outliers in the training data, which were possibly caused by map-matching failures. However, these would be insignificant in this experiment because the estimated mixing coefficient value for the 16th traffic state was zero or almost zero for almost all segments; that is, the 16th traffic state was ignored automatically.

We visualized the estimated model as shown in Figure 3 for a qualitative analysis. Because of space limitations, the figure shows the estimated model only for the morning rush hour on weekdays. The color of a segment indicates the expected values of average link speeds; their probability densities were obtained by the transformation of random variables using the estimated travel time distribution and the road length. The map tends to be red as the weather states become warmer with less snow, but the map of the 6th state is exceptional because it is almost blue. This suggests that the traffic tends to slow down almost everywhere when the temperature is nearly zero degrees and there is snow accumulation. Vehicles can be expected to slow down in such weather conditions because the accumulated snow melts and freezes at around zero degrees and makes the road surface condition very poor. On the other hand, some road links keep the mean travel speed over different weather states at a slow value. We speculate that the reason is queueing and waiting for traffic lights regardless of weather.

3.3 Model Evaluation

We evaluated two quantitative metrics: prediction error of the expected travel time and the cross entropy. For comparison, we also trained linear regression models for each segment as the baseline.

Expected Travel Time Prediction With the estimated model, we calculated the expected travel time for each road segment every hour. We used ML estimation in this experiment, so the expected value was given by Equation (15) with $\boldsymbol{\alpha} = \mathbf{1}$ as we described in the previous section. We regarded the sample mean of the actual travel times of vehicles in the test data as the ground truth in this experiment. Statistically, the number of samples should be large enough to be a

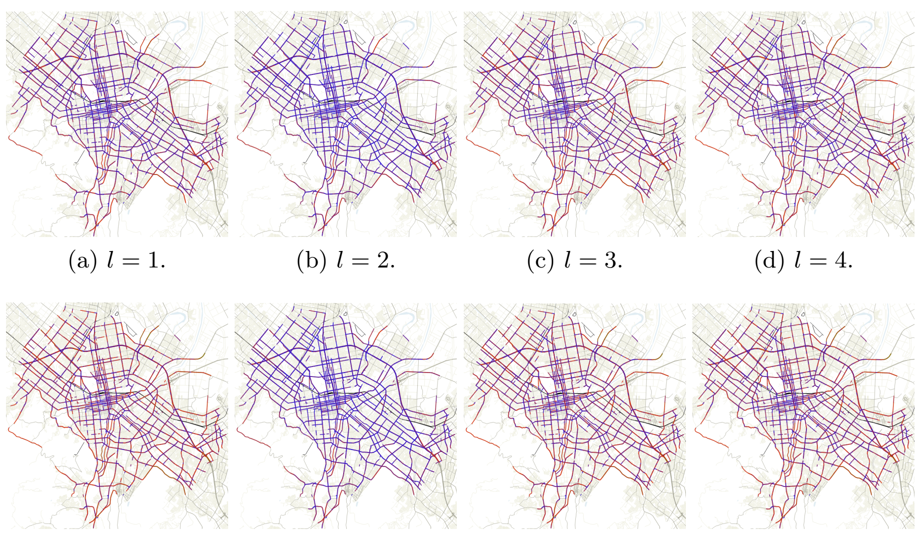
Table 2: Estimated mean parameters for each weather state distribution. The weather states are ordered by temperature.

Weather state l	1	2	3	4	5	6	7	8
Temperature [°C]	−3.6	−2.9	−2.4	−1.8	−1.4	1.1	8.3	8.8
Snowfall [cm/h]	2	1	0	3	0	0	0	0
Snow depth [cm]	65	56	66	47	56	17	0	0
Precipitation [mm/h]	1.1	0.5	0.0	2.4	0.6	0.0	1.4	0.0

Table 3: Estimated mean parameter for each traffic state distribution. The traffic states are ordered by mean speed.

Traffic state k	1	2	3	4	5	6	7	8
Mean speed [km/h]	1.2	3.0	5.0	7.2	10.0	13.4	17.6	23.3

Traffic state k	9	10	11	12	13	14	15	16
Mean speed [km/h]	31.0	39.4	48.1	56.9	65.6	75.4	96.7	295.5



(a) $l = 1$. (b) $l = 2$. (c) $l = 3$. (d) $l = 4$.

(e) $l = 5$. (f) $l = 6$. (g) $l = 7$. (h) $l = 8$.

Fig. 3: Estimated model for the morning rush hour (7–10 h) on weekdays. The color of a segment indicates the expected value of the average link speed. Green represents high speed (100 km/h), red is moderate speed (50 km/h), and blue is “almost stopped” (0 km/h). Each subfigure corresponds to one of the eight weather states. Base map tiles ©OpenStreetMap contributors, Who’s On First, and openstreetmapdata.com. Data are licensed under ODbL.

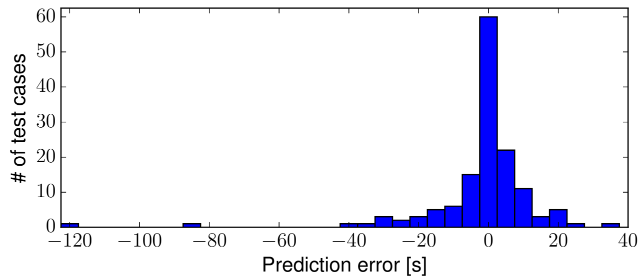
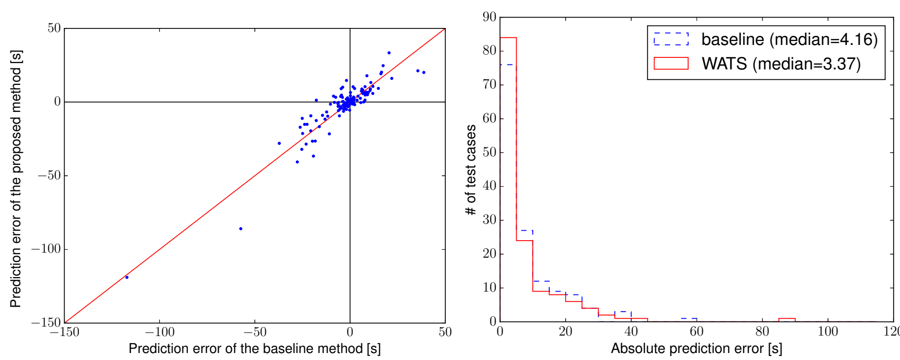


Fig. 4: Histogram of prediction errors for expected travel time calculated with the WATS model. Positive error means that the predicted travel time was longer than the ground truth, and vice versa.



(a) Prediction error of the two methods. (b) Absolute prediction error distribution.

Fig. 5: Comparison of the prediction performance of the expected travel time between the linear regression (baseline) method and the proposed (WATS) model.

reliable estimator of the ground truth. We therefore employed only the test cases that included 30 or more actual travel time observations per hour. We obtained 141 test cases from 68 segments.

Figure 4 shows the distribution of prediction error with the WATS model. Here it can be seen that the error distribution has a sharp peak around zero. The absolute prediction error was less than 2.5 seconds for more than 40% of the test cases and less than 20 seconds for 90%.

Figure 5 shows the comparison of the prediction error between the WATS model and the linear regression method. For this evaluation, the linear regression model was trained for each segment using the training data, with the hourly weather observation value $\boldsymbol{w}$ being the input value and with the mean travel time per hour being the target value. In the left figure, each point corresponds to a test case, and the error of the proposed method against the baseline method is shown. The plots are substantially along the identity line shown as a red line, and there are both improved and worse cases. The right figure shows the distribution of absolute prediction error of the two methods. The distribution seems to be exponential, and the median for the WATS model was less than that of the linear regression model. However, it is clear that the difference between two distributions is not very significant and the performances of the two methods are comparable.

Cross Entropy From the viewpoint of probabilistic modeling, cross entropy between the distribution estimated by the model and that of the training data set is also an interesting evaluation metric. It is equal to the average negative log-likelihood over the test data. Smaller entropy values indicate better explanations of the data set by the probabilistic model and therefore such a model is considered to be better at summarizing the data. The linear regression is regarded as a Gaussian model; i.e., the target variable follows a Gaussian distribution with mean represented by a linear formula and the variance is a constant. The variance estimator is the variance of the residual in the training data set.

In this evaluation, the linear regression model was trained for each segment, with the hourly weather observation value $\boldsymbol{w}$ being the input value and with the actual travel time being the target value, so that the cross entropy of the actual travel time data can be evaluated. The test data included 56,186,309 observation values from 31,891 segments. The cross entropy for the linear model was 4.57 while that for the WATS model was 3.35 (73% of the value for the linear model), calculated using natural logarithms. This shows that our method fitted the data better than did the linear regression method.

3.4 Discussion

Our model has confirmed that traffic patterns depend on weather conditions as well as time and location. It shows traffic smoothness through the travel time or travel speed distributions under different weather conditions, which are characterized by values such as temperature and snow depth, and the model shows the chronic or weather-sensitive bottlenecks. Road administrators and experts might use this knowledge to improve roads before disaster occurs, or for planning snow removals after a heavy snowfall. Thanks to the mixing coefficient for each road segment and for each weather state in our model, clustering analysis could also be applied over the segments. If it works, it will help us to understand the characteristics for each road segment and time period by grouping similar segments such as "susceptible to snow."

The prediction performance of the expected value using our WATS model was comparable to that with linear regression. This result suggests that statistics such as expected value can be learned by a simple model and that there is potentially a linear relationship between weather and traffic. However, the WATS model reduced the entropy of data and therefore fitted the data better than the linear regression model, which cannot explain the individual data points.

It is presumably caused by the multimodality of the travel time distribution. Typically, there are several modes in ordinary roads. Vehicles that pass through the road link without stopping are said to be in "smooth" mode. Vehicles may also stop at an intersection or a traffic signal, so the mean speed slows down in this case; such vehicles are in "stop" mode. Because of this multimodality, the expected value can take a value with a low probability density. In other words, the actual observed values tend to differ from the expected value, and therefore, it seems not to be reasonable to suppose that an observed value is unusual or anomalous just because it differs from the expected value.

Cross entropy is the average information of a data point. Less entropy indicates less information per data point, so that the data become more predictable. In other words, a low-entropy model can regard most of the data as usual. From the entropy evaluation, the proposed model regards larger amounts of data as more "usual" than does the linear regression. Further work is under way to conduct anomaly detection using the WATS model, whereby traffic incidents or sudden bottlenecks could be found in real time.

Our analysis is empirical but represents a first step in understanding the social–physical space. In this paper, we have used travel time as the feature value for the traffic on a road segment and assumed that it follows a gamma distribution. However, feature selection and probabilistic modeling are still open to discuss. Other features such as traffic flow and density, which might be obtained from data sources other than probe vehicles, would be worth considering. Knowledge information such as traffic signal timing will also be helpful to improve graphical model structure and probability distribution functions for mixture components. The model estimation algorithm also requires further development. Model selection, such as determining K and L, is an important open problem. Bayesian inference may also improve model estimation.

4 Conclusion

We have studied a probabilistic model to describe traffic observation data using both traffic and weather conditions. We proposed the WATS model, which is a latent variable model to relate traffic data to weather data by borrowing the idea of the Pachinko allocation model. We have conducted an empirical winter traffic analysis in the city of Sapporo, Japan, by applying the WATS model to real weather and traffic data. The model showed the relationship of the expected vehicular speed to weather conditions, and showed the potential bottleneck segments according to the weather conditions. The quantitative evaluation showed that the WATS model fits the data better than the linear regression model, which suggests the potential for anomaly detection from vehicular observation data.

Acknowledgment This work was supported by the CPS-IIP Project in the research promotion program for national-level challenges "Research and development for the realization of next-generation IT platforms" of the Ministry of Education, Culture, Sports, Science and Technology, Japan. Weather observation data were provided by the Japan Meteorological Agency. The authors received some information on winter traffic in Sapporo from specially appointed assistant professor Dr. Hajime Imura at Hokkaido University, Japan.

References

1. Bishop, C.M.: Mixture Models and EM. In: Pattern Recognit. Mach. Learn., chap. 9, pp. 423–460. Springer, New York, NY, USA (2006)
2. Cao, P., Miwa, T., Morikawa, T.: Modeling distribution of travel time in signalized road section using truncated distribution. In: Proc. 9th Int. Conf. Traffic Transp. Stud. (ICTTS 2014), vol. 138, pp. 137–147. Elsevier (2014). DOI 10.1016/j.sbspro.2014.07.189
3. City of Sapporo: Actual budget for the counterplan against snow. http://www.city.sapporo.jp/kensetsu/yuki/jigyou/budget.html. Accessed on Nov. 30, 2016.
4. Guessous, Y., Aron, M., Bhouri, N., Cohen, S.: Estimating travel time distribution under different traffic conditions. Transp. Res. Procedia **3**(July), 339–348 (2014). DOI 10.1016/j.trpro.2014.10.014
5. Hall, F.L., Barrow, D.: Effect of weather on the relationship between flow and occupancy on freeways. Transportation Research Record **1194**, 55–63 (1988)
6. Japan Meteorological Agency: Observations. http://www.jma.go.jp/jma/en/Activities/observations.html. Accessed on Nov. 23, 2016.
7. Keay, K., Simmonds, I.: The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia. Accident Analysis and Prevention **37**(1), 109–124 (2005). DOI 10.1016/j.aap.2004.07.005
8. Kinoshita, A., Takasu, A., Adachi, J.: Real-time traffic incident detection using a probabilistic topic model. Inf. Syst. **54**, 169–188 (2015). DOI 10.1016/j.is.2015.07.002
9. Lee, J., Hong, B., Lee, K., Jang, Y.J.: A prediction model of traffic congestion using weather data. In: 2015 IEEE Int. Conf. Data Sci. Data Intensive Syst., pp. 81–88. IEEE (2015). DOI 10.1109/DSDIS.2015.96
10. Li, W., McCallum, A.: Pachinko allocation: DAG-structured mixture models of topic correlations. In: Proc. 23rd Int. Conf. Mach. Learn. (ICML '06), pp. 577–584. ACM (2006). DOI 10.1145/1143844.1143917
11. Tanimura, R., Hiromori, A., Yamaguchi, H., Higashino, T., Umedu, T.: Prediction of deceleration amount of vehicle speed in snowy urban roads using weather information and traffic data. In: 2015 IEEE 18th Int. Conf. Intell. Transp. Syst., pp. 2268–2273. IEEE (2015). DOI 10.1109/ITSC.2015.366
12. Turksma, S.: The various uses of floating car data. In: Tenth Int. Conf. Road Transp. Inf. Control, pp. 51–55. IET, London, England, UK (2000). DOI 10.1049/cp:20000103
13. Xu, X., Su, B., Zhao, X., Xu, Z., Sheng, Q.Z.: Effective traffic flow forecasting using taxi and weather data. In: Proceedings of the 12th International Conference on Advanced Data Mining and Applications (ADMA '16), pp. 507–519. Springer International Publishing, Gold Coast, QLD, Australia (2016). DOI 10.1007/978-3-319-49586-6_35
14. Yang, Q., Wu, G., Boriboonsomsin, K., Barth, M.: Arterial roadway travel time distribution estimation and vehicle movement classification using a modified Gaussian mixture model. In: 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), pp. 681–685. IEEE, The Hague, The Netherlands (2013). DOI 10.1109/ITSC.2013.6728310