Probabilistic Weather Forecasting for Winter Road Maintenance

Veronica J. Berrocal, Adrian E. Raftery, Tilmann Gneiting, and Richard C. Steed[1]

Technical Report no. 511
Department of Statistics
University of Washington

April 3, 2007

[1]Veronica J. Berrocal (Email: veronica@stat.washington.edu) is a Graduate Research Assistant, Adrian E. Raftery (Email: raftery@u.washington.edu) is Blumstein-Jordan Professor of Statistics and Sociology, and Tilmann Gneiting (Email: tilmann@stat.washington.edu) is Associate Professor of Statistics, all at the Department of Statistics, University of Washington, Seattle, WA 98195-4322. Richard C. Steed is Research Meteorologist, Department of Atmospheric Sciences, University of Washington, Box 351640, Seattle, WA 98195-1640. This research was supported by the DoD MultiDisciplinary Research Initiative (MURI) program administered by the Office of Naval Research under Grant N00014-01-10745.

Abstract

Road maintenance is one of the main problems Departments of Transportation face during winter time. Anti-icing, i.e. applying chemicals to the road to prevent ice formation, is often used to keep the roads free of ice. Given the preventive nature of anti-icing, accurate predictions of road ice are needed. Currently, anti-icing decisions are usually based on deterministic weather forecasts. However the costs of the two kinds of error are highly asymmetric because the cost of a road closure due to ice is much greater than that of taking anti-icing measures. As a result, probabilistic forecasts are needed to optimize decision-making.

We propose two methods for forecasting the probability of ice formation. Starting with deterministic numerical weather predictions, they produce a joint predictive probability distribution of temperature and precipitation. This then yields the probability of ice formation, defined here as the occurrence of precipitation when the temperature is below freezing. In the first method, temperature and precipitation at different spatial locations are treated as conditionally independent given the numerical weather predictions. In the second method, spatial dependence between forecast errors at different locations is modeled. The model parameters are estimated using a Bayesian approach via Markov chain Monte Carlo. We evaluated both methods by comparing their probabilistic forecasts with observations of ice formation for Interstate Highway 90 in Washington State for the 2003–2004 and 2004–2005 winter seasons. Results showed that using probabilistic forecasts can save a considerable amount of money compared with deterministic forecasts. The method that takes account of spatial dependence improved the reliability of the forecast, but did not save more money.

Keywords: Numerical weather forecast; Predictive distribution; Spatial dependence; Markov chain Monte Carlo; Latent Gaussian process; Cost-loss ratio.

Contents

List of Tables

List of Figures

1 Introduction

Ice and snow on roads have large impacts. Failure to maintain roads in winter often leads to road closures and hence economic losses (Sherif and Hassan 2004). The Washington State Department of Transportation estimated that between 1992 and 2004, Snoqualmie Pass on Interstate-90 (I-90) had been closed on average 120 hours per year, causing an annual loss of at least $17.5 million dollars. Ice and snow also increase the risk of accidents (Norrman et al. 2000; Eriksson and Norrman 2001). The crash rate on the I-90 Mountains to Sound Greenway, Washington State's primary east-west bound highway, in the presence of snow is about five times the rate in clear conditions (Federal Highway Administration 2006).

Currently, two main strategies are used for winter road maintenance: de-icing, in which chemicals are used to melt ice and snow, and anti-icing, a preventive measure that reduces ice by hindering bonds between ice crystals and road pavement. Due to the use of chemicals, both strategies hurt the environment: soil, vegetation, streams, road surface and vehicles are damaged by the chemicals used in both de-icing and anti-acing (Shao and Lister 1996; Ramakrishna and Viraraghavan 2005). However, anti-icing is preferred to de-icing on roads with high traffic volumes because it reduces total chemical use and allows a higher level of service to the public.

The costs of winter road maintenance are high, but the losses due to road closures are much higher, so accurate ice forecasts are needed (Chapman et al. 2001a; Shao 1998). Currently, forecasts of road ice come primarily from numerical road prediction models or numerical weather prediction models. Numerical road prediction models take weather forecasts and road condition data as inputs, and forecast future road conditions using the surface energy-balance equation, which describes the fluxes of energy between the atmosphere and the road (Sass 1992). Numerical weather prediction models describe the atmosphere using seven differential equations, and forecast future weather by integrating them forward in time. Both kinds of forecast are deterministic and do not assess uncertainty, which is important in weather-related decision making (Palmer 2000; Richardson 2000; American Meteorological Society 2002; Gneiting and Raftery 2005; Roulston et al. 2006; National Research Council of the National Academies 2006).

The cost of failing to take anti-icing measures when ice does form on the road is much greater than that of anti-icing when no ice forms. As a result, it would be best to take anti-icing measures when the predicted probability of ice is greater than some threshold that is typically well below 50%. Thus a good estimate of the probability of ice is needed, and deterministic forecasts do not provide it. In this paper we present methods for estimating this, building on the work of Mass et al. (2003), who document the joint effort of the

Department of Atmospheric Sciences at the University of Washington and of the Washington State Department of Transportation to monitor and provide forecasts of road ice for I-90 by postprocessing numerical weather forecasts.

The physical process that produces ice on roads is complex and involves the interaction of weather with road surface conditions (Chapman et al. 2001a; Thornes et al. 2005). Here we simplify it, and we assume that ice will form at a point on the road if precipitation occurs and the air temperature is at or below freezing. Starting from numerical weather forecasts, we develop a model for the joint predictive probability distribution of temperature and precipitation, which yields the probability of ice formation.

The problem of winter road maintenance is intrinsically spatial. We examine whether spatial dependence should be modeled explicitly, by comparing two models, one that does not take account of spatial correlation and one that does. The latter produces joint predictive probability distributions of future temperature and precipitation over space that account for the spatial correlation of the forecast errors.

In Section 2, we describe the data and introduce the statistical models and the estimation methods that we use. In Section 3, we give verification results, and in Section 4 we discuss other approaches that have been developed for this problem, and possible limitations of our methodology.

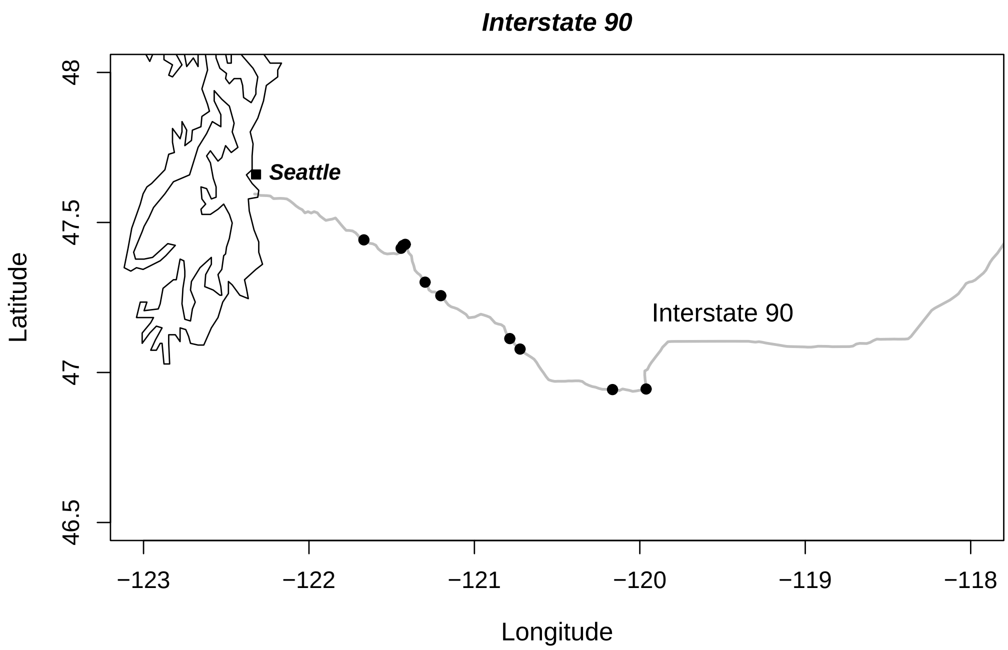
2 Data and methods

2.1 Road maintenance problem

The I-90 Mountains to Sound Greenway is one of the main arteries of the State of Washington, with a traffic volume of 20 million vehicles per year. The highway crosses the Cascade Mountains with large changes in altitude and connects the urban centers of Puget Sound with farmlands in Eastern Washington; it is an important route for the economy of the region. During the winter months from October to March, I-90 is often congested because of poor driving conditions due to snow accumulation and ice. To provide a safe transit to travelers, the Washington State Department of Transportation employs preventive anti-icing measures on I-90 during the winter season. In this study, we use data from ten meteorological stations along the section of I-90 that includes Snoqualmie Pass. Figure 1 shows the altitude profile for the portion of I-90 considered in this study.

The data span the three winter seasons of 2002–2003, 2003–2004, and 2004–2005, and include minimum temperature and precipitation during the three-hour time interval 1:00am–4:00am. Precipitation occurrence was recorded if there was at least 0.01 inch of precipitation. Geographical information (latitude, longitude and elevation) for the observation sites was

(a)



(b)

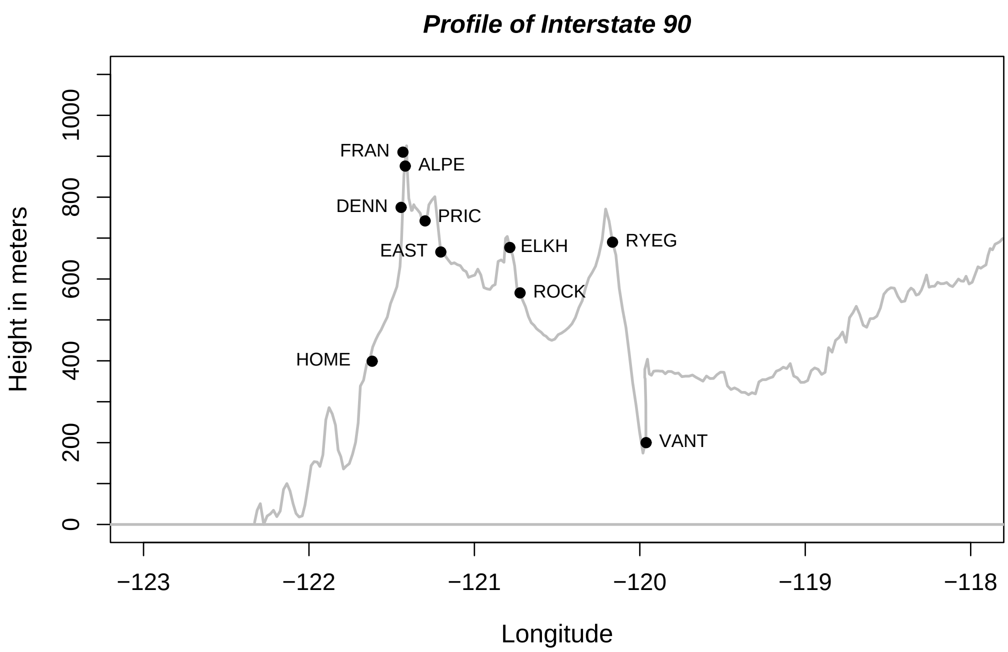


Figure 1: (a) Section of I-90 and stations along it. (b) Altitude profile of I-90 and stations along it.

also available. From now on, we will use the term "temperature" to refer to minimum temperature between 1:00am and 4:00am.

The data comprise 438 days of observations: 131 during the 2002–2003 winter season, 159 during the following winter season and 148 during the 2004–2005 season. The number of observations per station ranges from a minimum of 169 to a maximum of 363, with an average of 307 and a median of 323.

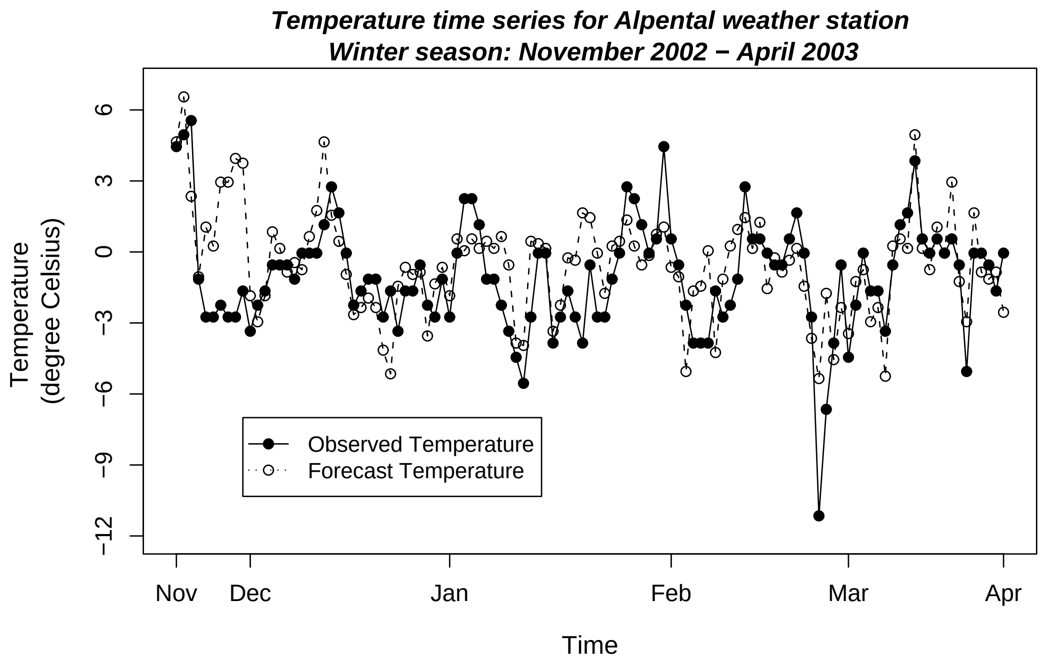
We use 12-hour-ahead forecasts of temperature and precipitation produced by MM5, the fifth-generation Pennsylvania State University—National Center for Atmospheric Research Mesoscale Model (Grell et al. 2004). The model was run by the University of Washington Department of Atmospheric Sciences with initial and boundary conditions supplied by the United Kingdom Meteorological Office (Grimit and Mass 2002; Eckel and Mass 2005). The forecasts were generated on a 12 km grid and then bilinearly interpolated to the observation sites. These forecasts are based on the information available at 4:00pm the preceding day, which is typically the most recent available to managers deciding whether to take anti-icing measures overnight. Figure 2 shows the weather data at the Alpental weather station. This is close to Snoqualmie Pass and is one of the highest stations on I-90. The MM5 model forecasted temperature quite accurately: the mean absolute error for the forecast at Alpental during the winter season 2002–2003 was $1.6°C$. The forecast of precipitation occurrence at Alpental was reasonably good during the 2002–2003 winter season. For example, in 59 of the 64 cases in which more than 0.01 inch of rain was predicted, precipitation was indeed observed. In 33 of the 50 cases in which the forecast was for no precipitation, the forecast was correct and there was no precipitation.

We will say that a point along I-90 requires anti-icing if the temperature is equal or less than $0°C$ and there is precipitation. Similarly, we will say that a section of I-90 requires anti-icing if there is at least one point in the section at which the temperature is at most $0°C$ and there is precipitation. Our goal is to produce probabilistic forecasts of ice formation along a section of I-90.

2.2 Statistical model

We now describe our statistical model for ice formation at a given time. The time is fixed, and so is not explicitly included in our notation. We denote by $\mathbf{Y}(s)$ and $\tilde{\mathbf{Y}}(s)$ the observed temperature and the 12-hour forecast of temperature at a point s along the portion $\mathcal{I}$ of I-90 shown in Figure 1. We follow Sloughter et al. (2007) in using the cube root of the forecast of accumulated precipitation as the basis for our model of precipitation. We let $\mathbf{W}(s) = 1$ if precipitation occurred at s and 0 if not. We denote by $\tilde{\mathbf{W}}(s)$ the cube root of the forecasted

(a)



(b)

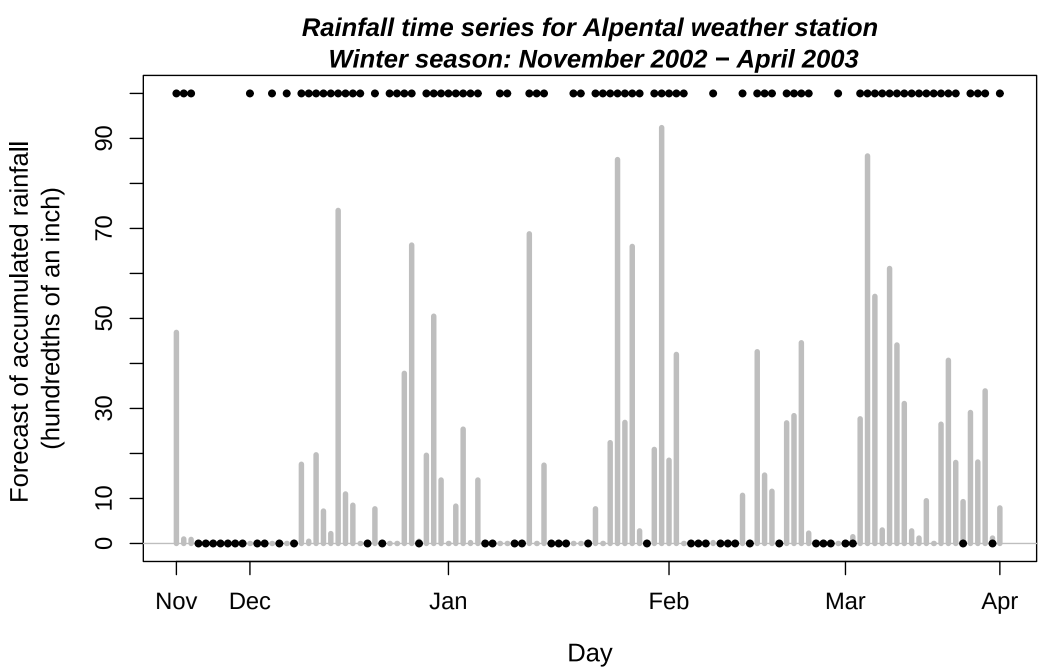


Figure 2: (a) Temperature time series for the Alpental weather station during the 2002–2003 winter season. The solid line refers to the observed temperature, while the dashed line refers to the forecast. (b) Time series of precipitation occurrence (black dots), and forecasts (grey) of accumulated precipitation during the three-hour interval 1:00am–4:00am.

accumulated precipitation amount at s. We will say that ice forms on the road at s if

$$\mathbf{Y}(s) \leq 0°C \text{ and } \mathbf{W}(s) = 1. \tag{1}$$

Our goal is to produce probabilistic forecasts of ice formation simultaneously at all sites s on a grid of points along $\mathcal{I}$.

Given our definition of ice formation, in order to produce probabilistic forecasts of ice, we need to specify a joint model for temperature and precipitation occurrence. We assessed the dependence between temperature and precipitation occurrence given the forecasts by dividing the observations into groups with similar values of the forecasts. Within each group we found no evidence of dependence. We therefore assume that temperature and precipitation occurrence are conditionally independent of one another given the forecasts. We now specify two models, each of them for the joint distribution of temperature and precipitation occurrence.

In our first model, which we call the marginal model, we assume that temperatures at different locations are independent of one another, as are occurrences or not of precipitation, given the forecasts of temperature and accumulated precipitation. To remove systematic nonstationary variation, we include a regression-based adjustment of the mean temperature field on latitude, longitude and elevation, similarly to Handcock and Wallis (1994) in their analysis of the average U.S. winter temperature. The marginal model for temperature is then

$$\mathbf{Y}(s) = \gamma_0 + \gamma_1 \tilde{\mathbf{Y}}(s) + \gamma_2 \text{lat}(s) + \gamma_3 \text{lon}(s) + \gamma_4 \text{ht}(s) + \epsilon(s), \tag{2}$$

where lat, lon and ht denote the centered latitude, longitude and elevation (measured in meters), $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ and γ_4 are regression coefficients, and $\epsilon(s)$ is a mean-zero Gaussian process with

$$\text{Cov}(\epsilon(s), \epsilon(t)) = \varsigma^2 \delta_{st}, \tag{3}$$

where $\delta_{st} = 1$ if $s = t$ and 0 otherwise.

We now describe our spatial model, which takes account of spatial correlation in temperature and precipitation. The mean temperature field is modeled using the same regression adjustment as in the marginal model. The temperature field is assumed to have a stationary, isotropic exponential covariance function with a nugget effect. The model is:

$$\mathbf{Y}(s) = \alpha_0 + \alpha_1 \tilde{\mathbf{Y}}(s) + \alpha_2 \text{lat}(s) + \alpha_3 \text{lon}(s) + \alpha_4 \text{ht}(s) + \xi_T(s), \tag{4}$$

where $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ and α_4 are regression coefficients, and $\xi_T(s)$ is a mean-zero stationary isotropic Gaussian process with covariance function

$$\text{Cov}(\xi_T(s), \xi_T(t)) = \rho^2 \delta_{st} + \tau^2 \exp\left(-\frac{|s-t|}{r}\right), \tag{5}$$

where $|s-t|$ is the distance between sites s and t. The covariance parameters are the nugget effect, ρ^2, the sill, $\sigma^2 = \rho^2 + \tau^2$, equal to the marginal variance, and the range, r, which specifies the rate at which the exponential correlation decays (Cressie 1993; Chilès and Delfiner 1999).

We now describe our models for precipitation occurrence. In the marginal model we assume conditional spatial independence for precipitation occurrence given the forecast of precipitation, so that

$$\mathbf{W}(s) = \begin{cases} 1 & \text{with probability } \pi(s), \\ 0 & \text{with probability } 1-\pi(s), \end{cases} \tag{6}$$

where

$$\log\left(\frac{\pi(s)}{1-\pi(s)}\right) = \lambda_0 + \lambda_1 \tilde{\mathbf{W}}(s) + \lambda_2 \mathrm{ht}(s), \tag{7}$$

and $\mathbf{W}(s)$ and $\mathbf{W}(t)$ are independent given $\tilde{\mathbf{W}}(s)$, $\tilde{\mathbf{W}}(t)$, $\mathrm{ht}(s)$ and $\mathrm{ht}(t)$. In this case, the inclusion of latitude and longitude made no difference and so they were not included in (7).

In specifying our spatial model for precipitation occurrence we follow Albert and Chib (1993), and extend their hierarchical model for independent binary data to spatially dependent binary data. We postulate a latent Gaussian process $\mathbf{Z}(s)$ that regulates precipitation occurrence. If the latent variable $\mathbf{Z}(s)$ is greater than 0, then there is precipitation at the site; otherwise there is no precipitation at s.

The mean of $\mathbf{Z}(s)$ is a linear combination of the cube root of the forecast, and elevation. The covariance structure of $\mathbf{Z}(s)$ is modeled using an exponential covariance function. To ensure identifiability, the marginal variance of $\mathbf{Z}(s)$ is set equal to 1, and the only covariance parameter is the range parameter θ, which gives the rate of decay of the spatial correlation.

The complete spatial model for precipitation occurrence is thus:

$$\mathbf{W}(s) = \begin{cases} 1 & \text{if } \mathbf{Z}(s) > 0, \\ 0 & \text{otherwise,} \end{cases} \qquad s \in \mathcal{I}, \tag{8}$$

$$\mathbf{Z}(s) = \beta_0 + \beta_1 \tilde{\mathbf{W}}(s) + \beta_2 \mathrm{ht}(s) + \xi_P(s), \tag{9}$$

where $\xi_P(s)$ is a mean-zero unit variance Gaussian process with covariance function

$$\mathrm{Cov}(\xi_P(s), \xi_P(t)) = \exp\left(-\frac{|s-t|}{\theta}\right). \tag{10}$$

2.3 Model fitting

For each day, the parameters of models (2) through (10) were estimated using data from a "sliding window" training period consisting of the previous N days. The parameters of the marginal model (2) and (3) were estimated by fitting a linear regression of observed temperature on the forecasted temperature, latitude, longitude and elevation. The parameters of the marginal model (6) and (7) were estimated by fitting a logistic regression of the observed precipitation occurrence on the cube root of the forecasted accumulated precipitation and elevation. The data used to fit both regressions were observations and forecasts for the ten stations located along $\mathcal{I}$ relative to the previous N days. The choice of N is explained in Section 2.4.

For the spatial model, the parameters for temperature are computed in two stages: first the regression coefficients $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ and α_4 and the marginal variance σ^2 are evaluated, then the covariance parameters, ρ^2, τ^2 and r are determined. Since the training period is a sliding window consisting of the N days preceding the verification time, only a limited amount of data is available. This can render the ordinary least squares estimates of the regression coefficients and of the marginal variance unstable, as Figure 3 shows. Therefore, to estimate the model parameters for temperature, we use a Bayesian framework and we adopt the posterior means as estimates of the regression coefficients, $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ and α_4, and of the marginal variance, σ^2.

Our estimates of $\alpha_0, \alpha_1, \alpha_2, \alpha_3$, α_4 and σ^2 are based on the following Bayesian model:

$$\mathbf{Y}(s) = \alpha_0 + \alpha_1 \tilde{\mathbf{Y}}(s) + \alpha_2 \text{lat}(s) + \alpha_3 \text{lon}(s) + \alpha_4 \text{ht}(s) + \xi'_T(s) \tag{11}$$

$$\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)^T | \sigma^2 \quad \sim \quad \mathcal{MVN}_5(\eta, \sigma^2 \Omega) \tag{12}$$

$$\sigma^2 \quad \sim \quad InvGamma(\nu, \psi) \tag{13}$$

where $\xi'_T(s)$ is a mean-zero Gaussian process with covariance function $\text{Cov}(\xi'_T(s), \xi'_T(t)) = \sigma^2 \delta_{st}$, η is a vector with components $(\eta_0, \ldots, \eta_4)$ and Ω is a diagonal matrix with diagonal elements $(\omega_0^2, \ldots, \omega_4^2)$. This is a simplification in that the spatial structure is not included in this model, but the verification results in Section 3 indicate that this did not hurt the out-of-sample predictions from the spatial model.

We based the prior distributions of $\alpha_0, \alpha_1, \ldots, \alpha_4$ and σ^2 on the data from the first winter season available to us, that of 2002–2003. For each day in the 2002–2003 winter season, we estimated the regression parameters by fitting a linear regression with data from the previous N days. The prior mean and prior standard deviation of α_0 were the mean and twice the standard deviation of the resulting estimates of the α_0; similarly for $\alpha_1, \ldots, \alpha_4$. Table 1 reports the values of the prior mean and the prior standard deviation of $\alpha_0, \alpha_1, \ldots, \alpha_4$ with training period of length $N = 20$.

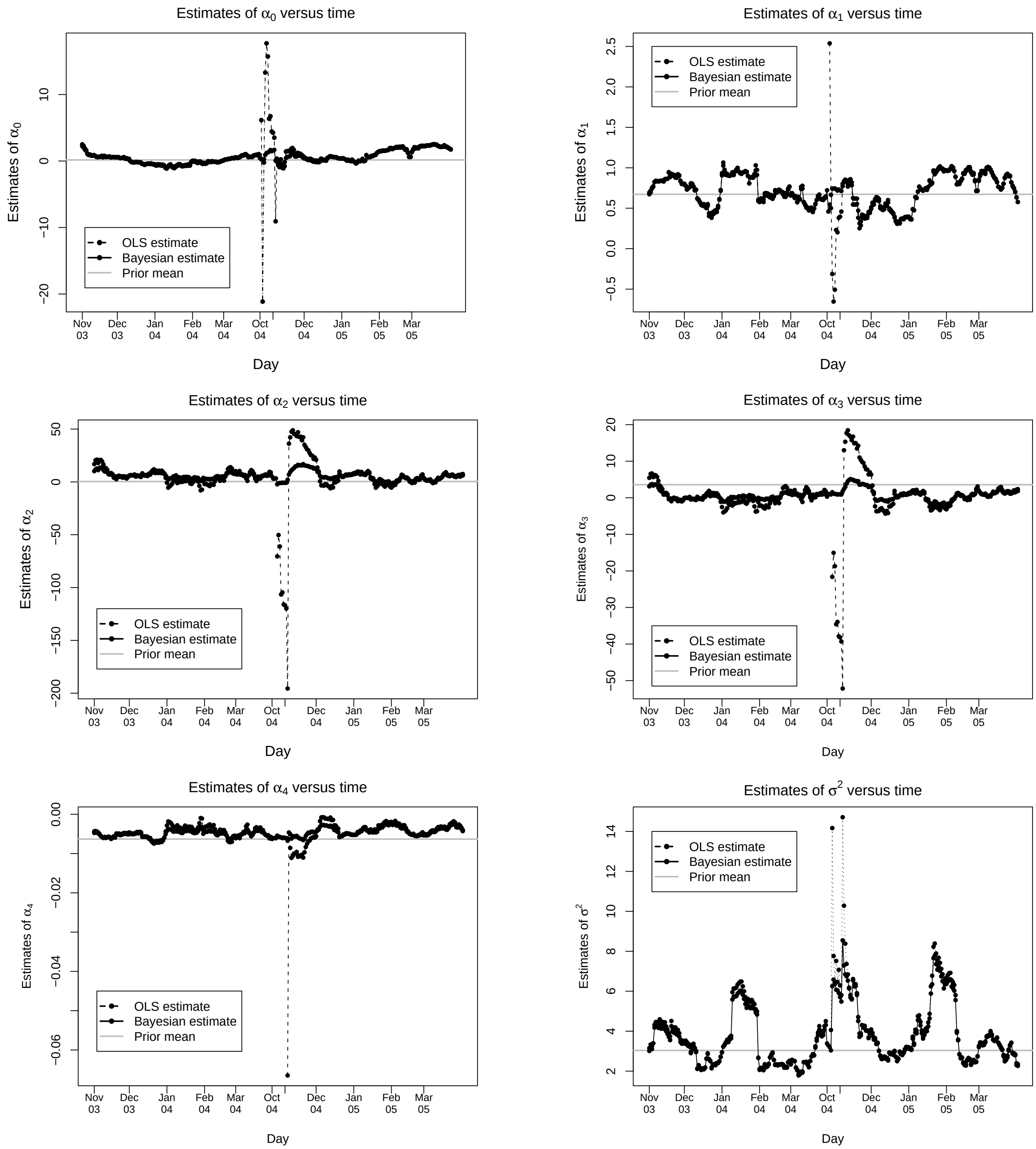


Figure 3: Bayesian estimates of $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4$ and σ^2 versus time (solid black line). In each panel, the dashed black line shows the ordinary least squares estimates of the parameter, and the solid grey line is the prior mean.

Table 1: Prior mean and prior standard deviation of α_0, α_1, α_2, α_3, α_4 for training period length $N = 20$.

	α_0	α_1	α_2	α_3	α_4
Prior mean	0.164	0.672	0.366	3.578	–0.006
Prior standard deviation	0.701	0.205	0.473	1.166	0.001

To specify the hyperparameters of the inverse Gamma prior distribution of σ^2, we minimized the sum of squared deviations between the cumulative distribution function of an inverse Gamma distribution and the empirical distribution of the estimates of σ^2. The resulting 10th, 50th and 90th percentiles of the prior distribution of σ were $1.36°C$, $1.67°C$ and $2.12°C$.

The posterior means of the regression parameters are given by standard closed form expressions (e.g. Gelman et al. 2004, Chapter 14). As can be seen from Figure 3, using a Bayesian approach stabilized the ordinary least squares estimates when they were unstable, but otherwise hardly changed them at all, which was our goal.

To estimate the covariance parameters ρ^2 and r, which are assumed to be constant in time, we used the data from the 2002–2003 winter season. We constructed the empirical variogram of the residuals of the linear regression of the observed temperature on the centered latitude, longitude and elevation, and on the forecast of temperature. We then fitted a parametric exponential variogram with nugget effect to the empirical variogram using Weighted Least Squares (Cressie 1993), where the weights were the numbers of times each pair of stations had been observed simultaneously. Figure 4 shows the empirical variogram of the residuals of temperature and the fitted parametric exponential variogram.

The marginal variance, σ^2, varies with time and is re-estimated for each day. We ascribe the variability in time of the marginal variance to the variability in time of the variance of the continuous component of (5). We thus model ρ^2, the small scale variability, as constant in time, while we allow τ^2 to change with time and re-estimate it for each day. From (5) it follows that, for each day, we can obtain a new estimate of τ^2 by simply subtracting the estimate of ρ^2 from the estimate of σ^2.

In estimating the parameters of the spatial model for precipitation occurrence, the parameters of the model given by (8), (9), and (10), were estimated for each day in a Bayesian way using data from a training period made up of the previous N days. We used the following prior for β and θ:

$$\beta = (\beta_0, \beta_1, \beta_2)^T \quad \sim \quad \mathcal{MVN}_3(\mu, \mathbf{V}), \tag{14}$$

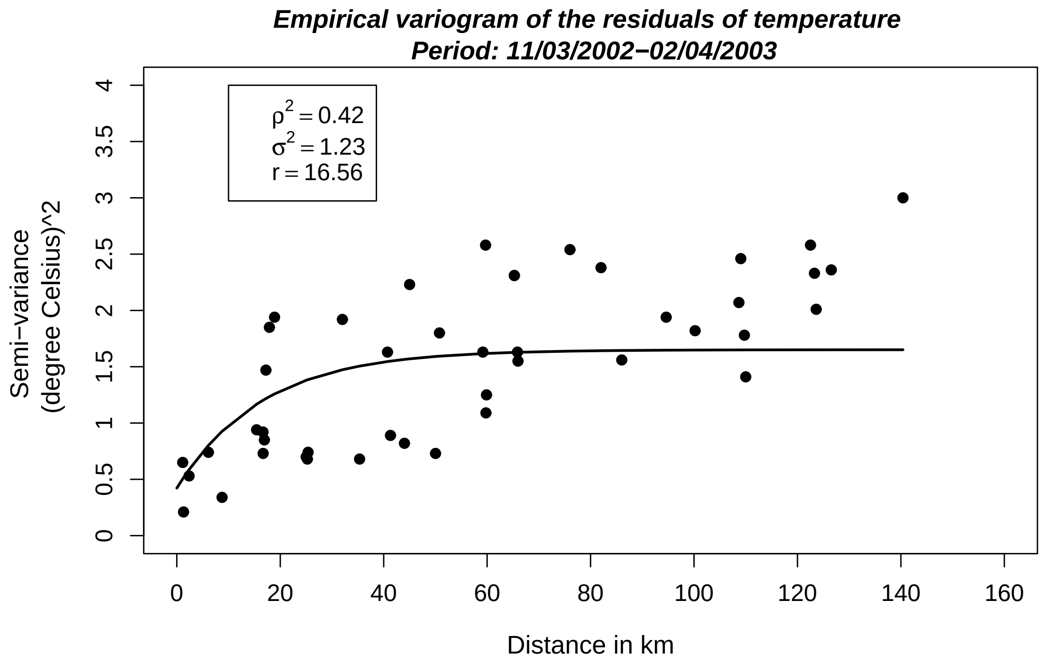


Figure 4: Empirical variogram of the residuals with fitted exponential variogram. In the plot, each dot represents a pair of stations.

$$\theta \propto \text{Unif}\,(0, 1000). \tag{15}$$

As for temperature, we used spread-out prior distributions with hyperparameters based on data from the 2002–2003 winter season. For each day in the 2002–2003 season, we fit a logistic regression of the observed precipitation occurrence on the cube root of the forecast of accumulated precipitation and on elevation to data from a training period made up of the previous N days. This yielded a set of estimates of β_0, β_1 and β_2. The prior mean μ then consisted of the means of the estimates of β_0, β_1 and β_2, while V was diagonal with diagonal elements equal to four times the variances of the estimates of β_0, β_1 and β_2.

For each day, the parameters of the spatial hierarchical model for precipitation, given by (8), (9), (10), (14) and (15), were estimated using a Markov Chain Monte Carlo algorithm (MCMC; Gelfand and Smith 1990). Since a closed form for the full conditional distributions of $\mathbf{Z}(s)$ and β is available, we used a Metropolis-Hastings step to update the covariance parameter θ, and a Gibbs sampler algorithm to update all other parameters.

2.4 Choice of training period

In choosing the length of the training period there is a trade-off: a shorter period allows changes in the atmosphere to be taken into account more promptly, but on the other hand

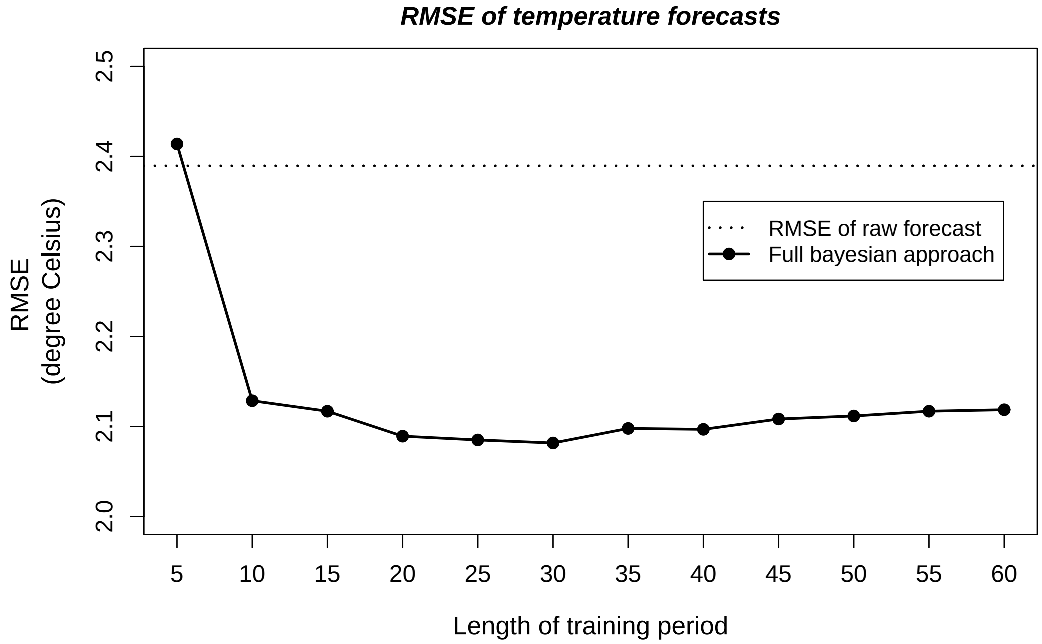


Figure 5: RMSE of temperature predictions versus training period length. Predictions are made only for the 2003–2004 and 2004–2005 winter seasons. The month of October is always excluded from the RMSE computation.

it reduces the amount of data available for estimation of the parameters.

To choose the length of the training period we used the Root Mean Square Error (RMSE) of the temperature predictions obtained using training periods of lengths $N = 5, 10, \ldots, 60$ days. Figure 5 shows the RMSE of the predictions as a function of the training period length. The magnitude of the error decreased noticeably as the length N of the training period increased up to 20 days. Beyond 20 days, there was not much gain in using a longer training period, and the quality of the predictions worsened slightly for training periods longer than 30 days.

We therefore used a training period of 20 days to estimate the model parameters for both temperature and precipitation. However, different training period lengths may be appropriate for different forecast lead times and regions.

2.5 Generating forecasts

To produce probabilistic forecasts of ice formation along the section $\mathcal{I}$ of I-90 in Figure 1, we discretized the problem and simulated from the joint predictive distribution of temperature and precipitation occurrence at 104 points along the highway. The distances between neighboring points range from 1.3 km to 2.0 km and cover a section of I-90 about 140 km long.

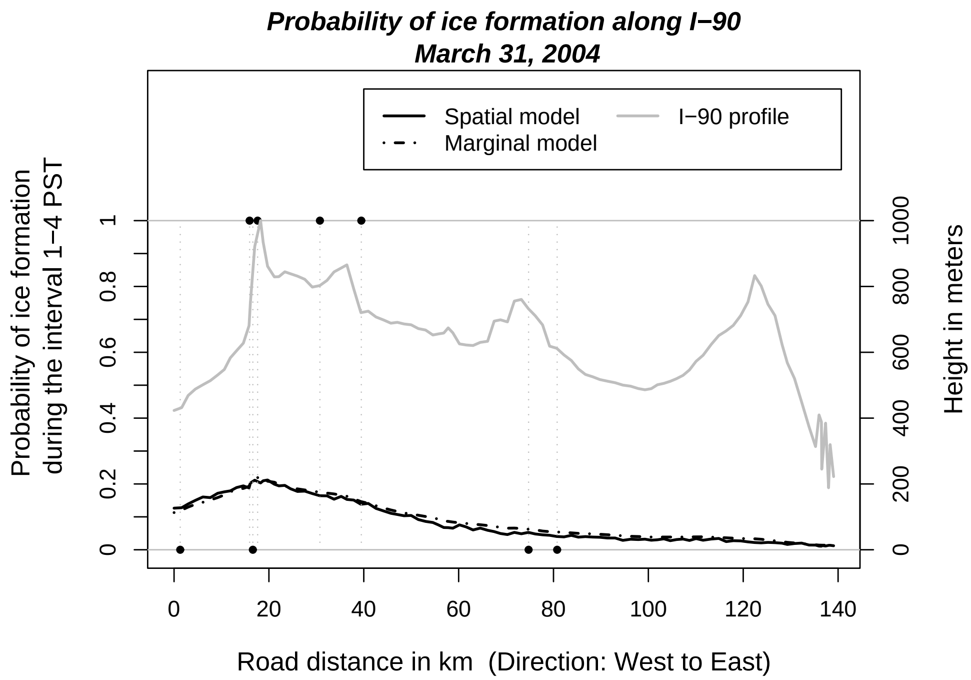


Figure 6: Probability of ice formation along I-90 on March 31, 2004, as forecast by the spatial model (solid black line) and the marginal model method (dot-dash black line). The profile of I-90 is shown in grey, while observations are represented with black dots.

Realizations of the temperature and precipitation occurrence fields were obtained by simulating from the stochastic processes given by (2) and (3), and (6) and (7) for the marginal model and from the stochastic processes given by (4) and (5), and (8), (9) and (10) for the spatial model, using parameters estimated from the training data. We defined a realization as forecasting ice at one of the 104 points if it gave both precipitation and temperature at or below freezing at that point on the road. The probability of ice on the road at that location was then the proportion of the realizations forecasting ice.

Figure 6 maps these probabilities for one day for both models. Not surprisingly, they are similar between the two models, because the real differences between the models are for the joint distribution of ice at different places, not for the probability of ice at one place.

To verify our probabilistic forecasts, we used the observations at the ten stations, for which the forecast probabilities can be computed analytically. We did not observe directly whether or not there actually was ice on the road, and instead we say that ice was observed if the temperature was at or below freezing and there was precipitation.

Comparisons of forecasts and observations at individual sites assess the performance of the predictive distributions only marginally. To evaluate the probabilistic forecasts of the ice field as a whole, we forecasted ice formation simultaneously at all ten observation sites. This was done using the procedure described at the beginning of this section, replacing the 104 points with the ten observation sites. For a given realization, we say that ice was predicted along the section $\mathcal{I}$ of I-90 if ice formation was forecasted at at least one of the ten sites. We say that ice was observed along $\mathcal{I}$ if temperature was at or below freezing and there was precipitation at at least one of the stations.

To distinguish between the two different types of verification — marginal and spatial — we will use the expression "ice at observation sites" for verification results relative to forecasts of ice at the individual sites, and "ice along $\mathcal{I}$" to refer to forecasts of ice simultaneously at the ten observation sites.

3 Results

We now compare the out-of-sample predictive performance of our probabilistic forecasting methods with that of the deterministic forecast for the winter seasons 2003–2004 and 2004–2005, for individual sites and for the entire section $\mathcal{I}$ of I-90. We refer to a deterministic forecast from the numerical weather prediction model as a "raw" forecast. We define this as predicting ice if it forecasts both freezing temperature and non-zero precipitation.

Our main measure of performance is the Brier score (Brier 1950), defined as:

$$\text{Brier score} = \frac{1}{M}\sum_{j=1}^{M}(o_i - f_i)^2, \tag{16}$$

where M is the total number of predictions, o_i is the ith observed event (1 if the event occurred, 0 otherwise) and f_i is the forecast probability that the ith event will occur. It is negatively oriented, that is, lower is better. The Brier score can be decomposed into uncertainty, reliability and resolution components, and is equal to uncertainty plus reliability minus resolution (Murphy 1973). The uncertainty component measures the inherent uncertainty in the observations and is independent of the forecast. The reliability component measures the deviation of the reliability curve from the diagonal. It addresses calibration, that is the statistical consistency between the forecasts and the observations, and is negatively oriented. The resolution component measures the ability of the forecast to distinguish between prior situations that will lead to the occurrence or nonoccurrence of the event, and is positively oriented.

Table 2 shows that for the probability that ice forms at individual locations, both probabilistic forecasting methods substantially outperformed the raw forecast. Not surprisingly, they performed similarly to one another for individual locations. Figures 7 and 8 show reliability diagrams for both probabilistic forecasts of ice formation. At individual locations along I-90, the two methods performed similarly, as expected. However, when we looked at the forecasted probability of the spatial aggregate "*ice formation along* $\mathcal{I}$", the spatial

Table 2: Brier scores for the probability of ice formation at observation locations and along the section $\mathcal{I}$ of I-90. The decomposition of the Brier score into its three components (uncertainty, reliability, resolution) is shown in parentheses. Note that the Brier score is equal to uncertainty plus reliability minus resolution. Period: November 1, 2003–March 31, 2004 and November 1, 2004–March 31, 2005.

	Observation locations	Along $\mathcal{I}$
Raw forecast	0.179 (0.138; 0.070; 0.030)	0.280 (0.234; 0.085; 0.040)
Marginal model	0.103 (0.138; 0.043; 0.078)	0.167 (0.234; 0.103; 0.171)
Spatial model	0.115 (0.138; 0.045; 0.069)	0.163 (0.234; 0.080; 0.151)

model was clearly more reliable: the marginal model method tended to overestimate the probability of ice formation.

We now compare the economic value of the different forecasting methods. In the case of the probabilistic forecasts, we assume that anti-icing measures, costing C, are taken whenever the probability of road ice is greater than a given threshold. A loss L is incurred when no anti-icing measures are taken but ice does form on the road. Various estimates of the economic loss L associated with closure of I-90 have been reported, ranging from $\$1,000,000$ to $\$18,000,000$ per day (Paulson 2001), while the cost C of anti-icing measures has been estimated at about $\$100,000$ per day.

The threshold probability used to decide whether to take anti-icing measures is then equal to the *cost-loss ratio*, $R = C/L$. Table 3 shows contingency tables cross-classifying action (anti-icing measures or not) against outcome (road ice or not) for each of four different forecasting methods when $R = 0.1$. We have already discussed three of the four forecasts: the raw forecast and the two probabilistic forecasts. The fourth forecast we consider, the naive forecast, is the same in every instance. If the marginal probability of occurrence of road ice is above the threshold probability R, then the naive forecast always predicts road ice; otherwise it always forecasts no road ice. The marginal probability of ice formation over the 2002–2003 winter season was 0.07 (9 instances out of 131 forecast events), and so the naive forecast always predicted ice if $R < 0.07$, and never predicted ice if $R \geq 0.07$.

The contingency tables in Table 3 enable us to estimate what the loss associated with winter road maintenance decisions would have been using each of the four forecasts during the two-year period. We denote by n_{00}, n_{01}, n_{10} and n_{11} the entries in contingency tables such as those in Table 3 where, for example, n_{01} is the number of times that ice was forecast but no ice formed. Then the total loss over the period considered would have been $Ln_{10} + C(n_{01} + n_{11})$.

Figure 9 shows the economic loss associated with each of the four forecasting methods

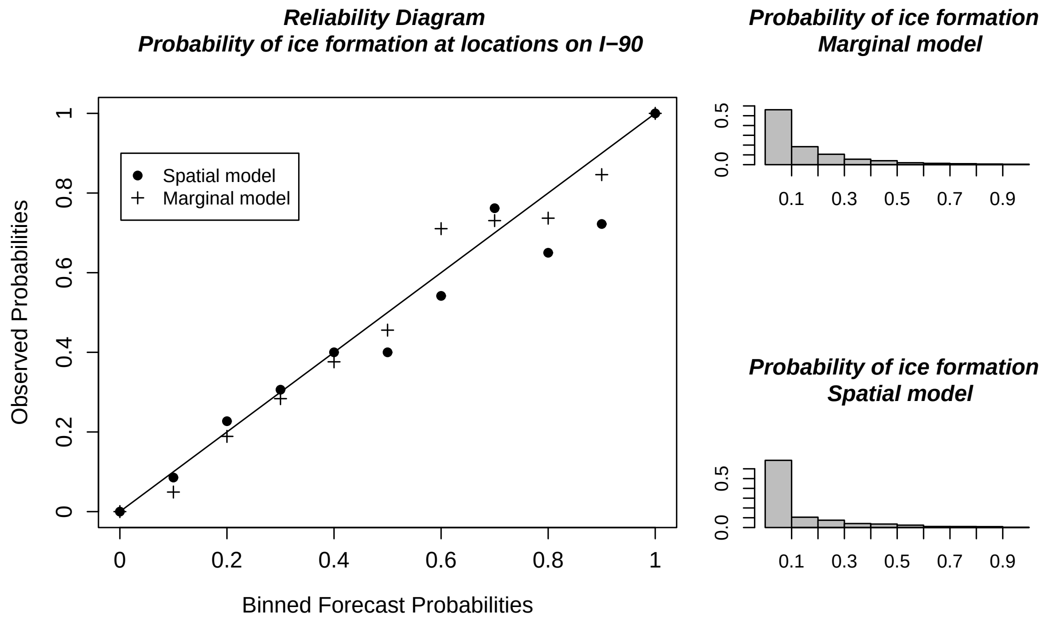


Figure 7: Reliability diagram for the probability of ice formation at observation locations as provided by the marginal model and the spatial model method, respectively, during the 2003–2004 and 2004–2005 winter seasons. Probabilities have been binned into ten bins, each of width 0.1. Histograms of the forecast probability of ice formation at observation locations from both methods are also shown.

Table 3: Forecasts and observations of ice formation for the naive forecast, the raw forecast, the marginal model and the spatial model with threshold probability $R = 0.1$. Period: November 1, 2003–March 31, 2004 and November 1, 2004–March 31, 2005. The raw forecast is deterministic, so the entries for it do not depend on R.

Forecast	Ice Forecast?	Ice Observed	Ice Not observed
Naive forecast	Yes	103	172
	No	0	0
Raw forecast	Yes	68	42
	No	35	130
Marginal model	Yes	101	111
	No	2	61
Spatial model	Yes	102	123
	No	1	49

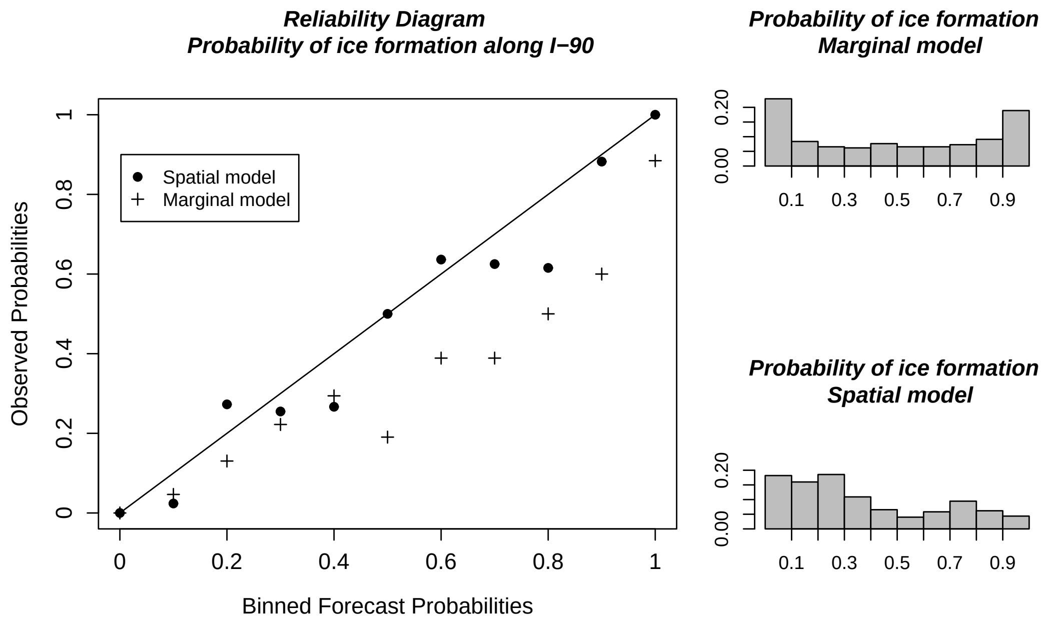


Figure 8: Reliability diagram for the probability of ice formation along the section $\mathcal{I}$ of I-90 from the spatial and marginal models during the 2003–2004 and 2004–2005 winter seasons. The probabilities have been binned into ten bins, each of width 0.1. The figure also shows histograms of the forecast probability of road maintenance along $\mathcal{I}$ provided by the two methods.

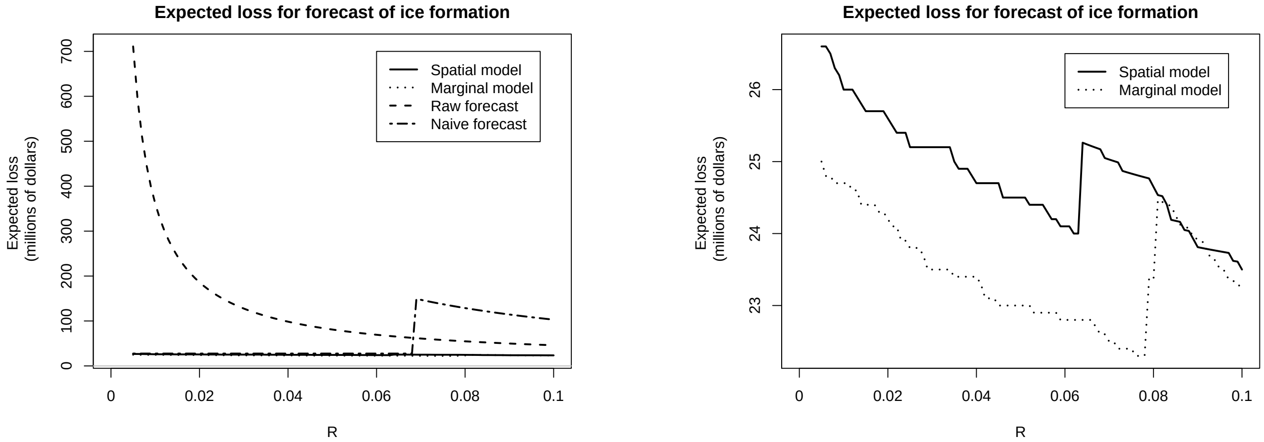


Figure 9: Economic loss associated with each of the four forecasts of ice formation plotted against the cost-loss ratio R, for the 2003–2004 and 2004–2005 winter seasons.

as a function of the cost-loss ratio R. The probabilistic forecasts led to a considerably lower economic loss than the raw forecast over all thresholds.

To get an idea of what these numbers mean in practice, consider the case where $C =$ \$100,000 and $L =$ \$1,000,000, so that $R = 0.1$, which is realistic for I-90. In that case, acting on the basis of the raw forecast would have yielded a loss of \$46M over the two-year period, while the spatial model would have yielded a loss of \$23.5M and the marginal model a loss of \$23.2M. Thus using probabilistic forecasts rather than deterministic ones would have lowered the overall economic loss by 49%, or by \$11.25M per year on just one mountain pass. Taking account of spatial dependence would have prevented one additional closure of I-90, but would have also led to anti-icing measures being taken unnecessarily 12 times more than if the marginal model method had been used.

Our test dataset consisted of 275 cases, so in our example with C = \$100,000, taking anti-icing measures every day would have cost \$27.5M. Basing decisions on the raw forecasts would have led to a loss considerably greater than this, showing the danger of relying on deterministic guidance when the costs and losses are as asymmetric as in this case. Using probabilistic forecasts of ice provided by the marginal model would have yielded a reduction in economic loss of about 15%, or \$2M per year compared with the strategy of always anti-icing.

4 Discussion

We have developed two ways of estimating the probability of ice forming on a roadway. The simpler one ignores spatial dependence, and the more complex one models spatial dependence explicitly. In our experiments, both probabilistic methods considerably outperformed the raw forecast we considered, and would have almost halved the total economic cost relative to relying on deterministic guidance. The two methods gave similar results because much of the spatial dependence in ice formation was accounted for by the numerical weather forecasts.

Our approach to the problem of predicting road ice goes beyond previous approaches by providing probabilistic forecasts rather than deterministic predictions. Commonly, road ice is forecast using mathematical models that reproduce the physical interactions between the road and the atmosphere (Best 1998; Sass 1992; Shao and Lister 1995; Crevier and Delage 2001; Chapman et al. 2001b; Korotenko 2002). Such models take into account meteorological parameters, such as air temperature, precipitation, wind direction, wind speed, humidity and dew point, and predict both road surface temperature and road conditions (Chapman et al. 2001b). However, despite the high level of detail, predictions are not always accurate (Shao 1998). For example, three numerical road models used in the UK were found to be negatively biased (Chapman et al. 2001b).

Statistical methods have been used to correct prediction errors, with a stress on choosing the weather and other variables that best predict road surface temperature (Shao and Lister 1995, 1996; Chapman et al. 2001a; Thornes et al. 2005). Unlike ours, these methods are all deterministic.

Some statistical postprocessing methods for forecasting road ice from numerical weather predictions or the outputs of road prediction models have been proposed. Shao (1998) used a back-propagation neural network to postprocess short-range forecasts of road surface temperature. Sherif and Hassan (2004) used stepwise regression to identify the more useful weather variables to predict pavement temperature. Once the meteorological parameters were selected, Sherif and Hassan (2004) generated predictions of road surface temperature by linear regression of the observed pavement temperature on the numerical forecasts of the selected meteorological variables. These methods are all deterministic in the sense that they yield point forecasts rather than predictive distributions.

Our marginal and spatial models are fairly simple, yet they provide big improvements over the raw forecast, both in terms of accuracy and economic value. The method for computing the predictive mean in our marginal model can be viewed as an instance of Model Output Statistics (MOS; Glahn and Lowry 1972; Wilks 2006). However, MOS was not previously applied to road ice forecasting, and MOS does not yield probabilistic forecasts. Our spatial model can be viewed as a generalization of the model of Gel et al. (2004) for probabilistic forecasting of temperature fields to simultaneous forecasting of temperature and precipitation fields.

In our spatial model, precipitation occurrence is modeled by a latent Gaussian random field with spatial dependence between sites. Similar models for modeling precipitation have been used by others (Bardossy and Plate 1992; Hughes and Guttorp 1994; Hutchinson 1995; Hughes et al. 1999; Guillot 1999; Sansò and Guenni 1999, 2000; Allcroft and Glasbey 2003; Sansò and Guenni 2004; Rappold et al. 2007). Our main contribution here is modeling temperature and precipitation jointly to generate forecasts of ice.

One limitation of our method is that it uses only a single deterministic forecast. Many probabilistic weather forecasts are based on an ensemble of such forecasts from different initial conditions, different numerical weather prediction models or different weather centers. Such ensembles can capture uncertainties due to the nonlinear dynamics of the atmosphere. One direction in which our approach could be expanded is the extension of the marginal and spatial models to a forecast ensemble, perhaps using a Bayesian model averaging approach (Raftery et al. 2005; Berrocal et al. 2007; Sloughter et al. 2007).

We developed our model using numerical weather forecasts of air temperature and rainfall amount, and we defined ice as the joint occurrence of air temperature below freezing and precipitation. A more accurate definition of road ice would be the joint occurrence of pavement temperature at or below freezing and precipitation. Numerical forecasts of pavement temperature and observations of road surface temperature were not available to us. However, both the marginal model and the spatial model method can be easily adapted and applied to forecasts and observations of pavement temperature rather than air temperature.

References

Albert, J. H. and Chib, S. (1993), "Bayesian analyis of binary and polychotomous response data," *Journal of the American Statistical Association*, 88, 669–679.

Allcroft, D. J. and Glasbey, C. A. (2003), "A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation," *Applied Statistics*, 52, 487–498.

American Meteorological Society (2002), "Enhancing weather information with probability forecasts," *Bulletin of the American Meteorological Society*, 83, 450–452.

Bardossy, A. and Plate, E. J. (1992), "Space-time model for daily rainfall using atmospheric circulation patterns," *Water Resources Research*, 28, 1247–1259.

Berrocal, V. J., Raftery, A. E., and Gneiting, T. (2007), "Combining spatial statistical and ensemble information in probabilistic weather forecasts," *Monthly Weather Review*, 1386–1402.

Best, M. J. (1998), “A model to predict surface temperatures,” *Boundary-Layer Meteorology*, 88, 279–306.

Brier, G. W. (1950), “Verification of forecasts expressed in terms of probability,” *Monthly Weather Review*, 78, 1–3.

Chapman, L., Thornes, J. E., and Bradley, A. V. (2001a), “Modelling of road surface temperature from a geographical parameter database. Part 1: Statistical,” *Meteorological Applications*, 8, 409–419.

— (2001b), “Modelling of road surface temperature from a geographical parameter database. Part 2: Numerical,” *Meteorological Applications*, 8, 421–436.

Chilès, J.-P. and Delfiner, P. (1999), *Geostatistics: Modeling Spatial Uncertainty*, Wiley, 695 pp.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, Wiley, revised ed., 900 pp.

Crevier, L. P. and Delage, Y. (2001), “METRo: A new model for road-condition forecasting in Canada,” *Journal of Applied Meteorology*, 40, 2026–2037.

Eckel, F. A. and Mass, C. F. (2005), “Aspects of effective mesoscale, short-range ensemble forecasting,” *Weather and Forecasting*, 20, 328–350.

Eriksson, M. and Norrman, J. (2001), “Analysis of station locations in a road weather information system,” *Meteorological Applications*, 8, 437–448.

Federal Highway Administration (2006), “Best practices for road weather management. Version 2.0,” Available online at http://ops.fhwa.dot.gov/Weather/best_practices/CaseStudies/029.pdf.

Gel, Y., Raftery, A. E., and Gneiting, T. (2004), “Calibrated probabilistic mesoscale weather field forecasting: the Geostatistical Output Perturbation (GOP) method (with discussion),” *Journal of the American Statistical Association*, 99, 575–588.

Gelfand, A. E. and Smith, A. F. M. (1990), “Sampling-based approaches to calculating marginal densities,” *Journal of the American Statistical Association*, 85, 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis*, Chapman and Hall/CRC, 2nd ed., 668 pp.

Glahn, H. R. and Lowry, D. A. (1972), “The use of model output statistics (MOS) in objective weather forecasting,” *Journal of Applied Meteorology*, 11, 1203–1211.

Gneiting, T. and Raftery, A. E. (2005), "Weather forecasting using ensemble methods," *Science*, 310, 248–249.

Grell, G. A., Dudhia, J., and Stauffer, D. R. (2004), "A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5)," NCAR Tech. Note NCAR/TN-398+STR, 121 pp. [Available from MMM Division, NCAR, P.O. Box 3000, Boulder, CO 80307.].

Grimit, E. P. and Mass, C. F. (2002), "Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest," *Weather and Forecasting*, 17, 192–205.

Guillot, G. (1999), "Approximation of Sahelian rainfall fields with meta-Gaussian random functions," *Stochastic Environmental Research and Risk Assessment*, 13, 100–112.

Handcock, M. S. and Wallis, J. R. (1994), "An approach to statistical spatial-temporal modeling of meteorological fields," *Journal of the American Statistical Association*, 89, 368–378.

Hughes, J. P. and Guttorp, P. (1994), "Incorporating spatial dependence and atmospheric data in a model of precipitation," *Journal of Applied Meteorology*, 33, 1503–1515.

Hughes, J. P., Guttorp, P., and Charles, S. P. (1999), "A non-homogeneous hidden Markov model for precipitation occurrence," *Applied Statistics*, 48, 15–30.

Hutchinson, M. F. (1995), "Stochastic space-time weather models from ground-based data," *Agricultural and Forest Meteorology*, 73, 237–264.

Korotenko, K. (2002), "An automated system for prediction of icing on the road," *Computational Science - ICSS 2002, Pt III Proceedings. Lecture Notes in Computer Science*, 2331, 1193–1200.

Mass, C., Albright, M., Ovens, D., Steed, R., MacIver, M., Grimit, E., Eckel, T., Lamb, B., Vaughan, J., Westrick, K., Storck, P., Colman, B., Hill, C., Maykut, N., Gilroy, M., Ferguson, S. A., Yetter, J., Sierchio, J. M., Bowman, C., Stender, R., Wilson, R., and Brown, W. (2003), "Regional environmental prediction over the Pacific Northwest," *Bulletin of the American Meteorological Society*, 84, 1353–1366.

Murphy, A. H. (1973), "A new vector partition of the probability score," *Journal of Applied Meteorology*, 12, 595–600.

National Research Council of the National Academies (2006), *Completing the forecast: characterizing and communicating uncertainty for better decisions using weather and climate forecasts*, The National Academies Press, 178 pp.

Norrman, J., Eriksson, M., and Lindqvist, S. (2000), "Relationships between road slipperiness, traffic accident risk and winter road maintenance activity," *Climate Research*, 15, 185–193.

Palmer, T. N. (2000), "Predicting uncertainty in forecasts of weather and climate," *Reports on Progress in Physics*, 63, 71–116.

Paulson, T. (2001), "In avalanche country, thinnest of defenses hangs tough," Appeared in Seattle Post-Intelligencer on Thursday December 21, 2001. Available at http://seattlepi.nwsource.com/local/52258_avalanche27.shtml.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005), "Using Bayesian Model Averaging to Calibrate Forecast Ensembles," *Monthly Weather Review*, 133, 1155–1174.

Ramakrishna, D. M. and Viraraghavan, T. (2005), "Environmental impact of chemical deicers – a review," *Water, Air, and Soil Pollution*, 166, 49–63.

Rappold, A., Gelfand, A. E., and Holland, D. M. (2007), "Understanding mercury deposition through a latent space-time process," Submitted.

Richardson, D. S. (2000), "Skill and relative economic value of the ECMWF ensemble prediction system," *Quarterly Journal of the Royal Meteorological Society*, 126, 649–667.

Roulston, M. S., Bolton, G. E., Kleit, A. N., and Sears-Collins, A. L. (2006), "A laboratory study of the benefits of including uncertainty information in weather forecasts," *Weather and Forecasting*, 21, 116–122.

Sansò, B. and Guenni, L. (1999), "Venezuelan rainfall data analysed by using a Bayesian space-time model," *Applied Statistics*, 48, 345–362.

— (2000), "A nonstationary multisite model for rainfall," *Journal of the American Statistical Association*, 95, 1089–1100.

— (2004), "A Bayesian approach to compare observed rainfall data to deterministic simulations," *Environmetrics*, 15, 597–612.

Sass, B. H. (1992), "A numerical model for prediction of road temperature and ice," *Journal of Applied Meteorology*, 31, 1499–1506.

Shao, J. (1998), "Improving nowcasts of road surface temperature by a backpropagation neural network," *Weather and Forecasting*, 13, 164–171.

Shao, J. and Lister, P. J. (1995), "The prediction of road surface-state and simulations of the shading effect," *Boundary-Layer Meteorology*, 73, 411–419.

— (1996), "An automated nowcasting model of road surface temperature and state for winter road maintenance," *Journal of Applied Meteorology*, 35, 1352–1361.

Sherif, A. and Hassan, Y. (2004), "Modelling pavement temperature for winter maintenance operations," *Canadian Journal of Civil Engineering*, 31, 369–378.

Sloughter, J. M., Raftery, A. E., Gneiting, T., and Fraley, C. (2007), "Probabilistic quantitative precipitation forecasting using Bayesian model averaging," *Monthly Weather Review*, in press.

Thornes, J. E., Cavan, G., and Chapman, L. (2005), "XRWIS: the use of geomatics to predict winter road surface temperatures in Poland," *Meteorological Applications*, 12, 83–90.

Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, Elsevier Academic Press, 2nd ed., 627 pp.