BAYESIAN HIERARCHICAL MODELS FOR ENVIRONMENTAL DATASETS

By

Jason Andrew Matney

A THESIS

Submitted
to Michigan State University
in partial fulfillment of the requirements
for the degree of

Geography—Master of Science

2014

ABSTRACT

BAYESIAN HIERARCHICAL MODELS FOR ENVIRONMENTAL DATASETS

By

Jason Andrew Matney

This thesis explores the applicability of Bayesian spatial models for predicting the occurrence of permafrost across Alaska, USA. Exploratory analysis of a large Alaska soil carbon database suggests the impact of some important environmental covariates on permafrost occurrence is non-linear. Also, exploratory analysis using non-spatial regression models shows that substantial spatial autocorrelation among residuals exists even after accounting for available covariates. Spatial regression models specifically designed to accommodate non-linearity between covariates and probability of permafrost are developed and tested. Results show the proposed models provide improved fit and predictive ability over conventional modeling techniques. Considerations for applying the proposed models to large spatial domains for creating high-resolution map permafrost products are also discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

GENERALIZED ADDITIVE PROCESS MODELS FOR ENVIRONMENTAL CONTROLS OF NEAR-SURFACE PERMAFROST

1.1 Introduction

Recently an influx of high-dimensional spatial data has shifted resources within the scientific community. Coping with the outgrowth of this development involves an understanding of large, dynamic data sets, and the ability to select for signal amongst noise. The origins of these data include NASA and other sources, and involve orders of magnitude growth in processing power as spatial ranges increase [8]. The technique employed to accomplish the analysis below will be a Bayesian hierarchical model, which, when leveraged correctly, will produce an adequate estimate of the presence or absence of permafrost across the entire span of Alaska, thus relying on large-scale processing techniques comprised of novel geostatistical developments [8].

Effects of climate change have been shown to be accentuated in the Arctic [24]. Analyzing the extent of permafrost then yields a signal for assessing the reach and severity of climate change. Presence and persistence of permafrost is a function of various climate variables, primarily mean annual temperature, and local scale substrate and topographic characteristics. Permafrost distribution is therefore strongly controlled by soil drainage patterns, but the presence of permafrost also limits local drainage, affecting organic matter decomposition, and nutrient release to plants, see, e.g., [19],[20]. Probability of permafrost can also serve as a proxy for determining mean temperature regime shifts in the Alaskan discontinuous zone. Moreover, permafrost thawing can potentially exacerbate global climate change by releasing methane and other hydrocarbons [24]. These features of arctic ecosystem shifts underscore the necessity of improving the accuracy of permafrost estimates.

In addition to topography and drainage, permafrost formation is determined by interactions be-

tween air temperature and organic layer thickness, see, e.g., [17], [25]. For example, over climatic gradients, permafrost probability decreases when mean annual temperature (MAT) increases, but this interaction is mostly limited to loamy and silty soils, whereas coarse textured soils maintain little permafrost regardless of MAT [17]. Other factors controlling the persistence of frozen soil include snow depth, vegetation, slope aspect, and fire disturbance, see, e.g., [18],[22]. Given the complexity of permafrost formation, models that predict permafrost extent are most useful when they leverage environmental variables that represent soil forming factors and incorporate robust estimates of uncertainty.

Taking these parameters into account, the following discussion will define and assess the predictive ability of a non-spatial, spatially-varying intercept, and spatially-varying coefficient Bayesian regression model [7]. Spatial random effects are specified to accommodate dependence of model residuals and the non-linear impact of environmental covariates. The independent variables used in these regression models are mean annual temperature, slope, heatload, and composite topographic index (CTI). Heatload is a standardized index of incoming solar radiation while CTI is a catchment wetness index [17]. These covariates were selected for their regular profile, continuity of deployment, and explanatory prowess.

The central addition in this analysis from past implementations of regression models assessing permafrost range lies in the ability to explicitly accommodate residual spatial dependence while also considering the space-varying impact of the covariates [9]. Ignoring spatial dependence and non-stationarity of covariates can lead to incorrect inference about model parameters and erroneous predictions [9]. Adding a spatially-varying model intercept via normally distributed spatial random effects can resolve this source of error by capturing spatial association. Moreover, hierarchical models are capable of acknowledging uncertainty and dependence across several sources, allowing for the estimation of richer models.

These contributions serve to inform the following research objectives. We wish to broadly improve the prediction accuracy of statistical models commonly used in geostatistical research while

also increasing the precision of our predictions. We accomplish this by working in a regression context to assess probability of permafrost and showing that the non-linear impact of covariates is accommodated via a gaussian process. This is a novel application of these methods and is akin to approaches taken in the Generalized Additive Model (GAM) literature, see, e.g., [26]. Flexible, function-based parameters will be shown to allow slope coefficients to change based on the value of the covariate. We also aim to update on the current methodological approaches by demonstrating how model fit and prediction improves upon capturing the residual dependent pattern. We plan to produce value-added applications for the spatial modeling community by composing high-resolution data products with associated uncertainty. These products will help us answer our core research question; which variables explain the variability in permafrost and how certain are we of those explanatory relationships?

The format of the manuscript is as follows. Section 1.2 provides an overview of the motivating data and exploratory analysis. Section 1.3 details the proposed model setup, Subsection 1.3.1 discusses implementation details including prediction and interpolation, whereas our model assessment approach is defined in Subsection 1.4.2. The permafrost analysis is offered in Section 1.4. Finally, Section 1.6 concludes the manuscript with an eye toward future work.

1.2 Permafrost data

We considered permafrost presence and absence data for 4646 sample locations across the U.S. state of Alaska (Figure A.1). These data were drawn from the Alaska soil carbon database described in Johnson et al. [16], and the National Resources Conservation Service (NRCS) pedon dataset for Alaska (NASIS 2010). The sample locations often follow natural features, such as a ridge line or river bed, mirroring the typical terrain found along the Alaskan discontinuous zone. These data identify presence/absence of permafrost within the first meter from the surface. Data were collected between 1998 and 2008 and only the most recent measurement was considered for

each location.

As described in Section 1.1, both broad scale climate and local scale soil variables contribute to permafrost formation. In the subsequent analyses, covariates include mean annual temperature (MAT), percent slope (SLOPE), a heat load index (HLI), and a soil wetness index called compound topographic index (CTI). MAT, shown in Figure A.3(b), was extracted from the 2 km Parameter-Elevation Regression on Independent Slopes Model for the years 1961-1990 (PRISM) [6]. SLOPE, HLI, and CTI, were calculated using a 30 m spatial resolution digital elevation model (DEM) [12]. Following Balice et al. [1], HLI was calculated as $\cos(\text{aspect}\ \pi/180))$ where higher values indicate more northerly slopes. CTI serves as a proxy for soil wetness and is a function of both the slope and contributing area per width orthogonal to the flow direction [13]. Previous research suggests that CTI is useful for mapping permafrost in interior Alaska [21]. Figures A.3(a), figHLI, and figCTI show maps of SLOPE, HLI, and CTI, respectively. The 2 km SLOPE variable was resampled using a cubic convolution algorithm to match the 30 m spatial resolution of the DEM derived variables [17].

1.2.1 Exploratory data analysis

Figure A.2 provides a summary of the empirical probability of permafrost occurrence and each covariate. Each histogram bar contains 100 observations; hence, greater bar width reflects greater sparsity of observations across an interval of the given covariate. The height of each bar is the average over the observed zeros and ones within the given interval. These figures are useful for exploring how covariate values might influence permafrost formation. Figure A.2(a) suggests the probability of permafrost generally increases with increasing slope from 0 to $\sim$5%. Beyond $\sim$5% the probability of permafrost decreases with increasing slope. This change from a positive to negative relationship was also identified by Johnson et al. [17]. For the extremes in observed MAT, Figure A.2(b) shows the expected trend between temperature and permafrost—lower mean annual temperature results in higher probability of permafrost. However, for intermediate temperatures,

the histogram suggests some counter-intuitive trends. For example, warming from ∼-5 to -3°C results in a moderate increase in permafrost occurrence, which is then followed by the expected approximately exponential decrease in permafrost beyond -3°C. Figure A.2(c) suggests there is very little relationship between HLI and permafrost, except, perhaps, on the most northerly facing sites where there is a slight increase in the probability of occurrence. Similarly, Figure A.2(d) shows little evidence of differentiation of permafrost occurrence across CTI values.

The exploratory analysis suggests that SLOPE, MAT, and perhaps HLI might impact the permafrost formation in a nonlinear way over their range of values. The strongest signal for nonlinearity is seen in percent slope and intermediate values of MAT. These observations encourage further exploration using a model framework that is able to estimate these potential patterns.

1.3 Generalized additive process models

As detailed in Section 1.2, the presence or absences of permafrost in the first meter of the surface along with a set of covariates were recorded at a set of locations $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ within the study area, where n=4646. At generic location $\mathbf{s}$, we assume the response variable $y(\mathbf{s})$, where $y(\mathbf{s}) = 1$ when permafrost is observed and 0 otherwise, follows a Bernoulli distribution and consider the model:

$$\pi(y(\mathbf{s}) \,|\, \eta(\mathbf{s})) \sim Ber(p(\eta(\mathbf{s}))), \tag{1.1}$$

where $p(\eta(\mathbf{s})) =$

$\eta(\mathbf{s}))/(1+\eta(\mathbf{s})))$ is the probability of permafrost at $\mathbf{s}$ and

$$\eta(\mathbf{s}) = \beta_0 + \sum_{k=1}^{p} g_k(x_k(\mathbf{s}); \theta_k) + w(\mathbf{s}), \tag{1.2}$$

where β_0 is the intercept and $g_k(\cdot)$ is some function that involves the kth covariate $x_k(\mathbf{s})$ and associated vector of parameters θ_k, and $w(\mathbf{s})$ provides spatial adjustment (with structured dependence) to the intercept. Here, we assume $w(\mathbf{s})$ follows a Gaussian process with mean zero and

covariance function $C_w(\cdot;\theta_w)$. For any two locations, say $\mathbf{s}_i$ and $\mathbf{s}_j$, we specify $C_w(\mathbf{s}_i,\mathbf{s}_j;\theta_w) = \sigma_w^2\rho(\mathbf{s}_i,\mathbf{s}_j;\phi_w)$, where $\theta_w = \{\sigma_w^2,\phi_w\}$ and $\rho(\cdot;\phi_w)$ is a *correlation function* with ϕ_w controlling the correlation decay and σ_w^2 representing the spatial variance. The function $\rho(\cdot;\phi_w)$ could be any *valid* spatial correlation function, see, e.g., Cressie [5], Chilés and Delfiner [4] or Finley et al. [8]. Here, we use an exponential function such that $C_w(\mathbf{s}_i,\mathbf{s}_j;\theta_w) = \sigma_w^2$

-$\phi_w||\mathbf{s}_i-\mathbf{s}_j||$), where $||\mathbf{s}_i-\mathbf{s}_j||$ is the Euclidean distance between locations $\mathbf{s}_i$ and $\mathbf{s}_i$.

If we assume a constant relationship between the probability of permafrost and the kth covariate then $g_k(x_k;\theta_k) = x_k\beta_k$ resulting in common linear and additive intensity $\sum_{k=1}^{p} x_k(\mathbf{s})\beta_k$. However, the exploratory data analysis results presented in Section 1.2.1 suggest that several of the covariates might have non-constant, or non-linear, impact on the probability of permafrost. For example, the probability of finding permafrost might increase with slope in the interval between $\sim$0.4-4% but then decrease with increasing slope beyond $\sim$4%, Figure A.2(a). Exploratory results also suggested the occurrence of permafrost might increase at a non-constant rate as mean annual temperature changes. Given these observations and findings from other studies noted in Section 1.1, we wish to explore a more flexible specification of $g_k(\cdot)$ that is able to accommodate non-linear impacts of the covariate on the probability of permafrost. Examples of such specifications abound in the Generalized Additive Model (GAM) literature, where penalized regression splines are used to approximate these unknown functions, see, e.g., Hastie and Tibshirani [15] and Wood [26].

In our setting a Bayesian modeling framework is attractive given our desire to include hierarchical spatial random effects and interest in predictive inference. Here we take an approach similar to a GAM specification, but rather than using a penalized regression spline, the unknown function $g_k(\cdot)$ is approximated using a *predictive process* over the covariate values, see, e.g., [2]. Specifically, for a given covariate observed at n locations we consider a set of m_k knots $x_k^* = \{x_{k,1}^*, x_{k,2}^*, \ldots, x_{k,m}^*\}$ that are distributed over the covariate's range and $m << n$ (in the subsequent development we suppress the covariate subscript index k to reduce clutter). We assume a Gaussian process over the knots $g^* \sim MVN(0, C^*(\theta))$, where g^* is a $m \times 1$ vector and $C^*(\theta)$ is a $m \times m$ correlation

matrix with i, jth element $C(x_i^*, x_j^*; \theta) = \rho(x_i^*, x_j^*; \theta)$. Here, again, we consider an exponential correlation function such that $\rho(x_i^*, x_j^*; \theta) = -\phi||x_i^* - x_j^*||)$ where $\theta = \{\sigma^2, \phi\}$. Analogous to the 2-dimensional spatial predictive process defined by Banerjee et al. [2], interpolation to any covariate value, say x_0, is given by $g(x_0; \theta) = c^\top(x_0; \theta)C^*(\theta)^{-1}g^*$, where $c(x_0; \theta)$ is the $m \times 1$ vector of correlations between the value to be interpolated and the knots, i.e., $C(x_0, x_j^*; \theta)$ for $j = 1, 2, \ldots, m$.

1.3.1 Implementation

To complete the Bayesian model specification, we assign prior distributions to the model parameters and inference proceeds by sampling from the posterior distribution of the parameters. For sub-models 1 and 2, we assumed β follows a $MVN(\mu_\beta, \Sigma_\beta)$ prior with mean vector $\mu_\beta = 0$ and diagonal covariance matrix Σ_β with elements equal to 10000. The variance parameters σ^2's for all models were assigned inverse-Gamma, $IG(a, b)$, priors. With a shape hyperparameter of $a = 2$ the IG prior distribution mean is equal to the scale b and has an infinite variance (see, e.g., IG definition in Gelman et al. [10]. Exploratory data analysis was useful for setting process specific b values. The process correlation decay parameters ϕ's followed a Uniform, $Unif(a, b)$, with support over the covariate's range. The support for the spatial decay parameter ϕ_w in sub-models 2 and 4 had support between 10 and 2500 km.

Using notations similar to Gelman et al. [10], the posterior distribution of the model parameters $p(\Omega \,|\, y)$, where $y = \{y(\mathbf{s}_1), y(\mathbf{s}_2), \ldots, y(\mathbf{s}_n)\}^\top$ and Ω is all the model parameters, is proportional to

$$\begin{aligned} \prod_{k=1}^{p} p(\theta_k) \times p(\theta_w) \times N(\beta_0 \,|\, \mu_0, \sigma_0^2) \\ \times \prod_{k=1}^{p} N(g_k^* \,|\, 0, C_k^*(\theta_k)) \times N(w \,|\, 0, C_w(\theta_w)) \\ \times \prod_{i=1}^{n} P(y(\mathbf{s}_i) = 1 \,|\, x(\mathbf{s}_i), \Omega)^{y(\mathbf{s}_i)} (1 - P(y(\mathbf{s}_i) = 1 \,|\, x(\mathbf{s}_i), \Omega))^{1 - y(\mathbf{s}_i)}, \end{aligned} \tag{1.3}$$

where $P(y(\mathbf{s}_i) = 1) = p(\eta(\mathbf{s}_i))$.

An adaptive Metropolis-within-Gibbs Markov chain Monte Carlo (MCMC) algorithm was used to draw posterior samples from Ω, see, e.g, [23]. MCMC samplers were implemented in C++ and Fortran and leveraged Intel's Math Kernel Library threaded BLAS and LAPACK routines for matrix computations. All analyses were conducted on a Linux workstation using two Intel Nehalem quad-Xeon processors.

1.4 Permafrost data analysis

1.4.1 Candidate models

Following the exploratory analyses detailed in Section 1.2.1 and the specification of $w(\mathbf{s})$ and $g(\cdot)$ we consider the following candidate specifications for $\eta(\mathbf{s})$ in (1.1):

$$\begin{aligned}
&\text{Sub-model 1: } \beta_0 + \sum_{k=1}^{p} x_k(\mathbf{s})\beta_k, \\
&\text{Sub-model 2: } \beta_0 + \sum_{k=1}^{p} x_k(\mathbf{s})\beta_k + w(\mathbf{s}) \\
&\text{Sub-model 3: } \beta_0 + \sum_{k=1}^{p} g_k(x_k(\mathbf{s});\theta_k), \\
&\text{Sub-model 4: } \beta_0 + \sum_{k=1}^{p} g_k(x_k(\mathbf{s});\theta_k) + w(\mathbf{s})
\end{aligned}$$

1.4.2 Model fit and predictive performance

Sub-models of (1.1) were assessed using both a formal model fit criterion and predictive qualities. We used the popular *Deviance Information Criterion* (DIC), to rank models in terms of how well they fit the data, see, e.g., [3]. This criterion is the sum of the Bayesian deviance (a measure of model fit) and the effective number of parameters p_D (a penalty for model complexity). Here, lower DIC indicates *better* fit. In addition to DIC, models' predictive performance was assessed

using a split-set validation approach. Here, 75% of the observations were used to estimate candidate models' parameters and the remaining 25% were used for subsequent prediction. Following Gneiting and Raftery [14], models were ranked based on scoring rules calculated using the holdout set observed and predicted values. A scoring rule provides a summary measure for evaluating a probabilistic prediction given the predictive distribution and the observed response. In our setting the scoring rule function is $S(r,i)$, where $r = (p(\eta(\mathbf{s})), 1-p(\eta(\mathbf{s})))$ and i is the index of the observed response, i.e., either 0 or 1. Given the posterior predictive distribution mean and observed response at the $n_0 = 1161$ holdout locations, we calculated a mean score $\hat{S} = \sum_{j=1}^{n_0} S(r_j, i_j)/n_0$. We considered the following scoring rules:

$$\text{Zero-one: } S(r,i) = \begin{cases} 1, & \text{if } r_i = \max\{r_1, r_2\} \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Quadratic: } S(r,i) = 2r_i - \sum_{j=1}^{2} r_j^2 - 1$$

$$\text{Spherical: } S(r,i) = \frac{r_i}{\left(\sum_{j=1}^{2} r_j^2\right)^{1/2}}$$

$$\text{Logarithmic: } S(r,i) = \log r_i.$$

For these rules, larger scores indicate superior predictive performance. That is, for the zero-one and spherical rules, scores closer to 1 indicate greater predictive performance, whereas for the quadratic and logarithmic rules, scores closer to 0 suggest improved performance. We also note, that with the exception of the zero-one, all of these scoring rules are *strictly proper*—a quality desirable in assessment of probabilistic prediction, see, e.g., [14] for more details.

Three MCMC chains were run for 25000 iterations each. The most demanding model required $\sim$1 hour to complete a single MCMC chain. Convergence was diagnosed using the `CODA` package in `R` by monitoring mixing of chains and the Gelman-Rubin statistic [11]. Satisfactory convergence was diagnosed within 10000 iterations for all models. Posterior inference was based on a post

burn-in sub-sample of 15000 iterations (5000 from each chain).

1.5 Results and discussion

Candidate model parameter estimates and fit criteria are provided in Table A.1. For sub-model 1, posterior summaries suggest that SLOPE, MAT, and CTI are useful for explaining variability in permafrost occurrence, i.e., regression coefficient 95% credible interval (CI) exclude zero. The addition of spatial random effects to sub-model 1—reducing spatial correlation among the model residuals—causes CTI in sub-model 2 to become not significant. The addition of the spatial random effects also improves model fit as indicated by the lower DIC for sub-model 2 compared to that of sub-model 1.

Sub-models 3 and 4 relax the assumption of a constant change in the log odds of permafrost occurrence for a unit increase in the corresponding covariate, holding the other covariates constant at a given value. The DIC values for the non-spatial sub-models 1 and 3, suggests the Gaussian additive process model's increased flexibility results in improved fit to the observed data. Similar improvements to data fit are seen when moving from sub-model 2 to 4.

Covariate specific process parameters σ^2's and ϕ's as well as the functional form of the covariates' impact on permafrost occurrence did not differ substantially between sub-models 3 and 4. The regression coefficient posterior summaries for sub-model 4 are shown in Figure A.7. Here, the median and associated 95% CI band for the SLOPE coefficient, Figure A.7(a), are consistent with patterns observed in the univariate exploratory analysis Figure A.2(a). Specifically, slight increases in slope from flat areas to $\sim$3% increase the probability of permafrost formation. This positive relationship continues, but at a decreasing rate from $\sim$3 to 10% slope. Moderate slopes of $\sim$12% have a negative impact on formation of permafrost. Beyond $\sim$12%, slope explains little variability in permafrost occurrence. Based on univariate analyses similar to those presented in Section 1.2.1, Johnson et al. [17] attribute the non-linear impact of slope on permafrost forma-

tion to changes in soil texture, i.e., percent silt, clay, and sand, at various topographic positions. Soil texture at a given slope determines water holding capacity and subsurface flow, both of which influence the formation and persistence of permafrost.

The posterior summary for the MAT coefficient is given in Figure A.7(b). As expected, there is a negative relationship between temperature and permafrost occurrence. This trend is clear when looking at the temperature extremes where the CI band does not include zero. Figures A.1 and A.3(b) show that permafrost formation in areas with MAT from $\sim$-8 to $0^{\circ}C$ is highly variable. This variability translates into a coefficient value indistinguishable from zero within the corresponding range within Figure A.7(b). This region of moderate below zero mean annual temperature covers the majority of the mid-latitude region of the study area.

Figures A.7(c) and A.7(d) confirm patterns seen in the univariate exploratory data analysis that suggested HLI and CTI are not very useful for discerning between the presence or absence of permafrost. The exception is some marginal support for increased permafrost occurrence on only the most northerly locations, as indicated in Figure A.7(c) by a positive coefficient at the largest HLI values.

DIC and parameter estimates suggest the flexibility of the proposed additive Gaussian process models do improve fit to the observed data. However, the improvement in fit could be the result of overfitting the observed data and hence result in reduced predictive performance. We assess the degree of overfitting using the split-set validation described in Section 1.4.2. The resulting holdout set prediction scores are given in Table A.2 for each candidate model. Here, larger scores indicate superior predictive performance. Results are consistent with model DIC values and suggest the additional complexity of sub-models 2 and 4 is warranted. Further, the addition of the spatial random effects improve model prediction. Although less appealing from a theoretical perspective, see, e.g., [14], the zero-one scoring rule can be interpreted simply as the prediction success rate. Moving from the sub-model 1 to 4 increases prediction accuracy from 58 to 75%.

Two 200 x 200 pixels clips at a 30 x 30 m resolution are included as examples of the mapped

permafrost probabilities. The locations of the clips span ecologically diverse regions of Alaska, and can be identified in Figure A.4. Clips 1, taken from the southwestern portion of the state and 6, taken from the Alaskan north, were used in the mapping. Both sub-Model 3 and sub-Model 4, representing nonspatial and spatial models, respectively, were then employed to produce the associated plots and compared based on their variance and predictive ability. The results demonstrate how the spatial random effect applies a local adjustment to the predictions, reducing the probability of permafrost in relation to the nonspatial model across both sites. The comparison of the variance maps demonstrate greater variance with the addition of spatial random effects. However, the increased variability is not sufficient enough to decline to the select the spatial model as superior in both fit and prediction.

1.6 Summary

We proposed a set of generalized additive process models to accommodate non-constant relationships between covariates and the probability of permafrost. The functional forms of these relationships were estimated using low-rank Gaussian predictive process models. The proposed modeling framework also accommodates the addition of spatial random effects to reduce spatial correlation among model residuals. Compared with a conventional logistic regression model, the increased flexibility afforded by the functional regression coefficients and spatial random effects improved both model fit to observed data and prediction at unobserved locations. Further, inference about covariates explanatory power at different values provides insights into environmental controls on permafrost formation and, perhaps, persistence under anticipated climate change. For example, one might explore how different temperature warming scenarios influence the extent and location of permafrost.

Future work will focus on developing statewide permafrost probability maps with associated uncertainty. Executing the prediction algorithm detailed in Section 1.3 is tractable for a few thou-

sand locations. However, applying this algorithm to millions of locations, e.g., over the extent of Alaska, is computationally onerous due to the required cubic order matrix operations. We are currently exploring both model-based and computing solutions to this challenge. Additionally we will consider how variable interactions could be accommodated in this model framework.

APPENDIX

	Sub-model			
Parameter	1	2	3	4
β_0	-0.97 (-1.21, -0.77)	-2.04 (-2.96, -1.47)	-0.23 (-1.57, 0.92)	-1.00 (-4.05, 0.85)
β_{SLOPE}	-0.04 (-0.06, -0.02)	-0.072 (-0.099, -0.049)	–	–
β_{MAT}	-0.20 (-0.24, -0.17)	-0.37 (-0.51, -0.29)	–	–
β_{HLI}	0.0003 (-0.0012, 0.0021)	0.0003 (-0.0020, 0.0023)	–	–
β_{CTI}	-0.0025 (-0.0050, -0.0001)	-0.0008 (-0.0039, 0.0023)	–	–
σ^2_{SLOPE}	–	–	0.41 (0.18, 1.01)	0.47 (0.19, 1.23)
σ^2_{MAT}	–	–	3.62 (1.53, 9.40)	4.79 (1.76, 14.97)
σ^2_{HLI}	–	–	0.21 (0.11, 0.47)	0.25 (0.12, 0.63)
σ^2_{CTI}	–	–	0.91 (0.33, 2.79)	0.55 (0.20, 1.81)
σ^2_w	–	3.84 (2.54, 5.79)	–	4.88 (3.51, 8.22)
ϕ_{SLOPE}	–	–	0.194 (0.086, 0.312)	0.170 (0.087, 0.309)
ϕ_{MAT}	–	–	0.196 (0.152, 0.409)	0.199 (0.151, 0.418)
ϕ_{HLI}	–	–	0.032 (0.020, 0.058)	0.034 (0.020, 0.059)
ϕ_{CTI}	–	–	0.082 (0.033, 0.117)	0.075 (0.033, 0.116)
ϕ_w	–	0.022 (0.016, 0.024)	–	0.023 (0.020, 0.024)
pD	5.08	119.03	35.67	257.14
DIC	4497.70	3546.97	4191.64	3230.50

Table A.1: Candidate model parameter posterior distribution percentiles 50 (2.5, 97.5) and model fit criterion. Sub-models 3 and 4 fit using $m = 50$ equally spaced knots.

Sub-model	Zero-one	Quadratic	Spherical	Logarithmic
1	0.58	-0.46	0.73	-0.65
2	0.74	-0.35	0.80	-0.52
3	0.67	-0.41	0.76	-0.59
4	0.75	-0.33	0.81	-0.50

Table A.2: Candidate model predictive performance using scoring rules summaries, $\hat{S}$, calculated using posterior predictive distribution mean and observed values at 1072 holdout locations. Sub-models 3 and 4 fit using $m = 50$ equally spaced knots.

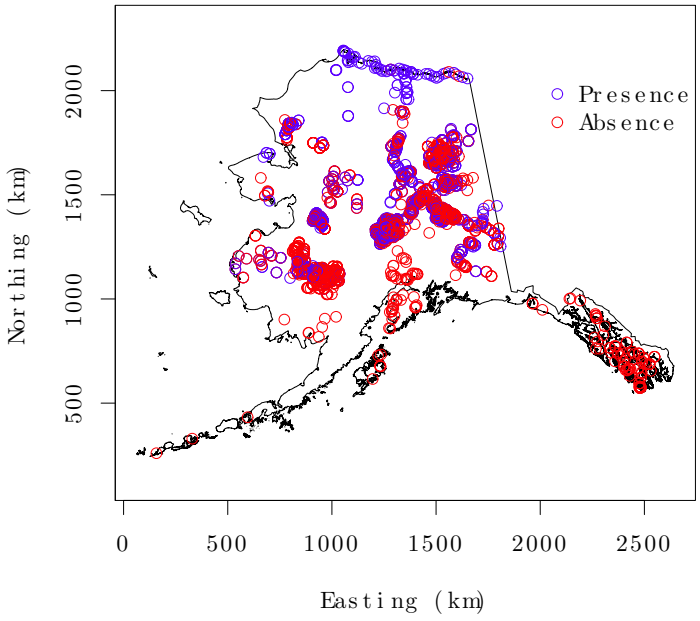
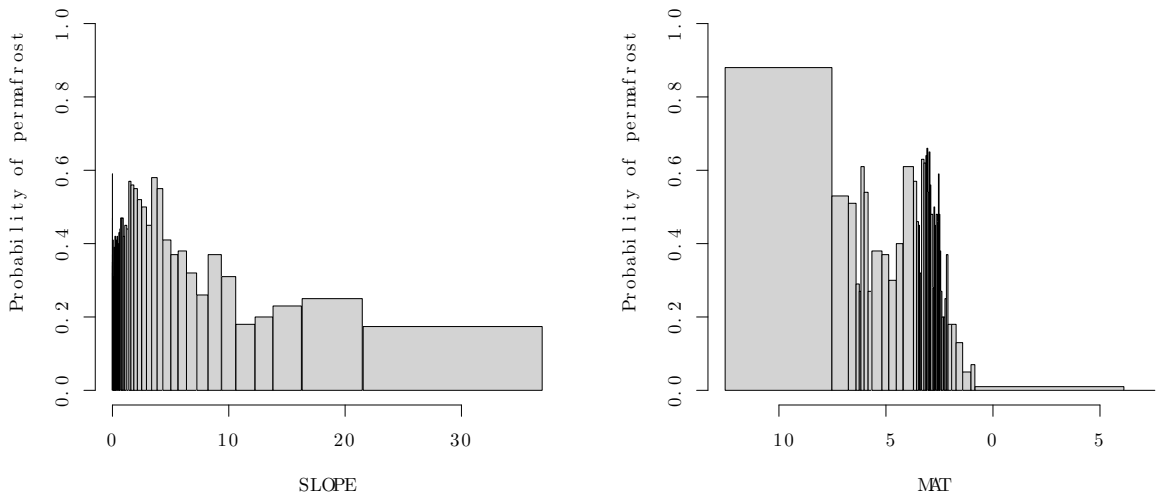
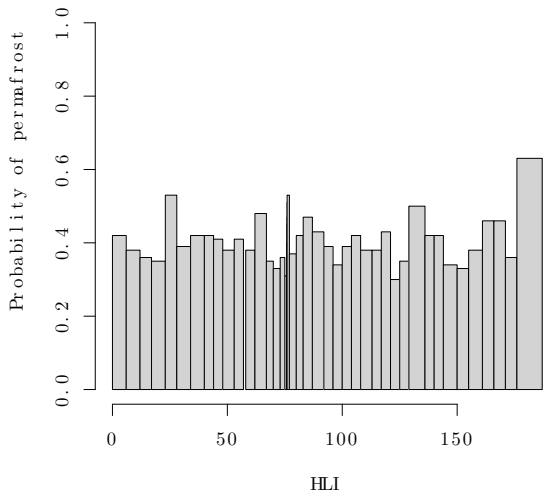


Figure A.1: Locations where permafrost presence and absence was recorded across the U.S. state of Alaska.

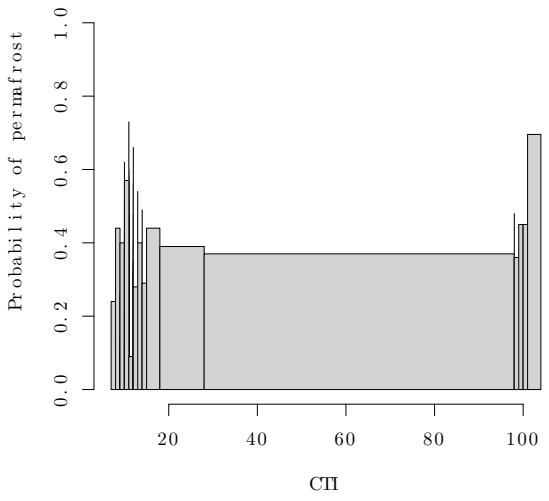


(a)

(b)

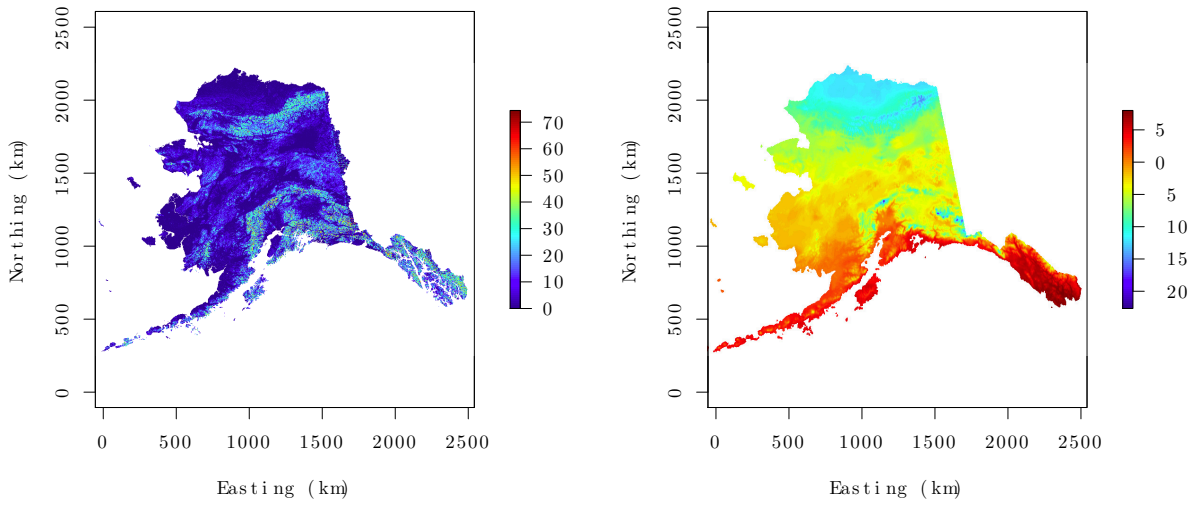


(c)



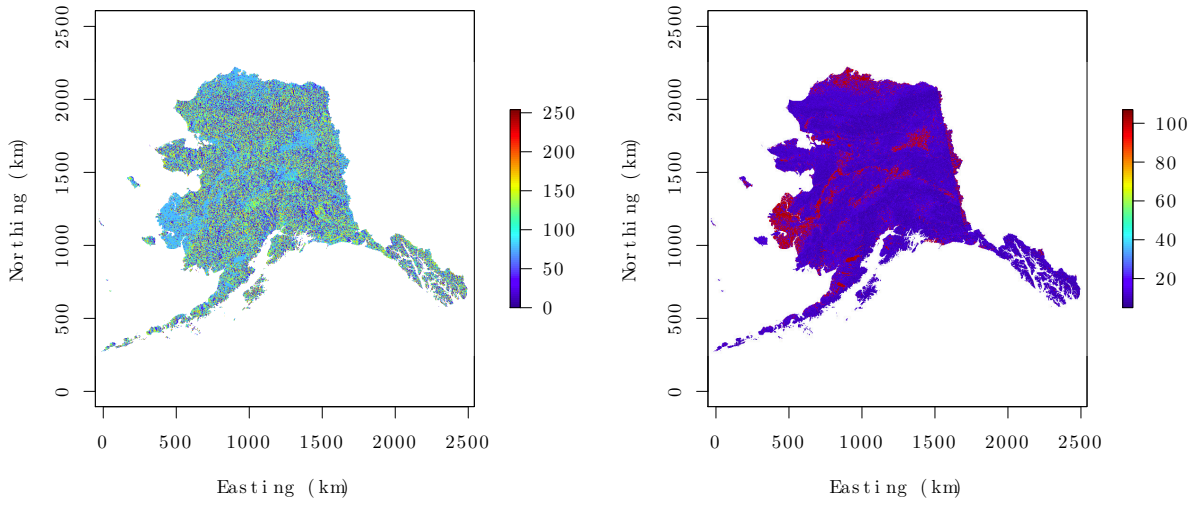
(d)

Figure A.2: Exploratory data analysis of the relationship between predictor variables and observed probability of permafrost. The height of each histogram bar was calculated using 100 observations. Wider bars represent sparser distributions of data.



(a) SLOPE %

(b) MAT



(c) HLI

(d) CTI

Figure A.3: Covariates used to explain permafrost occurrence.

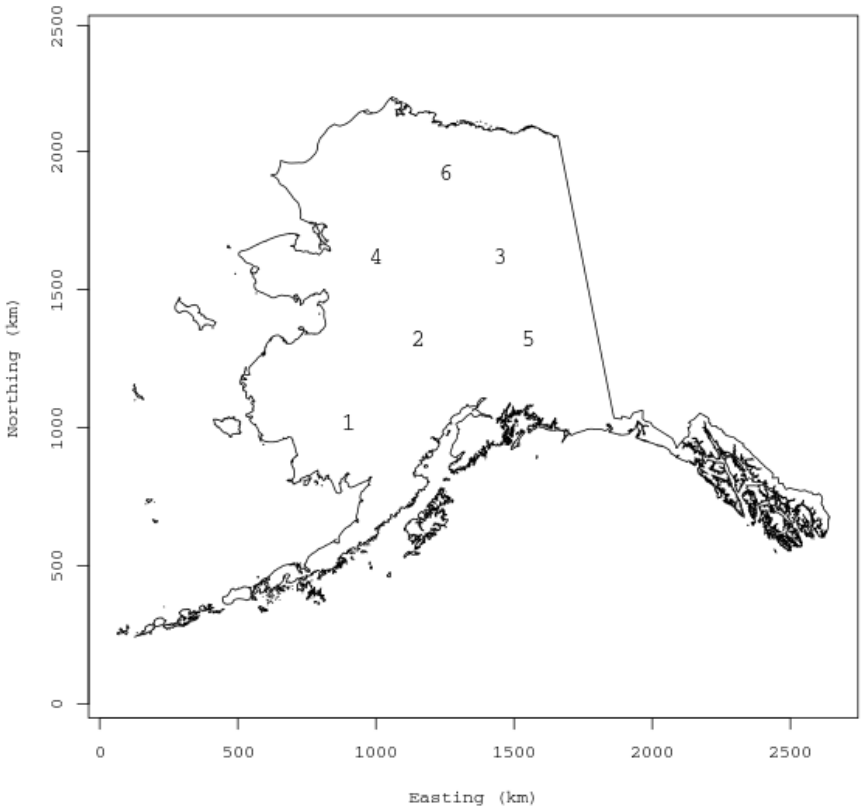
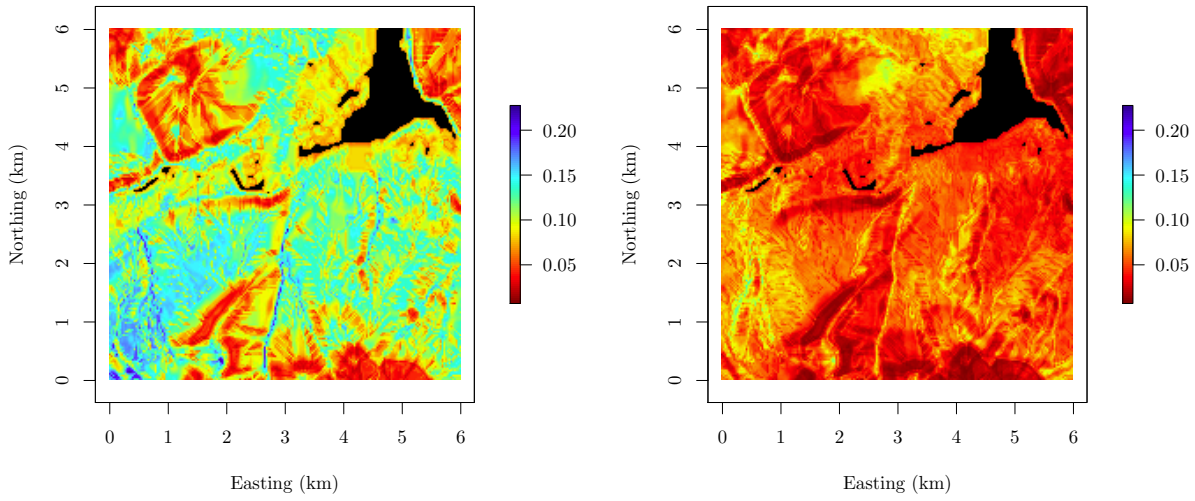
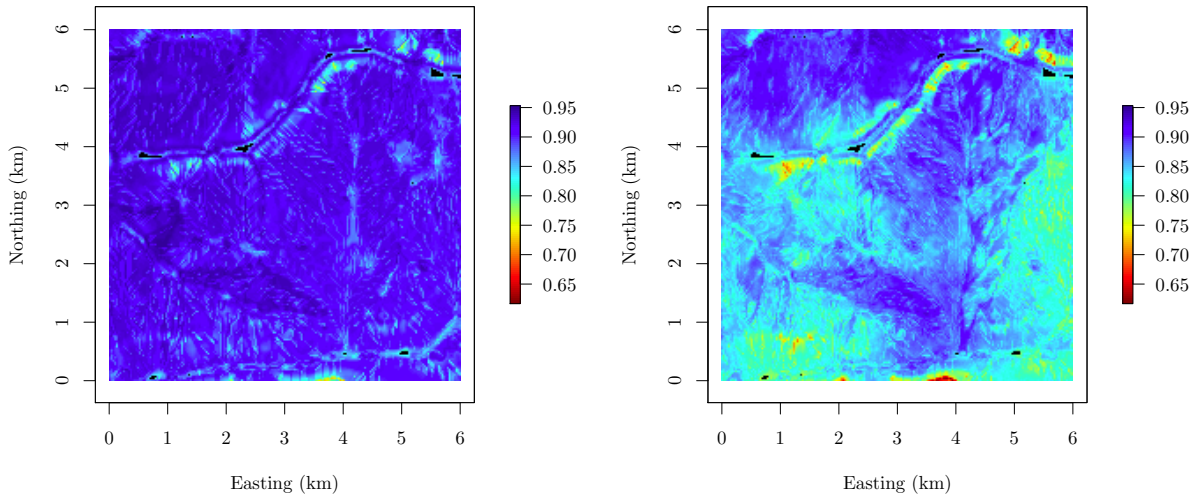


Figure A.4: Locations where permafrost presence was predicted over 200x200 pixel, 30x30 m resolution clips.



(a) nonspatial probability - site 1

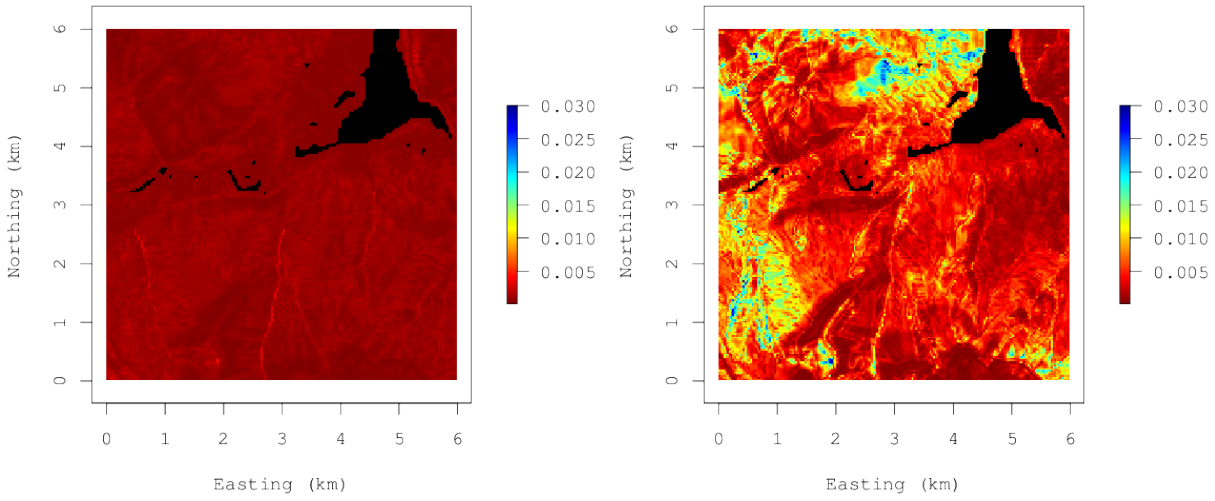
(b) spatial probability - site 1



(c) nonspatial probability - site 6

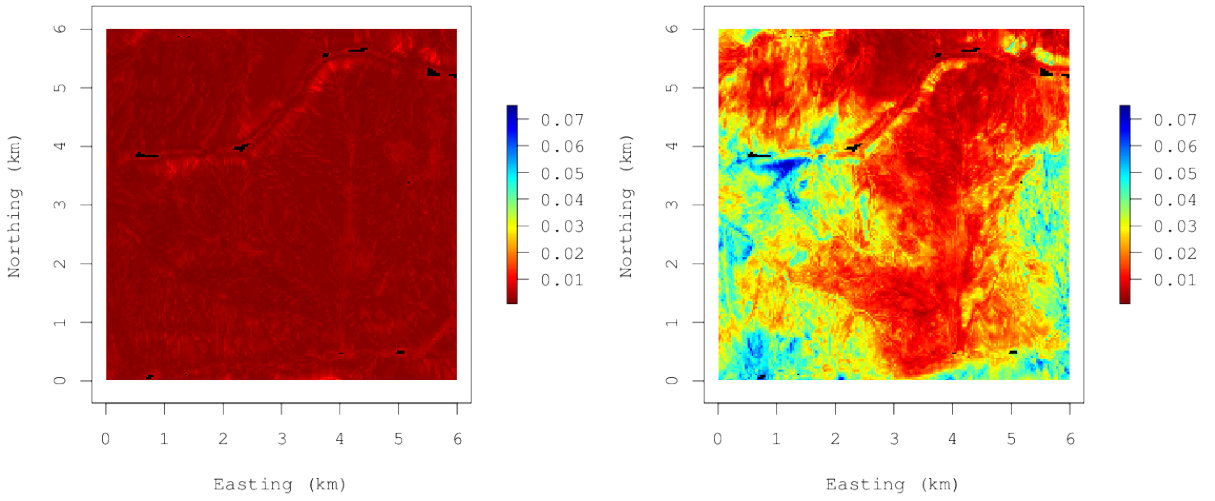
(d) spatial probability - site 6

Figure A.5: Probability permafrost, Sub-models 3 and 4.



(a) nonspatial variance - site 1

(b) spatial variance- site 1



(c) nonspatial variance - site 6

(d) spatial variance - site 6

Figure A.6: Probability permafrost variance, Sub-models 3 and 4.

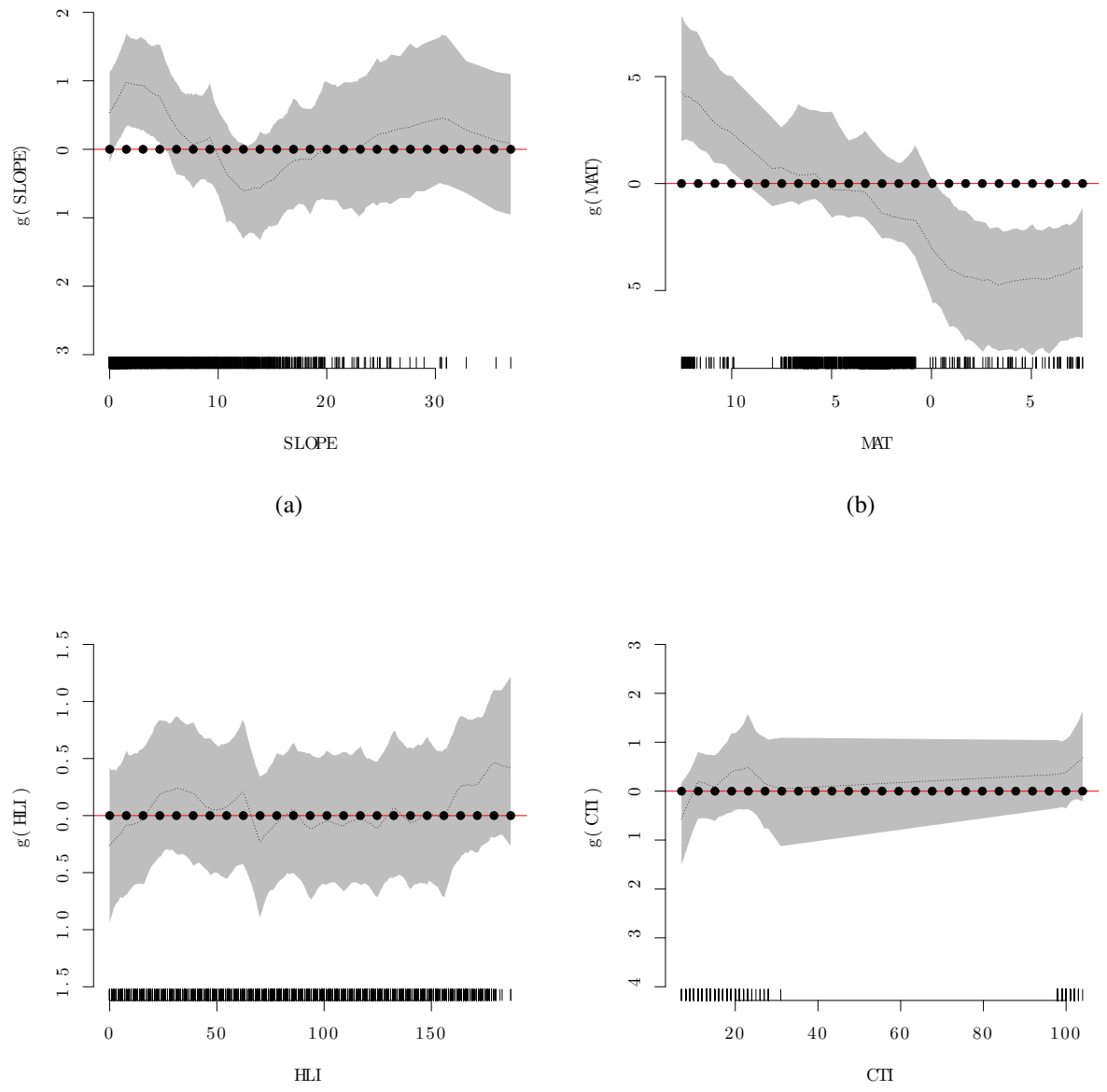


Figure A.7: Sub-model 4 regression coefficient estimates using $m = 50$ equally spaced knots. Black dotted line is the coefficient posterior median and gray region delineates the associated 95% credible interval. Black points along the red zero line indicate the location of predictive process knots across the covariate support. Rug dashes along the x-axis correspond to observed covariate values.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Randy G Balice, Jay D Miller, Brian P Oswald, Carl Edminster, and Stephen R Yool. Forest surveys and wildfire assessment in the los alamos region; 1998-1999. Technical report, Los Alamos National Lab, NM (US), 2000.

[2] Sudipto Banerjee, Alan E. Gelfand, Andrew O. Finley, and Huiyan Sang. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society Series*, 70:825–848, sep 2008.

[3] David J. Spiegelhalter Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64:583–639, 2002.

[4] JP Chilès. *P. deLFiner. 1999. Geostatistics: modeling spatial uncertainty*. Wiley Interscience, New York, 1999.

[5] Noel A. C. Cressie. *Statistics for Spatial Data*. Wiley, New York, NY, second edition, 1993.

[6] C Daly, GH Taylor, WP Gibson, TW Parzybok, GL Johnson, PA Pasteris, et al. High-quality spatial climate data sets for the united states and beyond. *Transactions of the ASAE-American Society of Agricultural Engineers*, 43(6):1957–1962, 2000.

[7] Peter Diggle, J. A. Tawn, and R. A. Moyeed. Model-based geostatistics. *Journal of the Royal Statistical Society*, 47:299–350, 1998.

[8] Andrew O Finley, Sudipto Banerjee, Bruce D Cook, and John B Bradford. Hierarchical bayesian spatial models for predicting multiple forest variables using waveform lidar, hyperspectral imagery, and large inventory datasets. *International Journal of Applied Earth Observation and Geoinformation*, 22:147–160, 2013.

[9] Andrew O. Finley, Huiyan Sang, Sudipto Banerjee, and Alan E. Gelfand. Improving the performance of predictive process modeling for large datasets. *Computational Statistics and Data Analysis*, 53:2873–2884, 2009.

[10] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. CRC press, 2013.

[11] Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992.

[12] Dean Gesch, Gayla Evans, James Mauck, John Hutchinson, and William J Carswell Jr. The national map: Elevation. *US geological survey fact sheet*, 3053(4), 2009.

[13] Paul E Gessler, ID Moore, NJ McKenzie, and PJ Ryan. Soil-landscape modelling and spatial prediction of soil attributes. *International Journal of Geographical Information Systems*, 9(4):421–432, 1995.

[14] Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[15] Trevor J Hastie and Robert J Tibshirani. *Generalized additive models*, volume 43. CRC Press, 1990.

[16] Kristofer D Johnson, Jennifer Harden, A David McGuire, Norman B Bliss, James G Bockheim, Mark Clark, Teresa Nettleton-Hollingsworth, M Torre Jorgenson, Evan S Kane, Michelle Mack, et al. Soil carbon distribution in alaska in relation to soil-forming factors. *Geoderma*, 167:71–84, 2011.

[17] Kristofer D Johnson, Jennifer W Harden, A David McGuire, Mark Clark, Fengming Yuan, and Andrew O Finley. Permafrost and organic layer interactions over a climate gradient in a discontinuous permafrost zone. *Environmental Research Letters*, 8(3):035028, 2013.

[18] MT Jorgenson and TE Osterkamp. Response of boreal ecosystems to varying modes of permafrost degradation. *Canadian Journal of Forest Research*, 35(9):2100–2111, 2005.

[19] ES Kane and JG Vogel. Patterns of total ecosystem carbon storage with changes in soil temperature in boreal black spruce forests. *Ecosystems*, 12(2):322–335, 2009.

[20] Michelle C Mack, Edward AG Schuur, M Syndonia Bret-Harte, Gaius R Shaver, and F Stuart Chapin. Ecosystem carbon storage in arctic tundra reduced by long-term nutrient fertilization. *Nature*, 431(7007):440–443, 2004.

[21] Neal J Pastick, M Torre Jorgenson, Bruce K Wylie, Burke J Minsley, Lei Ji, Michelle A Walvoord, Bruce D Smith, Jared D Abraham, and Joshua R Rose. Extending airborne electromagnetic surveys for regional active layer and permafrost mapping with remote sensing and ancillary data, yukon flats ecoregion, central alaska. *Permafrost and Periglacial Processes*, 24(3):184–199, 2013.

[22] Chien-Lu Ping, Richard D Boone, Marcus H Clark, Edmond C Packee, and David K Swanson. State factor control of soil formation in interior alaska. *AlaskaâĂŹs Changing Boreal Forest*, pages 21–38, 2006.

[23] Gareth O Roberts and Jeffrey S Rosenthal. Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

[24] Edward AG Schuur, James Bockheim, Josep G Canadell, Eugenie Euskirchen, Christopher B Field, Sergey V Goryachkin, Stefan Hagemann, Peter Kuhry, Peter M Lafleur, Hanna Lee, et al. Vulnerability of permafrost carbon to climate change: Implications for the global carbon cycle. *BioScience*, 58(8):701–714, 2008.

[25] David K Swanson. Susceptibility of permafrost soils to deep thaw after forest fires in interior alaska, usa, and some ecologic implications. *Arctic and Alpine Research*, pages 217–227, 1996.

[26] Simon Wood. *Generalized additive models: an introduction with R*. CRC Press, 2006.