Appendices

A Spatial Model Prior Distributions and Starting Values

A.1 Prior distributions

I chose prior distributions to be diffuse but proper, based on the mean values of temperature variance and Eady growth over 1949-1999 for the Northern hemisphere, $20° - 70°$ N, using Paciorek et al. (2002, Fig. 1). These climatological mean fields gave me a sense for the reasonable range of values for the parameters. In particular, for temperature variance, I take

$$\begin{aligned}
\mu_\alpha &\sim \mathrm{N}(1.2, 0.7^2)\\
\sigma_\alpha^2 &\sim \mathrm{IG}(0.1, 0.1)\\
\mu_\beta &\sim \mathrm{N}(0, 0.01^2)\\
\sigma_\beta^2 &\sim \mathrm{IG}(0.1, 1 \times 10^{-7})\\
\mu_\eta &\sim \mathrm{N}(-4.0, 3.0^2)\\
\sigma_\eta^2 &\sim \mathrm{IG}(0.1, 0.1)\\
\log \kappa_\phi &\sim \mathrm{N}(-5.4, 1.2^2)\\
\log \delta &\sim \mathrm{U}(-23.0, 2.3)\\
\nu_{\boldsymbol{Y}} &\sim \mathrm{U}(0.5, 15.0),
\end{aligned}$$

where κ_ϕ indicates that I use the same prior for $\phi \in \{\alpha, \beta, \eta\}$. For the inverse gamma distributions, these are parameterized such that the mean is $\frac{\beta}{\alpha-1}$. For the stationary model, I take $\log \kappa_{\boldsymbol{Y}} \sim \mathrm{N}(-5.4, 1.2^2)$. For the nonstationary model, $k = 1, \ldots, 9$, I take $\log \lambda_k \sim \mathrm{U}(-5.6, 7.8)$ for each of the two eigenvalues in the kth basis kernel matrix. I take the kth basis kernel matrix Givens angle, $\gamma_k \sim \text{U}(0, \pi)$. Finally, I take the weight decay parameter, $\log \kappa_{\boldsymbol{Y}} \sim \text{U}(-2.3, 1.6)$. The priors for parameters that affect various correlation scales are informed by the fact that I do not want the scale less than the smallest distance between grid points or much larger than the largest distance between grid points.

For Eady growth rate, I use the same parameters, with the following exceptions:

$$\begin{aligned}
\mu_\alpha &\sim \text{N}(0.5, 0.5^2) \\
\sigma_\alpha^2 &\sim \text{IG}(0.1, 0.001) \\
\mu_\beta &\sim \text{N}(0.0, 0.005^2) \\
\sigma_\beta^2 &\sim \text{IG}(0.1, 1 \times 10^{-8}) \\
\mu_\eta &\sim \text{N}(-7.0, 6.0^2).
\end{aligned}$$

For the wavelet models, $R_{\boldsymbol{Y}}$ is fixed, so I only have priors for the remaining parameters, which I take to be the same as for the kernel nonstationary model.

A.2 Starting values

For starting values for the hyperparameters of the $\alpha(\cdot), \beta(\cdot)$ and $\log \eta(\cdot)^2$ processes, I calculated approximate maximum likelihood estimates (MLEs) based on $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\eta}}^2$, namely the MLEs for the process values based on assuming independent locations, and used these approximate MLEs for the hyperparameters to come up with reasonable starting values. For temperature variance I use $\mu_\alpha = 1.2$, $\log \sigma_\alpha = -1.2$, $\log \kappa_\alpha = -1.3$, $\mu_\beta = 0.0$, $\log \sigma_\beta = -6.9$, $\log \kappa_\beta = -1.3$, $\mu_\eta = -4.0$, $\log \sigma_\eta = 0.0$, $\log \kappa_\eta = -1.3$, $\log \delta = -9.2$, and $\nu_{\boldsymbol{Y}} = 4.0$. For the stationary model, I use $\log \kappa_{\boldsymbol{Y}} = -2.2$ and for the nonstationary model, I take $\log \kappa_{\boldsymbol{Y}} = -1.2$ and started the basis kernel matrix eigenvalues at 1.0 and the Givens angles at $\frac{\pi}{2}$. I construct the process values, $\boldsymbol{\phi} = \mu_\phi + \sigma_\phi L_{\boldsymbol{\phi}} \boldsymbol{\omega}_\phi$, using $\boldsymbol{\omega}_\phi \sim \text{N}(0, I)$.

For the Eady growth models, I used the same initial values, with the following exceptions. I took $\mu_\alpha = 0.72$, $\log \sigma_\alpha = -1.8$, $\log \kappa_\alpha = -1.1$, $\mu_\beta = 0.00012$, $\log \sigma_\beta = -6.6$, $\log \kappa_\beta = -1.3$, $\mu_\eta = -5.2$, $\log \sigma_\eta = -0.7$, and $\log \kappa_\eta = -1.5$. For the stationary model, I used $\log \kappa_{\boldsymbol{Y}} = -2.38$.

For the wavelet models, I again used the same values, except for the parameters involved in $R_{\boldsymbol{Y}}$, which are not used.

B Notation

I have attempted to be consistent in my notation, both in my use of alphabets and cases, as well as my use of individual letters and symbols. However, in a work as large as this, I have needed in some situations to use the same symbol in different contexts, and there are also undoubtedly places where I have not been entirely consistent.

In general, I have indicated functions with both lower and upper case Arabic letters, matrices with upper case Arabic and Greek letters, vectors with bold Arabic and Greek letters, and parameters with lower-case Greek letters. For indices, I have used lower case Arabic letters.

For random variables that are parameters, I have been lax and used lower case Greek letters to indicate the random variable itself and realizations of the random variable.

In various places in the thesis, I need a vector-valued mean for a vector-valued random variable; as necessary I take $\mu = \mu\mathbf{1}$.

Next I list the notation and meanings, broadly grouped.

B.1 Parameters, stochastic processes, matrices, data, and covariates

$f(\cdot), \boldsymbol{f}, f_i, f(\boldsymbol{x_i})$: a regression function/process, a vector of values of the function evaluated at a finite set of covariates, the value of the function at the ith covariate value, the value of the function at $\boldsymbol{x_i}$

$\phi(\cdot), \boldsymbol{\phi}, \phi_i, \phi(\boldsymbol{x_i})$: a stochastic process, a vector of values of the process evaluated at a finite set of covariates, the value of the process at the ith covariate value, the value of the process at $\boldsymbol{x_i}$

$Z(\cdot), Z(\boldsymbol{x_i})$: a stochastic process, the value of the process at $\boldsymbol{x_i}$

$\alpha(\cdot), \beta(\cdot), \eta(\cdot)^2$: intercept, slope, and residual variance processes in the spatial model

η^2: error (noise) variance in the regression model

$\boldsymbol{Y}, \boldsymbol{y}$: vector of data values as a random variable, as a realization

$\boldsymbol{x}_i, \boldsymbol{x}_j$: two different covariate values, $\boldsymbol{x}_i \in \Re^P$

Σ_i : positive definite kernel matrix

R : correlation matrix

C : covariance matrix

Γ: eigenvector matrix

Λ: eigenvalue matrix

Q: quadratic form in nonstationary correlation function

τ : Euclidean distance

κ : correlation scale parameter, in units of distance

ν : smoothness parameter in Matérn correlation function

μ : mean of a stochastic process

σ^2 : variance of a stochastic process

ρ : angle or angular distance

$\theta, \boldsymbol{\theta}$: a parameter or vector of parameters

$\boldsymbol{\omega}$: a vector of values with a standard normal prior or drawn from a standard normal, or white noise values in general

$\psi, \boldsymbol{\psi}$: value(s) used in generating an MCMC proposal

v : proposal standard deviation in an MCMC

ϵ : tolerance in numerical calculations

$\boldsymbol{u}$: spatial location

S, s : scale parameter

W, w : spectral random variable

λ : eigenvalue

γ : parameter used in constructing eigenvectors

c: a constant

B.2 Indices

i : indexes training set values $(1, \ldots, n)$

j : indexes test set values $(1, \ldots, m)$

k : indexes MCMC draws or number of components in a model (1,...,K)

m : indexes derivatives $(1, \ldots, M)$

p : indexes dimension of the covariate space $(1, \ldots, P)$

t : indexes time in the spatial model $(1, \ldots, T)$

$\boldsymbol{x}_i, \boldsymbol{x}_j$: two different covariate values, $\boldsymbol{x}_i \in \Re^P$

$\boldsymbol{f_1}, \boldsymbol{f_2}$: training set values of f, test set values of f

B.3 Symbols, superscripts, and subscripts

$f^{(m)}$: the mth derivative of the function, f

$f_{(k)}$: kth MCMC draw

$\hat{\boldsymbol{\phi}}$: maximum likelihood estimate

$\check{f}$: the true value of a parameter

$\tilde{f}$: posterior mean

$\widetilde{f|\mu}$: conditional posterior mean

B.4 Functions

$R(\cdot), R(\cdot, \cdot)$: stationary correlation function, nonstationary correlation function

$C(\cdot), C(\cdot, \cdot)$: stationary covariance function, nonstationary covariance function

$g(\cdot)$: used to indicate functions in various contexts

$h(\cdot), H(\cdot)$: density function, distribution function

References

Abrahamsen, P. (1997), "A review of Gaussian random fields and correlation functions," Technical Report 917, Norwegian Computing Center.

Abramowitz, M. and Stegun, I. (1965), *Handbook of Mathematical Functions*, New York: Dover.

Adler, R. (1981), *The Geometry of Random Fields*, New York: John Wiley & Sons.

Anderson, T. W., Olkin, I., and Underhill, L. G. (1987), "Generation of random orthogonal matrices," *SIAM Journal on Scientific and Statistical Computing*, 8, 625–629.

Bakin, S., Hegland, M., and Osborne, M. (2000), "Parallel MARS algorithm based on B-splines," *Computational Statistics*, 15, 463–484.

Barnard, J., McCulloch, R., and Meng, X.-L. (2000), "Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage," *Statistica Sinica*, 10, 1281–1311.

Benjamini, Y. and Hochberg, Y. (1995), "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society B*, 57, 289–300.

Biller, C. (2000), "Adaptive Bayesian regression splines in semiparametric generalized linear models," *Journal of Computational and Graphical Statistics*, 9, 122–140.

Billingsley, P. (1995), *Probability and Measure* (third ed.), New York: John Wiley & Sons.

Bochner, S. (1959), *Lectures on Fourier Integrals*, Princeton, N.J.: Princeton University Press.

Breslow, N. E. and Clayton, D. G. (1993), "Approximate inference in generalized linear mixed models," *J. Am. Stat. Assoc.*, 88, 9–25.

Brockmann, M., Gasser, T., and Herrmann, E. (1993), "Locally adaptive bandwidth choice for kernel regression estimators," *J. Am. Stat. Assoc.*, 88, 1302–1309.

Brockwell, A. and Kadane, J. (2002), "Identification of regeneration times in MCMC simulation, with application to adaptive schemes," Technical Report 770, Department of Statistics, Carnegie Mellon University.

Bruntz, S., Cleveland, W., Kleiner, B., and Warner, J. (1974), "The dependence of ambient ozone on solar radiation, temperature, and mixing height," in *Symposium on Atmospheric Diffusion and Air Pollution*, ed. A. M. Society, pp. 125–128.

Buck, R. C. (1965), *Advanced Calculus* (second ed.), New York: McGraw-Hill.

Cambanis, S. (1973), "On some continuity and differentiability properties of paths of Gaussian processes," *Journal of Multivariate Analysis*, 3, 420–434.

Casella, G. and Berger, R. (1990), *Statistical Inference*, Belmont, California: Duxbury Press.

Chipman, H., George, E., and McCulloch, R. (1998), "Bayesian CART model search," *J. Am. Stat. Assoc.*, 93, 935–960.

Christensen, O., Møller, J., and Waagepetersen, R. (2000), "Analysis of spatial data using generalized linear mixed models and Langevin-type Markov chain Monte Carlo," Technical Report R-002009, Department of Mathematics, Aalborg University.

——— (2001), "Geometric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalized linear mixed models," *Methodology and Computing in Applied Probability*, 3, 309–327.

Christensen, O., Roberts, G., and Sköld, M. (2003), "Robust MCMC methods for spatial GLMMs," *in preparation*.

Christensen, O. F. and Waagepetersen, R. (2002), "Bayesian prediction of spatial count data using generalized linear mixed models," *Biometrics*, 58, 280–286.

Churchill, R. (1960), *Complex variables and applications* (second ed.), New York: McGraw-Hill.

Cleveland, W. and Devlin, S. (1988), "Locally-weighted regression: an approach to regression analysis by local fitting," *J. Am. Stat. Assoc.*, 83, 597–610.

Cohen, A. and Jones, R. H. (1969), "Regression on a random field," *J. Am. Stat. Assoc.*, 64, 1172–1182.

Cramér, H. and Leadbetter, M. (1967), *Stationary and Related Stochastic Processes*, New York: John Wiley & Sons.

Cressie, N. (1993), *Statistics for Spatial Data* (Revised ed.): Wiley-Interscience.

Damian, D., Sampson, P., and Guttorp, P. (2001), "Bayesian estimation of semi-parametric non-stationary spatial covariance structure," *Environmetrics*, 12, 161–178.

Daniels, M. and Kass, R. (1999), "Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models," *J. Am. Stat. Assoc.*, 94, 1254–1263.

Das, B. (2000), *Global Covariance Modeling: a Deformation Approach to Anisotropy*, unpublished Ph.D. dissertation, University of Washington, Department of Statistics.

Denison, D., Holmes, C., Mallick, B., and Smith, A. (2002), *Bayesian Methods for Nonlinear Classification and Regression*, Chichester, England: Wiley.

Denison, D., Mallick, B., and Smith, A. (1998a), "Automatic Bayesian curve fitting," *Journal of the Royal Statistical Society, Series B*, 60, 333–350.

——— (1998b), "Bayesian MARS," *Statistics and Computing*, 8, 337–346.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998), "Model-based geostatistics," *Applied Statistics*, 47, 299–326.

DiMatteo, I. (2001), *Bayesian Curve Fitting Using Free-Knot Spline*, unpublished Ph.D. dissertation, Carnegie Mellon University, Department of Statistics.

DiMatteo, I., Genovese, C., and Kass, R. (2002), "Bayesian curve-fitting with free-knot splines," *Biometrika*, 88, 1055–1071.

Donoho, D. L. and Johnstone, I. M. (1995), "Adapting to unknown smoothness via wavelet shrinkage," *J. Am. Stat. Assoc.*, 90, 1200–1224.

Doob, J. (1953), *Stochastic Processes*, New York: John Wiley & Sons.

Duane, S., Kennedy, A., Pendleton, B., and Roweth, D. (1987), "Hybrid Monte Carlo," *Physics Letters B*, 195, 216–222.

Friedman, J. (1991), "Multivariate adaptive regression splines," *Annals of Statistics*, 19, 1–141.

Fuentes, M. (2001), "A high frequency kriging approach for non-stationary environmental processes," *EnvironMetrics*, 12, 469–483.

Fuentes, M. and Smith, R. (2001), "A New Class of Nonstationary Spatial Models," Technical report, North Carolina State University, Department of Statistics.

Gaspari, G. and Cohn, S. (1999), "Construction of correlation functions in two and three dimensions," *Quart. J. Roy. Meteor. Soc.*, 125, 723–757.

Gelfand, A., Kim, H.-J., Sirmans, C., and Banerjee, S. (2003), "Spatial modelling with spatially varying coefficient processes," Technical report, Department of Statistics, University of Connecticut.

Gelfand, A., Sahu, S., and Carlin, B. (1996), "Efficient parametrizations for generalized linear mixed models," in *Bayesian Statistics 5*, eds. J. Bernardo, J. Berger, A. Dawid, and A. Smith, pp. 165–180.

Gelman, A. and Rubin, D. (1992), "Inference from iterative simulation using multiple sequences," *Statistical Science*, 7, 457–511.

Gibbs, M. (1997), *Bayesian Gaussian Processes for Classification and Regression*, unpublished Ph.D. dissertation, University of Cambridge.

Gibbs, M. and MacKay, D. (1997), "Efficient implementation of Gaussian processes," Technical report, University of Cambridge.

Gihman, I. and Skorohod, A. (1974), *The Theory of Stochastic Processes, I*, New York: Springer-Verlag.

Gneiting, T. (1999), "Correlation functions for atmospheric data analysis," *Quart. J. Roy. Meteor. Soc.*, 125, 2449–2464.

——— (2001), "Compactly supported correlation functions," *Journal of Multivariate Analysis*, 83, 493–508.

Golub, G. and van Loan, C. (1996), *Matrix Computations* (third ed.), Baltimore, Maryland: Johns Hopkins University Press.

Gradshteyn, I. and Ryzhik, I. (1980), *Tables of Integrals, Series and Products: Corrected and Enlarged Edition*, New York: Academic Press, Inc.

Green, P. (1995), "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711–732.

Green, P. and Silverman, B. (1994), *Nonparametric Regression and Generalized Linear Models*, Boca Raton: Chapman & Hall/CRC.

Handcock, M. and Stein, M. (1993), "A Bayesian analysis of kriging," *Technometrics*, 35, 403–410.

Hansen, M. and Kooperberg, C. (2002), "Spline adaptation in extended linear models," *Statistical Science*, 17, 2–51.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2001), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer-Verlag Inc.

Hastie, T. J. and Tibshirani, R. J. (1990), *Generalized Additive Models*, London: Chapman & Hall Ltd.

Higdon, D. (1998), "A process-convolution approach to modeling temperatures in the North Atlantic Ocean," *Journal of Environmental and Ecological Statistics*, 5, 173–190.

——— (2002), "Space and space-time modeling using process convolutions," in *Quantitative Methods for Current Environmental Issues*, ed. C. Anderson, London: Springer, pp. 37–54.

Higdon, D., Swall, J., and Kern, J. (1999), "Non-stationary spatial modeling," in *Bayesian Statistics 6*, eds. J. Bernardo, J. Berger, A. Dawid, and A. Smith, Oxford, U.K.: Oxford University Press, pp. 761–768.

Hilbert, D. and Cohn-Vassen, S. (1952), *Geometry and the Imagination*, New York: Chelsea Pub. Co., translated by P. Nemenyi. Translation of Anschauliche Geometrie.

Holland, D., Oliveira, V. D., Cox, L., and Smith, R. (2000), "Estimation of regional trends in sulfur dioxide over the eastern United States," *Environmetrics*, 11, 373–393.

Holmes, C. and Mallick, B. (2001), "Bayesian regression with multivariate linear splines," *Journal of the Royal Statistical Society, Series B*, 63, 3–17.

Holmes, C., Mallick, B., and Kim, H. (2002), "Analyzing non-stationary spatial data using piecewise Gaussian processes,", Poster presented at Bayesian Statistics 7.

Huerta, G., Sansó, B., and Guenni, L. (2001), "A spatio-temporal model for Mexico City ozone levels," Technical Report 2001-04, Centro de Estadística y Software Matemático,Universidad Simón Bolívar.

Hwang, J.-N., Lay, S.-R., Maechler, M., Martin, D., and Schimert, J. (1994), "Regression modeling in back-propagation and projection pursuit learning," *IEEE Transactions on Neural Networks*, 5, 342–353.

Kalnay, E. and Coauthors (1996), "The NCEP/NCAR 40-year reanalysis project," *Bull. Amer. Meteor. Soc.*, 77, 437–471.

Kennedy, M. C. and O'Hagan, A. (2001), "Bayesian calibration of computer models," *Journal of the Royal Statistical Society, Series B*, 63, 425–464.

Kitagawa, G. (1987), "Non-Gaussian state-space modeling of nonstationary time series," *J. Am. Stat. Assoc.*, 82, 1032–1063.

Lee, H. K. (1998), *Model Selection and Model Averaging for Neural Networks*, unpublished Ph.D. dissertation, Carnegie Mellon University, Department of Statistics.

Leithold, L. (1968), *The Calculus with Analytic Geometry*, New York: Harper & Row.

Loader, C. and Switzer, P. (1992), "Spatial covariance estimation for monitoring data," in *Statistics in the Environmental and Earth Sciences*, eds. A. Walden and P. Guttorp, London: Edward Arnold.

Lockwood, J. (2001), *Estimating Joint Distributions of Contaminants in U.S. Community Water System Sources*, unpublished Ph.D. dissertation, Carnegie Mellon University, Department of Statistics.

Lockwood, J., Schervish, M., Gurian, P., and Small, M. (2001), "Characterization of arsenic occurrence in source waters of U.S. community water systems," *J. Am. Stat. Assoc.*, 96, 1184–1193.

Loève, M. (1978), *Probability Theory II* (fourth ed.), New York: Springer-Verlag Inc.

MacKay, D. (1997), "Introduction to Gaussian Processes," Technical report, University of Cambridge.

Matérn, B. (1986), *Spatial Variation* (second ed.), Berlin: Springer-Verlag Inc.

McCullagh, P. and Nelder, J. (1989), *Generalized Linear Models* (second ed.), Boca Raton: Chapman & Hall.

McLeish, D. (1982), "A robust alternative to the normal distribution," *The Canadian Journal of Statistics*, 10, 89–102.

Meiring, W., Monestiez, P., Sampson, P., and Guttorp, P. (1997), "Developments in the modelling of nonstationary spatial covariance structure from space-time monitoring data," in *Geostatistics Wallongong 1996*, eds. E. Baafi and N. Schofield, Dordrecht: Kluwer, pp. 162–173.

Minka, T. P. and Picard, R. W. (1997), "Learning how to learn is learning with point sets,", http://citeseer.nj.nec.com/minka97learning.html.

Møller, J., Syversveen, A., and Waagepetersen, R. (1998), "Log Gaussian Cox processes," *Scandinavian Journal of Statistics*, 25, 451–482.

Neal, R. (1993), "Probabilistic Inference Using Markov Chain Monte Carlo Methods," Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto.

——— (1996), *Bayesian Learning for Neural Networks*, New York: Springer.

——— (1997), "Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification," Technical Report 9702, Department of Statistics, University of Toronto.

Nott, D. and Dunsmuir, W. (2002), "Estimation of nonstationary spatial covariance structure," *Biometrika*, 89, 819–829.

Nychka, D., Wikle, C., and Royle, J. (2001), "Multiresolution Models for Nonstationary Spatial Covariance Functions," Technical report, Geophysical Statistics Project, National Center for Atmospheric Research.

O'Connell, M. and Wolfinger, R. (1997), "Spatial regression models, response surfaces, and process optimization," *Journal of Computational and Graphical Statistics*, 6, 224–241.

Odell, P. and Feiveson, A. (1966), "A numerical procedure to generate a sample covariance matrix," *J. Am. Stat. Assoc.*, 61, 199–203.

Oehlert, G. (1993), "Regional trends in sulfate wet deposition," *J. Am. Stat. Assoc.*, 88, 390–399.

O'Hagan, A. (1978), "Curve fitting and optimal design for prediction," *Journal of the Royal Statistical Society, Series B*, 40, 1–42.

Olmsted, J. M. (1961), *Advanced Calculus*, New York: Appleton-Century-Crofts.

Paciorek, C., Risbey, J., Ventura, V., and Rosen, R. (2002), "Multiple indices of Northern Hemisphere Cyclone Activity, Winters 1949-1999," *J. Climate*, 15, 1573–1590.

Paige, R. and Butler, R. (2001), “Bayesian inference in neural networks,” *Biometrika*, 88, 623–641.

Papaspiliopoulos, O., Roberts, G., and Sköld, M. (2003), “Non-centered parameterisations for hierarchical models and data augmentation,” in *Bayesian Statistics 7*, eds. J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West, Oxford, U.K.: Oxford University Press, pp. 319–333.

Paulo, R. (2002), *Problems on the Bayesian-Frequentist Interface*, unpublished Ph.D. dissertation, Duke University, Institute of Statistics and Decision Sciences.

Rasmussen, C. (1996), *Evaluation of Gaussian Processes and Other Methods for Non-Linear Regression*, unpublished Ph.D. dissertation, University of Toronto, Graduate Department of Computer Science.

Rasmussen, C. and Ghahramani, Z. (2001), “Occam’s razor,” in *Advances in Neural Information Processing Systems 13*, eds. T. Leen, T. Dietterich, and V. Tresp, Cambridge, Massachusetts: MIT Press.

——— (2002), “Infinite mixtures of Gaussian process experts,” in *Advances in Neural Information Processing Systems 14*, eds. T. G. Dietterich, S. Becker, and Z. Ghahramani, Cambridge, Massachusetts: MIT Press.

Robert, C. and Casella, G. (1999), *Monte Carlo Statistical Methods*, New York: Springer-Verlag.

Roberts, G. O. and Rosenthal, J. S. (1998), “Optimal scaling of discrete approximations to Langevin diffusions,” *Journal of the Royal Statistical Society, Series B, Methodological*, 60, 255–268.

——— (2001), “Early sample measures of variability,” *Statistical Science*, 16(4), 351–367.

Ruppert, D. and Carroll, R. J. (2000), “Spatially-adaptive penalties for spline fitting,” *Australian and New Zealand Journal of Statistics*, 42(2), 205–223.

Sampson, P., Damian, D., and Guttorp, P. (2001), “Advances in modeling and inference for environmental processes with nonstationary spatial covariance,” in *GeoENV 2000: Geostatistics* *for Environmental Applications*, eds. P. Monestiez, D. Allard, and R. Froidevaux, Dordrecht: Kluwer, pp. 17–32.

Sampson, P. and Guttorp, P. (1992), "Nonparametric estimation of nonstationary spatial covariance structure," *J. Am. Stat. Assoc.*, 87, 108–119.

Sansó, B. and Guenni, L. (2002), "Combining Observed Rainfall and Deterministic Prediction Using a Bayesian Approch," Technical Report 2002-02, Centro de Estadística y Software Matemático,Universidad Simón Bolívar.

Schervish, M. (1995), *Theory of Statistics*, New York: Springer-Verlag.

Schmidt, A. and O'Hagan, A. (2000), "Bayesian Inference for Nonstationary Spatial Covariance Structure via Spatial Deformations," Technical Report 498/00, University of Sheffield, Department of Probability and Statistics.

Schoenberg, I. (1938), "Metric spaces and completely monotone functions," *Ann. of Math.*, 39, 811–841.

Schucany, W. R. (1995), "Adaptive bandwidth choice for kernel regression," *J. Am. Stat. Assoc.*, 90, 535–540.

Seeger, M. and Williams, C. (2003), "Fast forward selection to speed up sparse Gaussian process regression," in *Workshop on AI and Statistics 9*.

Skilling, J. (1989), "The eigenvalues of mega-dimensional matrices," in *Maximum entropy and Bayesian methods, Cambridge, England, 1988*, ed. J. Skilling, Dordrecht: Kluwer Academic Publishers, pp. 455–466.

——— (1993), "Bayesian numerical analysis," in *Physics and Probability*, eds. J. Grandy, W.T. and P. Milonni, Cambridge: Cambridge University Press, pp. 207–221.

Smith, M. and Kohn, R. (2002), "Parsimonious covariance matrix estimation for longitudinal data," *J. Am. Stat. Assoc.*, 97, 1141–1153.

Smith, R. (2001), “Environmental Statistics,” Technical report, Department of Statistics, University of North Carolina.

Smola, A. and Bartlett, P. (2001), “Sparse greedy Gaussian process approximation,” in *Advances in Neural Information Processing Systems 13*, eds. T. Leen, T. Dietterich, and V. Tresp, Cambridge, Massachusetts: MIT Press.

Stein, M. (1999), *Interpolation of Spatial Data : Some Theory for Kriging*, New York: Springer.

Swall, J. (1999), *Non-Stationary Spatial Modeling Using a Process Convolution Approach*, unpublished Ph.D. dissertation, Duke University, Institute of Statistics and Decision Sciences.

Vanmarcke, E. (1983), *Random Fields: Analysis and Synthesis*, Cambridge, Massachusetts: The MIT Press.

Ventura, V., Paciorek, C., and Risbey, J. (2003), “Controlling the proportion of falsely-rejected hypotheses when conducting multiple tests with geophysical data,” *in preparation*.

Vidakovic, B. (1999), *Statistical Modeling by Wavelets*, New York: John Wiley & Sons.

Vivarelli, F. and Williams, C. (1999), “Discovering hidden features with Gaussian processes regression,” in *Advances in Neural Information Processing Systems 11*, eds. M. Kearns, S. Solla, and D. Cohn.

Wahba, G. (1990), *Splines for observational data*, Philadelphia: Society for Industrial and Applied Mathematics.

Wand, M. (2003), “Smoothing and mixed models,” *Under Review*.

Wikle, C., Berliner, L., and Cressie, N. (1998), “Hierarchical Bayesian space-time models,” *Environmental and Ecological Statistics*, 5, 117–154.

Williams, C. (1997), “Prediction with Gaussian processes: from linear regression to linear prediction and beyond,” Technical Report NCRG/97/012, Neural Computing Research Group, Aston University.

Williams, C., Rasmussen, C., Schwaighofer, A., and Tresp, V. (2002), "Observations on the Nyström Method for Gaussian Process Prediction," Technical report, Gatsby Computational Neuroscience Unit, University College London.

Williams, C. and Seeger, M. (2001), "Using the Nyström Method to Speed Up Kernel Machines," in *Advances in Neural Information Processing Systems 13*, eds. T. Leen, T. Dietterich, and V. Tresp, Cambridge, Massachusetts: MIT Press.

Williams, C. K. I. and Barber, D. (1998), "Bayesian classification With Gaussian processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12), 1342–1351.

Williams, C. K. I. and Rasmussen, C. E. (1967), "Gaussian processes for regression," in *Advances in Neural Information Processing Systems, 8*, eds. D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, MIT Press.

Wood, S., Jiang, W., and Tanner, M. (2002), "Bayesian mixture of splines for spatially adaptive nonparametric regression," *Biometrika*, 89, 513–528.

Yaglom, A. (1987), *Correlation Theory of Stationary and Related Random Functions I: Basic Results*, New York: Springer-Verlag, Inc.

Yang, R. and Berger, J. O. (1994), "Estimation of a covariance matrix using the reference prior," *The Annals of Statistics*, 22, 1195–1211.

Zhou, S. and Shen, X. (2001), "Spatially adaptive regression splines and accurate knot selection schemes," *J. Am. Stat. Assoc.*, 96, 247–259.