

LOSS FUNCTION APPROACHES TO PREDICT A SPATIAL QUANTILE AND ITS EXCEEDANCE REGION

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Jian Zhang, Ph.D

* * * * *

The Ohio State University

2007

Dissertation Committee:

Peter F. Craigmile, Co-Adviser

Noel Cressie, Co-Adviser

Tailen Hsing

Approved by

Co-Adviser

Co-Adviser

Graduate Program in
Statistics

ABSTRACT

An important problem in spatial statistics is to predict a spatial quantile and its associated exceedance region. This has applications in environmental sciences, natural resources, and agriculture, since unusual events tend to have a strong impact on the environment. In this dissertation, we first review loss-function approaches to quantify exceedances. We then develop a method for the prediction of the spatial exceedance region involving a class of loss functions based on image metrics. We give special attention to Baddeley's loss function, for which we calibrate the choice of a tuning parameter. We then propose a joint-loss approach for the prediction of both a spatial quantile and its associated exceedance region. The optimal predictor is obtained by minimizing the posterior expected loss, given the spatial-trend, noise, and spatial-covariance parameters. In practice, the parameters are estimated and the minimization involves simulated annealing. We compare various predictors' performances through a simulation and apply our methodology to a spatial dataset of decadal temperature change over the Americas.

ACKNOWLEDGMENTS

I would like to thank my co-advisers, Professor Peter F. Craigmile and Professor Noel Cressie, for their guidance and support while working on this research. I also want to thank my wife, Xiang Ling, for her love and support, and my family back in China, for always being there.

VITA

November 17, 1975	Born, ShaoXing, China
July, 1997	B.S. in Environmental Science, Peking University, Beijing, China
July, 2000	M.S. in Environmental Science, Peking University, Beijing, China
2001 - 2003	Graduate research associate, GIS help-desk and consultant, Center for Mapping, The Ohio State University, Columbus, Ohio
March, 2003	M.S. in Statistics, The Ohio State University, Columbus, Ohio
2004 - present	Graduate teaching and research associate, Department of Statistics, The Ohio State University, Columbus, Ohio

PUBLICATIONS

Research Publications

N. Cressie, J. Zhang, and P. F. Craigmiller (2005). Geostatistical prediction of spatial extremes and their extent. In *Geostatistics for Environmental Applications, Proceedings of the Fifth European Conference on Geostatistics for Environmental Applications*, eds P. Renard, H. Demougeot-Renard, and R. Froidevaux. Dordrecht, Kluwer, 27-37.

N. Cressie, B.E. Buxton, C.A. Calder, P.F. Craigmiller, C. Dong, N.J. McMillan, M. Morara, T.J. Santner, K. Wang, G. Yong, and J. Zhang (2007). From Sources to Biomarkers: A Hierarchical Bayesian Approach for Human Exposure Modeling. *Journal of Statistical Planning and Inference*, in press.

FIELDS OF STUDY

Major Field: Statistics

TABLE OF CONTENTS

	Page
Abstract	ii
Acknowledgments	iii
Vita	iv
List of Tables	ix
List of Figures	x
Chapters:	
1. Introduction	1
1.1 Motivation	1
1.2 Model and parameter estimation	2
1.3 Optimal prediction of spatial exceedances and spatial exceedance regions	5
1.3.1 Loss functions for spatial exceedances	6
1.3.2 Other approaches for inferences on exceedances	8
1.4 Outline of dissertation	20
2. Loss function on exceedances	22
2.1 Introduction	22
2.2 Squared error loss	25
2.2.1 Loss function for kriging predictor	25
2.2.2 Loss function for covariance matching constrained kriging .	26
2.3 Integrated weighted quantile squared error loss	29
2.3.1 Properties	31

2.4	Simulation study	34
2.5	Discussion	42
3.	Prediction of the exceedance region above one threshold	45
3.1	Introduction	45
3.2	Image metrics and the resulting predictor	47
3.3	Baddeley's loss function and the resulting predictor	50
3.4	Other approaches to detect hotspots	52
3.4.1	Computation of the predictor via simulated annealing	54
3.5	Simulation study	57
3.5.1	Calibration of the taper t	58
3.5.2	Performance comparison	61
3.6	Discussion	67
4.	Joint prediction of a spatial quantile and its exceedance region	69
4.1	Joint loss functions and the resulting predictor	71
4.2	Choice of loss functions and the resulting predictor	72
4.3	Simulation studies	76
4.3.1	Calibration of λ	77
4.3.2	Comparison of optimal predictor with plug-in predictor	81
4.4	Discussion	84
5.	An application to decadal temperature change over the Americas	90
5.1	Spatial model and parameter estimates	91
5.2	Prediction of exceedance regions above fixed thresholds	94
5.3	Prediction of extremely positive changes in temperature and their exceedance regions	98
5.4	Discussion	101
6.	Future research	103
6.1	Loss functions for the prediction of multiple exceedance regions	104
6.1.1	Metrics for grayscale images as loss functions	105
6.1.2	An optimal Bayes predictor for multiple exceedance regions	110
6.1.3	Iterative approach for the prediction of multiple exceedance regions based on binary image metrics	111
6.2	Other future research	111

Appendices:

A.	Asymptotic calculations of Baddeley's loss function in 1-D and 2-D cases	113
A.1	Baddeley's loss function in the 1-D case, for small tapers.	114
A.2	Baddeley's loss function in the 2-D case, for small tapers.	121
	Bibliography	127

LIST OF TABLES

Table		Page
2.1	Levels of factors in simulation studies for the single-loss approach. . .	36
2.2	The choice of λ (to the nearest integer) in IWQSEL for minimizing the bias in estimating quantile α	39
3.1	Levels of factors in simulation studies for the single-loss approach . .	59
3.2	The choice of taper, t , in the Baddeley's loss function based on the relative efficiency, with the grid size, $\omega = 1$, and the maximum pairwise distance, $\xi = \sqrt{2}$	64
4.1	Levels of factors used in the simulation studies	78
4.2	The best choice of λ in WQSEL, based on minimizing the absolute bias in estimating the α -quantile of the SCDF of interest.	81

LIST OF FIGURES

Figure		Page
2.1	The ACDF, $F_D(y)$, versus the SCDF, $S_D(y; Y(\cdot))$. The solid curve is the ACDF, and the dashed curves are the SCDFs for two different realizations of the $Y(\cdot)$ process.	33
2.2	The bias and the RMSE of $S_D(q(\alpha); \hat{Y}(\cdot))$ from IWQSEL-based predictor versus λ for SNR = 2; Fig 2.2(a) shows results for $r = 0.5$, Figure 2.2(b) shows results for the $r = 1.5$, and Figure 2.2(c) shows results for $r = 5$. Dashed lines show the corresponding results for CMCK-based predictor.	40
2.3	The bias and the RMSE of $S_D(q(\alpha); \hat{Y}(\cdot))$ from IWQSEL-based predictor versus λ for SNR = 0.5; Fig 2.2(a) shows results for $r = 0.5$, Figure 2.2(b) shows results for $r = 1.5$, and Figure 2.2(c) shows results for $r = 5$. Dashed lines show the corresponding results for CMCK-based predictor.	41
2.4	The proportion of a predicted process below the inverse ACDF of the process; data were obtained from a single realization with $r = 0.5$, and SNR= 2. The quantile $\alpha = 0.9$ (and hence $\lambda = 6$) was chosen. The vertical solid line shows the target quantile 0.9. The horizontal solid line shows the 0.9-empirical quantile using the proportion of a predicted process below the target quantile of the ACDF. The 45° solid line represents the unbiased prediction of IWQSEL predictor for quantiles.	43
3.1	The relative efficiency of the Baddeley-loss-based predictor and the posterior-mean-based predictor, for different tapers, range parameters, r , and values of α (used to set the threshold), for SNR = 2.0. The solid line is for the sample-averaged relative efficiency. The dashed line is for the 95% confidence interval for the sample-averaged relative efficiency.	62

3.2	The relative efficiency of the Baddeley-loss-based predictor and the posterior-mean-based predictor, for different tapers, range parameters, r , and values of α (used to set the threshold), for $\text{SNR} = 0.5$. The solid line is for the sample-averaged relative efficiency. The dashed line is for the 95% confidence interval for the sample-averaged relative efficiency.	63
3.3	Boxplots of difference of the sensitivity and specificity between the Baddeley-loss-based predictor and the posterior-mean-based predictor, when the grid size is $M = 20$.	66
3.4	Boxplots of difference of the sensitivity and specificity between the Baddeley-loss-based predictor and the posterior-mean-based predictor, for a grid size of $M = 40$.	68
4.1	The bias in estimating the α -quantiles ($\alpha = 0.5, 0.75, 0.9$) of the SCDF using the WQSEL, versus the calibration quantity λ .	80
4.2	Boxplots of difference of bias in estimating α -quantiles using the joint-loss-based approach and the posterior-mean-based plug-in approach, and boxplot of difference of SSB values. The 20 values in each boxplot are generated by the 20 realizations of the $Y(\cdot)$ process. The signal-to-noise ratio, SNR , is 0.5 for this experiment.	85
4.3	Boxplots of difference of the sensitivity (Se), the specificity (Sp), and the average of Se and Sp with the α -quantile. The 20 values in each boxplot are generated by the 20 realizations of the $Y(\cdot)$ process. The signal-to-noise ratio, SNR , is 0.5 for this experiment.	86
4.4	Boxplots of difference of bias in estimating α -quantiles using the joint-loss-based approach and the posterior-mean-based plug-in approach, and boxplot of difference of SSB values. The 20 values in each boxplot are generated by the 20 realizations of the $Y(\cdot)$ process. The signal-to-noise ratio, SNR , is 2 for this experiment.	87
4.5	Boxplots of difference of the sensitivity (Se), the specificity (Sp), and the average of Se and Sp with the α -quantile. The 20 values in each boxplot are generated by the 20 realizations of the $Y(\cdot)$ process. The signal-to-noise ratio, SNR , is 2 for this experiment.	88

5.1	Spatial analysis of the average temperature change from the 1980s to the 1990s on the Celsius scale: (a) grayscale map of the temperature change over the prediction region D^* with observation sites B shown as crosses, (b) spatial trend of temperature change versus latitude, (c) spatial trend of temperature change versus longitude, (d) histogram of temperature change.	92
5.2	Spatial analysis of the detrended residuals of the observed averaged-temperature change in B from the 1980s to the 1990s on the Celsius scale : (a) spatial trend of detrended temperature change versus latitude, (b) spatial trend of detrended temperature change versus longitude, (c) histogram of detrended temperature change, (d) empirical and fitted semivariograms: circles for the empirical semivariogram, solid line for the fitted exponential model, and dashed line for the fitted spherical variogram.	95
5.3	Prediction of exceedance regions with a change in mean temperature above K , for $K = 0.05, 0.10, 0.15, 0.20^\circ\text{C}$. The left column shows the Baddeley-loss-function-based predictor and the right column shows the posterior-mean-based plug-in predictors.	97
5.4	Prediction of spatial regions with extremely positive mean-temperature change (0.9-quantile of the SCDF), as shown by the gray shading: (a) the joint-loss-based predictor, (b) the posterior-mean-based plug-in predictor, (c) the plug-in predictor from all the data over D^*	100
5.5	Prediction of spatial regions with mean-temperature change above the 0.75-quantile of the SCDF, as shown by the gray shading: (a) the joint-loss-based predictor, (b) the posterior-mean-based plug-in predictor, (c) the plug-in predictor from all the data over D^*	100

CHAPTER 1

INTRODUCTION

1.1 Motivation

An *exceedance region* is the set of all spatial locations where the process of interest has values larger than some threshold. Scientists pay close attention to the locations of exceedance events where they are may help answer why they occur Environmentalists are interested in how severe pollution is and where it is. Epidemiologists wish to know the location of large or small disease incidence rates. Meteorologists track strong winds and where the affected areas are. The adhoc, but usual way, to find the exceedances and their locations is to make predictions of the process of interest using spatial methods, and then plug in those predictions to obtain the exceedance region. The problems with plug-in methods are twofold: first, spatial methods such as kriging are known to smooth out the peaks and fill in the valleys of the distribution of the process (e.g., Louis, 1984), so that the prediction for the exceedance region above some high thresholds based on the predicted process is biased; second, optimality of a predicted surface is not preserved under nonlinear functionals of the surface, such as the exceedance region.

This dissertation uses a series of loss functions to tackle the problem of prediction of the spatial exceedance and the corresponding exceedance region. The first loss function of interest is the integrated weighted quantile squared error loss (IWQSEL) (Craigmile et al., 2006), which attempts to avoid the underdispersion of exceedances that kriging suffers from. Until now, the prediction of the exceedance region in the IWQSEL approach has been plug-in based. An important contribution of this dissertation is to introduce a loss function that depends directly on the exceedance region, which yields an optimal predictor. We feature Baddeley loss function in our research. The third loss function we consider combines both the integrated weighted quantile squared error loss (IWQSEL) for quantifying spatial quantiles of interest, and Baddeley's loss function for predicting the exceedance region. The optimal predictor based on this joint loss will predict the exceedances and locate the associated exceedance region simultaneously.

1.2 Model and parameter estimation

Suppose that $Y(\cdot) \equiv \{Y(\mathbf{s}) : \mathbf{s} \in D\}$ is a hidden geostatistical process of interest, defined in a domain $D \subset \mathbb{R}^d$ such that it has a positive d-dimensional volume, $|D|$. Let $Z(\cdot) \equiv \{Z(\mathbf{s}) : \mathbf{s} \in D\}$ be the process of observed and potentially observed values. Then the hierarchical model can be described by the following data model and process model.

The data model:

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \epsilon(\mathbf{s}); \quad \mathbf{s} \in D, \quad (1.1)$$

where $\epsilon(\cdot)$ is a zero-mean white noise process with measurement-error variance, $\sigma_\epsilon^2 > 0$. We assume that $\epsilon(\cdot)$ is independent of $Y(\cdot)$.

The process model:

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}); \quad \mathbf{s} \in D. \quad (1.2)$$

Here the function $\mu(\cdot)$ captures the large-scale, deterministic mean structure (*spatial trend*) of $Y(\cdot)$, and $\delta(\cdot)$ is a stationary stochastic process that models the small-scale spatial structure in $Y(\cdot)$. We assume that the spatial trend, $\mu(\cdot)$, follows a linear model given by

$$\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta}, \quad (1.3)$$

where the vector $\mathbf{x}(\mathbf{s}) \equiv (x_1(\mathbf{s}), \dots, x_p(\mathbf{s}))^T$ is a collection of covariates at \mathbf{s} , and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the *spatial-trend parameter*. We also assume that $\delta(\cdot)$ is a zero-mean, second-order stationary spatial process with a parametric covariance function, $C_{\boldsymbol{\theta}_Y}(\cdot)$. That is,

$$E[\delta(\mathbf{s})] = 0, \quad \text{and} \quad \text{cov}[\delta(\mathbf{s}), \delta(\mathbf{s}')] \equiv C_{\boldsymbol{\theta}_Y}(\mathbf{s} - \mathbf{s}'); \quad \mathbf{s}, \mathbf{s}' \in D, \quad (1.4)$$

where $\boldsymbol{\theta}_Y$ is the *spatial-covariance parameter* that at least consists of the *micro-scale variation* τ^2 , the *partial sill* σ^2 , and the *range parameter* r . Here τ^2 is assumed to be zero because it is impossible to estimate the micro-scale variation and noise variation separately from spatial data that have one observation per location. Henceforth, $\boldsymbol{\theta}_Y \equiv (\sigma^2, r)^T$.

Let e_K^Y denote the *exceedance region* of $Y(\cdot)$ above a threshold K , defined by

$$e_K^Y \equiv \{\mathbf{s} \in D : I(Y(\mathbf{s}) > K) = 1\}. \quad (1.5)$$

The process e_K^Y inherits its randomness from the process $Y(\cdot)$. Suppose that we have n observations, $\mathbf{Z} = \{Z(\mathbf{s}) : \mathbf{s} \in B\}$, taken at a collection of n locations in $B \equiv \{\mathbf{s}_1, \dots, \mathbf{s}_n\} \subseteq D$. Then, our interest is to predict the exceedance region, e_K^Y ,

which is a nonlinear functional of $Y(\cdot)$, based on the observations, \mathbf{Z} . This problem is formulated in terms of the hierarchical model described above, where we choose to estimate unknown parameters (an empirical Bayes approach). That is, we look for the optimal Bayes predictor for the exceedance region based on a given loss function, where the parameters for noise variance, spatial covariance, and spatial trend are estimated and plugged in.

Let $\boldsymbol{\phi} \equiv (\boldsymbol{\beta}^T, \sigma_\epsilon^2, \boldsymbol{\theta}_Y^T)^T$ be the parameters that characterize the hierarchical model in (1.1) and (1.2). If the noise variance, σ_ϵ^2 , and the spatial parameter, $\boldsymbol{\theta}_Y$, are known, then the *best linear unbiased estimator* (BLUE) of the spatial trend, $\boldsymbol{\beta}$, is

$$\tilde{\boldsymbol{\beta}} \equiv (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z}, \quad (1.6)$$

where $\mathbf{X} \equiv (\mathbf{x}(\mathbf{s}_0), \dots, \mathbf{x}(\mathbf{s}_n))^T$ is a $n \times p$ design matrix for the observation sites, and $\boldsymbol{\Sigma} \equiv \text{var}[\mathbf{Z}] = \text{var}[(Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^T] + \sigma_\epsilon^2 \mathbf{I}$ is the covariance matrix of the observations. Let \mathbf{h} denote the distance matrix between the observation sites, where the (i, j) -element $h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$, $i, j \in \{1, \dots, n\}$. Then $\text{var}[(Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^T] \equiv \mathbf{C}_{\boldsymbol{\theta}_Y}(\mathbf{h})$, and hence $\boldsymbol{\Sigma} = \mathbf{C}_{\boldsymbol{\theta}_Y}(\mathbf{h}) + \sigma_\epsilon^2 \mathbf{I}$, and is parameterized by $\boldsymbol{\theta}_Z \equiv (\sigma_\epsilon^2, \boldsymbol{\theta}_Y^T)^T$.

There are several estimators available for $\boldsymbol{\theta}_Z$, such as the *maximum likelihood estimator* (MLE), the *restricted maximum likelihood* (REML) estimator, and the *weighted least squares* (WLS) estimator. In this dissertation, we estimate $\boldsymbol{\theta}_Z$ through the WLS estimator based on the empirical *semivariogram* of the data (Cressie, 1993b, pp. 96-98). The semivariogram is a commonly used measure of spatial dependence. Let $\gamma_Z(\mathbf{s}, \mathbf{s}')$ denote the semivariogram of the data between $Z(\mathbf{s})$ and $Z(\mathbf{s}')$. Then $2\gamma_Z(\mathbf{s}, \mathbf{s}')$ is the *variogram* of the data, defined by

$$2\gamma_Z(\mathbf{s}, \mathbf{s}') \equiv \text{var}[Z(\mathbf{s}) - Z(\mathbf{s}')] = \text{var}(Z(\mathbf{s})) + \text{var}(Z(\mathbf{s}')) - 2\text{cov}[Z(\mathbf{s}), Z(\mathbf{s}')], \quad (1.7)$$

for sites \mathbf{s} and \mathbf{s}' in D . The spatial covariance function for $Z(\cdot)$ defined by (1.1) is

$$\text{cov}[Z(\mathbf{s}), Z(\mathbf{s}')] \equiv \begin{cases} \sigma_\epsilon^2 + \sigma^2, & \mathbf{s} = \mathbf{s}'; \\ C_{\boldsymbol{\theta}_Y}(\mathbf{s} - \mathbf{s}'), & \mathbf{s} \neq \mathbf{s}', \end{cases} \quad (1.8)$$

where $\sigma^2 \equiv C_{\boldsymbol{\theta}_Y}(0)$. Hence, the variogram of the data in this case is given by

$$2\gamma_Z(\mathbf{s}, \mathbf{s}') = \begin{cases} 0, & \mathbf{s} = \mathbf{s}'; \\ 2(\sigma_\epsilon^2 + \sigma^2 - C_{\boldsymbol{\theta}_Y}(\mathbf{s} - \mathbf{s}')), & \mathbf{s} \neq \mathbf{s}'. \end{cases} \quad (1.9)$$

We see that the semivariogram of the data, $\gamma_Z(\mathbf{s}, \mathbf{s}')$, is only a function of relative displacement $\mathbf{s} - \mathbf{s}'$; henceforth write it as $\gamma_Z(\mathbf{s} - \mathbf{s}')$. The nugget of the semivariogram of the data is the limit of $\gamma_Z(\mathbf{s} - \mathbf{s}')$ as $\mathbf{s} - \mathbf{s}'$ approaches zero, and is equal to σ_ϵ^2 in this case. Hence, the estimated nugget effect from observations \mathbf{Z} is the estimate of σ_ϵ^2 .

Suppose that $\hat{\boldsymbol{\theta}}_Z \equiv (\hat{\sigma}_\epsilon^2, \hat{\boldsymbol{\theta}}_Y^T)^T$ is the WLS estimator of $\boldsymbol{\theta}_Z$. In (1.6), we plug in $\hat{\boldsymbol{\theta}}_Z$ for $\boldsymbol{\theta}_Z$ in $\boldsymbol{\Sigma}$, to obtain $\hat{\boldsymbol{\Sigma}}$, and $\boldsymbol{\beta}$ is estimated by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{Z}, \quad (1.10)$$

where $\hat{\boldsymbol{\Sigma}} \equiv \mathbf{C}_{\hat{\boldsymbol{\theta}}_Y}(\mathbf{h}) + \hat{\sigma}_\epsilon^2 \mathbf{I}$.

1.3 Optimal prediction of spatial exceedances and spatial exceedance regions

There are several approaches to predict the spatial exceedance and its associated exceedance region. One approach is via loss functions. The other approach focuses on building probability models for the distribution of extreme values. The latter approach includes classical extreme value theory as well as a method called the threshold method.

1.3.1 Loss functions for spatial exceedances

Loss functions to quantify exceedances

The well known loss function for quantifying the hidden geostatistical process, $Y(\cdot)$, is the squared error loss (SEL) function. The kriging predictor (Matheron, 1963; Cressie, 1993b) is the best linear unbiased predictor (BLUP) in terms of squared error loss. With a Gaussian assumption for the hidden process and the error process, it can be shown that the (simple) kriging predictor minimizes the Bayes risk based on the squared error loss function (e.g., Gotway and Cressie, 1993). In the Bayesian context, the optimal Bayes predictor under squared error loss is always the posterior mean. These two predictors (kriging predictor and Bayes predictor) preserve their optimality when predicting linear functionals of the hidden process. If we are interested in prediction of some linear functional of $Y(\cdot)$, the same linear functional with the kriging predictor plugged in is the BLUP and the optimal Bayes predictor under squared error loss, under the same assumptions as given above.

The squared error loss function focuses on the center of the empirical distribution of $Y(\cdot)$, not the tail of empirical the distribution that characterizes exceedances. Louis (1984) and Ghosh (1992) show that the sample variance of the posterior mean is smaller than the posterior expected sample variance of $Y(\cdot)$. This is because the posterior mean has a tighter distribution, having smoothed out the tails of the distribution of $Y(\cdot)$. Cressie et al. (2000) and Wright et al. (2003) proposed the weighted ranks squared error loss (WRSEL) function, based on the order statistics of the observed process of interest. This was designed to avoid squared error loss' underdispersion of predictors where the setting was a finite-dimensional Markov random field.

Craigmile et al. (2006) extended the WRSEL to uncountable-index geostatistical processes. They proposed the integrated weighted quantile squared error loss (IWQSEL) function, which assigns more weight to values of $Y(\cdot)$ that are spatially extreme as determined by the averaged spatial cumulative distribution function (ACDF) of $Y(\cdot)$.

Loss functions to estimate the exceedance region

The exceedance region, as defined by (1.5), is a nonlinear functional of the hidden process $Y(\cdot)$. Although the posterior mean of $Y(\cdot)$ is the optimal Bayes predictor under squared error loss, using a plug-in predictor, based on plugging in for the exceedance region is no longer optimal. The estimated exceedance region is underestimated because of the shorter tails for the empirical distribution of the posterior mean as compared to the distribution of the hidden process. However, there is no theory to justify a plug-in predictor for the exceedance region based on IWQSEL function either, although it does have a more dispersed empirical distribution than that of the posterior mean.

There have been several predictors for nonlinear functionals of the process $Z(\cdot)$ in the geostatistic literature. Indicator kriging (Journel, 1983), indicator cokriging (Lajaunie, 1990), and disjunctive kriging (Matheron, 1976) are appropriate for nonlinear functionals at point level, like $I(Z(\mathbf{s}) \geq K)$, but they are not good at predicting the nonlinear functionals over blocks like $I(Z(A) \geq K)$, where A is a block in D . Constrained Bayes predictors (Louis, 1984; Cressie, 1988; Ghosh, 1992) are designed to predict nonlinear functionals of $Y(\cdot)$ at both point and block support. However, because of imposed variance constraints, these predictors have larger expected sum-of-squared-error loss than Bayes predictors. Moreover, they are designed to predict functionals of $Z(\cdot)$, not those of the hidden process $Y(\cdot)$. Aldworth and Cressie

(2003) proposed a constrained kriging method, covariance matching constrained kriging (CMCK), where both covariance and variance constraints augment the classical kriging equations. These constraints force the entries of the covariance matrix of a vector of linear predictors to match those from the corresponding predictands. Assuming the existence of the second derivative (generally, matrix derivative) of the nonlinear functionals, the CMCK predictor is approximately unbiased for prediction of these nonlinear functionals, and it has approximately optimal mean squared prediction error among predictors satisfying the given constraints. In Chapter 2, we show that the CMCK predictor is obtained by minimizing a loss function involving a squared error loss plus two constraints.

1.3.2 Other approaches for inferences on exceedances

Other approaches for inferences on exceedances are mostly based on modeling the probability distribution for extremes and exceedances. This section mainly reviews *generalized extreme value* (GEV) theory and the *threshold methods* that use probability distributions to describe extremes and exceedances for univariate and multivariate random quantities, and it mentions some extensions to an uncountable-index domain.

Most conventional statistical methods focus primarily on the characteristics of the center of the distribution of random variables (RV), and pay less attention to tails of the RVs. The GEV method was developed to explore the distribution of the maximum of RVs. Later, the threshold method and the related Poisson-Generalized Pareto Distribution (GPD) model was proposed to model exceedances over thresholds. It has been applied in fields of insurance and finance (Embrechts et al., 1997), engineering (Castillo, 1988), hydrology (Katz et al., 2002), and environmental sciences (Gilleland

and Nychka, 2005). The max-stable method (Smith, 1990; Schlather and Tawn, 2003; Cooley et al., 2006), an uncountable-index version of multivariate extreme value theory, has application to the extremes of dependent geostatistical processes, and we shall review this below.

Classical extreme value theory

Classical extreme value theory arises from the limiting distributions of the maximum or minimum of a series of independent random variables (RVs). This theory was first identified by Fisher and Tippett (1928), and then developed as a broadly used approach by Gumbel (1958). Out of the most, important results in the literature (the “three-types theorem”) shows that, under very general conditions, the maximum is approximately distributed as the generalized extreme value distribution. This distribution has three forms: (1) Gumbel (light tail), (2) Fréchet (heavy tail), and (3) Weibull (bounded tail).

More precisely, suppose we have a sequence of RVs, $\{R_1, \dots, R_n\}$, that are independently and identically distributed (IID) with a common distribution function $F(\cdot)$. Let $M_n \equiv \max\{R_1, \dots, R_n\}$. Then, by independence, the distribution of M_n is given by

$$\begin{aligned} P(M_n \leq r) &= P(R_1 \leq r, \dots, R_n \leq r) \\ &= P(R_1 \leq r) \times \dots \times P(R_n \leq r) \\ &= (F(r))^n. \end{aligned} \tag{1.11}$$

For each valid r in the domain of the distribution, $F(r)$ is between 0 and 1. Hence $(F(r))^n \rightarrow 0$ as $n \rightarrow \infty$, which means that the limiting distribution of the maximum of a sequence of IID random variables degenerates to a point mass. To obtain a

non-degenerate distribution, M_n needs to be shifted and scaled appropriately. In particular, the extremal-types theorem, equivalently, the “three-types theorem” (Coles, 2001, p. 46), says that if there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ that satisfy the limit,

$$P((M_n - b_n)/a_n \leq r) \rightarrow G(r), \quad \text{as } n \rightarrow \infty, \quad (1.12)$$

where $G(\cdot)$ is a non-degenerate distribution function, then $G(\cdot)$ belongs to one of the three following families.

1. Gumbel type:

$$G(r) = \exp\{-\exp(-r)\}, \quad -\infty < r < \infty. \quad (1.13)$$

2. Fréchet type:

$$G(r) = \begin{cases} 0, & \text{if } r < 0; \\ \exp\{-(r)^{-a}\}, & \text{if } 0 < r < \infty, \end{cases} \quad (1.14)$$

where a is some fixed positive constant.

3. Weibull type:

$$G(r) = \begin{cases} \exp\{-(-r)^a\}, & \text{if } -\infty < r < 0; \\ 1, & \text{if } r > 0, \end{cases} \quad (1.15)$$

where a is some fixed positive constant.

These three types of limiting distribution can be combined into a single family of distribution functions, given by

$$G(r; \mu_G, \psi, \xi) = \exp\left\{-\left(1 + \xi \frac{r - \mu_G}{\psi}\right)^{-1/\xi}\right\} I\left(1 + \xi \frac{r - \mu_G}{\psi} > 0\right). \quad (1.16)$$

This is known as the *generalized extreme value* (GEV) family of distributions. There are three parameters in the GEV family: a location parameter, $\mu_G \in (-\infty, \infty)$; a

scale parameter, $\psi \in (-\infty, \infty)$; and a shape parameter, $\xi \in (0, \infty)$. This shape parameter determines the nature of the tail of the distribution. The GEV family given by (1.16) has as a special case the Fréchet-type family defined by (1.14) with $a = 1/\xi$, and the Weibull-type family defined by (1.15) with $a = -1/\xi$. The Gumbel-type family defined by (1.13) corresponds to the limiting case of the GEV family as ξ tends to zero.

Estimation of the parameters in a GEV model can be obtained by maximum likelihood estimation (MLE). A numerical solution is required to find MLEs, with some care needed when ξ is close to zero because of numerical difficulties in the neighborhood of $\xi = 0$. Suppose $(\widehat{\mu}_G, \widehat{\psi}, \widehat{\xi})$ is the MLE for (μ_G, ψ, ξ) . With some limitations on ξ (Coles, 2001, pp. 54-55), the asymptotic distribution of the MLE is multivariate normal with mean (μ_G, ψ, ξ) and covariance matrix equal to the inverse of the Fisher information matrix. In practice, this is replaced by the observed Fisher information matrix.

This GEV theory for the maximum requires an assumption of independence for the underlying random process. However, there exist analogous results for non-IID processes. Leadbetter et al. (1983, p. 53) show that if a dependent stationary process satisfies the distributional mixing condition (“condition D ” or “condition $D(u_n)$ ”), the limiting distribution of the maximum of a dependent stationary process also follows the GEV distribution. The distributional mixing condition assumes that for sets of variables that are far enough apart, the difference between a joint distribution for all the random variables and a product of the CDFs of each separate set of variables is small.

In practice, to estimate the parameters in the GEV distribution of a maximum, multiple observations for the maximum (block maxima) are necessary. Suppose R_1, \dots, R_n is a sequence of possibly dependent observations. To generate block maxima, series of observations are blocked into b blocks with m observations in each block. Usually m is large enough to guarantee that the maximum in each block represents a realization of the maximum. Denoting these block maxima by $M_{m,1}, \dots, M_{m,b}$, then by classical GEV theory, they can be modeled as a sequence of IID observations from a GEV distribution. This practical approach requires the number of blocks to be balanced with the length of each block. Blocks with short lengths violate the independence assumption or the distributional mixing condition for these block maxima, and hence introduce biases in the parameter estimation. On the other hand, less blocks of longer length make the assumptions for the GEV distribution valid, but they introduce larger variation in the parameter estimation because less block maxima are available. The block choice is made by trading off bias and variance when estimating the GEV parameters (Coles, 2001, p. 54).

Threshold method

The threshold method was first developed by hydrologists (Todorovic and Zelenhasic, 1970, Todorovic and Rousselle, 1971). These methods assume that the number of exceedances of a sequence of IID observations above a fixed high threshold is a random variable distributed as a Poisson process. In addition, the levels of excesses, conditional on the number of exceedances, is assumed to be independently distributed as the generalized Pareto distribution (GPD) (Davison and Smith, 1990). More precisely, suppose that R is a random variable with distribution function $F(\cdot)$ that falls into the domain of attraction of the GEV distribution. Let R_K be the level of the

excesses above a threshold K conditional on the fact that R exceeds K . Then we have

$$P(R_K \leq r | R > K) = P(R - K \leq r | R > K) = \frac{F(K + r) - F(K)}{1 - F(K)}; \quad r > 0. \quad (1.17)$$

For large enough K , $P(R_K \leq r)$ can be approximated by (e.g., Coles, 2001, pp. 76-77)

$$H(r; \tilde{\psi}, \xi) = \left\{ 1 - \left(1 + \frac{\tilde{\xi} r}{\tilde{\psi}} \right)^{-1/\xi} \right\} I \left(\left(1 + \frac{\tilde{\xi} r}{\tilde{\psi}} \right) > 0 \right) I(r > 0), \quad (1.18)$$

where in the notation of the GEV distribution defined by (1.16), $\tilde{\psi} \equiv \psi + \xi(K - \mu_G)$. The distribution function in (1.18) is known as the *generalized Pareto distribution* (GPD). Notice that the GPD approximation for exceedances has limited utility because the GPD variable cannot be less than K .

This Poisson-GPD approach can be used to derive the generalized extreme value distribution. Let R_1, \dots, R_n denote a sequence of IID RVs. Let N be a RV for the number of exceedances above level K , and R_{K_1}, \dots, R_{K_N} be the corresponding excess values. Assume that, approximately, N has a Poisson distribution with mean λ and R_{K_1}, \dots, R_{K_N} are conditionally IID observations from the GPD given that $N > 1$. Then for n large, $r > K$ and by (1.17), we have

$$\begin{aligned} P(M_n \leq r) &= P(\max\{R_i : i = 1, \dots, n\} \leq r) \\ &\simeq Pr(N = 0) + \sum_{j=1}^{\infty} P(N = j, R_1 \leq r, \dots, R_N \leq r) \\ &= Pr(N = 0) + \sum_{j=1}^{\infty} P(N = j) \cdot P(R_1 \leq r, \dots, R_N \leq r | N = j) \\ &\simeq e^{-\lambda} + \sum_{j=1}^{\infty} \left[\frac{\lambda^j e^{-\lambda}}{j!} \cdot \left\{ \left(1 + \tilde{\xi} \frac{r - K}{\tilde{\psi}} \right)^{-1/\xi} \right\}^j \right] \\ &= \exp \left\{ -\lambda \left(1 + \tilde{\xi} \frac{r - K}{\tilde{\psi}} \right)^{-1/\xi} \right\}. \end{aligned} \quad (1.19)$$

The final expression in (1.19) is the the GEV distribution given in (1.16) if we reparameterize using $\tilde{\psi} = \psi + \tilde{\xi}(K - \mu_G)$ and $\lambda = \left(1 + \tilde{\xi}(K - \mu_G)/\psi\right)^{-1/\tilde{\xi}}$.

The Poisson-GPD model given above assumes that the underlying process is IID. For dependent processes, the occurrence of exceedances tends to appear in clusters. Hydrologists (e.g., North, 1980) model maxima of the dependent process using the “peaks over threshold” (POT) method, which separates out the maxima in each cluster and assumes the maxima are approximately independent. In addition, assuming that the number of exceedances occurs according to a Poisson distribution and the level of the excess is distributed as a GPD, then the Poisson-GPD model is valid for the maximum of dependent processes. Statisticians have also derived the asymptotic distribution for exceedances of dependent processes. For example, Leadbetter et al. (1983, pp. 102-103) and Hsing et al. (1988, p. 99) show that, under very general conditions, the number of exceedances over a high threshold in certain stationary processes occurs according to a a Poisson distribution. Azaïs and Mercadier (2003) show the asymptotic Poisson character of the occurrence of exceedances for certain non-stationary Gaussian processes.

We can use maximum likelihood or maximum profile likelihood to estimate parameters in the Poisson-GPD model. Bayesian inference for the parameters is possible after imposing a prior on the parameters and computing the posterior via Markov chain Monte Carlo (MCMC). One difficulty in fitting the GPD to data is the choice of the threshold K . The difficulty of this choice is analogous to the choice of block parameters in the GEV models. There should be a trade-off between bias and variance of the estimator. If a threshold is too low, it will violate the assumption of the model, leading to biased estimators. If a threshold is too high, it will result in fewer data

available for the model fitting and parameter estimation, leading to a large estimation variance. Some simple practical approaches for the choice of the threshold (e.g., Coles, 2001; Gilleland et al., 2006) are based on the fact that for a GPD distribution, the shape parameter $\tilde{\xi}$ and a certain linear transformation of the scale parameter $\tilde{\psi}$ are almost constant if threshold choice is sensible. Hence, there should be no significant change in the GPD parameters. Another way to select sensible thresholds is using a mean-residual-life plot (Coles, 2001, p. 79). This plot relies on the fact that mean of the excesses above each threshold is expected to change linearly with the threshold, if the GPD approximation is valid for that threshold.

Max-stable models for spatial extremes

In previous subsections, we reviewed the GEV and the Poisson-GPD models for univariate RVs with finite index. Analogous models in uncountable 1-D space are also possible. In what follows, I shall present an approach to apply max-stable models for the spatial exceedance in an uncountable-index spatial domain.

Max-stable processes arise from an uncountable-index generalization of extreme value theory (Smith, 1990; Coles, 2001). Let $\{M(\mathbf{s}) : \mathbf{s} \in D\}$ be a stochastic process with an uncountable-index space, D . Suppose that $M_1(\mathbf{s}), \dots, M_N(\mathbf{s})$ are N independent replications of the process at $\mathbf{s} \in D$. Then $M(\mathbf{s})$ is said to follow a max-stable variable if there exist constants $A_{N\mathbf{s}} > 0$ and $B_{N\mathbf{s}}$ (for $N \geq 1, \mathbf{s} \in D$) that make the following quantity,

$$M^*(\mathbf{s}) = \left(\max_{1 \leq n \leq N} M_n(\mathbf{s}) - B_{N\mathbf{s}} \right) / A_{N\mathbf{s}}; \quad \mathbf{s} \in D, \quad (1.20)$$

identical to $M(\mathbf{s})$ in probability. The idea is that if the maximum of a sequence of independent (transformed) variables is equivalent to the same sequence of variables

in probability, then the variable is max-stable. If $M(\mathbf{s})$ is max stable for each $\mathbf{s} \in D$, then the process $\{M(\mathbf{s}) : \mathbf{s} \in D\}$ is said to be max-stable. Both Smith (1990, pp. 5-6) and Coles (2001, pp. 180-181) derive the joint probability distribution of the max-stable process $\{M(\mathbf{s}) : \mathbf{s} \in D\}$. The calculation assumes that $M(\mathbf{s})$ is distributed marginally as a standard Fréchet distribution for each \mathbf{s} , and then shows that the joint distribution for $\{M(\mathbf{s}) : \mathbf{s} \in D\}$ is given by

$$P(M(\mathbf{s}) \leq K(\mathbf{s}), \forall \mathbf{s} \in D) = \exp \left\{ - \int_{\Lambda} \max_{\mathbf{s} \in D} \left(\frac{f(\mathbf{s}, \lambda)}{K(\mathbf{s})} \right) \nu(d\lambda) \right\}, \quad (1.21)$$

where (\mathbf{s}, λ) is a latent Poisson process on $D \times \Lambda$ (e.g., Smith, 1990). The function $f(\mathbf{s}, \lambda)$ is defined on $D \times \Lambda$, and satisfies

$$\int_{\Lambda} f(\mathbf{s}, \lambda) \nu(d\lambda) = 1, \quad \forall \mathbf{s} \in D, \quad (1.22)$$

where $\nu(d\lambda)$ is a positive intensity function on Λ . Note that if $|D| < \infty$, the joint distribution for a max-stable process given by (1.21) degenerates to the multivariate extreme-value distribution for the maximum. If $|D| = 1$ and $K(\mathbf{s})$ is a constant as \mathbf{s} ranges over D , the joint distribution for the max-stable process can be approximated by the classical GEV distribution for the univariate maximum (Smith, 1990). When measuring bivariate spatial extremes, the *extreme coefficient* is introduced by the max-stable representation of the distribution of bivariate extremes. In what follows, I will introduce the extreme coefficient for a bivariate extreme value distribution.

Let $R(\cdot) \equiv \{R(\mathbf{s}) : \mathbf{s} \in D\}$ denote a strictly stationary process in D . In addition, let $\mathbf{R}(\mathbf{s}) \equiv \{R_1(\mathbf{s}), \dots, R_{m_{\mathbf{s}}}(\mathbf{s})\}$ denote a collection of $m_{\mathbf{s}}$ independent replicates at site \mathbf{s} in D . For example, $\mathbf{R}(\mathbf{s})$ might be the sequence of daily temperatures at a particular station \mathbf{s} . Let $M(\mathbf{s}) \equiv \max\{\mathbf{R}(\mathbf{s})\}$ be the site-wise maximum at the site \mathbf{s} . Then by univariate extreme-value theory, $M(\mathbf{s})$ is asymptotically distributed as a

GEV distribution. Notice that a RV R , distributed as a GEV distribution given by (1.16), can be transformed to a RV R^* , distributed as a standard Fréchet distribution, where $R^* = \frac{\psi(R^\xi - 1)}{\xi} + \mu$. Hence, without loss of generality, let the marginal maximum $M(\mathbf{s})$ be distributed as the standard Fréchet distribution, with distribution function given by

$$P(M(\mathbf{s}) \leq K(\mathbf{s})) = \exp(-1/K(\mathbf{s})); \quad K(\mathbf{s}) > 0, \quad (1.23)$$

where $K(\mathbf{s})$ is a location-specific threshold. Then for a collection of all site-wise maxima $\{M(\mathbf{s} : \mathbf{s}) \in D\}$, the joint max-stable distribution is given by

$$P(\{M(\mathbf{s}) \leq K(\mathbf{s}) : \mathbf{s} \in D\}) = \exp \left\{ - \int_{\Lambda} \max_{\mathbf{s} \in D} \left(\frac{f(\mathbf{s}, \lambda)}{K(\mathbf{s})} \right) \nu(d\lambda) \right\}, \quad (1.24)$$

where the function $f(\mathbf{s}, \lambda)$ is defined on $D \times \Lambda$. Therefore, the joint probability for any bivariate maxima, $(M(\mathbf{s}), M(\mathbf{s} + \mathbf{h}))$, is given by

$$\begin{aligned} & P(M(\mathbf{s}) < K(\mathbf{s}), M(\mathbf{s} + \mathbf{h}) < K(\mathbf{s} + \mathbf{h})) \\ &= \exp \left[- \int \max \left\{ \frac{f(\mathbf{s}, \lambda)}{K(\mathbf{s})}, \frac{f(\mathbf{s} + \mathbf{h}, \lambda)}{K(\mathbf{s} + \mathbf{h})} \right\} \nu(d\lambda) \right]. \end{aligned} \quad (1.25)$$

Usually, we assume $K(\mathbf{s}) \equiv K; \mathbf{s} \in D$. Then, the probability in (1.25) becomes

$$P(M(\mathbf{s}) < K, M(\mathbf{s} + \mathbf{h}) < K) = \exp(-\theta(\mathbf{h})/K) = (F(K))^{\theta(\mathbf{h})}, \quad (1.26)$$

where $F(K) \equiv \exp(-\frac{1}{K})$ and $\theta(\mathbf{h})$ is given by,

$$\theta(\mathbf{h}) \equiv \int \max\{f(\mathbf{s}, \lambda), f(\mathbf{s} + \mathbf{h}, \lambda)\} \nu(d\lambda) \quad (1.27)$$

$$= \int \max\{f(\mathbf{0}, \lambda), f(\mathbf{h}, \lambda)\} \nu(d\lambda), \quad (1.28)$$

since $R(\cdot)$ is strictly stationary. The function $\theta(\mathbf{h})$, which depends on the spatial lag \mathbf{h} , is known as the *extremal coefficient*. The extremal coefficient belongs to the

interval $[1, 2]$. If two maxima are independent, the left-hand side of (1.26) becomes

$$P(M(\mathbf{s}) < K, M(\mathbf{s} + \mathbf{h}) < K) = P(M(\mathbf{s}) < K)P(M(\mathbf{s} + \mathbf{h}) < K) = F(K)^2, \quad (1.29)$$

since $M(\mathbf{s})$ is distributed as the standard Fréchet distribution for all \mathbf{s} in D . Therefore, $\theta(\mathbf{h})$ is equal to 2 if two maxima are independent. If two maxima are perfectly correlated; that is, $M(\mathbf{s}) = M(\mathbf{s} + \mathbf{h})$, the left-hand side of (1.26) becomes $P(M(\mathbf{s}) < K, M(\mathbf{s} + \mathbf{h}) < K) = P(M(\mathbf{s}) < K) = F(K)$, which implies that $\theta(\mathbf{h})$ is equal to 1.

This coefficient approximates the joint distribution of two correlated bivariate maxima by a power of the standard Fréchet distribution with the equivalent sample size equal to $\theta(\mathbf{h})$. However, the coefficient does not give complete information about the dependence structure of bivariate maxima with different thresholds.

Cooley et al. (2006) propose a statistic called the *madogram* to estimate $\theta(\mathbf{h})$. The madogram is the analog of the variogram in geostatistics (the variogram was defined in (1.7)), and is defined by

$$\nu_F(\mathbf{h}) = \frac{1}{2}E[|F(M(\mathbf{s} + \mathbf{h})) - F(M(\mathbf{s}))|]; \quad \mathbf{s} \in D, \quad (1.30)$$

where recall that $F(u) \equiv \exp(-1/u)$. Cooley et al. (2006) shows that the madogram is related to $\theta(\mathbf{h})$ by

$$\theta(\mathbf{h}) = \frac{1 + 2\nu_F(\mathbf{h})}{1 - 2\nu_F(\mathbf{h})}; \quad \mathbf{h} \in \mathbb{R}^d. \quad (1.31)$$

The moment-based estimator of $\nu_F(\mathbf{h})$ is given by

$$\hat{\nu}_F(\mathbf{h}) = \frac{1}{2|\mathcal{N}_h|} \sum_{\|\mathbf{s}_i - \mathbf{s}_j\| \in \mathcal{N}_h} |F(M(\mathbf{s}_i)) - F(M(\mathbf{s}_j))|, \quad (1.32)$$

where \mathcal{N}_h is the set of all sample pairs for which $\|\mathbf{s}_i - \mathbf{s}_j\| \in [h - \delta, h + \delta)$. Just as for the variogram, we can fit parametric families to the madogram estimates, and plug $\hat{\nu}(\mathbf{h})$ into (1.31) to obtain an estimate $\hat{\theta}(\mathbf{h})$ for extremal coefficient $\theta(\mathbf{h})$. This

should be done with care because $\nu_F(\mathbf{h})$ (or $\theta(\mathbf{h})$) has to satisfy certain conditions (see Cooley, 2005, p. 55, for details of the conditions).

For estimation of the dependence of spatial extremes, the madogram connects classical geostatistics with extreme value theory. However, it is only applicable to bivariate extremes. Spatial measures for exceedances, and spatial interpolation methods for maxima are still under development.

Applicability of extreme value theory to the prediction of exceedance regions

Extreme value theory and the threshold method is well suited to finite-index processes in 1-D and 2-D. But, are they useful for spatial maxima with appropriate extensions to 2-D uncountable-index geostatistical processes? The max-stable process model attempts to describe the spatial dependence among uncountable-index spatial maxima based on extreme value theory for the marginal maxima. The max-stable distribution with the extremal coefficient given by (1.27) is introduced to capture the dependence for any spatial bivariate maxima via a parametric joint distribution, but it is not available for uncountable-index spatial extremes. Furthermore, these extreme-value-theory-based approaches require replications at the same site (for example, replication derived from repeatability in time), to allow estimation of the parameters in the GEV or GPD for the marginal maxima. It makes more sense to compare extreme-value-theory-based approaches for spatial maxima with conventional space-time models, rather than the single-replicate geostatistical model introduced in Section (1.2). If we do apply extreme-value-theory-based approaches to the geostatistical model, we can imagine that there are only a few observations available for the inferring the spatial-extreme and the exceedance process. It is then almost

impossible to estimate parameters in the GPD or GEV model because of the lack of observations.

There are other attempts, such as the spatial GPD model (Gilleland et al., 2006), to introduce spatial dependence among uncountable-index spatial extremes by imposing dependence structure on location-specific GPD *parameters* over space. As an example of this approach, Gilleland et al. (2006) compared probability maps of the fourth-highest daily average (FHDA) ozone exceeding 84 ppb using a space-time model and spatial GPD model. The result shows that the probability that the FHDA exceeds 80 ppb in a year is similar whether one uses the extreme-value or a space-time modeling approach, at least for the study region in 1997. The similarity between results from the spatial GPD and the space-time model suggests that the spatial GPD probably still underestimates extremes and exceedances, since the space-time model focuses more on the central part of the distribution.

1.4 Outline of dissertation

In this dissertation, the problem of the prediction of exceedances and the exceedance region is tackled using a series of loss functions. In Chapter 2, I will present loss functions for the prediction of the hidden process, exceedances. These loss functions include squared error loss (SEL), covariance matching constrained kriging (CMCK), and integrated weighted quantile squared error loss (IWQSEL). The prediction of the exceedance region based on the predicted processes are plug-in and sub-optimal. In Chapter 3, I will introduce a loss function we call Baddeley's loss function, which used to obtain the optimal predictor of an exceedance region, given a

known threshold. Often, the threshold is itself random, so Chapter 4 presents a combined loss function for the joint prediction of a spatial quantile and the exceedance region above the quantile. The loss function for quantifying the spatial quantile is the IWQSEL, while the one for identifying the exceedance region is Baddeley's loss function, and the combined loss function is the product of the two. Chapter 5 applies the methodologies to a climate dataset generated from the Climate System Model (CSM) of the National Center for Atmospheric Research (NCAR), in order to detect decadal temperature changes over the Americas (e.g., Shen et al., 2002). Chapter 6 contains a discussion of this area of research, particularly function research that includes extensions of current methods to multiple thresholds where they might be fixed, random, or both. In the appendix, I expand the Baddeley's loss function in simple 1-D and 2-D cases using asymptotic calculations.

CHAPTER 2

LOSS FUNCTION ON EXCEEDANCES

2.1 Introduction

The spatial exceedance is defined as a collection of large values of a (hidden) spatial process. The difficulty of prediction of the spatial exceedance arises from the unobservability of the hidden process, as well as the rareness of the exceedance. Loss-function approaches are a well known method used for the prediction of spatial functionals (Cressie, 1993a, p. 108). Common choices include squared error loss and weighted squared error loss. The resulting predictors from the squared error loss are kriging-type predictors, such as simple kriging (SK), ordinary kriging (OK), universal kriging (UK) (e.g., Cressie, 1993b), and covariance matching constrained kriging (CMCK) (Aldworth and Cressie, 2003). The predictors from weighted square-error loss include weighted rank squared error loss (WRSEL) (Cressie et al., 2000, Wright et al., 2003) and integrated weighted quantile squared error loss (IWQSEL) (Craigmile et al., 2006) predictors. These predictors focus on different parts of the distribution of hidden process. For example, the SK, OK, and UK predictors mainly focus on the center of the distribution of the hidden process, while the IWQSEL predictor has the flexibility to focus on either the center or the tails of the distribution.

Before presenting the loss functions in this chapter, we first review the hierarchical model that we presented in Chapter 1, Section 1.2. We shall consider the same model here.

Recall that $Y(\cdot) \equiv \{Y(\mathbf{s}) : \mathbf{s} \in D\}$ is the hidden geostatistical process of interest, defined in $D \subset \mathbb{R}^d$, with a positive d -dimensional volume, $|D|$. Let $Z(\cdot) \equiv \{Z(\mathbf{s}) : \mathbf{s} \in D\}$ be the process of observed and potentially observed values. Then the hierarchical model can be described by the following data and process models.

The data model:

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \epsilon(\mathbf{s}); \quad \mathbf{s} \in D, \quad (2.1)$$

where $\epsilon(\cdot)$ is a zero-mean white noise process with measurement-error variance, $\sigma_\epsilon^2 > 0$. We assume that $\epsilon(\cdot)$ is independent of $Y(\cdot)$.

The process model:

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}); \quad \mathbf{s} \in D. \quad (2.2)$$

Here the function $\mu(\cdot)$ captures the large-scale, deterministic mean structure (*spatial trend*) of $Y(\cdot)$, and $\delta(\cdot)$ is a stationary stochastic process that models the small-scale spatial structure in $Y(\cdot)$. We assume that the spatial trend, $\mu(\cdot)$, follows a linear model given by

$$\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})' \boldsymbol{\beta}, \quad (2.3)$$

where the vector $\mathbf{x}(\mathbf{s}) \equiv (x_1(\mathbf{s}), \dots, x_p(\mathbf{s}))^T$ is a collection of covariates at \mathbf{s} , and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the *spatial-trend parameter*. We also assume that $\delta(\cdot)$ is a zero-mean second-order stationary spatial process with a parametric covariance function, $C_{\boldsymbol{\theta}_Y}(\cdot)$. Namely,

$$E[\delta(\mathbf{s})] = 0, \quad \text{and } \text{cov}[\delta(\mathbf{s}), \delta(\mathbf{s}')] \equiv C_{\boldsymbol{\theta}_Y}(\mathbf{s} - \mathbf{s}'); \quad \mathbf{s}, \mathbf{s}' \in D, \quad (2.4)$$

where $\boldsymbol{\theta}_Y \equiv (\sigma^2, r)^T$ is the *spatial-covariance parameter* that consists of the *partial sill*, σ^2 , and the *range parameter*, r . In this chapter, we assume that $\boldsymbol{\theta}_Y$ and σ_ϵ^2 are known, and we have n observations, $\mathbf{Z} = \{Z(\mathbf{s}) : \mathbf{s} \in B\}$, taken at a collection of n locations in $B \equiv \{\mathbf{s}_1, \dots, \mathbf{s}_n\} \subseteq D$. We are interested in predicting functionals of the hidden process, $Y(\cdot)$, based on the observations, \mathbf{Z} , using a criterion based on loss functions.

In general, a loss function, $L(\theta, a)$, is a real-valued function defined for all $(\theta, a) \in \Theta \times \mathcal{A}$, where Θ is the parameter space and \mathcal{A} is the action space that contains the parameter space; L is bounded below and $L(\theta, \theta) = 0$ (Berger, 1985). Sometimes, θ is considered to be random, where the randomness of θ is specified by a prior distribution on Θ . Specifically, a loss function for the prediction of $Y(\cdot)$ is defined on $\Theta \times \mathcal{A} = \mathbb{R}^d \times \mathbb{R}^d$. Given the loss function at hand, the optimal Bayes predictor is one that minimizes the expected loss over all possible predictors based on data. When predicting $Y(\cdot)$ by $\hat{Y}(\cdot)$, the *optimal Bayes predictor* (OBP) is a value that minimizes the so-called *Bayes risk* $E[L(Y(\cdot), \hat{Y}(\cdot))]$, where $E[\cdot]$ denotes expectation with respect to the joint distribution of $Y(\cdot)$ and \mathbf{Z} . By a well known result from Bayesian decision theory (e.g., Berger, 1985), the optimal predictor that minimizes Bayes risk is the same one that minimizes $E[L(Y(\cdot), \hat{Y}(\cdot)) | \mathbf{Z}]$, where $E[\cdot | \mathbf{Z}]$ denotes expectation with respect to the conditional distribution of $Y(\cdot) | \mathbf{Z}$.

We now define the squared error loss and the weighted squared error loss for the spatial process, $Y(\cdot)$, defined over the domain, D . Then we derive the optimal Bayesian predictors for each loss function.

2.2 Squared error loss

2.2.1 Loss function for kriging predictor

The *squared error loss* (SEL) for predicting $Y(\mathbf{s})$ by $\hat{Y}(\mathbf{s})$ is given by

$$L_S(Y(\mathbf{s}), \hat{Y}(\mathbf{s})) = \left(Y(\mathbf{s}) - \hat{Y}(\mathbf{s})\right)^2. \quad (2.5)$$

The Bayes risk of the SEL is then given by

$$E \left[L_S(Y(\mathbf{s}), \hat{Y}(\mathbf{s})) \right] = E \left[\left(Y(\mathbf{s}) - \hat{Y}(\mathbf{s})\right)^2 \right]. \quad (2.6)$$

By Bayesian decision theory, the optimal Bayesian predictor (OBP) based on the SEL for site $Y(\mathbf{s})$ is $E[Y(\mathbf{s}) | \mathbf{Z}]$. If the prediction of the whole process $Y(\cdot)$ by $\hat{Y}(\cdot)$ over D is of interest, we introduce the integrated squared error loss, given by

$$L_{IS}(Y(\cdot), \hat{Y}(\cdot)) = \int_D \left(Y(\mathbf{s}) - \hat{Y}(\mathbf{s})\right)^2 d\mathbf{s}. \quad (2.7)$$

The Bayes risk of the ISEL is

$$E \left[L_{IS}(Y(\cdot), \hat{Y}(\cdot)) \right] = E \left[\int_D \left(Y(\mathbf{s}) - \hat{Y}(\mathbf{s})\right)^2 d\mathbf{s} \right] \quad (2.8)$$

$$= \int_D E \left[\left(Y(\mathbf{s}) - \hat{Y}(\mathbf{s})\right)^2 \right] d\mathbf{s}. \quad (2.9)$$

Notice that the component-wise Bayes risk of ISEL inside the integral of equation (2.9) is the Bayes risk of the SEL for each site as given in (2.6). Hence, the optimal predictor at each site that minimizes the Bayes risk of ISEL given by (2.8) is the optimal predictor that minimizes the Bayes risk of SEL given by (2.6). Thus, without loss of generality, for squared error loss, we only need to consider prediction at a single site \mathbf{s} .

The optimal predictor $E[Y(\mathbf{s}) | \mathbf{Z}]$ coincides with the simple kriging predictor for $Y(\mathbf{s})$, if $Y(\cdot)$ and $\epsilon(\cdot)$ in the model given by (2.1) and (2.2) are Gaussian (Gotway

and Cressie, 1993). There are at least three well-known kriging predictors: *simple kriging* (SK), *universal kriging* (UK), and *ordinary kriging* (OK). Suppose that the model parameter vector, $\boldsymbol{\phi} \equiv (\boldsymbol{\beta}^T, \sigma_\epsilon^2, \boldsymbol{\theta}_Y^T)^T$, is known. Recall that the covariance matrix for the data, $\boldsymbol{\Sigma}$, is defined by $\boldsymbol{\Sigma} \equiv \text{var}(\mathbf{Z}) = C_{\boldsymbol{\theta}_Y}(\mathbf{h}) + \sigma_\epsilon^2 \mathbf{I}$, and the covariance vector between the prediction site \mathbf{s} and the data is given by $\mathbf{c}(\mathbf{s}) \equiv \text{cov}(\mathbf{Z}, Y(\mathbf{s})) = (C_{\boldsymbol{\theta}_Y}(\mathbf{s}_1 - \mathbf{s}), \dots, C_{\boldsymbol{\theta}_Y}(\mathbf{s}_n - \mathbf{s}))^T$. When the mean $\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})^T \boldsymbol{\beta}$ is known, the SK predictor for $Y(\mathbf{s})$ is given by

$$\hat{Y}_{SK}(\mathbf{s}) = \mu(\mathbf{s}) + \mathbf{c}(\mathbf{s}_o)' \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \boldsymbol{\mu}). \quad (2.10)$$

When $\boldsymbol{\beta}$ is unknown, it can be estimated and plugged into the SK predictor (2.10). The UK predictor for site \mathbf{s} is

$$\hat{Y}_{UK}(\mathbf{s}) = \mathbf{x}(\mathbf{s})^T \tilde{\boldsymbol{\beta}} + \mathbf{c}(\mathbf{s})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \mathbf{X} \tilde{\boldsymbol{\beta}}), \quad (2.11)$$

where recall that $\mathbf{x}(\mathbf{s}) \in \mathbb{R}^p$ is a collection of covariates at \mathbf{s} , the matrix $\mathbf{X} \equiv (\mathbf{x}(\mathbf{s}_0), \dots, \mathbf{x}(\mathbf{s}_n))^T$ is a $n \times p$ design matrix, and $\tilde{\boldsymbol{\beta}}$ is the BLUE for the spatial trend parameter $\boldsymbol{\beta}$, given by,

$$\tilde{\boldsymbol{\beta}} \equiv (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z}. \quad (2.12)$$

The OK predictor is a special case of the UK predictor with $\mu(\mathbf{s}) \equiv \mu$; that is a constant mean.

2.2.2 Loss function for covariance matching constrained kriging

Kriging predictors (SK, UK, and OK) match their first moment with that of the predictand. However, the variances of kriging predictors are smaller than that of the predictand. This implies that when predicting large or small quantiles of $Y(\cdot)$, which

are non-linear functionals of the predictand, $g(Y(\cdot))$, the plug-in predictor $g(\hat{Y}_{SK}(\mathbf{s}))$, can be biased. This motivates the *constrained kriging* (CK) predictor (Cressie, 1993a) and the *covariance-matching constrained kriging* (CMCK) predictor (Aldworth and Cressie, 2003). The CK predictor matches both the mean and the variance of the predicted process with those of the hidden process for each site; and the CMCK predictor matches the mean and the covariance matrix (including variance) of the predictor with those of the hidden process, $Y(\cdot)$.

Suppose that a collection of locations, $D^* \equiv \{\mathbf{s}_0^*, \dots, \mathbf{s}_m^*\} \subset D$, covers the domain D well, and $Y(\cdot)$ is discretized to $\mathbf{Y} \equiv \{Y(\mathbf{s}) : \mathbf{s} \in D^*\}$. Let $g(\mathbf{Y})$ denote a predictand of interest, where $g(\cdot)$ is some scalar-valued function that is smooth enough in its vector-valued argument, \mathbf{Y} , to possess two continuous derivatives; that is, the vector of the first derivatives, $g'(\mathbf{Y}) \equiv \left(\frac{\partial g(\mathbf{Y})}{\partial Y(\mathbf{s}_0)}, \dots, \frac{\partial g(\mathbf{Y})}{\partial Y(\mathbf{s}_m)} \right)^T$, and the matrix of second derivatives with (i, j) -element given by $g''_{ij}(\mathbf{Y}) = \frac{\partial^2 g(\mathbf{Y})}{\partial Y(\mathbf{s}_i) \partial Y(\mathbf{s}_j)}$, both exist and are continuous. In addition, let $\hat{\mathbf{Y}} \equiv \mathbf{A}^T \mathbf{Z}$ denote a linear predictor of \mathbf{Y} , where $\mathbf{A} \equiv (\mathbf{a}_1, \dots, \mathbf{a}_m)$ is an $n \times m$ matrix. Then the CMCK predictor minimizes a squared-error loss, given by

$$L_{CK}(\mathbf{Y}, \hat{\mathbf{Y}}) = \left(g(\mathbf{Y}) - g(\hat{\mathbf{Y}}) \right)^2 = \left(g(\mathbf{Y}) - g(\mathbf{A}^T \mathbf{Z}) \right)^2, \quad (2.13)$$

with two additional constraints:

$$E[\mathbf{A}^T \mathbf{Z}] = E[\mathbf{Y}], \quad \text{and} \quad \text{var}[\mathbf{A}^T \mathbf{Z}] = \text{var}[\mathbf{Y}]. \quad (2.14)$$

The Bayes risk of the loss function given by (2.13) for the CMCK predictor is

$$E \left[L_{CK}(\mathbf{Y}, \hat{\mathbf{Y}}) \right] = E \left[\left(g(\mathbf{Y}) - g(\mathbf{A}^T \mathbf{Z}) \right)^2 \right]. \quad (2.15)$$

Aldworth and Cressie (2003) showed that if \mathbf{A} satisfies constraints (2.14), then by the δ -method, the Bayes risk in (2.15) can be approximated by

$$E \left[L_{CK} \left(\mathbf{Y}, \hat{\mathbf{Y}} \right) \right] \simeq (g'(\boldsymbol{\mu}_m))^T \mathbf{M}_A g'(\boldsymbol{\mu}_m), \quad (2.16)$$

where $\boldsymbol{\mu}_m \equiv (\mu(\mathbf{s}_1), \dots, \mu(\mathbf{s}_m))^T$, and $\mathbf{M}_A \equiv \text{var}[\mathbf{A}^T \mathbf{Z} - \mathbf{Y}]$. Let $\boldsymbol{\Lambda}_1$ be a $p \times m$ matrix, and $\boldsymbol{\Lambda}_2$ be an $m \times m$ symmetric matrix. Then this constrained minimization problem to obtain the CMCK predictor is equivalent to minimizing an objective function using $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ as Lagrange multipliers, given by

$$f(\mathbf{A}) \equiv -2(g'(\boldsymbol{\mu}_m))^T \mathbf{A}^T \mathbf{C} g'(\boldsymbol{\mu}_m) - 2\mathbf{1}_p^T \boldsymbol{\lambda}_1 (\mathbf{A}^T \mathbf{X} - \mathbf{X}_m) \mathbf{1}_p - \mathbf{1}_m^T \boldsymbol{\lambda}_2 (\mathbf{A}^T \boldsymbol{\Sigma} \mathbf{A} - \boldsymbol{\Sigma}_m) \mathbf{1}_m, \quad (2.17)$$

with respect to \mathbf{A} . Here, $\mathbf{C} \equiv \text{cov}(\mathbf{Z}, \mathbf{Y}) = (\mathbf{c}(\mathbf{s}_1^*), \dots, \mathbf{c}(\mathbf{s}_m^*))$ is an $n \times m$ covariance matrix between the observed data and the hidden process of interest, $\boldsymbol{\Sigma}_m = \text{var}(\mathbf{Y})$ is the covariance matrix of the hidden process, $\mathbf{1}_p$ is a $p \times 1$ vector with all elements equal to 1, and $\mathbf{1}_m$ is an $m \times 1$ vector with all elements equal to 1. The $m \times p$ design matrix is given by $\mathbf{X}_m = (\mathbf{x}(\mathbf{s}_1^*), \dots, \mathbf{x}(\mathbf{s}_m^*))^T$ for all sites in D^* . Define matrices, \mathbf{P} and \mathbf{Q} ,

$$\mathbf{P} \equiv \text{var}[\mathbf{Y}] - \text{var}[\mathbf{X}_m \tilde{\boldsymbol{\beta}}], \quad \text{and} \quad \mathbf{Q} \equiv \text{var}[\hat{\mathbf{Y}}_{UK}] - \text{var}[\mathbf{X}_m \tilde{\boldsymbol{\beta}}],$$

where $\tilde{\boldsymbol{\beta}}$ is the BLUE for $\boldsymbol{\beta}$, given by (2.12). Suppose that \mathbf{P} , \mathbf{Q} , and $\boldsymbol{\Sigma} = \text{var}[\mathbf{Z}]$ are all positive-definite (p.d.), and \mathbf{P}_1 and \mathbf{Q}_1 are square root matrices that satisfy $\mathbf{P} = \mathbf{P}_1 \mathbf{P}_1$ and $\mathbf{Q} = \mathbf{Q}_1 \mathbf{Q}_1$ respectively. Let $\hat{\mathbf{Y}}_{CM}$ denote the CMCK predictor for \mathbf{Y} . Then $\hat{\mathbf{Y}}_{CM}$ is given by (Aldworth and Cressie, 2003),

$$\hat{\mathbf{Y}}_{CM} = \mathbf{X}_m \tilde{\boldsymbol{\beta}} + \mathbf{K}^T \mathbf{C}^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \mathbf{X} \tilde{\boldsymbol{\beta}}), \quad (2.18)$$

where $\mathbf{K} = \mathbf{Q}_1^{-1} \mathbf{P}_1$.

The CMCK predictor $\hat{\mathbf{Y}}_{CM}$ is the optimal linear predictor that minimizes Bayes risk for the loss given by (2.13), in the class of predictors that satisfy the constraints given by (2.14). However, the existence of the CMCK predictor requires \mathbf{P} , \mathbf{Q} , and $\mathbf{\Sigma}$ to be p.d. It is safe to assume $\mathbf{\Sigma}$ is p.d., but the p.d. condition for \mathbf{P} and \mathbf{Q} are not always satisfied. Hence, the CMCK predictor of \mathbf{Y} exists if

$$\mathbf{P} \equiv \text{var}[\mathbf{Y}] - \text{var}[\mathbf{X}_m \tilde{\boldsymbol{\beta}}] \text{ is p.d.}, \quad (2.19)$$

$$\mathbf{Q} \equiv \text{var}[\hat{\mathbf{Y}}_{UK}] - \text{var}[\mathbf{X}_m \tilde{\boldsymbol{\beta}}] \text{ is p.d.}. \quad (2.20)$$

If one of these two conditions are violated, one solution is to partition the process \mathbf{Y} into sub-vectors. Suppose that \mathbf{Y} is partitioned to r parts, $\mathbf{Y} \equiv (\mathbf{Y}_1^T, \dots, \mathbf{Y}_r^T)^T$. Then we can relax the constraints given by (2.19) and (2.20) such that only each of the r sub-vectors satisfies these constraints. There are several ways to partition \mathbf{Y} . For example, Aldworth and Cressie (2003) partitioned \mathbf{Y} into sub-vectors in such a way that each sub-vector not only has equal length, but also has the same geometry. Cressie and Johannesson (2001) proposed another partition algorithm, where each sub-vector consists of one element in \mathbf{Y} and a largest possible set of nearest observations to satisfy the constraints given by (2.19) and (2.20). Note that the number of the nearest observations may vary for each element in \mathbf{Y} .

2.3 Integrated weighted quantile squared error loss

Unlike the classical kriging-type predictors (OK, UK, SK) with tighter distributions than the predictand, the CMCK predictor stretches the prediction distribution by matching the first (mean) and second moments (covariance matrix) with the corresponding moments of the hidden process $Y(\cdot)$. There are other methods to stretch the distribution of kriging-type predictors. For example, Wright et al. (2003) introduce

weighted ranks squared error loss (WRSEL) for finite-index spatial processes. The WRSEL is a weighted squared error loss, where the weight is related to the rank of values of hidden process in a finite-index space. If larger values are more important, the ranks of those values are given more weight than for small values. The optimal predictor is obtained by minimizing the Bayes risk of the WRSEL.

Craigmile et al. (2006) extend the WRSEL to uncountable-index geostatistical processes, and introduce *integrated weighted quantile squared error loss* (IWQSEL). The weight in the IWQSEL is assigned based on the importance of a value determined by the averaged spatial cumulative distribution function (ACDF). Before presenting the IWQSEL, I will first define the ACDF and the inverse ACDF of a geostatistical process.

For a given spatial trend parameter β and a given spatial covariance parameter θ_Y , the ACDF of $Y(\cdot)$ in domain D , denoted by $F_D(y)$, is defined by

$$F_D(y) \equiv \frac{1}{|D|} \int_D P(Y(\mathbf{s}) < y) d\mathbf{s}, \quad (2.21)$$

and the inverse ACDF is defined by

$$F_D^{-1}(\alpha) \equiv \arg \min\{y : F_D(y) \geq \alpha\}, \quad \text{for } \alpha \in [0, 1]. \quad (2.22)$$

When predicting $Y(\cdot)$ by $\hat{Y}(\cdot)$, the IWQSEL function is defined by

$$L_{IW} \left(Y(\cdot), \hat{Y}(\cdot) \right) \equiv \int_D w_D(Y(\mathbf{s})) \left(Y(\mathbf{s}) - \hat{Y}(\mathbf{s}) \right)^2 d\mathbf{s}, \quad (2.23)$$

where the weight function, $w_D(Y(\mathbf{s}))$, is defined by

$$w_D(Y(\mathbf{s})) \equiv \int_0^1 w(p) I \left(Y(\mathbf{s}) \in [F_D^{-1}(p), F_D^{-1}(p + dp)] \right), \quad (2.24)$$

for some given "importance function", $w(\cdot) : [0, 1] \rightarrow [0, \infty)$. Here, $I(\cdot)$ is the indicator function. The IWQSEL function assigns weights to different values of $Y(\cdot)$ based on

their importance as determined by the ACDF. The Bayes risk based on the IWQSEL is given by

$$\begin{aligned} E \left[L_{IW} \left(Y(\cdot), \hat{Y}(\cdot) \right) \right] &= E \left[\int_D w_D(Y(\mathbf{s})) \left(Y(\mathbf{s}) - \hat{Y}(\mathbf{s}) \right)^2 d\mathbf{s} \right] \\ &= \int_D E \left[w_D(Y(\mathbf{s})) \left(Y(\mathbf{s}) - \hat{Y}(\mathbf{s}) \right)^2 d\mathbf{s} \right]. \end{aligned} \quad (2.25)$$

Then the IWQSEL predictor $\hat{Y}_{IW}(\mathbf{s})$, obtained by component-wise minimization of (2.25), is given by

$$\hat{Y}_{IW}(\mathbf{s}) \equiv \frac{E_{Y(\mathbf{s})|\mathbf{Z}}[w_D(Y(\mathbf{s}))Y(\mathbf{s})]}{E_{Y(\mathbf{s})|\mathbf{Z}}[w_D(Y(\mathbf{s}))]}, \quad (2.26)$$

where the expectation $E_{Y(\mathbf{s})|\mathbf{Z}}[\cdot]$ is with respect to the distribution of $Y(\mathbf{s}) | \mathbf{Z}$.

The IWQSEL and related optimal predictor depend on the choice of the “importance function”, $w(p)$. We choose a sigmoid-type importance function (Craigmile et al., 2006), given by

$$w(p) = \frac{1}{1 + e^{-\lambda(p-a)}}; \quad p \in [0, 1], \quad (2.27)$$

where $\lambda > 0$ is a scale parameter, and $1/2 \leq \alpha < 1$ is a quantile of the ACDF of interest. Then the IWQSEL-based predictor given by (2.26) for this importance function is:

$$\hat{Y}_{IW}(\mathbf{s}) = \frac{E_{Y(\mathbf{s})|\mathbf{Z}} \left[Y(\mathbf{s}) \int_0^1 \frac{1}{1+e^{-\lambda(p-a)}} I(Y(\mathbf{s}) \in [F_D^{-1}(p), F_D^{-1}(p+dp)]) \right]}{E_{Y(\mathbf{s})|\mathbf{Z}} \left[\int_0^1 \frac{1}{1+e^{-\lambda(p-a)}} I(Y(\mathbf{s}) \in [F_D^{-1}(p), F_D^{-1}(p+dp)]) \right]}. \quad (2.28)$$

2.3.1 Properties

The IWQSEL predictor given by (2.28) defines a class of predictors that depends on λ . For a given α , a larger λ puts more weight on values of $Y(\cdot)$ near or above the α -th quantile. When λ approaches zero, the sigmoid function approaches 0.5 for all p , which results in an ISEL function, for which the optimal predictor is the posterior

mean $Y(\mathbf{s})$ given the data \mathbf{Z} . When λ approaches ∞ , the sigmoid function approaches the step function, $I(p \geq \alpha)$. Therefore, λ controls the shrinkage of the IWQSEL-based predictor. In the finite-index domain, using the WRSEL, Wright et al. (2003) show that smoother weight functions yield less-biased predictors than do discontinuous weight functions like the step function. We will use simulation studies to investigate this question of bias for the continuous-index domain using the IWQSEL function.

We now define the *spatial cumulative distribution function* (SCDF) (Lahiri et al., 1999) of the process $Y(\cdot)$, namely

$$S_D(y; Y(\cdot)) \equiv \frac{1}{|D|} \int_D I(Y(\mathbf{s}) \leq y) d\mathbf{s}, \quad (2.29)$$

and the inverse SCDF

$$S_D^{-1}(p; Y(\cdot)) \equiv \arg \min\{y : S_D(y; Y(\cdot)) \geq p\}. \quad (2.30)$$

The SCDF is a random functional, where the randomness comes from the original process, $Y(\cdot)$. Note that the expectation of the SCDF, $S_D(y; Y(\cdot))$, with respect to $Y(\cdot)$, is equal to the ACDF, $F_D(y)$. Thus the SCDF of $Y(\cdot)$ is an empirical estimator of the ACDF. In Figure 2.1, we show an example of an ACDF, $F_D(y)$ and two realizations of the corresponding SCDF, $S_D(y; Y(\cdot))$. Note that for a given y , $S_D(y; Y(\cdot))$ changes as $Y(\cdot)$ changes. Hence, for a given quantile, the inverse SCDF, $S_D^{-1}(p; Y(\cdot))$, also changes as $Y(\cdot)$ changes. Let θ_1 denote the α th quantile of the SCDF of interest, defined by $\theta_1 \equiv S_D^{-1}(\alpha; Y(\cdot))$. Further, let $\delta_1(\mathbf{Z})$ denote a predictor for θ_1 based on data \mathbf{Z} . For prediction of only the α th quantile of the SCDF, we use *weighted quantile squared error loss* (WQSEL) function, defined by

$$L_W(\theta_1, \delta_W(\mathbf{Z})) \equiv w_D(\theta_1)(\theta_1 - \delta_1(\mathbf{Z}))^2, \quad (2.31)$$

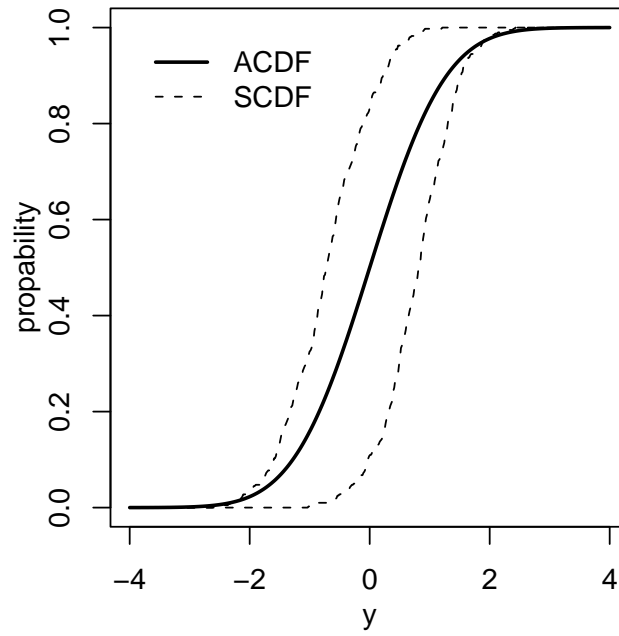


Figure 2.1: The ACDF, $F_D(y)$, versus the SCDF, $S_D(y; Y(\cdot))$. The solid curve is the ACDF, and the dashed curves are the SCDFs for two different realizations of the $Y(\cdot)$ process.

where the weight function, $w_D(\theta_1)$ is defined by

$$w_D(\theta_1) \equiv \int_0^1 w(p) I(\theta_1 \in [F_D^{-1}(p), F_D^{-1}(p + dp)]) , \quad (2.32)$$

and the corresponding “importance function”, $w(\cdot)$, is same as in the IWQSEL. The optimal Bayes predictor based on the WQSEL is

$$\begin{aligned} \tilde{\theta}_1 &\equiv \arg \min_{\delta_1(\mathbf{Z}) \in \mathbb{R}} L_W(\theta_1, \delta_1(\mathbf{Z})) \\ &= \frac{E_{\theta_1|\mathbf{Z}} \left[\theta_1 \int_0^1 w(p) I(\theta_1 \in [F_D^{-1}(p), F_D^{-1}(p + dp)]) \right]}{E_{\theta_1|\mathbf{Z}} \left[\int_0^1 w(p) I(\theta_1 \in [F_D^{-1}(p), F_D^{-1}(p + dp)]) \right]} \\ &= \frac{E_{Y(\cdot)|\mathbf{Z}} \left[S_D^{-1}(\alpha; Y(\cdot)) \int_0^1 w(p) I(S_D^{-1}(\alpha; Y(\cdot)) \in [F_D^{-1}(p), F_D^{-1}(p + dp)]) \right]}{E_{Y(\cdot)|\mathbf{Z}} \left[\int_0^1 w(p) I(S_D^{-1}(\alpha; Y(\cdot)) \in [F_D^{-1}(p), F_D^{-1}(p + dp)]) \right]} . \end{aligned} \quad (2.33)$$

2.4 Simulation study

We have designed a simulation study to compare performance of the kriging predictor, the CMCK predictor based on a partition with nearest observations, and the IWQSEL predictor (for different values of the tuning parameter λ in the IWQSEL predictor). The details of the simulation study are now presented.

We consider a unit square domain of $D = \{\mathbf{s} = (x, y) : 0 \leq x \leq 1; 0 \leq y \leq 1\}$ (i.e., $\omega = 1$). We will observe data and evaluate each predictor on the gridded domain $D^* = \{\mathbf{s}^* = (x, y) : x = 0 = \frac{0}{M-1}, \frac{1}{M-1}, \dots, 1; y = 0 = \frac{0}{M-1}, \frac{1}{M-1}, \dots, 1\}$ that covers D , with $M = 20$. Assuming no spatial trend in (2.2) (i.e., $\boldsymbol{\beta} = \mathbf{0}$), we let the hidden process \mathbf{Y} be an isotropic mean-zero Gaussian process with an exponential covariance function. Recall that this process is characterized solely by its variance σ^2 , and by the spatial range parameter, r . That is, the exponential covariance function is given

by

$$C_{\theta_Y}(\mathbf{s}_i - \mathbf{s}_j) = \sigma^2 \exp(-\|\mathbf{s}_i - \mathbf{s}_j\|/r), \quad (2.34)$$

where the Euclidean distance function, $\|\mathbf{s}_i - \mathbf{s}_j\|$, is given by

$$\|\mathbf{s}_i - \mathbf{s}_j\| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}.$$

In our experiments, we let $\sigma^2 = 1$, and vary r from 0.5 (weak spatial dependence), 1.5 (moderate spatial dependence), to 5 (very strong spatial dependence). Using a Cholesky decomposition for the covariance matrix, we simulate 100 realizations of $\mathbf{Y} = \{Y(\mathbf{s}) : \mathbf{s} \in D^*\}$ for each of the three levels (weak, moderate, strong) for the spatial dependence parameter. To each $Y(\mathbf{s})$ in D^* , we add independent Gaussian measurement error generated from $N(0, \sigma_\epsilon^2)$, to obtain the data over D^* . We then sample 1/4 of all sites in D^* as the collection of observation sites B for experiments, and henceforth fix B . We consider two levels of measurement-error variance: $\sigma_\epsilon^2 = 0.5$ and $\sigma_\epsilon^2 = 2$. Defining the signal-to-noise ratio (SNR) as the ratio of the partial sill in the hidden process to the measurement-error variance (i.e., $\text{SNR} \equiv \sigma^2/\sigma_\epsilon^2$), this corresponds to SNRs of 2 and 0.5. This definition of the signal-to-noise ratio ignores the amount of spatial dependence in the spatial process. In summary, the factors and levels for our 2-factor design are listed in Table 2.1.

Predictands and responses

In the simulation, given the α th quantile of the ACDF, $q(\alpha) = F_D^{-1}(\alpha)$, we are interested in predicting $\alpha = 0.5, 0.75, 0.9$ using a plug-in predictor $S_D(q(\alpha); \hat{Y}(\cdot))$. We consider two responses in the experiment: the *bias* and the *root mean square prediction error* (RMSE). The bias of $S_D(q(\alpha); \hat{Y}(\cdot))$ is

$$\text{Bias}[S_D(q(\alpha); \hat{Y}(\cdot))] \equiv E[S_D(q(\alpha); \hat{Y}(\cdot)) - \alpha], \quad (2.35)$$

Factor	Number of levels	Values of the levels
r	3	0.5
		1.5
		5
SNR	2	2
		0.5

Table 2.1: Levels of factors in simulation studies for the single-loss approach.

and the RMSE of $S_D(q(\alpha); \widehat{Y}(\cdot))$ is

$$\text{RMSE}[S_D(q(\alpha); \widehat{Y}(\cdot))] \equiv \sqrt{E[S_D(q(\alpha); \widehat{Y}(\cdot)) - \alpha]^2}. \quad (2.36)$$

The bias and RMSE of each predictor for all combinations of the factors and predicands of the experiment are obtained from the experimental RMSE and the experimental bias, given by averaging over the Monte Carlo samples of the experiment.

Computation in a discrete domain

When the domain of interest is discretized, we approximate integrals over D to calculate the ACDF and SCDF of the process. Recall that D^* is a discrete domain with D , defined by $D^* \equiv \{\mathbf{s}_0^*, \dots, \mathbf{s}_m^*\}$, and \mathbf{Y} is the discretized process for $Y(\cdot)$ in D^* , defined by $\mathbf{Y} \equiv \{Y(\mathbf{s}) : \mathbf{s} \in D^*\}$. Let $\widetilde{F}_D(y)$ denote the discrete version of the ACDF of $Y(\cdot)$ in D^* , defined by

$$\widetilde{F}_D(y) \equiv \frac{1}{m} \sum_{j=1}^m P(Y(\mathbf{s}_j) \leq y). \quad (2.37)$$

Then, for $\alpha \in [0, 1]$, the approximation for the inverse ACDF, denoted by $\widetilde{F}_D^{-1}(\alpha)$, is defined by

$$\widetilde{F}_D^{-1}(\alpha) \equiv \arg \min\{y : \widetilde{F}_D(y) \geq \alpha\}. \quad (2.38)$$

Let $\tilde{S}_D(y; \mathbf{Y})$ denote an analogous discrete version of the SCDF in D^* , defined by

$$\tilde{S}_D(y; \mathbf{Y}) \equiv \frac{1}{m} \sum_{k=1}^m I(Y(\mathbf{s}_k) \leq y). \quad (2.39)$$

Then the corresponding approximation for the inverse SCDF, $\tilde{S}_D^{-1}(\alpha)$, is

$$\tilde{S}_D^{-1}(\alpha; \mathbf{Y}) \equiv \arg \min\{y : \tilde{S}_D(y; \mathbf{Y}) \geq \alpha\}. \quad (2.40)$$

Given the discrete approximation for the ACDF and the integral approximation for p , we can approximate the weight function $w_D(y)$ by

$$\tilde{w}_D(Y(\mathbf{s})) = \int_0^1 w(p) I\left(Y(\mathbf{s}) \in [\tilde{F}_D^{-1}(p), \tilde{F}_D^{-1}(p + dp)]\right) \quad (2.41)$$

$$\simeq \frac{1}{G} \sum_{g=1}^G w\left(\frac{g}{G}\right) I\left(Y(\mathbf{s}) \in \left[\tilde{F}_D^{-1}\left(\frac{g-1}{G}\right), \tilde{F}_D^{-1}\left(\frac{g}{G}\right)\right]\right), \quad (2.42)$$

where G is the total number of values used in the discretization for p .

Gaussianity of both the true process $Y(\cdot)$ and the measurement-error process helps to simplify the probability calculations in the IWQSEL and the WQSEL predictors. The first simplification implies that the ACDF of $Y(\cdot)$ is $P(Y(\mathbf{s}) \leq y) = \Phi\left(\frac{y - x(\mathbf{s})^T \boldsymbol{\beta}}{\sigma}\right)$. Secondly, the conditional distribution of $\mathbf{Y} | \mathbf{Z}$ is a known multivariate Gaussian distribution, $N_m(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, with the parameters given by,

$$\boldsymbol{\mu}^* = \mathbf{X}_m \boldsymbol{\beta} - \mathbf{F}^* \boldsymbol{\Sigma}^{-1}(\mathbf{X} \boldsymbol{\beta} - \mathbf{Z}), \text{ and } \boldsymbol{\Sigma}^* = \mathbf{C}^* - \mathbf{F}^* \boldsymbol{\Sigma}^{-1}(\mathbf{F}^*)^T, \quad (2.43)$$

where \mathbf{F}^* is an $m \times n$ matrix with (i, j) -element; $\text{cov}(Y(\mathbf{s}_i^*), Z(\mathbf{s}_j)) = \text{cov}(Y(\mathbf{s}_i^*), Y(\mathbf{s}_j))$; \mathbf{C}^* is an $m \times m$ matrix with (i, j) -element $\text{cov}(Y(\mathbf{s}_i^*), Y(\mathbf{s}_j^*))$; and $\boldsymbol{\Sigma}$ is an $n \times n$ matrix with (i, j) -element $\text{cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j))$. Hence, both conditional expectations in the IWQSEL predictor and the WQSEL predictor can be approximated by averages over Monte Carlo samples from known Gaussian distributions.

Simulation results

In Figure 2.2, we show the relationship between the bias of the predictor $S_D(q(\alpha); \hat{Y}(\cdot))$ given by (2.35), and the tuning parameter λ in the importance function (2.27), as well as the relationship between the RMSE of the predictor given by (2.36) and λ , when SNR=2.0. We show the same relationships in Figure 2.3 when SNR=0.5. In bias plots of Figures 2.2 and 2.3, the circle represents the bias; the horizontal solid line represents the zero bias; and the dashed line is the bias for the CMCK predictor. In RMSE plots of Figures 2.2 and 2.3, the circle represents the RMSE; and the dashed line is the RMSE for the CMCK predictor.

We observe that the SNR is not an important factor in these experiments since the patterns between results with SNR= 2 and with SNR= 0.5 are almost the same. The biases of estimating the $S_D(q(0.5); Y(\cdot))$ using $S_D(q(0.5); \hat{Y}(\cdot))$ based on the SK and the CMCK predictors are very close to zero since both predictors are good at the prediction of the center of the distribution of the hidden process. Hence, based on the bias criterion, for $\alpha = 0.5$, we should choose $\lambda = 0$ in the IWQSEL, which corresponds to the SK predictor. For higher quantiles, the biases of $S_D(q(\alpha); \hat{Y}(\cdot))$ based on the SK predictor increase and are positive since the SK predictor tends to have tighter distribution in the center, and it has less extreme values. Therefore, the percentage of the predicted process values below a given high threshold (e.g., the 75th and 90th percentiles of the ACDF) is over-predicted using SK. Alternatively, CMCK stretches the distribution of predicted values so that there are more extreme values in the right (and left) tail of the distribution. For higher percentiles, we see an under-predicted percentage of the process below a given threshold, and as the percentile of interest,

α , increases, it gets worse. The RMSE of CMCK increases as the percentile increases due to the increase of the bias.

By changing λ in the IWQSEL, the bias of $S_D(q(\alpha); \hat{Y}(\cdot))$ based on the IWQSEL can be made small, zero in most cases. We observe that the choice of λ that minimizes the bias tends to increase as α increases, since for larger percentiles the IWQSEL predictor needs larger λ to put more weight on larger values. For the same level of α , λ tends to increase as the spatial dependence (expressed in terms of the range r) tends to be stronger: for stronger dependence, the process tends to have similar extreme values appearing in subregions of D . Therefore, IWQSEL-based predictor needs larger λ in order to discriminate large extreme values. We also observe that the RMSE of $S_D(q(\alpha); \hat{Y}(\cdot))$ increases as λ increases. Based on the bias criterion, we propose the following rule, listed in Table 2.2, for the choice of λ in the IWQSEL. In the table, r is the range parameter that captures the spatial dependence in the hidden process. Recall that $r = 0.5$ represents the “weak” dependence, $r = 1.5$ represents the “moderate” dependence, and $r = 5$ represents the “strong” dependence.

SNR	r	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 0.9$
2	0.5	0	2	6
	1.5	0	6	6
	5	0	6	18
0.5	0.5	0	2	6
	1.5	0	6	6
	5	0	6	18

Table 2.2: The choice of λ (to the nearest integer) in IWQSEL for minimizing the bias in estimating quantile α .

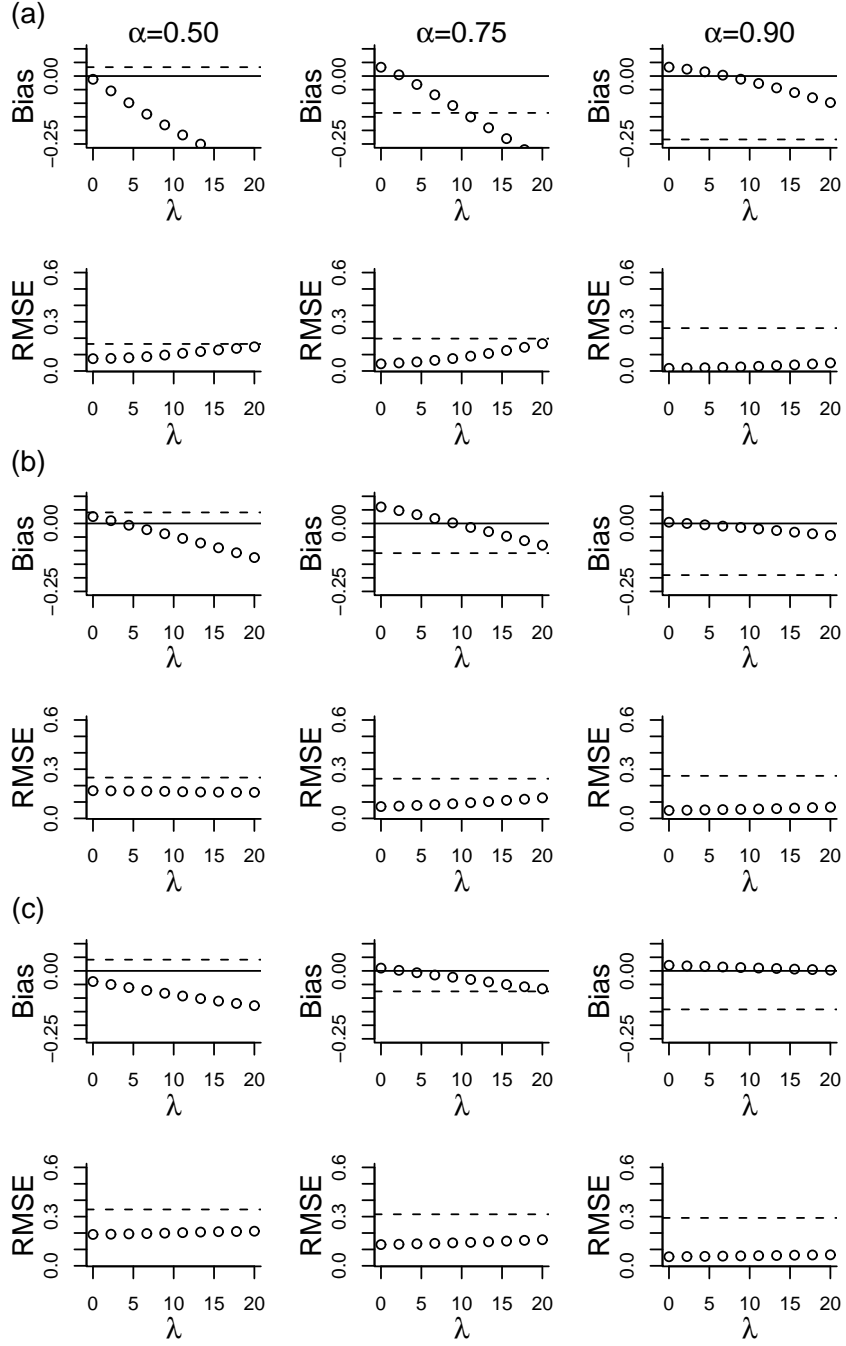


Figure 2.2: The bias and the RMSE of $S_D(q(\alpha); \hat{Y}(\cdot))$ from IWQSEL-based predictor versus λ for SNR = 2; Fig 2.2(a) shows results for $r = 0.5$, Figure 2.2(b) shows results for the $r = 1.5$, and Figure 2.2(c) shows results for $r = 5$. Dashed lines show the corresponding results for CMCK-based predictor.

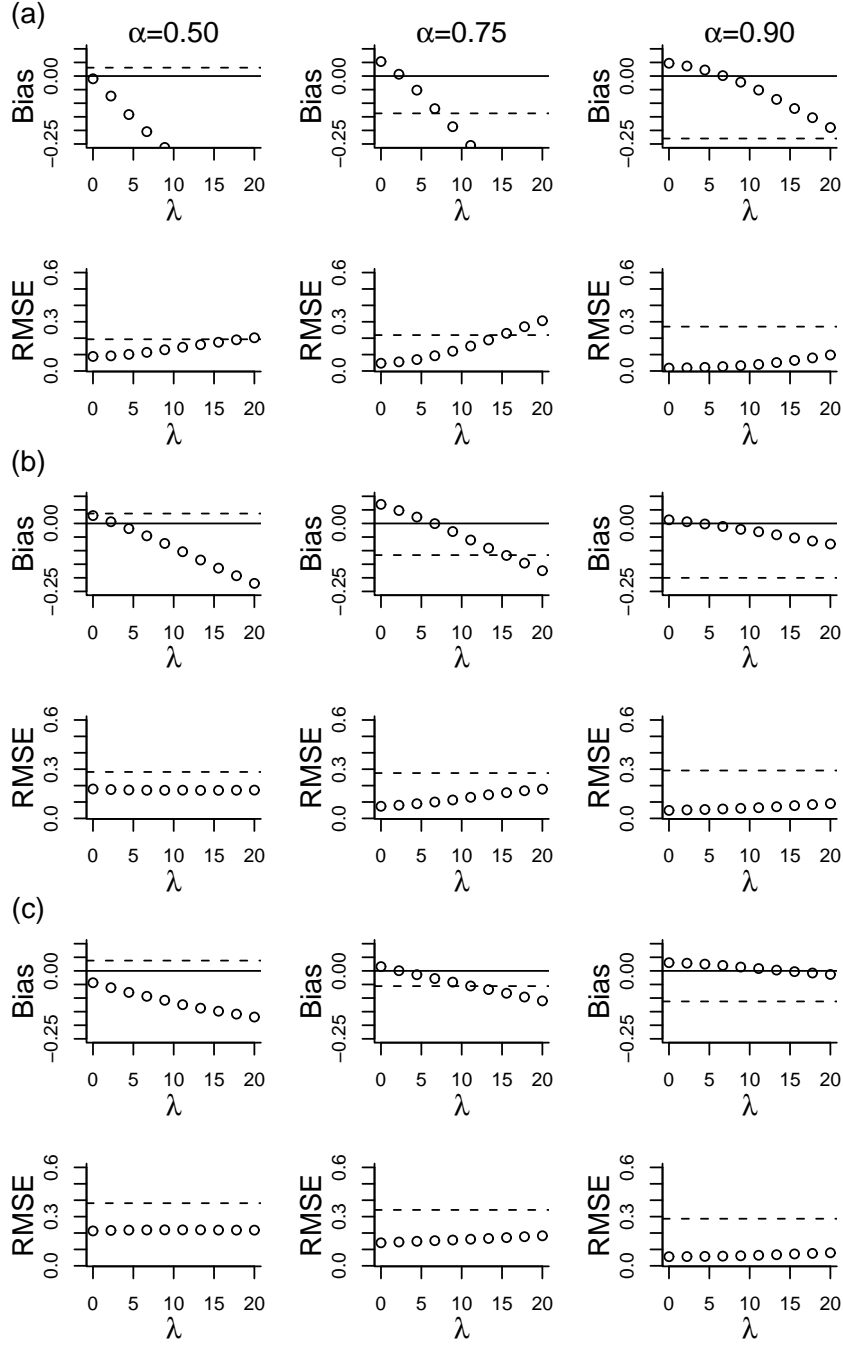


Figure 2.3: The bias and the RMSE of $S_D(q(\alpha); \hat{Y}(\cdot))$ from IWQSEL-based predictor versus λ for SNR = 0.5; Fig 2.2(a) shows results for $r = 0.5$, Figure 2.2(b) shows results for $r = 1.5$, and Figure 2.2(c) shows results for $r = 5$. Dashed lines show the corresponding results for CMCK-based predictor.

After choosing λ , we compute the empirical percentile of the ACDF based on the predicted process. To illustrate using simulations, we generate data from a spatial process that has a weak spatial dependence with r equal to 0.5, SNR equal to 2, and we set the target quantile equal to 0.9. Then from Table 2.2, we choose λ equal to 6.

Figure 2.4 shows the proportion of the predicted process below the inverse ACDF of the process for one particular realization of the process chosen at random. In Figure 2.4, we observe that the IWQSEL predictor with chosen λ is very good at estimating of the target quantile 0.9, shown as the horizontal and the vertical lines. Notice that the 45° line in Figure 2.4 represents the unbiased prediction of IWQSEL predictor for quantiles. Hence, the IWQSEL predictor underestimates quantiles less than the target quantile 0.9, and overestimates those larger than 0.9. The SK predictor overestimates the target quantile, but is best when the quantiles are below 0.7, since the SK predictor predicts the center of the distribution of the hidden process well. The CMCK predictor has larger predictions (for this particular realization), and hence underestimates the quantiles.

2.5 Discussion

In this chapter, we use a loss-function approach to predict exceedances. The SEL and the resulting kriging predictor mainly focus on the center of the distribution of the hidden process. Hence, the kriging-type predictors for the hidden process has tighter prediction distributions than the hidden process, resulting in biased predictors for the exceedance. The CMCK predictor does a better job of stretching the prediction distribution by matching the covariance matrix of the predictor and the predictand. However, the CMCK-based predictor for the exceedance is also biased because of

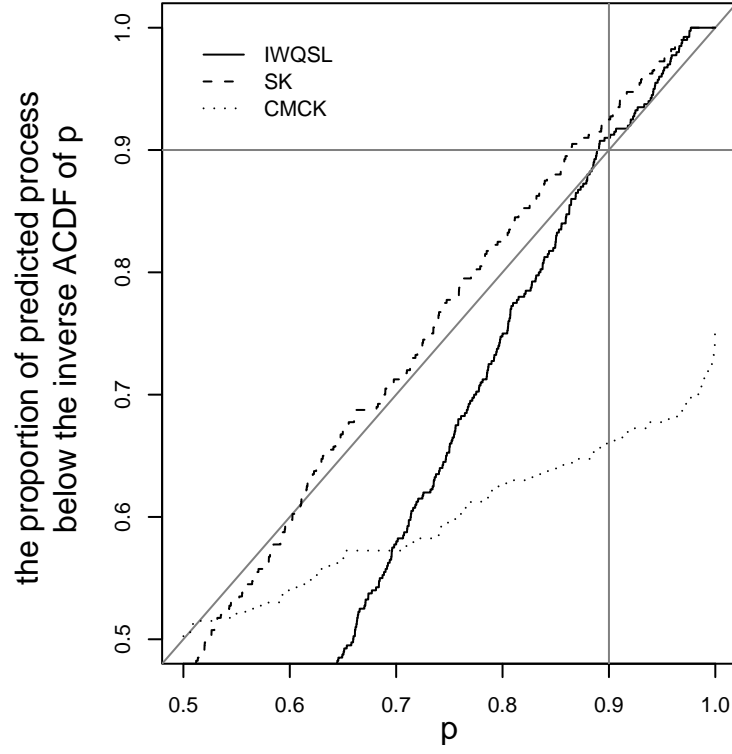


Figure 2.4: The proportion of a predicted process below the inverse ACDF of the process; data were obtained from a single realization with $r = 0.5$, and $\text{SNR} = 2$. The quantile $\alpha = 0.9$ (and hence $\lambda = 6$) was chosen. The vertical solid line shows the target quantile 0.9. The horizontal solid line shows the 0.9-empirical quantile using the proportion of a predicted process below the target quantile of the ACDF. The 45° solid line represents the unbiased prediction of IWQSEL predictor for quantiles.

the highly non-linear functionals in the exceedance. IWQSEL-based prediction is another way to stretch the prediction distribution by using a loss-function approach. Approximate unbiasedness of the IWQSEL-based predictor for the exceedance can be achieved by the choice of a tuning parameter, λ , in the weight function. We have given recommendations for the choice of λ for different exceedance and different spatial processes via simulation studies.

However, these loss functions are based directly on the process, and they do not address prediction of the exceedance region above some threshold. Ad-hoc prediction for the exceedance region using a plug-in approach is not optimal. In the next chapter, I shall present an approach to predict the exceedance region using a loss function we call Baddeley's loss function that is based directly on the exceedance region. The optimal predictor of the exceedance region is then obtained by minimizing the Bayes risk of Baddeley's loss function.

CHAPTER 3

PREDICTION OF THE EXCEEDANCE REGION ABOVE ONE THRESHOLD

3.1 Introduction

In Chapter 2, we have reviewed two classes of loss functions for prediction of the spatial exceedance, including squared error loss and weighted squared error loss. We demonstrated that the predictors based on squared error loss, such as the kriging-based predictor and the CMCK-based predictor, are biased for the prediction of a spatial exceedance. The IWQSEL-based predictor can achieve approximate unbiasedness for prediction of the exceedance for a suitable choice of a tuning parameter. Recall that an *exceedance region* (exceedance set) is the set of all spatial locations where the process of interest has values larger than some exceedance value. Even though the IWQSL is unbiased for the prediction of exceedance values, plugging in these predictors does not yield the optimal Bayes predictor of the exceedance region. In this chapter, we are interested in loss functions based directly on the exceedance region of a spatial process above a fixed threshold. We first review the hierarchical model that we presented previously. We shall consider the same model here.

Recall that $Y(\cdot) \equiv \{Y(\mathbf{s}) : \mathbf{s} \in D\}$ is the hidden geostatistical process of interest, defined in $D \subset \mathbb{R}^d$, with a positive d -dimensional volume, $|D|$. Let $Z(\cdot) \equiv \{Z(\mathbf{s}) : \mathbf{s} \in D\}$ be the process of observed and potentially observed values. Then the hierarchical model can be described by the following data and process models.

The data model:

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \epsilon(\mathbf{s}); \quad \mathbf{s} \in D, \quad (3.1)$$

where $\epsilon(\cdot)$ is a zero-mean white noise process with measurement-error variance, $\sigma_\epsilon^2 > 0$. We assume that $\epsilon(\cdot)$ is independent of $Y(\cdot)$.

The process model:

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}); \quad \mathbf{s} \in D. \quad (3.2)$$

Here the function $\mu(\cdot)$ captures the large-scale, deterministic mean structure (*spatial trend*) of $Y(\cdot)$, and $\delta(\cdot)$ is a stationary stochastic process that models the small-scale spatial structure in $Y(\cdot)$. We assume that the spatial trend, $\mu(\cdot)$, follows a linear model given by

$$\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})' \boldsymbol{\beta}, \quad (3.3)$$

where the vector $\mathbf{x}(\mathbf{s}) \equiv (x_1(\mathbf{s}), \dots, x_p(\mathbf{s}))^T$ is a collection of covariates at \mathbf{s} , and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the *spatial-trend parameter*. We also assume that $\delta(\cdot)$ is a zero-mean, second-order stationary spatial process with a parametric covariance function, $C_{\boldsymbol{\theta}_Y}(\cdot)$. Namely,

$$E[\delta(\mathbf{s})] = 0, \quad \text{and } \text{cov}[\delta(\mathbf{s}), \delta(\mathbf{s}')] \equiv C_{\boldsymbol{\theta}_Y}(\mathbf{s} - \mathbf{s}'); \quad \mathbf{s}, \mathbf{s}' \in D, \quad (3.4)$$

where $\boldsymbol{\theta}_Y \equiv (\sigma^2, r)^T$ is the *spatial-covariance parameter* that consists of the *partial sill*, σ^2 , and the *range parameter*, r . In this chapter, we assume that $\boldsymbol{\theta}_Y$ and σ_ϵ^2 are known. In practice they need to be estimated from the data (e.g., see Chapter 5).

Suppose henceforth that we have n observations $\mathbf{Z} \equiv \{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\}$ at a collection of observation sites, $B \equiv \{\mathbf{s}_1, \dots, \mathbf{s}_n\} \subset D$. A problem of interest is to use \mathbf{Z} to predict the exceedance region of the true process $Y(\cdot)$:

$$e_K^Y \equiv \{\mathbf{s} \in D : I(Y(\mathbf{s}) > K) = 1\}, \quad (3.5)$$

above a threshold value K . There is also an analogous problem where we wish to predict the region below a threshold; the development that follows can be easily adapted to this case.

There is a natural link between exceedance regions of a geostatistical process and a binary image. A binary image is a function $f : D \rightarrow \{0, 1\}$. It is in 1-1 correspondence with its foreground set:

$$e^f \equiv \{\mathbf{s} \in D : f(\mathbf{s}) = 1\}. \quad (3.6)$$

If we apply the thresholding function $f(\mathbf{s}) = I(Y(\mathbf{s}) > K)$ in (3.5) to each location in D , we obtain a binary image with the foreground set given by the exceedance set (3.5). Finally, let $\Omega = 2^D$ denote the set of all possible exceedance sets in D .

3.2 Image metrics and the resulting predictor

The link between spatial exceedance regions and binary images motivates us to use image metrics as loss functions to predict the exceedance region. A loss function, $L(\theta, a)$, is a real-valued function defined for all $(\theta, a) \in \Theta \times \mathcal{A}$, where Θ is the parameter space and \mathcal{A} is an action space that contains the parameter space; L is bounded below by 0 and $L(\theta, \theta) = 0$. Sometimes, θ is considered to be random, with the randomness of θ specified by a prior distribution on Θ . Hence, a loss function for the prediction of e_K^Y is defined on $\Theta \times \mathcal{A} = \Omega \times \Omega$. The true value of interest,

e_K^Y , has a prior distribution, which inherits its randomness from $Y(\cdot)$ (recall that ϕ is fixed and to be estimated). Metrics for binary images are usually defined through foregrounds (since there is a 1-1 map between a binary image and the corresponding foreground). Let ρ be a metric for binary images, where ρ is a function $\Omega \times \Omega$ to $\mathbb{R}^+ \equiv \{x : x \geq 0\}$. Then a metric (or distance function) on a set \mathcal{S} is a function $\rho : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}^+$, where for all s_1, s_2, s_3 in \mathcal{S} , ρ satisfies:

1. $\rho(s_1, s_2) \geq 0$ (non-negativity), and $\rho(s_1, s_2) = 0$ if $s_1 = s_2$,
2. $\rho(s_1, s_2) = \rho(s_2, s_1)$ (symmetry),
3. $\rho(s_1, s_3) \leq \rho(s_1, s_2) + \rho(s_2, s_3)$ (triangle inequality).

Since metrics for binary images are defined on $\Omega \times \Omega$ and satisfy the regularity conditions for loss functions, they can be used as a loss function for prediction of the exceedance region defined by (3.5).

An example of a metric for binary images is Baddeley's metric (Baddeley, 1992), which measures the discrepancy between two binary images through their corresponding foregrounds. Suppose f and g are two binary images on D with corresponding foreground sets e^f and e^g , respectively. Then Baddeley's metric is defined as:

$$\rho_p^B(e^f, e^g) \equiv \left[\frac{1}{|D|} \int_D |(u(d(\mathbf{s}, e^f)) - u(d(\mathbf{s}, e^g)))|^p d\mathbf{s} \right]^{1/p}, \quad (3.7)$$

for $0 \leq p < \infty$. Here the function $u : [0, \infty] \rightarrow [0, W]$ is any bounded concave function with $u(0) = 0$, and $d(\mathbf{s}, e)$ is a distance between location $\mathbf{s} \in D$ and foreground set $e \subset D$, defined by:

$$d(\mathbf{s}, e) \equiv \begin{cases} \min\{\rho(\mathbf{s}, \mathbf{s}') : \mathbf{s}' \in e\}, & \text{if } e \neq \emptyset; \\ \xi, & \text{if } e = \emptyset. \end{cases} \quad (3.8)$$

In (3.8), $\rho(\mathbf{s}, \mathbf{s}')$ is some metric defined between spatial locations \mathbf{s} and \mathbf{s}' (e.g., Euclidean, Manhattan), and ξ is the maximum pair-wise distance in the domain D ; that is, $\xi \equiv \max\{\rho(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{s}_i, \mathbf{s}_j \in D\}$. When $\rho(\mathbf{s}, \mathbf{s}') = \|\mathbf{s} - \mathbf{s}'\|$, where $\|\mathbf{s}_i - \mathbf{s}_j\|$ is the Euclidean metric between sites \mathbf{s}_i and \mathbf{s}_j , and $u(\cdot)$ is chosen to be the taper function defined as $u(x) \equiv \min(x, t)$, for a fixed taper $t > 0$ (Baddeley, 1992), the tapered version of Baddeley's metric for binary images is

$$\rho_p^B(e^f, e^g; t) \equiv \left[\frac{1}{|D|} \int_D |d_t(\mathbf{s}, e^f) - d_t(\mathbf{s}, e^g)|^p d\mathbf{s} \right]^{1/p}. \quad (3.9)$$

In (3.9), the *tapered distance* is $d_t(\mathbf{s}, e) \equiv \min(d(\mathbf{s}, e), t)$, and from (3.8),

$$d(\mathbf{s}, e) = \begin{cases} \min\{\|\mathbf{s} - \mathbf{s}'\| : \mathbf{s}' \in e\}, & \text{if } e \neq \emptyset; \\ \xi, & \text{if } e = \emptyset. \end{cases} \quad (3.10)$$

The tapered distance limits the effect of global background noise, since the contributions to the metric (3.9) for locations \mathbf{s} further than t units away from e^f and e^g are zero.

A special case of the tapered version of Baddeley's metric for binary images is the *site-wise misclassification metric* $\rho_1^B(e^f, e^g; 1)$, which is denoted by $\rho^E(e^f, e^g)$. Another metric for binary images is the Hausdorff metric (e.g., Rote, 1991) given by

$$\rho^H(e^f, e^g) \equiv \sup_{\mathbf{s} \in D} |d(\mathbf{s}, e^f) - d(\mathbf{s}, e^g)|, \quad (3.11)$$

where $d(\mathbf{s}, e^f)$ is defined by (3.10). The Hausdorff metric is the limiting case of $\rho_p^B(e^f, e^g; t)$ when p approaches ∞ with $t > \xi$ (Baddeley, 1992).

In the Bayesian framework, the optimal predictor of the exceedance set based on an image metric is obtained as follows. Suppose $\rho(e_K^Y, \hat{e})$ is a loss function for prediction of the exceedance region based on a metric $\rho(\cdot, \cdot)$ for binary images, where $\hat{e} \in \Omega$ is an action that is a function of observations \mathbf{Z} . Then, a Bayes rule can be

found by an action that minimizes the posterior expected loss (Berger, 1985, p.159). Suppose e_K^* is the Bayes rule (also known as the *optimal Bayes predictor*, or OBP), assuming the loss function $\rho(e_K^Y, \widehat{e})$; then e_K^* is given by:

$$e_K^* = \arg \min_{\widehat{e} \in \Omega} E_{Y(\cdot)|\mathbf{Z}} [\rho(e_K^Y, \widehat{e})], \quad (3.12)$$

where $E_{Y(\cdot)|\mathbf{Z}}[\cdot]$ is expectation with respect to the posterior distribution of $Y(\cdot)$ given the observations \mathbf{Z} .

In what follows, we shall demonstrate our methodology the tapered version of Baddeley's metric for binary images as defined by (3.9). Within this metric, the two calibration quantities t and p account for both the site-wise misclassification error and the localization error, where "localization" refers to some important local features, such as clusters. As we have noted before, it is implicit in all these calculations that the parameters, ϕ , of the joint distribution are known and given. In practice, they are estimated (see Chapter 5), and in this sense our approach is empirical Bayesian rather than fully Bayesian.

3.3 Baddeley's loss function and the resulting predictor

We define the Baddeley's loss function as the square of the tapered version of Baddeley's metric for binary images, with $p = 2$. That is,

$$B_t(e, \widehat{e}) \equiv \frac{1}{|D|} \int_D (d_t(\mathbf{s}, e) - d_t(\mathbf{s}, \widehat{e}))^2 d\mathbf{s}, \quad (3.13)$$

where $e \in \Omega$ is the true exceedance set, $\widehat{e} \in \Omega$ is an action predicting the true exceedance set based on observations \mathbf{Z} , and $d_t(\mathbf{s}, e)$ is the tapered distance (3.10) with taper parameter t . Clearly, Baddeley's loss function (3.13) is a well defined, non-negative loss function. It penalizes the discrepancy between the true set and the

predicted set; the more different the two sets are, the larger the loss. The loss is equal to zero if and only if $\hat{e} = e$.

Suppose e_K^* is the optimal Bayes predictor based on Baddeley's loss function. We wish to predict the exceedance region e_K^Y given by (3.5). From (3.12),

$$e_K^* = \arg \min_{\hat{e} \in \Omega} E_{Y(\cdot)|\mathbf{Z}} [B_t(e_K^Y, \hat{e})] \quad (3.14)$$

$$= \frac{1}{|D|} \arg \min_{\hat{e} \in \Omega} \int_D E_{Y(\cdot)|\mathbf{Z}} \left[(d_t(\mathbf{s}, e_K^Y))^2 + d_t(\mathbf{s}, \hat{e}) [d_t(\mathbf{s}, \hat{e}) - 2d_t(\mathbf{s}, e_K^Y)] \right] d\mathbf{s} \quad (3.15)$$

$$= \frac{1}{|D|} \arg \min_{\hat{e} \in \Omega} \int_D \left\{ E_{Y(\cdot)|\mathbf{Z}} \left[(d_t(\mathbf{s}, e_K^Y))^2 \right] d\mathbf{s} + \int_D E_{Y(\cdot)|\mathbf{Z}} [d_t(\mathbf{s}, \hat{e}) [d_t(\mathbf{s}, \hat{e}) - 2d_t(\mathbf{s}, e_K^Y)]] d\mathbf{s} \right\} \quad (3.16)$$

$$= \arg \min_{\hat{e} \in \Omega} \frac{1}{|D|} \int_D E_{Y(\cdot)|\mathbf{Z}} [d_t(\mathbf{s}, \hat{e}) [d_t(\mathbf{s}, \hat{e}) - 2d_t(\mathbf{s}, e_K^Y)]] d\mathbf{s} \quad (3.17)$$

$$= \arg \min_{\hat{e} \in \Omega} V_t(\hat{e}; K), \quad (3.18)$$

where

$$V_t(\hat{e}; K) \equiv \frac{1}{|D|} \int_D \{ d_t(\mathbf{s}, \hat{e}) \cdot (d_t(\mathbf{s}, \hat{e}) - 2E_{Y(\cdot)|\mathbf{Z}} [d_t(\mathbf{s}, e_K^Y)]) \} d\mathbf{s}. \quad (3.19)$$

Here, equation (3.15) holds because of definition (3.13). Equation (3.17) is valid because $E_{Y(\cdot)|\mathbf{Z}} [(d_t(\mathbf{s}, e_K^Y))^2]$ does not depend on \hat{e} . Equation (3.18) is true because $d_t(\mathbf{s}, \hat{e})$ is deterministic with respect to the conditional distribution $p(Y(\cdot) | \mathbf{Z})$. Since the solution is not available in closed form, we shall use numerical optimization (namely simulated annealing) to obtain the predictor e_K^* .

Baddeley's loss function given by (3.13) takes into account not only the site-wise misclassification error, but also the localization error, by using the taper t . In Appendix A, some asymptotic calculations in simple 1-D and 2-D cases shows that as t

approaches zero, the leading term of Baddeley’s loss function is t^2 times the site-wise misclassification error in simple 1-D and 2-D cases. Intuitively, given a threshold K , if $Y(\cdot)$ is less dependent, elements in the exceedance region above K tend to be separated from each other over the domain of interest. However, if $Y(\cdot)$ is highly dependent, elements in the exceedance region tend to be clustered together for the same threshold K . As a special case, consider independent $Y(\cdot)$ in D . Here it is plausible to use the site-wise misclassification error as a loss function to predict the exceedance region, since the prediction cannot borrow strength from other independent locations. Therefore, we expect to use a small taper in Baddeley’s loss function to predict the exceedance region for less dependent geostatistical processes. On the other hand, we need to use a large taper to predict the exceedance region for highly dependent geostatistical processes, because we expect to borrow strength from other regions of D . With this in mind, we suggest that the taper be chosen proportional to the range parameter r . Later in this chapter, I will use simulations based on different geostatistical processes and different thresholds, to justify the recommended choice of taper,

$$t = \min(r/2, \xi). \tag{3.20}$$

3.4 Other approaches to detect hotspots

There are other approaches to detecting spatial hotspots. One example is the upper level set (ULS) scan statistics by Patil and Taillie (2004).

Suppose that D is the field of interest and is discretized to a collection of sub-regions, defined by $D_a \equiv \{a_1, \dots, a_m\}$. Let Ω^* denote the set of all possible combinations of different sub-regions. In addition, suppose that there is a known “size”

parameter for each sub-region indexed by $a \in D_a$. For example, the number of people in the sub-region a , N_a . Denote a RV of interest in a sub-region a , such as count, by Y_a . Then the ULS method first models Y_a to be independently distributed as a binomial distribution with unknown success probability parameters $p_a \in [0, 1]$; that is, $Y_a \sim \text{Binomial}(N_a, p_a)$. Then the unknown hotspot $e \subseteq D_d$ consists of sub-regions with significant large values for p_a . Suppose that $p_a = p_1$ for all sub-regions in e , $p_a = p_0$ for all sub-regions not in e , and $p_0 < p_1$ for these two constants. If we consider e to be an unknown parameter, then the profile likelihood for e is given by

$$L(e) = \max_{p_0, p_1 \in [0, 1]} L(e, p_0, p_1) = L(e, \hat{p}_0, \hat{p}_1),$$

where \hat{p}_0 and \hat{p}_1 are the maximum likelihood estimates of p_0 and p_1 , respectively. Finally, the hotspot estimation requires to maximize $L(e)$ as e varies over Ω^* . However, maximizing the profile likelihood for the unknown hotspot is troublesome because the search space Ω^* is the configuration space of all the sub-regions in the field. Although Ω^* is a finite set, it is generally so large that maximizing the profile likelihood by an exhaustive search is computationally expensive.

The ULS scan statistic is an adaptive approach in which Ω^* is reduced to Ω_{ULS} by using the scan statistic G_a , for example, $G_a = Y_a/A_a$ for the success probability parameter in the Binomial model. For a given level g , an upper level set (ULS) based on the scan statistic is $U_g = \{a : G_a \geq g\}$, where the total number of levels, g , is finite. Then, Ω_{ULS} is the collection of all possible upper level set. The maximization of the profile likelihood for the hotspot is then done in the reduced space. The reduction is plausible if the reduced search space contains the MLE for the hotspot in the full search space because the local MLE is also the global MLE in this case.

In the problem of predicting the exceedance region of a geostatistical process, the original observations \mathbf{Z} are obviously not independent. Neither is the process $Y(\cdot)$ or the 0-1 value that indicates exceedance or not for each spatial location (see the definition of the exceedance region in (3.5)). If one does want to use the ULS scan statistics in this case, the scan statistic for $Y(\cdot)$ can not be guaranteed to be available for all the sites in D because \mathbf{Z} may not cover part of the field sometimes.

3.4.1 Computation of the predictor via simulated annealing

Consider the general goal of finding the $\boldsymbol{\theta} \in \Theta$ that minimizes the objective function, $h(\boldsymbol{\theta})$. We shall use simulated annealing to achieve this goal numerically (e.g., Robert and Casella, 2004). Consider a probability distribution of the form,

$$p_T(\boldsymbol{\theta}) \equiv (c(T))^{-1} \exp[-h(\boldsymbol{\theta})/T], \quad (3.21)$$

where $c(T)$ is a normalizing constant. The parameter $T > 0$ is called the “temperature” of the system. As $T \rightarrow \infty$, $p_T(\cdot)$ approaches a uniform distribution on Θ . On the other hand, when $T \rightarrow 0$, $p_T(\cdot)$ becomes concentrated on the global minimum of $h(\boldsymbol{\theta})$. If there are multiple global minima, $p_T(\cdot)$ approaches a uniform distribution over these multiple minima, as $T \rightarrow 0$.

Simulated annealing is a sequential procedure made up of a Metropolis sampler and an annealing schedule. The Metropolis sampler eventually simulates realizations with target probability distribution $p_T(\cdot)$ given in (3.21). Let $\hat{\boldsymbol{\theta}}_i \in \Theta$ denote the state of the samples at iteration i . We propose a new candidate state, $\boldsymbol{\zeta} \in \Theta$, where $\boldsymbol{\zeta}$ is chosen according to a proposal distribution; for example, $\boldsymbol{\zeta}$ is chosen uniformly in a neighborhood of $\hat{\boldsymbol{\theta}}_i$. We then choose $\hat{\boldsymbol{\theta}}_{i+1} = \boldsymbol{\zeta}$, with probability $\tau = \min \left\{ 1, p_T(\boldsymbol{\zeta})/p_T(\hat{\boldsymbol{\theta}}_i) \right\}$, or keep $\hat{\boldsymbol{\theta}}_{i+1} = \hat{\boldsymbol{\theta}}_i$, with probability $1 - \tau$. After canceling

out the normalizing constant $c(T)$ and the probability of symmetric proposals from a uniform distribution, the expression for the probability τ can be simplified to $\tau = \min \left\{ 1, \exp \left[- \left(h(\boldsymbol{\zeta}) - h(\widehat{\boldsymbol{\theta}}_i) \right) / T \right] \right\}$.

The annealing schedule is simply the rate at which T , which we write as $T(i)$ to emphasize the iteration number, approaches zero as $i \rightarrow \infty$. Under certain assumptions on $T(i)$, the simulated-annealing algorithm is guaranteed to find the global minimum (Geman and Geman, 1984). For example, the annealing schedule,

$$T(i) = \frac{T_0}{\log(1+i)}; \quad i = 1, 2, \dots, \quad (3.22)$$

where T_0 is a tuning constant, yields the global minimum. This is called logarithmic cooling. Fast convergence can be achieved with linear cooling $T(i) = T_0/(1+i)$, for $i = 1, 2, \dots$. However, for linear cooling the solution may not be the global minimum; in our context, results for linear cooling are still reasonable according to the simulation study in Section 3.5.

In general cases, when the objective function we want to minimize or maximize involves expressions without analytic forms, a Monte Carlo algorithm within simulated annealing might be necessary to evaluate certain expressions numerically. For example, the objective function that we want to minimize for the Bayes rule, e_K^* , is $V_t(\widehat{e}; K)$ defined by (3.19); here we may need a simulation procedure (e.g., Markov Chain Monte Carlo, or MCMC) to obtain the posterior expectation, $E_{Y(\cdot)|\mathbf{z}}[d_t(\mathbf{s}, e_K^Y)]$. This can be simplified for models given by (3.1) and (3.2), in which both the true process $Y(\cdot)$, and the measurement-error process $\epsilon(\cdot)$ are Gaussian. Suppose that we can always discretize D to a set of locations $D^* \equiv \{\mathbf{s}_1^*, \dots, \mathbf{s}_m^*\}$ that cover D well. Therefore, let $\mathbf{Y} \equiv (Y(\mathbf{s}_1^*), \dots, Y(\mathbf{s}_m^*))^T$ denote a vector of the corresponding hidden values at $\{\mathbf{s}_1^*, \dots, \mathbf{s}_m^*\}$, and $\mathbf{X}_m \equiv (\mathbf{x}(\mathbf{s}_1^*), \dots, \mathbf{x}(\mathbf{s}_m^*))$ denote the $p \times m$ design matrix

for predictions over D^* . Now, the conditional distribution $p(\mathbf{Y} | \mathbf{Z})$, has a known multivariate Gaussian distribution that we can exploit. Recall $C_{\boldsymbol{\theta}_Y}(\cdot)$ is the covariance function of $Y(\cdot)$, where $\boldsymbol{\theta}_Y$ is the vector of covariance parameters. By a well known property of the multivariate Gaussian distribution, the conditional distribution of interest, $p(\mathbf{Y} | \mathbf{Z})$, is $N_m(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, with

$$\boldsymbol{\mu}^* = (\mathbf{X}_m)^T \boldsymbol{\beta} - \mathbf{F}^* \boldsymbol{\Sigma}^{-1} (\mathbf{X}^T \boldsymbol{\beta} - \mathbf{Z}), \text{ and } \boldsymbol{\Sigma}^* = \mathbf{C}^* - \mathbf{F}^* \boldsymbol{\Sigma}^{-1} (\mathbf{F}^*)^T, \quad (3.23)$$

where $\mathbf{X} \equiv (\mathbf{x}(\mathbf{s}_1), \dots, \mathbf{x}(\mathbf{s}_n))$ is a $p \times n$ design matrix for the observations; \mathbf{F}^* is an $m \times n$ matrix with (i, j) -element $\text{cov}(Y(\mathbf{s}_i^*), Z(\mathbf{s}_j)) = \text{cov}(Y(\mathbf{s}_i^*), Y(\mathbf{s}_j)) = C_{\boldsymbol{\theta}_Y}(\|\mathbf{s}_i^* - \mathbf{s}_j\|)$; \mathbf{C}^* is an $m \times m$ matrix with (i, j) -element $\text{cov}(Y(\mathbf{s}_i^*), Y(\mathbf{s}_j^*)) = C_{\boldsymbol{\theta}_Y}(\|\mathbf{s}_i^* - \mathbf{s}_j^*\|)$; and $\boldsymbol{\Sigma}$ is an $n \times n$ matrix with (i, j) -element $\text{cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = C_{\boldsymbol{\theta}_Y}(\|\mathbf{s}_i - \mathbf{s}_j\|)$.

We now describe the simulated annealing algorithm that yields to optimal Bayes predictor based on Baddeley's loss function given by (3.17). Recall that the geostatistical model includes a data model given by (3.1) and a process model given by (3.2). Suppose that $\boldsymbol{\phi} \equiv (\boldsymbol{\beta}^T, \sigma_\epsilon^2, \boldsymbol{\theta}_Y^T)^T$ are already estimated using $\hat{\boldsymbol{\phi}} \equiv (\hat{\boldsymbol{\beta}}^T, \hat{\sigma}_\epsilon^2, \hat{\boldsymbol{\theta}}_Y^T)^T$. Then we can directly generate samples from $p(\mathbf{Y} | \mathbf{Z})$ based on the multivariate Gaussian distribution given in (3.23). Suppose that there are J realizations sampled from $p(\mathbf{Y} | \mathbf{Z})$. Then we can easily approximate $E_{Y(\cdot) | \mathbf{Z}} [d_t(\mathbf{s}, e_K^Y)]$ in $V_t(\hat{e}; K)$ defined by (3.19), by the average of the Monte Carlo samples from $E_{\mathbf{Y} | \mathbf{Z}} [d_t(\mathbf{s}, e_K^Y)]$. If we denote the approximation by $\hat{\zeta}_{\mathbf{s}}$, then

$$\hat{\zeta}_{\mathbf{s}} = \frac{1}{J} \sum_{j=1}^J d_t(\mathbf{s}, e_K^{Y^{(j)}}),$$

where $\mathbf{Y}^{(j)}$ is the j th realization from the posterior distribution $p(\mathbf{Y} | \mathbf{Z}, \hat{\boldsymbol{\phi}})$. Then, we do simulated annealing to solve for e^* with $E_{Y(\cdot) | \mathbf{Z}} [d_t(\mathbf{s}, e_K^Y)]$ replaced by $\hat{\zeta}_{\mathbf{s}}$ in (3.17). The actual algorithm to solve for e_K^* , is as follows:

1. Start with an initial exceedance set $\widehat{e}_K^{(0)}$. (Some suggestions for this starting value will be given below.)
2. Suppose that at the i th stage, the current exceedance set is $\widehat{e}_K^{(i)}$. Then propose a new exceedance set \tilde{e}_K by adding a uniformly-randomly selected site $\mathbf{s}_{(i)}$ in D to $\widehat{e}_K^{(i)}$, or dropping $\mathbf{s}_{(i)}$ from $\widehat{e}_K^{(i)}$: if $\mathbf{s}_{(i)}$ is excluded from $\widehat{e}_K^{(i)}$ in previous stage, \mathbf{s}_r is added to $\widehat{e}_K^{(i)}$ for \tilde{e}_K ; if $\mathbf{s}_{(i)}$ is included in $\widehat{e}_K^{(i)}$, $\mathbf{s}_{(i)}$ is dropped from $\widehat{e}_K^{(i)}$ for \tilde{e}_K .
3. Accept the proposed $\widehat{e}^{(i+1)} = \tilde{e}$ with probability

$$\min \left\{ 1, \exp \left[-\frac{1}{T(i)} (V_t(\tilde{e}; K) - V(\widehat{e}^{(i)}; K)) \right] \right\}.$$

4. Repeat previous steps until $\widehat{e}_K^{(i)}$ converges.

We also considered the simulated annealing with particles (Amzal et al., 2006). The idea is to start multiple chains that are called particles. Each particle updates one step just as the classical simulated annealing does. Then all the updated particles are re-sampled based on their corresponding sampling weights. Those particles with smaller losses being minimized have larger weights, which results in particles that are eventually closer to the global optimum of interest. In simulated annealing with particles, the cooling schedule is usually linear. However, the solution of simulated annealing with particles was comparable with the classical simulated annealing when both methods have a same number of iteration steps.

3.5 Simulation study

In this section, we evaluate the performance of the Baddeley-loss-based predictor of the exceedance region in the case of a fixed threshold K , as defined by 3.5. Recall

that the predictor of the exceedance region, e_K^* , defined by (3.14), depends on the taper size, t . We begin this section, by verifying the rule for the choice of the taper as suggested by (3.20), on the basis of a Monte Carlo experiment. We then follow up with comparison of the Baddeley-loss-based predictor of the exceedance region, with the often-used posterior-mean-based, plug-in predictor.

Recall that we have observations \mathbf{Z} at a fixed set B of locations in the domain D . For any location $\mathbf{s} \in D$, the posterior-mean predictor of the hidden process, $Y(\mathbf{s})$ is $\hat{Y}(\mathbf{s}) \equiv E_{Y(\mathbf{s})|\mathbf{Z}}[Y(\mathbf{s})]$. The plug-in predictor of the exceedance region above a fixed threshold K , based on this predictor of $Y(\mathbf{s})$, is then defined by

$$\hat{e}_K \equiv \left\{ \mathbf{s} \in D : \hat{Y}(\mathbf{s}) > K \right\}. \quad (3.25)$$

3.5.1 Calibration of the taper t

We consider a unit square domain $D \equiv \{\mathbf{s} = (x, y) : 0 \leq x \leq 1; 0 \leq y \leq 1\}$. We conduct the simulation and evaluate each predictor on the gridded domain $D^* \equiv \{\mathbf{s}_i^* = (x, y) : x = 0 = \frac{0}{M-1}, \frac{1}{M-1}, \dots, 1; y = 0 = \frac{0}{M-1}, \frac{1}{M-1}, \dots, 1\}$ for $M = 20$. Assuming no spatial trend in (3.2) (i.e., $\boldsymbol{\beta} = \mathbf{0}$), we let the hidden process $Y(\cdot)$ be an isotropic mean-zero Gaussian process with an exponential covariance function. This process is characterized solely by its variance σ^2 and by the spatial range parameter r . In our experiments, we let $\sigma^2 = 1$, and vary r among values 0.5 (corresponding to weak spatial dependence), 1.5 (moderate spatial dependence), and 5 (very strong spatial dependence). Using a Cholesky decomposition for the covariance matrix, we simulate 20 realizations of $\mathbf{Y} = \{Y(\mathbf{s}^*) : \mathbf{s}^* \in D^*\}$ for each of the three levels (weak, moderate, strong) for the spatial dependence parameter. To each $Y(\mathbf{s}^*)$, we add 100 independent Gaussian measurement error, $N(0, \sigma_\epsilon^2)$, to obtain the data process. We

then randomly sample 1/4 of all sites in D^* as the collection of observation sites B for experiments, and henceforth fix B . The corresponding data then become observations \mathbf{Z} . We consider two levels of measurement-error variance: $\sigma_\epsilon^2 = 0.5$ and $\sigma_\epsilon^2 = 2$. Defining the signal-to-noise ratio (SNR) as the ratio of the partial sill in the hidden process to the measurement-error variance (i.e., $\text{SNR} \equiv \sigma^2/\sigma_\epsilon^2$), this corresponds to SNRs of 2 and 0.5. The threshold $K = F_D^{-1}(\alpha)$ is fixed as the α the quantile of the ACDF of $Y(\cdot)$, with $\alpha = 0.5, 0.75, 0.9$. Notice that when $\beta = \mathbf{0}$ and $\sigma^2 = 1$, $F_D(\cdot)$ is the standard normal CDF, $\Phi(\cdot)$. Hence, $K = 0.00, 0.67, 1.28$. In summary, the factors for our 3-factor design are listed in Table 3.1. For each factor combination, there are 20 replicates of $Y(\cdot)$, and within each replicate there are 100 replicates of measurement errors. When defining the response to the simulation experiment, we average over the 100 replications of measurement error and display the 20 responses corresponding to the replicates of $Y(\cdot)$.

Factor	Number of levels	Values of levels
r	3	0.5
		1.5
		5
SNR	2	2
		0.5
α	3	0.50
		0.75
		0.90

Table 3.1: Levels of factors in simulation studies for the single-loss approach.

We define *relative efficiency* of e_K^* to \hat{e}_K to verify the choice of taper given by (3.20). Let $RE_B(e_K^*, \hat{e}_K; t)$ denote the relative efficiency with taper t , defined by,

$$RE_B(e_K^*, \hat{e}_K; t) \equiv \frac{1/E[B_t(e_K^*, e_K^Y)]}{1/E[B_t(\hat{e}_K, e_K^Y)]} = \frac{E[B_t(\hat{e}_K, e_K^Y)]}{E[B_t(e_K^*, e_K^Y)]}, \quad (3.26)$$

where $E[\cdot]$ is the expectation with respect to \mathbf{Z} and $Y(\cdot)$. Then in the Monte Carlo experiment conducted in D^* , the expectations with respect to \mathbf{Z} and $Y(\cdot)$ are approximated by the corresponding averages over realizations of the measurement error and realizations of \mathbf{Y} .

The relative efficiency, RE_B , is the ratio of the mean Baddeley's loss function for one predictor to another predictor, in which both predictors are compared to the true exceedance region. Here $RE_B(e_K^*, \hat{e}_K)$ is larger than 1 if and only if predictor e_K^* has smaller average Baddeley's loss function compared to the true exceedance region than the predictor \hat{e}_K has. We expect to see a smaller averaged Baddeley's loss function for e_K^* than for \hat{e}_K because \hat{e}_K plugs in underdispersed predictions, while e_K^* is the optimal predictor of the exceedance set in terms of Baddeley's loss function.

Suppose for each experiment that there are a total of n_Y realizations for \mathbf{Y} and n_ϵ realizations of the measurement error. Then for each experiment, equation (3.26) can be approximated by the sample average

$$\widetilde{RE}_B(e_K^*, \hat{e}_K; t) = \frac{\sum_{i=1}^{n_Y} \sum_{j=1}^{n_\epsilon} B_t(\hat{e}_K^{(i,j)}, e_K^{Y(i)})}{\sum_{i=1}^{n_Y} \sum_{j=1}^{n_\epsilon} B_t(e_K^{*(i,j)}, e_K^{Y(i)})}, \quad (3.27)$$

where $e_K^{Y(i)}$ is the true exceedance set defined by (3.5) based on the j th sample of \mathbf{Y} , while $\hat{e}_K^{(i,j)}$ and $e_K^{*(i,j)}$ are corresponding predictors of \hat{e}_K defined by (3.25) and e_K^* defined by (3.14), based on the i th sample of \mathbf{Y} and j th sample of the measurement error.

In Figure 3.1, we plot the sample-averaged relative efficiency, $\widetilde{RE}_B(e^*, \widehat{e}_K; t)$ (the solid line), along with the 95% confidence interval (the dashed line), against t , for simulation experiments over all factor combinations (2 levels of SNR, 3 levels of range parameter, r , and 3 levels of thresholds, K). From Figure 3.1, it is not surprising to see that all relative efficiency curves are above 1, since on average, the Baddeley-loss-based predictor, e^* , for the exceedance region minimizes the Bayes risk of the Baddeley's loss function, while the posterior-mean-based, plug-in predictor, \widehat{e}_K , is not the optimal Bayes predictor for Baddeley's loss function. Based on the relationship between the desirable taper t and the spatial dependence of the true process $Y(\cdot)$, shown in Figure 3.1, we choose the taper t that has the large and stable relative efficiency. Recall that the spatial dependence of the true process $Y(\cdot)$ is only determined by the range parameter, r , in all simulation experiments. For the weakly ($r = 0.5$), and the moderately ($r = 1.5$) dependent processes, the desirable taper appears to be approximately half of the range parameter; for the strongly dependent ($r = 5$), the desirable taper appears to be approximately equal to $\xi = \sqrt{2}$, the maximum of the pairwise distance of D . For a given range, this choice of t remains same across different combinations of other two factors, the SNR and the threshold, K . In Table 3.2, we list the choice of t for all combinations of experiment factors. Henceforth, we use the rule, $t = \min(r/2, \xi)$, for the choice of t .

3.5.2 Performance comparison

After verifying the rule for the choice of taper, we then conduct a simulation study to compare the performance of Baddeley-loss-based predictor, e_K^* , with the posterior-mean-based, plug-in predictor, \widehat{e}_K .

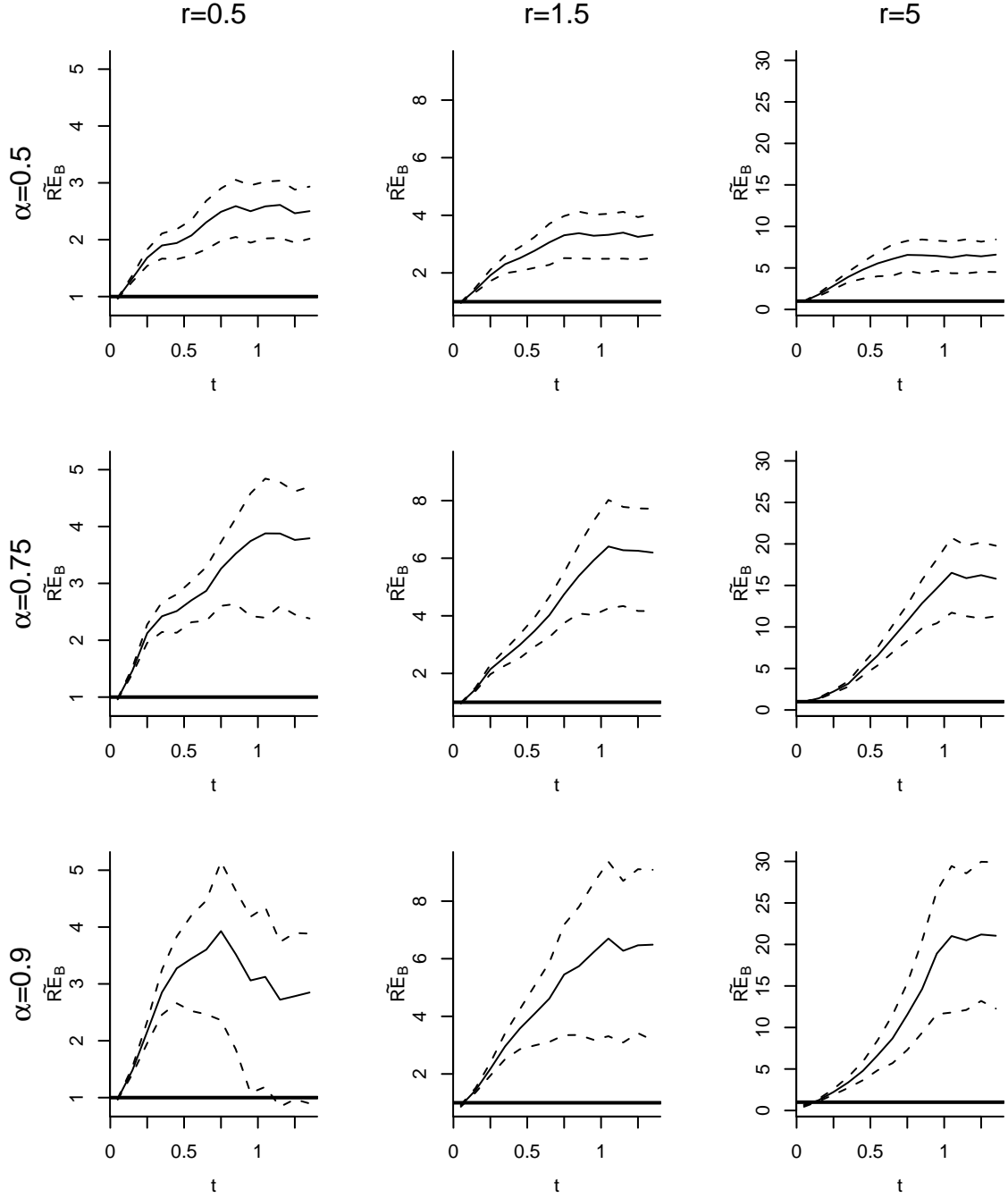


Figure 3.1: The relative efficiency of the Baddeley-loss-based predictor and the posterior-mean-based predictor, for different tapers, range parameters, r , and values of α (used to set the threshold), for $\text{SNR} = 2.0$. The solid line is for the sample-averaged relative efficiency. The dashed line is for the 95% confidence interval for the sample-averaged relative efficiency.

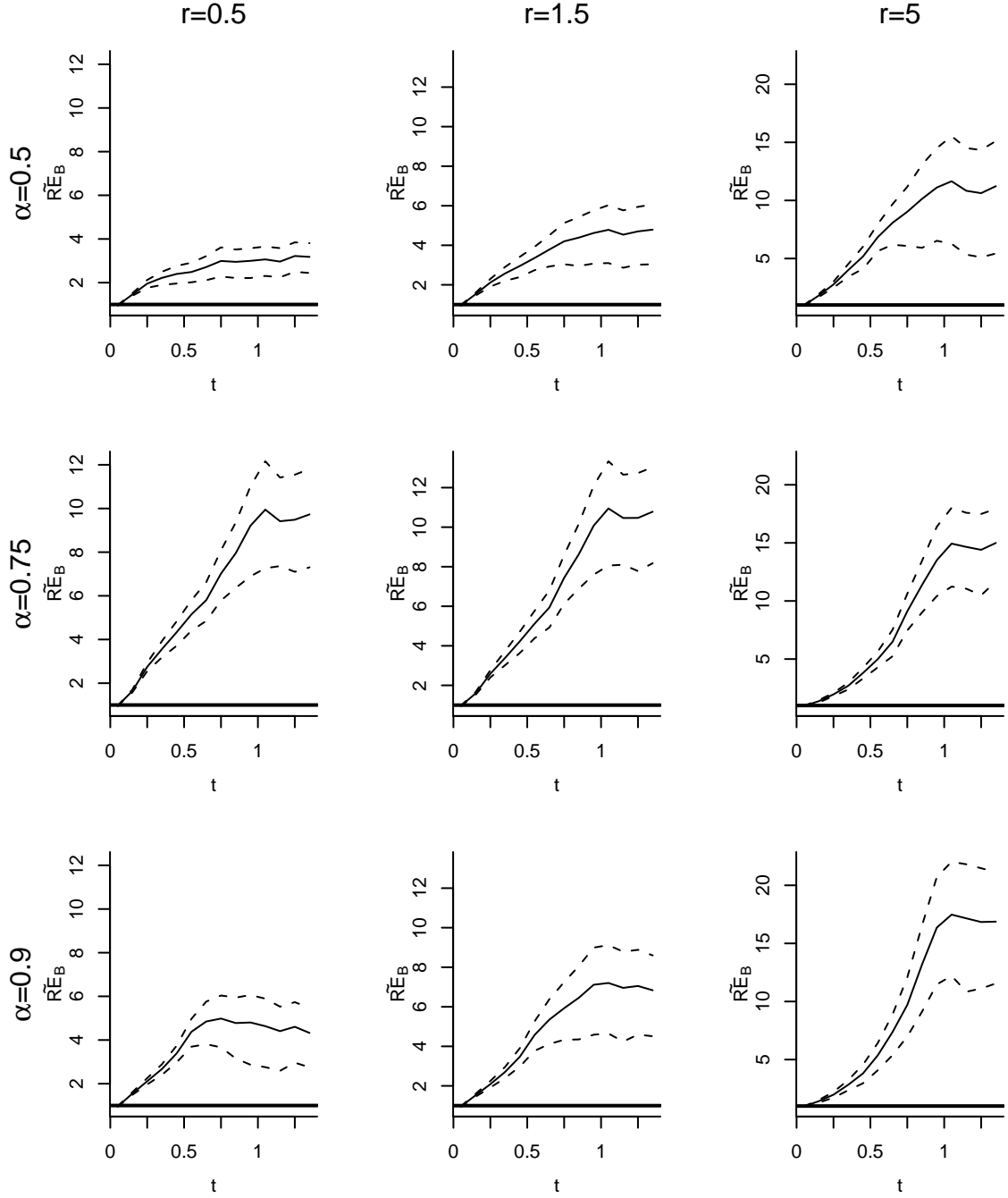


Figure 3.2: The relative efficiency of the Baddeley-loss-based predictor and the posterior-mean-based predictor, for different tapers, range parameters, r , and values of α (used to set the threshold), for $\text{SNR} = 0.5$. The solid line is for the sample-averaged relative efficiency. The dashed line is for the 95% confidence interval for the sample-averaged relative efficiency.

SNR	r	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 0.9$
2	0.5	0.25	0.25	0.25
	1.5	0.75	0.75	0.75
	5	1.25	1.25	1.25
0.5	0.5	0.25	0.25	0.25
	1.5	0.75	0.75	0.75
	5	1.25	1.25	1.25

Table 3.2: The choice of taper, t , in the Baddeley’s loss function based on the relative efficiency, with the grid size, $\omega = 1$, and the maximum pairwise distance, $\xi = \sqrt{2}$.

The domain of interest is the same unit square domain D that we discretized to D^* in Section 3.5.1. We let the model parameters have same values as the simulation in Section 3.5.1; that is, $\boldsymbol{\beta} = \mathbf{0}$, $\sigma^2 = 1$, and r varies among values $r = 0.5$ (corresponding to weak spatial dependence), $r = 1.5$ (moderate spatial dependence), and $r = 5$ (very strong spatial dependence). We also use a Cholesky decomposition for the covariance matrix, and simulate 20 realizations of $\mathbf{Y} = \{Y(\mathbf{s}^*) : \mathbf{s}^* \in D^*\}$ for each of the three levels (weak, moderate, strong) for the spatial dependence parameter; to each $Y(\mathbf{s}^*)$, we add 100 independent Gaussian measurement error, $N(0, \sigma_\epsilon^2)$, to obtain the data process. We then randomly sample 1/4 of all sites in D^* as the collection of observation sites B for experiments, and henceforth fix B . The corresponding data then become observations \mathbf{Z} . We consider two levels of measurement-error variance: $\sigma_\epsilon^2 = 0.5$ and $\sigma_\epsilon^2 = 2$. This corresponds to SNRs of 2 and 0.5. The threshold $K = F_D^{-1}(\alpha)$ is fixed as the α th quantile of the ACDF of $Y(\cdot)$, with $\alpha = 0.5, 0.75, 0.9$. Hence, $K = 0.00, 0.67, 1.28$. In summary, the factors are the same as in Table 3.1.

Notice that in the simulation for the performance comparison, for each factor combination there are 20 replicates of $Y(\cdot)$, and within each replicate there are 100 replicates of measurement errors. When defining the response to the simulation experiment for the performance comparison, we average over the 100 replications of measurement errors.

Suppose that e is the true exceedance region, and \hat{e} is a predicted exceedance region. Let $Se(e, \hat{e})$ denote a *sensitivity measure*, and $Sp(e, \hat{e})$ denote a *specificity measure* of the predictor \hat{e} , defined by:

$$Se(e, \hat{e}) = \frac{|e \cap \hat{e}|}{|\hat{e}|}, \quad \text{and} \quad Sp(e, \hat{e}) = \frac{|(D \setminus e) \cap (D \setminus \hat{e})|}{|D \setminus \hat{e}|}. \quad (3.28)$$

Then, $Se(e, \hat{e})$ calculates the area-proportion of correctly predicted exceedance points (“true positives”), while $Sp(e, \hat{e})$ calculates the area-proportion of correctly predicted non-exceedance points (“true negatives”).

Figures 3.3 shows boxplots of the *difference* of the sensitivity and specificity measures for the Baddeley-loss-based predictor (e_K^*) and the posterior-mean-based, plug-in predictor (\hat{e}_K). We observe that for the weakly ($r = 0.5$), moderately ($r = 1.5$) and strongly ($r = 5$) dependent geostatistical processes, e_K^* is more likely to predict the same number or more sites correctly for the true exceedance regions than for \hat{e}_K for all thresholds. The Baddeley-loss-based predictor e_K^* also tends to predict the same number or more sites correctly for the true non-exceedance regions than \hat{e}_K , except for the case of $\alpha = 0.5$, $r = 0.5$, and SNR=2, where \hat{e}_K is considered to be good, since $Y(\cdot)$ is less dependent and relative strong to the error process, and the threshold is close to the center of the distribution of $Y(\cdot)$.

We conducted a similar simulation experiment in a finer domain with $M = 40$. As shown in Figure 3.4, the Baddeley-loss-based predictor e_k^* performs similarly to

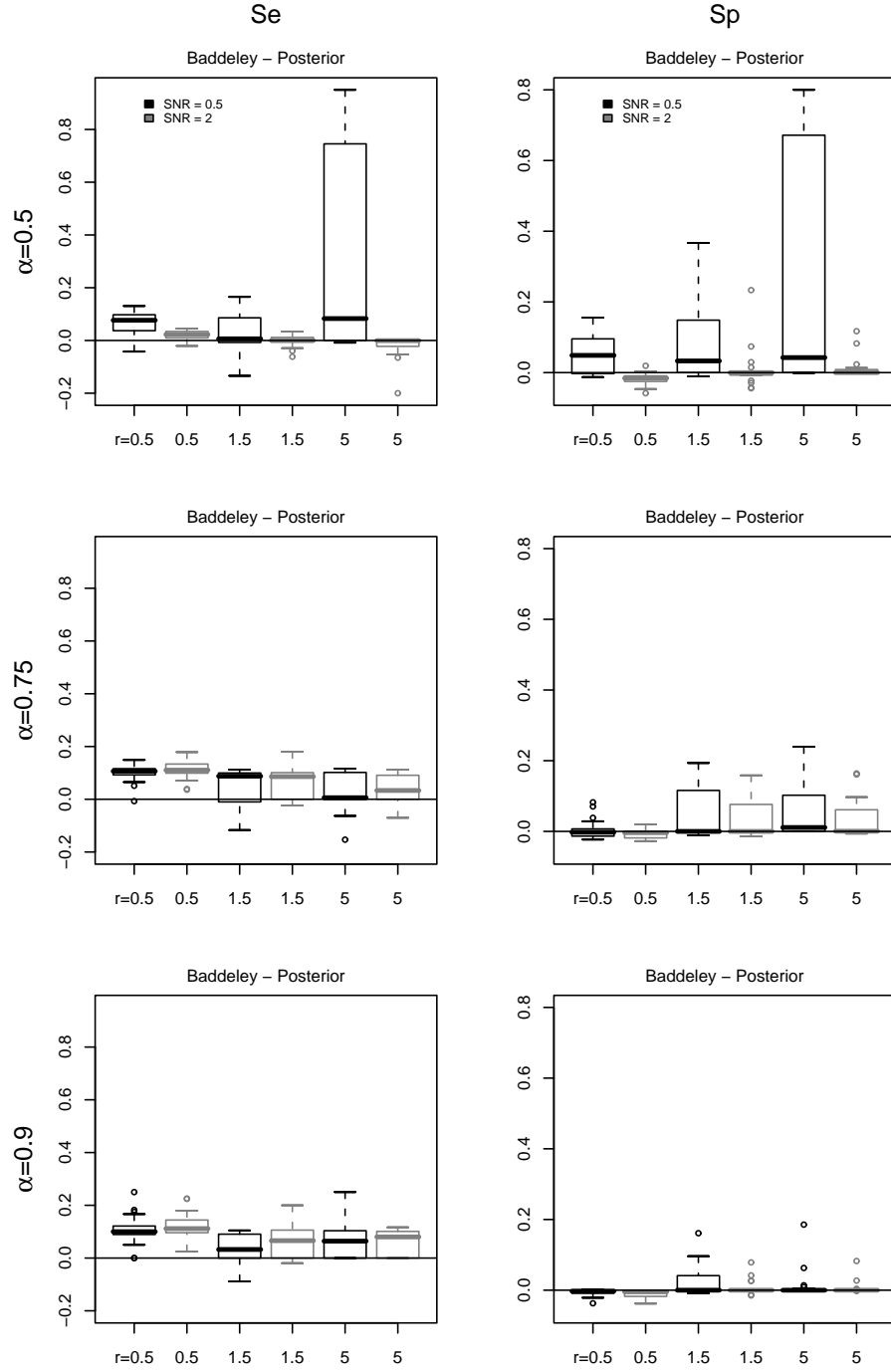


Figure 3.3: Boxplots of difference of the sensitivity and specificity between the Baddeley-loss-based predictor and the posterior-mean-based predictor, when the grid size is $M = 20$.

\hat{e}_K for the threshold close to the center of the distribution ($\alpha = 0.5$). For higher thresholds, we observe consistently similar results as we saw for $M = 20$.

3.6 Discussion

In this chapter, we introduce image metrics as loss functions for the prediction of an exceedance region above a known threshold. We then propose a framework to solve for the optimal Bayes predictor for the exceedance region with a general image metrics. We demonstrate our methodology with special attention to the Baddeley's loss function since the Baddeley's loss function is well suited for discriminating between local and global structures in the spatial exceedance region.

To obtain the Baddeley-loss-based predictor for the exceedance region, numerical solution (simulated annealing) is required. It is not surprising that the process of finding optimal predictors with squared error loss or weighted squared error loss is much less computationally intensive than those with Baddeley's loss because of the possibility of componentwise optimizations in squared error loss or weighted squared error loss. The extra cost for the Baddeley-loss-based predictor, involves the iteration number for simulated-annealing solution. Suppose there are m prediction sites and n_T iterations for simulated-annealing solution. Then the extra computation cost is $O(m^2 n_T)$.

In practice, we need to pre-estimate the spatial trend and covariance parameters. We then plug in these estimates to predict the exceedance region. It is empirical Bayesian approach, not hierarchical Bayesian. Throughout, we assume that the threshold K is given. In practice, this is not always true. In next chapter, we will use a joint loss approach to predict the threshold and exceedance region simultaneously.

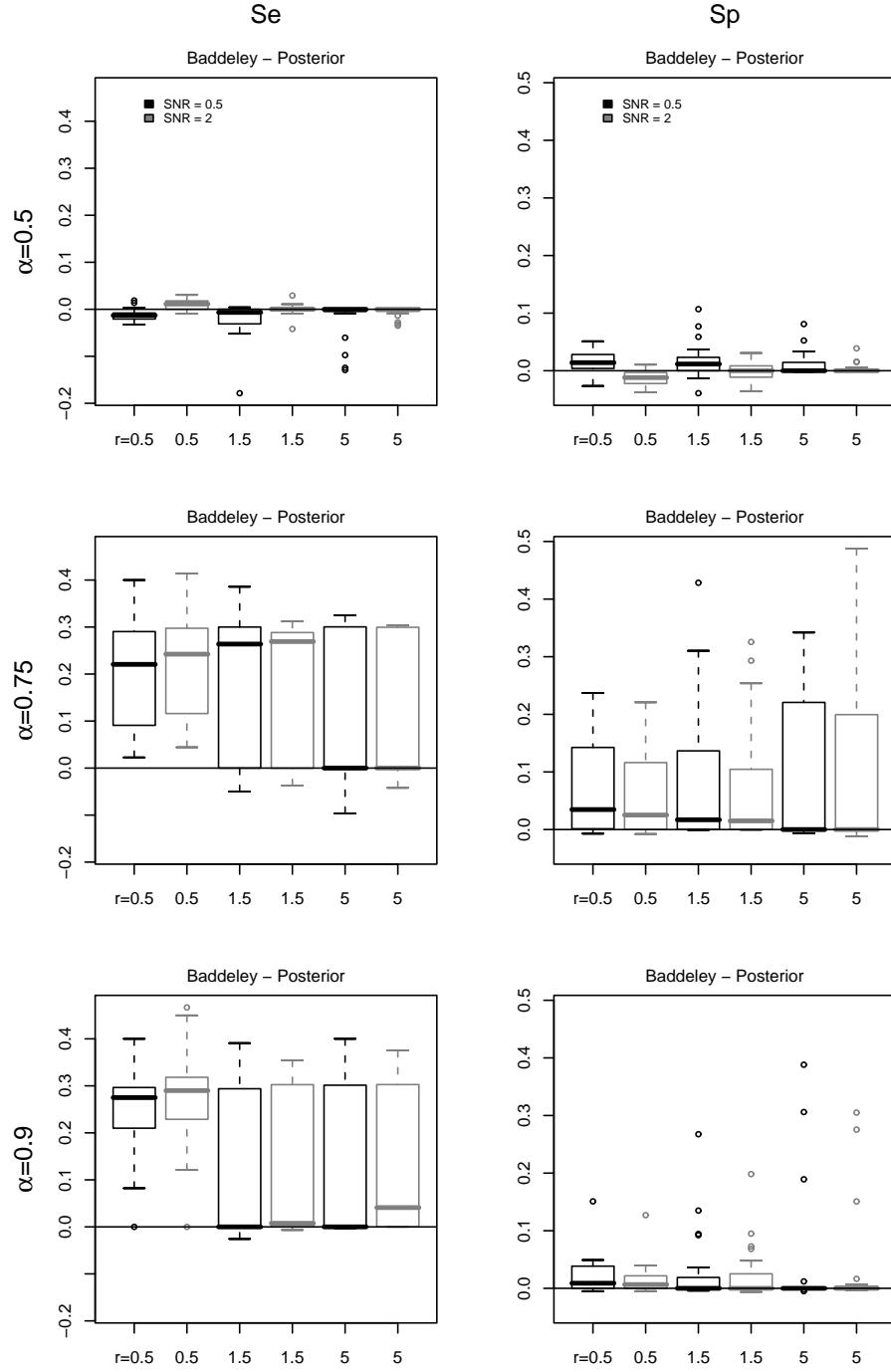


Figure 3.4: Boxplots of difference of the sensitivity and specificity between the Baddeley-loss-based predictor and the posterior-mean-based predictor, for a grid size of $M = 40$.

CHAPTER 4

JOINT PREDICTION OF A SPATIAL QUANTILE AND ITS EXCEEDANCE REGION

In Chapter 3, we introduced image metrics as loss functions to predict the spatial exceedance region for a given threshold K . The exceedance region e_K^Y in (3.5) is defined as those locations above a fixed threshold K . However, it is not always the case that the threshold K is known, such as when a spatial quantile is used as the threshold. In this section, we consider the joint prediction of a spatial quantile (i.e., a quantile of the *spatial cumulative distribution function* (SCDF), as defined by Lahiri et al., 1999), and its associated exceedance region.

Before presenting the approach for joint prediction, we shall briefly review the spatial model of interest. (It is the same model that was used in Chapters 2 and 3). Recall that $Y(\cdot) \equiv \{Y(\mathbf{s}) : \mathbf{s} \in D\}$ is the hidden geostatistical process of interest, defined on $D \subset \mathbb{R}^d$, with a positive d -dimensional volume, $|D|$. Let $Z(\cdot) \equiv \{Z(\mathbf{s}) : \mathbf{s} \in D\}$ be the process of observed and potentially observed values. Then the hierarchical model can be described by the following data and process models.

The data model:

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \epsilon(\mathbf{s}); \quad \mathbf{s} \in D, \quad (4.1)$$

where $\epsilon(\cdot)$ is a zero-mean, white noise process with measurement-error variance, $\sigma_\epsilon^2 > 0$. We assume that $\epsilon(\cdot)$ is independent of $Y(\cdot)$.

The process model:

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}); \quad \mathbf{s} \in D. \quad (4.2)$$

Here the function $\mu(\cdot)$ captures the large-scale, deterministic mean structure (*spatial trend*) of $Y(\cdot)$, and $\delta(\cdot)$ is a stationary stochastic process that models the small-scale spatial structure in $Y(\cdot)$. We assume that the spatial trend, $\mu(\cdot)$, follows a linear model given by

$$\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})' \boldsymbol{\beta}, \quad (4.3)$$

where the vector $\mathbf{x}(\mathbf{s}) \equiv (x_1(\mathbf{s}), \dots, x_p(\mathbf{s}))^T$ is a collection of covariates at \mathbf{s} , and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the *spatial-trend parameter*. We also assume that $\delta(\cdot)$ is a zero-mean, second-order stationary spatial process with a parametric covariance function, $C_{\boldsymbol{\theta}_Y}(\cdot)$. Specifically,

$$E[\delta(\mathbf{s})] = 0, \quad \text{and } \text{cov}[\delta(\mathbf{s}), \delta(\mathbf{s}')] \equiv C_{\boldsymbol{\theta}_Y}(\mathbf{s} - \mathbf{s}'); \quad \mathbf{s}, \mathbf{s}' \in D, \quad (4.4)$$

where $\boldsymbol{\theta}_Y \equiv (\sigma^2, r)^T$ is the *spatial-covariance parameter* that consists of the *partial sill*, σ^2 , and the *range parameter*, r . In this chapter, we assume that $\boldsymbol{\theta}_Y$ and σ_ϵ^2 are known, although in practice they are estimated from the data.

Suppose henceforth that we have n observations $\mathbf{Z} \equiv \{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\}$ at a collection of observation sites, $B \equiv \{\mathbf{s}_1, \dots, \mathbf{s}_n\} \subset D$. The problem of interest in this chapter is to use the data \mathbf{Z} to predict a spatial quantile and its associated exceedance region of the true process $Y(\cdot)$.

4.1 Joint loss functions and the resulting predictor

Recall that $S_D(y; Y(\cdot))$ is the spatial cumulative distribution function (SCDF) of $Y(\cdot)$ on the domain D ; that is,

$$S_D(y; Y(\cdot)) \equiv \frac{1}{|D|} \int_D I(Y(\mathbf{s}) \leq y) d\mathbf{s}, \quad (4.5)$$

where $I(\cdot)$ is the indicator function. Let $S_D^{-1}(\alpha; Y(\cdot))$ denote the inverse SCDF of $Y(\cdot)$, defined for $0 \leq \alpha \leq 1$ by:

$$S_D^{-1}(\alpha; Y(\cdot)) \equiv \arg \min\{y \in \mathbb{R} : S_D(y; Y(\cdot)) \geq \alpha\}. \quad (4.6)$$

Let θ_1 denote the α -quantile of the SCDF of $Y(\cdot)$, and θ_2 denote the spatial exceedance region of points in D that are above the quantile; that is,

$$\theta_1 = S_D^{-1}(\alpha; Y(\cdot)), \quad (4.7)$$

$$\theta_2 = e_{\theta_1}^Y \equiv \{\mathbf{s} \in D : I(Y(\mathbf{s}) > \theta_1) = 1\}. \quad (4.8)$$

Then, given the predictor, $\boldsymbol{\delta} \equiv (\delta_1(\mathbf{Z}), \delta_2(\mathbf{Z}))^T$, of $\boldsymbol{\theta} \equiv (\theta_1, \theta_2)^T$ that depends on the data vector \mathbf{Z} , we define the joint loss function, $L(\boldsymbol{\theta}, \boldsymbol{\delta})$ to be

$$L((\theta_1, \theta_2), (\delta_1(\mathbf{Z}), \delta_2(\mathbf{Z}))) = L_1(\theta_1, \delta_1(\mathbf{Z})) \times L_2(\theta_2, \delta_2(\mathbf{Z})), \quad (4.9)$$

which is of product form. Here, $L_1(\theta_1, \delta_1(\mathbf{Z})) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^+$ and $L_2(\theta_2, \delta_2(\mathbf{Z})) : \Omega \times \Omega \rightarrow \mathbb{R}^+$ are loss functions for θ_1 and θ_2 , respectively. Notice the joint loss $L((\theta_1, \theta_2), (\delta_1(\mathbf{Z}), \delta_2(\mathbf{Z})))$ is a nonnegative-valued function defined on $(\mathbb{R} \times \Omega) \times (\mathbb{R} \times \Omega)$. Suppose that (δ_1^*, δ_2^*) is the optimal Bayes predictor (OBP) for (θ_1, θ_2) based on the joint loss function defined by (4.9). Then, the joint OBP, (δ_1^*, δ_2^*) , is given by:

$$(\delta_1^*, \delta_2^*) = \arg \min_{(\delta_1, \delta_2) \in (\mathbb{R} \times \Omega)} E_{\theta_1, \theta_2 | \mathbf{Z}} [L((\theta_1, \theta_2), (\delta_1, \delta_2))], \quad (4.10)$$

where $E_{\theta_1, \theta_2 | \mathbf{Z}}[\cdot]$ is the expectation based on the conditional distribution of (θ_1, θ_2) given the observations \mathbf{Z} . We can simplify the calculation for $E_{\theta_1, \theta_2 | \mathbf{Z}}[\cdot]$ using the fact that θ_1 and θ_2 are functions of $Y(\cdot)$. Then,

$$E_{\theta_1, \theta_2 | \mathbf{Z}}[L((\theta_1, \theta_2), (\delta_1(\mathbf{Z}), \delta_2(\mathbf{Z})))] = E_{Y(\cdot) | \mathbf{Z}}[L((\theta_1, \theta_2), (\delta_1(\mathbf{Z}), \delta_2(\mathbf{Z})))], \quad (4.11)$$

where $E_{Y(\cdot) | \mathbf{Z}}[\cdot]$ is the expectation based on the conditional distribution of $Y(\cdot)$ given the observations \mathbf{Z} . Thus the joint OBP of (θ_1, θ_2) given data \mathbf{Z} is

$$(\delta_1^*, \delta_2^*) = \arg \min_{(\delta_1, \delta_2) \in (\mathbb{R} \times \Omega)} E_{Y(\cdot) | \mathbf{Z}}[L((\theta_1, \theta_2), (\delta_1, \delta_2))]. \quad (4.12)$$

4.2 Choice of loss functions and the resulting predictor

In principle, the loss function for a spatial quantile, $L_1(\theta_1, \delta_1(\mathbf{Z}))$, can be any loss function defined on $\mathbb{R} \times \mathbb{R}$. Possible examples include squared error loss, weighted squared error loss, or absolute error loss. Here, we propose a weighted quantile squared error loss (WQSEL) for $L_1(\cdot, \cdot)$, since it is designed for prediction of the tails of the SCDF (Craigmile et al., 2006). The specific loss function we shall use for $\theta_1 = S_D^{-1}(\alpha; Y(\cdot))$ is

$$L_1(\theta_1, \delta_1(\mathbf{Z})) = w_D(\theta_1)(\theta_1 - \delta_1(\mathbf{Z}))^2, \quad (4.13)$$

where the weight function $w_D(\theta_1)$ assigns weights based on the values of θ_1 . For example, in the problem of prediction of large spatial exceedances of $Y(\cdot)$, we might use the following weight function proposed by Craigmile et al. (2006):

$$w_D(\theta_1) \equiv \int_0^1 w(p) I(\theta_1 \in [F_D^{-1}(p), F_D^{-1}(p + dp))) , \quad (4.14)$$

where $F_D(y)$ is the spatially averaged cumulative distribution function (ACDF) of $Y(\cdot)$, defined by

$$F_D(y) \equiv \frac{1}{|D|} \int_D Pr\{Y(\mathbf{s}) \leq y\} d\mathbf{s}. \quad (4.15)$$

Consequently, the inverse ACDF is $F_D^{-1}(p) \equiv \arg \min\{y \in \mathbb{R} : F_D(y) \geq p; 0 \leq p \leq 1\}$. Notice that $F_D(y)$ is the expectation of $S_D(y; Y(\cdot))$ with respect to $Y(\cdot)$. The "importance function" $w : [0, 1] \rightarrow [0, \infty)$ in (4.14) is pre-specified. Craigmire et al. (2006) suggested using the sigmoid-type function,

$$w(p) = \frac{1}{1 + e^{-\lambda(p-\alpha)}}; \quad p \in [0, 1]. \quad (4.16)$$

Here, $1/2 \leq \alpha < 1$ is a target quantile of the SCDF of $Y(\cdot)$. In Section 4.3, we will demonstrate that the calibration quantity $\lambda > 0$ depends on the spatial dependence in the hidden process and the value of α . We calibrate λ to minimize the absolute bias in estimating the spatial quantiles using the WQSEL. Based on the results of a simulation study that we will present, we recommend large λ for high exceedances (α near 1) and strongly dependent spatial processes, and small λ for low exceedances (α near 0.5) and weakly dependent spatial processes.

The loss function for the spatial exceedance region, $L_2(\theta_2, \delta_2(\mathbf{Z})) : \Omega \times \Omega \rightarrow \mathbb{R}^+$, can be obtained from a binary-image metric, as developed in Chapter 3, Section 3.2. Specifically, we choose Baddeley's loss function (3.13), since it is well suited to discriminating between both local and global structures, as discussed in Chapter 3, Section 3.2. Recall that Baddeley's loss function is given by,

$$B_t(\theta_2, \delta_2(\mathbf{Z})) = \frac{1}{|D|} \int_D (d_t(\mathbf{s}, \theta_2) - d_t(\mathbf{s}, \delta_2(\mathbf{Z})))^2 d\mathbf{s}, \quad (4.17)$$

where $\theta_2 \in \Omega$ is the associated exceedance region above the spatial quantile, θ_1 , $\delta_2(\mathbf{Z}) \in \Omega$ is an action predicting the exceedance region, $d_t(\mathbf{s}, \theta_2)$ is the tapered

distance between site $\mathbf{s} \in D$ and set $\theta_2 \in \Omega$, given by

$$d_t(\mathbf{s}, \theta_2) = \begin{cases} \min_{\mathbf{s}' \in \theta_2} \{ \|\mathbf{s} - \mathbf{s}'\|, t \}, & \text{if } \theta_2 \neq \emptyset; \\ \min\{\xi, t\}, & \text{if } \theta_2 = \emptyset. \end{cases} \quad (4.18)$$

The calibration quantity (the taper) $t > 0$ depends on the spatial dependence of the hidden process. We will follow the same recommendation given in Chapter 3, Section 3.3; that is, $t = \min(r/2, \xi)$, where r is the range parameter; and ξ is the maximum pairwise distance in the domain D .

This choice of loss function for $L_1(\theta_1, \delta_1(\mathbf{Z}))$ and $L_2(\theta_2, \delta_2(\mathbf{Z}))$ results in the product-form loss function:

$$L((\theta_1, \theta_2), (\delta_1(\mathbf{Z}), \delta_2(\mathbf{Z}))) = w_D(\theta_1)(\theta_1 - \delta_1(\mathbf{Z}))^2 B_t(\theta_2, \delta_2(\mathbf{Z})), \quad (4.19)$$

where θ_1 , θ_2 , $w_D(\cdot)$, and $B_t(\cdot, \cdot)$ are defined by (4.7), (4.8), (2.24), and (3.13), respectively.

Given observations \mathbf{Z} , the choice of λ in the “importance function”, and the choice of taper t , the optimal Bayes predictor for (θ_1, θ_2) based on the joint loss function (4.19) can be calculated from:

$$\begin{aligned} (\delta_1^*, \delta_2^*) &= \arg \min_{(\delta_1, \delta_2) \in (\mathbb{R} \times \Omega)} E_{\theta_1, \theta_2 | \mathbf{Z}} [w_D(\theta_1)(\theta_1 - \delta_1)^2 B_t(\theta_2, \delta_2)] \\ &= \arg \min_{(\delta_1, \delta_2) \in (\mathbb{R} \times \Omega)} E_{Y(\cdot) | \mathbf{Z}} [w_D(\theta_1)(\theta_1 - \delta_1)^2 B_t(\theta_2, \delta_2)]. \end{aligned} \quad (4.20)$$

We could solve for the predictors in the joint space, $\mathbb{R} \times \Omega$, using simulated annealing. Instead, the following proposition shows that we can restrict the search for the optimal Bayes predictor to a reduced space, which simplifies the computation considerably.

Proposition 1 *For any given $\delta_2 \in \Omega$, define $\hat{\delta}_1(\delta_2)$ by*

$$\hat{\delta}_1(\delta_2) = \frac{E_{Y(\cdot) | \mathbf{Z}} [\theta_1 w_D(\theta_1) B_t(\theta_2, \delta_2)]}{E_{Y(\cdot) | \mathbf{Z}} [w_D(\theta_1) B_t(\theta_2, \delta_2)]}. \quad (4.21)$$

If θ_1 and θ_2 are defined by (4.7) and (4.8), respectively, then we have

$$\delta_1^* = \widehat{\delta}_1(\delta_2^*), \quad (4.22)$$

$$\delta_2^* = \arg \min_{\delta_2 \in \Omega} E_{Y(\cdot)|\mathbf{Z}} \left[w_D(\theta_1)(\theta_1 - \widehat{\delta}_1(\delta_2))^2 B_t(\theta_2, \delta_2) \right]. \quad (4.23)$$

Proof: We state and prove Lemma 1 and Lemma 2 below, which are used to prove Proposition 1.

Lemma 1 Suppose that θ_1 and θ_2 are defined by (4.7) and (4.8), respectively. Then for any given $\delta_2 \in \Omega$,

$$\widehat{\delta}_1(\delta_2) \equiv \frac{E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1) B_t(\theta_2, \delta_2) \theta_1]}{E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1) B_t(\theta_2, \delta_2)]},$$

minimizes $E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)(\theta_1 - \delta_1)^2 B_t(\theta_2, \delta_2)]$.

Proof: Differentiating $E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)(\theta_1 - \delta_1)^2 B_t(\theta_2, \delta_2)]$ with respect to δ_1 , setting the result equal to zero, and solving for δ_1 , we obtain the result.

Lemma 2 If (δ_1^*, δ_2^*) minimizes $E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)(\theta_1 - \delta_1)^2 B_t(\theta_2, \delta_2)]$, then $\delta_1^* = \widehat{\delta}_1(\delta_2^*)$, where the function $\widehat{\delta}_1(\cdot)$ is defined in Lemma 1.

Proof: By Lemma 1, $\widehat{\delta}_1(\delta_2^*)$ minimizes $E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)(\theta_1 - \delta_1)^2 B_t(\theta_2, \delta_2^*)]$. Therefore,

$$E_{Y(\cdot)|\mathbf{Z}} \left[w_D(\theta_1) \left(\theta_1 - \widehat{\delta}_1(\delta_2^*) \right)^2 B_t(\theta_2, \delta_2^*) \right] \leq E_{Y(\cdot)|\mathbf{Z}} \left[w_D(\theta_1) (\theta_1 - \delta_1^*)^2 B_t(\theta_2, \delta_2^*) \right].$$

Since (δ_1^*, δ_2^*) achieves the minimum of the (joint) Bayes risk, we have

$$E_{Y(\cdot)|\mathbf{Z}} \left[w_D(\theta_1) \left(\theta_1 - \widehat{\delta}_1(\delta_2^*) \right)^2 B_t(\theta_2, \delta_2^*) \right] = E_{Y(\cdot)|\mathbf{Z}} \left[w_D(\theta_1) (\theta_1 - \delta_1^*)^2 B_t(\theta_2, \delta_2^*) \right].$$

Hence, $\delta_1^* = \widehat{\delta}_1(\delta_2^*)$ is a solution.

Now we are ready to prove Proposition 1. On the right-hand side of equation (4.23), we have

$$\begin{aligned}
& \min_{(\delta_1, \delta_2) \in (\mathbb{R} \times \Omega)} E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)(\theta_1 - \delta_1)^2 B_t(\theta_2, \delta_2)] \\
&= \min_{\delta_2 \in \Omega} \left[\min_{\delta_1 \in \mathbb{R}} E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)(\theta_1 - \delta_1)^2 B_t(\theta_2, \delta_2)] \right] \\
&= \min_{\delta_2 \in \Omega} E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)(\theta_1 - \widehat{\delta}_1(\delta_2))^2 B_t(\theta_2, \delta_2)], \tag{4.24}
\end{aligned}$$

by Lemma 1. Therefore, δ_2^* defined in (4.23) achieves the minimum for δ_2 in the (joint) Bayes risk. By Lemma 2, δ_1^* defined by (4.22) achieves the other minimum for δ_1 in the (joint) Bayes risk.

We use simulated annealing to solve for the optimal Bayes predictor δ_2^* . From (3.13) and (4.23), we have that

$$\begin{aligned}
\delta_2^* = \arg \min_{\delta_2 \in \Omega} & \left[\int_D \left\{ d_t^2(\mathbf{s}, \delta_2) \mu_1 - 2d_t(\mathbf{s}, \delta_2) \mu_2(\mathbf{s}) + (\widehat{\delta}_1(\delta_2))^2 \mu_3(\mathbf{s}) \right. \right. \\
& + (\widehat{\delta}_1(\delta_2))^2 (d_t(\mathbf{s}, \delta_2))^2 \mu_4 - 2(\widehat{\delta}_1(\delta_2))^2 d_t(\mathbf{s}, \delta_2) \mu_5(\mathbf{s}) - 2\widehat{\delta}_1(\delta_2) \mu_6(\mathbf{s}) \\
& \left. \left. - 2\widehat{\delta}_1(\delta_2) (d_t(\mathbf{s}, \delta_2))^2 \mu_7 + 4\widehat{\delta}_1(\delta_2) d_t(\mathbf{s}, \delta_2) \mu_8(\mathbf{s}) \right\} d\mathbf{s} \right]. \tag{4.25}
\end{aligned}$$

Here $\mu_1 \equiv E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)\theta_1^2]$, $\mu_2(\mathbf{s}) \equiv E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)\theta_1^2 d_t(\mathbf{s}, \theta_2)]$, $\mu_3(\mathbf{s}) \equiv E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)d_t^2(\mathbf{s}, \theta_2)]$, $\mu_4 \equiv E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)]$, $\mu_5(\mathbf{s}) \equiv E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)d_t(\mathbf{s}, \theta_2)]$, $\mu_6(\mathbf{s}) \equiv E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)\theta_1 d_t^2(\mathbf{s}, \theta_2)]$, $\mu_7 \equiv E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)\theta_1]$, and $\mu_8(\mathbf{s}) \equiv E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)\theta_1 d_t(\mathbf{s}, \theta_2)]$. This expression is then substituted into (4.22) to obtain the optimal Bayes predictor, δ_1^* .

4.3 Simulation studies

The joint optimal Bayes predictor, (δ_1^*, δ_2^*) , depends on the weight function $w_D(\cdot)$, which is determined by the tuning parameter λ in the sigmoid-type “importance

function” $w(\cdot)$, as shown in (2.24). Hence, in this section, we first present a side simulation study to calibrate quantity λ for the WQSEL in the joint loss given by (4.19). We then evaluate the performance of the joint predictor of a spatial quantile and its associated exceedance region, defined in Section 4.1. We compare the joint-loss-based predictor of the spatial quantile and the exceedance region, with the posterior-mean-based plug-in predictor, which we now define. Let $\hat{Y}(\cdot) \equiv \{E_{Y(\mathbf{s})|\mathbf{z}}[Y(\mathbf{s})] : \mathbf{s} \in D\}$, the posterior-mean-based predictor for $Y(\cdot)$. Then, the posterior-mean-based plug-in predictor is:

$$\hat{\theta}_1 \equiv S_D^{-1}(\alpha; \hat{Y}(\cdot)), \quad \text{and} \quad \hat{\theta}_2 \equiv \{\mathbf{s} \in D : \hat{Y}(\mathbf{s}) > \hat{\theta}_1\}. \quad (4.26)$$

Before we outline the results of the simulation studies, we discuss the experimental design used in the simulations.

4.3.1 Calibration of λ

In the simulation study to calibrate λ , we consider a unit square domain $D \equiv \{\mathbf{s} = (x, y) : 0 \leq x \leq 1; 0 \leq y \leq 1\}$ (i.e., $\omega = 1$). We conduct the simulation and evaluate each predictor on the gridded domain $D^* \equiv \{\mathbf{s}_i^* = (x, y) : x = 0 = \frac{0}{M-1}, \frac{1}{M-1}, \dots, 1; y = 0 = \frac{0}{M-1}, \frac{1}{M-1}, \dots, 1\}$ for $M = 20$. Assuming no spatial trend in (4.2) (i.e., $\boldsymbol{\beta} = \mathbf{0}$), we let the hidden process $Y(\cdot)$ be an isotropic, mean-zero, Gaussian process with an exponential covariance function. This process is characterized solely by its variance σ^2 and by the spatial range parameter r . In our experiments, we let $\sigma^2 = 1$, and we vary r among values 0.5 (corresponding to weak spatial dependence), 1.5 (moderate spatial dependence), and 5 (very strong spatial dependence). We use a Cholesky decomposition for the covariance matrix, and we simulate 100 realizations of $\mathbf{Y} = \{Y(\mathbf{s}^*) : \mathbf{s}^* \in D^*\}$ for each of the three levels (weak, moderate, strong)

for the spatial dependence parameter; to each $Y(\mathbf{s}^*)$, we add $\epsilon(\mathbf{s}^*) \sim N(0, \sigma_\epsilon^2)$, an independent Gaussian measurement error, $N(0, \sigma_\epsilon^2)$ term to obtain the data process. We then randomly sample 1/4 of all sites in D^* and declare them to be the observation sites B ; B is subsequently held fixed throughout all the simulations. Then the data are $\mathbf{Z} = \{z(\mathbf{s}^*) : \mathbf{s}^* \in B\}$. We consider two levels of measurement-error variance: $\sigma_\epsilon^2 = 0.5$ and $\sigma_\epsilon^2 = 2$. Defining the signal-to-noise ratio (SNR) as the ratio of the partial sill (variance of the hidden process) to the measurement-error variance (i.e., $\text{SNR} \equiv \sigma^2/\sigma_\epsilon^2$), then in the simulation experiment, we have SNRs of 2 and 0.5. The spatial quantiles of interest are the 0.5-, 0.75-, 0.9- quantiles of the SCDF of $Y(\cdot)$. The factors and their levels are summarized in Table 4.1.

Factor	Number of levels	Values of levels
r	3	0.5
		1.5
		5
SNR	2	2
		0.5
α	3	0.50
		0.75
		0.90

Table 4.1: Levels of factors used in the simulation studies.

Let $\hat{\theta}_{1,W}$ be WQSEL-based optimal Bayes predictor for $\theta_1 = S_D^{-1}(\alpha; y(\cdot))$, defined by

$$\hat{\theta}_{1,W} \equiv \arg \min_{\delta_1 \in \mathbb{R}} E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)(\theta_1 - \delta_1)^2] \quad (4.27)$$

$$= \frac{E_{Y(\cdot)|\mathbf{Z}} [\theta_1 w_D(\theta_1)]}{E_{Y(\cdot)|\mathbf{Z}} [w_D(\theta_1)]}. \quad (4.28)$$

Then, in a similar manner to the simulation to calibrate λ in the IWQSEL in Chapter 3, Section 2.4, we calibrate λ in $w_D(\theta_1)$ using the following bias criterion:

$$\text{Bias}[S_D(\hat{\theta}_{1,W}; Y(\cdot))] \equiv E[S_D(\hat{\theta}_{1,W}; Y(\cdot)) - \alpha]. \quad (4.29)$$

The $E[\cdot]$ in (4.29) is obtained in the simulation by averaging over the samples taken of $Y(\cdot)$ and $\epsilon(\cdot)$.

In Figure 4.1, we show the bias in estimating the α -quantile of the SCDF of $Y(\cdot)$ using WQSEL for different factor combinations. By changing λ in the WQSEL, the bias of $S_D(q_\lambda(\alpha); \hat{Y}(\cdot))$, where $q_\lambda(\alpha) \equiv \hat{\theta}_{1,W}$ is the WQSEL-based OBP of $\theta_1 = S_D^{-1}(\alpha; Y(\cdot))$, can be made small or close to zero. For both $\text{SNR} = 0.5$ and $\text{SNR} = 2$, we observe that based on the bias criterion, the best λ becomes to increase as α increases, since for larger α the WQSEL predictor needs larger λ to put more weights on larger values. For the same level of $\alpha = 0.9$, λ tends to increase as the spatial dependence (expressed in terms of the range r) tends to be stronger: for stronger dependence, the process tends to have similar extreme values appearing in subregions of D . Therefore, the WQSEL-based predictor needs to be larger in order to deal with these larger extreme values. Based on the bias criterion, we propose the following rule, listed in Table 4.2, for the choice of λ in the WQSEL predictor. In the table, r is the range parameter that captures the spatial dependence in the hidden process

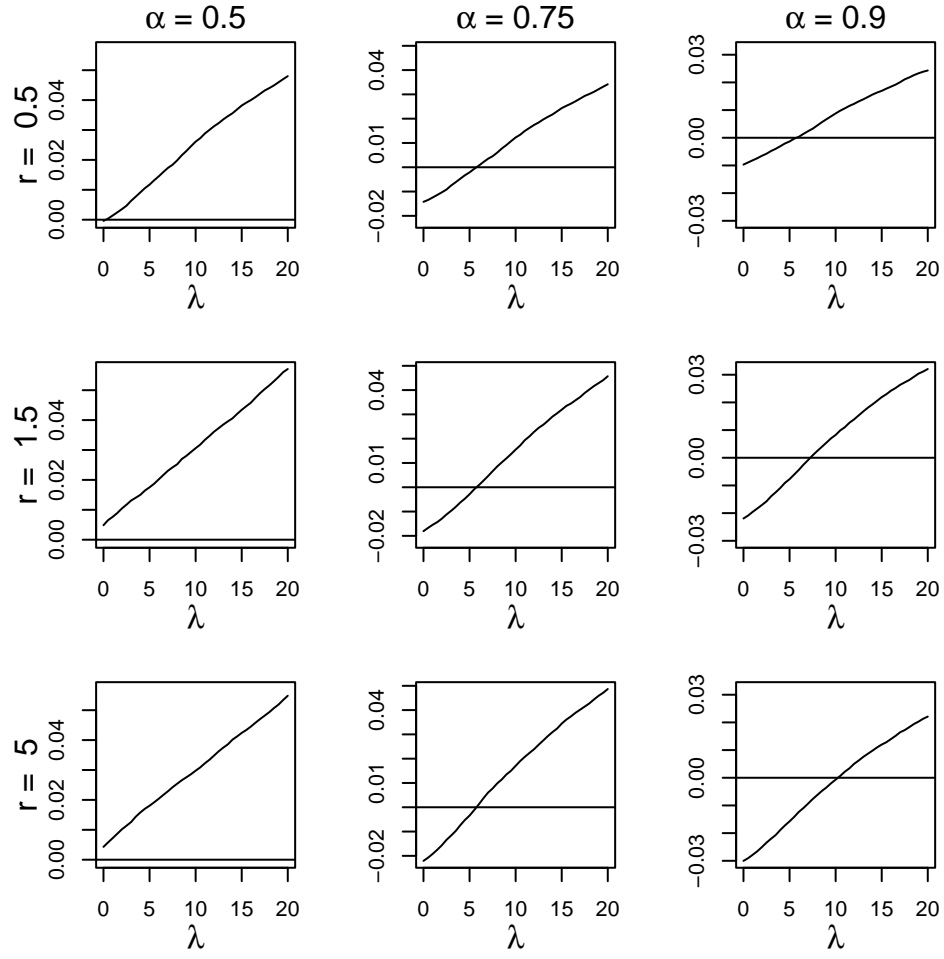


Figure 4.1: The bias in estimating the α -quantiles ($\alpha = 0.5, 0.75, 0.9$) of the SCDF using the WQSEL, versus the calibration quantity λ .

($r = 0.5$ represents “weak” dependence, $r = 1.5$ represents “moderate” dependence, and $r = 5$ represents “strong” dependence).

r	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 0.9$
0.5	0	6	6
1.5	0	6	7.5
5	0	6	10.5

Table 4.2: The best choice of λ in WQSEL, based on minimizing the absolute bias in estimating the α -quantile of the SCDF of interest.

4.3.2 Comparison of optimal predictor with plug-in predictor

We now describe a simulation study to compare the performance of joint-loss-based predictor, (δ_1^*, δ_2^*) given by (4.22), (4.23), with the posterior-mean-based plug-in predictor, $(\hat{\theta}_1, \hat{\theta}_2)$, given by (4.26). Similar to the simulation study used to calibrate the taper t , the domain of interest is the unit square domain D that is discretized to D^* , where $D^* = \{\mathbf{s}_i^* = (x, y) : x = 0 = \frac{0}{M-1}, \frac{1}{M-1}, \dots, 1; y = 0 = \frac{0}{M-1}, \frac{1}{M-1}, \dots, 1\}$ for $M = 20$. Assuming no spatial trend in (4.2) (i.e., $\boldsymbol{\beta} = \mathbf{0}$), we let the hidden process $Y(\cdot)$ be an isotropic mean-zero Gaussian process with an exponential covariance function. This process is characterized solely by its variance σ^2 and by the spatial range parameter r . In our experiments, we let $\sigma^2 = 1$, and vary r among values 0.5 (weak spatial dependence), 1.5 (moderate spatial dependence), and 5 (very strong spatial dependence). We use a Cholesky decomposition for the covariance matrix, and simulate 20 realizations of $\mathbf{Y} = \{Y(\mathbf{s}^*) : \mathbf{s}^* \in D^*\}$ for each of the three levels (weak, moderate, strong) for the spatial dependence parameter; to each $Y(\mathbf{s}^*)$,

we add 100 independent Gaussian measurement-error terms, $N(0, \sigma_\epsilon^2)$, to obtain the data process. We then randomly sample 1/4 of all sites in D^* and declare them to be the observation sites B for experiments; B is subsequently held fixed throughout all the simulations. Then the data are $\mathbf{Z} \equiv \{Z(\mathbf{s}^*) : \mathbf{s}^* \in B\}$. We consider two levels of measurement-error variance: $\sigma_\epsilon^2 = 0.5$ and $\sigma_\epsilon^2 = 2$. Defining the signal-to-noise ratio (SNR) as the ratio of the partial sill (variance of the hidden process $Y(\cdot)$) to the measurement-error variance (i.e., $\text{SNR} \equiv \sigma^2/\sigma_\epsilon^2$), then in the simulation experiment, we have SNRs of 2 and 0.5. The spatial quantiles of interest are the 0.5-, 0.75-, 0.9-quantiles of the SCDF of $Y(\cdot)$. In summary, the factors are the same as in Table 4.1. When defining the response to the simulation experiment, we average over the 100 replications of measurement error and display the 20 responses corresponding to the replicates of $Y(\cdot)$.

We define a measure to evaluate the overall performance of different predictor of (θ_1, θ_2) . Suppose that $(\tilde{\theta}_1, \tilde{\theta}_2)$ is a joint predictor for the true pair (θ_1, θ_2) of interest. Then the measure, $SSB((\theta_1, \theta_2), (\tilde{\theta}_1, \tilde{\theta}_2))$, is defined by

$$SSB\left((\theta_1, \theta_2), (\tilde{\theta}_1, \tilde{\theta}_2)\right) = \frac{1}{2} \left[Se\left(\theta_2, \tilde{\theta}_2\right) + Sp\left(\theta_2, \tilde{\theta}_2\right) \right] \left(1 - |\text{Bias}(S_D(\tilde{\theta}_1))|\right), \quad (4.30)$$

where $Se(\cdot, \cdot)$ and $Sp(\cdot, \cdot)$ are the sensitivity and specificity measures defined by (3.28), and $\text{Bias}(S_D(\tilde{\theta}_1; Y(\cdot))) \equiv S_D(\tilde{\theta}_1; Y(\cdot)) - \alpha$. Hence, the performance measure SSB quantifies the performance of $(\tilde{\theta}_1, \tilde{\theta}_2)$ on both the exceedance region, the non-exceedance region, and the α -quantile. Note that the joint measure $SSB \in [0, 1]$, and larger SSB corresponds to better performance.

We summarize our simulation results for $\text{SNR} = 0.5$ in Figures 4.2 and 4.3, and for $\text{SNR} = 2$ in Figures 4.4 and 4.5. In Figure 4.2, we show boxplots of the bias

in estimating α -quantile using the joint-loss-based predictor, δ_1^* , the bias using the posterior-mean-based plug-in predictor, $\hat{\theta}_1$, and the *difference* of the joint measure SSB between the paired joint-loss-based predictor, (δ_1^*, δ_2^*) and the paired posterior-mean-based plug-in predictor, $(\hat{\theta}_1, \hat{\theta}_2)$. In the first column of Figure 4.2, we observe that the joint-loss-based predictor δ_1^* tends to estimate the spatial quantile well, given most boxplots are centered around zero. For $\alpha = 0.9$, δ_1^* still estimates well the 0.9-quantile. In the second column of Figure 4.2, we observe that the posterior-mean-based plug-in predictor $\hat{\theta}_1$ estimates the 0.5-quantile quite well. However, $\hat{\theta}_1$ tends to underestimate the 0.75- and the 0.9- quantiles. In the third column of Figure 4.2, we show that for the 0.75- and the 0.9- quantiles and their associated exceedance regions, the paired joint-loss-based predictor, (δ_1^*, δ_2^*) , has better performance than the paired posterior-mean-based plug-in predictor, $(\hat{\theta}_1, \hat{\theta}_2)$, in terms of the SSB measure. For the 0.5-quantile, the SSB measure shows that (δ_1^*, δ_2^*) and $(\hat{\theta}_1, \hat{\theta}_2)$ have comparable performance in terms of SSB . In Figure 4.3, we show boxplots of the *difference* of the sensitivity measure (Se), the specificity measure (Sp), and the average of Se and Sp , $0.5(Se + Sp)$, for the joint-loss-based predictor (δ_2^*) and the posterior-mean-based plug-in predictor ($\hat{\theta}_2$) for the exceedance region. In the first column of Figure 4.3, we observe that for the weakly ($r = 0.5$), moderately ($r = 1.5$) and strongly ($r = 5$) dependent geostatistical processes, δ_2^* and $\hat{\theta}_2$ have comparable performance for the prediction of the spatial exceedance regions above the 0.5- and the 0.75- quantiles, while for the exceedance region above the 0.9-quantile, δ_2^* tends to predict more percentage of sites correctly for the true exceedance regions than $\hat{\theta}_2$. In the second column of Figure 4.3, we observe that joint-loss-based predictor tends to predict same or more percentage of sites correctly for the true non-exceedance regions than the

posterior-mean-based plug-in predictor. In the third column of Figure 4.3, we show that the joint-loss-based predictor and the posterior-mean-based plug-in predictor have the comparable average of Se and Sp for the exceedance region and the non-exceedance region associated with the 0.5- and 0.75- quantiles; while for the 0.9-quantile, the joint-loss-based predictor tends to have the larger average of Se and Sp than for the posterior-mean-based plug-in predictor.

In Figure 4.4 and 4.5, we observe similar simulation results when $SNR = 2$. In Figure 4.4, we observe that the performance of the joint-loss-based predictor and the posterior-mean-based predictor are comparable when the measurement-error is small ($SNR = 2$) for $\alpha = 0.5$ and $\alpha = 0.75$ in terms of SSB . When the measurement-error is small, the posterior mean is considered to be a good prediction for the center portion of the distribution of the true process. Hence, compared to the posterior-mean-based plug-in predictor, the joint-loss-based predictor does not have significant improvement on jointly estimating the spatial quantile and the associated exceedance region. We observe that this improvement of using joint-loss-based predictor increases when $\alpha = 0.9$ since the posterior mean still does not predictor well the upper tail of the distribution of the true process, even the measurement-error is small.

4.4 Discussion

In this chapter, we develop a joint loss for joint prediction of a spatial quantile and its associated exceedance region. We use a product-form loss function, namely the WQSEL function for the spatial quantile and Baddeley's loss function for the exceedance region. From this joint loss function, we obtain the optimal Bayes predictor for the spatial quantile and the exceedance region. We conduct a side simulation study

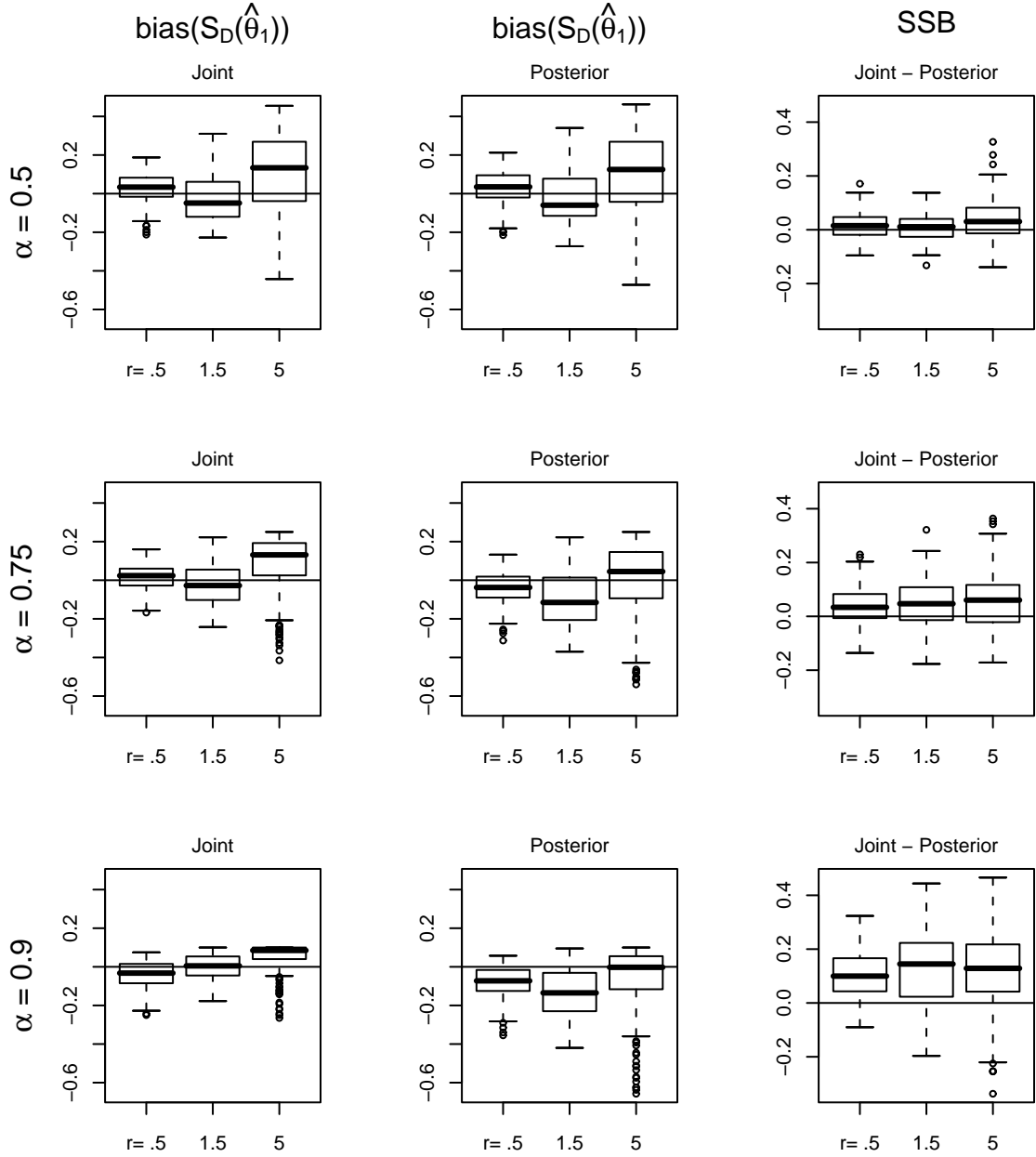


Figure 4.2: Boxplots of difference of bias in estimating α -quantiles using the joint-loss-based approach and the posterior-mean-based plug-in approach, and boxplot of difference of SSB values. The 20 values in each boxplot are generated by the 20 realizations of the $Y(\cdot)$ process. The signal-to-noise ratio, SNR, is 0.5 for this experiment.

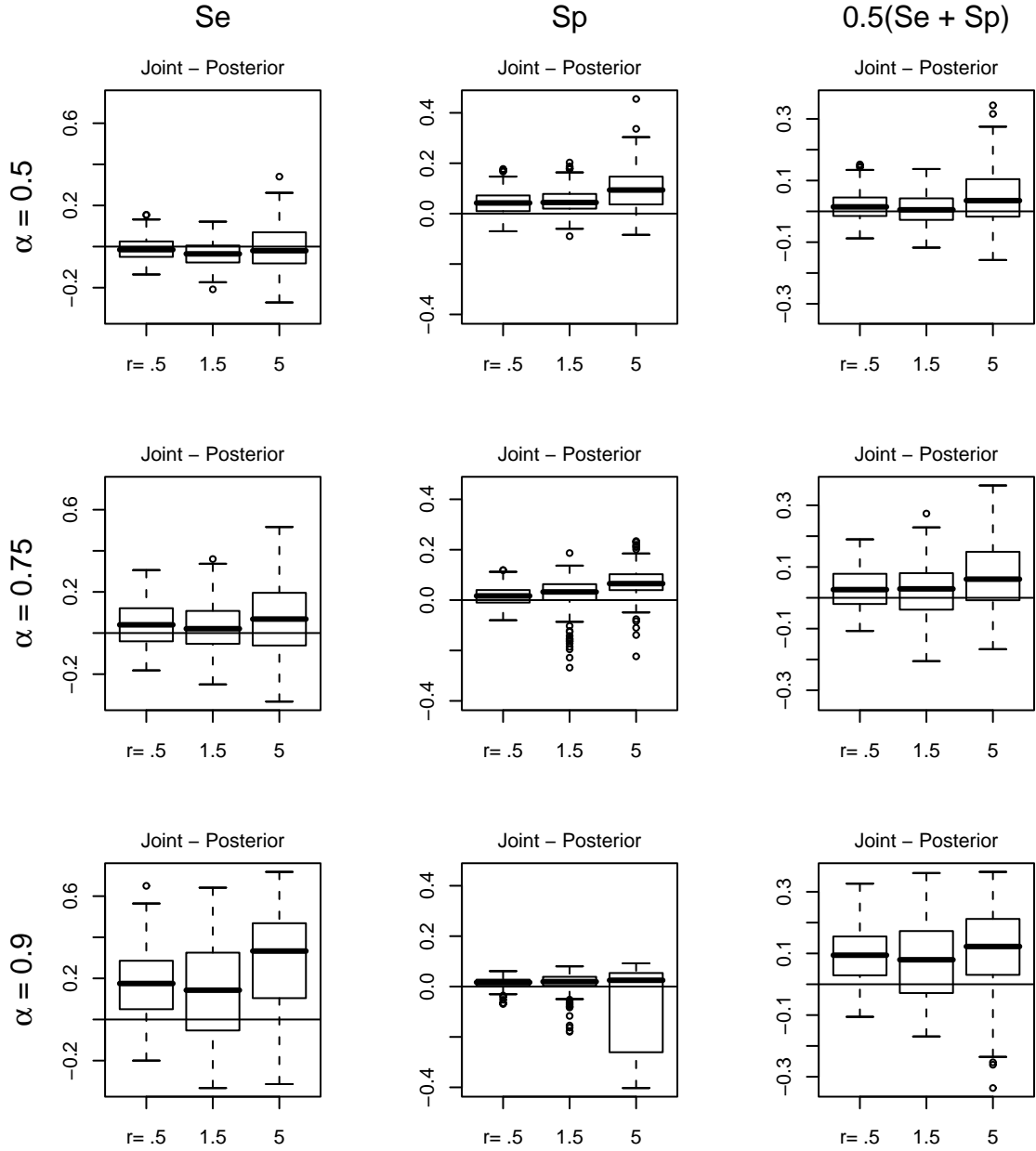


Figure 4.3: Boxplots of difference of the sensitivity (Se), the specificity (Sp), and the average of Se and Sp with the α -quantile. The 20 values in each boxplot are generated by the 20 realizations of the $Y(\cdot)$ process. The signal-to-noise ratio, SNR, is 0.5 for this experiment.

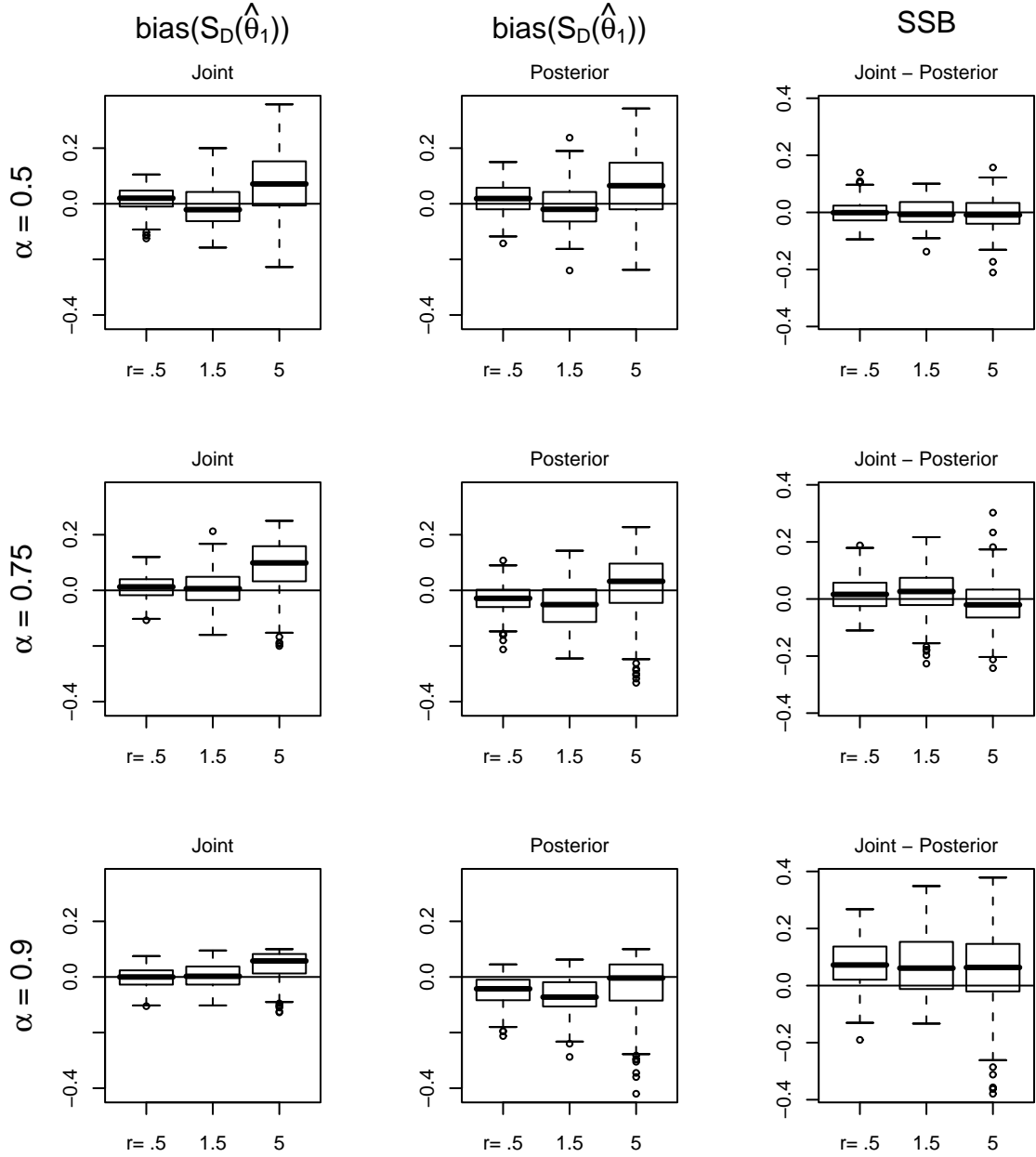


Figure 4.4: Boxplots of difference of bias in estimating α -quantiles using the joint-loss-based approach and the posterior-mean-based plug-in approach, and boxplot of difference of SSB values. The 20 values in each boxplot are generated by the 20 realizations of the $Y(\cdot)$ process. The signal-to-noise ratio, SNR, is 2 for this experiment.

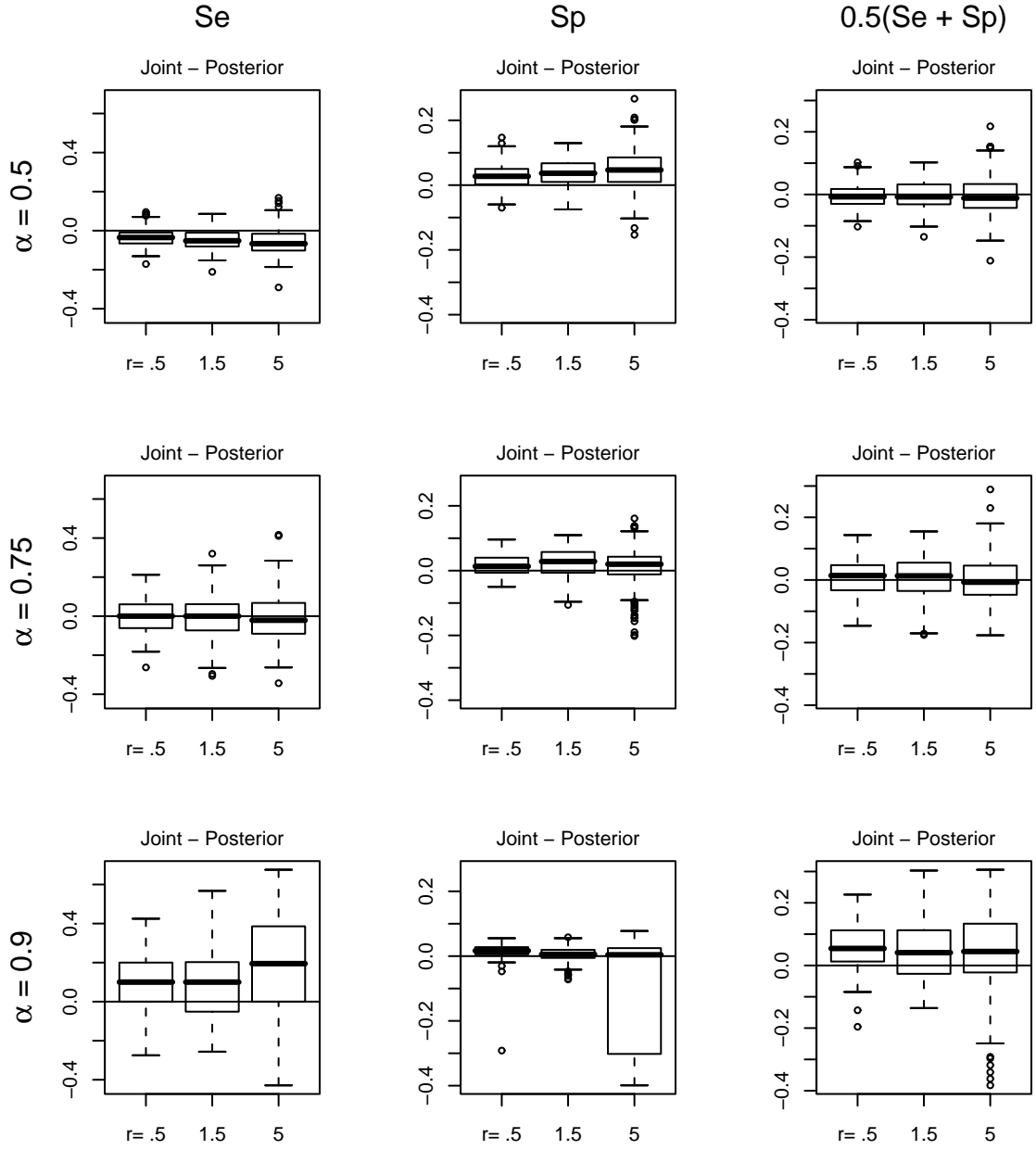


Figure 4.5: Boxplots of difference of the sensitivity (Se), the specificity (Sp), and the average of Se and Sp with the α -quantile. The 20 values in each boxplot are generated by the 20 realizations of the $Y(\cdot)$ process. The signal-to-noise ratio, SNR, is 2 for this experiment.

to choose the calibration quantity λ in the WQSEL function. Our simulation study shows that when the measurement-error is relatively large ($\text{SNR} = 0.5$), the joint-loss-based predictor generally has better overall performance than posterior-mean-based plug-in predictor for the prediction of the exceedance region, the non-exceedance region, and the associated spatial quantile, as quantified by the performance measure SSB . When the measurement-error is small ($\text{SNR} = 2$), the joint-loss-based predictor and the posterior-mean-based plug-in predictor have comparable results in terms of SSB for $\alpha = 0.5, 0.75$; we observe that the improvement gained by using joint-loss-based predictor increases when $\alpha = 0.9$, since the posterior mean is a poor predictor of the upper tail of the spatial distribution of the true process, even when the measurement-error is small.

Again, we note that throughout this chapter, we have taken an empirical Bayesian approach. That is, we assume that the spatial trend, and the covariance parameters are known; in practice, they are estimated from the data \mathbf{Z} . In the application in Chapter 5, Section 5.1, we show how this can be done. As we remarked at the end of Chapter 4, for exceedance regions associated with different spatial quantiles, the joint-loss-based predictor for each quantile does not guarantee nested structure in predicting exceedance regions. By nesting correctly we mean that if $\theta_1(\alpha)$ and $\theta_2(\alpha)$ denote the α -quantile and its associated exceedance region, then for $\alpha_a > \alpha_b$, $\theta_2(\alpha_a) \subset \theta_2(\alpha_b)$. If OBPs are obtained separately for each of α_a and α_b , then there is no guarantee of the set inclusion above. We discuss how this nesting of exceedance regions might be dealt with in Chapter 6.

CHAPTER 5

AN APPLICATION TO DECADAL TEMPERATURE CHANGE OVER THE AMERICAS

In this chapter, we shall apply the methodologies that we developed in this dissertation to an application that involves decadal temperature change over the Americas. Recall that we take an empirical Bayesian approach: After estimation of model parameters, we first use the loss-function-based predictor given in Chapter 3 to highlight spatial exceedance regions for which the change in decadal temperature is greater than several given temperature differences of interest. We then apply the joint-loss-based predictor shown in Chapter 4 to predict spatial quantiles and their associated spatial exceedance regions. Before presenting results, we shall briefly describe the decadal-temperature-change dataset.

The global temperature dataset that we study is obtained from the Climate System Model (CSM) at the National Center for Atmospheric Research (NCAR). The dataset consists of yearly averages of 2-meter air temperatures for the period of 1980 through 1999, over the whole globe on 128×64 equi-angular longitude-latitude grid cells. Each cell is roughly 2.8° in longitude by 2.8° in latitude. Using a procedure called the Enhanced False Discovery Rate (EFDR), Shen et al. (2002) found evidence of a temperature change between the two decades from this dataset, which occurred

for the most part in the central part of the USA and in coastal regions of South America, at about 20°S latitude. Like them, we focus on a reduced spatial domain D^* , for us, D^* is a 28×28 grid cells containing the majority of North and South America (from 40°N to 40°S and 125°W to 45°W). For each grid cell, we calculate the average temperature in the 1980s and subtract it from the average temperature in the 1990s. The resulting spatial dataset of the change in average temperatures (°C) between the two decades over D^* is shown as the grayscale map in Figure 5.1, panel (a). In order to study the performance of our methodology with missing data, we subsample regularly every third grid cell in latitude and longitude, starting from the cell at the south-west corner, to generate a collection of 10×10 observation sites B , shown as the crosses in Figure 5.1, panel (a). In Figure 5.1, panels (b) and (c), we show the difference in B as a function of latitude and of longitude, respectively. The temperature change appears to have a non-constant trend over the field. The histogram of the change in temperature in Figure 5.1, panel (d) shows that most of the observations have values from -0.4°C to 0.4°C . With this in mind, we propose a spatial model to describe this dataset.

5.1 Spatial model and parameter estimates

Let $Y(\cdot) \equiv \{Y(\mathbf{s}) : \mathbf{s} = (x, y) \in D^*\}$ be the hidden (true) temperature-change process in D^* . Suppose that $Z(\cdot) \equiv \{Z(\mathbf{s}) : \mathbf{s} \in D^*\}$ is the data process for the temperature-change process. Then we model the data as the true process of interest plus IID Gaussian measurement-error terms, given by

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \epsilon(\mathbf{s}); \quad \mathbf{s} \in D^*, \quad (5.1)$$

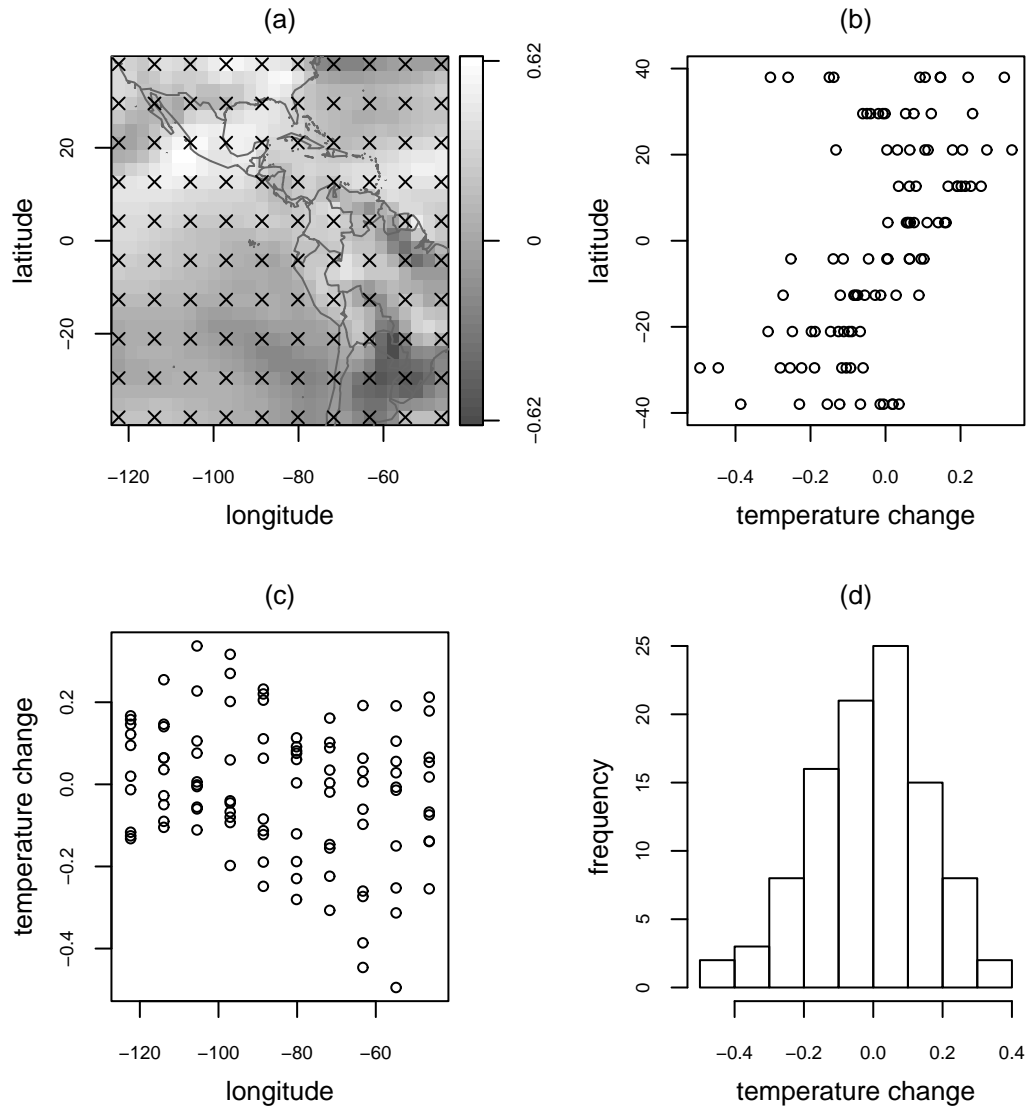


Figure 5.1: Spatial analysis of the average temperature change from the 1980s to the 1990s on the Celsius scale: (a) grayscale map of the temperature change over the prediction region D^* with observation sites B shown as crosses, (b) spatial trend of temperature change versus latitude, (c) spatial trend of temperature change versus longitude, (d) histogram of temperature change.

where $\epsilon(\cdot)$ is made up of IID RVs that are $N(0, \sigma_\epsilon^2)$ with measurement-error variance, $\sigma_\epsilon^2 > 0$. We assume that $\epsilon(\cdot)$ is independent of $Y(\cdot)$. The true process has a large-scale spatial trend $\mu(\cdot)$ plus a small-scale spatial structure $\delta(\cdot)$, given by

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}); \quad \mathbf{s} \in D^*. \quad (5.2)$$

To capture the non-linear trend on latitude, we model the spatial trend by $\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})^T \boldsymbol{\beta}$, where $\mathbf{x}(\mathbf{s}) = (1, x, y, y^2, y^3)^T$, for $\mathbf{s} = (x, y)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T$. Further, we assume that $\delta(\cdot)$ is a zero-mean, second-order stationary spatial process with parametric covariance function, $C_{\boldsymbol{\theta}_Y}(\cdot)$, where the spatial-covariance parameter, $\boldsymbol{\theta}_Y$, consists of the partial sill σ^2 , and the range parameter r . Recall that we obtained a collection of observations $\mathbf{Z} \equiv \{Z(\mathbf{s}) : \mathbf{s} \in B\}$ by regularly subsampling the domain D^* . We shall estimate $\boldsymbol{\beta}$, $\boldsymbol{\theta}_Y$, and σ_ϵ^2 based on the observations \mathbf{Z} .

We use linear regression to estimate the spatial trend parameter $\boldsymbol{\beta}$ based on the observation \mathbf{Z} and their coordinates. The ordinary least squares estimate of the coefficient of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (-0.1214, -0.00178, -0.0093, -0.000073, -0.0000055)^T$. The estimated spatial trend decreases linearly in longitude and cubically in latitude. In Figure 5.2, panel (a) and (b), we show the detrended residuals of the observed change in temperature as a function of latitude and longitude, respectively. We observe that the detrended residuals have almost a constant (zero) mean in both latitude and longitude. The histogram of the residuals in Figure 5.2, panel (c) suggests that a zero-mean normal distribution is a reasonable approximation for the distribution of the residuals. We then estimate the spatial parameter $\boldsymbol{\theta}_Y = (\sigma^2, r)^T$ and the measurement-error variance σ_ϵ^2 from the detrended residuals. The circles in Figure 5.2, panel (d), show the semivariogram of the residuals. We fit the semivariogram using both an exponential model, shown as the solid line in Figure 5.2, panel (d), and a spherical model,

shown as the dashed line in Figure 5.2, panel (d). The parameter estimates for the measurement-error variance σ_ϵ^2 , the partial sill σ^2 , and the range r , for both models are obtained using the weighted-least-squares (WLS) method (Cressie, 1993b, p.99). Based on the WLS criterion, we finally choose the exponential model. The estimates of the parameters in the exponential variogram are $(\hat{\sigma}_\epsilon^2, \hat{\sigma}^2, \hat{r}) = (0.0057, 0.0086, 25.0)$. Notice that the estimated SNR, the ratio of $\hat{\sigma}^2$ to $\hat{\sigma}_\epsilon^2$, is about 1.5. Further, based on the choice of taper in (3.20), we use $t = 12.5$. Assuming the exponential spatial covariance structure for $Y(\cdot)$ with $\hat{\sigma}^2$ and \hat{r} plug-in, we obtain the vector of the standard errors for $\hat{\beta}$ equal to $(0.1157, 0.001202, 0.002099, 0.00003677, 0.000001249)^T$.

5.2 Prediction of exceedance regions above fixed thresholds

Let $e_K^Y = \{\mathbf{s} \in D^* : Y(\mathbf{s}) > K\}$ denote the spatial exceedance region for which the mean change in temperature between the 1980s and the 1990s is greater than a threshold K measured in degrees Celsius. We choose the threshold K to be 0.05, 0.10, 0.15, and 0.20, which range from the center to the upper tail of the histogram of the data. Hence we are interested in prediction of the spatial region where the decadal temperature change from the 1980s to the 1990s is greater than 0.05°C , 0.10°C , 0.15°C , and 0.20°C , respectively. Given the data model in (5.1) and the process model in (5.2), we shall apply the Baddeley-loss-function-based approach (Chapter 3) to predict e_K^Y for $K = 0.05, 0.10, 0.15, 0.20$. Let e_K^* denote the OBP for the exceedance region above K . We shall compare e_K^* with the posterior-mean-based plug-in predictor: For any location $\mathbf{s} \in D^*$, the posterior-mean predictor of the hidden temperature-change process, $Y(\mathbf{s})$, is $\hat{Y}(\mathbf{s}) \equiv E_{Y(\mathbf{s})|\mathbf{Z}}[Y(\mathbf{s})]$. Then the

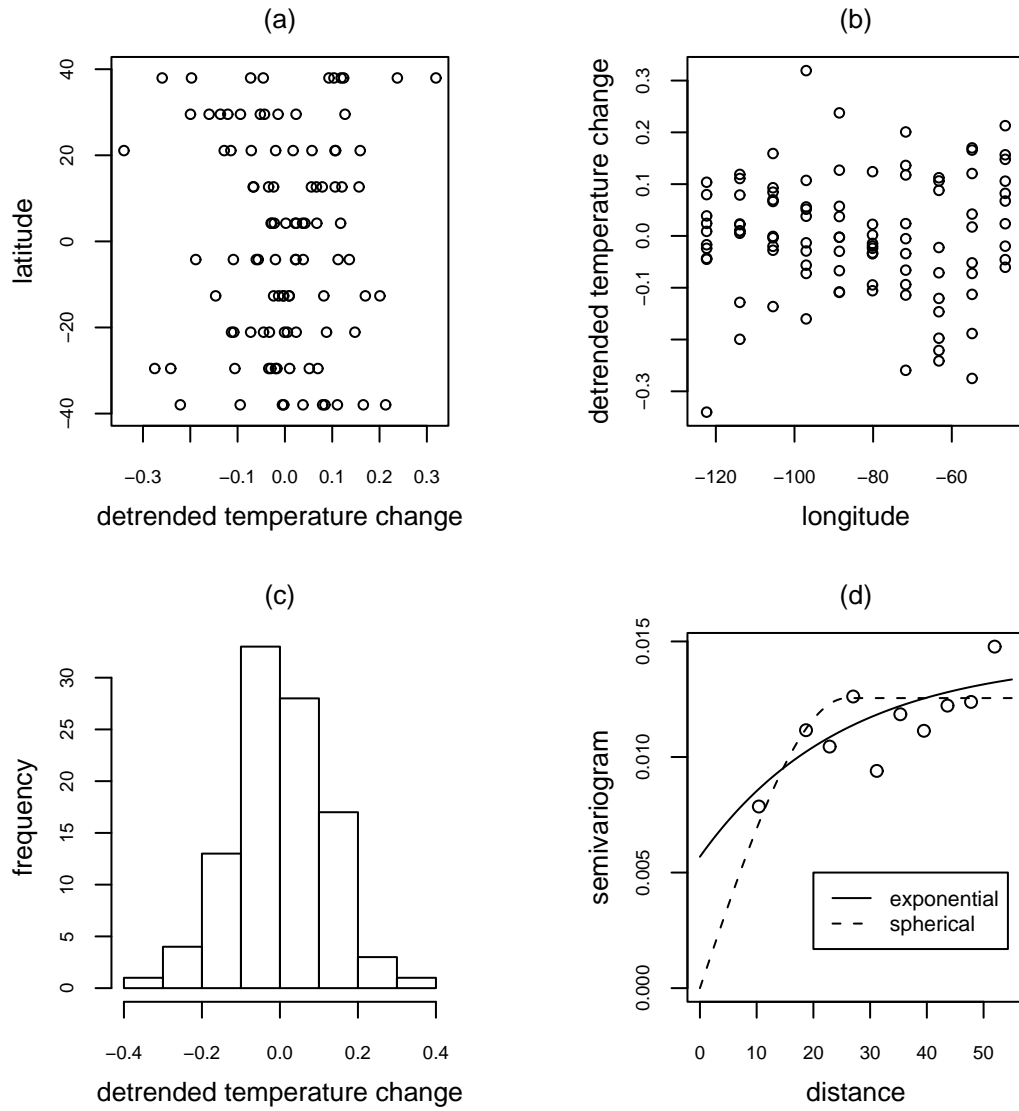


Figure 5.2: Spatial analysis of the detrended residuals of the observed averaged-temperature change in B from the 1980s to the 1990s on the Celsius scale : (a) spatial trend of detrended temperature change versus latitude, (b) spatial trend of detrended temperature change versus longitude, (c) histogram of detrended temperature change, (d) empirical and fitted semivariograms: circles for the empirical semivariogram, solid line for the fitted exponential model, and dashed line for the fitted spherical variogram.

posterior-mean-based plug-in predictor for the exceedance region above K , is $\widehat{e}_K \equiv \left\{ \mathbf{s} \in D^* : \widehat{Y}(\mathbf{s}) > K \right\}$.

In the left and the right columns of Figure 5.3, we show the Baddeley-loss-function-based predictor e_K^* and the posterior-mean plug-in predictor, \widehat{e}_K for $K = 0.05, 0.10, 0.15, 0.20$, respectively. In summary, we observe that for the mean temperature increases more than 0.05°C from the 1980s to the 1990s ($K = 0.05$), the two predictors have almost the same exceedance region, which covers most of the norther region except the north-west corner of the Atlantic Ocean. For the decadal temperature increase greater than 0.10°C ($K = 0.1$), both approaches largely agree; there is one large area that consists of the central part of the USA, Mexico, Gulf of Mexico, and the south-west Pacific coastal region of Mexico; and there is one small area in the Atlantic coastal region of Brazil. For the decadal temperature increase more than 0.15°C from the 1980s to the 1990s ($K = 0.15$), both approaches highlight the region over the central part of the USA, Mexico, Gulf of Mexico, the south-west Pacific coastal region of Mexico as the exceedance region. The Baddeley-loss-function-based predictor tends to have larger exceedance region than the posterior-mean-based plug-in predictor. For the decadal temperature change greater than 0.2°C ($K = 0.2$), the posterior-mean-based plug-in predictor consists of a small region over the south of Mexico. The Baddeley-loss-function-based prediction for the same exceedance region covers the south Mexico, and some small regions on the south-west Pacific coastal region of Mexico.

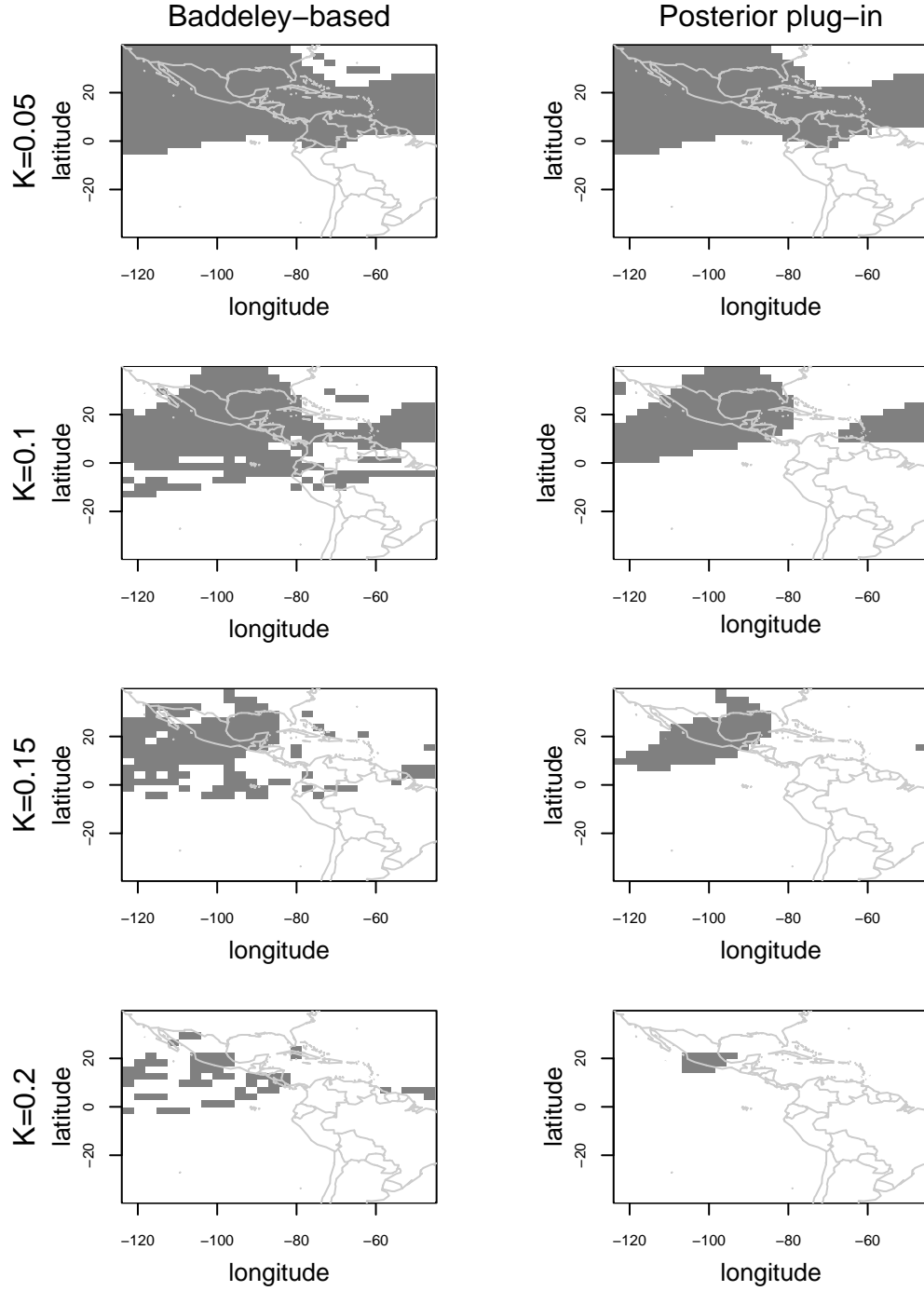


Figure 5.3: Prediction of exceedance regions with a change in mean temperature above K , for $K = 0.05, 0.10, 0.15, 0.20^\circ\text{C}$. The left column shows the Baddeley-loss-function-based predictor and the right column shows the posterior-mean-based plug-in predictors.

5.3 Prediction of extremely positive changes in temperature and their exceedance regions

In this section, we are interested in prediction of the exceedance region for which the decadal mean change in temperature is extremely positive. We define an extremely positive change in temperature to be the 0.9-quantile of the SCDF of the hidden temperature-change process, $Y(\cdot)$, as measured by $\theta_1 = S_{D^*}^{-1}(0.9)$. Let θ_2 denote the associated exceedance region, where the hidden mean-temperature-change process has value larger than this quantile, given by $\theta_2 = e_{\theta_1}^Y \equiv \{\mathbf{s} \in D^* : I(Y(\mathbf{s}) > \theta_1) = 1\}$. We apply the joint-loss approach, as introduced in Chapter 4, to this joint prediction problem. Since the hidden temperature-change process is weakly dependent ($\hat{r} = 25$ and D has a length of 80 in both latitude and longitude), we choose the calibration quantity $\lambda = 6$ based on the recommendation given by Table 4.2. The prediction of the spatial exceedance region with extremely positive mean-temperature change is shown in Figure 5.4, in gray shading. In Figure 5.4, panel (a), we see that the joint-loss-based prediction for the the spatial exceedance region associated with the 0.9-quantile consists of two major subregions, a large subregion that covers the central part of the USA, Mexico, Gulf of Mexico and the south-west Pacific coastal region of Mexico, and a small subregion covering a part of the Atlantic ocean, east of Puerto Rico and north of Brazil. In Figure 5.4, panel (b), the posterior-mean-based plug-in predictor for this spatial exceedance region covers a smaller portion of the central part of the USA, Mexico, Gulf of Mexico and the south-west Pacific coastal region of Mexico. The joint-loss-based predictor gives 0.166°C for the 0.9-quantile of the SCDF of the hidden temperature-change process over the prediction region D^* . The

posterior-mean-based plug-in predictor gives 0.15°C , which (as expected) is biased downwards relative to value of 0.166°C given above.

In Figure 5.5, the gray shading shows the exceedance region above a less extremely positive change, the 0.75-spatial quantile of the hidden temperature-change process, using the joint-loss-based approach and the posterior-mean-based plug-in approach. We choose $\lambda = 6$ according to the recommendation given by Table 2.2 with weak dependence ($r = 0.5$) and $\alpha = 0.75$. It is not surprising to see that both predicted exceedance regions are larger than the respective exceedance regions above the 0.9-quantile of the SCDF, since we are considering a smaller quantile. For $\alpha = 0.75$, both approaches result in the exceedance regions that largely agree; there is one main area in the north-west region that covers the central part of the USA, Mexico, Gulf of Mexico, and the south-west Pacific coastal region of Mexico; and there is one smaller area coving the Atlantic coastal region of Brazil. The joint-loss-based predictor gives 0.106°C for the 0.75-quantile of the SCDF and the posterior-mean-based plug-in prediction gives 0.098°C . Again, the latter prediction is smaller (as expected) and biased relative to the value of 0.106°C given above.

Notice that we actually have temperature-change data available over all of D^* . We apply the same spatial trend (linear in longitude and cubic in latitude) to all the data in D^* ; and the ordinary-least-squares estimates of the spatial-trend parameters are $\hat{\beta} = (-0.1419, -0.0020, 0.0091, -0.000087, -0.0000051)^T$. Further, we fit the semivariogram of the detrended residuals using the exponential model, and the weighted-least-squares estimates of the semivariogram parameters are σ_{ϵ}^2 , σ^2 , and r are $(\hat{\sigma}_{\epsilon}^2, \hat{\sigma}^2, \hat{r}) = (0.0, 0.0123, 10.0284)$. Notice that the estimated measurement-error

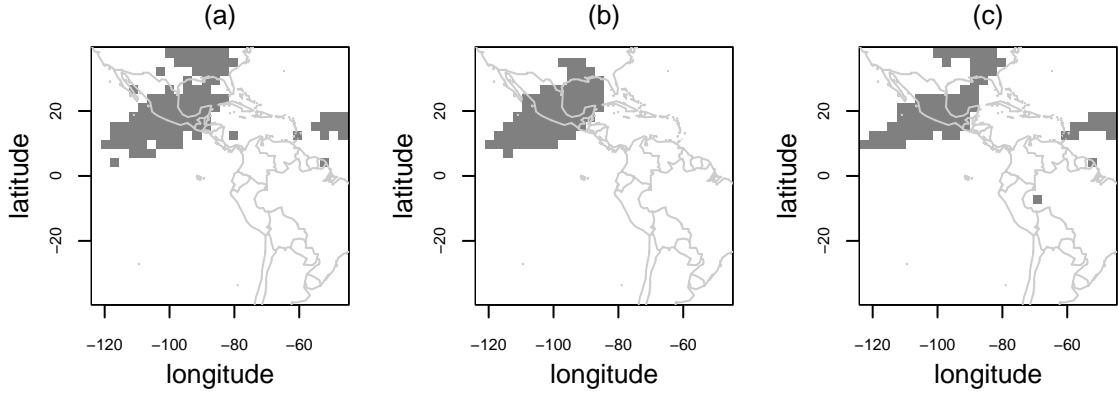


Figure 5.4: Prediction of spatial regions with extremely positive mean-temperature change (0.9-quantile of the SCDF), as shown by the gray shading: (a) the joint-loss-based predictor, (b) the posterior-mean-based plug-in predictor, (c) the plug-in predictor from all the data over D^* .

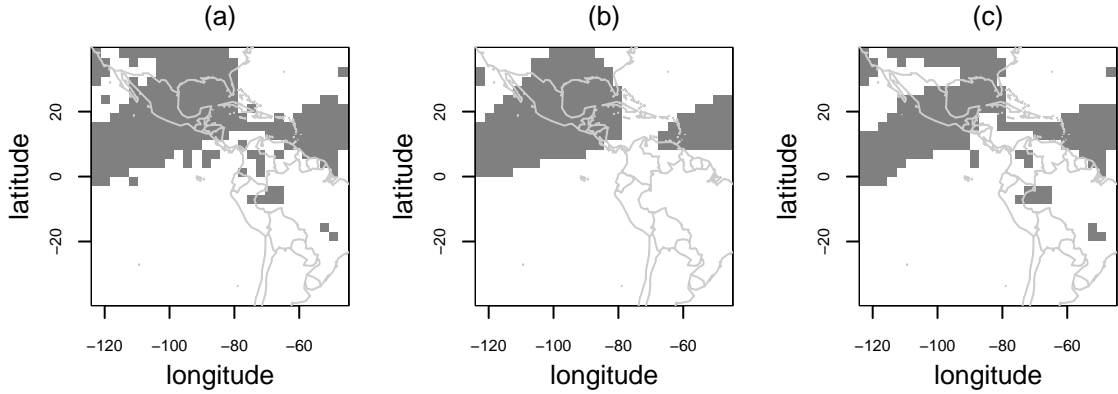


Figure 5.5: Prediction of spatial regions with mean-temperature change above the 0.75-quantile of the SCDF, as shown by the gray shading: (a) the joint-loss-based predictor, (b) the posterior-mean-based plug-in predictor, (c) the plug-in predictor from all the data over D^* .

variance $\widehat{\sigma}_\epsilon^2 = 0$, which means that $Z(\cdot) \equiv Y(\cdot)$. Hence, we can obtain the true 0.9-quantile, which is equal to 0.1992°C, and the true 0.75-quantile is equal to 0.1019°C. Moreover, the true associated exceedance regions can be obtained and are shown in Figures 5.4(c) and 5.5(c), respectively. The visual agreement between our proposed joint-loss-based prediction of the exceedance regions (Figures 5.4(a) and 5.5(a)) and the true exceedance regions (Figures 5.4(c) and 5.5(c)) is quite remarkable. Because we know the true process $Y(\cdot)$, we can calculate the performance measure SSB given by (4.30). For the 0.9-quantile, the $SSBs$ are 0.7504 and 0.9050 for the posterior-mean-based plug-in and the joint-loss-based predictors, respectively; for the 0.75-quantile, the corresponding $SSBs$ are 0.7912 and 0.9374. Clearly, for both the 0.9- and 0.75- quantiles, the joint-loss-based predictor has better performance than the posterior-mean-based plug-in predictor.

5.4 Discussion

In this chapter, we have applied the methodologies of Chapter 3 and 4 to a decadal temperature-change dataset over the Americas, to highlight exceedance regions where the potential positive mean change in temperature might happen. We first estimate model parameters including those associated with the spatial trend, the spatial covariance, and the measurement-error variance. We then predict the exceedance regions above fixed temperature-change thresholds using our Baddeley-loss-function-based approach. Our results show that for higher fixed thresholds, the extent of the exceedance region tends to be underestimated if we use the simple posterior-mean-based plug-in predictor (for Gaussian processes, the kriging predictor is in our case

equivalent to the posterior-mean-based predictor). This can have substantial implications in setting environmental-remediation regulations and in declaring climate change in regions of the globe. Further, we undertake joint prediction for an extreme positive change in temperature (quantified by a quantile of the SCDF) and its associated exceedance region above the quantile, using a joint-loss approach. As expected, the predicted *spatial quantiles* from the posterior-mean-based plug-in predictor are smaller than those from the joint-loss-based predictor. Compared to the joint-loss-based predictor, the posterior-mean-based plug-in predictor for *exceedance regions* above high quantiles also tend to be underestimated.

The application also show some issues that need to be investigated. We observe that a nested structure for exceedance regions based on ordered thresholds, is not always preserved. We shall direct our attention to this issue in Chapter 6.

CHAPTER 6

FUTURE RESEARCH

We have developed a loss function based on an image metric to predict the exceedance region above a known threshold in Chapter 3; we also have built a joint loss to predict a spatial quantile and its associated exceedance region in Chapter 4. In Chapter 5, we have illustrated the methodologies in an application. As we discussed before, there are extensions to our method that warrant future research. Suppose that we wish to predict the exceedance regions based on different thresholds. Calculating the OBP of the region for each threshold does not guarantee that the regions nest in the proper way. In future research, we shall consider a loss-based approach for simultaneously predicting multiple exceedance regions that preserve this nesting structure. Our approach is to use grayscale image metrics as the loss function for the simultaneous prediction of exceedance regions obtained from multiple fixed thresholds. Of course, this extension should preserve the nested structure of the exceedance regions with multiple thresholds. Finally, we shall briefly discuss the possibility of using a joint loss function that is more general than the product-form loss function we used, as well as the possibility of using a fully Bayesian approach to estimate the data-model and process-model parameters.

6.1 Loss functions for the prediction of multiple exceedance regions

In Chapter 3, we have demonstrated that metrics for binary images are well suited to loss functions for the exceedance set with one threshold, since applying one threshold on process $Y(\cdot)$ on a domain will turn the domain into a binary image with a foreground defined by (3.5). This motivates us to consider metrics for grayscale images as potential loss functions for exceedance regions associated with multiple thresholds.

Recall that $Y(\cdot) \equiv \{Y(\mathbf{s}) : \mathbf{s} \in D\}$ is the hidden geostatistical process of interest, defined on $D \subset \mathbb{R}^d$, with a positive d -dimensional volume, $|D|$. Let $Z(\cdot) \equiv \{Z(\mathbf{s}) : \mathbf{s} \in D\}$ be the process of observed and potentially observed values. Then, as we have seen before, the hierarchical model can be described by the following data and process models.

The data model:

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \epsilon(\mathbf{s}); \quad \mathbf{s} \in D, \quad (6.1)$$

where $\epsilon(\cdot)$ is a zero-mean white noise process with measurement-error variance, $\sigma_\epsilon^2 > 0$. We assume that $\epsilon(\cdot)$ is independent of $Y(\cdot)$.

The process model

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}); \quad \mathbf{s} \in D. \quad (6.2)$$

Here the function $\mu(\cdot)$ captures the large-scale, deterministic mean structure (*spatial trend*) of $Y(\cdot)$, and $\delta(\cdot)$ is a stationary stochastic process that models the small-scale spatial structure in $Y(\cdot)$. We assume that the spatial trend, $\mu(\cdot)$, follows a linear model given by

$$\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta}, \quad (6.3)$$

where the vector $\mathbf{x}(\mathbf{s}) \equiv (x_1(\mathbf{s}), \dots, x_p(\mathbf{s}))^T$ is a collection of covariates at \mathbf{s} , and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the spatial-trend parameter. We also assume that $\delta(\cdot)$ is a zero-mean, second-order stationary spatial process with a parametric covariance function, $C_{\boldsymbol{\theta}_Y}(\cdot)$. Precisely,

$$E[\delta(\mathbf{s})] = 0, \quad \text{and } \text{cov}[\delta(\mathbf{s}), \delta(\mathbf{s}')] \equiv C_{\boldsymbol{\theta}_Y}(\mathbf{s} - \mathbf{s}'); \quad \mathbf{s}, \mathbf{s}' \in D, \quad (6.4)$$

where $\boldsymbol{\theta}_Y \equiv (\sigma^2, r)^T$ is the spatial-covariance parameter that consists of the partial sill, σ^2 , and the range parameter, r . Suppose henceforth that we have n observations $\mathbf{Z} \equiv \{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\}$ at a collection of observation sites, $B \equiv \{\mathbf{s}_1, \dots, \mathbf{s}_n\} \subset D$. Then, the problem we shall now discuss is to predict well the exceedance regions of the hidden process $Y(\cdot)$ associated with threshold values K_1, \dots, K_ℓ , based on the observations \mathbf{Z} , where $\ell \geq 1$ is the number of thresholds.

6.1.1 Metrics for grayscale images as loss functions

Without loss of generality, we assume $K_1 < \dots < K_\ell$. Given the data model (6.1) and the process model (6.2), the exceedance region for each threshold is then defined by

$$e_{K_i}^Y \equiv \begin{cases} D, & \text{for } i = 0 \\ \{\mathbf{s} \in D : I(Y(\mathbf{s}) > K_i) = 1\}, & \text{for } i \in \{1, \dots, \ell\}. \end{cases} \quad (6.5)$$

From (6.5), the exceedance regions are nested; that is, $e_{K_\ell}^Y \subseteq e_{K_{\ell-1}}^Y \subseteq \dots \subseteq e_{K_0}^Y$. The sequence of exceedance regions has a 1-1 map to the following collection of sets:

$$e_{K_\ell}^Y \equiv \{e_i^Y; i = 0, \dots, \ell\} \quad \text{and} \quad e_i^Y \equiv \begin{cases} e_{K_i}^Y \setminus e_{K_{i+1}}^Y, & \text{for } i \in \{0, \dots, \ell - 1\}; \\ e_{K_\ell}^Y, & \text{for } i = \ell. \end{cases} \quad (6.6)$$

Notice that if there is only one threshold K , that is $\ell = 1$, then both $e_{K_1}^Y$ given in (6.5) and e_1^Y given in (6.6) are the exceedance set defined by (3.5).

For a process $Y(\cdot)$ in D , if we define a function $\kappa_\ell : \mathbb{R} \rightarrow \{0, \dots, \ell\}$ such that for any $\mathbf{s} \in D$,

$$\kappa_\ell(Y(\mathbf{s})) = \begin{cases} 0, & \text{if } Y(\mathbf{s}) \leq K_1; \\ i, & \text{if } K_i < Y(\mathbf{s}) \leq K_{i+1}, \text{ where } i \in \{1, \dots, \ell-1\}; \\ \ell, & \text{if } Y(\mathbf{s}) > K_\ell; \end{cases} \quad (6.7)$$

then the sets defined by (6.6) have the following representation based on κ_ℓ :

$$e_i^Y = \{\mathbf{s} \in D : \kappa_\ell(Y(\mathbf{s})) = i\}, \text{ where } i \in \{0, \dots, \ell\}. \quad (6.8)$$

There is a link between multiple exceedance regions defined by (6.5) and a grayscale image. A grayscale image is a function $f_\ell : D \rightarrow \{0, \dots, \ell\}$, where $D \subset \mathbb{R}^2$ and $\ell \geq 1$. After applying κ_ℓ on the process $Y(\cdot)$ in D , D turns into a grayscale image where the image function $f_\ell = \kappa_\ell$. Now the problem to predict well the exceedance regions defined by (6.5), or equivalently defined by (6.6), given a set of thresholds K_1, \dots, K_ℓ , is equivalent to predict well the grayscale levels induced by function κ_ℓ given in (6.7). Henceforth, our discussion will be based on the general grayscale image, f_ℓ , associated with multiple thresholds. Without loss of generality, we shall scale levels in f_ℓ by a factor of ℓ to let the largest level be 1; that is, $f_\ell : D \rightarrow \{0, 1/\ell, \dots, 1\}$. Notice that f_ℓ , for $\ell = 1$, is the binary image f . Let e_i^f denote the i -th foreground of the grayscale image, f_ℓ , defined by

$$e_i^f \equiv \{\mathbf{s} \in D : f_\ell(\mathbf{s}) = i/\ell\}, \quad \text{for } i \in \{1, \dots, \ell\}. \quad (6.9)$$

Finally, let $\Omega_\ell = (\ell + 1)^D$ denote the set of all possible exceedance sets in D .

The link between spatial exceedance regions and grayscale images motivates us to use image metrics for grayscale images, as loss functions to predict the exceedance region. Recall that a loss function, $L(\theta, a)$, is a real-valued function defined for all $(\theta, a) \in \Theta \times \mathcal{A}$, where Θ is a parameter space and \mathcal{A} is an action space that contains

the parameter space; L is bounded below by 0 and $L(\theta, \theta) = 0$. Sometimes, θ is considered to be random. The randomness of θ is specified by a prior distribution on Θ . Hence, a loss function for the prediction of $e_{K_\ell}^Y$ is defined on $\Theta \times \mathcal{A} = \Omega_\ell \times \Omega_\ell$. Now the true value of interest, $e_{K_\ell}^Y$, has a prior distribution that inherits its randomness from $Y(\cdot)$. Let ρ be a metric for grayscale images, where ρ is a function $\Omega_\ell \times \Omega_\ell$ to $\mathbb{R}^+ \equiv \{x : x \geq 0\}$. Since metrics for grayscale images are defined on $\Omega_\ell \times \Omega_\ell$ and satisfy the regularity conditions for loss functions, they can be used as a loss function for prediction of the multiple exceedance regions defined by (6.5).

We begin with a review of metrics for grayscale images. Some of the metrics are defined through the random-set representation of grayscale images. Given a scaled grayscale image f_ℓ , the corresponding random set is defined by

$$F_U \equiv \{\mathbf{s} \in D : U \leq f_\ell(\mathbf{s})\}, \quad (6.10)$$

where $U \in [0, 1]$ is a continuous random variable. The distribution of U determines the weighting associated with each gray level. If we assume a *uniformly weighted* random-set model that assumes U follows a uniform distribution on $[0, 1]$, then we can easily show a simple relation between the distribution of F_U and the image f_ℓ as follows:

$$P(\mathbf{s} \in F_U) = P(U \leq f_\ell(\mathbf{s})) = f_\ell(\mathbf{s}). \quad (6.11)$$

Friel and Molchanov (1998b) proposed an *integral metric* for two grayscale images based on the uniformly weighted random-set representation for grayscale images given by (6.10). Suppose that F_U and G_V are the corresponding random sets generated from f_ℓ and g_ℓ respectively, where U and V are independent random variables defined on

$[0, 1]$. The integral metric for grayscale images is defined as

$$\rho^I(f_\ell, g_\ell) \equiv \sup_{h^{\mathbf{a}} \in \mathcal{H}} \left| \int_0^1 \sup_{\mathbf{s} \in H_u^{\mathbf{a}}} f_\ell(\mathbf{s}) du - \int_0^1 \sup_{\mathbf{s} \in H_u^{\mathbf{a}}} g_\ell(\mathbf{s}) du \right|, \quad (6.12)$$

where $\mathcal{H} = \{h^{\mathbf{a}}(\mathbf{s}) : \mathbf{a} \in D\}$ is a collection of non-negative, location-specific functions for all \mathbf{a} in D , and $H_u^{\mathbf{a}}$ is the set of locations where the function, $h^{\mathbf{a}}(\mathbf{s}), \mathbf{s} \in D$, has value equal or larger than a threshold u ; that is, $H_u^{\mathbf{a}} \equiv \{\mathbf{s} \in D : h^{\mathbf{a}}(\mathbf{s}) \geq u\}$. An example of $h^{\mathbf{a}}(\mathbf{s})$ is given below in (6.13).

The Hausdorff metric in (3.11) is a special case of the integral metric given in (6.12). A binary image is a special case of a grayscale image with only two levels. Recall that the grayscale image f_ℓ with $\ell = 1$ is the binary image f ; that is, $f_1 = f$. Similarly, $g_1 = g$. In addition, if we choose

$$h^{\mathbf{a}}(\mathbf{s}) = 1 - \|\mathbf{s} - \mathbf{a}\|/t, \quad (6.13)$$

where t is sufficiently large to satisfy

$$t \geq \xi \equiv \sup_{\mathbf{s}, \mathbf{s}' \in D} \|\mathbf{s} - \mathbf{s}'\|, \quad (6.14)$$

then we can prove that the Hausdorff metric for two binary images, f and g , is a special case of the integral metric for the same binary images. The proof first notes that $\rho^I(f, g) = \rho^I(f_1, g_1)$ since $f_1 = f$ and $g_1 = g$. Then, applying the definition in (6.12), we have

$$\begin{aligned} \rho^I(f, g) &\equiv \sup_{h^{\mathbf{a}} \in \mathcal{H}} \left| \int_0^1 \sup_{\mathbf{s} \in H_u^{\mathbf{a}}} f(\mathbf{s}) du - \int_0^1 \sup_{\mathbf{s} \in H_u^{\mathbf{a}}} g(\mathbf{s}) du \right| \\ &= \sup_{h^{\mathbf{a}} \in \mathcal{H}} \left| \sup_{\mathbf{s} \in e^f} h^{\mathbf{a}}(\mathbf{s}) - \sup_{\mathbf{s} \in e^g} h^{\mathbf{a}}(\mathbf{s}) \right|. \end{aligned}$$

This last equation holds because

$$\int_0^1 \sup_{\mathbf{s} \in H_u^{\mathbf{a}}} f(\mathbf{s}) du = \int_0^{\sup\{h^{\mathbf{a}}(\mathbf{s}) : \mathbf{s} \in e^f\}} du = \sup_{\mathbf{s} \in e^f} h^{\mathbf{a}}(\mathbf{s}),$$

for any binary image f (and g), where e^f is the foreground set of f defined by (3.6).

Hence, upon choosing $h^{\mathbf{a}}(\mathbf{s})$ as in (6.13) and (6.14),

$$\rho^I(f, g) = t^{-1} \sup_{\mathbf{a} \in D} |d(\mathbf{a}, e^f) - d(\mathbf{a}, e^g)| = t^{-1} \rho^H(f, g),$$

because of (3.11). Consequently, if the grayscale image only has two levels (a binary image), the integral metric is the Hausdorff metric, modulo a normalizing constant, if we choose $h^{\mathbf{a}}(\mathbf{s})$ in (6.12) to have the form given in (6.13) and (6.14).

The *double-integral metric* (Friel and Molchanov, 1998b) is another metric for grayscale images, defined as follows:

$$\rho^{DI}(f_\ell, g_\ell) \equiv \left[\int_D \left| \int_0^1 \sup_{\mathbf{s} \in H_u^{\mathbf{a}}} f_\ell(\mathbf{s}) du - \int_0^1 \sup_{\mathbf{s} \in H_u^{\mathbf{a}}} g_\ell(\mathbf{s}) du \right|^p d\mathbf{a} \right]^{1/p}. \quad (6.15)$$

Compared to the integral metric given in (6.12), the double-integral metric replaces the supremum in (6.12) with an L^p average over all points in D .

In practice, and for $h^{\mathbf{a}}$ given by (6.13), the double-integral metric is usually discretized as follows:

$$\rho^{DI}(f_\ell, g_\ell) = \frac{1}{t+1} \left[\int_D \left| \sum_{c=0}^t ((f_\ell \oplus B_c)(\mathbf{a}) - (g_\ell \oplus B_c)(\mathbf{a})) \right|^p d\mathbf{a} \right]^{1/p}, \quad (6.16)$$

where

$$(f_\ell \oplus B_c)(\mathbf{a}) = \sup_{\mathbf{s} \in H_{1-c/t}^{\mathbf{a}}} f_\ell(\mathbf{s}) \quad (6.17)$$

is the largest image value at \mathbf{s} within a window around \mathbf{a} given by the ball of radius c . Note that $(f_\ell \oplus B_1)(\mathbf{a}) = f_\ell(\mathbf{a})$ if $t = 1$ unit in the discretized version of the double-integral metric defined by (6.16). This implies that the double-integral metric given in (6.16) includes a special case for the L^p average of the sitewise error for grayscale images, defined by

$$\rho^{EG}(f_\ell, g_\ell) \equiv \left[\int_D |f_\ell(\mathbf{a}) - g_\ell(\mathbf{a})|^p d\mathbf{a} \right]^{1/p}. \quad (6.18)$$

The third metric for grayscale images is a direct extension of Baddeley's metric for binary images in (3.9), due to Frigessi and Rue (1997). They called it *Baddeley's metric for grayscale images*. Suppose that f_ℓ and g_ℓ are two grayscale images with ℓ levels. Then Baddeley's metric for grayscale images is defined as

$$\rho_p^M(f_\ell, g_\ell) \equiv \left(\frac{1}{\ell|D|} \int_D \sum_{i=1}^{\ell} \left| u(d(\mathbf{s}, e_i^f)) - u(d(\mathbf{s}, e_i^g)) \right|^p d\mathbf{s} \right)^{1/p}, \quad (6.19)$$

where e_i^f and e_i^g are the foreground set for level $i \in \{1, \dots, \ell\}$ in f_ℓ and g_ℓ , respectively. Here the function $u : [0, \infty] \rightarrow [0, W]$ is the same bounded concave function as in Baddeley's metric for binary image, with $u(0) = 0$. If $\ell = 1$, then Baddeley's metric for grayscale images defined by (6.19) reduces to Baddeley's metric for binary images defined by (3.9).

6.1.2 An optimal Bayes predictor for multiple exceedance regions

Suppose $\rho(e_{K_\ell}^Y, \hat{e}_\ell)$ is a loss function for the prediction of multiple exceedance regions based on a metric $\rho(\cdot, \cdot)$ for grayscale images, where $\hat{e}_\ell \in \Omega_\ell$ is an action that is a function of observations \mathbf{Z} . Then, an optimal Bayes predictor can be found by an action that minimizes the posterior expected loss. Suppose $e_{K_\ell}^*$ is the OBP assuming the loss function $\rho(e_{K_\ell}^Y, \hat{e}_\ell)$; that is,

$$e_{K_\ell}^* = \arg \min_{\hat{e}_\ell \in \Omega_\ell} E_{Y(\cdot)|\mathbf{Z}} [\rho(e_{K_\ell}^Y, \hat{e}_\ell)], \quad (6.20)$$

where recall that $E_{Y(\cdot)|\mathbf{Z}}[\cdot]$ is expectation with respect to the posterior distribution of $Y(\cdot)$ given the observations \mathbf{Z} . Then a numerical solution is required to solve for $e_{K_\ell}^*$ since there are no analytic forms for the solution. The convergence to the optimal solution is slow because of the number of grayscale levels.

6.1.3 Iterative approach for the prediction of multiple exceedance regions based on binary image metrics

Another way to approach the prediction of multiple exceedance regions is to use image metrics for binary images, as follows. By way of illustration, we assume there are only two thresholds K_1 and K_2 , where $K_1 < K_2$. Then the collection of exceedance regions defined by (6.5) is $e_{K_0}^Y = D$, $e_{K_1}^Y = \{\mathbf{s} \in D : I(Y(\mathbf{s}) \geq K_1) = 1\}$, and $e_{K_2}^Y = \{\mathbf{s} \in D : I(Y(\mathbf{s}) \geq K_2) = 1\}$. Notice that the collection of exceedance regions are nested: $e_{K_2}^Y \subset e_{K_1}^Y \subset e_{K_0}^Y$. Consider $e_{K_0}^Y \setminus e_{K_1}^Y$ and $e_{K_1}^Y$ to be the background and the foreground of the binary image based on threshold K_1 . Then, we can use the same approach as in Chapter 3, where we use image metrics for binary image as loss functions, to predict $e_{K_1}^Y$. Suppose \widehat{e}_{K_1} is the predictor for the exceedance region $e_{K_1}^Y$ associated with the lowest threshold K_1 . Then $\widehat{e}_{K_1} \setminus e_{K_2}^Y$ and $e_{K_2}^Y$ are the background and the foreground of another binary image, based on the threshold K_2 . Conditional on \widehat{e}_{K_1} , we can repeat the approach given in Chapter 3 to iteratively use the image metrics for binary as loss functions to predict $e_{K_2}^Y$. Although the solution from this iterative approach of using image metrics for binary image will not be optimal from the point of view using a grayscale image metric as loss function, it does preserve the nesting structure among the predicted exceedance regions.

6.2 Other future research

For the prediction of a spatial quantile and its exceedance region, in Chapter 4 we propose a product-form joint loss function. There are other possible forms for a joint loss function, such as a sum-form or a form with both products and sums. When the joint loss function involves sum forms, we should calibrate the relative importance

of each loss functions for the prediction of the spatial quantile and the associated exceedance region; for example, see Stern and Cressie (1999), for the case of a finite spatial index.

In this dissertation, we pre-estimate model parameters using data; then we plug in these estimates to obtain the OBPs for both the exceedance region above a given threshold and a spatial quantile and its exceedance region. This is an empirical Bayesian approach. We could use a fully Bayesian approach that requires us to specify prior distributions on the unknown model parameters. For example, we might assume gamma priors for both the measurement-error-variance parameter, σ_ϵ^2 , and the partial-sill (process-variance) parameter, σ^2 . For the range parameter r , we could use a uniform prior distribution that is defined on a given finite interval on $[0, \infty)$. For the spatial-trend parameter β , a multivariate normal with diagonal covariance matrix is a possible prior. In the fully Bayesian approach, MCMC is necessary to obtain posterior samples for the hidden process given the data, before we use numerical optimization (e.g., simulated annealing) to solve for the OBPs of interest.

APPENDIX A

ASYMPTOTIC CALCULATIONS OF BADDELEY'S LOSS FUNCTION IN 1-D AND 2-D CASES

Baddeley's loss function takes into account not only the sitewise misclassification error $\rho_1^B(e^f, e^g; 1)$, but also the localization error. Here "localization" means some important local features, such as clusters, hotspots and so on. The taper parameter t controls the range of the localization error. In what follows we show that as t approaches 0 on a discrete grid, $B_t(e_K^Y, \hat{e}) \sim t^2 \cdot \rho^E(e_K^Y, \hat{e})$ in simple 1-D and 2-D cases, where $\rho^E(e_K^Y, \hat{e})$ is the sitewise misclassification error $\rho_1^B(e^f, e^g; 1)$. We derive the result using the Euclidean distance in the 1-D case, and the Manhattan's distance in the 2-D case, for the distance between two sites in (3.10). Manhattan's distance caters to the following asymptotic calculation for 2-D cases, and yet is approximately equal to Euclidean distance as $t \rightarrow 0$.

In both 1-D and 2-D cases, the threshold K is chosen to be large, so that it is plausible to assume that the exceedance region e_K^Y only consists of one single interval in the 1-D case, or one single square region in the 2-D case.

A.1 Baddeley's loss function in the 1-D case, for small tapers.

In the 1-D case, $\|\mathbf{s} - \mathbf{s}'\|$ be the Euclidean distance. The sets are intervals. In the following discussion, let

- $t \equiv$ the taper parameter in Baddeley's loss function;
- $\ell \equiv$ the total length of the field;
- $\Delta\ell \equiv$ the length of a grid cell and $n = \ell/\Delta\ell$;
- $x_i \equiv$ the coordinate of grid cell i ;
- $e_K^Y \equiv [a_1, b_1]$; $\widehat{e} \equiv [a_2, b_2]$.

We consider a number of cases, depending on the taper value and how the sets intersect (or not).

1. If $t \geq \ell$, then $d_t(x_i, e_K^Y) = d(x_i, e_K^Y)$.

(a) Suppose e_K^Y and \widehat{e} intersect, with $a_1 < a_2 < b_1 < b_2$. Then we have

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{1}{n} \sum_{i=1}^n [d_t(x_i, e_K^Y) - d_t(x_i, \widehat{e})]^2 \\
&= \frac{\Delta\ell}{\ell} \left[\frac{a_1}{\Delta\ell} (a_1 - a_2)^2 + \frac{\ell - b_2}{\Delta\ell} (b_1 - b_2)^2 \right. \\
&\quad \left. + \sum_{k=1}^{\frac{a_2 - a_1}{\Delta\ell}} (\Delta\ell \cdot k)^2 + \sum_{k=1}^{\frac{b_2 - b_1}{\Delta\ell}} (\Delta\ell \cdot k)^2 + 0 \right] \\
&= \frac{a_1}{\ell} (a_1 - a_2)^2 + \frac{\ell - b_2}{\ell} (b_1 - b_2)^2 \\
&\quad + \frac{(\Delta\ell)^3}{\ell} \left[\frac{1}{6} \frac{a_2 - a_1}{\Delta\ell} \left(\frac{a_2 - a_1}{\Delta\ell} + 1 \right) \left(\frac{2(a_2 - a_1)}{\Delta\ell} + 1 \right) \right. \\
&\quad \left. + \frac{1}{6} \frac{b_2 - b_1}{\Delta\ell} \left(\frac{b_2 - b_1}{\Delta\ell} + 1 \right) \left(\frac{2(b_2 - b_1)}{\Delta\ell} + 1 \right) \right] \\
&= \frac{a_1}{\ell} (a_1 - a_2)^2 + \frac{\ell - b_2}{\ell} (b_1 - b_2)^2 \\
&\quad + \frac{1}{3\ell} \left[(a_2 - a_1)(a_2 - a_1 + \Delta\ell) \left(a_2 - a_1 + \frac{\Delta\ell}{2} \right) \right. \\
&\quad \left. + (b_2 - b_1)(b_2 - b_1 + \Delta\ell) \left(b_2 - b_1 + \frac{\Delta\ell}{2} \right) \right].
\end{aligned}$$

As $\Delta\ell \rightarrow 0$ (equivalently, $n \rightarrow \infty$), the infill limit for Baddeley's loss function is

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{a_1}{\ell} (a_1 - a_2)^2 + \frac{\ell - b_2}{\ell} (b_1 - b_2)^2 + \frac{(a_2 - a_1)^3}{3\ell} + \frac{(b_2 - b_1)^3}{3\ell} \\
&= \frac{(a_2 - a_1)^2}{3\ell} (2a_1 + a_2) + \frac{(b_2 - b_1)^2}{3\ell} [2(\ell - b_2) + (\ell - b_1)].
\end{aligned}$$

(b) Suppose e_K^Y and \widehat{e} intersect, with $a_2 < a_1 < b_2 < b_1$. Then we have

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{1}{n} \sum_{i=1}^n [d_t(x_i, e_K^Y) - d_t(x_i, \widehat{e})]^2 \\
&= \frac{\Delta\ell}{\ell} \left[\frac{a_2}{\Delta\ell} (a_1 - a_2)^2 + \frac{\ell - b_1}{\Delta\ell} (b_1 - b_2)^2 \right. \\
&\quad \left. + \sum_{k=1}^{\frac{a_1 - a_2}{\Delta\ell}} (\Delta\ell \cdot k)^2 + \sum_{k=1}^{\frac{b_1 - b_2}{\Delta\ell}} (\Delta\ell \cdot k)^2 + 0 \right] \\
&= \frac{a_2}{\ell} (a_1 - a_2)^2 + \frac{\ell - b_1}{\ell} (b_1 - b_2)^2 \\
&\quad + \frac{1}{3\ell} \left[(a_1 - a_2)(a_1 - a_2 + \Delta\ell) \left(a_1 - a_2 + \frac{\Delta\ell}{2} \right) \right. \\
&\quad \left. + (b_1 - b_2)(b_1 - b_2 + \Delta\ell) \left(b_1 - b_2 + \frac{\Delta\ell}{2} \right) \right].
\end{aligned}$$

As $\Delta\ell \rightarrow 0$ (equivalently, $n \rightarrow \infty$), the infill limit for Baddeley's loss function is

$$B_t(e_K^Y, \widehat{e}) = \frac{(a_1 - a_2)^2}{3\ell} (a_1 + 2a_2) + \frac{(b_1 - b_2)^2}{3\ell} [2(\ell - b_1) + (\ell - b_2)].$$

(c) Suppose e_K^Y and \widehat{e} do not intersect, with $a_1 < b_1 \leq a_2 < b_2$. Then we have

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{1}{n} \sum_{i=1}^n [d_t(x_i, e_K^Y) - d_t(x_i, \widehat{e})]^2 \\
&= \frac{\Delta\ell}{\ell} \left\{ \frac{a_1}{\Delta\ell} (a_2 - a_1)^2 + \frac{\ell - b_2}{\Delta\ell} (b_2 - b_1)^2 \right. \\
&\quad + \sum_{k=1}^{\frac{b_1 - a_1}{\Delta\ell}} (a_2 - a_1 - k \cdot \Delta\ell)^2 \\
&\quad + \sum_{k=1}^{\frac{a_2 - b_1}{\Delta\ell}} [k \cdot \Delta\ell - (a_2 - b_1 - k \cdot \Delta\ell)]^2 \\
&\quad \left. + \sum_{k=1}^{\frac{b_2 - a_1}{\Delta\ell}} (a_2 - b_1 + k \cdot \Delta\ell)^2 \right\} \\
&= \frac{a_1}{\ell} (a_2 - a_1)^2 + \frac{\ell - b_2}{\ell} (b_2 - b_1)^2 \\
&\quad + \left[\frac{(a_2 - a_1)^2 (b_1 - a_1)}{\ell} - \frac{(a_2 - a_1)(b_1 - a_1)(b_1 - a_1 + \Delta\ell)}{\ell} \right. \\
&\quad \left. + \frac{(b_1 - a_1)(b_1 - a_1 + \Delta\ell)(b_1 - a_1 + \frac{\Delta\ell}{2})}{3\ell} \right] \\
&\quad + \left[\frac{(a_2 - b_1)^2 (a_2 - b_1)}{\ell} + \frac{2(a_2 - b_1)(a_2 - b_1 + \Delta\ell)}{\ell} \right. \\
&\quad \left. + \frac{4(a_2 - b_1)(a_2 - b_1 + \Delta\ell)(a_2 - b_1 + \frac{\Delta\ell}{2})}{3\ell} \right] \\
&\quad + \left[\frac{(a_2 - b_1)(b_2 - a_2)}{\ell} + \frac{(a_2 - b_1)(b_2 - a_2)(b_2 - a_2 + \Delta\ell)}{\ell} \right. \\
&\quad \left. + \frac{(b_2 - a_2)(b_2 - a_2 + \Delta\ell)(b_2 - a_2 + \frac{\Delta\ell}{2})}{3\ell} \right].
\end{aligned}$$

As $\Delta\ell \rightarrow 0$, the infill limit for Baddeley's loss function is

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{a_1(a_2 - a_1)^2}{\ell} + \frac{\ell - b_2}{\ell}(b_2 - b_1)^2 \\
&\quad + \frac{(a_2 - a_1)^2(b_1 - a_1) + (a_2 - b_1)^2(b_2 - a_2)}{\ell} \\
&\quad + \frac{(a_2 - b_1)(b_2 - a_2)^2 + 3(a_2 - b_1)^3 - (a_2 - a_1)(b_1 - a_1)^2}{\ell} \\
&\quad + \frac{(b_1 - a_1)^3 + 4(a_2 - b_1)^3 + (b_2 - a_2)^3}{3\ell} \\
&= \frac{(a_2 - a_1)^2 a_1}{\ell} + \frac{(\ell - b_2)(b_2 - b_1)^2}{\ell} \\
&\quad + \frac{(a_2 - a_1)(b_1 - a_1)(a_2 - b_1)}{\ell} + \frac{(a_2 - b_1)(b_2 - a_2)(b_2 - b_1)}{\ell} \\
&\quad + \frac{(b_1 - a_1)^3 + 13(a_2 - b_1)^3 + (b_2 - a_2)^3}{3\ell}.
\end{aligned}$$

(d) Suppose e_K^Y and \widehat{e} do not intersect, with $a_2 < b_2 \leq a_1 < b_1$. Then we have

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{1}{n} \sum_{i=1}^n [d_t(x_i, e_K^Y) - d_t(x_i, \widehat{e})]^2 \\
&= \frac{\Delta\ell}{\ell} \left\{ \frac{a_2}{\ell} (a_1 - a_2)^2 + \frac{\ell - b_1}{\Delta\ell} (b_1 - b_2)^2 \right. \\
&\quad + \sum_{k=1}^{\frac{a_2 - b_1}{\Delta\ell}} (a_1 - a_2 - k \cdot \Delta\ell)^2 \\
&\quad + \sum_{k=1}^{\frac{a_1 - b_2}{\Delta\ell}} [k \cdot \Delta\ell - (a_1 - b_2 - k \cdot \Delta\ell)]^2 \\
&\quad \left. + \sum_{k=1}^{\frac{b_1 - a_1}{\Delta\ell}} (a_1 - b_2 + k \cdot \Delta\ell)^2 \right\}.
\end{aligned}$$

As $\Delta\ell \rightarrow 0$, the infill limit for Baddeley's loss function is

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{a_2(a_1 - a_2)^2}{\ell} + \frac{(\ell - b_1)(b_2 - b_1)^2}{\ell} \\
&\quad + \frac{(a_1 - a_2)(b_2 - a_2)(a_1 - b_2)}{\ell} \\
&\quad + \frac{(a_1 - b_2)(b_1 - a_1)(b_1 - b_2)}{\ell} \\
&\quad + \frac{(b_2 - a_2)^3 + 13(a_1 - b_2)^3 + (b_1 - a_1)^3}{3\ell}.
\end{aligned}$$

Notice that for $t \geq \ell, m$ in all the cases (a)-(d), the asymptotic value of Baddeley's loss function as $\Delta\ell \rightarrow 0$ does not depend on the taper parameter t .

2. Now suppose that $t < \ell$. Define $n_t \equiv \frac{t}{\Delta\ell}$ and assume that $n_t > 1$, which means that the taper parameter is at least bigger than the length of the grid cell. Then $t \rightarrow 0$ implies $\Delta\ell \rightarrow 0$. Here we also assume that t is so small that dilated circles with radius t never intersect each other.

(a) Suppose e_K^Y and \widehat{e} intersect, with $a_1 < a_2 < b_1 < b_2$. Then we have

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{1}{n} \sum_{i=1}^n [d_t(x_i, e_K^Y) - d_t(x_i, \widehat{e})]^2 \\
&= \frac{\Delta\ell}{\ell} \left[\sum_{k=1}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 + \frac{(a_2 - t) - a_1}{\Delta\ell} t^2 + \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell)^2 \right. \\
&\quad \left. + \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell)^2 + \frac{b_2 - (b_1 + t)}{\Delta\ell} t^2 + \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 \right] \\
&= \frac{\Delta\ell}{\ell} \left[\frac{t^2}{\Delta\ell} (a_2 - t - a_1 + b_2 - 2 - b_1 - t) \right. \\
&\quad \left. + 4(\Delta\ell)^2 \sum_{k=0}^{\frac{t}{\Delta\ell}} k^2 + \frac{2t^3}{\Delta\ell} - 4\Delta\ell t \sum_{k=0}^{\frac{t}{\Delta\ell}} k \right] \\
&= \frac{t^2}{\ell} (a_2 - a_1 + b_2 - b_1 - 2t) \\
&\quad + \frac{4}{3\ell} t(t + \Delta\ell) \left(t + \frac{\Delta\ell}{2} \right) - 2t^2 \frac{\Delta\ell}{\ell}.
\end{aligned}$$

As $\Delta\ell \rightarrow 0$, the infill limit for Baddeley's loss function is

$$B_t(e_K^Y, \widehat{e}) = \frac{t^2}{\ell} \left[[(a_2 - a_1) + (b_2 - b_1)] - \frac{2t}{3} \right].$$

(b) Suppose e_K^Y and \widehat{e} intersect, with $a_2 < a_1 < b_2 < b_1$. Then we have

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{1}{n} \sum_{i=1}^n [d_t(x_i, e_K^Y) - d_t(x_i, \widehat{e})]^2 \\
&= \frac{\Delta\ell}{\ell} \left[\sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 + \frac{(a_1 - t) - a_2}{\Delta\ell} t^2 + \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell)^2 \right. \\
&\quad \left. + \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell)^2 + \frac{b_1 - (b_2 + t)}{\Delta\ell} t^2 + \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 \right] \\
&= \frac{t^2}{\ell} [(a_1 - a_2 - t) + (b_1 - b_2 + t)] \\
&\quad + \frac{4}{3\ell} t(t + \Delta\ell) \left(t + \frac{\Delta\ell}{2} \right) - \frac{2t^2\Delta\ell}{\ell}.
\end{aligned}$$

As $\Delta\ell \rightarrow 0$, the infill limit for Baddeley's loss function is

$$B_t(e_K^Y, \widehat{e}) = \frac{t^2}{\ell} \left[[(a_1 - a_2) + (b_1 - b_2)] - \frac{2t}{3} \right].$$

(c) Suppose that e_K^Y and \widehat{e} do not intersect, with $a_1 < b_1 < a_2 < b_2$. Then we have

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{1}{n} \sum_{i=1}^n [d_t(x_i, e_K^Y) - d_t(x_i, \widehat{e})]^2 \\
&= \frac{\Delta\ell}{\ell} \left[\sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 + \frac{b_1 - a_1}{\Delta\ell} t^2 + \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 + 0 \right. \\
&\quad \left. + \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell)^2 + \frac{b_2 - a_2}{\Delta\ell} t^2 + \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 \right] \\
&= \frac{t^2}{\ell} [(b_1 - a_1) + (b_2 - a_2)] \\
&\quad + \frac{4}{\ell} \left[\frac{t}{3} (t + \Delta\ell) \left(t + \frac{\Delta\ell}{2} \right) - t^3 - 2t^2\Delta\ell \right].
\end{aligned}$$

As $\Delta\ell \rightarrow 0$, the infill limit for Baddeley's loss function is

$$B_t(e_K^Y, \widehat{e}) = \frac{t^2}{\ell} \left[(b_1 - a_1) + (b_2 - a_2) - \frac{8}{3}t \right].$$

(d) Suppose that e_K^Y and \widehat{e} do not intersect, with $a_2 < b_2 < a_1 < b_1$. Then we have

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{1}{n} \sum_{i=1}^n [d_t(x_i, e_K^Y) - d_t(x_i, \widehat{e})]^2 \\
&= \frac{\Delta\ell}{\ell} \left[\sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 + \frac{b_2 - a_2}{\Delta\ell} t^2 + \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 + 0 \right. \\
&\quad \left. + \sum_{k=0}^{\frac{t}{\Delta\ell}} + \frac{b_1 - a_1}{\Delta\ell} t^2 + \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 \right] \\
&= \frac{t^2}{\ell} [(b_2 - a_2) + (b_1 - a_1)] \\
&\quad + \frac{4}{\ell} \left[\frac{t}{3} (t + \Delta\ell) \left(t + \frac{\Delta\ell}{2} \right) - t^3 - 2t^2 \Delta\ell \right].
\end{aligned}$$

As $\Delta\ell \rightarrow 0$ (equivalently, $n \rightarrow \infty$), the infill limit for Baddeley's loss function is

$$B_t(e_K^Y, \widehat{e}) = \frac{t^2}{\ell} \left[(b_1 - a_1) + (b_2 - a_2) - \frac{8}{3}t \right].$$

Notice that for $t < \ell$, in all the cases (a)-(d), the asymptotic value of Baddeley's loss function $\Delta\ell \rightarrow 0$, has the form

$$t^2 \frac{|(e_K^Y \cup \widehat{e} - e_K^Y \cap \widehat{e})|}{|D|},$$

where $|\cdot|$ is the volume of a set. In the 1-D case, the volume of a set (interval) is the length of the interval.

A.2 Baddeley's loss function in the 2-D case, for small tapers.

In the 2-D case, let $\|\mathbf{s} - \mathbf{s}'\|$ be the Manhattan distance, defined as

$$d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|.$$

The set here is chosen to be square region, $((x_l, y_l), (x_u, y_u))$, with the bottom-left corner (x_l, y_l) and upper-right corner (x_u, y_u) . In the following discussion, let

- $t \equiv$ the taper parameter in Baddeley's loss function;
- $\ell \equiv$ the total length in each direction;
- $\Delta\ell \equiv$ the length of a grid cell and $n = \ell/\Delta\ell$;
- $\mathbf{s}_i \equiv (i_1 \cdot \Delta\ell, i_2 \cdot \Delta\ell)$, $i_1, i_2 \in \{1, \dots, n\}$, are all the grid points over field D ;
- $e_K^Y \equiv ((a_1, b_1), (c_1, d_1))$, $\widehat{e} \equiv ((a_2, b_2), (c_2, d_2))$. Then $A_1 \equiv (c_1 - a_1)(d_1 - b_1)$ is the area of e_K^Y , $A_2 \equiv (c_2 - a_2)(d_2 - b_2)$ is the area of \widehat{e} , $P_1 \equiv 2(c_1 - a_1 + d_1 - b_1)$ is the perimeter of e_K^Y , $P_2 \equiv 2(c_2 - a_2 + d_2 - b_2)$ is the perimeter of \widehat{e} .

In the 2-D case, we shall only consider cases with $t < \ell$, since we have shown in the 1-D case that the asymptotic value of Baddeley's loss function as $\delta\ell \rightarrow 0$ is independent of the taper parameter, t . In cases with $t < \ell$, define $n_t \equiv \frac{t}{\Delta\ell}$ and assume that $n_t > 1$, which means that the taper parameter is at least longer than the length of the grid unit. Then if $n_t \rightarrow \infty$ and $t \rightarrow 0$, we must have $\Delta\ell \rightarrow 0$. Here we also assume that t is so small that dilated circles with radius t never intersect each other.

1. Suppose that e_K^Y and \widehat{e} do not intersect, with $c_1 < a_2$ and $d_1 < b_2$. Then we have

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n \{d_t[(i \cdot \Delta\ell, j \cdot \Delta\ell), e_K^Y] - d_t[(i \cdot \Delta\ell, j \cdot \Delta\ell), \widehat{e}]\}^2 \\
&= \frac{(\Delta\ell)^2}{\ell^2} \left\{ \left[4 \sum_{k=0}^{\frac{t}{\Delta\ell}} \sum_{m=0}^{\frac{t}{\Delta\ell}-k} (k \cdot \Delta\ell + m \cdot \Delta\ell - t)^2 \right. \right. \\
&\quad + 2 \frac{d_1 - b_1}{\Delta\ell} \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 \\
&\quad + 2 \frac{c_1 - a_1}{\Delta\ell} \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 \\
&\quad \left. + \frac{d_1 - b_1}{\Delta\ell} \frac{c_1 - a_1}{\Delta\ell} t^2 \right] \\
&\quad + \left[4 \sum_{k=0}^{\frac{t}{\Delta\ell}} \sum_{m=0}^{\frac{t}{\Delta\ell}-k} (k \cdot \Delta\ell + m \cdot \Delta\ell - t)^2 \right. \\
&\quad + 2 \frac{d_2 - b_2}{\Delta\ell} \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 \\
&\quad \left. + 2 \frac{c_2 - a_2}{\Delta\ell} \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 + \frac{d_2 - b_2}{\Delta\ell} \frac{c_2 - a_2}{\Delta\ell} t^2 \right] \Big\} \\
&= \frac{(\Delta\ell)^2}{\ell^2} \left\{ \frac{A_1 + A_2}{(\Delta\ell)^2} t^2 \right. \\
&\quad + \frac{P_1 + P_2}{(\Delta\ell)^2} \left[t^3 - t^2(t + \Delta\ell) + \frac{1}{3}t(t + \Delta\ell) \left(t + \frac{\Delta\ell}{2} \right) \right] \\
&\quad + 8 \left[t^2 \sum_{k=1}^{\frac{t}{\Delta\ell}+1} k - 2t \cdot \Delta\ell \sum_{k=1}^{\frac{t}{\Delta\ell}+1} k(k-1) \right. \\
&\quad \left. \left. + (\Delta\ell)^2 \sum_{k=1}^{\frac{t}{\Delta\ell}+1} k(k-1)^2 \right] \right\}
\end{aligned}$$

$$\begin{aligned}
&= \frac{t^2}{\ell^2}(A_1 + A_2) + \frac{P_1 + P_2}{\ell^2} \left[\frac{1}{3}t(t + \Delta\ell) \left(t + \frac{\Delta\ell}{2} \right) - \Delta\ell t^2 \right] \\
&\quad + \frac{8(\Delta\ell)^2}{\ell^2} \left[(t^2 + 2t \cdot \Delta\ell + (\Delta\ell)^2) \sum_{k=1}^{\frac{t}{\Delta\ell}+1} k \right. \\
&\quad \quad \left. - 2(t \cdot \Delta\ell + (\Delta\ell)^2) \sum_{k=1}^{\frac{t}{\Delta\ell}+1} k^2 + (\Delta\ell)^2 \sum_{k=1}^{\frac{t}{\Delta\ell}+1} k^3 \right] \\
&= \frac{A_1 + A_2}{\ell^2} t^2 + \frac{P_1 + P_2}{\ell^2} \left[\frac{1}{3}t(t + \Delta\ell) \left(t + \frac{\Delta\ell}{2} \right) \right] \\
&\quad + \frac{8}{\ell^2} \left[(t + \Delta\ell)^2 \frac{(t + \Delta\ell)(t + 2\Delta\ell)}{2} \right. \\
&\quad \quad \left. - \frac{1}{2}(t + \Delta\ell)(t + \Delta\ell)(t + 2\Delta\ell) \left(t + \frac{\Delta\ell}{2} \right) \right. \\
&\quad \quad \left. + \frac{1}{4}(t + \Delta\ell)^2(t + 2\Delta\ell)^2 \right].
\end{aligned}$$

As $\Delta\ell \rightarrow 0$, the infill limit for Baddeley's loss function is

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{A_1 + A_2}{\ell^2} t^2 + \frac{P_1 + P_2}{\ell^2} \frac{t^3}{3} + \frac{10t^4}{3\ell^2} \\
&= \frac{t^2}{\ell^2} \left[(A_1 + A_2) + \frac{P_1 + P_2}{3} t + \frac{10}{3} t^2 \right].
\end{aligned}$$

2. Suppose e_K^Y and \widehat{e} intersect, with $a_1 < a_2 < c_1$ and $b_1 < b_2 < d_1$. Then we have

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n \{d_t[(i \cdot \Delta\ell, j \cdot \Delta\ell), e_K^Y] - d_t[(i \cdot \Delta\ell, j \cdot \Delta\ell), \widehat{e}]\}^2 \\
&= \frac{(\Delta\ell)^2}{\ell^2} \left\{ \left[3 \sum_{k=0}^{\frac{t}{\Delta\ell}} \sum_{m=0}^{\frac{t}{\Delta\ell}-k} (k \cdot \Delta\ell + m \cdot \Delta\ell - t)^2 \right. \right. \\
&\quad + \sum_{k=0}^{\frac{t}{\Delta\ell}} \sum_{m=0}^{\frac{t}{\Delta\ell}-k} (k \cdot \Delta\ell + m \cdot \Delta\ell)^2 \\
&\quad + \frac{(d_1 - b_1) + (c_1 - a_1) + (b_2 - b_1 - t) + (a_2 - a_1 - t)}{\Delta\ell} \\
&\quad \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 + \frac{(d_1 - b_2) + (c_1 - a_2)}{\Delta\ell} \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell)^2 \\
&\quad \left. + \frac{(c_1 - a_1)(d_1 - b_1) - (d_1 - b_2 + t)(c_1 - a_2 + t)}{(\Delta\ell)^2} t^2 \right] \\
&\quad + \left[3 \sum_{k=0}^{\frac{t}{\Delta\ell}} \sum_{m=0}^{\frac{t}{\Delta\ell}-k} (k \cdot \Delta\ell + m \cdot \Delta\ell - t)^2 \right. \\
&\quad + \sum_{k=0}^{\frac{t}{\Delta\ell}} \sum_{m=0}^{\frac{t}{\Delta\ell}-k} (k \cdot \Delta\ell + m \cdot \Delta\ell)^2 \\
&\quad + \frac{(d_2 - b_2) + (c_2 - a_2) + (d_2 - d_1 - t) + (c_2 - c_1 - t)}{\Delta\ell} \\
&\quad \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - t)^2 + \frac{(d_1 - b_2) + (c_1 - a_2)}{\Delta\ell} \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell)^2 \\
&\quad \left. + \frac{(d_2 - b_2)(c_2 - a_2) - (d_1 - b_2 + t)(c_1 - a_2 + t)}{(\Delta\ell)^2} t^2 \right] \\
&\quad \left. + 2 \sum_{k=0}^{\frac{t}{\Delta\ell}} \sum_{m=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell - m \cdot \Delta\ell)^2 \right\}
\end{aligned}$$

$$\begin{aligned}
&= \frac{(\Delta\ell)^2}{\ell^2} \left\{ \left[8(\Delta\ell)^2 \sum_{k=1}^{\frac{t}{\Delta\ell}+1} k(k-1)^2 - 12t \cdot \Delta\ell \sum_{k=0}^{\frac{t}{\Delta\ell}+1} k(k-1) \right. \right. \\
&\quad \left. \left. + 6t^2 \sum_{k=0}^{\frac{t}{\Delta\ell}+1} k \right] \right. \\
&\quad \left. + \frac{2(d_2 + c_2 - a_1 - b_1 - 2t) + 2[(d_1 - b_2) + (c_1 - a_2)]}{\Delta\ell} \right. \\
&\quad \left. \sum_{k=0}^{\frac{t}{\Delta\ell}} (k \cdot \Delta\ell)^2 - \frac{2(d_2 + c_2 - a_1 - b_1 - 2t)}{\Delta\ell} \sum_{k=0}^{\frac{t}{\Delta\ell}} 2t \cdot \Delta\ell \cdot k \right. \\
&\quad \left. + \frac{(\Delta\ell)^2}{3} \frac{t^2}{(\Delta\ell)^2} \left(\frac{t}{\Delta\ell} + 1 \right) \left(\frac{t}{\Delta\ell} - 1 \right) + \frac{A_1 + A_2}{(\Delta\ell)^2} t^2 \right. \\
&\quad \left. - \frac{2(d_1 - b_2 + t)(c_1 - a_2 + t) + \Delta\ell \cdot 2(d_2 + c_2 - a_1 - b_1 - 2t)}{(\Delta\ell)^2} t^2 \right\} \\
&= \frac{1}{\ell^2} \left\{ 2[(t + \Delta\ell)(t + 2\Delta\ell)]^2 - \frac{16\Delta\ell + 12t}{3} (t + \Delta\ell)(t + 2\Delta\ell)(t + 1.5\Delta\ell) \right. \\
&\quad \left. + (4(\Delta\ell)^2 + 6t \cdot \Delta\ell + 3t^2)(t + \Delta\ell)(t + 2\Delta\ell) \right. \\
&\quad \left. - 2(d_2 + c_2 - a_1 - b_1 - 2t)t^2(t + \Delta\ell) \right. \\
&\quad \left. + \frac{2t[(d_2 + c_2 - a_1 - b_1 - 2t) + (d_1 - b_2) + (c_1 - a_2)]}{3} (t + \Delta\ell) \cdot \right. \\
&\quad \left(t + \frac{\Delta\ell}{2} \right) + t^2[(A_1 + A_2) - 2(d_1 - b_2 + t)(c_1 - a_2 + t) \\
&\quad \left. + 2(d_2 + c_2 - a_1 - b_1 - 2t)\Delta\ell] + \frac{1}{3}t^2(t + \Delta\ell)(t - \Delta\ell) \right\}.
\end{aligned}$$

As $\Delta\ell \rightarrow 0$, the infill limit for Baddeley's loss function is

$$\begin{aligned}
B_t(e_K^Y, \widehat{e}) &= \frac{t^2}{\ell^2} \left\{ 2t^2 - \frac{4t}{3} [(c_2 + d_2) - (a_1 + b_1)] \right. \\
&\quad \left. + [A_1 + A_2 - 2(d_1 - b_2)(c_1 - a_2)] \right\}.
\end{aligned}$$

Notice that in both 1-D and 2-D cases, the leading term in the infill limit of Baddeley's loss is $t^2 \cdot \rho^E(e_K^Y, \widehat{e})$, as the taper parameter $t \rightarrow 0$.

BIBLIOGRAPHY

- Aldworth, J. and N. Cressie (2003). Prediction of nonlinear spatial functionals. *Journal of Statistical Planning and Inference* 112, 3–41.
- Amzal, B., F. Bois, E. Parent, and C. Robert (2006). Bayesian-optimal design via interacting particle systems. *Journal of the American Statistical Association* 101, 773–785.
- Ancona-Navarrete, M. and J. Tawn (2002). Diagnostics for pairwise extremal dependence in spatial process. *Extremes* 5, 271–285.
- Azaïs, J. and C. Mercadier (2003). Asymptotic poisson character of extremes in nonstationary Gaussian model. *Extremes* 6, 301–318.
- Baddeley, A. (1992). Errors in binary images and an \mathbb{L}^p version of the Hausdorff metric. *Nieuw Archief voor Wiskunde* 10, 157–183.
- Baddeley, A. and I. Molchanov (1998). Averaging of random sets based on their distance functions. *Journal of Mathematical Imaging and Vision* 8, 79–92.
- Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis* (second edition). New York: Springer.
- Castillo, E. (1988). *Extreme Value Theory in Engineering*. New York: Academic Press.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer.
- Coles, S. and J. Tawn (1996). Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society, Series B* 58, 329–347.
- Cooley, D. (2005). *Statistical Analysis of Extremes Motivated by Weather and Climate Studies; Applied and Theoretical Advances*. Ph. D. thesis, Applied Math Department, University of Colorado, Boulder, CO.
- Cooley, D., P. Naveau, and V. Jomelli (2006). A Bayesian hierarchical extreme value model for Lichenometry. *Environmetrics* 17, 555–574.
- Cooley, D., P. Naveau, and P. Poncet (2006). Variograms for spatial max-stable random fields. In B. P., D. P., and S. P. (Eds.), *Dependence in Probability and Statistics*, Springer Lecture Notes in Statistics. New York: Springer.

- Craigmile, P. F., N. Cressie, T. Santner, and Y. Rao (2006). A loss function approach to identifying environmental exceedances. *Extremes* 8, 143–159.
- Cressie, N. (1988). When are census counts improved by adjustment. *Survey Methodology* 14, 191–208.
- Cressie, N. (1993a). Aggregation in geostatistical problems. In A. Soares (Ed.), *Geostatistics Troia 1992*, Volume 1, pp. 25–36. Dordrecht: Kluwer Academic Publishers.
- Cressie, N. (1993b). *Statistics for Spatial Data* (revised edition). New York: Wiley.
- Cressie, N. and G. Johannesson (2001). Kriging for cut-offs and other difficult problems. In P. Monestiez, D. Allard, and R. Froidevaux (Eds.), *geoENV III - Geostatistics for Environmental Applications*, pp. 299–310. Dordrecht: Kluwer.
- Cressie, N., H. Stern, and D. Wright (2000). Mapping rates associated with polygons. *Journal of Geographical Systems* 2, 61–69.
- Cressie, N., J. Zhang, and P. Craigmile (2005). Geostatistical prediction of spatial extremes and their extent. In P. Renard, H. Demougeot-Renard, and R. Froidevaux (Eds.), *Geostatistics for Environmental Applications, Proceedings of the Fifth European Conference on Geostatistics for Environmental Applications*, pp. 27–37. Dordrecht: Kluwer.
- Davison, A. and R. Smith (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society, Series B* 52, 393–442.
- de Haan, L. (1984). A spectral representation for max-stable processes. *Annals of Probability* 12, 1194–1204.
- de Oliveira, V. and M. Ecker (2002). Bayesian hot spot detection in the presence of a spatial trend: application to total nitrogen concentration in Chesapeake Bay. *Environmetrics* 13, 85–101.
- Deutsch, C. V. and A. G. Journel (1992). *GSLIB: Geostatistical Software Library and User's Guide*. New York: Oxford University Press.
- Dixon, M., J. Tawn, and J. Vassie (1998). Spatial modelling of extreme sea-levels. *Environmetrics* 9, 283–301.
- Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer-Verlag.
- Ferrari, P., A. Frigessi, and P. Gonzaga de Sa (1995). Fast approximate maximum a posteriori restoration of multicolor images. *Journal of the Royal Statistical Society, Series B* 57, 485–500.
- Fisher, R. and L. Tippett (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* 24, 180–190.

- Friel, N. and I. Molchanov (1998a). Distances between grey-scale images. In H. Heijmans and J. Roerdink (Eds.), *Mathematical Morphology and its Applications to Image and Signal processing*, pp. 283–291. Dordrecht: Kluwer.
- Friel, N. and I. Molchanov (1999). A new thresholding technique based on random sets. *Pattern Recognition* 32, 1507–1517.
- Friel, N. and I. S. Molchanov (1998b). A class of error metrics for grey-scale image comparison. In J. L. D. F. Prêteux and E. R. Dougherty (Eds.), *Mathematical Modelling and Estimation Techniques in Computer Vision*, Volume 3457 of *Proceedings of SPIE*, pp. 194–201. Bellingham, WA: SPIE.
- Frigessi, A. and H. Rue (1997). Bayesian image classification with Baddeley’s delta loss. *Journal of Computational and Graphical Statistics* 6, 55–73.
- Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721–741.
- Ghosh, M. (1992). Constrained Bayes estimation with applications. *Journal of the American Statistical Association* 87, 533–540.
- Gilleland, E. and D. Nychka (2005). Statistical models for monitoring and regulating ground-level ozone. *Environmetrics* 16, 535–546.
- Gilleland, E., D. Nychka, and U. Schneider (2006). Spatial models for the distribution of extremes. In J. Clark and A. Gelfand (Eds.), *Computational Statistics: Hierarchical Bayes and MCMC Methods in the Environmental Sciences*. Oxford: Oxford University Press.
- Glendinning, R. (1989). An evaluation of the ICM algorithm for image reconstruction. *Journal of Statistical Computation and Simulation* 31, 169–185.
- Gotway, C. and N. Cressie (1993). Improved multivariate prediction under a general linear model. *Journal of Multivariate Analysis* 45, 56–72.
- Gumbel, E. (1958). *Statistics of Extremes*. New York: Columbia University Press.
- Hsing, T. (1988). On the extreme order statistics for a stationary sequence. *Stochastic Processes and their Applications* 29, 155–169.
- Hsing, T., J. Husler, and M. Leadbetter (1988). On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields* 78, 97–112.
- Journal, A. (1983). Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology* 15, 445–468.
- Journal, A. and C. J. Huijbregts (1978). *Mining Geostatistics*. London: Academic Press.
- Katz, R. W., M. B. Parlange, and P. Naveau (2002). Statistics of extremes in hydrology. *Advances in Water Resources* 25, 1287–1304.

- Lahiri, S. N. (1999). Asymptotic distribution of the empirical spatial cumulative distribution function predictor and prediction bands based on a subsampling method. *Probability Theory and Related Fields* 114, 55–84.
- Lahiri, S. N., M. S. Kaiser, N. Cressie, and N.-J. Hsu (1999). Prediction of spatial cumulative distribution functions using subsampling. *Journal of the American Statistical Association* 94, 86–97.
- Lajaunie, C. (1990). Comparing some approximate methods for building local confidence intervals for predicting regionalized variables. *Mathematical Geology* 22, 123–144.
- Leadbetter, M., G. Lindgren, and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Series*. New York: Springer Verlag.
- Louis, T. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *Journal of the American Statistical Association* 79, 393–398.
- Matheron, G. (1963). Principles of geostatistics. *Economic Geology* 58, 1246–1266.
- Matheron, G. (1976). A simple substitute for conditional expectation: the disjunctive kriging. In *Advanced Geostatistics in the Mining Industry*, pp. 221–236. Dordrecht: Reidel.
- Muller, P. (1998). Simulation based optimal design. In J. Berger, J. Bernardo, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 6*, pp. 459–474. Oxford: Oxford University Press.
- Nandram, B., J. Sedransk, and L. Pickle (2000). Bayesian analysis and mapping of mortality rates for chronic obstructive pulmonary disease. *Journal of the American Statistical Association* 95, 1110–1118.
- North, M. (1980). Time-dependent stochastic models of floods. *American Society of Civil Engineers, Journal of the hydraulics division* 106, 649–655.
- Patil, G. and C. Taillie (2004). Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics* 11, 183–197.
- Robert, C. and G. Casella (2004). *Monte Carlo Statistical Methods* (second edition). New York: Springer-Verlag.
- Rote, G. (1991). Computing the minimum Hausdorff distance between two point sets on a line under translation. *Information Processing Letters* 38, 123–127.
- Rue, H. (1995). New loss functions in Bayesian imaging. *Journal of the American Statistical Association* 90, 900–908.
- Rue, H. (1999). Delta metric, Baddeley’s. In S. Kotz, C. Read, and D. Banks (Eds.), *Encyclopaedia of Statistical Science*, pp. 158–162. New York: Wiley.

- Rue, H. and A. Syversveen (1998). Bayesian object recognition with Baddeley's delta loss. *Advances in Applied Probability* 30, 64–84.
- Schlather, M. and J. A. Tawn (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika* 90, 129–156.
- Serra, J. (1982). *Image Analysis and Mathematical Morphology*. London: Academic Press.
- Shen, X., H.-C. Huang, and N. Cressie (2002). Nonparametric hypothesis testing for a spatial signal. *Journal of the American Statistical Association* 97, 1122–1140. (Correction: 2005, 100:716-718).
- Singpurwalla, N. (1972). Extreme values from a lognormal law with applications to air pollution problems. *Technometrics* 14, 703–711.
- Smith, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript, Department of Mathematics, University of Surrey, Guildford GU2 5XH, England.
- Smith, R. L. (2003). Statistics of extremes, with applications in environment, insurance and finance. In B. Finkenstadt and H. Rootzen (Eds.), *Extreme Values in Finance, Telecommunications and the Environment*, Chapter 1, pp. 1–78. London: Chapman and Hall/CRC Press.
- Stein, A., K. Turkman, P. Bermudez, R. van Heerd, and P. de Bruijn (1999). In search of spatial extremes. In V. Barnett, A. Stein, and K. Turkman (Eds.), *Statistics for the Environment 4: Statistical Aspects of Health and the Environment*, pp. 8–26. Chichester: Wiley.
- Stern, H. and N. Cressie (1999). Inference for extremes in disease mapping. In A. Lawson, A. Biggeri, D. Böhning, E. Lissaffre, and R. Bertollini (Eds.), *Disease mapping and risk assessment for public health*, Chapter Inference for Extremes in Disease Mapping, pp. 63–84. Chichester: Wiley.
- The Natural Environment Research Council (1975). *The Flood Studies Report*. London: The Natural Environment Research Council.
- Todorovic, P. and J. Rousselle (1971). Some problems of flood analysis. *Water Resources Research* 7, 1144–1150.
- Todorovic, P. and E. Zelenhasic (1970). A stochastic model for flood analysis. *Water Resources Research* 6, 1641–1648.
- Waller, L. A. and C. A. Gotway (2004). *Applied Spatial Analysis for Public Health Data*. New York: Wiley.
- Wikle, C. and N. Cressie (1999). A dimension reduced approach to space-time Kalman filtering. *Biometrika* 86, 815–829.

- Wilson, D., A. Baddeley, and R. Owens (1997). A new metric for grey-scale image comparison. *International Journal of Computer Vision* 24, 5–17.
- Wright, D. L., H. S. Stern, and N. Cressie (2003). Loss functions for estimation of extrema with an application to disease mapping. *The Canadian Journal of Statistics* 31, 251–266.
- Zhu, J., S. N. Lahiri, and N. Cressie (2002). Asymptotic inference for spatial CDFs over time. *Statistica Sinica* 12, 843–861.
- Zirschky, J. and D. Harris (1986). Geostatistical analysis of hazardous waste site data. *Journal of Environmental Engineering* 112, 770–783.