

Efficient Posterior Inference and Prediction of Space-Time Processes Using Dynamic Process Convolutions

Catherine A. Calder
Department of Statistics
The Ohio State University
1958 Neil Avenue
Columbus, OH 43221
Ph. 614-688-0004; Fax 614-292-2096
E-mail: calder@stat.ohio-state.edu

In the Proceedings of the Sixth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences and the Fifteenth Annual Conference of TIES, the International Environmetrics Society

Portland, Maine

June 28 - July 1, 2004

Key Words: Bayesian, factor analysis, dimension reduction, air pollution

Abstract

Bayesian dynamic process convolution models provide an appealing approach for modeling both univariate and multivariate spatial temporal data. Their structure can be exploited to significantly reduce the dimensionality of a complex spatial temporal process. This results in efficient Markov chain Monte Carlo (MCMC) algorithms required for full Bayesian inference. In addition, the dynamic process convolution framework readily handles both missing data and misaligned multivariate space-time data without the need for imputation. We review the dynamic process convolution framework and discuss these and other computational advantages of the approach. We present an application involving the modeling of air pollutants to demonstrate how this approach can be used to effectively model a space-time process and provide predictions along with corresponding uncertainty statements.

1 Introduction

Large space-time datasets arise in disciplines such as atmospheric science, geology, hydrology, epidemiology, economics, and oceanography. In order to learn about the process that generated such data and predict the process either at locations where data was not collected or in the future, it is often necessary to create a statistical model that realistically describes the process. In the space-time setting, observations are often misaligned (not observed at the same locations) and stationarity and isotropy assumptions are clearly violated; a flexible modeling approach is needed to handle these nonstandard situations. We review a class of models for point-referenced space-time data, which is based on the discrete process convolution approach to modeling spatial data (Higdon, 1998), and highlight both the flexible nature of these models in addition to their computational advantages. The spatial moving average form of these dynamic process convolution (DPC) (Calder et al., 2002; Higdon, 2002; Calder, 2003) models guarantees valid (*i.e.*, positive-definite)

cross-covariance functions, and, as a result, spatial nonstationarity and anisotropy can be handled in a straightforward manner as in Higdon (1998) and Higdon et al. (1998). This structure also allows full Bayesian inference procedures to remain computationally tractable even for extremely large datasets, and misaligned data can be handled in way that does not require imputation. In addition, this class of models allows exploration of the latent structure governing the relationships between the multiple space-time processes.

DPC models fall into the general category of reduced-dimension space-time Kalman filters as defined by Cressie and Wikle (2002). Each of the space-time components in these models is formed by smoothing an underlying dynamic process defined on a grid covering the spatial region of interest. The dimension of these DPC models is ‘reduced’ since the inference procedure scales with the dimension of the underlying grid, not with the number of observations. A variety of other reduced-dimension space-time Kalman filters have been proposed including Mardia et al. (1998)’s ‘kriged Kalman filter,’ Wikle and Cressie (1999)’s space-time model based on empirical orthogonal functions (EOFs), and the locally weighted linear regressions considered by Stroud et al. (2001). An advantage of the DPC approach is the ease of dealing with nonstationarity and anisotropy and with misaligned and missing data.

The DPC models developed in Calder et al. (2002) and Higdon (2002) were extended to the multivariate setting in Calder (2003) using ideas from Bayesian dynamic factor analysis (Aguilar, 1998; Aguilar and West, 2000). The use of factor analysis and latent variable techniques to specify the cross-covariance structure of multivariate spatial processes is not new (see Christensen and Amemiya (2001) and Christensen and Amemiya (2002) for a discussion). In most cases these models are developed explicitly for purely spatial data but have straightforward extensions in the space-time arena. In the Calder (2003) approach, the dynamic factor framework is inserted into the dynamic process convolution model and inference on both the factor weights and underlying latent processes is performed jointly. The form of the factor loadings matrix is chosen to allow the latent factor processes to be interpretable as the factors driving the data processes. As a result, these dynamic factor models can be used in an exploratory sense to discover possible underlying mechanisms or structure in the data in addition to being used for predictive and interpolative purposes.

In the following sections, we review the univariate DPC model and the multivariate versions proposed in Calder (2003). We highlight the flexibility of the modeling approach in addition to the utility of the dimension reduction features. Finally, we conclude by presenting an analysis of a multivariate pollution dataset taken from the Environmental Protection Agency’s (EPA) Clean Air Status and Trends Network (CASTNet) database.

2 Modeling Approach

2.1 Discrete Process Convolution Models

We begin by briefly reviewing the discrete process convolution model upon which the DPC approach is based. Traditional models for spatial point-referenced data specify that the data arise from a Gaussian process over the spatial domain of interest \mathcal{D} . The covariance structure of the Gaussian process is usually specified directly by choosing a valid covariance function. An alternative, constructive approach to specifying a Gaussian process is to use a process convolution model (Thiébaux and Pedder, 1987; Barry and Ver Hoef, 1996; Higdon, 2002). A Gaussian process, ψ , is created by convolving a continuous white noise process x defined over \mathcal{D} with a smoothing kernel κ :

$$\psi(s|x) = \int_D \kappa(\omega - s)x(\omega)d\omega, \text{ for } s \in D. \tag{1}$$

The moving average construction of ψ guarantees that the covariance structure of the Gaussian process is valid (*i.e.*, positive-definite); this assurance leads to increased flexibility in modeling spatial process. By choosing an appropriate kernel κ features such as nonstationarity anisotropy can be incorporated into the model (Higdon, 1998; Higdon et al., 1998). In addition, non-Gaussian spatial fields can be constructed by convolving underlying processes other than white noise as in Ickstadt and Wolpert (1999) and Lee et al. (2002).

Higdon (1998) proposes approximating a Gaussian process over \mathcal{D} using a discrete version of the process convolution model where the underlying process is a discrete white noise process defined at locations $\{\omega_1, \omega_2, \dots, \omega_M\}$. This approximation to a Gaussian process ψ at any location $s \in \mathcal{D}$ can be written

$$\psi(s | x) = \sum_{i=1}^M \kappa(\omega_i - s)x(\omega_i), \quad (2)$$

where $x(\omega_i) \stackrel{\text{i.i.d.}}{\sim} N(0, \lambda_x)$ and M is the number of locations where the underlying process is defined. The vector of the values of the process at all sites of interest in \mathcal{D} can be written as

$$\boldsymbol{\psi} = \mathbf{K}\mathbf{x} \quad (3)$$

where $\mathbf{x} = [x(\omega_1), x(\omega_2), \dots, x(\omega_M)]'$ and \mathbf{K} is a matrix with rows $\mathbf{K}(s) = [\kappa(\omega_1 - s), \kappa(\omega_2 - s), \dots, \kappa(\omega_M - s)]$. Finding the value of ψ at any new location s' is simply a matter of calculating $\mathbf{K}(s') = [\kappa(\omega_1 - s'), \kappa(\omega_2 - s'), \dots, \kappa(\omega_M - s')]$ and multiplying it with \mathbf{x} .

The computational advantage of using this specification of an underlying spatial process when specifying a statistical model for spatial data is that the number of locations where the discrete white noise process is defined, M , is usually less than N , the number of locations where data are observed. (See Kern (2000) for a discussion of choosing M .) In addition, the appropriate choice of M scales with the area of the region being modeled (usually a subspace of \mathcal{R}^2), not with N . Thus, discrete process convolution models often have fewer unknown parameters than traditional kriging-based models (see Cressie (1993)).

2.2 Univariate Dynamic Process Convolution Models

The discrete process convolution approach presented in Equation 2 can be extended to model space-time processes in two ways. One extension is proposed in Higdon (1998) for modeling ocean temperatures in the North Atlantic. In this approach, a three-dimensional underlying process $x_{\omega,t}$ is convolved with a three dimensional smoothing kernel. The $x_{\omega,t}$ s are *a priori* independent (across both space and time) Gaussian random variables, and the smoothing kernel κ is a function of both distance in space and distance in time. A disadvantage of this approach is that unless the smoothing kernel is given a separable form, *i.e.*,

$$\kappa(\Delta s, \Delta t) = \kappa_s(\Delta s)\kappa_t(\Delta t),$$

posterior inference procedures become computationally difficult even for moderately sized datasets since inversion of an $MT \times MT$ matrix is required, where T is the number of timesteps.

An alternative approach is to specify an underlying temporally dependent (spatially independent) process on a lattice which is smoothed spatially at each time step using two-dimensional kernels. A space-time model based on this approach would have the form

$$Y(s, t|x) = \sum_{i=1}^M \kappa(\omega_i - s)x(\omega_i, t) + \epsilon_{s,t} \quad (4)$$

$$x(\omega_m, t) = G(x(\omega_m, t-1)) + \nu_{m,t}, \quad \forall m$$

where $\epsilon_{s,t} \stackrel{i.i.d.}{\sim} N(0, \lambda_\epsilon)$, $\nu_{m,t} \stackrel{i.i.d.}{\sim} N(0, \lambda_\nu)$, and $x(\omega_m, 0) \sim N(m_0, C_0)$ (Calder et al., 2002; Higdon, 2002). An advantage of this dynamic process convolution (DPC) framework is that the temporal evolution of the space-time process can be examined by incorporating static parameters in the function $G(\cdot)$. Alternatively, external information based on physical theory can be incorporated into the model through the specification of the function $G(\cdot)$ analogously to the approach taken in Xu et al. (2003). Inference is also much more computationally feasible due to the dynamic linear framework of the model (see West and Harrison (1997)). In particular, by embedding a Forward Filtering Backward Sampling algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994) within a Markov chain Monte Carlo (MCMC) algorithm, inversion of an $MT \times MT$ matrix is no longer required.

2.3 Multivariate Dynamic Process Convolution Models

We now discuss a multivariate extension of the DPC framework developed in Calder (2003). For simplicity and notational convenience, we assume that we have I ‘types’ of observations at each of N spatial locations observed over time and will discuss later how our approach can handle the more general situation where the different types of data are not all observed at the same locations. The $N \times I$ matrix \mathbf{Y}_t denotes values of the observed spatial temporal process at time t and consists of the I different types of data observed at the N locations at time t . The multivariate extension of Equation 4 relates the data to I different underlying processes defined at M spatial locations, and, by construction, the underlying processes are independent both across the I types and the M locations. For $t = 1, 2, \dots, T$, \mathbf{Y}_t is modeled as

$$\begin{aligned} \mathbf{Y}_t &= \underset{(N \times I)}{\mathbf{K}} \underset{(N \times M)(M \times I)}{\mathbf{X}_t} + \underset{(N \times I)}{\boldsymbol{\mu}} + \underset{(N \times I)}{\boldsymbol{\epsilon}_t} \\ \mathbf{X}_t &= \underset{(M \times I)}{G(\mathbf{X}_{t-1})} + \underset{(M \times I)}{\boldsymbol{\nu}_t} \end{aligned} \quad (5)$$

where $\boldsymbol{\epsilon}_t \stackrel{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \text{diag}(\boldsymbol{\lambda}_\epsilon), \boldsymbol{\Psi})$ and $\boldsymbol{\nu}_t \stackrel{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \text{diag}(\boldsymbol{\lambda}_\nu), \mathbf{I}_{I \times I})$. The notation $\text{diag}(\cdot)$ denotes a diagonal matrix with its argument on the diagonal. The I -dimensional vector $\boldsymbol{\lambda}_\epsilon$ represents the measurement error variances for the different types of data and the I -dimensional vector $\boldsymbol{\lambda}_\nu$ denotes the process error of the I latent \mathbf{X} processes. The notation $\mathcal{N}(\cdot, \cdot, \cdot)$ represents the matrix normal distribution as described in Dawid (1981). We assume the function $G(\cdot)$ operates on the elements of \mathbf{X}_{t-1} individually so that $\mathbf{X}_t(m, i)$ is directly related only to $\mathbf{X}_{t-1}(m, i)$. By construction, the multivariate space-time model defined by Equation 5 relates the data to I different underlying processes defined at M spatial locations, which are independent both across the I types and the M locations.

As in the univariate space-time process convolution models (Equation 4), the data \mathbf{Y}_t given the underlying process at time t , \mathbf{X}_t , are independent across spatial locations, i.e. the rows are independent. We do not, however, want to assume independence across the columns of \mathbf{Y}_t , that is, across the different types of measurements. The parameter $\boldsymbol{\Psi}$ in the distribution of the measurement error variance captures this between-type dependence. Instead of specifying $\boldsymbol{\Psi}$ directly, the covariance between the columns of \mathbf{Y}_t conditional on the values of the underlying process is induced a through latent factor framework (to be discussed).

As mentioned above, for notational simplicity we assume that each of the different types of measurements is observed at the same locations and that these locations remain constant over time. Misaligned or cross-sectional multivariate spatial temporal data can easily be modeled using this DPC framework by adjusting the \mathbf{K} matrix at each time-step. Another assumption we made that is easily relaxed is that the I types of measurements are smoothed using the same convolution kernel. This is by no means a necessary assumption as we can allow different convolution kernels (which imply different \mathbf{K} matrices) for each of the different types of data. In this way, we can jointly model the different components of a multivariate space-time process while allowing the individual components to have different spatial properties.

To further reduce the dimension of a space-time process using the multivariate DPC model, Calder (2003) borrowed ideas from dynamic factor analysis (Aguilar, 1998; Aguilar and West, 2000). Instead of relating the data to I different underlying processes, the data are taken to be linear combinations of J different underlying processes where $J < I$. For $t = 1, 2, \dots, T$, \mathbf{Y}_t is now modeled as

$$\mathbf{Y}_t = \underset{(N \times I)}{\mathbf{K}} \underset{(N \times M)(M \times J)}{\mathbf{X}_t} \underset{(J \times I)}{\mathbf{F}'} + \underset{(N \times I)}{\boldsymbol{\mu}} + \underset{(N \times I)}{\boldsymbol{\epsilon}_t} \quad (6)$$

$$\mathbf{X}_t = \underset{(M \times J)}{G(\mathbf{X}_{t-1})} + \underset{(M \times J)}{\boldsymbol{\nu}_t} .$$

In order to ensure identifiability of the model parameters, Calder (2003) proposed making restrictions on the matrix \mathbf{F} in a similar fashion to Aguilar (1998) and Aguilar and West (2000):

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \dots & & 0 \\ w_{2,1} & 1 & 0 & \dots & 0 \\ w_{3,1} & w_{3,2} & 1 & 0 & \dots & 0 \\ \vdots & & & & & \vdots \\ w_{k,1} & w_{k,2} & \dots & & & 1 \\ w_{k+1,1} & w_{k+1,2} & \dots & & & w_{k+1,K} \\ \vdots & & & & & \vdots \\ w_{I,1} & w_{I,2} & \dots & & & \dots w_{I,K} \end{bmatrix} . \quad (7)$$

This particular restriction on \mathbf{F} insures that the likelihood is invariant under invertible linear transformations of the factors and that \mathbf{F} is of full rank. Another restriction that is derived in Calder (2003) is that J , the number of underlying processes, must satisfy

$$\frac{1}{2}JN(JN + 1) - [(IJ - \frac{1}{2}J(J + 1)) + J + I] \geq 0,$$

which poses no realistic restrictions on J for most N and I .

An advantage of specifying \mathbf{F} constrained as in Equation 7 is that the underlying processes \mathbf{X} are interpretable. The first factor processes, the first row of \mathbf{X} , can be interpreted as the driving force underlying the first series since the first series places all of its weight on the first factor process. Similarly, the part of the second factor processes not explained by the first factor process, the second row of \mathbf{X} , places all of its weight on the second factor. In this way, the model structure allows us to learn about the relationships between the I different space-time data processes. This interpretability feature of the factor DPC model is explored in the next section.

3 CASTNet Example

To illustrate the multivariate factor DPC model, we analyze data from the EPA's Clean Air Status and Trends Network (CASTNet), which was established to monitor the impact of the Clean Air Act of 1990. Specifically, we model the concentration levels of five pollutants: sulfur dioxide (SO₂), nitric acid (HNO₃), particulate sulfate (SO₄²⁻), particulate nitrate (NO₃⁻), and particulate ammonium (NH₄⁺). Readings of these pollutants were obtained from 42 monitoring stations taken across the eastern United States on a weekly basis from 1990-2000. Figure 1 shows the locations of the monitoring network.

A factor DPC model with $J = 2$ and $M = 30$ was fit to the CASTNet data after removing a yearly cycle. The locations of the underlying processes are also shown in Figure 1. The temporal evolution of the underlying process was taken to be a random walk, *i.e.*, the function $G(\cdot)$ was taken to be the identity,

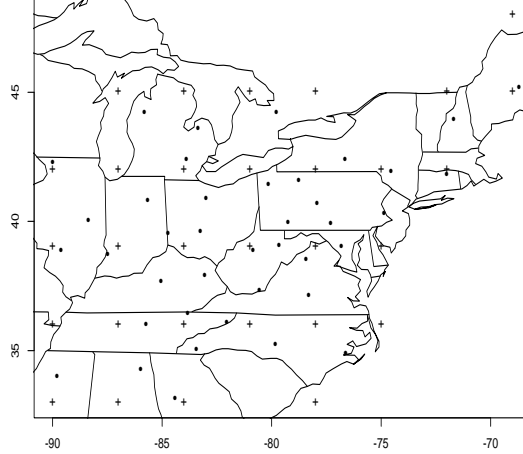


Figure 1: Locations of the monitoring stations and underlying grid process. The locations of the 42 monitoring stations are represented by circles. Crosses denote the 30 locations of the underlying grid process.

and the evolution error variances were fixed at 0.1. The dimension reduction and dynamic linear model framework are particularly useful in analyzing this dataset as there are over 160,000 data points.

As discussed in Section 2.3, the factor DPC framework is also useful in terms of gaining understanding about the relationships between the five pollutants. Figure 2 shows the average (taken over the M locations) of the posterior means of the underlying processes. The first latent process shows a steady decline from 1990-2000, while the second process remains fairly steady. The posterior means and 95 percent credible intervals for the parameters in the factor loadings matrix \mathbf{F} are

$$\hat{\mathbf{F}} = \begin{pmatrix} 1 & 0 \\ 0.224 (0.178, 0.270) & 1 \\ 0.678 (0.644, 0.715) & 0.686 (0.590, 0.792) \\ 0.268 (0.235, 0.307) & 0.883 (0.716, 0.920) \\ 0.148 (0.0871, 0.190) & 1.03 (0.904, 1.18) \end{pmatrix}.$$

This shows that the second (HNO_3), fourth (NO_3^-) and fifth (NH_4^+) pollutants are placing more weight on the second latent process and the third (SO_4^{2-}) pollutant is spitting its weight about equally between the two latent processes. Based on average trends in the emissions of SO_2 and NO_x during the time-frame of the study as reported in the EPA's 2000 Annual Report on CASTNet (Lavery et al., 2002), these results are reasonable; SO_2 emissions declined steadily over the period while NO_x remained fairly constant. The posterior means and 95 percent credible intervals of the measurement error variances are

$$\hat{\lambda}_\epsilon = \begin{pmatrix} 3.01 (2.80, 3.23) \\ 0.540 (0.455, 0.623) \\ 1.83 (1.65, 1.99) \\ 0.551 (0.476, 0.630) \\ 0.419 (0.368, 0.476) \end{pmatrix}.$$

4 Discussion

As illustrated by the analysis of the CASTNet data, the modeling approach can prove useful in revealing structure in large space-time datasets and is extremely flexible. In addition, the Bayesian specification of the model is appealing in that it is always straightforward to report uncertainty estimates of model parameters. This honest acknowledgment of uncertainty is especially important when using a space-time model for predictive purposes since estimates reported without uncertainty statements are, in most situations, completely meaningless. Thus, while performing inference on DPC models can remain computationally intensive even with the dimension reduction features, we argue the framework is extremely useful in analyzing space-time processes.

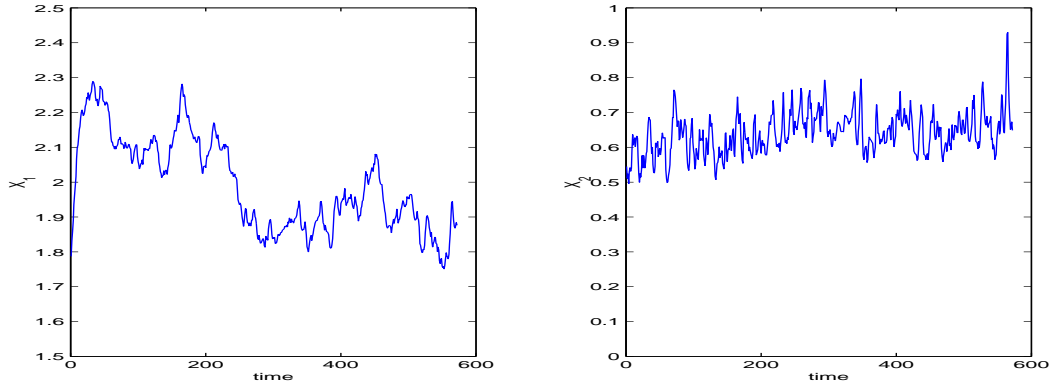


Figure 2: Average (taken over all locations of the underlying process) posterior means of the underlying process X .

Acknowledgments

The author would like to thank Dave Higdon of Los Alamos National Laboratory and Merlise Clyde and Michael Lavine of Duke University for many helpful discussions related to this work.

References

- Aguilar, O. (1998). “Latent structure in Bayesian multivariate time series models.” Ph.D. thesis, Duke University, Durham, NC 27708.
- Aguilar, O. and West, M. (2000). “Bayesian dynamic factor models and variance matrix discounting for portfolio allocation.” *Journal of Business & Economic Statistics*, 18, 338–357.
- Barry, R. and Ver Hoef, J. (1996). “Blackbox kriging: Spatial prediction without specifying variogram models.” *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 297–322.
- Calder, C. (2003). “Exploring latent structure in spatial temporal processes using process convolutions.” Ph.D. thesis, Duke University, Durham, NC 27708.
- Calder, C., Holloman, C., and Higdon, D. (2002). “Exploring space-time structure in ozone concentration using a dynamic process convolution model.” In *Case Studies in Bayesian Statistics*, vol. VI, 165–176. Springer.

- Carter, C. and Kohn, R. (1994). “On Gibbs sampling for state space models.” *Biometrika*, 81, 3, 541–553.
- Christensen, W. and Amemiya, Y. (2001). “Generalized shift-factor analysis method for multivariate geo-referenced data.” *Mathematical Geology*, 33, 7, 801–824.
- (2002). “Latent variable analysis of multivariate spatial data.” *Journal of the American Statistical Association*, 97, 457, 302–317.
- Cressie, N. (1993). *Statistics for spatial data*. New York: John Wiley.
- Cressie, N. and Wikle, C. (2002). *Encyclopedia of Environmetrics*, chap. Space-time Kalman filtering. John Wiley & Sons.
- Dawid, A. (1981). “Some matrix-variate distribution theory: notational considerations and a Bayesian application.” *Biometrika*, 68, 1, 265–274.
- Frühwirth-Schnatter, S. (1994). “Data augmentation and dynamic linear models.” *Journal of Time Series Analysis*, 15, 183–202.
- 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*, eds. C. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-Shaarawi, 37–56. Springer Verlag.
- Higdon, D., Swall, J., and Kern, J. (1998). “Non-stationary spatial modeling.” In *Bayesian Statistics 6*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 761–768. Oxford University Press.
- Ickstadt, K. and Wolpert, R. L. (1999). “Spatial regression for marked point processes.” In *Bayesian Statistics 6*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 323–341. Oxford University Press.
- Kern, J. (2000). “Bayesian process-convolution approaches to specifying spatial dependence structure.” Ph.D. thesis, Duke University, Durham, NC 27708.
- Lavery, T., Rogers, C., Howell, H., Burnett, M., and Wanta, C. (2002). “Clean Air Status and Trends Network (CASTNet): 2000 annual report.” Tech. rep., United States Environmental Protection Agency. Prepared by: Harding ESE.
- Lee, H., Holloman, C., Calder, C., and Higdon, D. (2002). “Flexible Gaussian processes via convolution.” Tech. Rep. 02-09, Institute of Statistics and Decision Sciences, Duke University.
- Mardia, K., Goodall, C., Redfern, E., and Alonso, F. (1998). “The kriged Kalman filter.” *Sociedad de Estadística e Investigación Operativa Test*, 7, 2, 217–285.
- Stroud, J., Müller, P., and Sanó, B. (2001). “Dynamic models for spatio-temporal data.” *Journal of the Royal Statistical Society, Series B*, 63, 673–689.
- Thiébaux, H. J. and Pedder, M. A. (1987). *Spatial objective analysis with applications in atmospheric science*. London: Academic Press.
- West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models*. 2nd ed. New York: Springer-Verlag.

- Wikle, C. and Cressie, N. (1999). “A dimension-reduced approach to space-time Kalman filtering.” *Biometrika*, 86, 4, 815–829.
- Xu, B., Wikle, C., and Fox, N. (2003). “A kernel-based spatio-temporal dynamical model for nowcasting radar precipitation.” In review.