STAT 517 Winter 2024 Final Project

Wenhao Pan
Department of Statistics
University of Washington
Seattle, WA 98195
wenhaop@uw.edu

Abstract

This is the written report on the final project of STAT 517 – Stochastic Modeling Of Scientific Data II in Winter 2024, taught by Professor Zaid Harchaoui. The project connects the material encountered in the course to recent methodological research on stochastic processes. We select the paper *Constructing Priors that Penalize the Complexity of Gaussian Random Fields* from Fuglstad et al. [2019]. In the report, we summarize the main research problem in the selected paper, connect the main contribution in the selected paper with materials studied in STAT 517, propose the simplified version of the main contribution, and conduct the simulation study on the simplified version.

1 Main Research Problem

The main research problem addressed in the paper *Constructing Priors that Penalize the Complexity of Gaussian Random Fields* Fuglstad et al. [2019] is the challenge of selecting appropriate prior distributions for Gaussian Random Fields (GRFs) in Bayesian hierarchical models. Fuglstad et al. [2019] focused on one-dimensional, two-dimensional, and three-dimensional GRFs with Matérn covariance functions with fixed smoothness, extending to nonstationary covariance structures. The Matern covariance function creates a ridge in the joint likelihood of the range and the marginal variance parameters [Warnes and Ripley, 1987], and no consistent estimator exists for them under in-fill asymptotics when the dimension of the GRF is three or lower [Stein, 2012, Zhang, 2004].

To the knowledge of Fuglstad et al. [2019] back then, only Berger et al. [2001] introduced a principled approach to prior selection for GRFs. However, they derived reference priors for a GRF partially observed without noise. In contrast, GRFs are often embedded in Bayesian hierarchical models in an over-complex way to derive the reference priors [Fuglstad et al., 2019]. And they did not provide guidance on which hyperparameters should be selected for the prior. Therefore, the Fuglstad et al. [2019] proposed a principled joint prior for the range and marginal variance of Matérn GRFs, which is weakly informative and penalizes complexity by shrinking toward a *base model* with infinite range and zero marginal variance through hyperparameters that indicate how strongly the user wishes to shrink toward the base model. Specifically, Fuglstad et al. [2019] used the penalized complexity (PC) prior framework [Klein and Kneib, 2016, Simpson et al., 2017] to construct a joint prior, which is independent of the observation process, for the range and the marginal variance parameters of a Matern GRF. To argue that their approach is valid, Fuglstad et al. [2019] answered the following three questions:

- Is the PC prior framework suitable for infinite-dimensional model components?
- How can we deal with the fact that the KLD between Matérn GRFs in general is infinite?
- How can we construct a multivariate PC prior that properly accounts for the intrinsic link between range and marginal variance due to the ridge in the likelihood?

Preprint. Under review.

Moreover, Fuglstad et al. [2019] show that the PC prior developed for the stationary Matern GRF can be extended further to a prior for a nonstationary GRF, where the nonstationarity is controlled by covariates. We do not discuss this extension in the report for the sake of time and space, but we strongly encourage the audience to study this piece of work if interested.

The research problem is significant as the choice of prior distribution profoundly impacts the behavior of the posterior of the parameters, especially under in-fill asymptotics where the likelihood provides limited information about the covariance structure. The proposed approach aims to provide a more principled alternative to reference priors, allowing practitioners to include expert knowledge in an interpretable way.

2 Simplification of the Main Contribution

The core part of the main contribution in [Fuglstad et al., 2019] is their *Theorem 2.6* as follows:

Theorem 1. *(PC prior for the Matern(ρ, σ)). Let u be a GRF defined on $\mathbb{R}^d$, where $d \leq 3$, with a Matern covariance function with parameters σ, ρ, ν. Assume ν is fixed. Then, the joint PC prior corresponding to a base model with infinite range and zero marginal variance is*

$$\pi(\sigma, \rho) = \frac{d}{2}\tilde{\lambda}_1\tilde{\lambda}_2\rho^{-d/2-1}\exp\left(-\tilde{\lambda}_1\rho^{-d/2} - \tilde{\lambda}_2\sigma\right), \quad \sigma > 0, \quad \rho > 0$$

where $\mathbb{P}(\rho < \rho_0) = \alpha_1$ and $\mathbb{P}(\sigma > \sigma_0) = \alpha_2$ are achieved by

$$\tilde{\lambda}_1 = -\log(\alpha_1)\rho_0^{d/2} \quad \textit{and} \quad \tilde{\lambda}_2 = -\frac{\log(\alpha_2)}{\sigma_0}.$$

In this PC prior, $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ are two hyperparameters jointly control the marginal tailed probability of ρ and σ at ρ_0 and σ_0 with α_1-level and α_2-level. To make this theorem more self-contained, we present Fuglstad et al. [2019]'s definition of a Matern covariance function.

Definition 1. (Matern covariance function) A Matern covariance function $c : [0, \infty] \to \mathbb{R}$ can be parameterized through a marginal standard deviation σ, a range parameter ρ, and a smoothness parameter ν, and is given by

$$c_\nu(r; \sigma, \rho) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{8\nu}\frac{r}{\rho}\right)^\nu K_\nu\left(\sqrt{8\nu}\frac{r}{\rho}\right)$$

where K_ν is the modified Bessel function of the second kind, order ν.

2.1 Derivation

We start the derivation of Theorem 1 by introducing the PC prior framework [Simpson et al., 2017].

2.1.1 PC Prior Framework

The first step is to design a distance metric from the base model to its extension using the Kullback-Leibler divergence (KLD).

Definition 2. (Kullback-Leibler divergence). Let P_0 and P be measures over the set $\mathcal{X}$, where P is absolutely continuous with respect to P_0, then the Kullback-Leibler divergence from P_0 to P is defined as

$$\mathrm{KL}(P||P_0) = \int_{\mathcal{X}} \log\frac{dP}{dP_0}dP,$$

where $\frac{dP}{dP_0}$ is the Radon-Nikodym derivative of P with respect to P_0.

Let P_0 denote the Gaussian measure of the base model for the GRF and P denote the Gaussian measure of the flexible extension, then the defined distance metric is $\mathrm{dist}(P||P_0) = \sqrt{2\mathrm{KL}(P||P_0)}$.

The second step is to define the prior on $\mathrm{dist}(P||P_0)$ based on three principles:

- Occam's razor: The prior penalizes more strongly when the flexible extension is further from the base model

- Constant-rate penalization: The prior on the distance, t, satisfies

$$\frac{\pi(t+\delta)}{\pi(t)} = r^{\delta}, \qquad t, \delta > 0,$$

for a constant decay rate $0 < r < 1$. The only continuous distribution with this property is the exponential distribution $\pi(t) = \lambda \exp(-\lambda t)$ for $t > 0$.

- User-defined scaling: The prior has a hyperparameter λ that has an interpretable way for the user to set its value.

2.1.2 Adaptation of the Framework

To adapt the PC Prior Framework for GRF with the dimension of three or lower, Fuglstad et al. [2019] started by providing an alternative parameterization of the Matern covariance function in Definition 1.

Definition 3. (Alternative parameterization of the Matern covariance function). Assume that the base space is $\mathbb{R}^d$ and introduce

$$\kappa = \sqrt{8\nu}/\rho \quad \text{and} \quad \tau = \sigma\kappa^{\nu}\sqrt{\frac{\Gamma(\nu + d/2)(4\pi)^{d/2}}{\Gamma(\nu)}}.$$

This alternative parameterization benefits from its description that τ can and κ cannot be consistently estimated under in-fill asymptotic when the dimension of the base space ≤ 3 [Fuglstad et al., 2019]. Then, if we separate the joint prior $\pi(\tau, \kappa)$ by $\pi(\tau|\kappa)$ and $\pi(\kappa)$, the PC prior for $\tau|\kappa$ must be derived based on a finite-dimensional observation.

Theorem 2. *(PC prior for $\tau|\kappa$). Let u be a GRF defined on $D \subset \mathbb{R}^d$ with a Matern covariance function with parameters τ, κ, ν. If the GRF is observed on $s_1, s_2, \ldots, s_n \in \mathcal{D}$, then conditionally on κ the PC prior for τ with base model $\tau = 0$ is*

$$\pi(\tau_\kappa) = \lambda \exp\left(-\lambda\tau\right), \qquad \tau > 0$$

where $\lambda > 0$ is a hyperparameter.

To control $\mathbb{P}(\sigma > \sigma_0|\kappa)$, the upper tail probability of the marginal standard deviation σ exceeding σ_0, at α, we can set

$$\lambda(\kappa) = -\kappa^{-\nu}\sqrt{\frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(4\pi)^{d/2}}}\frac{\log(\alpha)}{\sigma_0}.$$

Next, Fuglstad et al. [2019] contrusted the PC prior for κ using the infinite-dimensional GRF.

Theorem 3. *(PC prior for κ). Let u be a GRF defined on $D \subset \mathbb{R}^d$, where $d \leq 3$, with a Matern covariance function with parameters τ, κ, ν. The PC prior for κ with base model $\kappa = 0$ is*

$$\pi(\kappa) = \frac{d}{2}\lambda\kappa^{d/2-1}\exp\left(-\lambda\kappa^{d/2}\right), \qquad \kappa > 0$$

where $\lambda > 0$ is a hyperparameter.

Similarly, to control $\mathbb{P}(\rho < \rho_0)$, the upper tail probability of the range ρ below ρ_0, at α, we can set

$$\lambda = -\left(\frac{\rho_0}{\sqrt{8\nu}}\right)^{d/2}\log(\alpha).$$

Thus, combining the PC priors for $\tau|\kappa$ and κ provides the joint PC prior for (κ, τ) which further provides the joint PC prior for (ρ, σ) stated in Theorem 1 by reparameterization.

2.2 Code Implementation

The coding implementation of the simplified main contribution can be found in this Github repository. We thank Fuglstad et al. [2019] for distributing their original coding implementation here, though we were unable to run the code or replicate the simulation study in [Fuglstad et al., 2019] due to errors in the code. We suspect that it was due to the lack of maintenance. As a result, we instead use the method, `krige.bayes()`, from the package, `geoR` [Ribeiro Jr et al., 2007], to conduct the Bayesian analysis on the GRF. Since `krige.bayes()` cannot control the prior distribution on the marginal variance, we treat it as a known parameter in the algorithm. The marginal prior distribution of the range, ρ, can be easily derived from the joint prior in Theorem 1 as

$$
\begin{aligned}
\pi(\rho) &= \int_0^\infty \pi(\sigma,\rho) d\sigma \\
&= \int_0^\infty \frac{d}{2}\tilde{\lambda}_1\tilde{\lambda}_2\rho^{-d/2-1}\exp\left(-\tilde{\lambda}_1\rho^{-d/2}-\tilde{\lambda}_2\sigma\right)d\sigma \\
&= \frac{d}{2}\tilde{\lambda}_1\tilde{\lambda}_2\rho^{-d/2-1}\exp\left(-\tilde{\lambda}_1\rho^{-d/2}\right)\int_0^\infty \exp\left(-\tilde{\lambda}_2\sigma\right)d\sigma \\
&= \frac{d}{2}\tilde{\lambda}_1\tilde{\lambda}_2\rho^{-d/2-1}\exp\left(-\tilde{\lambda}_1\rho^{-d/2}\right)\frac{1}{\tilde{\lambda}_2} \\
&= \frac{d}{2}\tilde{\lambda}_1\rho^{-d/2-1}\exp\left(-\tilde{\lambda}_1\rho^{-d/2}\right)
\end{aligned}
$$

2.3 Connection to the Course Materials

The PC prior framework designed by Fuglstad et al. [2019] has a very strong connection to the course materials on GRF. In multiple homework problems (Homework 1, Problem 1(d); Homework 2, Problem 2(b),2(d); Homework 3, Problem 1), we were asked to fit GRFs to different data. Though we briefly learned the Bayesian estimation framework of GRF in the course, we did not learn the appropriate way to design the prior distribution. As an unfortunate result, most of us went with the Frequentist estimation framework.

Now, equipped with a modern method to construct the prior distribution that penalized the model complexity with an interpretable hyperparameter, we can fit GRFs in those homework problems through the Bayesian approach. It may result in a more principled and interpretable fitted GRF.

3 Simulation Study

We consider a two-dimensional GRF with the spatial design of 100 locations generated uniformly randomly on $[0,1]^2$. The spatial design is shown in Figure 1.

We use the similar exponential covariance function in [Fuglstad et al., 2019] $c(r) = \exp(-r)$ to model the GRF. The covariance function implies that the *true* range, ρ_T, and true standard deviation, σ_T, values are 0.1 and 1. The mean vector is 0. The nugget effect parameter is also 0. In short, this is a relatively simple GRF.

For the PC prior hyperparameters. we fix $\alpha = 0.05$ and use $\rho_0 = 0.025\rho_T, \rho_0 = 0.1\rho_T, \rho_0 = 0.4\rho_T, \rho_0 = 1.6\rho_T$. This covers a prior where ρ_0 is much smaller than the true range, two priors where ρ_0 is smaller than the true range, but not far away, and one prior where ρ_0 is higher than the true range [Fuglstad et al., 2019]. The density curve of $\pi(\rho)$ corresponding to different ρ_0 is shown in Figure 2. As ρ_0 increases, $\pi(\rho)$ shifts to the right and becomes more spread. This is expected as $\mathbb{P}(\rho < \rho_0) = \alpha_1$.

We simulate the GRF 100 times. Each time, we use `likfit()` to extract the maximum likelihood estimator on the range parameter and use `bayes.krige()` to extract the maximum a posteriori (MAP) estimator on the range parameter. The initial range parameter value in the maximum likelihood estimator (MLE), controlled by the argument `ini.cov.pars` in `likfit()`, is set to $0.5\rho_T$ or $1.5\rho_T$ randomly. To compare the performance of two estimators, we compute their empirical 95% confidence interval and their mean squared errors. The results are shown in the Table 1.

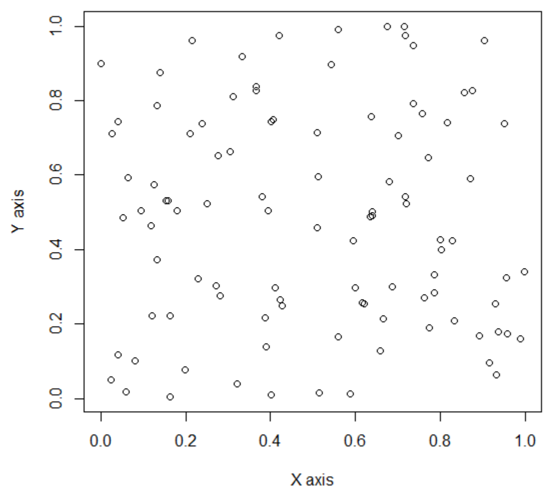

Figure 1: Spatial design for the simulation study.

	Confidence Interval	Confidence Interval Width	Mean Squared Error
MLE	$(0.0564, 0.156)$	0.0997	0.00097155
MAP ($\rho_0 = 0.025\rho_T$)	$(0.0519, 0.1553)$	0.1033	0.00101570
MAP ($\rho_0 = 0.1\rho_T$)	$(0.0548, 0.1595)$	0.1047	0.00104922
MAP ($\rho_0 = 0.4\rho_T$)	$(0.0604, 0.1780)$	0.1175	0.00131938
MAP ($\rho_0 = 1.6\rho_T$)	$(0.0873, 0.2779)$	0.1907	0.00677579

Table 1: 95% empirical confidence interval, the width of 95% empirical confidence interval, and the mean square error of MLE and MAP (with different ρ_0 values).

Comparing MAP with different ρ_0 values, we find a monotonically increasing relationship between ρ_0 and the width of its 95% empirical confidence interval. We also find a monotonically increasing relationship between ρ_0 and the width of the mean square error.

In general, 95% empirical confidence intervals of both MLE and MAP cover the true range parameter $\rho_T = 0.1$. However, MLE outperforms all MAP with different ρ_0 values in terms of both 95% empirical confidence intervals and mean square error. MLE has the most centered 95% empirical confidence interval at ρ_T and has the narrowest 95% empirical confidence interval. Also, MLE has the smallest mean square error. Observations here are interesting and not expected. As ρ_T is known in the Bayesian estimation framework, we expect at least one of the MAPs outperforms the MLE. There are several potential reasons for these surprises. First, the model assumptions of our GRF are not complex enough for the Bayesian estimation framework to stand out. The current mean vector of the GRF is 0. We can make it more complex by setting it as a linear, quadratic, or higher-order function of the location coordinates. Also, the marginal variance is 1, and there is no nugget effect. We can increase the marginal variance and introduce the nugget effect to add more noise to the observed GRF. Second, we treat the marginal variance as a known parameter, making the estimation problem easier. As a result, the Bayesian estimation with PC prior might not be as beneficial as when we need to jointly estimate the marginal variance and range parameters.

4 Discussion

In this report, we introduce an innovative framework, Penalized Complexity Prior, that aims to address the difficulty of selecting appropriate prior distributions for the marginal variance and range parameters in Gaussian Random Fields. This framework not only penalizes the complexity of the GRF

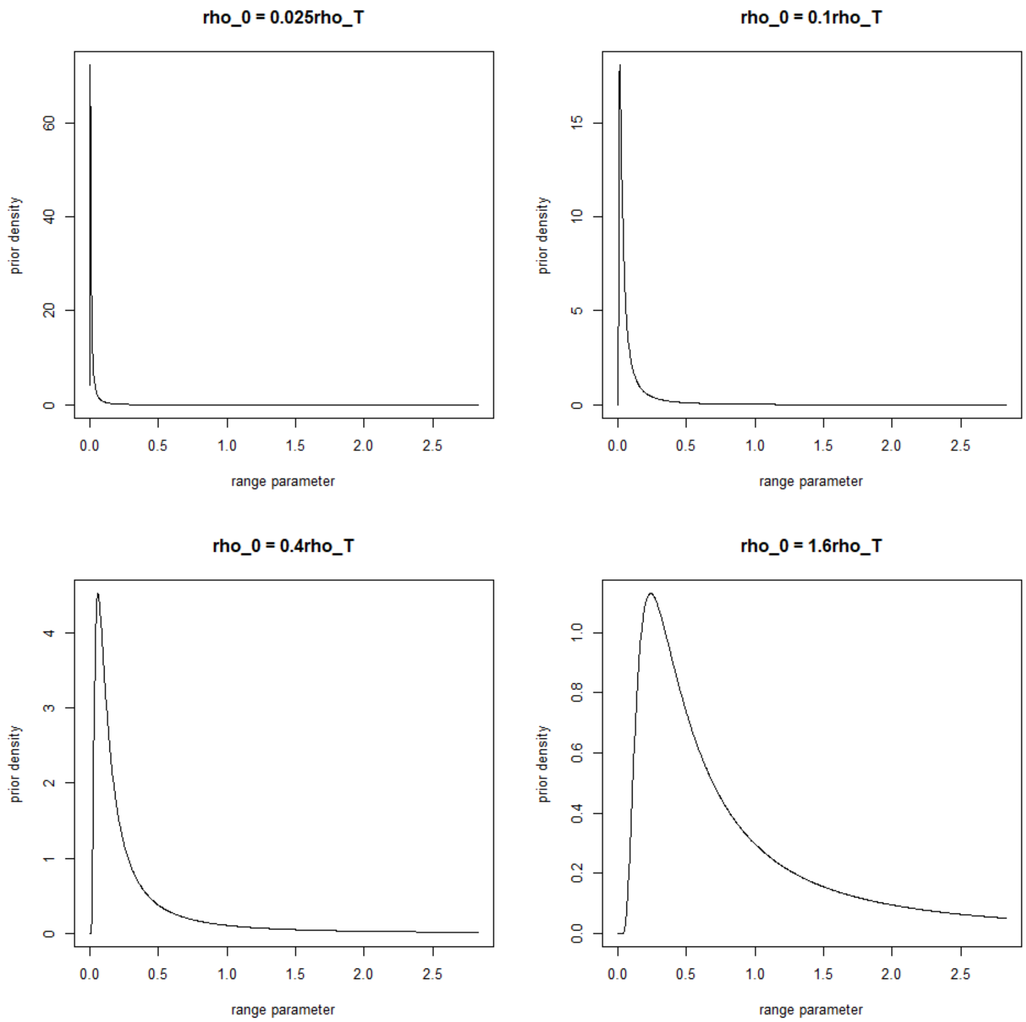

Figure 2: The density curve of the prior distribution of the range parameter corresponding to different ρ_0.

but also provides an interpretable way to select the hyperparameters in the framework. We implement the simplified framework by treating the marginal variance as fixed. We conduct a simulation study that compares the maximum a posteriori estimator with Penalized Complexity Prior and the maximum likelihood estimator in terms of their 95% empirical confidence intervals and mean square errors on a simple Gaussian Random Field. To further and better compare the two estimation approaches, we plan to implement the full framework and use it on more complex Gaussian Random Field in the future.

References

J. O. Berger, V. De Oliveira, and B. Sansó. Objective bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96(456):1361–1374, 2001.

G.-A. Fuglstad, D. Simpson, F. Lindgren, and H. Rue. Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association*, 114(525):445–452, 2019.

N. Klein and T. Kneib. Scale-dependent priors for variance parameters in structured additive distributional regression. 2016.

P. J. Ribeiro Jr, P. J. Diggle, M. P. J. Ribeiro Jr, and M. Suggests. The geor package. *R news*, 1(2): 14–18, 2007.

D. Simpson, H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye. Penalising model component complexity: A principled, practical approach to constructing priors. 2017.

M. L. Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.

J. J. Warnes and B. D. Ripley. Problems with likelihood estimation of covariance functions of spatial gaussian processes. *Biometrika*, 74(3):640–642, 1987.

H. Zhang. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261, 2004.