

Specification of Basis Spacing for Process Convolution Gaussian Process Models

Waley W. J. Liang and Herbert K. H. Lee

University of California, Santa Cruz

November 27, 2019

Abstract

Gaussian process (GP) models have been widely used for statistical modeling of point-referenced data in many scientific applications, including regression, classification, and clustering problems. Standard specification of GP models is computationally inefficient for applications with a large sample size. One solution is to construct the GP by convolving a smoothing kernel with a discretized White noise process, which requires choosing the number of bases. The distance between adjacent bases plays a key role in model accuracy. In this paper, we perform a series of simulations to find a general rule for the basis spacing required for accurate representation of a discrete process convolution GP model. Under certain common conditions, we find that using a basis spacing of one-quarter the practical range of the process works well in practice.

Keywords: Gaussian Processes; Process Convolutions; Spatial Modeling

1 Introduction

A common approach in spatial modeling is to represent the process of interest as a univariate spatial Gaussian process (GP) $\{z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$, which is a collection of random variables indexed by points \mathbf{s} in space \mathbb{R}^d . Any finite collection of these random variables $\{z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)\}$ is distributed as multivariate Gaussian with a certain mean function and covariance matrix Σ . In this paper, we limit our study to isotropic GP where the correlation between two points $z(\mathbf{s}_i)$ and $z(\mathbf{s}_j)$ depends only on their separating Euclidean distance $\|\mathbf{s}_i - \mathbf{s}_j\|$. Under this isotropy assumption,

the covariance matrix Σ can be factored into $\sigma^2\mathbf{C}$, where σ^2 denotes the marginal variance and \mathbf{C} is a correlation matrix whose elements \mathbf{C}_{ij} are pre-computed from a correlation function that depends on $\|\mathbf{s}_i - \mathbf{s}_j\|$ and other parameters. Common correlation functions include the Gaussian, Exponential, Matérn, and the Spherical class. More details on GP and associated correlation functions can be found in Cressie (1991), Stein (1999), Banerjee et al. (2003), and Paciorek and Schervish (2006). Parameter inference under this standard specification often requires inversion of the covariance matrix whose computational complexity grows at $O(n^3)$, where n denotes the data sample size. This makes standard GP impractical for applications with a large n , which is fairly common nowadays. This issue becomes even more undesirable for Bayesian inference with Markov chain Monte Carlo (MCMC) which requires such matrix inversion at each of thousands of MCMC iterations or more. Some methods can ameliorate this issue, such as partitioning the spatial domain and fitting a separate GP in each (Kim et al., 2005; Gramacy and Lee, 2008), or reducing the occurrence of such $O(n^3)$ computations for a GP with a single-parameter correlation function (Yang et al., 2014). Another approach for reducing computational cost due to large n is Discrete Process Convolutions (DPC) (Higdon, 1998, 2002), which formulates the GP $z(\mathbf{s})$ by convolving a symmetric kernel $k(\mathbf{u} - \mathbf{s}; \mathbf{Q})$ with a discretized latent process $x(\mathbf{u})$ indexed at a grid of bases $\{\mathbf{u}_j\}_{j=1}^m$:

$$z(\mathbf{s}) = \sum_{j=1}^m k(\mathbf{u}_j - \mathbf{s}; \mathbf{Q})x(\mathbf{u}_j), \quad \mathbf{s} \in \mathcal{S} \subseteq \mathbb{R}^d.$$

Here, \mathbf{Q}^{-1} denotes the covariance matrix of the kernel, $x(\mathbf{u})$ is usually given a White noise process, and m denotes the number of bases which gives a basis spacing of $1/(m-1)$. Note that DPC arises as an approximation (required in practice) to continuous process convolutions: $z(\mathbf{s}) = \int_{\mathbb{R}^d} k(\mathbf{u} - \mathbf{s}; \mathbf{Q})x(\mathbf{u})d\mathbf{u}$, where the correlation function for z is given by the convolution of the kernel with itself. For instance, when the kernel is a Gaussian density (a common choice in practice), it results in a Gaussian correlation function. We may safely assume that given enough bases, a GP constructed by DPC has a correlation structure that is a close approximation to its counterpart in continuous process convolutions. Figure 1 provides an one-dimensional (1-D) example of a GP constructed by DPC, where the left panel shows a set of six 1-D Gaussian kernels centered at evenly spaced bases on the interval $[-5, 5]$, and the center panel shows the resulting GP obtained as the sum of these kernels. In a two-dimensional (2-D) domain, bases are specified on a rectangular grid

with even spacing in the same dimension, and there can be more bases in one dimension than the other. The right panel of Figure 1 shows an example of bases specified on a 2-D domain. The computational complexity of DPC grows at $O(m^3)$, where m denotes the number of bases. Hence, DPC is computationally efficient for applications in low dimension where m is much smaller than n . For this reason, DPC is useful for many environmental applications such as geology and climatology whose spatial domain is naturally 2-D and data size is often huge due to advance in technology. Given a bounded domain, having a smaller distance between adjacent bases (thus resulting in more bases) allows DPC to better describe local features in the process of interest. Bases are selected by the user before modeling and the optimal basis spacing depends on the application. This paper presents a simulation study to establish a rule-of-thumb for choosing the basis spacing for a GP formulated by DPC. Simulation setup is discussed in detail in Section 2, simulation results are presented in Section 3, and a summary of our research is given in Section 4. In the rest of this paper, the term DPCGP is used to refer to a GP formulated by DPC.

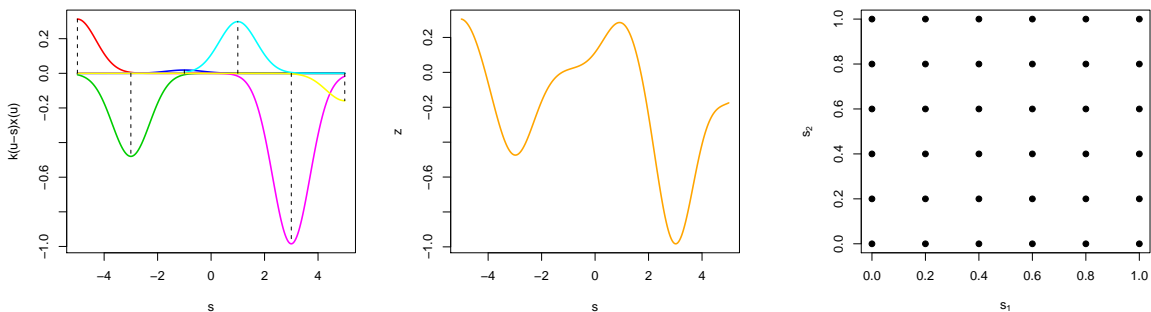


Figure 1: Left panel shows 1-D Gaussian kernels centered at six evenly spaced bases on the interval $[-5, 5]$, and center panel shows resulting GP obtained as the sum of these kernels. Right panel shows an example of bases specified on a 2-D domain.

2 Simulation Setup

As DPCGP is computational efficient only in low dimension, our study focuses on 1-D and 2-D domains. Without loss of generality, 1-D domain is specified as the unit interval $[0, 1]$ and 2-D domain is specified as the unit area $[0, 1]^2$. To simplify the experimental design, the number of

bases is used as the design points instead of basis spacing, because different spacing values may correspond to the same number of bases at slightly different locations which leads to a large number of cases to evaluate. Using the number of bases as design points limits the size of our study and the final results can be re-interpreted to arrive at a rule-of-thumb for basis spacing. In the rest of this paper, more bases is equivalent to smaller basis spacing, and vice versa. Our study develops a series of DPCGP models on simulated data with bases spaced evenly along each dimension, where the number of bases ranges from 5 to 40 in a step size of 1 for each dimension. For example, if the number of bases is 10 in each dimension of the 2-D domain, then there is a total of $10 \times 10 = 100$ bases. Previous experience with DPCGP shows that using more bases (smaller basis spacing) tends to improve model accuracy, but this effect diminishes as the number of bases is beyond a certain threshold. This threshold varies with respect to the dependence range in the unknown process of interest that DPCGP tries to model. A short range means that a sample point is correlated closely with its neighbors and less with points far away, which requires more bases than a longer range. Our study develops each DPCGP model on simulated data under different combinations of correlation functions and practical ranges (PR). As a property of the correlation function, PR is typically defined as the distance from the origin at which the correlation is 0.05. PR can be thought of as the maximum distance between two points having non-negligible correlation. Our goal is to obtain a rule-of-thumb for basis spacing associated with this threshold with respect to PR. Details on data generation is given in Section 2.1, followed by model specification in Section 2.2.

2.1 Data Generation

Data is simulated from three different correlation functions: Gaussian, Matérn with smoothness parameter $\kappa = 4$, and Exponential, which represent real scenarios where the process of interest is very smooth, somewhat smooth, and non-smooth, respectively. Figure 2 shows these three correlation functions along with one that is induced by the Bézier kernel which will be discussed later. Table 1 summarizes the differentiability of the resulting processes from these correlation functions. For each correlation function (except the one induced by Bézier kernel), three PR's: 0.1, 0.2, and 0.3 are considered for response generation. A smaller PR results in a response with more local features, which can be better modeled by DPCGP with finer basis spacing, up to a certain limit. Table 2 shows all combinations of correlation function and PR considered. Two sets

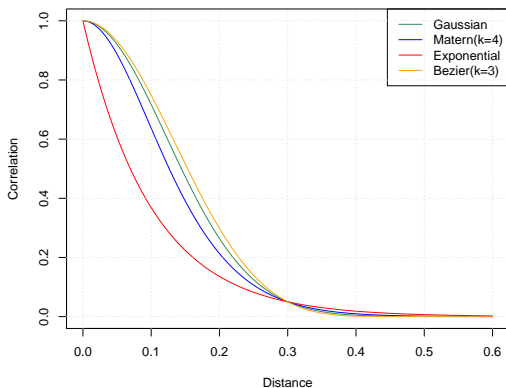


Figure 2: Gaussian, Matérn($\kappa = 4$), Exponential, and Bézier($\kappa = 3$) induced correlation functions. The correlation function induced by a Gaussian kernel is not shown here as it overlaps with the true Gaussian correlation function.

Table 1: Correlation functions and differentiability of their resulting responses.

Correlation Function	Differentiability of Resulting Response
Gaussian	Infinitely mean-square differentiable
Matérn ($\kappa = 4$)	$\lceil \kappa \rceil - 1 = 3$ times mean-square differentiable
Exponential	Not mean-square differentiable
Bézier ($\kappa = 3$) kernel induced	$\lceil \kappa \rceil = 3$ times mean-square differentiable

of response are simulated for each combination: one over the 1-D unit interval $[0, 1]$ and the other over the 2-D unit area $[0, 1]^2$. Each response is a random draw of n samples from a multivariate Gaussian distribution with mean zero and covariance matrix equal to the marginal variance times the correlation matrix determined from the correlation function. Here, we let $n = 1000$ for 1-D and $n = 2000$ for 2-D. In all cases, the marginal variance is given a value of 0.25, which gives a marginal standard deviation (SD) of $\sqrt{0.25} = 0.5$. Figure 3, 4, and 5 display all responses simulated for each combination of correlation function and PR for 1-D (top) and 2-D (bottom). Data is generated by adding zero-mean Gaussian error with $SD = 0.05$ to the response. This level of noise is 10% of the response marginal SD. Only one noise level is considered here to limit the size of our study. Results obtained from this noise level should remain useful for cases with a lower noise level as fewer number of bases is needed for less noisy data in general. Studying higher noise level is less meaningful as model can become inaccurate.

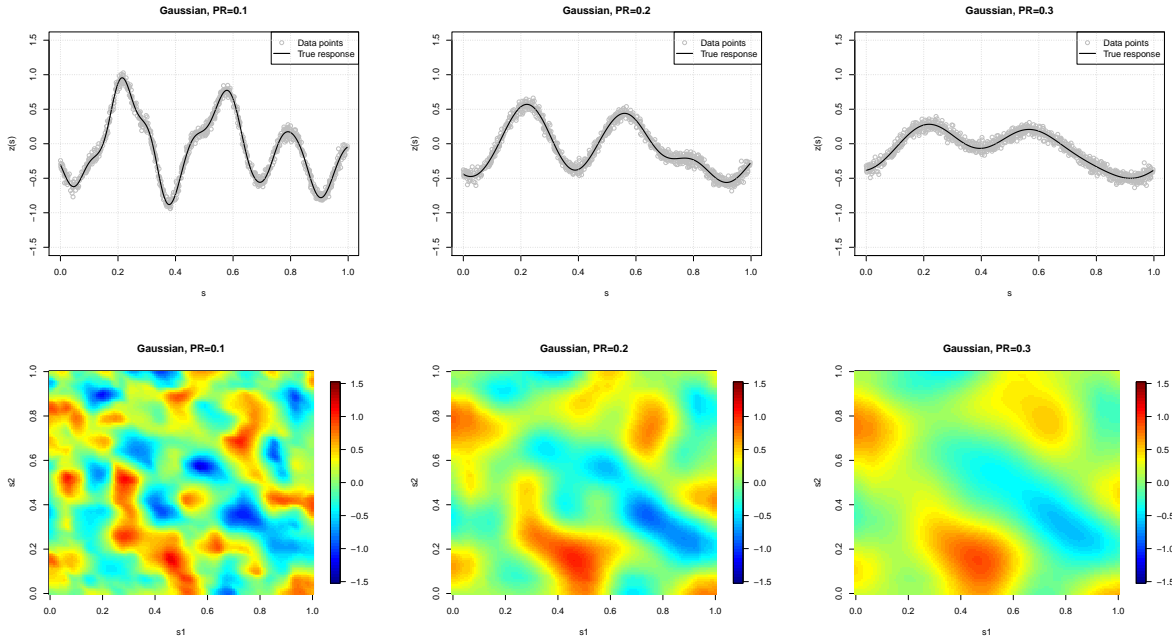


Figure 3: Top: Gaussian response and data on the 1-D unit interval $[0, 1]$. Bottom: Gaussian response on the 2-D unit area $[0, 1]^2$.

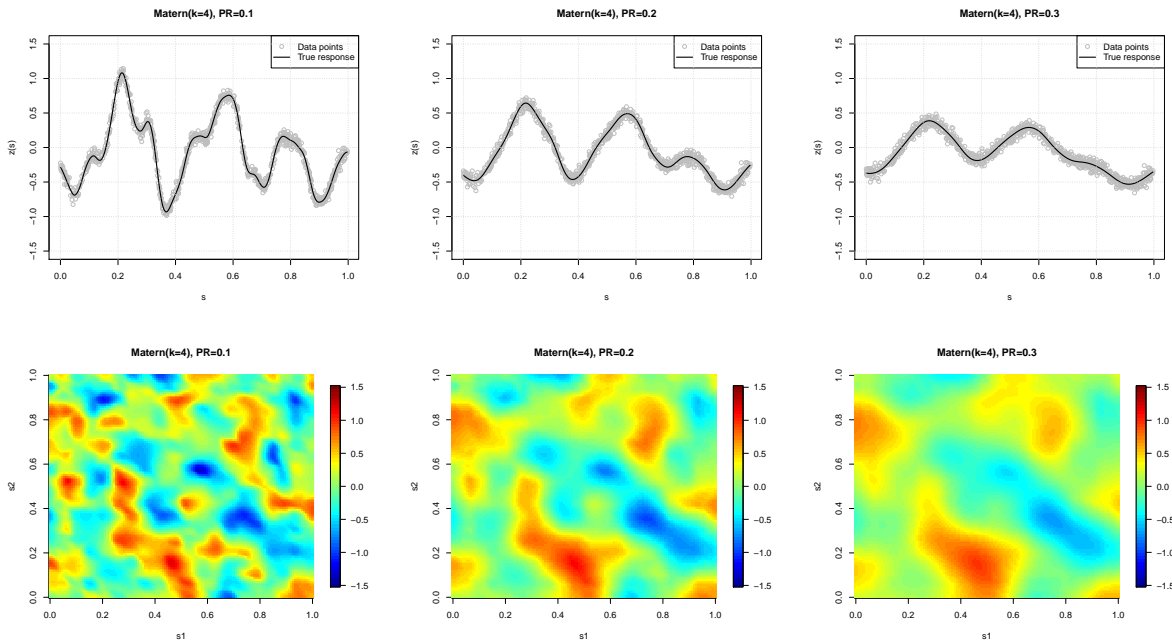


Figure 4: Top: Matérn($\kappa = 4$) response and data on the 1-D unit interval $[0, 1]$. Bottom: Matérn($\kappa = 4$) response on the 2-D unit area $[0, 1]^2$.

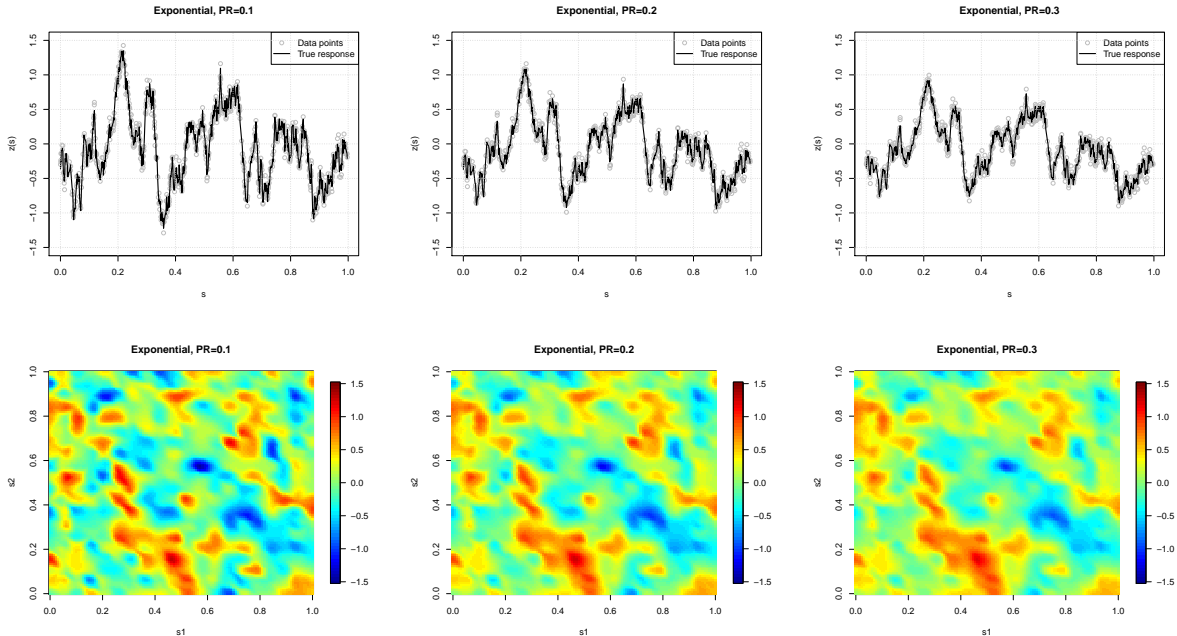


Figure 5: Top: Exponential response and data on the 1-D unit interval $[0, 1]$. Bottom: Exponential response on the 2-D unit area $[0, 1]^2$.

2.2 Model Specification

The correlation structure and smoothness of DPCGP mainly depends on the kernel type. In the literature, the Gaussian density is a common kernel choice which induces a Gaussian correlation structure in continuous process convolutions. The resulting GP is infinitely mean-square differentiable, which is a very smooth process. DPC is an approximation of its continuous counterpart, therefore, the resulting correlation structure is also approximately Gaussian given sufficiently fine basis spacing. When the unknown process of interest is less smooth, it requires a kernel supporting adjustment of smoothness, and one such choice is a compactly supported density (Lemos and Sansó, 2009) defined as follows:

$$k(\mathbf{u} - \mathbf{s}; \mathbf{Q}) = \begin{cases} (1 - D_M^2)^\kappa & \text{if } D_M < 1 \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where $D_M = \sqrt{(\mathbf{u} - \mathbf{s})^\top \mathbf{Q} (\mathbf{u} - \mathbf{s})}$ denotes the Mahalanobis distance with covariance matrix \mathbf{Q}^{-1} . Since our study concerns isotropic processes only, \mathbf{Q}^{-1} is specified as a diagonal matrix with identical diagonal elements. This isotropic form is called the Bézier kernel whose compact support

has a radius equal to the square-root of the diagonal elements of \mathbf{Q}^{-1} . Increasing κ for a Bézier kernel increases the smoothness of the resulting GP. According to Brenning (2001), the resulting GP is $\lfloor \kappa \rfloor$ times mean-square differentiable. An example of the induced correlation function from a Bézier ($\kappa = 3$) kernel is shown in Figure 2. Our study evaluates both the Gaussian and Bézier kernels with a diagonal \mathbf{Q}^{-1} having the same values on the diagonal. For each simulated dataset, a different kernel standard deviation is used such that the kernel induced correlation function has a PR equal to the PR of the correlation function in data generation. This kernel standard deviation is termed practical kernel size (PKS) in the rest of this paper. For example, when PR=0.1, 0.2, and 0.3, the corresponding PKS for a Gaussian kernel is 0.029, 0.058, and 0.087, respectively. Table 2 summaries all configurations in the experimental design. The same configuration is applied to 1-D and 2-D separately. Each dataset is randomly divided into a training set and a validation set of equal size. That is, there are 500 samples for each set in 1-D and 1000 samples for each set in 2-D. Each DPCGP model is developed on the training set and evaluated on the validation set. Training is performed using the *lme()* function from the *nlme* R package, which treats DPCGP as a Linear Mixed-Effects Model. RMSE (root-mean-squared-error) against the true response is computed on the validation set to assess model performance. The number of bases at which RMSE starts to saturate is recorded. Sensitivity to kernel size is analyzed by evaluating two (correlation function, kernel) combinations: (Gaussian, Gaussian) and (Matern($\kappa = 4$), Bézier($\kappa = 3$)) with kernel sizes at 80%, 90%, 100%, 110%, and 120% of PKS.

3 Simulation Results

Simulation results are shown in Figure 6, where each curve corresponds to a specific (correlation function, kernel) combination as indicated in the legend, and it is formed by a series of DPCGP models whose number of bases are shown on the x-axis and the resulting validation RMSE on the y-axis. In the 2-D study (bottom row), number of bases on the x-axis is for a single dimension; total number of bases is the square of this value. The left, center, and right panels correspond to PR of 0.1, 0.2, and 0.3, respectively. In general, validation RMSE starts to saturate when the number of bases reaches a certain threshold. A larger threshold (more bases) is associated with a smaller PR, which is expected since a smaller basis spacing is needed to better describe local features. Within each PR, threshold changes across different (correlation function, kernel)

Table 2: All configurations in the experiment. Same configuration is applied to 1-D and 2-D separately.

Dataset	Correlation Function	Practical Range	DPCGP Kernel	Practical Kernel Size	Number of Bases: m	Basis Spacing: $1/(m - 1)$
1	Gaussian	0.1	Gaussian	0.029	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
	Gaussian	0.1	Bézier ($\kappa = 3$)	0.086	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
2	Gaussian	0.2	Gaussian	0.058	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
	Gaussian	0.2	Bézier ($\kappa = 3$)	0.174	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
3	Gaussian	0.3	Gaussian	0.087	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
	Gaussian	0.3	Bézier ($\kappa = 3$)	0.261	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
4	Matérn ($\kappa = 4$)	0.1	Gaussian	0.029	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
	Matérn ($\kappa = 4$)	0.1	Bézier ($\kappa = 3$)	0.086	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
5	Matérn ($\kappa = 4$)	0.2	Gaussian	0.058	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
	Matérn ($\kappa = 4$)	0.2	Bézier ($\kappa = 3$)	0.174	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
6	Matérn ($\kappa = 4$)	0.3	Gaussian	0.087	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
	Matérn ($\kappa = 4$)	0.3	Bézier ($\kappa = 3$)	0.261	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
7	Exponential	0.1	Gaussian	0.029	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
	Exponential	0.1	Bézier ($\kappa = 3$)	0.086	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
8	Exponential	0.2	Gaussian	0.058	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
	Exponential	0.2	Bézier ($\kappa = 3$)	0.174	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
9	Exponential	0.3	Gaussian	0.087	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256
	Exponential	0.3	Bézier ($\kappa = 3$)	0.261	5, 6, ..., 39, 40	0.25, 0.2, ..., 0.0263, 0.0256

combinations. In general, fewer bases is needed when the kernel induced correlation function is the same or close to that of the data, e.g., (Gaussian, Gaussian) and (Gaussian, Bézier ($\kappa = 3$)). More bases are needed when the kernel induced and data correlation functions are moderately different, e.g., (Matern ($\kappa = 4$), Gaussian) and (Matern ($\kappa = 4$), Bézier ($\kappa = 3$)). Finally, when the difference in correlation structure is large, e.g., (Exponential, Gaussian) and (Exponential, Bézier ($\kappa = 3$)), the model completely misses the true response as indicated by the large validation RMSE which is much larger than the noise SD. This case is excluded from further analysis since the associated models are inaccurate. Table 3 summarizes the near-optimal number of bases and the corresponding basis spacing under each scenario. These numbers are based on visual inspection from Figure 6, which may or not may be the exact optimal, but close enough for us to establish a rule-of-thumb for choosing the basis spacing. Results are roughly the same between 1-D and 2-D except for PR = 0.1, where DPCGP has good 1-D model fit but struggles to fit the 2-D true response resulting in validation RMSE being slightly higher than noise SD. Figure 7 and 8 illustrate the effect of kernel size on model performance for two cases: (Gaussian, Gaussian) and (Matern ($\kappa = 4$), Bézier ($\kappa = 3$)), which represent matching and moderately different correlation structure, respectively. Here, kernel size is varied at 80%, 90%, 100%, 110%, and 120% of PKS. These results show that kernel size has limited effect on model performance given enough bases (small enough

basis spacing), and the saturation point in each case roughly stays the same. In practice, DPCGP is mostly used for 2-D applications where the kernel induced correlation function is unlikely to be a perfect match to the true correlation structure. A rule-of-thumb for choosing the basis spacing should be robust enough to work in practice. We consider establishing the rule-of-thumb based on the 2-D case where the correlation function is Matern ($\kappa = 4$) and the kernel is Gaussian or Bézier ($\kappa = 3$) due to reasons described above. According to Table 3, the near-optimal basis spacing in this case is 0.0345 for $PR = 0.1$, 0.0526 for $PR = 0.2$, and 0.0714 for $PR = 0.3$. The relationship between basis spacing and PR appears to be linear: $\text{basis spacing} = \alpha \times PR$ for $PR > 0.1$, where the least squares estimate of α is found to be around 0.2528. Hence, a reasonable rule-of-thumb for basis spacing can be established as $PR/4$ for $PR \geq 0.1$. Note that $PR < 0.1$ is excluded because DPCGP tends to be unreliable in this region for 2-D. Given this rule-of-thumb, the general steps for setting up a DPCGP model proceed as follows:

1. Scale the spatial domain to $[0, 1]$ or $[0, 1]^2$ and estimate the dependence range from data.
2. Select a kernel whose induced correlation function best describes the process of interest.
3. Specify a kernel size such that the PR of the kernel induced correlation function equals to the estimated dependence range in step 1.
4. Calculate basis spacing as: estimated dependence range / 4, and obtain the corresponding number of bases.

The estimated dependence range may be inaccurate and lead to an over/underestimated kernel size. This is acceptable provided that inaccuracy is not too large, because sensitivity analysis shows that model performance is not very sensitive to kernel size. Overestimation of dependence range can lead to an overestimated basis spacing. This issue can be alleviated by further reducing the basis spacing given by the rule-of-thumb as appropriate.

4 Summary

Research presented in this paper aims to obtain a rule-of-thumb for choosing the basis spacing for DPCGP models. A series of experiments is performed on simulated data based on Gaussian, Matérn ($\kappa = 4$), and Exponential correlation functions. Three different PR 's (0.1, 0.2, and 0.3)

Table 3: Near-optimal number of bases and the corresponding basis spacing for each experiment.

Dimension	Correlation Function	DPCGP Kernel	Near-Optimal (Number of Bases, Spacing)		
			PR=0.1	PR=0.2	PR=0.3
1-D	Gaussian	Gaussian	(30, 0.0345)	(15, 0.0714)	(10, 0.1111)
	Gaussian	Bézier ($\kappa = 3$)	(30, 0.0345)	(15, 0.0714)	(10, 0.1111)
	Matérn ($\kappa = 4$)	Gaussian	(30, 0.0345)	(20, 0.0526)	(15, 0.0714)
	Matérn ($\kappa = 4$)	Bézier ($\kappa = 3$)	(35, 0.0294)	(20, 0.0526)	(15, 0.0714)
2-D	Gaussian	Gaussian	(25, 0.0417)	(15, 0.0714)	(10, 0.1111)
	Gaussian	Bézier ($\kappa = 3$)	(25, 0.0417)	(15, 0.0714)	(10, 0.1111)
	Matérn ($\kappa = 4$)	Gaussian	(30, 0.0345)	(20, 0.0526)	(15, 0.0714)
	Matérn ($\kappa = 4$)	Bézier ($\kappa = 3$)	(30, 0.0345)	(20, 0.0526)	(15, 0.0714)

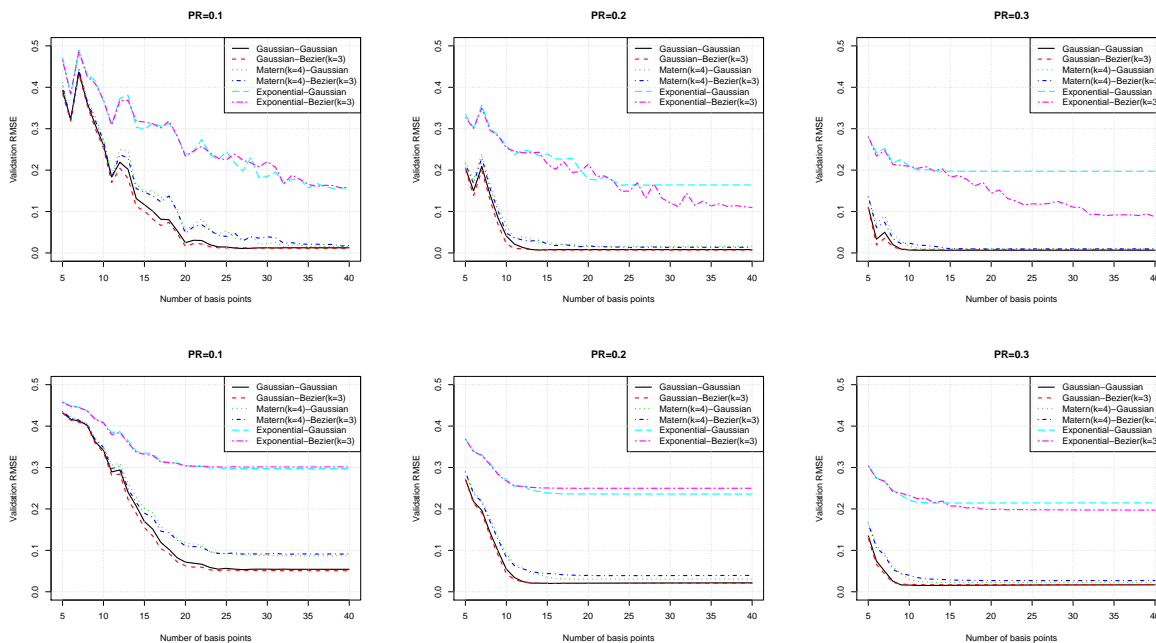


Figure 6: Validation RMSE vs. number of bases for each (correlation function, kernel) combination. Top row shows results for 1-D and bottom row shows results for 2-D.

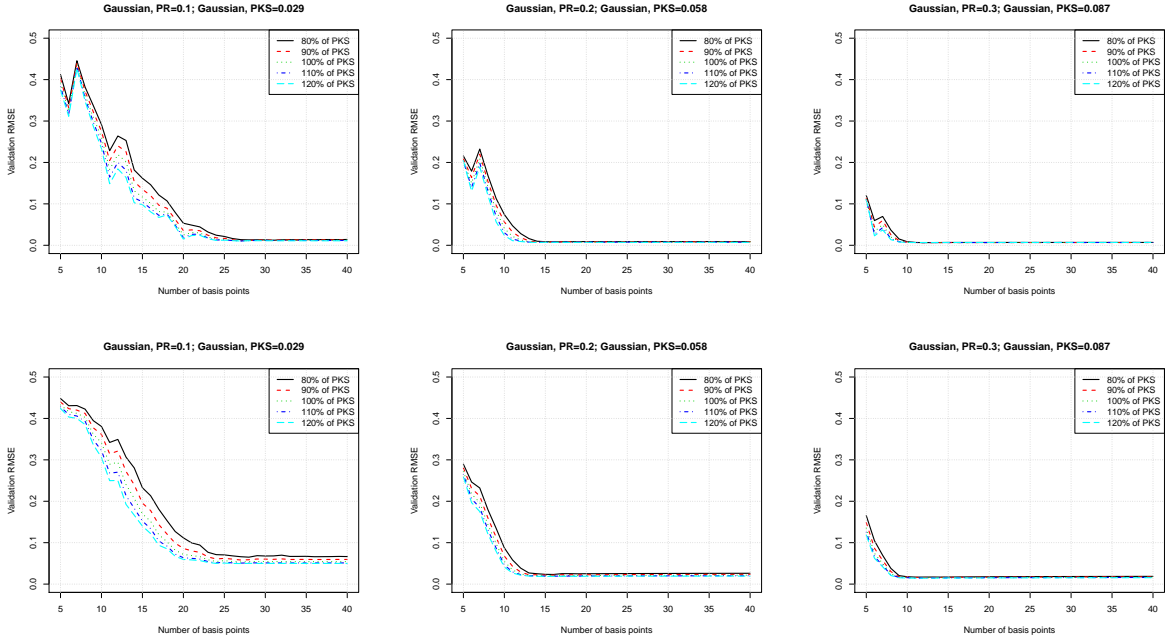


Figure 7: Sensitivity to kernel size under Gaussian correlation function and Gaussian kernel (1-D at the top and 2-D at the bottom). Kernel SD is varied at 80%, 90%, 100%, 110%, and 120% of PKS.

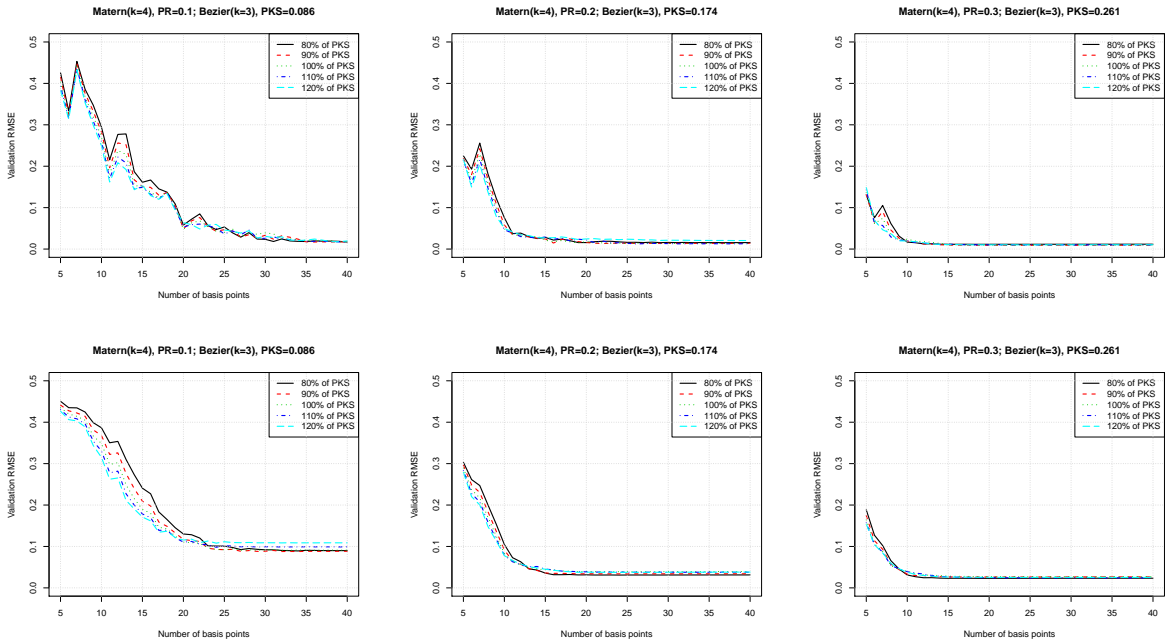


Figure 8: Sensitivity to kernel size under Matérn ($\kappa = 4$) correlation function and Bézier ($\kappa = 3$) kernel (1-D at the top and 2-D at the bottom). Kernel SD is varied at 80%, 90%, 100%, 110%, and 120% of PKS.

are evaluated for each case. DPCGP models under the Gaussian and Bézier ($\kappa = 3$) kernels are developed on the training data using different number of bases ranging from 5 to 40. In each case, the kernel size is specified to match the dependence range of the resulting GP to the PR in data generation. Model performance is assessed via RMSE on validation data, and the near-optimal basis spacing is obtained for each PR. A rule-of-thumb for basis spacing is established as PR/4.

References

- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2003), *Hierarchical modeling and analysis for spatial Data*, Chapman & Hall, Boca Raton, FL.
- Brenning, A. (2001), “Geostatistics without stationarity assumptions within geographical information systems,” *Freiberg Online Geoscience*, 6, 1–108.
- Cressie, N. (1991), *Statistics for spatial data*, Wiley, New York.
- Gramacy, R. B. and Lee, H. K. (2008), “Bayesian treed Gaussian process models with an application to computer modeling,” *Journal of the American Statistical Association*, 103(483), 1119–1130.
- Higdon, D. (1998), “A process-convolution approach to modeling temperatures in the North Atlantic Ocean,” *Journal of Environmental and Ecological Statistics*, 5(2), 173–190.
- Higdon, D. (2002), “Space and space-time modeling using process convolutions,” in *Quantitative methods for current environmental issues*, eds. C. Anderson, V. Barnett, P. Chatwin, and A. El-Shaarawi, pp. 37–54, London, Springer.
- Kim, H.-M., Mallick, B. K., and Holmes, C. C. (2005), “Analyzing nonstationary spatial data using piecewise Gaussian processes,” *Journal of the American Statistical Association*, 100, 653–668.
- Lemos, R. T. and Sansó, B. (2009), “Spatio-temporal model for mean, anomaly and trend fields of north Atlantic sea surface temperature,” *Journal of the American Statistical Association*, 104, 5–18.
- Paciorek, C. and Schervish, M. J. (2006), “Spatial modelling using a new class of nonstationary covariance functions,” *Environmetrics*, 17, 483–506.

Stein, M. L. (1999), *Interpolation of spatial data*, Springer, New York, NY.

Yang, H., Liu, F., Ji, C., and Dunson, D. (2014), “Adaptive sampling for Bayesian geospatial models,” *Statistical Computing*, 24, 1101–1110.