A Kernel-Based Spatio-Temporal Dynamical Model for Nowcasting Weather Radar Reflectivities

Ke Xu, Christopher K. Wikle, and Neil I. Fox

A good short-period forecast of heavy rainfall is essential for many meteorological and hydrological applications. Traditional deterministic and stochastic nowcasting methodologies have been inadequate in their characterization of pixelwise rainfall reflectivity propagation, intensity, and uncertainty. The methodology presented herein uses an approach that efficiently parameterizes spatio-temporal dynamic models in terms of integro-difference equations within a hierarchical framework. The approach accounts for the uncertainty in the prediction and provides relevant distributional information concerning the nowcast. An application is presented that shows the effectiveness of the technique and its potential for nowcasting weather radar reflectivities.

KEY WORDS: Bayesian; Dilation; Dynamic; Forecast; Hierarchical; Integro-difference equations; Precipitation; Space time; Translation.

1. INTRODUCTION

For many hydrological applications, especially flash flood warnings and urban drainage management, a good short-term forecast (i.e., a “nowcast”) of heavy rainfall is required (Wilson, Crook, Mueller, Sun, and Dixon 1998). Typically, this involves forecasting the movement and/or development of radar-based estimates of precipitation over fairly short time scales (on the order of 1 hour or less). For example, Figure 1 shows a sequence of pixelated radar reflectivities near Sydney, Australia, with a time interval of 10 minutes. In the nowcasting problem, one would seek to forecast the pattern and intensity of such radar reflectivities at times into the immediate future, given the sequence of observed images.

One can divide the currently used approaches to nowcasting weather radar reflectivities into two general categories. The first is the use of simple extrapolation techniques. Such approaches analyze a series of radar reflectivity fields to identify areas of high reflectivity and track coherent structures from one image to the next. These methods then simply extrapolate the motion of individual high-reflectivity areas in a linear fashion to predict where the areas will be in the future. Examples of these methods are found in such operational systems as the Thunderstorm Identification, Tracking, Analysis and Nowcasting (TITAN) system (Dixon and Weiner 1993) and the Storm Cell Identification and Tracking (SCIT) system (part of the National Severe Storms Laboratory Warning Decision Support System) (Johnson et al. 1993). Extrapolations can also be made of the size and shape of the high-reflectivity area, as well as the implied intensity of the rainfall (e.g., Mecklenburg, Joss, and Schmid 2000).

The second method of nowcasting weather radar reflectivities uses physical (deterministic) approaches. These tend to combine knowledge of the motion of the rain area (often determined through extrapolation methods) with that of the atmospheric environment. Systems such as Gandolf (Pierce, Collier, Hardaker, and Haggett 2000) and Nimrod (Golding 2000) attempt to predict the motion and development of the implied precipitation detected by the radar using numerical model wind fields and atmospheric thermodynamic parameters.

Both extrapolation and deterministic methods for nowcasting have limitations. For example, rainfall areas do not typically move in a linear fashion, and so extrapolation-based forecasts are useful for only short periods; in the best case this may be 30 minutes, but often errors are apparent after 10 minutes. Another problem fundamental to the forecasting of convective rain is that development and dissipation of individual storms often occur over a short period, somewhere between 30 and 40 minutes. This means that forecasts of rainfall intensity by extrapolation methods cannot be reliable at longer times. The development of new cells and the splitting and merging of cells cannot be predicted by extrapolation methods. Such development can, in principle, be forecasted with deterministic approaches; however, the deterministic nowcasting of thunderstorm motion and development has proved a difficult problem in meteorology, because of the limited spatial and temporal continuity of convective systems. In many instances the model resolution is too large to capture the fine-scale patterns that influence convective storm properties, and the storms themselves can modify the environment, rendering the model parameters inaccurate. Furthermore, systems that attempt to model storm development (e.g., Pierce et al. 2000) are very sensitive to the data used to parameterize the convective model (Sleigh, Fox, Collier, and Pierce 2000; Pierce et al. 2004).

Both extrapolation and deterministic approaches do not realistically account for the forecast uncertainty. Because all forecasts have an inherent error, any attempt to use a forecast of precipitation in a hydrological application without reference to the range of outcomes and their impact on the hydrological situation will inevitably result in poor hydrological forecasts. To emphasize this, Smith and Austin (2000) stated that “forecast products, particularly those for hydrological applications, need to be statistical in nature, giving (for example) a range and a likelihood for falling within that range rather than just a best estimate of the value.” Thus it is critical that nowcasting procedures provide useful measures of forecast uncertainty.

Recent developments in the field of weather radar nowcasting include some systems that can be run to produce limited measures of uncertainty. The Spectral Prognosis (S–PROG) system (Seed 2003) is undergoing operational testing in a number of countries. S–PROG uses a spatial cascade technique to separate
the spatial scales of the implied precipitation areas observed by the radar. Using known relationships between spatial scale and temporal persistence, S–PROG then decays smaller objects in shorter times to indicate the lack of certainty in their persistence. It systematically disperses high-intensity, small-scale features in recognition of the inherent unpredictability of such features.

Less explicit means of representing a loss of predictability have also been presented by Germann and Zawadzki (2002, 2004), whose methods also recognize the predictability limits for different scales of precipitation in that they attempt to convey the loss of certainty by increasing the size of the resolved features with forecast time. Thus, although these recent approaches have recognized the need to address the forecast uncertainty issues, the approaches are somewhat ad hoc, although scientifically based. There is still a need for an approach to characterize uncertainty that accounts for uncertainty in data, process, and parameters and thus can provide a coherent and statistically meaningful uncertainty assessment.

The method presented herein has a statistical basis but retains the flexibility to incorporate physical knowledge to inform or constrain the nowcasts such that they would be consistent with our knowledge of atmospheric dynamics. This is accomplished through the Bayesian hierarchical paradigm, in which a flexible spatio-temporal dynamical model is efficiently parameterized within a kernel-based integro-difference equation (IDE) framework. This allows for the production of reasonable nowcasts of reflectivity location and intensity and the realistic characterization of uncertainty of such nowcasts. The advantage of this approach is that one can retain a distribution of nowcasts that are physically realistic and meaningful while avoiding unrealistic determinism. Ultimately, this provides a suite of forecasts that can provide meteorologists and hydrologists with realistic assessments of nowcast variability.

Section 2 presents some background material on efficient parameterization of spatio-temporal dynamical models, with an emphasis on IDE models. Section 3 follows with a discussion of the hierarchical Bayesian nowcast model. Section 4 presents the implementation of this model for the nowcasting of radar-observed convective precipitation around Sydney, Australia, on November 3, 2000. Section 5 presents a final discussion.

2. EFFICIENT PARAMETERIZATION OF SPATIO–TEMPORAL DYNAMICAL MODELS

Although the development and propagation of precipitation is likely a nonlinear process, it is reasonable to model the evolution of such processes linearly at the short time scales of interest in nowcasting. Let the unobservable state of the system be denoted by \( Y_t(s_i) \), where \( t \) is the time index and \( s_i, i = 1, \ldots, n \), are spatial locations. In the radar nowcasting problem, time is discrete \((t = 1, 2, \ldots)\) because we typically have access to images at fixed time increments and spatial locations are discrete pixels in a rectangular domain. Let \( \mathbf{Y}_t = [Y_t(s_1), \ldots, Y_t(s_n)]' \) be an arbitrary vectorization of the spatial locations at time \( t \). We then define a first-order spatio-temporal dynamic model as

\[
\mathbf{Y}_t = \mathbf{H}\mathbf{Y}_{t-1} + \mathbf{\eta}_t, \tag{1}
\]

where \( \mathbf{H} \) is an \( n \times n \) propagator matrix and \( \mathbf{\eta}_t \) is a noise process that is assumed to be independent in time and correlated in space, with variance–covariance matrix \( \Sigma_\eta \). Typically, the noise process is assumed to be Gaussian as well. In the spatio-temporal setting, the propagator matrix is critical; the \( i \)th row of the matrix \( \mathbf{H} \) contains the evolution coefficients that describe how the process at the \( n \) spatial locations at time \( t - 1 \) are related to locations \( s_i \) at time \( t \). This is just a vector autoregressive process of order 1. Estimation of the parameters and forecasting for such models is fairly simple (see, e.g., West and Harrison 1997). However, in the spatio-temporal context presented here, the number of spatial locations \( n \) is much larger than the number of times \( (T) \) for which we have data. This is a common situation with environmental data. In this case it is not possible to estimate directly the \( n^2 \) parameters in \( \mathbf{H} \) or the \((n + 1)n/2 \) unique parameters in \( \Sigma_\eta \). Rather, we must parameterize the matrices.

2.1 Parameterizations of the Noise Covariance Matrix

Several simple parameterizations for \( \Sigma_\eta \) can be considered. The simplest is to assume a white noise structure, \( \Sigma_\eta = \sigma^2_{\eta} \mathbf{I} \). Although such a structure can lead to a fairly complicated
marginal covariance for the $Y$-process in the presence of a complicated $H$ matrix, such forcing is not realistic for most environmental processes.

Alternatively, a more complicated parameterization for the spatial covariance matrix $\Sigma_\eta$ is to assume that it belongs to a family of flexible covariance functions, such as the Matérn class (e.g., Stein 1999). That is, we assume that the covariance matrix depends on parameters $\alpha_\eta$, $\Sigma_\eta(\alpha_\eta)$. Other alternatives would be to assume that the error process is a first-order Markov random field and specify the appropriate marginal covariance (see, e.g., Cressie 1993). One point to consider here is that we need not specify a nonstationary covariance model for this noise process if the propagator matrix parameterization $H$ is fairly flexible, because the marginal covariance of the $Y$-process is determined by both $H$ and $\Sigma_\eta$.

2.2 Parameterization of the Propagator Matrix

The dynamics of the $Y$-process are dependent primarily on the propagator matrix $H$. The simplest parameterization of $H$ is to assume that $H = I$, corresponding to a multivariate random walk (e.g., Stroud, Müller, and Sansó 2001). Although simple, such processes are fairly unrealistic for most environmental processes that exhibit diffusion or propagation through space (such as the precipitation processes fundamental to now-casting). Similarly, modeling the process as correlated univariate autoregressions, $H = hI$, or the nonseparable version, $H = \text{diag}(h)$, does not allow for the spatio-temporal interactions necessary to describe most environmental processes. At a minimum, one needs at least lagged (in time) dependence on “nearest” neighbors to accommodate realistic dynamics. That is, $Y_t(s_j)$ must be related to its neighbors $Y_{t-1}(s_j), j \in N_i$, where $N_i$ is the neighborhood of location $i$ (which must be defined specifically for a given problem) at the previous time. Such dependence structures are inherent in first-order space–time autoregressive (STAR) models (see, e.g., the review in Cressie 1993). In cases such as the nowcasting problem of interest here, one expects these lagged nearest-neighbor relationships to vary with space; that is, the parameters $\theta$ that describe the interaction should depend on location, $\theta_i, i = 1, \ldots, n$. This can present a problem in terms of estimation from a classical perspective. Wikle, Berliner, and Cressie (1998) showed that the Bayesian hierarchical framework is useful in these cases because the parameters $\theta$ can be modeled at the next stage of the hierarchy, thus reducing the effective number of parameters in the model.

Wikle, Milliff, Nychka, and Berliner (2001) and Wikle (2003) showed that one could use the underlying partial differential equations (PDEs) for some spatio-temporal processes to facilitate the choice of prior parameter models for $H$ and $\Sigma_\eta$. For example, Wikle (2003) discretized a reaction–diffusion equation and showed that the structure of $H$ follows the lagged nearest-neighbor structure, but the parameterization is greatly facilitated by simply allowing the associated diffusion parameters to be modeled as a spatial random field at the next level of the hierarchy. Such approaches work well when one has a fairly simple process that can be explained reasonably well by a deterministic PDE (although the implementation is by no means deterministic). However, as mentioned in Section 1, deterministic models for precipitation are very complicated, and one cannot simply discretize the system of PDEs and use them in this statistical context. Yet we know qualitatively that there are physically based spatial-temporal features, such as growth, diffusion, and propagation, in this process. Thus we seek alternative approaches to efficiently model the known qualitative dynamical features without direct specification of the underlying governing equations. Another potential limitation of PDE-based discretization is the inherent local nature of the lagged dependence. Although such dependence is needed to model propagation, it is not sufficient for modeling a wide variety of dynamics, including long-range dependence. As illustrated later, the IDE framework meets these specifications. In particular, IDE models suggest efficient, potentially nonlocal parameterizations for the propagator matrix $H$.

2.3 Integro-Difference Equation Models

Consider the stochastic linear IDE model

$$Y_{t+1}(s) = \gamma \int_D k_t(r; \theta) Y_t(r) \, dr + \eta_{t+1}(s),$$

where $Y_t(s)$ is the spatio-temporal process of interest; $s$ and $r$ are spatial locations in the spatial domain of interest, $D$; $k_t(r; \theta)$ is a redistribution kernel that describes how the process $Y$ at time $t$ is redistributed in space at time $t + 1$; $\theta$’s are parameters that control the redistribution (and may vary spatially); $\eta_{t+1}(s)$ is a mean-0 spatially colored noise process, assumed to be independent across time and typically Gaussian; and $\gamma$ is a growth parameter. Although the PDE and IDE frameworks share the notion of continuous space, the IDE framework differs in that it is formulated in discrete time. There is a substantial literature on deterministic IDEs in the mathematical ecology literature (e.g., Kot, Lewis, and van den Driessche 1996), and the notions are similar to the discrete-time spatial contact models of Mollison (1977).

Although the IDE model has been considered implicitly in the statistical modeling of spatio-temporal processes, as described by Wikle and Cressie (1999) and Brown, Karesen, Roberts, and Tonellato (2000), only recently has it been recognized that such a framework can efficiently model relatively complicated dynamical processes. The key to modeling dynamical processes is the redistribution kernel, $k_t(r; \theta_t)$, Kot et al. (1996) showed that such a framework can model diffusive wave fronts. Specifically, the shape and speed of diffusion depends on the kernel width and tail behavior. In fact, Kot et al. (1996) showed that the deterministic IDE framework yields the same analytical solution as Fisher’s (1937) solution to a reaction–diffusion PDE for modeling the speed of propagation for a diffusive wave. More recently, Wikle (2001, 2002a) demonstrated that the IDE framework can also model the movement (i.e., propagation) of spatial features through time. Such extradifusive propagation can be modeled via the relative displacement (i.e., translation) of the kernel. As shown schematically in Figure 2, if the kernel is shifted (translated) relative to spatial location $s$, then the propagation at that location at the next time is in the opposite direction of the shift and the speed of propagation is related to the magnitude of the shift (see, e.g., Wikle 2002a).

This framework is even more powerful when one recognizes that by allowing the kernel translation and dilation to be spatially dependent, one can model quite complicated nonseparable spatio-temporal behavior. This is analogous to Higdon’s
Figure 2. (a) Illustration Showing That if a Symmetric Kernel Is Translated to the Left of the Spatial Location (10), Then Propagation Is to the Right; (b) Propagation Is to the Left if the Kernel Is Translated to the Right of Location (10).

(1998) approach to modeling nonstationary spatial random fields by convolving white noise with kernels whose properties vary with space. In the IDE context, if the spatially varying kernels can be modeled efficiently, then this framework can be used to represent high-dimensional and complicated dynamical processes with relatively few parameters.

In the context of nowcasting radar reflectivities, the kernel-based IDE model is very flexible. The short time-scale forecast of precipitation is largely an advection problem; that is, one is most interested in where existing precipitation is likely to be in the very near future. Of course, there are diffusion (spread and weakening) of precipitation, as well as convection (increasing intensities and/or new development), to contend with. Although important, these later processes are less critical for basic nowcasting over small time scales. The IDE model is ideally suited to model advection through the kernel shift (translation) parameters. It allows for diffusion through the kernel spread (dilation) parameters and convection through the growth parameter ($\gamma$). Although such processes can be accommodated through PDE-based dynamics (e.g., Wikle 2003), the IDE model allows more flexibility. This is primarily because the PDE parameterization is often based on discretizations of the relevant partial derivatives, and this implies a fixed kernel width. This effectively suggests a knowledge of the scale of interaction, lagged in time; that is, typically there is only a lagged nearest-neighbor dependence in the PDE-based discretization.

The IDE framework allows for longer-range dependence, depending on the width of the kernel. Furthermore, because the kernel parameters are modeled as spatial random fields, this lagged-in-time dependence can vary locally in space.

2.4 Spectral Representation of an Integro-Difference Equation State Process

The kernel-based IDE equation (2) is often more practical if considered from an equivalent spectral representation. Such a representation serves two purposes. First, it allows for a natural way in which to handle the integral relationship; second, it facilitates computation via dimension reduction and the use of fast spectral transforms.

Consider the spectral expansion of the kernel and process in terms of orthonormal spectral basis functions, $\Phi_j(s)$,

$$ k_s(r; \theta_s) = \sum_{j=1}^{J} b_j(s; \theta_s) \phi_j(r) \quad (3) $$

and

$$ Y_t(s) = \sum_{j=1}^{J} a_j \phi_j(s), \quad (4) $$

where $b_j(s; \theta_s)$ and $a_j$ are the random spectral coefficients for the kernel and process. As shown by Wikle (2002a) for one-dimensional space, substituting (3) and (4) into (2), assuming Gaussian noise, and making use of the orthonormality of the basis functions gives

$$ Y_{t+1} = \gamma B\Phi a_t + \tilde{\eta}_{t+1}, \quad \tilde{\eta}_t \sim N(0, \Sigma_\eta), \quad (5) $$

where

$$ Y_{t+1} = [Y_{t+1}(s_1), \ldots, Y_{t+1}(s_n)]', $$

$$ a_t = [a_{t1}, \ldots, a_{tn}]', $$

$$ \Phi = [\phi_1, \ldots, \phi_n], $$

$$ \phi_j = [\phi_j(s_1), \ldots, \phi_j(s_n)]', $$

$$ B_\Phi = [B'(\theta) \ 0], $$

$$ B(\theta) = [b(s_1, \theta_{s_1}), \ldots, b(s_n, \theta_{s_n})]', $$

$$ b(s_j, \theta_{s_k}) = [b_1(s_j, \theta_{s_k}), \ldots, b_J(s_j, \theta_{s_k})]', $$

$\theta$ is an $n \times (J - 1)$ matrix of 0’s, and $\Sigma_\eta$ is the variance-covariance matrix of the $\tilde{\eta}$ noise process at locations $s_1, \ldots, s_n$. Making use of (4) and the complete and orthonormal properties of the basis functions, we obtain the state equation for the spatio-temporal process in spectral form,

$$ a_{t+1} = \gamma \Phi B_\Phi a_t + \eta_{t+1}, \quad \eta_t \sim N(0, \Sigma_\eta), \quad (6) $$

where $\Sigma_\eta = \Phi^T \Sigma_\Phi \Phi$ and, because of completeness, $J = n$.

Thus, in the context of the discussion of spatio-temporal dynamical models in Section 2.2, the propagator matrix for the spectral coefficients is a function of $\gamma$ and $\theta = [\theta_{s_1}', \ldots, \theta_{s_n}'][T]$, so that $H(\gamma, \theta) = \gamma \Phi^T B_\Phi$, which is parameterized fairly simply in terms of the growth parameter $\gamma$, the spatially varying kernel parameters $\theta$, and the (known) basis function matrix $\Phi$. In fact, $H(\gamma, \theta)$ in this context can be quite sparse and, depending on the choice of kernel and basis functions, can have a well-defined structure.

2.4.1 Choice of Basis Functions. In principle, any orthogonal set of basis functions can be chosen for $\phi_j(s)$. However, the choice of Fourier basis functions is advantageous with respect to dimension reduction and computation. If we let $\phi_j(s)$, $j = 1, \ldots, n$, be Fourier basis functions, then in the case of a Gaussian kernel, the kernel spectral coefficients $b_j(s; \theta_s)$ are known (given the parameters $\theta_s$), because the Fourier transform of the Gaussian kernel is its characteristic function, that is, $b_j(s; \theta_s) = \exp[i\omega_j(\mu(\theta_s) + s) - \frac{1}{2}\sigma^2(\theta_s)] |\omega_j|$, where $\omega_j$ is the spatial frequency pair and $i = \sqrt{-1}$ here. Thus, for a given
set of spatial frequencies, $B_\theta$ is completely defined if we know the parameters $\theta$. In the weather radar application, we choose the complete set of spatial frequencies used for the discrete Fourier transform of the pixelated rectangular radar image. That is, for an image with $M$ pixels in the $x$-direction and $N$ pixels in the $y$-direction, the frequencies correspond to $(\omega_p = 2\pi p/M, \omega_q = 2\pi q/N)$, where $p = 0, 1, \ldots, M/2$ and $q = 0, 1, \ldots, N/2$ ($N, M$ even) (see Royle and Wikle 2005).

If the spatio-temporal region of influence is not small so that the kernel dilation is nontrivial (i.e., the kernel is not a delta function), then most of the high-frequency Fourier coefficients are very close to 0, and thus it can be shown that $I \ll n$ (e.g., Wikle 2002a). In this case the number of 0’s in $B_\theta$ is large, and sparse matrix operations can significantly increase computational efficiency. Computational efficiency is also enhanced by the use of fast Fourier transform algorithms. Note that for some vector $g$ is an inverse discrete Fourier transform of $g$, and $\Phi(x)$ is the discrete Fourier transform of $x$. Thus one can use fast transform algorithms when implementing (6).

2.4.2 IDE Kernel Parameterization. For this application we consider two-dimensional kernels, because the radar reflectivity process of interest is represented in two-dimensional Euclidean space. The Gaussian kernel for location $s$ is given by

$$k_s(r, \theta_s) = \frac{1}{2\pi |\Sigma(\theta_s)|^{1/2}} \exp\{-\frac{1}{2}(r - s - \mu(\theta_s))^T \Sigma(\theta_s)^{-1}(r - s - \mu(\theta_s))\}. \quad (7)$$

In this case kernel translation is controlled by $\mu(\theta_s) = [\mu_x(s), \mu_y(s)]^T$, and dilation and orientation are controlled by the three possibly unique elements of $\Sigma(\theta_s)$. Alternatively, to facilitate the eventual hierarchical formulation, we consider an equivalent parameterization in terms of the foci of an ellipse (e.g., Higdon 1998; Higdon, Swall, and Kern 1999). As shown in Figure 3, the two foci coordinates, $(\psi^a_s, \psi^b_s)$ and $(\psi^b_s, \psi^b_s)$, and a scaling parameter, $\tau(s)$, describe the kernel. Thus in (7), $\theta_s = (\psi^a_s(s), \psi^b_s(s), \psi^b_s(s), \psi^b_s(s), \tau(s))^T$, and $\mu(\theta_s)$ and $\Sigma(\theta_s)$ are functions of these ellipse parameters, that is,

$$\mu(\theta_s) = \left(\begin{array}{c} \frac{1}{2} \psi^a_s(s) + \frac{1}{2} \psi^b_s(s) \\ \frac{1}{2} \psi^a_s(s) + \frac{1}{2} \psi^b_s(s) \end{array}\right)$$

and

$$\Sigma(\theta_s)^{1/2} = \tau(s)^T \left(\begin{array}{cc} |4A^2 + ||\psi_s||^4/(2\pi) + ||\psi_s||^4/2|^{1/2} & 0 \\ 0 & |4A^2 + ||\psi_s||^4/(2\pi) - ||\psi_s||^4/2|^{1/2} \end{array}\right) \times \left(\begin{array}{cc} \cos \alpha_s & \sin \alpha_s \\ -\sin \alpha_s & \cos \alpha_s \end{array}\right),$$

where $\alpha_s = \tan^{-1}((\psi^a_s(s) - \psi^b_s(s))/(\psi^b_s(s) - \psi^b_s(s)))$, $||\psi_s|| = \sqrt{((\psi^a_s(s) - \psi^b_s(s))^2 + ((\psi^a_s(s) - \psi^b_s(s))^2)^2}$, and $A$ is the ellipse area (assumed fixed).

3. BAYESIAN HIERARCHICAL MODEL FOR NOWCASTING

We now describe the hierarchical model that uses the IDE-based spatio-temporal process model.

3.1 Data Model

Given the $n \times 1$ data vector $Z_t$ (corresponding to $n$ measurement locations $s_1, \ldots, s_n$), the data model is

$$Z_t | a_t, \sigma^2 = \text{indep} \mathcal{N}(\Phi(a_t, \sigma^2 I), 1). \quad t = 1, \ldots, T, \quad (8)$$

where we assume that the true process $Y_t$ at $n$ spatial locations is given by $Y_t = \Phi(a_t)$. In the nowcast situation we have data for all pixels at each time. The data model could easily be modified to accommodate missing data (e.g., Wikle et al. 2001).

The variance component, $\sigma^2$, in (8) is the measurement error variance. Observations of radar-derived rainfall are prone to errors. Most of these are concerned with the algorithmic retrieval of physically meaningful atmospheric properties (e.g., rainfall rate or liquid water content) from the reflectivity data received. For this reason, we choose to work in reflectivity units rather than derived rainfall rates. The measurement of reflectivity also has associated errors, due primarily to instrumental limitations (i.e., error in the measured power return to the receiver) and a calibration error concerned with the conversion of the measured power to the equivalent standard reflectivity scale. However, over the relatively small domain of interest in this study, the homogeneous white noise assumption for the measurement error is reasonable.

3.2 Process Model

The process model describes the evolution of the spectral coefficients $a_t$. This is simply the IDE spectral model (6) described previously. In distributional notation,

$$a_t | \theta_{s_1}, \ldots, \theta_{s_n}, \psi_s, \gamma, \alpha_s \sim \mathcal{N}(\gamma \Phi B_{\psi_s} a_{t-1}, \sigma^2 \Phi D_s (\alpha_s)), \quad t = 1, \ldots, T, \quad (9)$$

with the initial process value given the prior $a_0 \sim \mathcal{N}(\bar{a}_0, \bar{\Sigma}_0)$.

---

**Figure 3.** Ellipse Representation of a Two-Dimensional Gaussian Kernel.
We write the noise covariance matrix from (6), $\Sigma_\eta = \sigma^2_\eta D_\eta(\alpha_\eta)$, where $D_\eta(\alpha_\eta)$ is a diagonal matrix. This form is obtained by first assuming that the noise forcing has variance $\sigma^2_\eta$ and a correlation that follows the Matérn class with parameters $\alpha_\eta$ (e.g., Stein 1999). It is well known that the Fourier transform of such a stationary covariance function yields an approximately diagonal covariance matrix. Thus, following Wikle (2002b) and Royle and Wikle (2005), we assume that the diagonal elements are given by the analytical Fourier transform of the Matérn class. Using this spectral representation greatly facilitates computation. Specifically, the diagonal matrix $D_\eta(\alpha_\eta)$ is specified to correspond to an exponential correlation model with unknown spatial dependence parameter $\alpha_\eta$. That is, we assume that $\text{cov}(\eta(s), \eta(r)) = \exp(-\alpha_\eta d)$, where $d$ is the Euclidean distance between locations $s$ and $r$. Furthermore, we assume that the distance between adjacent pixels in the rectangular grid is one unit.

3.3 Parameter Models

We now specify the distributions for the parameters defined at previous stages. We discuss the sensitivity of the nowcasts to hyperparameters and priors in the next section.

The radar used to collect the data used in this study is a well-characterized instrument. Specifically, as described in detail by Keenan et al. (1998), the radar is situated at Badgerys Creek, west of Sydney, and is operated by the Australian Bureau of Meteorology Research Centre (BMRC). It has measured error characteristics related to noise in the transmitted power, antenna gain, and received power, which combine to produce an uncertainty in the measured reflectivity of 1.7 decibels (Keenan et al. 1998). This value is typical for a weather radar; however, most radars are not so well characterized. Thus, assuming that this uncertainty represents two standard deviations, the measurement error variance $\sigma^2_\eta$ is assigned an inverse gamma (IG) prior, IG($7.18, .225$), corresponding to a prior mean of $.72$ and a relatively narrow variance of $.1$.

As discussed earlier, the kernel parameters can be modeled as spatial fields. For example, the $x$-values of the “$a$” foci field $\psi^a_x(s)$, as described in Section 2.4.2, are assigned a multivariate normal distribution with a specified covariance matrix. However, the dimensionality of this random field can be quite high and thus can be difficult to implement. Alternatively, as with the error process in (9), one can model such a process efficiently in the spectral domain (Wikle 2002b; Royle and Wikle 2005), that is, write

$$\psi^a_x = \Phi \beta^a_x,$$

where $\psi^a_x = (\psi^a_x(s_1), \ldots, \psi^a_x(s_n))^T$. The matrix $\Phi$ represents the $n \times n$ matrix of Fourier basis functions corresponding to $n$ spatial locations of the underlying dynamical process. Correspondingly, $\beta^a_x$ is the $n \times 1$ Fourier coefficient vector. As noted earlier, (10) can be computed via the inverse discrete Fourier transform if one seeks to avoid the more time-consuming matrix multiplication. Analogous parameterizations are made for the other foci fields, $\psi^b_y$, $\psi^c_z$, and $\psi^b_y$.

We then assign a distribution for the $\beta$ coefficients,

$$\beta^l_{b,c} \sim \text{N}(0, D^b_{\beta,c}),$$

where $i = (a, b, c) = (x, y)$, and the four $\beta$’s are assumed to be conditionally independent. As before for the noise-forcing covariance, $D^b_{\beta,c} = D_\eta(\alpha_\eta)$ is a diagonal matrix specified by the spectral transform of an exponential correlation model with a relatively high degree of spatial dependence; that is, we assume that $\text{cov}(\psi^i_x(s), \psi^i_x(r)) = \exp(-.1d)$, where $d$ is the Euclidean distance between locations $s$ and $r$ (where we assume that the distance between adjacent pixels in the rectangular grid is one unit). Given our prior belief that the reflectivity propagation should be relatively similar across the domain, we have used the prior with fairly large spatial dependence. We assume the same prior spatial dependence for each foci coordinate. Implicit in this representation is that the foci parameters have unit variance. This is sufficient for our application, because we do not expect the kernel to be centered too far away from the grid location for which it controls the propagation. But we note that $D^b_{\beta,c}$ is not identity in this case, because the Fourier transform of a stationary correlation function gives variances that vary with spatial frequency (e.g., Stein 1999).

Alternative nonstationary spectral spatial models for parameters, such as those given by Nychka, Wikle, and Royle (2002) and Fuentes (2002), could be considered if desired. In the case of the former, one would use wavelet basis functions for $\phi$. Sensitivity analyses in the application presented here suggests that there is little advantage in using nonstationary covariance matrices for these prior models.

The choices of hyperparameter distributions are made for computational simplicity and generally reflect our lack of certainty at the lower stages of the model hierarchy. We assume that the kernel scaling factor $\tau$ does not vary spatially, for the sake of simplicity. We put a relatively vague uniform prior distribution on $\tau, \tau \sim \text{U}(.01, 10)$. The growth parameter $\gamma$ is assumed to have a truncated normal prior, $\text{N}([1, 1.5], (1, 1))$. This allows for some degree of explosive growth ($\gamma > 1$), if necessary. The spatial dependence for the $\eta$ noise process, $\alpha_\eta$, is not known and is assumed to follow a uniform distribution, $\text{U}(.001, 3)$. Again we note that the distance between adjacent pixels is considered to be one unit in this application. The associated variance for this error process is $\sigma^2_\eta$ and is assumed to have a relatively vague IG prior, $\text{IG}(22, .0024)$, corresponding to a prior mean of 20 and a prior variance of 20.

4. NOWCASTING SYDNEY CONVECTIVE PRECIPITATION

4.1 Data

The data used in this study were collected during the World Weather Research Programme (WWRP) Sydney 2000 Forecast Demonstration Project (FDP). This was a project designed to demonstrate and test the current state-of-the-art nowcasting techniques (Keenan et al. 2003; Fox, Sleigh, and Pierce 2001). These data allow comparison with results from other nowcasting systems, such as those based on extrapolation and more physical models, as described in Section 1. The data represent a horizontal cross-section of the atmosphere at an altitude of 2000 meters above the surface. The data were mapped to a 45 x 45 grid of 2.5-km pixels centered on the radar location. For purposes of this study, a cropped region of dimension 28 × 40, corresponding to a 70-km × 100-km domain was considered.
This limited area clearly leads to edge problems when precipitation areas move into the domain from outside; however, for the limited time period considered here, this is not a serious problem.

The data values are in radar reflectivity units of decibels (dBZ). This is a measure of the power scattered back to the radar by precipitation particles in the atmosphere. Power is a function of the size distribution of raindrops. Meteorologists commonly convert from reflectivity units to a rainfall rate. This is seen as one major application of meteorological radar, because these values can be used in models and to forecast floods. However, the conversion is dependent on the exact nature of the distribution of raindrop sizes (which has been investigated extensively; see, e.g., Battan 1973) and is not without error. Therefore, although one potential application of this methodology would be a forecast of rainfall total, our motivation is to forecast movement and intensity in reflectivity units for general use. If desired, these forecasted reflectivities can be converted to rainfall amounts for specific posterior analyses. Brown, Diggle, Lord, and Young (2001) considered spatio-temporal models for calibrating radar rainfall to gauge measurements. Our analysis focuses strictly on forecasting the movement of images in reflectivity units.

4.2 Implementation

The model was implemented with data from six consecutive radar images (spaced at 10-minute intervals) starting at 08:25Z on November 3, 2000, and forecast distributions were obtained for four consecutive 10-minute intervals beyond this (i.e., 09:25Z–09:55Z). Note that it is standard in meteorology to report all times in “Z-time,” which is Greenwich mean time. We follow this convention here. These data are shown in the left column of Figure 4.

The model analysis was performed using standard Markov chain Monte Carlo (MCMC) algorithms. In particular, Gibbs sampling steps were used for the $\alpha_i$, $\beta$, $\sigma_i^2$, and $\tau^2$ parameters, and Metropolis–Hastings steps within the Gibbs sampler were used for the spatial dependence parameter $\alpha_0$, kernel scaling parameter $\tau$, and foci parameters $\beta^0$. After a 2,500 iteration burn-in, 5,000 iterations were considered for the posterior analysis. Simple diagnostic checks were performed, and no evidence was found that the chain had not converged. Such checks included evaluating the Markov chains for the various parameters under various starting conditions, both visually and with the diagnostic of Gelman and Rubin (1992). This was by no means proof of convergence, especially with a relatively low number of iterations and a relatively large number of parameters. But the effective number of parameters is not all that large, given the spatial and temporal dependence in the model and data. Furthermore, practical “real-time” implementation of this procedure requires relatively fast methodologies. Given the dimensionality of the spatial domains of potential interest, it will not be possible in the near future to accommodate extremely long MCMC runs.

One purpose of this study is to determine whether useful results can be obtained despite the practical limitation of having relatively few iterations.  

4.2.1 Sensitivity Analysis. We considered several sensitivity analyses relative to the priors and hyperparameters reported earlier. First, the kernel dimension reduction was such that we assumed that only 20% of the Fourier spectral coefficients were needed to adequately reproduce the Gaussian kernels. Sensitivity analyses showed that increasing the percentage beyond this level did not alter the results significantly. As mentioned previously, sensitivity analyses were also performed to consider nonstationary (i.e., wavelet-based) spatial processes for the foci fields. Again, the posterior predictions were not significantly altered by this choice. Similarly, tests in which the $\eta$ process was assumed to have smoother correlation functions than exponential (i.e., models from the general Matérn class) did not significantly alter the predictive results.

The model results were sensitive to the fixed prior on the spatial dependence for the foci fields. In particular, if the foci field was assumed to be uncorrelated, then the portions of the spatial domain in which there was no precipitation suggested the distribution of raindrop sizes (which has been investigated extensively; see, e.g., Battan 1973) and is not without error. Therefore, although one potential application of this methodology would be a forecast of rainfall total, our motivation is to forecast movement and intensity in reflectivity units for general use. If desired, these forecasted reflectivities can be converted to rainfall amounts for specific posterior analyses. Brown, Diggle, Lord, and Young (2001) considered spatio-temporal models for calibrating radar rainfall to gauge measurements. Our analysis focuses strictly on forecasting the movement of images in reflectivity units.

4.3 Results

Posterior parameter summaries for the univariate parameters are given in Table 1. In addition, Figure 5 shows the propagation directions implied by the spatially varying kernels from the posterior. Clearly, this plot implies relatively coherent propagation to the east with a strong northeast orientation in the upper half of the domain.

Figure 4 shows a sequence of 10 radar images, spaced 10 minutes apart. The left panel shows the radar data in which the first six are the sequence of images used to perform the Bayesian estimation and the final four are true images for which we compare the model predictive output. Initially, there were two relatively intense areas of precipitation moving to the north and east. Apparently, these two “cells” merged together by 09:55Z. A developing cell in the southern portion of the domain is present by 08:55Z, and this cell also moves to the northeast. The second column contains the model posterior mean, with the last four images being true posterior predictions (i.e., data for these periods were not used when running the MCMC). In addition to the posterior means, the Bayesian hierarchical implementation also gives samples from the posterior and uncertainty estimates of the reflectivity fields for each time period. The last column of Figure 4 shows the posterior standard deviation (pixelwise), and columns three and four show two realizations from the MCMC simulation. The intensity differences in the sample fields for the prediction (i.e., nowcast) time periods are quite substantial. For example, the intensity difference during the last period suggests a large variation in the reflectivity intensity associated with the developing cell in the southern portion of the domain. Such a range in the implied precipitation intensity could have a significant impact in hydrological
Figure 4. Estimation and Nowcast of Storms in Sydney on November 3, 2000. The left column shows consecutive 10-minute returns (data); the second column shows the associated model posterior means. (Note that the last four panels are true nowcasts.) The third and fourth columns show realizations from the posterior distribution, and the last column shows the pixelwise posterior standard deviations. All units are in dBZ.

applications. The posterior standard deviation associated with each of these predictions reflects the greater uncertainty as one predicts further into the future.

Figure 4 shows that, qualitatively, the forecast is reasonable. There is a fairly good representation of the storm cell motion, and the nowcast realistically retains areas of high reflectivity for the most part. There is obvious “blurring” of the features as the forecast lead time increases. Such blurring is a property of IDE-based models as described by Brown et al. (2000). Figure 6 shows a more detailed comparison of the forecast precipitation fields to the actual radar observations for the four time periods for which nowcasts were made, as well as the model posterior standard deviation for the predictions at those time points. On this closer inspection, we find that the nowcast loca-
Table 1. Posterior Summary for Univariate Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>2.5th percentile</th>
<th>50th percentile</th>
<th>97.5th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_\epsilon^2$</td>
<td>.95</td>
<td>.62</td>
<td>.97</td>
<td>1.31</td>
</tr>
<tr>
<td>$\sigma_\eta^2$</td>
<td>22.74</td>
<td>20.15</td>
<td>23.04</td>
<td>24.38</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1.13</td>
<td>1.09</td>
<td>1.13</td>
<td>1.16</td>
</tr>
<tr>
<td>$\tau$</td>
<td>.61</td>
<td>.55</td>
<td>.61</td>
<td>.66</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>2.23</td>
<td>1.74</td>
<td>2.29</td>
<td>2.56</td>
</tr>
</tbody>
</table>

tion is displaced slightly from the actual location of the storm, but the main difference is in the reflectivity intensity. It is important to note that if the associated forecast error suggested by the standard deviations gives a reasonable representation of the true forecast error, then this would provide good guidance to the meteorologist. However, before we can make such a claim, we must consider additional methods of verification, examples of which are discussed in Section 4.4.

4.4 Nowcast Verification

We note that it is essential to verify the results of our (or any) nowcast model. This is a challenging problem given that one is verifying reflectivity intensities in space and in time. Small differences in shape, translation, and dilation of reflectivity “cells” (i.e., large clusters associated with intense precipitation) can dramatically affect traditional measures of forecast error (e.g., root mean squared error). This difficulty is compounded by the distributional nature of the output (and only one realization of the “truth”). The study of such issues is an active area of research in the meteorological community, and a complete discussion is beyond the scope of this study. However, we do consider some relatively simple verification procedures, one statistical, one subjective (yet scientific), and one based on a standard meteorological verification approach.

4.4.1 Space–Time Validation Statistics. For a quantitative measure of the model’s precision and accuracy, we consider the following three validation statistics modified from Cressie (1993, p. 102), Carroll and Cressie (1996), and Wikle and
Cressie (1999):
\[
CR_1(t) = \frac{(1/n) \sum_{j=1}^{n} (Z_t(s_j) - \hat{Z}_t(s_j))}{(1/n) \sum_{j=1}^{n} \hat{\sigma}^2_Z(s_j; t)^{1/2}},
\]

(12)
\[
CR_2(t) = \left( \frac{(1/n) \sum_{j=1}^{n} (Z_t(s_j) - \hat{Z}_t(s_j))^2}{(1/n) \sum_{j=1}^{n} \hat{\sigma}^2_Z(s_j; t)} \right)^{1/2},
\]

(13)
\[
CR_3(t) = \left( \frac{1}{n} \sum_{j=1}^{n} (Z_t(s_j) - \hat{Z}_t(s_j))^2 \right)^{1/2},
\]

(14)

where \(\hat{Z}_t(s_j)\) is the prediction (e.g., posterior predictive mean) of the reflectivities \(Z\) at location \(s_j\) and time \(t\) and \(\hat{\sigma}_Z(s_j; t)\) is the corresponding posterior predictive standard deviation at time \(t\) and location \(s_j\). As discussed by Carroll and Cressie (1996), the statistic \(CR_1(t)\) provides an estimate of the unbiasedness of the predictor at each time (averaged over location) and should be close to 0; \(CR_2(t)\) gives a measure of the accuracy of the predictive standard deviations and should be close to 1; and \(CR_3(t)\) is a measure of the “goodness of prediction.” We would like \(CR_3(t)\) to be small so that the predicted values are close to the true values.

Table 2 gives the CR statistics for the posterior mean nowcast at forecast lead times in 10-minute intervals out to 60 minutes, along with the average for those times. There is a very slight negative bias in the nowcasts across all lead times, as shown in the \(CR_1\) statistics. In addition, the predictive standard deviations are a bit too large, as indicated by the \(CR_2\) statistics being less than 1. As expected, \(CR_3\) increases with lead time, indicating that the nowcast accuracy decreases with lead time.

4.4.2 Kernel Shift Fields. One might assume that the spatial distribution of the kernel shift field shown in Figure 5 is a manifestation of the vertically averaged wind field in the domain, because advection of precipitation is greatly influenced by the spatial distribution of the winds. Such an association is reasonable, but there are many meteorological reasons why we would not expect such fields to match. Nevertheless, it is instructive to consider such a comparison as an independent scientific verification. Figure 7 shows the average winds for 09:05Z on November 3, 2000 as obtained from a high-resolution numerical model that has assimilated radar information. Note the similarity to the kernel shift plot in Figure 5 in terms of the general southwest-to-northeast flow. In addition, the west-to-east flow in the center part of the domain is in fairly good agreement. However, there are many differences as well. Most notably, the implied flow from southwest to northeast in the upper left corner of Figure 5 is not present in the derived wind field. In addition, the derived wind field is much smoother. This smoothness is expected from the analysis procedure that derives the winds. In addition, the procedure that calculates the winds does so on a much larger domain than that presented here, and the flow field must be dynamically consistent throughout the entire domain. Thus one cannot make too much out of small regional differences. More important, as mentioned earlier, the storm motion, as suggested by the kernel shift field, is not simply a function of the background wind, but also includes influences related to the three-dimensional circulation, particularly the vertical component of the wind and thermodynamic variables. That being said, the fact that there is general agreement between the kernel shift field and the derived winds is encouraging.

4.4.3 Comparison With Other Nowcast Methods. It is important to compare our Bayesian IDE nowcasting approach with existing methods. Although a thorough comparison is beyond the scope of this analysis, a comparison using this Bayesian IDE methodology was considered by Fox and Wikle (2005). In their analysis, Fox and Wikle also used data from the Sydney 2000 Forecast Demonstration Project as considered here, although they considered a much larger proportion of the available data and considered many more forecasts with hyperparameters for one forecast being given by the posterior from

<table>
<thead>
<tr>
<th>Time</th>
<th>(CR_1)</th>
<th>(CR_2)</th>
<th>(CR_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:25Z</td>
<td>-.008</td>
<td>.640</td>
<td>5.660</td>
</tr>
<tr>
<td>09:35Z</td>
<td>-.052</td>
<td>.665</td>
<td>6.863</td>
</tr>
<tr>
<td>09:45Z</td>
<td>-.083</td>
<td>.623</td>
<td>7.619</td>
</tr>
<tr>
<td>10:05Z</td>
<td>-.113</td>
<td>.621</td>
<td>8.199</td>
</tr>
<tr>
<td>10:15Z</td>
<td>-.099</td>
<td>.628</td>
<td>8.125</td>
</tr>
<tr>
<td>Average (all times)</td>
<td>-.066</td>
<td>.638</td>
<td>7.510</td>
</tr>
</tbody>
</table>
the Bayesian analysis of the previous forecast period. They compared their results with other nowcasting methods that were recently compared on the same dataset by Ebert et al. (2004). The results for the IDE methodology were encouraging in that it was shown to be competitive with or better than the state-of-the-art methods and demonstrated skill above persistence and extrapolation out to at least 30 minutes. Note that none of the methods showed much skill at lead times of 40 minutes, which illustrates why this is an important topic of research.

5. DISCUSSION

Spatio-temporal processes in the environmental and ecological sciences can be quite complicated. Often there is available scientific theory that enhances our understanding of the process. Although traditional methods are not well-suited to modeling such processes, the hierarchical approach can accommodate complicated processes through a series of conditional models. One can use PDE and IDE priors suggested by the science in this framework to facilitate the modeling. In particular, in situations where the governing PDE dynamics are not well known or are too complicated, the kernel-based stochastic IDE model can model complicated dynamics with relatively few parameters. Furthermore, spectral implementations are useful in this setting, because they facilitate computation and effective dimension reduction.

The hierarchical IDE formulation for nowcasting appears promising based on the results presented here. The chief advantages to this approach are the incorporation of efficient parameterizations of realistic dynamical movements on a pixelwise basis and the realistic characterization of uncertainty. In addition, an extensive comparison of this methodology to other nowcast systems by Fox and Wikle (2005) suggests that the nowcasts from the hierarchical IDE model are generally as good or superior to those currently produced by other systems.

The hierarchical IDE model does not account for additional physical variables that one would expect might improve the nowcast. However, one primary advantage of the hierarchical approach to nowcasting is that one could incorporate such physical/meteorological variables into the models for kernel parameters. For instance, a relatively simple model would allow the kernel shift to be related to the background wind direction (e.g., Fig. 7). More complicated formulations could model convective initiation based on other meteorological variables.

One can imagine the need for time-varying dynamics in situations with multiple weather regimes. From a statistical perspective, one can consider time-varying kernels for such a case,

\[ a_{t+1} = \gamma \Phi a_t + \eta_{t+1}, \]

where \( \theta \) varies over time and space, \( \theta_{k,j} = G \theta_{k,j-1} + \nu_j \). One must be careful with such a formulation so as not to just replace one complicated spatio-temporal problem with several others of equal complexity! However, it is likely that the dynamical behavior of the parameters in such a model is much less complicated than the precipitation process!

[Received July 2003. Revised May 2005.]


