INVITED PAPER

A general science-based framework for dynamical spatio-temporal models

Christopher K. Wikle · Mevin B. Hooten

Abstract Spatio-temporal statistical models are increasingly being used across a wide variety of scientific disciplines to describe and predict spatially-explicit processes that evolve over time. Correspondingly, in recent years there has been a significant amount of research on new statistical methodology for such models. Although descriptive models that approach the problem from the second-order (covariance) perspective are important, and innovative work is being done in this regard, many real-world processes are dynamic, and it can be more efficient in some cases to characterize the associated spatio-temporal dependence by the use of dynamical models. The chief challenge with the specification of such dynamical models has been related to the curse of dimensionality. Even in fairly simple linear, first-order Markovian, Gaussian error settings, statistical models are often over parameterized. Hierarchical models have proven invaluable in their ability to deal to some extent with this issue by allowing dependency among groups of parameters. In addition, this framework has allowed for the specification of science based parameterizations (and associated prior distributions) in which classes of deterministic dynamical models (e.g., partial differential equations (PDEs), integro-difference equations (IDEs), matrix models, and agent-based models) are used to guide specific parameterizations. Most of the focus for the application of such models in statistics has been in the linear case. The problems mentioned above with linear dynamic models are com-

C.K. Wikle (✉)
Department of Statistics, University of Missouri, Columbia, MO 65211, USA
e-mail: wiklec@missouri.edu

M.B. Hooten
USGS Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80523, USA
e-mail: mevin.hooten@colostate.edu

pounded in the case of nonlinear models. In this sense, the need for coherent and sensible model parameterizations is not only helpful, it is essential. Here, we present an overview of a framework for incorporating scientific information to motivate dynamical spatio-temporal models. First, we illustrate the methodology with the linear case. We then develop a general nonlinear spatio-temporal framework that we call general quadratic nonlinearity and demonstrate that it accommodates many different classes of scientific-based parameterizations as special cases. The model is presented in a hierarchical Bayesian framework and is illustrated with examples from ecology and oceanography.

1 Introduction

Spatio-temporal statistics has blossomed as a thriving sub-discipline in the last decade. The motivating force behind this increase in attention has been the pace of scientific inquiry. That is, in fields as seemingly disparate as human biology, ecology, meteorology, and economics, it is the case that the controlling processes often occur in space and time, and it is increasingly unrealistic to assume that realizations or aggregates across time (spatial processes) or across space (time series) are sufficient for describing such process. Perhaps more importantly, there has been a somewhat collective realization that there exists uncertainty not only in observations, but in our understanding of the behavior of the processes as well. Indeed, the over-reaching goal of spatio-temporal modeling in statistics is related to the characterization of the process of interest in the presence of uncertain and (often) incomplete observations and system knowledge. This includes prediction in space (interpolation), prediction in time (forecasting), assimilation of observations and deterministic model output, and to a lesser extent, inference on the controlling parameters of the process.

Statisticians have increasingly recognized the view that most spatial, temporal, or spatio-temporal processes are in fact "hidden" or "latent" (i.e., there is observational uncertainty), but have, to some extent, adopted the view that descriptive characterizations of processes based on the data are in some sense more "objective" and thus, preferable. Granted, this view is more prevalent in spatial and spatio-temporal statistics than in time series. One could argue that in this regard we have somewhat forgotten the lessons of some of the pioneers in our discipline, and that the development of statistical models should make use of the knowledge obtained throughout the development of the subject matter science at hand. This paper makes use of this view in the context of spatio-temporal dynamical models with the purpose being to illustrate how one can efficiently parameterize classes of linear and nonlinear spatio-temporal processes given scientific insight, particularly in the environmental sciences.

The idea of using fundamental scientific relationships to facilitate statistical modeling is not new! Indeed, Yule (1927) used the differential equation governing pendulum motion as motivation for an autoregressive time-series model for the Wolfer sunspot data (see the discussion in Bisgaard and Kulahci 2005). Perhaps a more rel-

evant example is Hotelling (1927), in which he used *approximations* of differential equations to model U.S. population growth. In this paper, Hotelling makes the following comment:

> "Indeed the use of differential equations supplies the statistician with a powerful tool, replacing the purely empirical fitting of arbitrary curves by a reasonable resultant of general considerations with particular data. But this growing statistical use of differential equations must inevitably face the fact that our a priori knowledge can never supply us with a definite relation between a variable and its rate of change, but only with a correlation." (Hotelling 1927, p. 283)

This quote states quite clearly the fundamental notion that one should make use of the information in the developed science (i.e., differential equations), but should also recognize that there are uncertainties associated with that knowledge. This is where statistics plays its biggest role.

Rather than use differential equation-based knowledge directly in a model of the underlying process, Heine (1955) used stochastic (partial) differential equations to develop general classes of covariance functions for spatial processes. The obvious advantage of such an approach is that one is able to characterize realistic second-order spatial properties for a variety of processes. That is, the process-based differential equation view is somewhat tied to the specific formulations appropriate for that process, whereas general model classes developed from general differential equation forms are more broadly applicable. This idea was extended by Whittle (1954, 1962). More recently, Jones and Zhang (1997) developed classes of non-separable spatio-temporal covariance models from this perspective. In contrast, statistical time-series analysis has developed in more of a process-based direction (e.g., the unification and development of ARIMA models by Box and Jenkins 1970), although the fundamental processes driving the science are sometimes ignored in such settings (e.g., the prevalence of random walks in environmental processes).

Although there is no doubt that descriptive, covariance-motivated approaches to spatial statistics have been critical to its utility (e.g., kriging), it is less clear that such approaches are as helpful for spatio-temporal statistics. The two principal limitations with such approaches are (1) lack of realism, and (2) the curse of dimensionality. In the first context, it is increasingly obvious that real-world spatio-temporal processes are non-separable (and often non-stationary), with very complicated spatio-temporal interactions (e.g., nonlinearity). However, we note that many new spatio-temporal model classes have been developed in recent years, and this remains an active and vital area of research (e.g., Cressie and Huang 1999; Gneiting 2002; Ma 2003; Stein 2005; Fuentes et al. 2008; Gregori et al. 2008; Zastavnyi and Porcu 2009). In the case of the second issue, for problems in the environmental sciences, the size of covariance matrices necessary to describe the joint spatio-temporal structure of the underlying process is often prohibitive. It could certainly be argued that as computational power increases, this becomes less of a concern. However, it is also the case that several different physical processes could imply the same marginal covariance structure, implying that the use of a descriptive approach is most useful when knowledge of the process is limited.

An alternative to descriptive, covariance-based formulations of spatio-temporal processes is a conditional, or dynamic approach. In this case, the current values of

a process at a location are assumed to evolve, at least partially, from the past values of the process at various locations. Such models are closer to the etiology of the phenomena of interest and thus are most useful when there is some a priori knowledge available concerning process behavior. That is, in many instances it is easier to specify realistic conditional models rather than the full joint structure, particularly when one is trying to account for specific types of process behavior. In this sense, the use of dynamic models for spatio-temporal processes is somewhat of a practical decision. A large portion of dynamic spatio-temporal models have followed the state-space approach developed in the engineering literature (e.g., Jazwinski 1970) and further refined in statistics (e.g., see the overviews in West and Harrison 1997; Shumway and Stoffer 2006). Such approaches are very powerful, and although some of the issues related to the issue of lack of realism and the curse of dimensionality are addressed in this framework, there has often been a tendency in the environmental sciences to forget the science-based motivation suggested by Hotelling (1927).

Of course, there is a connection between the underlying science-based dynamical behavior of the process and the associated spatio-temporal covariance and the marginal spatial and temporal covariances. For example, consider the release of a pollutant into the atmosphere. If there were no predominant wind direction (i.e., no advection by the background flow), then pure diffusion would dominate the distribution of the pollutant with time. In that case, the marginal spatial covariance function would most likely be isotropic. However, if there was a predominant wind direction, then the concentration of the pollutant over time would be higher in the down wind direction. In that case, the marginal spatial covariance function would likely be anisotropic. Similarly, a diffusion process with constant (i.e., non-spatial) diffusion coefficients will imply a separable spatio-temporal covariance structure, but the presence of spatially varying diffusion coefficients will imply non-separable dependence. More generally, Jones and Zhang (1997) derive the spatio-temporal covariance function for a class of stochastic partial differential equations (e.g., an injection-diffusion equation). One can show that a sensibly discretized form of these differential equations leads to conditional dynamical specification that gives the same marginal spatio-temporal covariance function as the time-discretization constant gets smaller. Such equivalencies motivate the consideration of dynamical models for spatio-temporal processes.

1.1 Dynamical spatio-temporal models (DSTMs)

Before we get into the details, we first present the basic framework for DSTMs, along with the notation used in this paper. Consider the $m_t \times 1$ observation vectors $\mathbf{Z}_t \equiv (Z(\mathbf{r}_1; t), \ldots, Z(\mathbf{r}_{m_t}; t))'$ where $Z(\mathbf{r}_i; t)$ is an observation of a spatio-temporal process at location $\mathbf{r}_i \in D \subset \mathbb{R}^2$ and time $t \in \mathcal{T} \subset \mathbb{R}$. In addition, say we are interested in a true, but hidden, spatio-temporal process $\{Y(\mathbf{s}; t)\}$ for $\mathbf{s} \in D \subset \mathbb{R}^2$. In principle, this process can be considered in higher dimensional space, but we will focus on $\mathbb{R}^2$ in this paper. In particular, assume we are interested in predicting this process at n spatial locations, and thus consider the $n \times 1$ process vector $\mathbf{Y}_t \equiv (Y(\mathbf{s}_1; t), \ldots, Y(\mathbf{s}_n; t))'$. We note that the observation locations $\{\mathbf{r}_i : i = 1, \ldots, m_t\}$ may, but need not, coincide with some or all of the prediction locations $\{\mathbf{s}_j : j = 1, \ldots, n\}$. Thus, the focus here is on statistical models that are discrete in

time and space, but may be associated with processes that are continuous in both. We recognize that discretization in both the spatial and temporal dimensions can be due to aggregation or discrete sampling/prediction of a continuous interval or domain. Here, as in many environmental problems, we consider that there is likely to be discrete (and typically, regular) sampling in time $t = 1, 2, 3, \ldots$, and that we are interested in discrete prediction in space. We note that missing observations in space and time are allowed (see Sect. 2.1).

Perhaps the most critical aspect of dynamic modeling is how one exploits conditional independence. That is, we can partition the joint distribution of the observed data and the true process of interest into the following conditional distributions:

$$[\mathbf{Z}_1, \ldots, \mathbf{Z}_T, \mathbf{Y}_1, \ldots, \mathbf{Y}_T] = [\mathbf{Z}_1, \ldots, \mathbf{Z}_T | \mathbf{Y}_1, \ldots, \mathbf{Y}_T][\mathbf{Y}_1, \ldots, \mathbf{Y}_T] \qquad (1)$$

$$= \prod_{t=1}^{T} [\mathbf{Z}_t | \mathbf{Y}_t][\mathbf{Y}_1, \ldots, \mathbf{Y}_T] \qquad (2)$$

$$= \prod_{t=1}^{T} [\mathbf{Z}_t | \mathbf{Y}_t][\mathbf{Y}_t | \mathbf{Y}_{t-1}][\mathbf{Y}_0], \qquad (3)$$

where the square bracket notation '[·]' refers to a probability distribution. Clearly, (1) follows from simple rules of probability, but (2) makes the assumption that the observations are independent, conditioned on knowing the true process, and (3) makes the additional (Markov) assumption that, given the most recent past values of the process, the current process in conditionally independent of the past, $[\mathbf{Y}_t | \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots] = [\mathbf{Y}_t | \mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_{t-p}]$. Specifically, as shown here for simplicity of presentation, we will typically assume that the first-order Markov assumption is reasonable in this paper, i.e., $[\mathbf{Y}_t | \mathbf{Y}_{t-1}]$, which can accommodate a wide variety of dynamical processes in the vector setting. We note, however, that there are certainly spatio-temporal processes for which second-order (or higher) Markov dependence is appropriate. Thus, from our perspective, the key to spatio-temporal state-space modeling is the specification of the observation distribution $[\mathbf{Z}_t | \mathbf{Y}_t]$ and the process distribution $[\mathbf{Y}_t | \mathbf{Y}_{t-1}]$. Taken together, the associated model forms for these distributions are often referred to as a "state-space model." In practice, there are also parameters $\{\boldsymbol{\theta}_d, \boldsymbol{\theta}_p\}$ associated with each of these component models: i.e., $[\mathbf{Z}_t | \mathbf{Y}_t, \boldsymbol{\theta}_d]$, $[\mathbf{Y}_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}_p]$. These parameters might be estimated in a maximum likelihood or EM-algorithm context or, given prior distributions $[\boldsymbol{\theta}_d, \boldsymbol{\theta}_p]$, updated through Bayes' rule: e.g.,

$$[\mathbf{Y}_1, \ldots, \mathbf{Y}_T, \boldsymbol{\theta}_d, \boldsymbol{\theta}_p | \mathbf{Z}_1, \ldots, \mathbf{Z}_T] \propto \prod_{t=1}^{T} [\mathbf{Z}_t | \mathbf{Y}_t, \boldsymbol{\theta}_d][\mathbf{Y}_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}_p][\mathbf{Y}_0][\boldsymbol{\theta}_d, \boldsymbol{\theta}_p]. \quad (4)$$

It is well-established how to do inference and prediction on the parameters and unobserved state process in cases where m_t and n are fairly small, or in situations when the parameters $\boldsymbol{\theta}_d, \boldsymbol{\theta}_p$ are known (e.g., see Shumway and Stoffer 2006 for an overview). However, in the spatio-temporal context, the dimensions m_t and, more critically, n, are often very large. Indeed, one of the biggest challenges in dynamical spatio-temporal modeling is effectively reducing the dimensionality associated

with the parameter space and/or the process in a manner that accommodates realistic dynamics. In this sense, the key issues involve the specification of the data, process, and parameter models (Berliner 1996). In general, the data model is written: $\mathbf{Z}_t = \mathcal{H}(\mathbf{Y}_t; \boldsymbol{\theta}_d; \boldsymbol{\epsilon}_t)$, where $\mathcal{H}$ is some observation function that maps the hidden process $\{\mathbf{Y}_t\}$ to the data, and $\{\boldsymbol{\epsilon}_t\}$ is a noise process. Similarly, the process evolution model is written $\mathbf{Y}_t = \mathcal{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_p; \boldsymbol{\eta}_t)$, where $\mathcal{M}$ is the evolution operator that accommodates the dynamics and $\{\boldsymbol{\eta}_t\}$ is a noise process. In both cases, these functions can be linear or nonlinear, and the error/noise processes can be additive or multiplicative, Gaussian or non-Gaussian, or, if the focus is on deterministic data or process models, the error/noise processes can be set to zero.

The goal of this paper is to illustrate how one can use process knowledge as *motivation* for particular structures and parameterizations in DSTMs. Specifically, we will discuss motivating structures for linear dynamical spatio-temporal models in Sect. 2, followed by a presentation of a general class of spatio-temporal nonlinear dynamical models, focusing on what we call "general quadratic nonlinearity," in Sect. 3. We will emphasize that this model class provides a useful framework in which to characterize various science-based parameterizations and is a convenient mechanism for implementation of fairly complicated science-based models. In addition, we will address the fundamental issues of lack of realism and the curse of dimensionality, as well as computational considerations. In Sect. 4 we illustrate the quadratic nonlinear spatio-temporal dynamic model in a reduced-dimension parameterization of monthly Pacific sea surface temperature (SST) for the purposes of long-lead forecasting. Finally, we will conclude with a discussion in Sect. 5. Overall, this paper serves as an overview of statistical approaches to spatio-temporal dynamic modeling, but with a focus on model specifications driven by mathematical forms of scientific knowledge.

2 Linear dynamic spatio-temporal models

As an illustration of issues associated with dynamic spatio-temporal models, we focus in this section primarily on the linear case with additive Gaussian errors. This is by far the most common setting in the literature and it certainly can present significant modeling challenges for situations in which the process or data vectors are of high dimension. We briefly discuss deviations from these assumptions as well.

2.1 Observation model

In the linear Gaussian case, the data (observation) model can often be written:

$$\mathbf{Z}_t = \mathbf{d}_t + \mathbf{H}_t \mathbf{Y}_t + \boldsymbol{\epsilon}_t, \quad t = 1, \dots, T, \tag{5}$$

where, as defined previously, $\mathbf{Z}_t$ and $\mathbf{Y}_t$ are the $m_t \times 1$ and $n \times 1$ data and process vectors, respectively, $\mathbf{d}_t$ is an additive bias term, $\mathbf{H}_t$ is the observation mapping matrix, and the additive errors $\boldsymbol{\epsilon}_t \sim \text{Gau}(\mathbf{0}, \mathbf{R}_t)$, are assumed to be independent across time and independent of $\mathbf{Y}_t$. This error process represents measurement error and/or small-scale (in space and/or time) noise processes. In this case, $\mathbf{d}_t$ might be parameterized in terms of some known covariates (e.g., $\mathbf{d}_t = \mathbf{X}_t \boldsymbol{\beta}$) and may accommodate

large-scale spatio-temporal features such as seasonality, time trends, and exogenous processes; it may or may not be time-varying. We note that seasonality can also be accounted for in the state process (e.g., see West and Harrison 1997), however, we prefer to decouple the typically more deterministic seasonality from the underlying dynamics. For environmental processes, seasonality is often a function of processes that can be considered exogenous. In principle, $\mathbf{d}_t$ can be a random process as well, but must be restricted in some way to avoid identifiability issues with the error process and the hidden process, $\mathbf{Y}_t$. The matrix $\mathbf{H}_t$, often assumed to be known, can accommodate missing observations as an incidence matrix (e.g., Wikle et al. 1998). We note that although $\mathbf{Y}_t$ need not be spatial in this case (since it is just a hidden random process), the product $\mathbf{H}_t\mathbf{Y}_t$ does correspond to the m_t spatial locations in $\mathbf{Z}_t$.

More generally, one can also use the observation mapping matrix to accommodate dimension reduction of the state process. For example, assume we can decompose the spatial process into two random components,

$$\mathbf{Y}_t = \mathbf{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{v}_t, \tag{6}$$

where $\mathbf{\Phi}$ is an $n \times p$ matrix, $\boldsymbol{\alpha}_t$ is a $p \times 1$ vector, and $\boldsymbol{v}_t$ is an $n \times 1$ vector (e.g., see Wikle 1996; Wikle and Cressie 1999). In this case, we assume that $\boldsymbol{\alpha}_t$ is correlated in time (e.g., a dynamic process) and $\{\boldsymbol{v}_t\}$ represents spatial processes that are independent across time (e.g., $\boldsymbol{v}_t \sim \text{iid Gau}(\mathbf{0}, \Sigma_v)$). Note, as mentioned above, if $\mathbf{Y}_t$ is spatially referenced, $\{\boldsymbol{\alpha}_t\}$ does not need to be a spatial process, as long as $\{\mathbf{\Phi}\boldsymbol{\alpha}_t\}$ is. Substituting (6) into (5), we get:

$$\mathbf{Z}_t = \mathbf{d}_t + \mathbf{H}_t\mathbf{\Phi}\boldsymbol{\alpha}_t + \mathbf{H}_t\boldsymbol{v}_t + \boldsymbol{\epsilon}_t = \mathbf{d}_t + \mathbf{K}_t\boldsymbol{\alpha}_t + \boldsymbol{\gamma}_t, \tag{7}$$

where $\mathbf{K}_t \equiv \mathbf{H}_t\mathbf{\Phi}$, $\boldsymbol{\gamma}_t \sim \text{iid Gau}(\mathbf{0}, \mathbf{R}_{\gamma,t})$, and $\mathbf{R}_{\gamma,t} \equiv \mathbf{H}_t\Sigma_v\mathbf{H}_t' + \mathbf{R}_t$. Obviously, critical to this decomposition is that the dimension of $\boldsymbol{\alpha}_t$ must be much smaller than that of $\mathbf{Y}_t$, i.e., $p \ll n$, or, the decomposition should be such that $\mathbf{R}_{\gamma,t}$ has very simple structure (see below). That is, the specification of $\mathbf{\Phi}$, or equivalently, $\mathbf{K}_t$ and $\mathbf{R}_{\gamma,t}$ in (7) requires some attention.

Perhaps most important for this discussion is the specification of the mapping matrix, $\mathbf{K}_t$. There is an extensive literature on various ways to specify $\mathbf{K}_t$. One option is to use basis functions $\mathbf{\Phi}$ such as orthogonal polynomials, empirical orthogonal functions, wavelets, process normal modes, or splines (e.g., Wikle 1996; Mardia et al. 1998; Wikle and Cressie 1999; Berliner et al. 2000; Stroud et al. 2001; Wikle et al. 2001; Huang and Hsu 2004; Xu et al. 2005; Johannesson et al. 2007; Cressie et al. 2009). In these formulations, the process is typically not spatially indexed, but rather represents some "spectral" coefficients (in the sense that they represent the projection of the true spatial process onto the set of p basis functions). An additional advantage to this type of expansion is that the spectral decomposition may act, to some extent, as a decorrelator, leading to near-diagonal structure in $\mathbf{R}_{\gamma,t}$ (e.g., Wikle 2002b; Royle and Wikle 2005; Paciorek 2007). Alternatively, one can consider the $\{\boldsymbol{\alpha}_t\}$ process to be spatially indexed, but on a different scale than $\{\mathbf{Y}_t\}$. For example, Wikle and Berliner (2005) show how one can use $\mathbf{K}_t$ to accommodate spatio-temporal change of support/alignment. In addition, $\mathbf{K}_t$ can represent spatial convolution kernels in the spirit of Barry and Ver Hoef (1996) and Higdon (1998), in which case $\boldsymbol{\alpha}_t$ are the nodes

of the discrete convolution process (e.g., Calder et al. 2002; Calder 2007; Lemos and Sansó 2009). In this case, one might specify the kernel parameters or estimate them (i.e., one is parameterizing $\mathbf{K}_t$). In that sense, one can also try to "estimate" $\mathbf{K}_t$ more fully, as in the linear model of coregionalization (e.g., Sansó et al. 2008) and factor analysis (e.g., Lopes et al. 2008). In the case where one is parameterizing the mapping matrix and non-trivial process dynamics, one typically must specify strong prior structure on the parameters *and/or* the associated evolution model (see Sect. 2.2) in order to satisfy practical identifiability issues.

Typically, one must choose a fairly simple structure for $\mathbf{R}_{\gamma,t}$ since the dimensionality is high. By "simple," we mean that there is no dependence (i.e., a diagonal structure) or, at least, that this covariance matrix is effectively parameterized in terms of a very few number of parameters. Thus, one faces the same issues as in purely spatial modeling, and the choice for this parameterization depends, in part, on the dimensionality of the observation vectors, as well as the type of spatial data one is considering (e.g., lattice, "geostatistical", etc.). One approach that is useful and computationally efficient in cases where it can be assumed that $\boldsymbol{v}_t$ is composed primarily of expansions on additional basis functions orthogonal to $\boldsymbol{\Phi}$ is demonstrated in a Bayesian context in Berliner et al. (2000) and in an EM-algorithm context in Xu and Wikle (2007). Alternatively, this additional error process can be further decomposed in terms of another basis function expansion to account for non-trivial spatial or temporal dependence. In this case, the associated expansion coefficients are also modeled at the next stage of the model hierarchy. For example, Wikle et al. (2001) include two basis function expansions, where the primary dynamical process was expanded in terms of physically-meaningful system normal mode basis functions and the secondary process was expanded in terms of multiresolution spatial wavelet basis functions in order to accommodate turbulent-scale dynamics.

Although the discussion above is concerned with Gaussian data models, it is fairly straightforward to apply the generalized linear mixed model framework (e.g., see Mc-Culloch and Searle 2001 for an overview) in which the data are, conditional upon the true spatio-temporal process, independent and follow a distribution from the exponential family. This was formulated in the spatial literature in Diggle et al. (1998) and has been considered by numerous authors in the context of state-space models (e.g., Kitagawa 1987; Fahrmeir and Kaufmann 1991; Carlin et al. 1992; Fahrmeir 1992; Gamerman 1998). The use of such data models in spatio-temporal processes is common (e.g., Sansó and Guenni 1999; Brix and Diggle 2001; Wikle 2003; Wikle and Hooten 2006; Hooten et al. 2007; Hooten and Wikle 2008). In these contexts, one assumes that the data are from an exponential family and some transformation of the true spatio-temporal dynamic process is Gaussian. In some cases with non-Gaussian data, the usual exponential family/canonical link formulations may not be the most realistic, and one has to consider data models that are more appropriate for the process (e.g., Cangelosi and Hooten 2009). Similarly, nonlinear mapping functions could also be considered in data models. Since the focus of this paper is on the process dynamics, we will not consider non-Gaussian and nonlinear data models explicitly.

It is important to note that the general application of the linear data model presented here applies whether one considers linear or nonlinear process models. For example, an incidence matrix can accommodate missing observations regardless of

the dynamic structure on $\mathbf{Y}_t$ since, in the data model, the observations are conditional on a known $\mathbf{Y}_t$. Indeed, one of the strengths of the hierarchical conditional partitioning of data-given-process and process models is that the process model can be considered separately from the data.

2.2 Process evolution model

Given the observation model (5), the classical first-order evolution model for linear processes with additive Gaussian errors is given by

$$\mathbf{Y}_t = \mathbf{M}\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \tag{8}$$

where $\boldsymbol{\eta}_t \sim \text{iid Gau}(\mathbf{0}, \mathbf{Q})$ (and $\boldsymbol{\eta}_t$ is independent of $\mathbf{Y}_{t-1}$). Clearly, the dynamics are controlled by the propagator (evolution or transition) matrix $\mathbf{M}$ and the conditional covariance matrix, $\mathbf{Q}$. In general, the propagator and conditional covariance matrices can vary with time, but we consider the simpler case for the moment. As mentioned previously, in traditional time-series modeling, it has long been known how to perform parameter estimation and state process prediction (as well as filtering and smoothing). However, there are some fundamental differences in the context of spatio-temporal processes. In particular, spatio-temporal dynamics are a result of the *interaction* of the process across space and time and/or across scales of variability. Furthermore, some types of interaction make sense for some processes, and some do not. In general, process knowledge should not be ignored if it is available. Another aspect of modeling spatio-temporal dynamics is that it is often the case that n is large and one cannot reliably estimate $\mathbf{M}$ and $\mathbf{Q}$ without additional information to help reduce the dimension of the parameter space. In what follows, we present some typical approaches that have been considered to address this issue.

The dynamical interaction in $\{\mathbf{Y}_t\}$ is largely controlled by the propagator matrix $\mathbf{M}$. Perhaps the simplest parameterization of this matrix is $\mathbf{M} = \mathbf{I}$, corresponding to a multivariate random walk (e.g., Gelfand et al. 2005). There is obviously a tremendous reduction of the parameter space in this case, but this comes with the limitation that such specifications may be unrealistic for many spatio-temporal processes, particularly those that exhibit spatial spread (diffusion) or propagation through space (advection). In addition, modeling the process as group of correlated univariate autoregressions, $\mathbf{M} = m\mathbf{I}$ (e.g., Huang and Cressie 1996), or the nonseparable version, $\mathbf{M} = \text{diag}(\mathbf{m})$, where $\mathbf{m} \equiv (m_1, \ldots, m_n)'$ (i.e., spatially varying parameters; e.g., Wikle et al. 1998; Gelfand et al. 2003), does not allow for the spatio-temporal interactions necessary to describe such processes. Such parameterizations are simple to implement, and may be helpful when modeling time-varying parameters, but are probably not sufficient to accommodate the dynamics of the process itself.

At a minimum, to accommodate realistic spatio-temporal dynamics, one typically needs at least lagged (in time) dependence on "nearest" neighbors. That is, $Y(\mathbf{s}_i; t)$ must be related to its neighbors $Y(\mathbf{s}_j; t), j \in \mathcal{N}_i$, where $\mathcal{N}_i$ 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 general, one expects

these lagged nearest-neighbor relationships to vary with space; that is, the parameters $\boldsymbol{\theta}$ that describe the interaction should be dependent on location, $\boldsymbol{\theta}_i$, $i = 1, \ldots, n$. Although this can present a problem in terms of estimation from a classical perspective, Wikle et al. (1998) showed that the Bayesian hierarchical framework is useful in these cases because the parameters $\boldsymbol{\theta}_i$ can be modeled at the next stage of the hierarchy and spatial dependence can easily be accommodated, thus reducing the effective number of parameters in the model. In such cases, it is sometimes difficult to specify which parameters should vary spatially, and in what way. Science-based mathematical model formulations can help provide a structure in which to make such decisions in an informed manner.

2.2.1 Scientific-based parameterization

As discussed above, one must often parameterize the propagator and covariance matrices. It is natural that *a priori* scientific process information can be useful to motivate such parameterizations. For example, many real-world spatio-temporal processes exhibit linear dynamics that can be classified as "diffusion" and "advection." Advection simply refers to the process being moved along by the background "flow" (e.g., wind, or ocean current) and diffusion on the macro-scale refers to the process spreading out over a larger and larger area with time (and thus decreasing its "concentration" over time). Note that on the micro-scale, diffusion is a result of the random motion of individual particles.

In the differential equation context, one could write a simple advection/diffusion process as

$$\frac{\partial Y}{\partial t} = \nabla \cdot (aY) + \nabla \cdot (b\nabla Y),$$

where ∇ is the multivariate gradient operator and a is the advection parameter (typically representing the flow process) and the diffusion coefficient b controls the rate of spread. In one spatial dimension, this reduces to

$$\frac{\partial Y}{\partial t} = \frac{\partial (aY)}{\partial s} + \frac{\partial}{\partial s}\left(b\frac{\partial Y}{\partial s}\right).$$

In both the advection and diffusion cases, the associated parameters may vary with space (i.e., $a(s)$, $b(s)$), and of course, they can be considered separately so the process is solely an advection or diffusion process. As shown in Wikle (2003), Xu and Wikle (2007), Malmberg et al. (2008) and Stroud et al. (2010), finite-difference discretizations of these PDEs imply a lagged nearest-neighbor parameterization of $\mathbf{M}$ (i.e., sparse, with tri- (penta-) diagonal structure in the one (two) dimensional case), where the parameterization is controlled by the advection and diffusion parameters. In cases where these parameters are spatially varying, these parameters can be assigned spatial random fields at a lower level of a general hierarchical model framework.

We also note that the form of the PDE might suggest higher-order Markovian dependence in the discretized statistical model. For example, a hyperbolic wave equation such as $\partial^2 Y/\partial t^2 = c\partial^2 Y/\partial s^2$ would imply a second-order difference equation if discretized by centered finite differences. In general, the type of discretization one

considers (e.g., finite differences, Runge–Kutta, explicit, implicit, etc.) and the order of the discretization can motivate a different spatio-temporal statistical model. As in all statistical modeling, one should chose the most parsimonious parameterization that can accommodate the dynamics in the process and still be informed by the observations. As mentioned in the introduction, in more simple parameterizations (e.g., fixed parameters) with stochastic error processes, PDEs can be shown to imply classes of fairly complicated marginal spatio-temporal covariance functions (e.g., Heine 1955; Whittle 1954, 1962; Jones and Zhang 1997).

One might also motivate the spatio-temporal dynamics by a discrete time, continuous space, linear integro-difference equation (IDE) model, such as:

$$Y_t(s) = \int_D m_s(r; \theta_s) Y_{t-1}(r) \, dr, \tag{9}$$

where D is the spatial domain of interest. The key component of (9) is the *redistribution kernel* $m_s(r; \theta_s)$ which potentially has a different shape for each spatial location s (controlled by the parameters θ_s). In particular, as reviewed in Wikle (2002a) and Xu et al. (2005), the shape and speed of diffusive wave fronts are a function of the kernel width and tail behavior, and non-diffusive propagation (e.g., advection) is a function of the degree of kernel skewness relative to the reference location s. In the deterministic setting, such models have been used to describe the evolution of ecological processes (e.g., Kot 1992; Kot et al. 1996) and stochastic versions have been considered in statistics and engineering (e.g., Wikle 1996; Wikle and Cressie 1999; Wikle 2002a; Huang and Hsu 2004; Xu et al. 2005; Dewar et al. 2009; Scerri et al. 2009). As with the PDE-motivated models, one typically has to discretize the integral equation, or use a spectral (Galerkin) expansion to reformulate the model. The key advantage of this is that on a discretized grid (e.g., $\{s_1, \ldots, s_n\}$), the kernel defined for location s_i, $m_{s_i}(r; \theta_{s_i})$, corresponds to the ith row of the propagator matrix $\mathbf{M}$. Thus, again, we see that the effective "weights" applied to the linear combination of previous process values control the dynamics. As with the PDE discretization, it is crucial to be able to allow the parameters to be random (typically, random processes) at a lower level of a model hierarchy in order to model many environmental processes. In addition, we note that like the stochastic PDEs, stochastic versions of the IDE (with fairly simple parameter structure) can be related to classes of marginal space-time covariance functions (e.g., Brown et al. 2000; Storvik et al. 2002). Note that the expansion of the redistribution kernel and process in terms of linear combinations of basis functions is similar to geostatistical functional analysis, especially that associated with co-kriging spatio-temporal data (e.g., see Goulard and Voltz 1993; Giraldo et al. 2009, and the overview in Delicado et al. 2009). However, in those cases, the time variation is considered to be a continuous function rather than a Markovian dynamical process as considered here.

In the case of a dimension-reduced data model (7), one effectively considers the process equation:

$$\boldsymbol{\alpha}_t = \mathbf{M}_\alpha \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_{\alpha,t}, \tag{10}$$

where the $\boldsymbol{\eta}_{\alpha,t} \sim \text{iid} \, \text{Gau}(\mathbf{0}, \mathbf{Q}_\alpha)$ (with the usual independence assumptions). In this case, if the dimension of p is sufficiently small, one might be able to use standard approaches (e.g., see Shumway and Stoffer 2006) for estimating the full propagator and

noise covariance matrices, $\mathbf{M}_\alpha$, $\mathbf{Q}_\alpha$, respectively (subject to the identifiability issues discussed previously for the case when $\mathbf{K}_t$ is not known a priori.) However, in many cases, the dimensionality of p is still too large to accommodate direct estimation, and the propagator and covariance matrices must again be parameterized. Although one could again utilize naive time-series parameterizations (i.e., random walks, simple univariate autoregressions, etc.), one should still try to ensure that such processes are realistic relative to the underlying science. In particular, one can use science-based models to motivate parameterizations for spectral representations of spatio-temporal processes (e.g., (10)) as with spatial-domain models. For example, Wikle et al. (2001) use equatorial wave theory normal modes and wavelets as basis vectors to parameterize large-scale and turbulence-scale dynamics in a linear tropical wind model, and Wikle (2002a), Huang and Hsu (2004), and Xu et al. (2005) use integro-difference equation (IDE) models (i.e., continuous space and discrete time) to parameterize diffusive and advective behavior in a dimension-reduced spectral framework.

This discussion of science-based parameterizations suggests that linear dynamics can easily (and quite efficiently) accommodate realistic types of processes (e.g., advection, diffusion, etc.). In general, if one considers the neighborhood of importance for lagged values of the process to be a "stencil", then the "width" of the stencil (i.e., how many neighbors are important) controls the rate of spread and the degree of "asymmetry" in the stencil controls the speed of the "object's" propagation. Furthermore, more complicated behavior such as long-range dependence can be accommodated by multimodal stencils. These behaviors can be combined and local variations (non-separable and non-stationary behavior) can be accommodated by spatially (or temporally) varying parameters. Again, we emphasize the purpose here is not to model PDEs or IDEs exactly, or to necessarily perform parameter estimation. Rather, the goal is simply to use these models to motivate realistic parameterizations of statistical models for dynamic spatio-temporal processes. Then, we let the data speak!

3 Nonlinear dynamic spatio-temporal models

Although the linear dynamical specifications mentioned in the previous section are quite powerful and useful, there are many real-world processes that exhibit nonlinear spatio-temporal interactions. The focus here is on the process evolution model. For example, many processes exhibit state-dependent (or "density"-dependent growth), e.g., $\partial Y/\partial t = Y g(Y; \boldsymbol{\theta})$ for some nonlinear growth function g (e.g., logistic, Ricker, Beverton-Holt, etc.). In addition, many processes exhibit what is sometimes referred to as nonlinear advection, e.g., in one spatial dimension, $\partial Y/\partial t = Y \partial Y/\partial s$. More general nonlinear spatio-temporal dynamic models must be able to accommodate such processes, among others. Again, the key point is that nonlinear behavior in spatio-temporal systems generally arises from *interactions* across various spatio-temporal scales of variability and across multiple processes.

The issues of dimensionality and efficient parameterization surrounding linearly evolving DSTMs are complicated. Not surprisingly, these issues are even more critical for nonlinearly evolving spatio-temporal systems. In this case, one must have a strong sense of the basic dynamical form or a very large amount of data in order

to perform scientific inference and/or prediction. A very general class of dynamical models in this context is the nonlinear autoregressive model:

$$\mathbf{Y}_t = \mathcal{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_t; \boldsymbol{\eta}_t), \tag{11}$$

where $\mathcal{M}$ is a nonlinear function of the process $\{\mathbf{Y}_t\}$ at the previous time, $\{\boldsymbol{\theta}_t\}$ can be thought of as a collection of parameters, and $\{\boldsymbol{\eta}_t\}$ is a noise/error process. Such a model is quite general. However, without a specified form for $\mathcal{M}$, (11) is not particularly useful for statistical modeling, other than as a indicator that there is first-order conditional dependence in the state process. Of course, higher-order lag terms could be added, as with the linear AR model, but we shall focus on the first-order lag dependence here for simplicity.

The general first-order nonlinear AR process can be simplified by considering the nonlinear state-dependent model:

$$\mathbf{Y}_t = \mathbf{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_t)\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \tag{12}$$

where $\mathbf{M}$ is an $n \times n$ matrix that depends on the value of the state process at the previous time, as well as possibly time-varying parameters, $\{\boldsymbol{\theta}_t\}$, and $\{\boldsymbol{\eta}_t\}$ is a spatially-referenced noise term. Such models are not new to the time-series literature (e.g., see overviews in Priestly 1988; Tong 1990; Chatfield 2004; Fan and Yao 2005). In some cases, with linear evolution matrices and additive errors, recursive algorithms can be used for estimation (e.g., see the review in Young 2000).

In general, state-dependent models are quite powerful and include many nonlinear models as special cases, such as exponential autoregressive models, bi-linear models, and threshold autoregressive models. For example, consider the threshold AR model:

$$\mathbf{Y}_t = \mathbf{M}_t\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t = \begin{cases} \mathbf{M}_1\mathbf{Y}_{t-1} + \boldsymbol{\eta}_{1,t}, & \text{if } f_1(\xi_t) \in c_1 \\ \vdots & \vdots \\ \mathbf{M}_k\mathbf{Y}_{t-1} + \boldsymbol{\eta}_{k,t}, & \text{if } f_k(\xi_t) \in c_k, \end{cases} \tag{13}$$

where $f_i(\xi_t)$ is some function of a time-varying variable ξ_t, and $f_i(\xi_t) \in c_i; i = 1, \ldots, k$, gives the condition under which the ith equation is appropriate. Note that ξ_t can be multivariate and can depend on the state process and/or other processes and parameters. Thus, depending on the value of ξ_t (often unknown or a hidden process), one has a different linear evolution operator and potentially, a different noise process. Clearly, this is related to the state-dependent model in the case where ξ_t is a function of $\mathbf{Y}_{t-1}$, but it is often simpler to implement. Hughes and Guttorp (1994) and Berliner et al. (2000) use such models in atmospheric and ocean science applications, respectively, and Hooten and Wikle (2007) employed threshold models for an ecological application.

In the case of most spatio-temporal problems, the state dimension is quite high, leading to difficulties in parameter estimation. In addition, there may be certain types of process behavior, known *a priori*, that one wishes to model (e.g., advection, diffusion, repulsion, density-dependent growth, etc.). In that case, more efficient estimation can be accomplished by parameterizing the state-dependent transition matrices

consistent with these processes (as with the PDE and IDE examples presented earlier). In the hierarchical context, these parameterizations can be made simpler by allowing complicated dependence at lower levels of the model hierarchy.

3.1 General quadratic nonlinearity

In order to provide a framework that can accommodate science-based parameterizations for nonlinear DSTMs we start with the following rather general nonlinear DSTM:

$$\mathbf{Y}_t = \mathbf{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_{m,t})\mathbf{Y}_{t-1} + \mathbf{F}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_{f,t})\boldsymbol{\eta}_t, \tag{14}$$

where $\mathbf{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_{m,t})$ is a propagator matrix depending on the state $\mathbf{Y}_{t-1}$ and some (possibly) time-varying parameters $\boldsymbol{\theta}_{m,t}$. Furthermore, for sake of generality, we have introduced the matrix, $\mathbf{F}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_{f,t})$, which may also depend on the state process and parameters $\boldsymbol{\theta}_{f,t}$. This matrix controls the dependence structure in the noise process, and can allow it to be state-dependent. However, our focus here is on parameterization of the evolution matrix $\mathbf{M}(\cdot)$. Specifically, a general form for this matrix that can accommodate many realistic nonlinear spatio-temporal dynamical processes is given by

$$\mathbf{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_{m,t}) = \mathbf{A}(\boldsymbol{\theta}_{a,t}) + \mathbf{B}(\boldsymbol{\theta}_{b,t})\mathbf{G}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_{g,t})\mathbf{C}(\boldsymbol{\theta}_{c,t}), \tag{15}$$

where $\boldsymbol{\theta}_{m,t} \equiv \{\boldsymbol{\theta}_{a,t}, \boldsymbol{\theta}_{b,t}, \boldsymbol{\theta}_{c,t}, \boldsymbol{\theta}_{g,t}\}$. Some special cases of processes that can be accommodated by this structure are:

- $\mathbf{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_{m,t}) = \mathbf{A}(\boldsymbol{\theta}_a)$: This corresponds to the first-order linear DSTM described in Sect. 2.
- $\mathbf{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_{m,t}) = \mathbf{B}(\boldsymbol{\theta}_b)\mathbf{G}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_g)$: This formulation of the model reduces to the spatio-temporal "matrix model" commonly used in ecological population dynamics (Caswell 2001; Hooten et al. 2007). That is, $\mathbf{G}$ accounts for growth and $\mathbf{B}$ accounts for diffusion or spread. For example, $\mathbf{B}$ can be parameterized in terms of a kernel function (e.g., Hooten et al. 2007) and the functional form of the growth process defines the structure in $\mathbf{G}$.
- $\mathbf{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_{m,t}) = \mathbf{A}(\boldsymbol{\theta}_a) + \mathbf{G}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_g)$: This form of the model accommodates discretized linear PDE dynamics with a nonlinear growth term. Typically, $\mathbf{G}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_g)$ is diagonal in this case (e.g., Wikle and Hooten 2006; Hooten and Wikle 2008).
- $\mathbf{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_{m,t}) = \mathbf{A}(\boldsymbol{\theta}_a) + \mathbf{G}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_g)\mathbf{C}(\boldsymbol{\theta}_c)$: This form of the model accommodates a discretized PDE with nonlinear advection.

Perhaps the most notable feature of (14) and (15) is the *quadratic* nature of the nonlinear interaction. That is, the second term on the right-hand side (RHS) of (15) includes a function of $\mathbf{Y}_{t-1}$, and this propagator is multiplied by $\mathbf{Y}_{t-1}$ in (14). This form leads us to consider a particular type of nonlinear spatio-temporal model that is fairly robust and allows us to accommodate the types of dynamical behavior often encountered in the environmental sciences. Thus, we focus on a special subclass of

such spatio-temporal models characterized by what we term *general quadratic non-linearity* (GQN):

$$Y_t(\mathbf{s}_i) = \sum_{j=1}^{n} a_{ij} Y_{t-1}(\mathbf{s}_j) + \sum_{k=1}^{n} \sum_{l=1}^{n} b_{i,kl} Y_{t-1}(\mathbf{s}_k) g(Y_{t-1}(\mathbf{s}_l); \boldsymbol{\theta}_g) + \eta_t(\mathbf{s}_i), \quad (16)$$

for $i = 1, \ldots, n$, where the first term on the RHS of (16) contains a linear combination of the process at the previous time (as in the linear model presented earlier), and the double sum on the RHS of (16) contains quadratic interactions of the process and (potentially) some transformation of the lagged process, at the previous time. The term "general" comes in because of the flexibility of the model to accommodate transformations of the process through the function $g(\cdot)$, where this function might depend on parameters $\boldsymbol{\theta}_g$. Note, there are $O(n^3)$ parameters (i.e., a's, b's, $\boldsymbol{\theta}_g$) in this model, which could be quite massive for the typical n one might see in spatio-temporal applications. We note that if $g(\cdot)$ is the identity function, then there are $n(n-1)/2$ unique dyadic interactions for each $i = 1, \ldots, n$; otherwise, there are n^2 such interactions for each i. We also note that further generality could be obtained if we allowed the lagged process in the first (linear) term on the RHS of (16) to also be transformed. Similarly, we could allow the untransformed lagged process in the second term on the RHS of (16) to also be transformed. Such extra generality might be useful for some processes, but the single transformation given by $g(\cdot)$ is typically sufficient for our exposition. We note that the use of quadratic interaction models is not new in the context of time-series modeling (e.g., Fan and Yao 2005, Chap. 8) or for characterizing geophysical processes (e.g., see the overview in Majda and Wang 2006), but its use as a framework for characterizing scientific-based parameterizations of statistical spatio-temporal dynamic models has not been explored extensively.

Aside from the general specification presented in (14), there are other ways to write (16) in matrix notation. We will consider a couple of useful forms. First, let

$$\mathbf{B}_i \equiv \{b_{i,kl}\}_{k,l=1,\ldots,n}, \quad (17)$$

be an $n \times n$ matrix for $i = 1, \ldots, n$. Then,

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \left(\mathbf{I}_n \otimes g(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_g)'\right)\mathbf{B}\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \quad (18)$$

where the $n^2 \times n$ matrix $\mathbf{B}$ is given by

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_n \end{pmatrix}. \quad (19)$$

Thus, in terms of the previously presented general state-dependent model (12), we have

$$\mathbf{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_m) = \mathbf{A} + \left(\mathbf{I}_n \otimes g(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_g)'\right)\mathbf{B}, \quad (20)$$

where the parameters are given by $\boldsymbol{\theta}_m \equiv \{\mathbf{A}, \mathbf{B}, \boldsymbol{\theta}_g\}$.

We can also write the model:

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{B}_v\big(\mathbf{Y}_{t-1} \otimes g(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_g)\big) + \boldsymbol{\eta}_t, \tag{21}$$

where the $n \times n^2$ matrix $\mathbf{B}_v$ is given by

$$\mathbf{B}_v = \begin{pmatrix} \text{vec}(\mathbf{B}_1')' \\ \text{vec}(\mathbf{B}_2')' \\ \vdots \\ \text{vec}(\mathbf{B}_n')' \end{pmatrix}, \tag{22}$$

and the $n^2 \times 1$ vector $\mathbf{Y}_{t-1} \otimes g(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_g)$,

$$\big(\mathbf{Y}_{t-1} \otimes g(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_g)\big)' = \big(Y_{t-1}(\mathbf{s}_1)g\big(Y_{t-1}(\mathbf{s}_1)\big), \ldots, Y_{t-1}(\mathbf{s}_1)g\big(Y_{t-1}(\mathbf{s}_n)\big),$$
$$Y_{t-1}(\mathbf{s}_2)g\big(Y_{t-1}(\mathbf{s}_1)\big), \ldots, Y_{t-1}(\mathbf{s}_n)g\big(Y_{t-1}(\mathbf{s}_n)\big)\big),$$

contains all of the pair-wise (dyadic) interaction terms (note that we omitted the parameter dependence in the g function for notational clarity).

For unstructured spatio-temporal processes, $\mathbf{A}$, $\mathbf{B}$, and/or $\boldsymbol{\theta}_g$ have too many parameters to estimate reliably. We have already discussed some science-based parameterizations for the linear dynamic spatio-temporal model that suggested realistic ways to reduced the dimensionality of the parameter space. The GQN structure accommodates many science-based parameterizations as well. We give several examples in the next section, followed by an application in Sect. 4.

3.2 GQN parameterization: examples

We illustrate several GQN parameterizations characterized by differing amounts of available scientific information. Specifically, we first illustrate the case where one knows the general form of the nonlinear function $\mathcal{M}(\cdot)$. This shows how a second-order Taylor expansion of such a function fits into the GQN framework. This is followed by two real-world scientific examples, one corresponding to a reaction-diffusion process and one a highly nonlinear fluid dynamics process. Finally, we illustrate how one can also use a spectral reduction in the state process in this framework, with a further reduction in the dynamical interactions through scale analysis.

3.2.1 Example: second-order Taylor expansion

Historically, a common approach to accommodating nonlinear dynamics when the form of the nonlinearity is known, is by a local-linear approximation. For greater accuracy, this is often extended to a second-order approximation of the nonlinear function through the truncated Taylor series expansion (e.g., Jazwinski 1970; West and Harrison 1997). Such an expansion fits into the GQN framework presented here.

Consider the deterministic vector-valued nonlinear function:

$$\mathbf{Y}_t = \mathcal{M}(\mathbf{Y}_{t-1}), \tag{23}$$

where $\mathcal{M}(\mathbf{Y}_{t-1}) \equiv (m_1(\mathbf{Y}_{t-1}), \dots, m_n(\mathbf{Y}_{t-1}))'$, with $m_i(\cdot)$ being scalar-valued non-linear functions. Then, the Taylor expansion about $\mathbf{w}_t$ can be written:

$$\mathcal{M}(\mathbf{Y}_t) = \mathcal{M}(\mathbf{w}_t) + \mathbf{M}_t(\mathbf{Y}_t - \mathbf{w}_t) + \frac{1}{2}\big(\mathbf{I}_n \otimes (\mathbf{Y}_t - \mathbf{w}_t)'\big)\mathbf{H}_t(\mathbf{Y}_t - \mathbf{w}_t) + \cdots, \quad (24)$$

where $\mathbf{M}_t$ is the matrix of first partial derivatives of $\mathcal{M}(\cdot)$ evaluated at $\mathbf{w}_t$,

$$\mathbf{M}_t = \begin{pmatrix} \frac{\partial m_1}{\partial Y_t(\mathbf{s}_1)} & \frac{\partial m_1}{\partial Y_t(\mathbf{s}_2)} & \cdots & \frac{\partial m_1}{\partial Y_t(\mathbf{s}_n)} \\ \vdots & & & \vdots \\ \frac{\partial m_n}{\partial Y_t(\mathbf{s}_1)} & \frac{\partial m_n}{\partial Y_t(\mathbf{s}_2)} & \cdots & \frac{\partial m_n}{\partial Y_t(\mathbf{s}_n)} \end{pmatrix}_{\mathbf{Y}_t = \mathbf{w}_t}, \quad (25)$$

where we have temporarily suppressed the dependence of $m_i(\cdot)$ on $\mathbf{Y}_t$, and

$$\mathbf{H}_t = \begin{pmatrix} \mathbf{H}_{1t}(\mathbf{Y}_t) \\ \vdots \\ \mathbf{H}_{nt}(\mathbf{Y}_t) \end{pmatrix}_{\mathbf{Y}_t = \mathbf{w}_t}, \quad (26)$$

where $\{\mathbf{H}_{it}(\mathbf{Y}_t)\}$ are the Hessian matrices,

$$\mathbf{H}_{it}(\mathbf{w}_t) = \begin{pmatrix} \frac{\partial^2 m_i}{\partial Y_t(\mathbf{s}_1)\partial Y_t(\mathbf{s}_1)} & \frac{\partial^2 m_i}{\partial Y_t(\mathbf{s}_1)\partial Y_t(\mathbf{s}_2)} & \cdots & \frac{\partial^2 m_i}{\partial Y_t(\mathbf{s}_1)\partial Y_t(\mathbf{s}_n)} \\ \vdots & & & \vdots \\ \frac{\partial^2 m_i}{\partial Y_t(\mathbf{s}_n)\partial Y_t(\mathbf{s}_1)} & \cdots & \cdots & \frac{\partial^2 m_i}{\partial Y_t(\mathbf{s}_n)\partial Y_t(\mathbf{s}_n)} \end{pmatrix}_{\mathbf{Y}_t = \mathbf{w}_t}, \quad (27)$$

where, again, we have temporarily suppressed the dependence of $m_i(\cdot)$ on $\mathbf{Y}_t$.

Clearly, this equation is in the GQN form (with respect to $\mathbf{Y}_t - \mathbf{w}_t$), with $\mathbf{M}_t$ and $\mathbf{H}_t$ corresponding to $\mathbf{A}$ and $\mathbf{B}$, respectively. Of course, this requires complete knowledge of the nonlinear functions (i.e., $m_i(\cdot)$), which is not commonly available for most spatio-temporal processes of interest to statisticians.

3.2.2 Example: reaction–diffusion models

Many processes in ecology are motivated by a reaction-diffusion equation (e.g., see the overview in Hastings 1997). For example, after a successful introduction, an invasive species will tend to spread (diffuse) and its population abundance will increase at some growth rate until the carrying capacity of the environment is achieved. Such processes are well-described by reaction-diffusion PDEs. For example, consider the equation

$$\frac{\partial Y}{\partial t} = \frac{\partial}{\partial x}\left(\delta(x, y)\frac{\partial Y}{\partial x}\right) + \frac{\partial}{\partial y}\left(\delta(x, y)\frac{\partial Y}{\partial y}\right) + \gamma_0(x, y)Y \exp\left(1 - \frac{Y}{\gamma_1(x, y)}\right),$$

where $\mathbf{s} \equiv (x, y)'$ and the diffusion coefficient δ, growth parameter γ_0 and carrying capacity γ_1 are spatially varying (in general). In this case, the growth (reaction) process is assumed to follow a logistic formulation. Simple finite-difference

discretization of this PDE (Wikle and Hooten 2006) implies the following vector difference equation:

$$\mathbf{Y}_t \approx \mathbf{A}(\boldsymbol{\delta})\mathbf{Y}_{t-1} + \mathbf{B}_v\big(\mathbf{Y}_{t-1} \otimes g(\mathbf{Y}_{t-1}; \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1)\big), \qquad (28)$$

where $\mathbf{A}(\boldsymbol{\delta})$ is a sparse $n \times n$ matrix with five non-zero diagonals with elements of $\boldsymbol{\delta}$ (accommodating the diffusion portion of the model, as discussed in Sect. 2 for the linear DSTM), and $\mathbf{B}_v$ is a *very* sparse matrix such that the $\mathbf{B}_i$ submatrices are each zero matrices except for 1 in the (i, i)th location, and the ith element of $g(\cdot)$ is $\gamma_0(\mathbf{s}_i) \exp(1 - Y_{t-1}(\mathbf{s}_i)/\gamma_1(\mathbf{s}_i))$. Now, this form would serve as "motivation" for a statistical model, which would typically have an additive or multiplicative error process, with random parameters $(\boldsymbol{\delta}, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1)$ assigned distributions (e.g., spatial random fields) at the next level of a hierarchical model. For details of such a discretization and hierarchical model applied to the prediction of the spread of the invasive Eurasian Collared Dove, see Wikle and Hooten (2006) and Hooten and Wikle (2008). In particular, Wikle and Hooten (2006) includes a general algorithm for implementation, as well as R-code.

Note, other forms of the reaction portion of the model can be considered. For example, the Beverton–Holt (1957) growth formulation of the reaction–diffusion PDE is given by

$$\frac{\partial Y}{\partial t} = \frac{\partial}{\partial x}\left(\delta(x, y)\frac{\partial Y}{\partial x}\right) + \frac{\partial}{\partial y}\left(\delta(x, y)\frac{\partial Y}{\partial y}\right) + \frac{\gamma_0(x, y)Y}{1 + [(\gamma_0(x, y) - 1)Y/\gamma_1(x, y)]},$$

and the discretized form is exactly the same as (28) with the exception of $g(\mathbf{Y}_{t-1}; \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1)$, in which the ith element of this function is $\gamma_0(\mathbf{s}_i)/\{1 + [(\gamma_0(\mathbf{s}_i) - 1)Y_{t-1}(\mathbf{s}_i)/\gamma_1(\mathbf{s}_i)]\}$. Ricker (1954) growth processes can be handled in an analogous fashion.

3.2.3 Example: nonlinear processes in geophysics

The governing processes of geophysical fluid dynamics are inherently nonlinear, due primarily to nonlinear advection, unstable growth, and nonlinear forcing. It can be shown (e.g., Pedlosky 1987; Holton 2004) that a useful approximation to many important processes in the atmosphere and ocean is given by so-called quasi-geostrophic dynamics. In the case of modeling ocean stream function (a function of two-dimensional fluid motion such that lines of constant stream function are everywhere tangent to the local flow), the quasi-geostrophic dynamics are captured in the following PDE:

$$\left(\nabla^2 - \frac{1}{r^2}\right)\frac{\partial\psi}{\partial t} = -J\big(\psi, \nabla^2\psi\big) - \beta\frac{\partial\psi}{\partial x} + \frac{1}{\rho H}\,\mathrm{curl}_z\tau - \gamma\nabla^2\psi - a_H\nabla^4\psi, \quad (29)$$

where ψ is the stream function, $J(a, b) = \partial a/\partial x\,\partial b/\partial y - \partial b/\partial x\,\partial a/\partial y$ is the Jacobian (nonlinear in ψ), τ the wind stress, and $r, \beta, \rho, H, \gamma, a_H$ are parameters related to the flow (e.g., see Berliner et al. 2003 for additional explanation).

As shown in Berliner et al. (2003), simple finite-difference approximations to the derivatives in (29) yield the following vector difference equation:

$$\boldsymbol{\psi}_t \approx \left\{ \mathbf{I} + \delta_t \left(\mathbf{L} - r^{-2}\mathbf{I} \right)^{-1} \left(-\beta \mathbf{D}_x - \gamma \mathbf{L} + a_h \mathbf{L}^2 \right) \right\} \boldsymbol{\psi}_{t-1}$$

$$+ \delta_t \left(\mathbf{L} - r^{-2}\mathbf{I} \right)^{-1} \left(-\mathcal{J}(\boldsymbol{\psi}_{t-1}) + \frac{1}{\rho H} \mathbf{c}(\mathbf{u}_t, \mathbf{v}_t) \right) + \mathbf{W}\boldsymbol{\psi}_{b,t}, \qquad (30)$$

where $\boldsymbol{\psi}_t \equiv (\psi_t(\mathbf{s}_1), \dots, \psi_t(\mathbf{s}_n))'$ represents a discretization of the stream function at n spatial (grid) locations, $\mathbf{I}$ is the $n \times n$ identity matrix, $\mathcal{J}(\boldsymbol{\psi}_t)$ is a discretized Jacobian evaluated at $\boldsymbol{\psi}_t$, $\mathbf{c}(\mathbf{u}_t, \mathbf{v}_t)$ is a vector representing the contribution of the discretized wind-stress curl (forcing) from the atmospheric winds (i.e., north-south wind components, $\mathbf{v}_t$ and east-west components, $\mathbf{u}_t$), $\mathbf{D}_x$ is the matrix operator corresponding to the finite differences in the x-direction, $\mathbf{L}$ is the matrix operator for a discretized Laplacian ∇^2 (i.e., if a 5-point stencil is used to approximate the Laplacian, then $\mathbf{L}$ has five diagonals), $\mathbf{W}\boldsymbol{\psi}_{b,t}$ corresponds to the boundary effects (e.g., see Berliner et al. 2003), and δ_t is the time-discretization constant.

Now, from (30) we see that the linear portion of the GQN formulation suggests that

$$\mathbf{A} = \mathbf{I} + \delta_t \left(\mathbf{L} - r^{-2}\mathbf{I} \right)^{-1} \left(-\beta \mathbf{D}_x - \gamma \mathbf{L} + a_h \mathbf{L}^2 \right). \qquad (31)$$

It is a little more difficult to see how the $\mathbf{B}$ matrix in the GQN formulation is parameterized in this context. In this case, the quadratic nonlinearity is due to the Jacobian term. One can see the quadratic structure when the discretized Jacobian term is written

$$\mathcal{J}(\boldsymbol{\psi}_{t-1}) = \left(\text{diag}(\mathbf{D}_y \mathbf{L} \boldsymbol{\psi}_{t-1}) \mathbf{D}_x + \text{diag}(\mathbf{D}_x \mathbf{L} \boldsymbol{\psi}_{t-1}) \mathbf{D}_y \right) \boldsymbol{\psi}_{t-1}, \qquad (32)$$

where again, $\mathbf{L}$ is the matrix operator for a discretized Laplacian, $\mathbf{D}_x$ is the matrix operator corresponding to the x-direction finite-difference operation, and $\mathbf{D}_y$ is the matrix operator corresponding to the y-direction finite-difference operation. Given a five-point Laplacian discretization and centered difference approximations to horizontal derivatives, one can show that the Jacobian results in quadratic interactions in a relatively near neighborhood of a grid point, say (i, j), which has nearest neighbor to the right $(i+1, j)$, north $(i, j+1)$, etc. For example, for (i, j) an interior point on the discretization grid, this contains dyadic interactions between the stream function at locations given by the Cartesian products,

$$\big((i-1, j), (i+1, j)\big) \times \big((i, j), (i, j+1), (i, j+2), (i+1, j+1), (i-1, j+1),$$

$$(i, j-1), (i, j-2), (i+1, j-1), (i-1, j-1)\big), \quad (33)$$

$$\big((i, j+1), (i, j-1)\big) \times \big((i, j), (i+1, j), (i+2, j), (i+1, j-1), (i+1, j+1),$$

$$(i-1, j), (i-2, j), (i-1, j-1), (i-1, j+1)\big). \quad (34)$$

Note that the specific dyadic interactions for non-interior points can be derived as well (somewhat tediously).

The coefficients of the matrix $\mathbf{B}_v$ in this case are functions of the discretization parameters (e.g., grid point spacings) and model parameters (e.g., β, H, r, γ, a_H). In practice, one would probably allow some or all of these parameters to be random and/or add additional random parameters to account for model uncertainty. Furthermore, an additive error process would most likely be included.

Another feature of the model given in (30) that is outside the general framework presented here is the exogenous atmospheric wind-stress curl forcing term, $\frac{1}{\rho H}\mathbf{c}(\mathbf{u}_t, \mathbf{v}_t)$, and the boundary effect term, $\mathbf{W}\boldsymbol{\psi}_{b,t}$. Within a hierarchical framework, both of these terms are known, conditionally. Indeed, conditioning on other processes in the hierarchical modeling paradigm is one of the most efficient ways to consider boundary effects and multivariate spatial or spatio-temporal processes (e.g., Wikle et al. 2003; Royle and Berliner 1999; Berliner et al. 2003). Obviously, the GQN framework in (16) could be augmented by an extra term on the RHS that accommodated a (conditionally) known additive effect to account for boundary and multivariate process behavior. Since the focus in this paper is on the parameterization of the dynamical propagator, we do not pursue such general formulations here.

3.2.4 Example: reduced–dimension state process

As we mentioned in the case of linear dynamics, there are situations in which the motivating dynamics are not known or they are more complicated than can be specified in a simple (e.g., PDE) equation, and thus, we cannot achieve dimension reduction in the parameters by the motivating science. In such cases, it is often helpful to consider dimension reduction of the state process itself. Recall the dimension-reduced form given in (6). Clearly, the GQN framework could apply to the $p \times 1$ dimensional state vector, $\boldsymbol{\alpha}_t$, in this decomposition as well (recall, we assume $p \ll n$):

$$\alpha_t(i) = \sum_{j=1}^{p} a_{ij}\alpha_{t-1}(j) + \sum_{k=1}^{p}\sum_{l=1}^{p} b_{i,kl}\alpha_{t-1}(k)g\big(\alpha_{t-1}(l); \boldsymbol{\theta}_g\big) + \eta_t(i), \qquad (35)$$

for $i = 1, \ldots, p$. There are on the order of p^3 parameters in this model. If p is very small, we can proceed with "uninformed" estimation. However, in most cases, p is not small and we must make additional simplifying assumptions and/or perform model selection to chose particular interactions that are important (e.g., de Luna and Genton 2005). Here, we consider simplification via a further dimension reduction through scale analysis.

In many problems it is reasonable to further decompose (6):

$$\mathbf{Y}_t = \boldsymbol{\Phi}^{(1)}\boldsymbol{\alpha}_t^{(1)} + \boldsymbol{\Phi}^{(2)}\boldsymbol{\alpha}_t^{(2)} + \boldsymbol{\nu}_t, \qquad (36)$$

where $\boldsymbol{\alpha}_t^{(i)}$ is of dimension $p_i \times 1$ and where $p_i \ll n$. In this case, say that the spectral coefficients $\boldsymbol{\alpha}_t^{(1)}$ correspond in some sense to "large-scale" modes, and $\boldsymbol{\alpha}_t^{(2)}$ correspond to "small-scale" modes. Now, very loosely motivated by the notion of "Reynolds averaging" in the study of turbulence (e.g., see Holton 2004, Chap. 5), assume that the dyadic interactions between components of $\boldsymbol{\alpha}_t^{(1)}$ are explicit, but those among the "small-scale" components $\boldsymbol{\alpha}_t^{(2)}$ are "noise" and the interactions between

the components of $\boldsymbol{\alpha}_t^{(1)}$ and $\boldsymbol{\alpha}_t^{(2)}$ imply random coefficients in a linear combination of the large scale modes. As a simple example, consider the case where $g(\cdot)$ is the identity and $\boldsymbol{\alpha}_t^{(1)} \equiv (\alpha_{1,t}^{(1)}, \alpha_{2,t}^{(1)})'$ and $\boldsymbol{\alpha}_t^{(2)} \equiv (\alpha_{1,t}^{(2)}, \alpha_{2,t}^{(2)}, \alpha_{3,t}^{(2)})'$. Thus, there are three "resolved" dyadic interactions corresponding to the three unique combinations of $\alpha_{i,t}^{(1)}, i = 1, 2$. Furthermore, the six unique dyadic interactions between the "resolved" $\alpha_{i,t}^{(1)}$ and "unresolved" $\alpha_{j,t}^{(2)}, i = 1, 2; j = 1, 2, 3$ coefficients can be thought of as implying linear terms in the large scale modes with unknown but *random* coefficients (corresponding to the components of $\mathbf{A}$ in the GQN formulation). Finally, the six dyadic interactions between the unresolved small scales $\alpha_{j,t}^{(2)}, j = 1, 2, 3$, are assumed to correspond to the noise process. More generally, assuming the $g(\cdot)$ function is identity, we could write the GQN model as:

$$\boldsymbol{\alpha}_t^{(1)} = \mathbf{A}\boldsymbol{\alpha}_{t-1}^{(1)} + \big(\mathbf{I}_{p_1} \otimes \boldsymbol{\alpha}_{t-1}^{(1)'}\big)\mathbf{B}\boldsymbol{\alpha}_{t-1}^{(1)} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \text{Gau}(\mathbf{0}, \mathbf{Q}), \tag{37}$$

where the structure in $\mathbf{Q}$ derives, at least partially, from the unresolved scale interactions in the dyadic combinations of the elements of $\boldsymbol{\alpha}_t^{(2)}$.

3.3 GQN implementation

For very small dimensions n or p, one might be able to use a full parameterization and, if the $g(\cdot)$ functions are identity, estimate the state process and parameters directly in a recursive state-space framework (e.g., Young 2000). However, as mentioned above, in most cases in which the GQN spatio-temporal model is appropriate, one seeks the benefits of allowing the parameters to be random and specifies structured (often dependent) models for them. This extra stage of the model hierarchy suggests that the hierarchical Bayesian paradigm is often preferred for model implementation.

In the hierarchical Bayesian setting, by far the most popular approach to finding the posterior distribution of interest has been Markov Chain Monte Carlo (MCMC). Though, perhaps not the most efficient algorithm, it is robust and pairs nicely with the hierarchical nature of Markovian models. In the case of a nonlinear process model, as we have in the GQN specifications, $\mathbf{Y}_t \sim [\mathbf{Y}_t | \mathcal{M}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}_t)]$, the full-conditional distributions of the latent state variables, $\mathbf{Y}_t$, will not be conjugate (even under Gaussian additive error assumptions) and thus, some form of Metropolis–Hastings updates must be employed. Metropolis updates are inherently less efficient than Gibbs updates with analytical full-conditional distributions and they can also make for a delicate algorithm, potentially requiring extensive user-informed tuning. For these reasons, the strict Markovian nonlinearity has sometimes been approximated using second-order temporal dependence. Specifically, the MCMC algorithm employed to fit the ecological reaction-diffusion model described by Hooten and Wikle (2008) contains an approximation to a quadratic evolution equation whereby the second-order term $\theta \, \text{diag}(\mathbf{Y}_{t-1})\mathbf{Y}_{t-2}$ is used rather than $\theta \, \text{diag}(\mathbf{Y}_{t-1})\mathbf{Y}_{t-1}$ (where, θ is a model parameter). Under normality, this slight second-order change in the evolution equation allows one to analytically obtain the full-conditional distributions pertaining to the latent state variables $\mathbf{Y}_t$ resulting in much faster, more efficient computation via Gibbs sampling. In many cases, one would not want to change the dependence structure in the process model, but this approach could still be used to generate

useful proposals in a block Metropolis update algorithm. Under the GQN umbrella, such modifications are only possible when the function $g(\cdot)$ is linear, and thus are only applicable in certain situations. Note that many other approximations have been proposed so that the efficiency and robustness of the Gibbs sampler can be relied upon in the implementation of nonlinear state-space models (e.g., Wakefield et al. 1994), sometimes through additional rejection algorithms (e.g., Carlin et al. 1992; Jungbacker and Koopman 2007).

There has been considerable recent work on the development and application of sophisticated algorithms for Bayesian inference in general high-dimensional, highly nonlinear models. For example, Haario et al. (2006) discuss the advantages of a "delayed-rejection adaptive Metropolis" (DRAM) sampling algorithm for improving convergence in Metropolis–Hastings steps for highly nonlinear models. Villagran et al. (2008) show the effectiveness of this and other (e.g., "multiple very fast simulated annealing" (MVFSA)) algorithms in the context of parameter estimation in climate models. Further improvements are suggested in the "differential evolution adaptive Metropolis" (DREAM) algorithm of Vrugt et al. (2009). As discussed in the overview of Higdon et al. (2009), these algorithms and other similar ones can be quite effective in improving convergence for nonlinear spatial and spatio-temporal processes. However, they show that in many cases, relatively simple single component Metropolis updates are often as efficient as these more complicated sampling algorithms. This remains an active area of research with increasing development of promising methods (e.g., Christen and Fox 2010).

In cases where the function $g(\cdot)$ is more complicated and Metropolis updates are undesirable due to extensive tuning or efficiency, one may wish to abandon MCMC for an alternative computational method. One alternative class of methods that stand out in the nonlinear setting are sequential Monte Carlo methods (i.e., particle or bootstrap filtering and ensemble Kalman filtering). Approaches involving independent simulation, or importance sampling are appealing because of their simplicity and ease of implementation in a parallel computing environment. Specifically, sequential importance sampling, or particle filtering (Gordon et al. 1993), is based on sampling–importance–resampling methods (Rubin 1987) and can be employed for state-estimation in nonlinear dynamic models. In such cases, the algorithm proceeds by sampling from the process model, calculating weights based on the likelihood, and then resampling the state variables according to those weights. Unfortunately, as with most importance sampling methods, particle filtering can suffer from sample impoverishment (i.e., a few samples carrying most of the weight) in practice. Moreover, these methods have been criticized for not utilizing the current data, $\mathbf{Z}_t$, in the sampling step (Liu 2004). Various modifications to the resampling algorithm can help alleviate the sample degeneracy issue and a particle filter can often be constructed to run much faster than an MCMC algorithm for the same problem. For general discussion, see Doucet et al. (2001).

The ensemble Kalman filter (EnKF; Evensen 1994), by contrast, is a sequential Monte Carlo algorithm that relies on Gaussian assumptions to approximate the distributions of the unobserved state variables $\{\mathbf{Y}_t\}$. In general, the algorithm works like the particle filter, with a set of particles being propagated forward with the dynamic model, though without the resampling that is used in sequential importance sampling

algorithms. Instead of computing weights and resampling, the integrated state vectors are updated using the Kalman gain matrix (as in traditional Kalman filtering) but with the sample forecast mean and covariance matrix from the ensemble of particles. The use of ensemble covariance is the key feature that distinguishes these methods from Kalman filtering and allows the EnKF to be more efficient than a particle filter (due to the lack of resampling). Although EnKF algorithms are stable for high-dimensional state variables and do not have the same sample impoverishment problems that particle filters do, they still often require the use of disjoint parameter and state-estimation procedures (with the parameters commonly estimated via maximum likelihood conditional on the process). Stroud et al. (2010) provide a nice discussion of such methods applied to the sediment transport modeling in Lake Michigan using an advection-diffusion model (which fits into our GQN framework) to motivate the latent dynamical process.

Another method for implementing latent Gaussian models that is recently enjoying great popularity is known as integrated nested Laplace approximation (i.e., INLA; Rue et al. 2009). The INLA approach is a numerically implemented analytical solution for approximating posterior marginals in hierarchical models with latent Gaussian processes. Such models are most certainly a subset of the GQN class we present herein, and thus, an INLA approach may be of use for certain GQN specifications. The advantages of INLA over other stochastic solutions (e.g., importance sampling, MCMC) are primarily that INLA algorithms can be orders of magnitude faster and automatic, thus facilitating model checking and reduction while reducing necessary algorithm tuning. Having used the INLA approach to find marginal posterior distributions for models that can be specified as Gaussian Markov random fields (GMRFs; Rue and Held 2005), we see little reason for using any other currently available approach. That is, in practice, INLA methods appear to be extremely fast and accurate and, thanks to the availability of INLA software (e.g., Martino and Rue 2009), a wide selection of models that meet the INLA criteria can be easily fit using a set of preconstructed routines (similar to a "black box"). Unfortunately, to create more sophisticated INLA implementations, one needs a substantially greater amount of programming skills than is required for many stochastic methods. Thus, for models that have a large number of non-Gaussian hyperparameters and/or process nonlinearities or spatial structures that cannot be coerced into a GMRF, it remains to be shown whether INLA methods are directly helpful. We also feel that the combination of analytical and stochastic procedures shows great promise for implementing models with more general assumptions.

In terms of other useful methods for fitting GQN models, when only characteristics of posterior distributions are of interest, specific numerical optimization procedures have been developed to estimate posterior modes in the nonlinear state-space modeling framework (So 2003). Another approach that may prove to be useful is referred to as variational Bayesian learning (e.g., Raiko et al. 2006). These methods, which are largely dependent on additive Gaussian error in both the observation and process models, can be employed in situations where there exists significant nonlinearities in the first-order structure (i.e., mean) of the hierarchical model components. In essence, these methods work by approximating the intractable posterior distribution of interest with a tractable parametric distribution and then "correcting" the estimation through

the use of a cost function that is based on the discrepancy or misfit between the distributions (often Kullback–Leibler divergence). These methods have proven valuable in applications involving control theory (e.g., Raiko and Tornio 2009), and as they continue to develop, may prove to be useful for environmental and ecological applications as well.

Finally, a promising avenue for fitting GQN models may involve hybrid sequential Monte Carlo and MCMC approaches (e.g., Andrieu et al. 2010). Such methods appear to allow one to construct efficient proposal distributions for high-dimensional, and highly correlated, processes by using sequential importance sampling within the MCMC algorithm in a manner that ensures convergence. Similarly, hybrid methods using MCMC and more flexible Metropolis algorithms for nonlinear state-space models are also promising (Niemi and West 2010).

4 Example: long-lead prediction of tropical SST

As a simple example of the use of the GQN model with dimension reduction, we consider the problem of seasonal forecasting sea surface temperature (SST). A full treatment of this problem is beyond the scope of this paper. Thus, this example is not intended to be comprehensive, but is included just to illustrate how this methodology could be implemented on a problem of real-world importance.

Sea surface temperature is a complicated process associated with atmosphere/ocean interactions on a variety of temporal and spatial scales. Its variability on interannual and interdecadal time scales is strongly tied to changes in atmospheric processes across the globe (e.g., see the overview in Philander 1990). On interannual time scales, the dominant feature in SST is the episodic warming and cooling that occurs with periods of approximately 3–5 years, the so-called El Niño and La Niña phenomena. As observational networks and data assimilation approaches have improved in recent years, the ability to provide skillful forecasts of SST with time leads of 6–12 months has improved dramatically. In addition, because of its complexity, and uncertainties related to the interaction of the atmosphere and ocean, this is one of the few processes in oceanography in which "statistical" long-lead forecast models are competitive with deterministic models (e.g., Barnston et al. 1999). In fact, linear state-space models in a reduced-dimensional space (e.g., based on empirical orthogonal function (EOF) projections) have proven to be quite effective over the years (e.g., Penland and Magorian 1993). Nevertheless, there is evidence that the dynamics of El Niño and La Niña are nonlinear (e.g., Hoerling et al. 1997; Burgers and Stephenson 1999; Kondrashov et al. 2005). Several nonlinear statistical models for SST have been developed in recent years, and their forecast performance is comparable to or better than linear models (e.g., Tangang et al. 1998; Berliner et al. 2000; Tang et al. 2000; Timmermann et al. 2001; Kondrashov et al. 2005.)

The nonlinear SST models in Berliner et al. (2000) and Kondrashov et al. (2005) are most relevant to the modeling framework presented here. Berliner et al. (2000) demonstrated that a hierarchical Bayesian threshold vector AR model in EOF-reduced-dimensional space was able to capture fairly realistic SST magnitudes and

patterns in prediction experiments on the 1997–1998 extreme El Niño and La Niña events. The critical aspect of that model was that the propagator matrix was assumed to vary with time depending on the current and likely SST regimes (i.e., a state-dependent vector AR model, where the future state was informed by a measure of the near-surface wind). An additional novel feature of this model was the incorporation of residual spatial dependence in the data model conditional covariance matrix, based on remaining components of the EOF expansion. In contrast, Kondrashov et al. (2005) consider a quadratically nonlinear model (in terms of a polynomial regression) on EOF-reduced-dimensional space, but with temporally correlated errors. They showed that the quadratic model did out-perform the linear model when predicting the magnitude of extreme SST anomalies. Given these results, we consider the GQN formulation of SST evolution in an EOF-reduced-dimensional space to demonstrate potential skill in long-lead forecasting. Our model is implemented in a hierarchical Bayesian setting. We consider a quadratic formulation similar to Kondrashov et al. (2005), but critically, we allow the process to be treated as random, and we account for the truncation in the data stage similar to Berliner et al. (2003).

4.1 Data and hierarchical model

The data are monthly averaged Climate Analysis Center SST anomalies (where the anomalies are with respect to a January 1970–December 1985 climatology). The data are gridded at a $2°$ by $2°$ resolution over the tropical Pacific from $29°$S–$29°$N latitude and $124°$E–$70°$W longitude for the period January 1970–October 1997. There are 2261 oceanic locations in the dataset and 334 time periods. For a more complete description of these data, see Berliner et al. (2000). Figure 1 shows a sequence of 12 consecutive SST images for January 1997 through December 1997. This illustrates the onset of the extreme 1997 El Niño event.

Let $\mathbf{Z}_t$ denote the $n \times 1$ vector ($n = 2261$) of oceanic SST observations at time t. We then consider the following data stage in our hierarchical model:

$$\mathbf{Z}_t = \mathbf{\Phi}^{(1)} \boldsymbol{\alpha}_t^{(1)} + \boldsymbol{\gamma}_t, \quad \boldsymbol{\gamma}_t \sim \text{Gau}(\mathbf{0}, \mathbf{R}_\gamma), \tag{38}$$

where $\mathbf{\Phi}^{(1)}$ is an $n \times p_1$ (where $p_1 = 10$) matrix of leading EOF basis vectors calculated from the SST anomalies for the period January 1970–March 1997. These first 10 EOFs account for about 88% of the variance in the SST observations. Figure 2 shows image plots of these EOFs. As in Berliner et al. (2000) and Xu and Wikle (2007), we account for the spatial structure of the next p_2 remaining EOFs in the parameterization of the covariance matrix $\mathbf{R}_\gamma$:

$$\mathbf{R}_\gamma \equiv k \left(c_e \mathbf{I} + \sum_{j=1}^{p_2} \lambda_j^{(2)} \boldsymbol{\phi}_j^{(2)} \boldsymbol{\phi}_j^{(2)\prime} \right), \tag{39}$$

where $\boldsymbol{\phi}_j^{(2)}$, $j = 1, \ldots, p_2$ are the $n \times 1$ EOF basis vectors from $p_1 + 1$ to $p_1 + p_2$ (note: $p_2 = 10$ in this application), and $\lambda_j^{(2)}$ are the corresponding eigenvalues. Note that these extra 10 EOFs account for an additional 10% of the variance in the SST observations. Obviously, an $n \times n$ covariance matrix based on these 10 additional
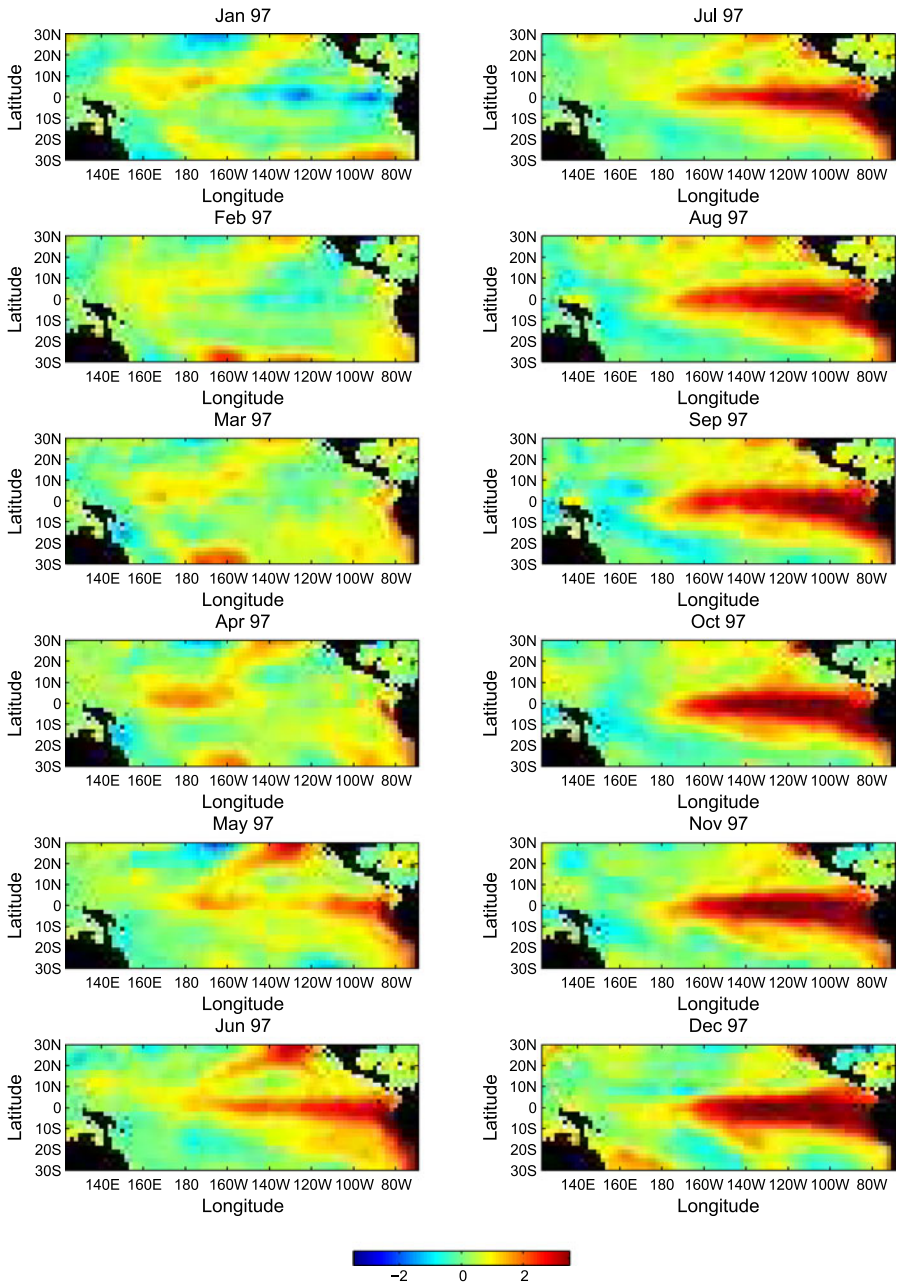
Fig. 1 Sequence of Pacific SST monthly anomalies for January 1997—December 1997

EOFs would be severely rank deficient. Hence, the formulation in (39) includes the additive fixed constant c_e on the diagonal to ensure that the covariance is invertible. In addition, the quantity k is assumed to be random to allow some degree of flexi-
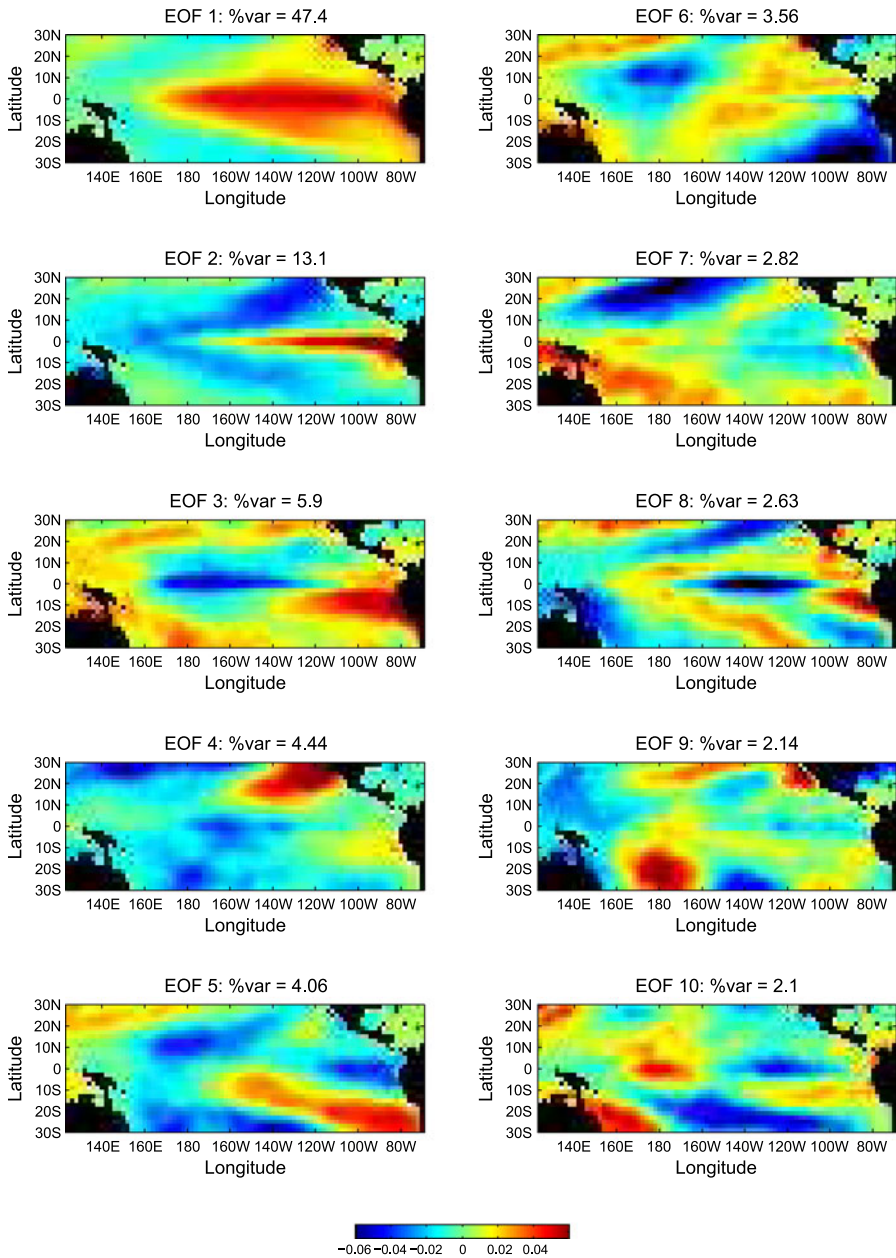
Fig. 2 Image plots of the first 10 EOFs from the SST monthly anomaly data set for the period January 1970—March 1997

bility. As outlined in Berliner et al. (2000), this covariance formulation is efficient to implement computationally as its inverse is analytical, and simple (e.g., see Graybill 1969, p. 426).

The process model is simply the reduced-dimension GQN model (37), with $g(\cdot)$ the identity function:

$$\boldsymbol{\alpha}_{t+\tau}^{(1)} = \mathbf{A}\boldsymbol{\alpha}_t^{(1)} + \left(\mathbf{I}_{p_1} \otimes \boldsymbol{\alpha}_t^{(1)\prime}\right)\mathbf{B}\boldsymbol{\alpha}_t^{(1)} + \boldsymbol{\eta}_{t+\tau}, \quad \boldsymbol{\eta}_{t+\tau} \sim \text{Gau}(\mathbf{0}, \mathbf{Q}). \qquad (40)$$

Note that in this example, the time lead is τ, corresponding to a 7 month-ahead prediction in our case (see the discussion in Berliner et al. 2000). We then specify conjugate parameter models for $\mathbf{A}$, $\mathbf{B}$, $\mathbf{Q}$, and the parameter k in the matrix $\mathbf{R}_\gamma$. Specifically, we specify $\text{vec}(\mathbf{A}) \sim \text{Gau}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_A)$, $\text{vec}(\mathbf{B}) \sim \text{Gau}(\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B)$, $\mathbf{Q}^{-1} \sim$ Wishart$((\nu\mathbf{S})^{-1}, \nu)$, and $k \sim \text{IG}(q, r)$. In our example, we specified the following hyperparameters: $c_e = 0.005$, $\boldsymbol{\mu}_A = \text{vec}(0.9\mathbf{I})$, $\boldsymbol{\Sigma}_A = 100\mathbf{I}$, $\boldsymbol{\mu}_B = \mathbf{0}$, $\boldsymbol{\Sigma}_B = 100\mathbf{I}$, $\mathbf{S} = \mathbf{I}$, $\nu = p_1$, $q = 12$, $r = 9.091 \times 10^{-4}$ (corresponding to a mean of 100 and a variance of 1000). The MCMC algorithm can be written so that $\text{vec}(\mathbf{A})$ and $\text{vec}(\mathbf{B})$ are multivariate normal Gibbs updates, k is an inverse gamma update, and $\mathbf{Q}^{-1}$ is a Wishart update. As discussed in Sect. 3.3, the challenge is the update of the nonlinear state process, $\boldsymbol{\alpha}_t$. In this example, we use a block Metropolis update (updating the p_1 components in a block for each t) with acceptance percentages of about 50%. For the results presented here, the algorithm was run for 4000 iterations beyond a 1000 iteration burn-in. We also compare the GQN forecast to a forecast from the equivalent linear evolution model. In this case, we keep all of the model specifications the same except that we fix $\mathbf{B} = \mathbf{0}$. Note that we did do some preliminary runs to chose $p_1 = 10$ (based on the forecast mean-square prediction error), but as this is simply an illustrative example and not a comprehensive analysis, we have not made an exhaustive attempt to optimize the choice of the number of modes for p_1, p_2, the choice of c_e, nor have we spent much effort checking for sensitivity to prior specifications.

4.2 Results

One of the "gold standards" for evaluating SST prediction models is the ability of the model to capture the onset of the extreme 1997 El Niño. As a simple test of our GQN model, we used data from January 1970–March 1997 to train the model and then generated a forecast distribution for October 1997. For comparison, we also performed such a prediction with an analogous linear evolution model. Figure 3 shows the results for both models, with the GQN model results on the left panels and the linear model results on the right panels. The top row (in both panels) shows the observed anomalies for October 1997, the second row shows the mean of the posterior predictive distribution for the GQN and linear models, the third row shows the pixel-wise 2.5 percentiles from the posterior distribution, and the last row shows the pixel-wise 97.5 percentiles from the posterior distribution. These results suggest that both the GQN and linear models captured the structure of the El Niño, and the GQN model was better able to predict the magnitude of the El Niño. Certainly, the GQN model is better at capturing the uncertainty in the predictive distribution in that it accommodates the extreme nature of this El Niño. Note also that the GQN model better captures the intensity of the SSTs off the western coast of North America.

Obviously, this one forecast does not prove that the GQN model is "better" in general (there may be cases in which it fails to predict the SST as well as the linear model). However, it does illustrate the potential of such models to capture the
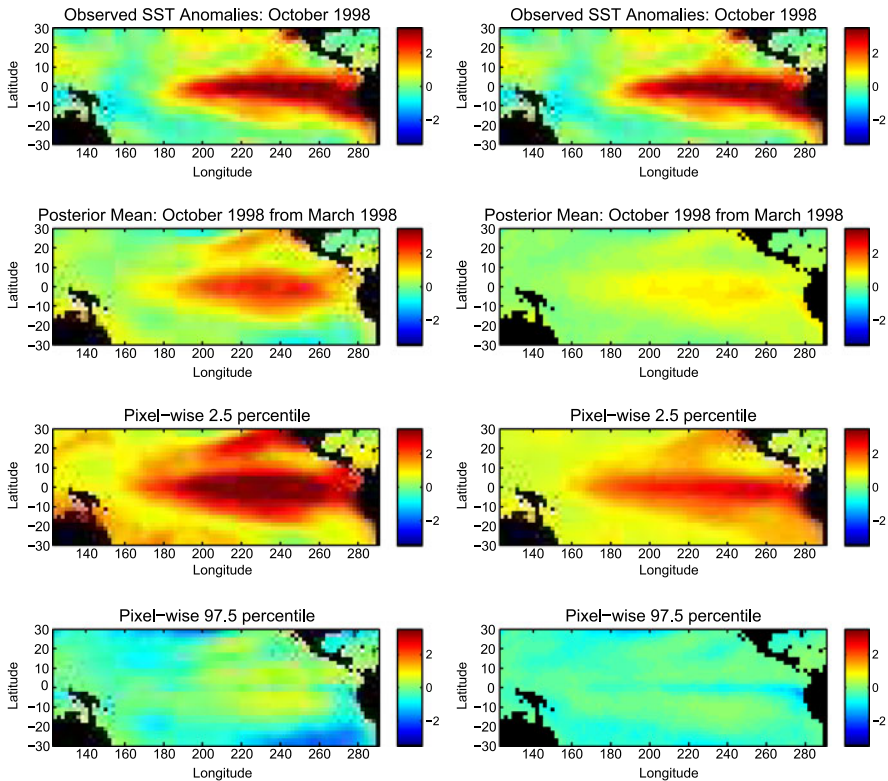
Fig. 3 Forecast of Pacific tropical SST anomalies for October 1997 given observations up to March 1997. *Left panel* shows results for the GQN model, with *top figure* corresponding to the observed anomalies in October 1997, the *second panel* corresponding to the posterior mean prediction from the model, and the *bottom two figures* corresponding to the lower 2.5 and upper 97.5 pixel-wise percentiles from the posterior predictive distribution. The figures on the *right panel* are analogous, but correspond to the linear version of the model (i.e., no dyadic interactions)

nonlinear behavior in the SST evolution, even when little thought has gone into prior specification. Much more investigation into the GQN model's ability to predict across a range of SST conditions (e.g., weaker El Niño's, La Niña's, etc.) and model specifications is needed, but such an investigation is beyond the scope of this paper.

5 Discussion

Spatio-temporal processes in the environmental and physical sciences are almost always the result of some underlying dynamical process. As such, it is important to recognize that statistical models for such processes are inherently different than the dynamical models one might encounter in other areas, such as economics. There has been a tendency in the past to think of dynamic as just "time-varying." Although that is certainly one perspective, it is also the case that dynamic does not *just* mean time-varying—that is, the process matters. As outlined here, even if one recognizes that

there are interactions across space and time, the formulation of dynamical spatio-temporal statistical models is often limited by the curse of dimensionality. In that regard, we suggest that one can often achieve dimension reduction in the parameter space by recognizing the underlying dynamical process, and using that to motivate the statistical model parameterization.

We illustrate how typical PDE and IDE models can motivate the structure for linear evolution models, and that the properties of the lagged neighbor "stencil" accommodates certain dynamical behavior. For example, the width of the stencil corresponds to the rate of diffusion and the skewness of the stencil weights corresponds to advective behavior. Critical to such parameterizations is that the parameters that control such behavior can themselves be dependent random processes (particularly spatial processes in these models). This gives the framework tremendous flexibility and allows the models to fit more complicated real-world processes than described by the motivating dynamical equations.

We use the notion of motivating dynamical equations to suggest a useful parameterization for nonlinear spatio-temporal processes. We term this general quadratic nonlinearity (GQN) and, although not a new idea in time-series modeling, its novelty in the spatio-temporal setting is that it serves as a unifying framework in which to characterize different motivating nonlinear processes (and, of course, linear dynamics is a special case as well). In a reduced-dimensional spectral setting, this model can be implemented directly, but due to the curse of dimensionality, often requires further simplifications in terms of parameter specification, model selection, or, as illustrated here, a scale-reduction in the actual state process.

Obvious extensions to the GQN model are to consider state process transformation functions (i.e., $g(\cdot)$) on the other components (e.g., in the linear portion of the model, and the other state process component of the dyadic interactions). In addition, one might consider additional processes in the model (e.g., Berliner et al. 2003). Finally, there is no reason why models with higher-order (e.g., triad) interactions could not be considered, as well as higher-order lags in the conditional process specification. However, the curse of dimensionality can make such implementations problematic, and alternative state and parameter dimension reduction strategies must be utilized in such a setting.

Clearly, there is always room for improvement in the implementation of nonlinear models in a Bayesian framework. Our SST example considered just a basic implementation, and was not particularly efficient. As discussed in Sect. 3.3, there are numerous alternatives in the literature that could be considered, and the investigation of "optimal" sampling strategies is clearly going to be a major area of research in coming years.

It is important to emphasize here that in the nonlinear example, just as with the linear model, we do not expect the processes that we are modeling to follow the physical/mathematical models exactly. Rather, we expect that the implied statistical model to be flexible enough to adapt to realistic dynamics. This is the primary notion behind what Berliner (2003) terms "physical–statistical" modeling. One potential drawback of such a modeling framework is that one might lose the ability to interpret parameters relative to the motivating physical/mathematical model. This is a natural consequence of having random parameters and allowing the model to adapt to data. Thus,

the model's flexibility to overcome general process misspecification relative to the data also prevents its interpretation in the context of this motivating model. Whether this is a problem depends on the purpose of the modeling. If the goal is retrospective interpolation, it is not critical. If the goal is forecasting, it is only critical if such adaptations force the model into an "explosive" or "non-physical" regime, in which case forecasts might be meaningless. Clearly, if the goal is to actually fit a given physical/mathematical model to data, then this issue is critical. In such situations, it might be better to constrain the parameters and/or to leave off the additive noise term in the process stage. Intuitively, the reason for this is due to the effect of latent stochasticity in the process model. Given that the presence of process error implies that a chosen process model is not known exactly, one cannot expect estimated process parameters to coincide with a specific deterministic evolution equation. In some sense, the process model itself *and* the parameters are compromising by exploiting their flexibility to best fit the data. Thus, in the situation where the process uncertainty is minimal, we would expect parameter estimates to have a more meaningful interpretation with regards to the motivating deterministic evolution equation.

References

Andrieu C, Doucet A, Holenstein R (2010) Particle Markov chain Monte Carlo methods. J R Stat Soc, Ser B 72:1–33

Barnston AG, Glantz MH, He Y (1999) Predictive skill of statistical and dynamical climate models in forecasts of SST during the 1998–1997 El Niño episode and the 1998 La Niña onset. Bull Am Meteorol Soc 80:217–244

Barry RP, Ver Hoef J (1996) Blackbox kriging: spatial prediction without specifying variogram models. J Agric Biol Ecol Stat 1:297–322

Berliner LM (1996) Hierarchical Bayesian time series models. In: Hanson KM, Silver RN (eds) Maximum entropy and Bayesian methods. Kluwer Academic, Dordrecht, pp 15–22

Berliner LM (2003) Physical-statistical modeling in geophysics. J Geophys Res Atmos 108 (D24). doi:10.1029/2002JD002865

Berliner LM, Wikle CK, Cressie N (2000) Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. J Climate 13:3953–3968

Berliner LM, Milliff RF, Wikle CK (2003) Bayesian hierarchical modeling of air-sea interaction. J Geophys Res, Oceans 108 (C4)

Beverton RJH, Holt SJ (1957) On the dynamics of exploited fish populations. In: Fisheries investigations, series 2, vol 19. HM Stationery Office, London

Bisgaard S, Kulahci M (2005) Quality quandaries: interpretation of time series models. Qual Eng 17:653–658

Box GEP, Jenkins GM (1970) Time series analysis, forecasting, and control. Holden–Day, Oakland

Brix A, Diggle PJ (2001) Spatiotemporal prediction for log-Gaussian Cox processes. J R Stat Soc, Ser B 63:823–841

Brown PE, Karesen KF, Roberts GO, Tonellato S (2000) Blur-generated non-separable space-time models. J R Stat Soc, Ser B 62:847–860

Burgers G, Stephenson DB (1999) The "normality" of El Niño. Geophys Res Lett 26:1027–1030

Calder CA (2007) A dynamic process convolution approach to modeling ambient particulate matter concentrations. Environmetrics 19:39–48

Calder CA, Holloman C, Higdon D (2002) Exploring space-time structure in ozone concentration using a dynamic process convolution model. In: Case studies in Bayesian statistics, vol 6. Springer, Berlin, pp 165–176

Cangelosi AR, Hooten MB (2009) Models for bounded systems with continuous dynamics. Biometrics 65:850–856

Carlin BP, Polson NG, Stoffer DS (1992) A Monte Carlo approach to nonnormal and nonlinear state space modeling. J Am Stat Assoc 87:493–500

Caswell H (2001) Matrix population models, 2nd edn. Sinauer Associates, Sunderland

Chatfield C (2004) The analysis of time series: an introduction, 6th edn. Chapman & Hall/CRC Press, Boca Raton

Cressie NAC (1993) Statistics for spatial data, revised edn. Wiley, New York

Cressie N, Huang N-C (1999) Classes of nonseparable, spatiotemporal stationary covariance functions. J Am Stat Assoc 94:1330–1340

Cressie N, Shi T, Kang EL (2009) Fixed rank filtering for spatio-temporal data. Technical report number 819, Department of Statistics, The Ohio State University

Christen JA, Fox C (2010) A general purpose sampling algorithm for continuous distributions (the t-walk). Bayesian Anal 5:1–20

de Luna X, Genton MG (2005) Predictive spatio-temporal models for spatially sparse environmental data. Stat Sin 15:547–568

Delicado P, Giraldo R, Comas R, Mateu J (2009) Statistics for spatial functional data; some recent contributions. Envirometrics. doi:10.1102/env.1003

Dewar M, Scerri K, Kadirkamanathan V (2009) Data-driven spatio-temporal modeling using the integro-difference equation. IEEE Trans Signal Process 57(1):83–91

Diggle PJ, Tawn JA, Moyeed RA (1998) Model-based geostatistics (with discussion). Appl Stat 47:299–350

Doucet A, de Freitas N, Gordon N (eds) (2001) Sequential Monte Carlo methods in practice. Springer, New York

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. J Geophys Res 99:10143–10162

Fan J, Yao Q (2005) Nonlinear time series. Springer, New York

Fahrmeir L (1992) Posterior mode estimation by extended Kalman filtering for a multivariate dynamic generalized linear models. J Am Stat Assoc 87:501–509

Fahrmeir L, Kaufmann H (1991) On Kalman filtering, posterior mode estimation and Fisher scoring in dynamic exponential family regression. Metrika 38:37–60

Fuentes M, Chen L, Davis JM (2008) A class of nonseparable and nonstationary spatio temporal covariance functions. Environmetrics 19:487–507

Gamerman D (1998) Markov chain Monte Carlo for dynamic generalized linear models. Biometrika 85:215–227

Gelfand AE, Kim H-J, Sirmans CF, Banerjee S (2003) Spatial modeling with spatially varying coefficient processes. J Am Stat Assoc 98:387–396

Gelfand AE, Banerjee S, Gamerman D (2005) Spatial process modelling for univariate and multivariate dynamic spatial data. Environmetrics 16:465–479

Giraldo R, Delicado P, Mateu J (2009) Continuous time-varying kriging for spatial prediction of functional data: An environmental application. J Agric Biol Environ Stat. doi:10.1198/jabes.2009.08051

Gneiting T (2002) Nonseparable, stationary covariance functions for space-time data. J Am Stat Assoc 97:590–600

Gordon NJ, Salmond DJ, Smith AFM (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc F Radar Signal Process 140:107113

Goulard M, Voltz M (1993) Geostatistical interpolation of curves: a case study in soil science. In: Soares A (ed) Geostatistics Troia 92, vol 2. Kluwer Academic, Dordrecht, pp 805–816

Graybill FA (1969) Matrices with applications in statistics. Wadsworth, Belmont

Gregori P, Porcu E, Mateu J, Sasvàri Z (2008) Potentially negative space time covariances obtained as a sum of products of marginal ones. Ann Inst Stat Math 60:865–882

Haario H, Laine M, Mira A, Saksman E (2006) DRAM: efficient adaptive MCMC. Stat Comput 16:339–354

Hastings A (1997) Population biology: concepts and models. Springer, New York

Heine V (1955) Models for two-dimensional stationary stochastic processes. Biometrika 42:170–178

Higdon D (1998) A process-convolution approach to modelling temperatures in the North Atlantic ocean. Environ Ecol Stat 5:173–190

Higdon D, Reese CS, Moulton JD, Vrugt JA, Fox C (2009) Posterior exploration for computationally intensive forward models. Technical report LAUR 08-05905, Statistical Sciences Group, Los Alamos National Lab

Holton JR (2004) An introduction to dynamic meteorology, 4th edn. Elsevier, Boston

Hooten MB, Wikle CK (2007) Shifts in the spatio-temporal growth dynamics of shortleaf pine. Environ Ecol Stat 14:207–227

Hooten MB, Wikle CK (2008) A Hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. Environ Ecol Stat 15:59–70

Hooten MB, Wikle CK, Dorazio RM, Royle JA (2007) Hierarchical spatio-temporal matrix models for characterizing invasions. Biometrics 63:558–567

Hoerling MP, Kumar A, Zhong M (1997) El Niño, La Niña, and the nonlinearity of their teleconnections. J Climate 10:1769–1786

Hotelling H (1927) Differential equations subject to error, and population estimates. J Am Stat Assoc 22:283–314

Huang H-C, Cressie N (1996) Spatio-temporal prediction of snow water equivalent using the Kalman filter. Comput Stat Data Anal 22:159–175

Huang H-C, Hsu N-J (2004) Modeling transport effects on ground-level ozone using a non-stationary space-time model. Environmetrics 15:5–25

Hughes JP, Guttorp P (1994) A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. Water Resour Res 30:1535–1546

Jazwinski AH (1970) Stochastic processes and filtering theory. Academic Press, New York

Johannesson G, Cressie NAC, Huang H-C (2007) Dynamic multi-resolution spatial models. Environ Ecol Stat 14:5–25

Jones RH, Zhang Y (1997) Models for continuous stationary space-time processes. In: Gregoire TG, Brillinger DR, Diggle PJ, Russek-Cohen E, Warren WG, Wolfinger RD (eds) Modelling longitudinal and spatially correlated data. Lecture Notes in Statistics, vol 122. Springer, Berlin, pp 289–298

Jungbacker B, Koopman SJ (2007) Monte Carlo estimation for nonlinear non-Gaussian state space models. Biometrika 94(4):827–839

Kitagawa G (1987) Non-Gaussian state-space modeling of non-stationary time series. J Am Stat Assoc 82:1032–1041

Kondrashov D, Kravtsov S, Robertson AW, Ghil M (2005) A hierarchy of data-based ENSO models. J Climate 18:4425–4444

Kot M (1992) Discrete-time travelling waves: Ecological examples. J Math Biol 30(4):413–436

Kot M, Lewis MA, van den Briessche P (1996) Dispersal data and the spread of invading organisms. Ecology 77(7):2027–2042

Lemos RT, Sansó B (2009) A spatio-temporal model for mean, anomaly and trend fields of North Atlantic sea surface temperature. J Am Stat Assoc 104:5–25

Liu JS (2004) Monte Carlo strategies in scientific computing. Springer, New York

Lopes HF, Salazar E, Gamerman D (2008) Spatial dynamic factor analysis. Bayesian Anal 3(4):759–792

Ma C (2003) Families of spatio-temporal stationary covariance models. J Stat Plan Inference 116:489–501

Majda AJ, Wang X (2006) Nonlinear dynamics and statistical theories for basic geophysical flows. Cambridge University Press, Cambridge

Malmberg A, Arellano A, Edwards DP, Flyer N, Nychka D, Wikle CK (2008) Interpolating fields of carbon monoxide data using a hybrid statistical-physical model. Ann Appl Stat 2:1231–1248

Mardia K, Goodall C, Redfern E, Alonso F (1998) The kriged Kalman filter. Test 7:217–285 (with discussion)

Martino S, Rue H (2009) INLA: Functions which allow to perform a full Bayesian analysis of structured additive models using integrated nested laplace approximaxion, R Package

McCulloch CE, Searle SR (2001) Generalized, linear and mixed models. Wiley, New York

Niemi J, West M (2010) Adaptive mixture modeling Metropolis methods for Bayesian analysis of nonlinear state-space models. J Comput Graph Stat 19:260–280

Paciorek CJ (2007) Computational techniques for spatial logistic regression with large datasets. Comput Stat Data Anal 51:3631–3653

Pedlosky J (1987) Geophysical fluid dynamics, 2nd edn. Springer, New York

Penland C, Magorian T (1993) Prediction of Niño 3 sea surface temperatures using linear inverse modeling. J Climate 6:1067–1076

Philander SG (1990) El Niño, La Niña, and the southern oscillation. Academic Press, San Diego

Priestly MB (1988) Non-linear and non-stationary time series analysis. Academic Press, London

Raiko T, Tornio M (2009) Variational Bayesian learning of nonlinear hidden state-space models for model predictive control. Neurocomputing 72:3704–3712

Raiko T, Tornio M, Honkela A, Karhunen J (2006) State inference in variational Bayesian nonlinear state-space models. In: Proceedings of the 6th international conference on independent component analysis and blind source separation (ICA, 2006), Charelston, South Carolina, March 2006, pp 222–229

Ricker WE (1954) Stock and recruitment. J Fish Res Board Can 11:559–623

Royle JA, Berliner LM (1999) A hierarchical approach to multivariate spatial modeling and prediction. J Agric Biol Environ Stat 4:29–56

Royle JA Wikle CK, (2005) Efficient statistical mapping of avian count data. Environ Ecol Stat 12:225–243

Rubin DB (1987) A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. J Am Stat Assoc 52:543–546

Rue H, Held L (2005) Gaussian Markov random fields—theory and applications. Monographs on statistics and applied probability, vol 104. Chapman & Hall, London

Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc, Ser B 71:319–392

Sansó B, Guenni L (1999) Venezuelan rainfall data analyzed by using a Bayesian space-time model. Appl Stat 48:345–362

Sansó B, Schmidt A, Nobre A (2008) Spatio-temporal models based on discrete convolutions. Can J Stat 36(2):239–258

Scerri K, Dewar M, Kadirkamanathan K (2009) Estimation and model selection for an IDE-based spatio-temporal model. IEEE Trans Signal Process 57(2):482–492

Shumway RH, Stoffer DS (2006) Time series analysis and its applications with R examples, 2nd edn. Springer, New York

So MKP (2003) Posterior mode estimation for nonlinear and non-Gaussian state space models. Stat Sin 13:255–274

Stein M (2005) Space-time covariance functions. J Am Stat Assoc 100:310–321

Storvik G, Frigessi A, Hirst D (2002) Stationary space-time Gaussian random fields and their time autoregressive representation. Stat Model 2:139–161

Stroud J, Mueller P, Sansó B (2001) Dynamic models for spatio-temporal data. J R Stat Soc, Ser B 63:673–689

Stroud JR, Stein ML, Lesht BM, Schwab DJ, Beletsky D (2010) An ensemble Kalman filter and smoother for satellite data assimilation. J Am Stat Assoc. doi:10.1198/jasa.2010.ap07636

Tang BY, Hsieh WW, Monahan AH, Tangang FT (2000) Skill comparisons between neural networks and canonical correlation analysis in predicting the equatorial Pacific sea surface temperatures. J Climate 13:287–293

Tangang FT, Tang B, Monahan AH, Hsieh WW (1998) Forecasting ENSO events: A neural network-extended EOF approach. J Climate 11:29–41

Timmermann A, Voss HU, Pasmanter R (2001) Empirical dynamic system modeling of ENSO using nonlinear inverse techniques. J Phys Oceanogr 31:1579–1598

Tong H (1990) Non-linear time series: a dynamical systems approach. Oxford University Press, Oxford

Villagran A, Huerta G, Jackson CS, Sen MK (2008) Computational methods for parameter estimation in climate models. Bayesian Anal 3:823–850

Vrugt JA, ter Braak CJF, Dilks CGH, Robinson BA, Hyman JM, Higdon D (2009) Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. Int J Nonlinear Sci Numer Simul 10(3):273–290

Wakefield JC, Smith AFM, Racine-Poon A, Gelfand AE (1994) Bayesian analysis of linear and non-linear population models by using the Gibbs sampler. Appl Stat 43(1):201–221

West M, Harrison J (1997) Bayesian forecasting and dynamic models, 2nd edn. Springer, New York

Whittle P (1954) On stationary processes in the plane. Biometrika 44:434–449

Whittle P (1962) Topographic correlation, power-law covariance functions, and diffusion. Biometrika 49:305–314

Wikle CK (1996) Spatio-temporal statistical models with applications to atmospheric processes. PhD dissertation, Iowa State University, Ames, IA

Wikle CK (2002a) A kernel-based spectral model for non-Gaussian spatial processes. Stat Model, Int J 2:299–314

Wikle CK (2002b) Spatial modeling of count data: a case study in modelling breeding bird survey data on large spatial domains. In Lawson A, Denison D (eds) Spatial cluster modelling. Chapman & Hall, London, pp 199–209

Wikle CK (2003) Hierarchical Bayesian models for predicting the spread of ecological processes. Ecology 84:1382–1394

Wikle CK, Berliner LM (2005) Combining information across spatial scales. Technometrics 47:80–91

Wikle CK, Cressie N (1999) A dimension-reduced approach to space-time Kalman Filtering. Biometrika 86:815–829

Wikle CK, Hooten MB (2006) Hierarchical Bayesian spatio-temporal models for population spread. In: lark JS, Gelfand A (eds) Applications of computational statistics in the environmental sciences: hierarchical Bayes and MCMC methods. Oxford University Press, Oxford, pp 145–169

Wikle CK, Berliner LM, Cressie N (1998) Hierarchical Bayesian space-time models. Environ Ecol Stat 5:117–154

Wikle CK, Milliff R, Nychka D, Berliner LM (2001) Spatiotemporal hierarchical Bayesian modeling: tropical ocean surface winds. J Am Stat Assoc 96:382–397

Wikle CK, Berliner LM, Milliff RF (2003) Hierarchical Bayesian approach to boundary value problems with stochastic boundary conditions. Mon Weather Rev 131:1051–1062

Xu K, Wikle CK (2007) Estimation of parameterized spatio-temporal dynamic models. J Stat Plan Inference 137:567–588

Xu K, Wikle CK, Fox NI (2005) A kernel-based spatio-temporal dynamical model for nowcasting radar precipitation. J Am Stat Assoc 100(472):1133–1144

Young P (2000) Stochastic, dynamic modelling and signal processing: time variable and state dependent parameter estimation. In: Fitzgerald WJ, Smith RL, Walden AT, Young P (eds) Nonlinear and non-stationary signal processing. Cambridge University Press, Cambridge, pp 74–114

Yule GU (1927) On a method of investigating periodicities in disturbed series, with reference to Wolfer's sunspot numbers. Philos Trans R Soc Lond, Ser A 226:267–298

Zastavnyi VP, Porcu E (2009) Space-time covariance functions with compact support. arXiv:0902.3656v1