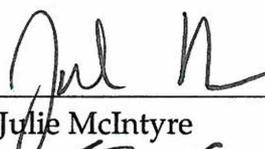


GAUSSIAN PROCESS CONVOLUTIONS FOR BAYESIAN
SPATIAL CLASSIFICATION

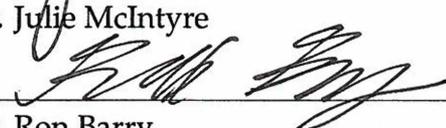
By

John K. Best Jr.

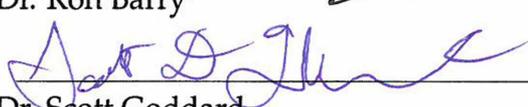
RECOMMENDED:



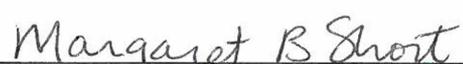
Dr. Julie McIntyre



Dr. Ron Barry

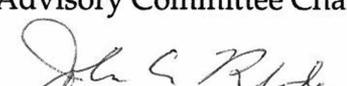


Dr. Scott Goddard



Dr. Margaret Short

Advisory Committee Chair



Dr. John Rhodes

Chair, Department of Mathematics and Statistics

GAUSSIAN PROCESS CONVOLUTIONS FOR BAYESIAN SPATIAL
CLASSIFICATION

A
PROJECT

Presented to the Faculty
of the University of Alaska Fairbanks
in Partial Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

By

John K. Best Jr., B.S.

Fairbanks, Alaska

May 2016

Abstract

We compare three models for their ability to perform binary spatial classification. A geospatial data set consisting of observations that are either permafrost or not is used for this comparison. All three use an underlying Gaussian process. The first model considers this process to represent the log-odds of a positive classification (i.e. as permafrost). The second model uses a cutoff. Any locations where the process is positive are classified positively, while those that are negative are classified negatively. A probability of misclassification then gives the likelihood. The third model depends on two separate processes. The first represents a positive classification, while the second a negative classification. Of these two, the process with greater value at a location provides the classification. A probability of misclassification is also used to formulate the likelihood for this model. In all three cases, realizations of the underlying Gaussian processes were generated using a process convolution. A grid of knots (whose values were sampled using Markov Chain Monte Carlo) were convolved using an anisotropic Gaussian kernel. All three models provided adequate classifications, but the single and two-process models showed much tighter bounds on the border between the two states.

1 Introduction

Subsurface conditions are always important when planning and building infrastructure. There are additional challenges at high latitudes, where permafrost and massive ice can greatly complicate or even rule out construction at a site. The standard method for evaluating potential sites is to drill a series of boreholes and then have an expert reconstruct subsurface conditions based on the information they provide. However, this is expensive and time consuming.

Reducing the number of boreholes required for a construction project is one goal of the Cold Regions Research and Engineering (CRREL) Permafrost Decision Support System project. One aspect of this entails using borehole information as efficiently as possible. Classifying regions of a transect into different categories is one aspect of this. We focus on classifying areas by permafrost state (a binary classification), but also consider a method that could be extended to classify locations into three or more categories.

Geostatistical models can also aid integration of other available information. Surface features can provide important initial indications of subsurface conditions. Geophysical data such as resistivity can also be acquired, typically much less expensively than boreholes. Bringing these information sources together will result in a more complete picture of subsurface conditions, and may help in planning borehole placement.

Binary spatial classification is typically accomplished by considering an

underlying Gaussian process and using a link function such as the inverse logit to provide a probabilistic classification at any point. A Gaussian process is a model where any finite collection of points has a multivariate normal distribution. The covariance matrix of this multivariate normal distribution results from the spatial correlation of the various locations within the region of interest.

If we have p observations of the process and m locations to be predicted, then the covariance matrix of this multivariate Gaussian distribution has dimension $p + m$. With many locations to predict, as when making a map, this quickly becomes very large. Finding the probability density of this multivariate normal distribution requires inverting this $(p + m) \times (p + m)$ covariance matrix. This is computationally expensive, typically increasing as $\mathcal{O}((p + m)^3)$.

One strategy to lower the dimensionality of the model is to use a set of knot locations with normally-distributed knot values. These can be convolved with a kernel to find a value at any point in the region. Any points in the region are then linear combinations of Gaussian random variables, and thus Gaussian themselves. This then meets our definition of a Gaussian process. This was explored in detail by Higdon (2001) and Lee et al. (2002).

Spatial correlation structure is induced by the convolution kernel. In our case, we expect observations to be more highly correlated horizontally than vertically due to the underlying geological processes. This is captured as anisotropy in the convolution kernel.

The 2×2 matrix describing the distance scale and anisotropy of the convolution kernel is the only matrix inversion required for this method. Besides this, the number of knots is chosen to reduce the number of parameters that must be sampled. These combine to make full Bayesian inference of spatial prediction computationally feasible.

The Gaussian process accounts for dependence in the data through a covariance function. Covariance determines the effect of nearby points on a point's predicted value. The specific parameterization of the covariance function determines how much a fitted surface will change on different length scales, and how smooth those changes will be. The parameterization of these covariance functions is where variogram methods are typically used. A variogram is a function describing the correlation of measurements with nearby data as a function of distance; it is based on the idea that nearby data will be more similar than more distant data.

The standard approach to binary prediction like this is a logistic regression model, where a log-odds surface is fit to the data. The value of this surface at a prediction location indicates the probability of a "success" at that location. However, this model implicitly allows state changes at small lags. If two locations separated by a short distance both have an 80% probability of "success", then they also have a 20% chance of failure, and thus the state could plausibly change multiple times over short distances under this model. The processes we are considering tend to be much smoother with contiguous areas of each state.

The other two models we consider use a simple classification criterion to generate a binary classification map. These predictions can then be compared to the observations we have to formulate our likelihood. Allowing for some misclassification, a simple Markov Chain Monte Carlo sampler can explore the space of knot values.

Each MCMC iteration represents a plausible map of the subsurface conditions. A well-mixed chain of knot value samples can then be considered to represent a sample from the posterior distribution of classification maps. Taking a point-wise mean across these maps then gives an uncertainty-based probability of permafrost at that point.

Section 2 will describe the data used in this study, the three likelihoods that are compared, and how the models were fit. Also described is a simulation study, used to verify that the models can classify more complicated structure than appears in the observed data. Section 3 presents the results of the model fits. Next, in section 4, is a discussion of the relative effectiveness of the different models. Finally, section 5 presents our conclusions and notes areas for future work.

2 Methods

2.1 Data Description

Between August 19 and September 5 2014, the CRREL team drilled a series of 34 boreholes along a 100m transect near the Cold Climate Housing Research

Center (CCHRC) in College, AK. The boreholes were drilled every 3m from 50m along a transect to 149m. Every 10cm along each borehole presence of permafrost, amount of visible ice, and soil type were recorded. Every 30cm a sample was taken and the gravimetric moisture content of the soil was measured.

The location of the transect can be seen in Figure 1. Figure 2 shows the extent of the transect. This was the 3rd transect in this area, so boreholes were recorded as “T3-xxx”. The three digit number denotes the distance along the transect where the borehole was taken. Note that the transect distance increases from right to left, so all subsequent images of the transect are orientated so that the viewer is facing generally South.

This work concentrates specifically on classifying areas of the transect as permafrost or not, and so makes exclusive use of those data. Further work incorporating the other data is ongoing. The spatial region addressed in this study extends horizontally along the transect, and vertically from ground level down approximately 8m. The permafrost/no permafrost observations from the boreholes can be seen in Figure 3.

2.2 Process convolutions

A grid of knot values was overlayed on the transect every 5m horizontally and 1m vertically. The locations of the knots can be seen in Figure 3. Horizontally, the grid extended from 45m on the transect to 155m, so that there was at least 5m to either side of the boreholes. Vertically, the grid extended



Figure 1: Map showing the general location of the borehole transect in relation to the University of Alaska Fairbanks campus and surrounding area.



Figure 2: Map showing the positions of individual boreholes.

from 125m to 135m above sea level. This includes 1m below the deepest borehole and 2–3m above ground level.

Then for n knot values x_1, \dots, x_n at locations $\mathbf{s}_1^*, \dots, \mathbf{s}_n^*$, the value of the spatial process Z at location \mathbf{s} is given by

$$Z(\mathbf{s}) = \sum_{i=1}^n x_i k(\mathbf{s}, \mathbf{s}_i^*; \mathbf{V}),$$

where k is the convolution kernel and \mathbf{V} is a matrix that determines the shape of the kernel.

For a given location (\mathbf{s}_0), we define the vector $\mathbf{w} \in \mathbb{R}^n$ as

$$\mathbf{w} = (k(\mathbf{s}_0, \mathbf{s}_i^*, \mathbf{V}) : i = 1, \dots, n)'$$

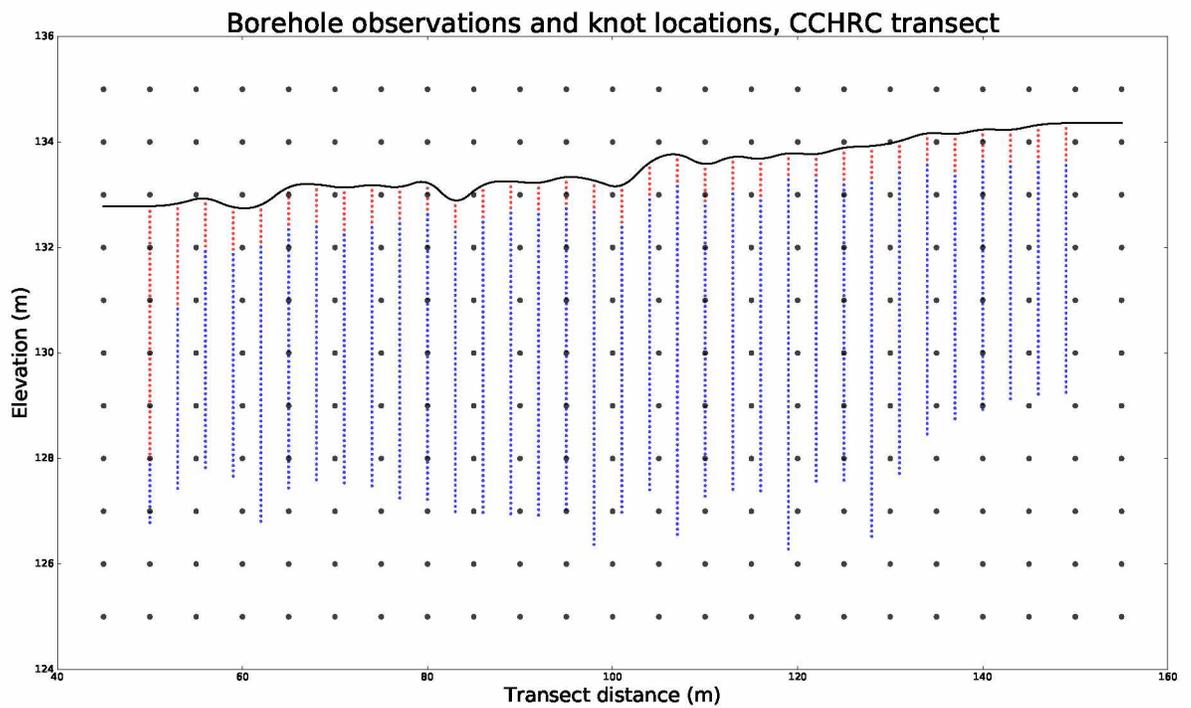


Figure 3: Permafrost/no permafrost observations from the borehole transect and knot locations. Red points represent observations of no permafrost. Blue points represent observations of permafrost. Black points are knot locations. The black line is the ground surface elevation.

These are the convolution weights for \mathbf{s}_0 . If the corresponding knot values are in a vector $\mathbf{x} \in \mathbb{R}^n$, then the value of Z at \mathbf{s} is

$$Z(\mathbf{s}_0) = \mathbf{w}'\mathbf{x}.$$

Extending this to a set of p locations $\mathbf{s}^{(1)}, \dots, \mathbf{s}^{(p)}$, we can precalculate the convolution weights for each of these as $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(p)}$. Then if we let $\mathbf{W} = \begin{bmatrix} \mathbf{w}^{(1)} & \mathbf{w}^{(2)} & \dots & \mathbf{w}^{(p)} \end{bmatrix}$ (thus $\mathbf{W} \in \mathbb{R}^{n \times p}$), the value of Z at these locations is the vector

$$\mathbf{Z}(\mathbf{s}^{(1)}, \dots, \mathbf{s}^{(p)}) = \mathbf{W}'\mathbf{x}.$$

Computationally this is very efficient, as it can take advantage of highly optimized matrix multiplication in libraries such as OpenBLAS or MKL.

2.3 Convolution kernel

The spatial process is a moving average of the knot values, with weights determined by a convolution kernel. In this case, we use a Gaussian convolution kernel. This induces a Gaussian spatial covariance in the represented surface (Higdon, 2001). Because we are dealing with geological processes, we expect that horizontal correlations will occur at longer distances than vertical correlations. In order to account for this anisotropy, we use a diagonal matrix to shape the convolution kernel appropriately. For knot location \mathbf{s}_i^* , location to be predicted \mathbf{s}_0 , and anisotropy matrix \mathbf{V} , the weight assigned to that knot

value is given by

$$k(\mathbf{s}_0, \mathbf{s}_i^*; \mathbf{V}) = \exp\left(-\frac{1}{2}(\mathbf{s}_0 - \mathbf{s}_i^*)' \mathbf{V}^{-1}(\mathbf{s}_0 - \mathbf{s}_i^*)\right).$$

In this case, we let

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix},$$

where $\sigma_1^2 = 14$ and $\sigma_2^2 = 1$, indicating correlation between points separated by $\sqrt{14}$ m horizontally was the same as correlation between points separated by 1m vertically. An early exploratory model fit these length parameters as part of the model. These values were the mean of the fitted values, and we have carried them through the present analysis.

2.4 Log-odds model

Three likelihoods were used and compared in this study. For each of them we consider different relationships between the value of Z and the presence of permafrost at a given location along the transect. As such, while fitting the model we only need to calculate the value of Z at the locations where we have observations. Each of these observation locations is classified as either permafrost or not, so for convenience we define

$$\text{PF}(\mathbf{s}) = \begin{cases} 1 & \text{if permafrost is present at } \mathbf{s}, \\ 0 & \text{if permafrost is } \textit{not} \text{ present at } \mathbf{s}. \end{cases}$$

The first, which we will call the “log-odds” (LO) model, considered the surface represented by Z to be the log-odds of a particular point being permafrost. It uses the standard inverse logit link function, so that for a spatial process value of $Z(\mathbf{s})$, the probability of permafrost at \mathbf{s} is

$$\begin{aligned} \Pr [\text{PF}(\mathbf{s}) = 1] &= \text{logit}^{-1} [Z(\mathbf{s})] \\ &= \frac{\exp [Z(\mathbf{s})]}{1 + \exp [Z(\mathbf{s})]}. \end{aligned}$$

Thus, the likelihood is

$$\begin{aligned} \mathcal{L}_{\text{LO}}(\mathbf{x}) &= \prod_{i=1}^p \{ \text{logit}^{-1} [Z(\mathbf{s}^{(i)})] \}^{\text{PF}(\mathbf{s}^{(i)})} \times \\ &\quad \{ 1 - \text{logit}^{-1} [Z(\mathbf{s}^{(i)})] \}^{1 - \text{PF}(\mathbf{s}^{(i)})}. \end{aligned}$$

This likelihood has some advantages. It is differentiable, and so can be sampled more efficiently using Hamiltonian Monte-Carlo and other gradient-based samplers. These characteristics were not taken advantage of in this study, but could be an area of exploration going forward.

2.5 Single-process model

The log-odds model gives only a probability of permafrost at a given location. These probabilities are assumed spatially correlated, but there is still room

for a “nugget” effect, where moving a small distance could result in a changed state even if the probability does not change. This does not match our expectation; the states will probably be patchy, with large contiguous areas of the same state. In some respect, our real concern is locating the border between the two states. Thus, it makes more sense to assign a discrete state to each point, and then calculate the likelihood from those. This is the approach that the other two likelihoods take.

The first of these uses a simple cutoff value, in this case zero. At location \mathbf{s} , if $Z(\mathbf{s}) \geq 0$, that point is predicted to be permafrost. If $Z(\mathbf{s}) < 0$, then permafrost is not predicted for that location. We call this the single-process (1P) model.

By comparing these predictions to the observations from the boreholes, we can formulate a likelihood based on a probability that a borehole observation was misclassified. This misclassification is not necessarily because we expect incorrect observations. However, it is required if we expect to be able to sample from the model.

If we allowed no misclassification, the likelihood would be equal to zero if a single observations was misclassified, jumping to one when they are all correctly classified. In a model with hundreds of knots (and an equally large-dimensional parameter space), the volume of the parameter space with zero likelihood is so much larger than that with likelihood one that it would be virtually impossible to find a set of parameter values that correctly predicts every observation. Allowing for some misclassification allows for incremen-

tal improvements; a set of parameter values where 90% of observations are classified correctly *should* be more likely than one that correctly classifies only 80% of observations. The magnitude of allowed misclassification is only important in terms of the MCMC iterations. The misclassification rate determines how often a proposed parameter value that “loses” a correctly classified observation would be accepted. Besides, allowing for a 0.001 chance of misclassification translates to an expectation of one observation in 1,000 being misidentified, which seems not unreasonable.

Then if we let m be the misclassification rate, and define

$$Z_{1P}^*(\mathbf{s}) = \begin{cases} 1 & \text{if } Z(\mathbf{s}) \geq 0, \\ 0 & \text{if } Z(\mathbf{s}) < 0, \end{cases}$$

then the likelihood is

$$\mathcal{L}_{1P}(\mathbf{x}) = \prod_{i=1}^p (1 - m)^{\mathbb{I}[Z_{1P}^*(\mathbf{s}^{(i)}) = \text{PF}(\mathbf{s}^{(i)})]} \times m^{\mathbb{I}[Z_{1P}^*(\mathbf{s}^{(i)}) \neq \text{PF}(\mathbf{s}^{(i)})]}.$$

2.6 Two-process model

The third model uses two separate processes, one for permafrost and one for absence of permafrost. This “two-process” (2P) model doubles the number of parameters, but provides a useful jumping-off point for prediction of categorical responses with more than two states. For the same reasons as above,

this likelihood also allows for some misclassification. The process with the maximum value at a spatial location provides the prediction for that location.

So if we let Z_{PF} be the process representing the presence of permafrost and Z_N be the process representing the absence of permafrost, then

$$Z_{2\text{P}}^*(\mathbf{s}) = \begin{cases} 1 & \text{if } Z_{\text{PF}}(\mathbf{s}) \geq Z_N(\mathbf{s}), \\ 0 & \text{if } Z_{\text{PF}}(\mathbf{s}) < Z_N(\mathbf{s}), \end{cases}$$

and the likelihood is the same as above, so

$$\mathcal{L}_{2\text{P}}(\mathbf{x}) = \prod_{i=1}^p (1 - m)^{\mathbb{I}[Z_{2\text{P}}^*(\mathbf{s}^{(i)}) = \text{PF}(\mathbf{s}^{(i)})]} \times m^{\mathbb{I}[Z_{2\text{P}}^*(\mathbf{s}^{(i)}) \neq \text{PF}(\mathbf{s}^{(i)})]}.$$

2.7 Priors

In all cases, the parameters to be sampled were the knot values. These were given the same prior, so that

$$x_i \stackrel{\text{iid}}{\sim} \text{Normal}(0, 1).$$

This was done specifically to regularize the knot values. Because the likelihood does not change based on the magnitude of the underlying process for the single- and two-process models, there is little to prevent the knot values from increasing or decreasing to arbitrary magnitudes.

2.8 Implementation

All models were fitted using a Metropolis sampler with an independent Normal proposal function, as described in Gelman et al. (2014, p. 278). The first 5000 iterations were used to adapt the variance of the proposal distribution. During this stage, every 200 iterations the acceptance rate was calculated for each knot value. If the acceptance rate was less than 30%, the proposal width for that knot value was halved. If the acceptance rate was greater than 60%, the proposal width was doubled. Proposal widths were held constant for the subsequent 1,000,000 samples.

At the beginning of each MCMC iteration, the order of knot value updates was randomly permuted. This was done so that knot values were not consistently updated early or late in the process. As each knot value was updated, the unnormalized log-posterior was calculated, and the knot value accepted or rejected according to its Metropolis ratio.

All of the models and the Metropolis random walk sampler were coded in the Julia programming language (Bezanson et al., 2014, 2012). The Distributions.jl package was used to find log densities. The PyPlot.jl package was used to access the Matplotlib library (Hunter, 2007) for plotting. Unnormalized posterior densities were calculated on the log scale in order to avoid underflows and other numerical issues.

A single chain of 1 million samples was run for each of the three models. This was then thinned, keeping every 100 in order to make storage and processing of the samples reasonable. All predictions are based on these

10,000 samples.

2.9 Post-processing

Each set of knot values represents a realization of the underlying spatial Gaussian processes. These were interpolated on a 1025×94 grid within the bounds of the knot locations. Thus, the surface was interpolated roughly every 10cm. The summary statistics used to create the final maps were different depending on the model used.

The log-odds model requires some choice in how to summarize the represented surfaces. The value of the process was calculated at the interpolation locations. The mean and the median values of this surface was found, and these were then transformed using the inverse logit function to get a probability of permafrost. The mean and median values were very similar, so only the mean values are considered.

For the single-process model, the value of the spatial process was calculated at each interpolation location. Anywhere the spatial process was greater than 0 represented a classification of permafrost was present at that location. The proportion of MCMC samples that classified a location as permafrost was then used as the probability of permafrost at that location.

The classification was similar for the two-process model. The values of the permafrost and non-permafrost spatial processes were interpolated at each location. These were then compared, and the process with greater value provided the classification. Again, the proportion of processes that classified

a location as permafrost was used as the probability of permafrost at that location.

2.10 Synthetic data study

The original data has a fairly simple structure. There is a layer of non-permafrost, and everything below this is permafrost. We altered a number of observations to test the model’s ability to fit more complicated structure.

Observations of permafrost were modified in two areas. The first was an ellipse centered at 75.5m along the transect and an elevation of 129.5. The horizontal semiaxis length was 14m and the vertical semiaxis length was 1.25m. All the observations within this ellipse were permafrost. These were changed to no permafrost.

The second ellipse was centered at 130m along the transect and 129m elevation. Its horizontal semiaxis was 9.5m and its vertical semiaxis was 2m. Again, all observations inside this ellipse were changed from permafrost to no permafrost.

These “holes” introduce some large-scale complexity to the observations. The same three models were fit using these modified data. Again, 1 million iterations of the Metropolis sampler were run, and every 100th sample was taken for final inferences and prediction.

3 Results

3.1 Original data

Figure 4 shows the maps of probability of permafrost at a given location along the transect using all three models. At the top is the mean of the log-odds surface, then the single-process prediction, and finally the two-process prediction.

Figure 5 shows maps of the Bernoulli variance of the predictions. This is one way to represent the uncertainty at a location, as predicting that permafrost will be present with 50% probability essentially states that we have maximum uncertainty. Thus, if the predicted probability of permafrost is p , then the Bernoulli variance of that prediction is $p(1 - p)$, which takes values between 0 and 0.25.

3.2 Synthetic data study

Figure 6 uses the same models as the used in Figure 4, but with observations modified as described in section 2.10. The two ellipses with altered observations are clearly visible. The modified observations do not seem to have changed the classifications in the unmodified area.

Again, the log-odds model shows the greatest uncertainty as to the locations of borders between states. This is visible in Figure 6, and strikingly clear in Figure 7, with a thick band of high variance around each modified oval. The other two models again provide very similar maps. Notably, they

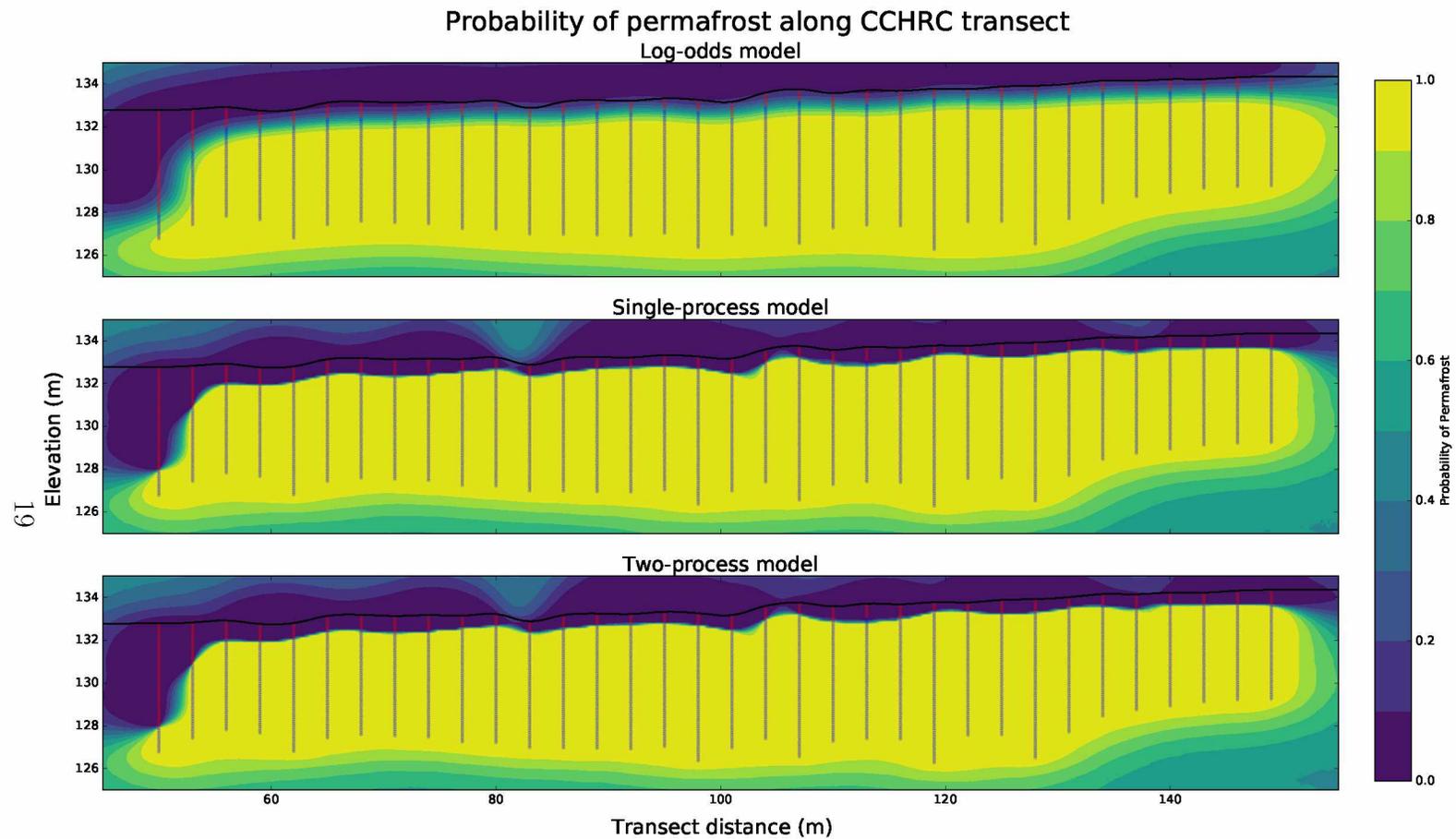


Figure 4: Probability of finding permafrost along the CCHRC transect under the log-odds, single process, and two-process models. Red points represent observations of no permafrost, while blue points are locations where permafrost was observed.

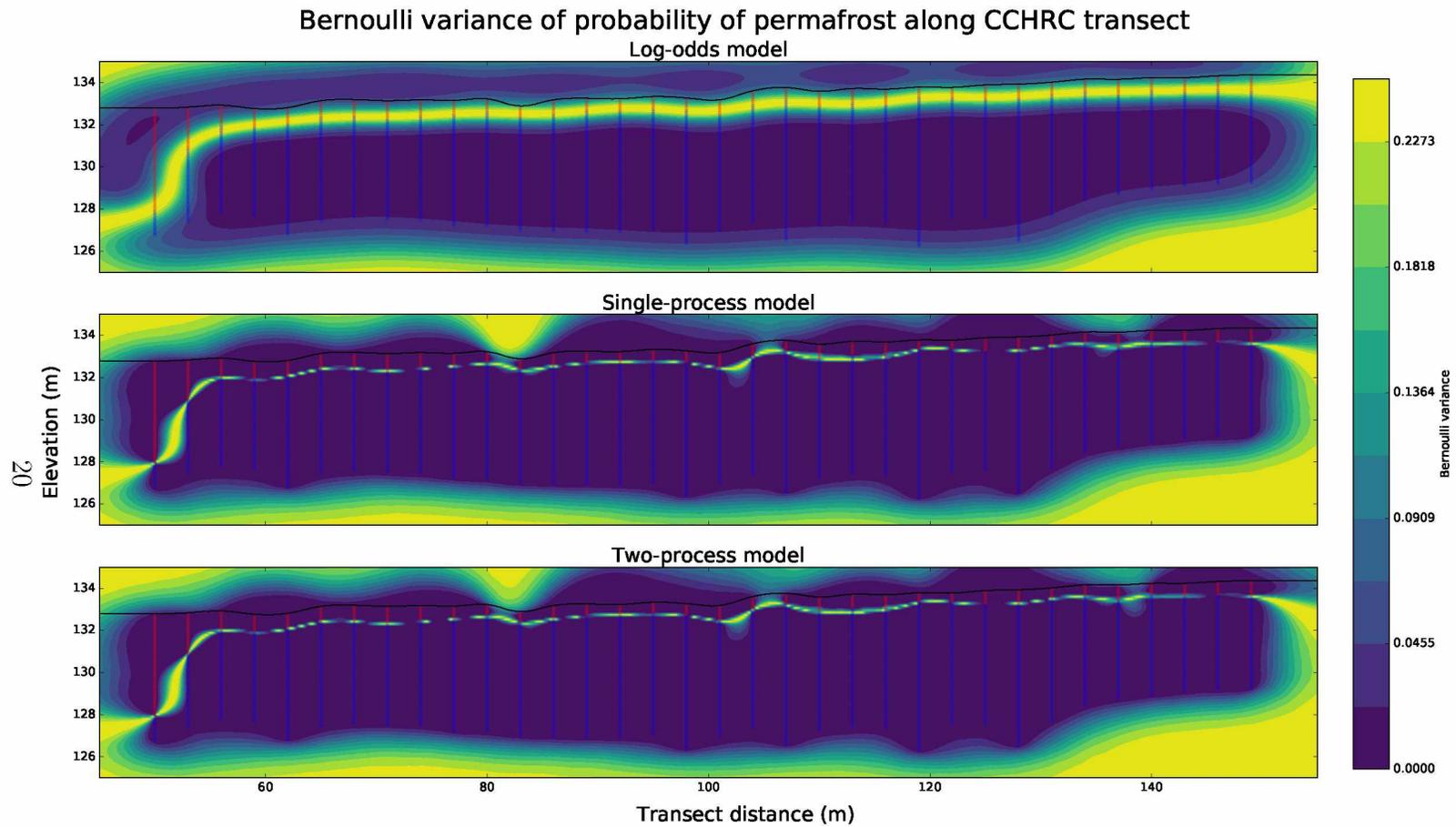


Figure 5: Variance of permafrost predictions. Lower variance indicates a probability close to either zero or one.

both show some uncertainty as to the location of the edge of permafrost at the edge of the oval around transect distance 120m. This uncertainty is appropriate given the observations.

4 Discussion

All three of the likelihood models demonstrate an ability to classify transect areas as permafrost or not consistent with observations. The modified set of observations used in the simulation study demonstrate this even with more complicated subsurface structure. The main difference between the models is that the log-odds model provides much less well-defined borders between types.

This is visible in the rapid transition between low and high probabilities of permafrost where observations change in figures 4 and 6. It is also apparent when looking at the Bernoulli variance of the predictions in figures 5 and 7.

The log-odds model does have some advantages over the other two models however. The likelihood is smooth and differentiable, and so more efficient MCMC algorithms could be used. It is also easier to formulate interpretable priors for additional effects. For example, if one believes that a specific surface condition is associated with permafrost at a given depth, a prior can be formulated by considering how much the additional information is expected to change the log-odds of the state. In the single-process model, one must formulate a prior by considering how frequently the information should

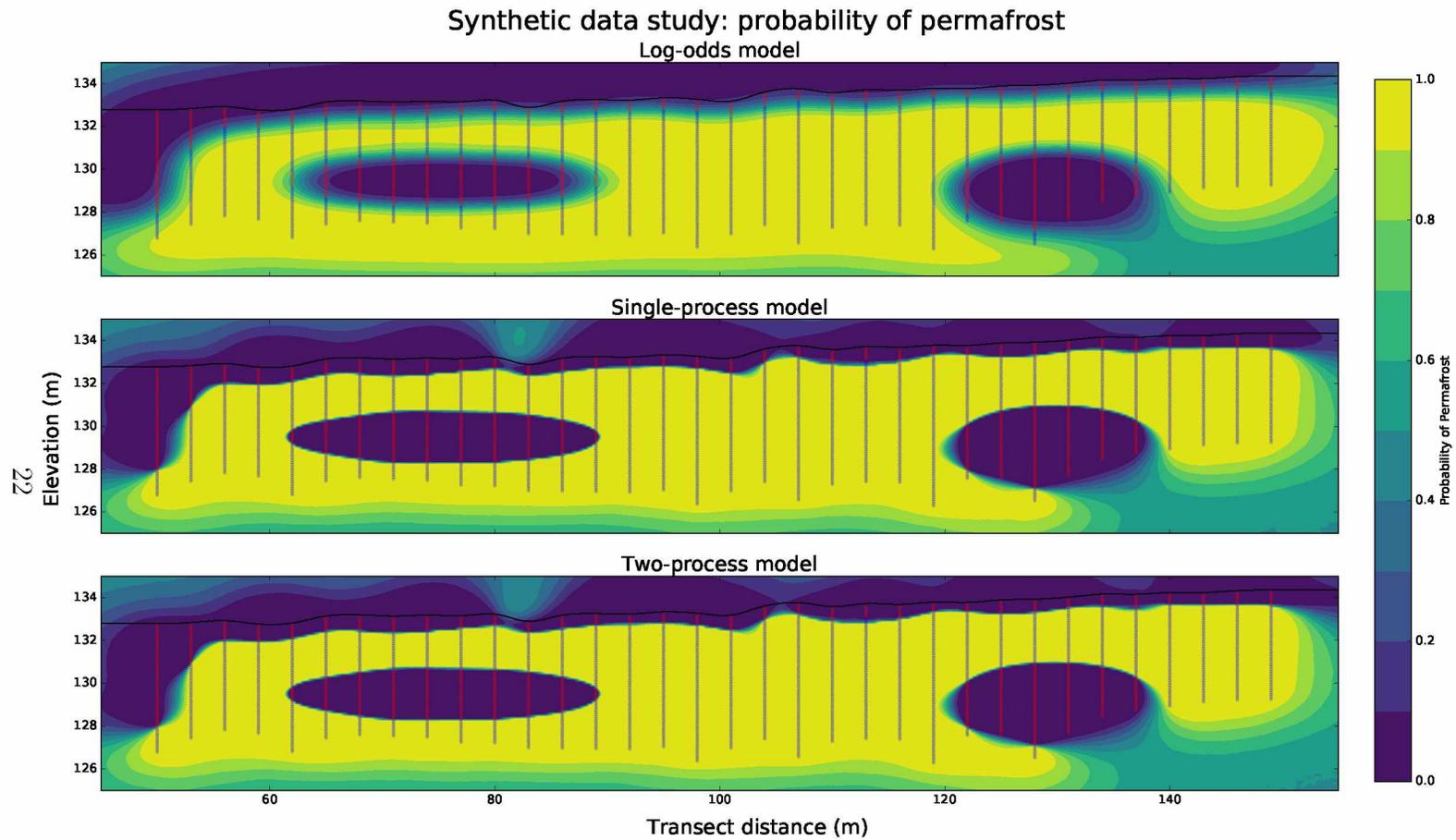


Figure 6: Probability of permafrost predictions for modified observations in the simulation study.

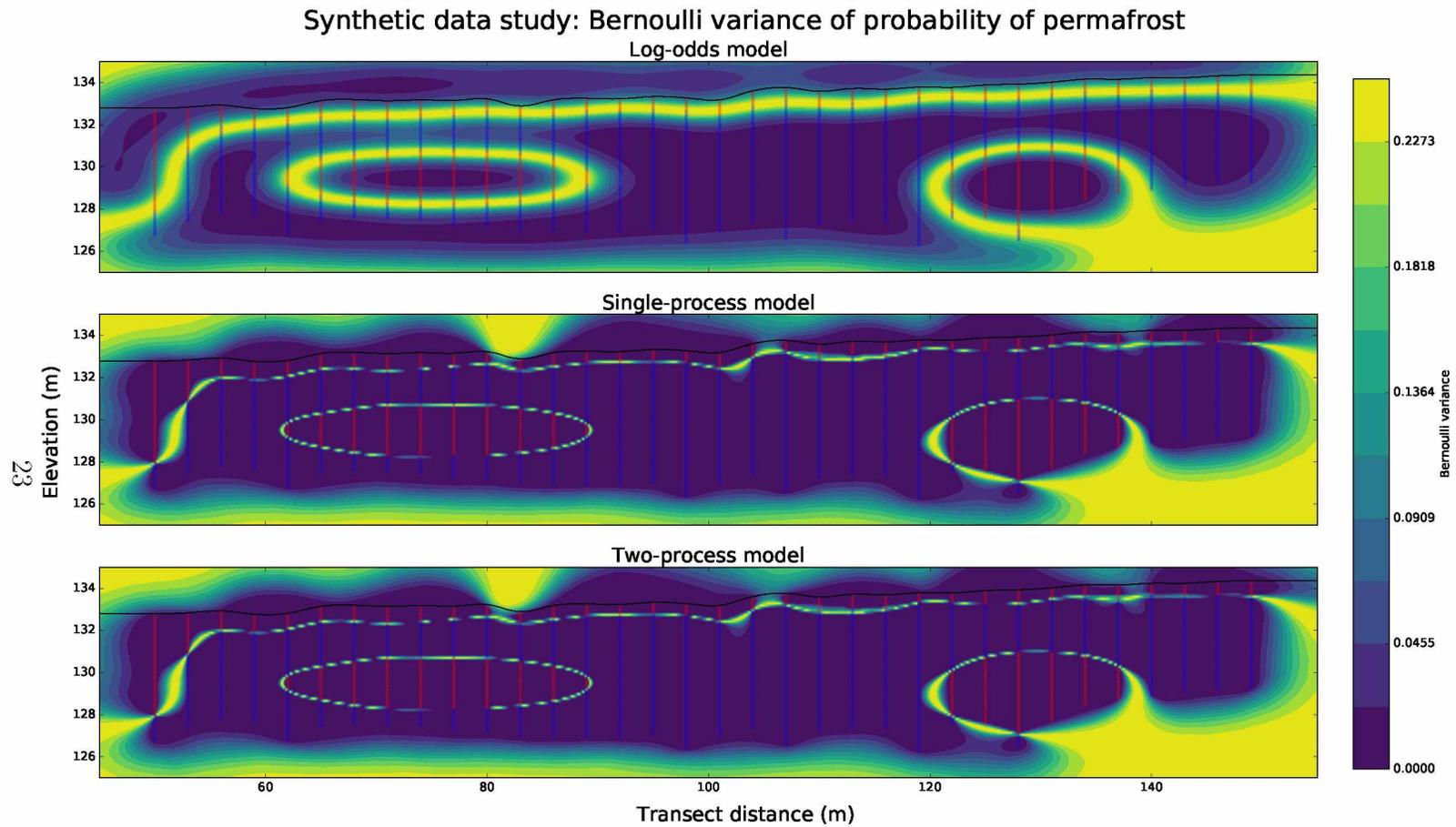


Figure 7: Variance of probability predictions for modified observations in the simulation study.

move the surface above or below the cutoff. The two-process model then requires setting a prior for both processes, again considering how frequently it should be positive or negative and with what magnitude.

Because the value of the spatial process in the log-odds model has a one-to-one correspondence with the probability of finding permafrost, the log-odds model may also be more sensitive to the prior. The lower certainty demonstrated in this model may have been a result of the $\text{Normal}(0, 1)$ prior on the knot values. This may have reduced the magnitude of the log-odds surface near the border of the two states.

In this case, the classification maps resulting from the single- and two-process models appear very similar. They also appear equally effective. Predictions follow observations very closely, and uncertainty increases farther from the observations. The border between the permafrost and non-permafrost regions are well-defined, and much sharper than that given by the log-odds model.

The simulation study shows a similar effect. The log-odds model shows a band of uncertainty as to which points are or are not permafrost (c.f. figure 7). This does not appear in either of the other two models. There is uncertainty around 120m. This is to be expected, as the observations support a border between the two states anywhere between the two boreholes.

In this case, the two-process model appears to be almost identical to the single-process model. This provides some evidence that this model could be extended from predicting a binary response to a categorical response with

three or more categories by modeling a corresponding number of spatial processes and taking the processes with maximum value at a location as the classification at that location.

5 Conclusion

Process convolutions can represent Gaussian processes in a very computationally efficient manner. They reduce the dimension of the parameter space, and allow for easy interpolation without refitting the model. The single and two-process models provided more effective classifications than the log-odds model.

All of these models provide a basis for further work on transect profiles for infrastructure planning. They can be used to determine the number of boreholes needed, and can serve as the basis for integrating additional sources of information. Some version of these models may be integrated into a Decision Support System in the future, where users can receive interactive guidance about building in areas where there is concern about permafrost.

It appears that the two-process model could be extended to an k -process model for observations with k categories. This may be useful for this project if mapping soil types becomes important. Otherwise, it may have applications in other classification tasks. We expect that when it is appropriate, the k -process model will outperform an analogous multi-logit model.

5.1 Future work

There are many paths forward in this work. Using a k -fold cross-validation scheme, it will be interesting to quantify the difference in predictive ability between the models presented here. This could be done using a Brier score (Brier, 1950) or an area under ROC curve.

Because the calculations were fast, it was reasonable to run the MCMC chain for a large number of iterations and then thin it heavily. It would be useful to find a more efficient sampler so that fewer samples are necessary. The log-odds model can be implemented in MCMC software (e.g. Stan) that implement more efficient samplers because they take advantage of the differentiability of the posterior in order to more efficiently sample from the parameter space. This may be difficult for the single- and two-process models because they are not differentiable everywhere.

Fitting the length parameters for the convolution kernel rather than just setting them as constant would allow for some inferences about the underlying processes, but may not improve prediction much. This would likely require many more MCMC iterations, and so may be predicated on a more efficient sampler.

One goal of the project is to integrate surface features and resistivity data. These have always provided important information when formulating a subsurface reconstruction, and so informative priors should be used as initial guesses at subsurface structure.

Extending the two-process model to many processes in order to predict

a categorical state. These models could be very effective for this type of predictions given how flexible they are. It may be possible that each process representing a particular category could have different spatial structure (i.e. a different convolution kernel). This could allow for the prediction of things like massive ice wedges intruding vertically into horizontally-bedded soil layers. This data set has soil type data, which could be approached this way.

There are also questions about how the knot density relates to the level of detail that can be captured by these models. This could guide future modeling efforts, especially because knots do not have to be placed regularly. Thus more knots could be placed in areas with more detailed structure.

References

- Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2014). Julia: A fresh approach to numerical computing.
- Bezanson, J., Karpinski, S., Shah, V. B., and Edelman, A. (2012). Julia: A fast dynamic language for technical computing.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*. Taylor and Francis/CRC Press, third edition.

- Higdon, D. (2001). Space and space-time modeling using process convolutions. *Working Paper, Duke University Department of Statistical Science.*
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, 9(3):90–95.
- Lee, H. H. K., Holloman, C. H., Calder, C. A., and Higdon, D. M. (2002). Flexible Gaussian processes via convolution. *Working paper, Duke University Department of Statistical Science.*