

CARNEGIE MELLON UNIVERSITY

NONSTATIONARY GAUSSIAN PROCESSES
FOR REGRESSION AND SPATIAL MODELLING

A DISSERTATION
SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
for the degree
DOCTOR OF PHILOSOPHY
in
STATISTICS

by
Christopher Joseph Paciorek

Department of Statistics
Carnegie Mellon University
Pittsburgh Pennsylvania 15213

May 2003

©Copyright 2003 by Christopher Joseph Paciorek
All Rights Reserved

Abstract

Recent work in the areas of nonparametric regression and spatial smoothing has focused on modelling functions of inhomogeneous smoothness. In the regression literature, important progress has been made in fitting free-knot spline models in a Bayesian context, with knots automatically being placed more densely in regions of the covariate space in which the function varies more quickly. In the spatial statistics literature, attention has focused on using nonstationary covariance structures to account for inhomogeneity of the spatial field.

In this dissertation, I use nonstationary covariance functions in Gaussian process (GP) prior distributions over functions to perform both nonparametric regression and spatial smoothing in a Bayesian fashion. I extend the kernel convolution method of Higdon et al. (1999) to create a class of nonstationary covariance functions. I prove that the nonstationary covariance functions retain the differentiability properties of the stationary correlation functions on which they are based, provided there is sufficient smoothness in the underlying kernel structure used to generate the nonstationarity. The stationary Matérn covariance function has desirable differentiability properties; the generalized kernel convolution method developed here provides a Matérn-based nonstationary covariance function.

I develop a generalized nonparametric regression model and assess difficulties in identifiability and in fitting of the model using Markov Chain Monte Carlo (MCMC) algorithms. Of particular note, I show how to improve MCMC performance for non-Gaussian data based on an approximate conditional posterior mean. The modelling approach produces a flexible response surface that responds to inhomogeneity while naturally controlling overfitting. For Gaussian errors, on test datasets in one dimension, the GP model performs well, but not as well as the free-knot spline method. However, in two and three dimensions, the nonstationary GP model seems to outperform

both free-knot spline models and a stationary GP model. Unfortunately, as implemented the method is not feasible for datasets with more than a few hundred observations because of the computational difficulties involved in fitting the model.

The nonstationary covariance model can also be embedded in a spatial model. In particular, I analyze spatiotemporal climate data, using a nonstationary covariance matrix to model the spatial structure of the residuals. I demonstrate that the nonstationary model fits the covariance structure of the data better than a stationary model, but any improvement in point predictions relative to a stationary model or to the maximum likelihood estimates is minimal, presumably because the data are very smooth to begin with. My comparison of various correlation models for the residuals highlights the difficulty in fitting high-dimensional covariance structures.

Acknowledgements

I would like to acknowledge the advice, suggestions, support, and friendship of a number of people who helped me during the writing of this thesis and the rest of my time as a graduate student in the Department of Statistics. First, I would like to thank my advisor, Mark Schervish, for his ongoing involvement in this work. In his understated way, Mark offered hands-on advice without hands-on management, giving suggestions that often guided me in a better direction. He has always been willing to delve into the details of the work and talk through issues, sometimes long past a reasonable meeting length. I would also like to thank the other members of my committee. James Risbey got me started on the Advanced Data Analysis project that led to this thesis. I thank him for his collaboration, advice, and friendship. I thank Doug Nychka for an enjoyable and productive visit to NCAR as well as for his various suggestions and his assistance with the wavelet model. I thank Valérie Ventura for her support and for her collaboration on two related projects. Larry Wasserman offered an open door, of which I did not take as much advantage as I would have liked toward the close of my graduate work. The statistics department staff and the rest of the faculty in the department have helped me in various ways, and I thank them for making the department a great place to do graduate work. I'd also like to thank Chris Holmes for the Matlab code on his website.

I want to thank the graduate students in the department. In particular, Fang Chen, my longtime officemate, has tolerated with good humor my often foul moods during the writing of this thesis.

Seven years is a long time to be in graduate school; I'm glad Susan Davidson was there for many of them. I'd also like to thank old friends who do not need to be named to know their importance to me. Little did we know in grade school and high school that I would be 31 when I finally finished school. John, how many seconds have I been in school?

Finally, but most importantly, I thank my parents for setting me on the road with what I would need to get to this point.

Contents

Abstract	iii
Acknowledgements	v
1 Introduction	1
1.1 Problem Definition	1
1.2 Gaussian Processes and Covariance Functions	2
1.3 Spatial Smoothing Methods	5
1.4 Nonparametric Regression Methods	11
1.4.1 Gaussian process methods	11
1.4.2 Other methods	13
1.4.3 Smoothing as local linear averaging	19
1.4.4 Modelling non-Gaussian data	21
1.5 Thesis Outline	21
1.6 Contributions	22
2 Theoretical Development	25
2.1 Introduction	25
2.2 Nonstationary Covariance Functions Using Convolutions of Kernels	25
2.3 Generalized Kernel Convolution Covariance Functions	28
2.4 Nonstationary Covariance on the Sphere	32
2.5 Smoothness Properties of Covariance Functions	33

2.5.1	Overview	33
2.5.2	Theoretical framework	34
2.5.3	Lemmas for proofs	39
2.5.4	Smoothness properties of isotropic correlation functions	43
2.5.4.1	Mean square differentiability and scale mixtures	43
2.5.4.2	Sample path differentiability and scale mixtures	45
2.5.4.3	Application of results to specific correlation functions	46
2.5.4.4	Mean square analyticity	47
2.5.5	Smoothness properties of kernel convolution covariance functions	48
2.5.5.1	Mean square differentiability	48
2.5.5.2	Sample path differentiability	50
2.5.5.3	Implications for nonstationary modelling	53
2.6	Discussion	53
3	Methodological Development	57
3.1	Introduction	57
3.2	Parameterizing the Kernels for Nonstationary Covariance Models	58
3.2.1	Two-dimensional foci approach	58
3.2.2	Cholesky decomposition	59
3.2.3	Eigendecomposition	60
3.2.3.1	Givens angles	60
3.2.3.2	Overparameterized eigenvectors	62
3.2.3.3	Simplified eigendecomposition model	63
3.2.4	Basis kernel model	64
3.3	Numerical Sensitivity of GP Models	65
3.4	GP Parameterizations	70
3.4.1	Centered vs. noncentered parameterizations	70
3.4.2	Discretized parameterization	76
3.5	Model Dimensionality and Parameter Identifiability	79
3.5.1	Smoothing and dimensionality in the GP model	79

3.5.1.1	Occam's razor	79
3.5.1.2	Model dimension	80
3.5.1.3	Covariate selection	82
3.5.2	Parameter identifiability and interpretability	83
3.5.2.1	Uncentered parameterization	83
3.5.2.2	Centered parameterization	84
3.6	MCMC Sampling Schemes for GP Models	86
3.6.1	Integrating the process out of the model	86
3.6.2	Methods for sampling the process values	88
3.6.2.1	Derivative-based methods	89
3.6.2.2	Posterior mean centering	92
3.6.2.3	Application of posterior mean centering to non-Gaussian data	96
3.7	Computational Challenges of GP Models	104
3.7.1	Local methods	105
3.7.2	Sparse methods	105
3.7.3	Approximate matrix calculations	107
3.7.4	Parallel processing	108
3.7.5	Non-Bayesian approaches	109
3.7.6	Fast Fourier transform	109
3.7.7	Overview	109
4	Regression Model Results	111
4.1	Introduction	111
4.2	Model Structure	111
4.3	MCMC Sampling Scheme	115
4.4	Evaluation Procedures	118
4.4.1	Mean squared error	118
4.4.2	Predictive density	119
4.4.3	Kullback-Leibler divergence	120
4.5	Test Datasets and Competing Methods	121

4.6	Results	124
4.6.1	One-dimensional assessment	124
4.6.2	Higher-dimensional assessment	137
4.6.3	Convergence	143
4.6.4	Example using non-Gaussian data	143
4.7	Discussion	148
5	Spatial Model Results	151
5.1	Introduction	151
5.2	Scientific Problem	152
5.3	Basic Spatial Model	156
5.4	Covariance Modelling	160
5.4.1	Correlation models used in this work	160
5.4.1.1	Independence model	160
5.4.1.2	Stationary Matérn correlation	161
5.4.1.3	Nonstationary smoothed empirical correlation	161
5.4.1.4	Kernel-based Matérn nonstationary correlation	163
5.4.2	Alternate approaches not used in this work	165
5.4.2.1	Smoothing the MLE fields	165
5.4.2.2	Smoothing the empirical covariance toward a stationary covariance	165
5.4.2.3	Fitting an unconstrained data covariance	166
5.4.2.4	Covariances modelled indirectly through basis function regression	166
5.5	Fitting the Model	167
5.5.1	Data and data-dependent model specifications	167
5.5.2	MCMC sampling procedure	169
5.5.2.1	Posterior mean centering	169
5.5.2.2	Sampling steps	170
5.5.2.3	Running the chain	173
5.6	Model Comparison Criteria	174

5.7	Results	175
5.7.1	Convergence	175
5.7.2	Model estimates	180
5.7.3	Model comparison	196
5.7.4	Trend significance	199
5.8	Discussion	204
5.8.1	Covariance modelling assessment	204
5.8.2	Other approaches	207
6	Conclusions and Future Work	209
6.1	Summary and Contributions	209
6.2	Shortcomings and Potential Improvements	212
6.3	Future Work	215
	Appendices	218
A	Spatial Model Prior Distributions and Starting Values	219
A.1	Prior distributions	219
A.2	Starting values	220
B	Notation	221
B.1	Parameters, stochastic processes, matrices, data, and covariates	221
B.2	Indices	223
B.3	Symbols, superscripts, and subscripts	223
B.4	Functions	223
	References	224

List of Tables

2.1	Smoothness properties of Gaussian processes parameterized by various correlation functions. The asterisk indicates that the sample path part of this statement is a conjecture. In Section 2.5.4.2 I prove only that the Matérn is $\lceil \frac{\nu}{2} - 1 \rceil$ times sample path differentiable	44
3.1	Number of iterations that can be completed in the same time as a single iteration of the uncentered scheme.	100
3.2	Effective sample size (ESS) by sampling scheme for key model parameters. \bar{f} is the mean ESS for the function values, averaging over $f(x)$ for all 100 values of x	101
3.3	ESS by sampling scheme for key parameters, adjusted for computational speed. \bar{f} is the mean ESS for the function values, averaging over $f(x)$ at all 100 values of x	103
4.1	MSE for training and test sets for the four methods on a portion of the Wood dataset.	139
4.2	LPD for training and test sets for the four methods on a portion of the Wood dataset, averaged across observations.	141
4.3	MSE for training and test sets for the four methods, as well as linear regression and a generalized additive model (GAM) on the ozone dataset.	142
4.4	LPD for training and test sets for the four methods on the ozone data, averaged across observations.	143
4.5	Effective sample size (ESS) by sampling scheme for key model parameters for the Tokyo rainfall dataset. \bar{f} is the mean ESS for the function values, averaged over 10 randomly sampled calendar days, and $\bar{\lambda}$ is the mean ESS for the log of the kernel eigenvalues, averaged over 10 randomly sampled calendar days.	145

- 5.1 Log predictive density comparisons of the six models on test data for the two datasets. Values in parentheses are ranges based on blocked values from the posterior simulations. 197
- 5.2 MSE comparisons of the six models on test data for the two datasets. Values in parentheses are ranges based on blocked values from the posterior simulation. . . . 198

List of Figures

2.1	(a) Correlation of $f(-0.5)$ with the function at all other points. (b) Correlation of $f(0.5)$ with the function at all other points. (c) Kernel variance as a function of x . (d) Five sample functions drawn from the Gaussian process distribution; notice that the functions are least smooth at the location of the sharp change in the kernel size.	29
3.1	Sample function drawn from $f(\cdot) \sim \text{GP}(0, R(\kappa = 0.17, \nu = 25))$ with an inaccurate generalized Cholesky factor. Note the jaggedness in the function at several points.	69
3.2	Effect of covariate order on κ_λ mixing in a regression problem with $f(x) = \sin \frac{1}{x}$, $x \in [0.1, 0.7]$; the plots are based on subsampling every 10th iteration from chains of length 100,000. I ran two runs, the first with $x_1 = 0.7$, $x_{100} = 0.1$, and proposal standard deviation of 0.24 (acceptance rate of 34%), and the second with $x_1 = 0.1$, $x_{100} = 0.7$, and proposal standard deviation of 0.28 (acceptance rate of 41%). (a) time series plot with $x_1 = 0.7$, (b) time series plot with $x_1 = 0.1$, (c) ACF with $x_1 = 0.7$, (d) ACF with $x_1 = 0.1$	71

3.3 Sample function values from an MCMC in a Bernoulli data example with two different values of κ : (a) Sample function (solid line) and conditional posterior mean (dashed line) with $\kappa = 0.70$. (b) Proposing $\kappa^* = 0.30$ and \mathbf{f}^* conditional on κ^* using the PMC proposal induces the PMC sample function proposal (solid line) and conditional posterior mean (dashed line). The dotted line is the sample function that would be proposed based on a joint proposal for (κ, \mathbf{f}) without posterior mean centering. Notice that the function proposed without PMC is more extreme than the PMC proposal. Also notice that the conditional posterior mean and sample function proposal are less smooth in (b), but the deviations of the sample function in (a) and the PMC sample function proposal in (b) about their conditional means have similar structure. 95

3.4 Time series plots of μ , σ , and κ for the five basic sampling schemes. 102

4.1 Directed acyclic graph for the normal likelihood nonstationary Gaussian process regression model. Bold letters indicate vectors. 113

4.2 Boxplots of (a) MSE and (b) KL divergence for the three methods over 50 simulated datasets of example 1: Bayesian adaptive regression splines (BARS), nonstationary GP (NSGP), and stationary GP (SGP). 125

4.3 (a) BARS and (b) NSGP fit to one data sample in which NSGP has lower MSE than BARS. (c) BARS and (d) NSGP fit to a second data sample in which BARS has lower MSE. The thick dashed line is the true function, the solid line is the posterior mean estimate, and the thin dashed lines are 95% pointwise credible intervals. . . . 126

4.4 95% confidence intervals for the mean MSE over simulated datasets of example 1. SARS, M-D (Modified-DMS) and B10 (BARS) are based on 10 sample datasets as calculated in DiMatteo et al. (2002), while B50 (BARS), NSGP (nonstationary GP) and SGP (stationary GP) are based on 50 sample datasets as calculated here. . . 127

4.5 Boxplots of (a) MSE and (b) KL divergence for the three methods over 50 simulated datasets of example 2: Bayesian adaptive regression splines (BARS), nonstationary GP (NSGP), and stationary GP (SGP). For SGP, one outlier with KL divergence of 0.58 is not plotted. 128

- 4.6 (a) BARS and (b) nonstationary GP fit to one data sample of example 2. The thick dashed line is the true function, the solid line is the posterior mean estimate, and the thin dashed lines are 95% pointwise credible intervals. 129
- 4.7 95% confidence intervals for the mean MSE over simulated datasets of example 2. SARS, M-D (Modified-DMS) and B10 (BARS) are based on 10 sample datasets as calculated in DiMatteo et al. (2002), while B50 (BARS), NSGP (nonstationary GP) and SGP (stationary GP) are based on 50 sample datasets as calculated here. . . 130
- 4.8 Boxplots of (a) MSE and (b) KL divergence for the three methods over 50 simulated datasets of example 3: Bayesian adaptive regression splines (BARS), nonstationary GP (NSGP), and stationary GP (SGP). One outlying value for SGP is omitted in both plots (MSE=0.87, KL=0.54). The two outliers for NSGP appear to be datasets for which the MCMC did not mix well because the proposal variances were the same for the MCMCs for all 50 data samples. 131
- 4.9 (a) BARS and (b) nonstationary GP fit to one data sample of example 3. The thick dashed line is the true function, the solid line is the posterior mean estimate, and the thin dashed lines are 95% pointwise credible intervals. 132
- 4.10 95% confidence intervals for the mean MSE over simulated datasets of example 3. SARS, M-D (Modified-DMS) and B10 (BARS) are based on 10 sample datasets as calculated in DiMatteo et al. (2002), while B50 (BARS), NSGP (nonstationary GP) and SGP (stationary GP) are based on 50 sample datasets as calculated here. . . 133
- 4.11 (a) Kernel size (standard deviation of kernels) as a function of the covariate for one posterior sample for example 3. (b) Implicit smoothing kernels at six covariate values, with different line types and/or widths for each implicit kernel. 133
- 4.12 (a) Sample posterior regression function with $\nu_f = 1.3$ (solid line) and true function (dashed line) from example 2. (b) Kernel size (standard deviation of kernel) as a function of the covariate for the sample shown in (a). (c) Sample posterior regression function with $\nu_f = 4.5$ (solid line) and true function (dashed line). (d) Kernel size (standard deviation of kernel) as a function of the covariate for the sample shown in (c). 135

4.13 (a) Sample posterior regression function from one data sample (solid line) and true function (dashed line) for example 2. (b) Kernel size (standard deviation of kernel) as a function of the covariate for the sample shown in (a). Notice the lack of smoothness in the function for $-2 < x < -0.5$, where the kernel sizes are large but variable. (c) Sample posterior regression function for a different data sample of example 2 (solid line) and true function (dashed line). (d) Kernel size (standard deviation of kernel) as a function of the covariate for the sample shown in (c). Notice that the nonintuitive sharp increase in the kernel size is what allows the model to capture the function jump at $x = 0$ 136

4.14 Sample posterior regression function from example 2 using a stationary GP model; here $\nu_f = 0.69$ and $\kappa_f = 4.0$ 137

4.15 Perspective plots of (left) true Hwang function and (right) posterior mean function using the nonstationary GP model. Note that these plots involve interpolation by the `interp` and `persp` functions in the R statistical package; the `interp` function is found in the `akima` library. 139

4.16 Contour plots of (a) true Hwang function, (b) posterior mean function using the nonstationary GP model, and (c) difference between the true function and the estimated function. Note that these plots involve interpolation by the `interp` and `contour` functions in the R statistical package; the `interp` function is found in the `akima` library. 140

4.17 Boxplots of FVU for (a) training covariates and (b) test covariates over 50 simulated datasets of the Hwang function for the four methods 140

4.18 Boxplots of KL divergence for (a) training covariates and (b) test covariates over 50 simulated datasets of the Hwang function for the four methods. 141

4.19 Time series plots for the Tokyo rainfall dataset for model log likelihood, log prior density, degrees of freedom of the conditional posterior mean function, and hyperparameters. 146

4.20 Time series plots for the Tokyo rainfall dataset for function values, $f(\cdot)$ (first row), and log of kernel eigenvalues, $\log \lambda(\cdot)$ (second row), at four covariate values. . . . 146

- 4.21 Posterior mean estimate of $p(\cdot)$, the probability of rainfall as a function of calendar day, with 95% pointwise credible intervals. Dots are empirical probabilities of rainfall based on the two binomial trials. 147
- 4.22 Posterior geometric mean kernel size as a function of calendar day. The kernel sizes are plotted as the square roots of the geometric means of the kernel eigenvalues, and hence can be thought of as correlation scale parameters, with units of days. . . 147
- 5.1 Plots of residual correlation of storm activity between all locations and each of two focal locations, each marked with an ‘F’: (left) 50° N, 330° E (in the North Atlantic) and (right) 30° N, 30° E (in Egypt). The residual correlation is calculated after removing location-specific linear trends. Gray shading indicates the value of the correlation coefficient, with darker colors indicating values large in magnitude. Negative values are indicated by horizontal hatching. The high topography of the Himalayas is blacked out. 152
- 5.2 Directed acyclic graph of nonstationary spatial model. Bold letters indicate vectors. 157
- 5.3 Map of the Northern hemisphere, $20^\circ - 75^\circ$ N, with $5^\circ \times 5^\circ$ grid overlaid as dotted lines and Pacific (P) and Atlantic (A) region boundaries indicated by the thick dark lines of longitude. 168
- 5.4 Time series plots of the log posterior density for temperature variance for the four Bayesian models: (a) stationary, (b) kernel nonstationary, (c) wavelet-empirical, and (d) wavelet-smooth. 177
- 5.5 Time series plots of the hyperparameters, μ (first column), σ (second column), and κ (third column) for α (first row), β (second row), and η^2 (third row) from the kernel nonstationary model fit to the temperature variance data by MCMC. 178
- 5.6 Time series plots of three process values of α (first row), β (second row) and η^2 (third row) for the kernel nonstationary model fit to the temperature variance by MCMC. 179
- 5.7 Maps of estimated β values for temperature variance for (a) MLE model, and posterior means from (b) stationary model, (c) kernel nonstationary model, (d) wavelet-smoothed covariance model, and (e) wavelet-empirical covariance model. . 181

- 5.8 Maps of estimated η^2 values for temperature variance for (a) MLE model, and posterior means from (b) stationary model, (c) kernel-based nonstationary model, (d) wavelet-smoothed covariance model, and (e) wavelet-empirical covariance model. 182
- 5.9 Scatterplots of model estimates (posterior means) of intercept (column 1), slope (column 2), and residual variance (column 3) fields compared to the MLE values for the four models: stationary (row 1), kernel nonstationary (row 2), wavelet-smoothed (row 3) and wavelet-empirical (row 4) for temperature variance. 184
- 5.10 (a) Plot of standardized residuals (defined in text) as a function of location for temperature variance; these residuals are calculated based on the posterior mean parameters, but using the MLEs for the residual variances. (b) Plot of the ratio of the posterior mean residual variance estimates to the MLE variance estimates. . . . 185
- 5.11 Scatterplots of model estimates (posterior means) of intercept (column 1), slope (column 2), and residual variance (column 3) fields compared to the MLE values for the models for Eady growth rate: stationary (row 1) and kernel nonstationary (row 2). 187
- 5.12 Mean basis kernels for (a) temperature variance in the Pacific region and (b) Eady growth rate in the Atlantic. For each posterior sample, I represent the kernel as a constant density ellipse from a normal density with covariance matrix equal to the basis kernel matrix. Mean ellipses are then plotted using the average distances from the ellipse origin to the ellipse itself at 44 different angles, averaged over the posterior samples. Note that since the kernels are plotted on a latitude-longitude grid, distances toward the top of the plot are exaggerated and the true basis kernels there are smaller in size than represented here. 189
- 5.13 Plot of posterior mean correlation structure from the kernel nonstationary model for temperature variance between each of nine focal locations and all 288 locations. Correlation structures at the nine focal locations are overlaid on the same plot because correlations are less than 0.20 except in the bullseye areas. The nine focal locations are at the centers of the bullseyes and are the same locations as the centers of the basis kernels, as listed in the text. 190

- 5.14 Plot of posterior mean correlation structure from the kernel nonstationary model for Eady growth between each of nine focal locations and all 288 locations. Plots are overlaid because correlations are less than 0.20 except in the bullseye areas. The nine focal locations are at the centers of the bullseyes and are the same locations as the centers of the basis kernels, as listed in the text. 190
- 5.15 Plot of posterior mean correlation structure from the stationary model for temperature variance between each of nine focal locations and all 288 locations. Correlation structure appears different at different latitudes because of the distortion induced by the latitude-longitude grid. Other details are as in Figure 5.13. 191
- 5.16 Plot of posterior mean correlation structure from the stationary model for Eady growth between each of nine focal locations and all 288 locations. Correlation structure appears different at different latitudes because of the distortion induced by the latitude-longitude grid. Other details are as in Figure 5.14. 191
- 5.17 Plots of empirical correlations for temperature variance between each of the nine focal locations and all 288 locations. Each subplot displays the correlation structure for one focal location (marked by ‘X’) with latitude and longitude increasing from bottom to top and left to right respectively: (a) 150° E, 60° N, (b) 190° E, 60° N, (c) 230° E, 60° N, (d) 150° E, 45° N, (e) 190° E, 45° N, (f) 230° E, 45° N, (g) 150° E, 30° N, (h) 190° E, 30° N, (i) 230° E, 30° N. 192
- 5.18 Plots of empirical correlations for Eady growth between each of the nine focal locations and all 288 locations. Each subplot displays the correlation structure for one focal location (marked by ‘X’) with latitude and longitude increasing from bottom to top and left to right respectively: (a) 150° E, 60° N, (b) 190° E, 60° N, (c) 230° E, 60° N, (d) 150° E, 45° N, (e) 190° E, 45° N, (f) 230° E, 45° N, (g) 150° E, 30° N, (h) 190° E, 30° N, (i) 230° E, 30° N. 193
- 5.19 Plots of wavelet-empirical model correlations between each of nine focal locations and all 288 locations for temperature variance. Details are as in Figure 5.17. 194
- 5.20 Plots of wavelet-smooth model correlations between each of nine focal locations and all 288 locations for temperature variance. Details are as in Figure 5.17. 195

- 5.21 Scatterplot of standard error estimates as a function of the point estimates for the linear trends in both the MLE and Bayesian nonstationary models for the 288 locations of temperature variance in the Pacific. Points in the areas toward the outer sides of the plot relative to the nearly vertical lines are individually significant based on the point estimates being at least two standard errors away from zero. 200
- 5.22 Scatterplot of standard error estimates as a function of the point estimates for the linear trends in both the MLE and Bayesian nonstationary models for the 288 locations of Eady growth rate in the Atlantic. Points in the areas toward the outer sides of the plot relative to the nearly vertical lines are individually significant based on the point estimates being at least two standard errors away from zero. 202