

Long-Lead Prediction of Pacific SSTs via Bayesian Dynamic Modeling

L. MARK BERLINER

Department of Statistics, The Ohio State University, Columbus, Ohio

CHRISTOPHER K. WIKLE

Department of Statistics, University of Missouri, Columbia, Missouri

NOEL CRESSIE

Department of Statistics, The Ohio State University, Columbus, Ohio

(Manuscript received 1 November 1999, in final form 7 February 2000)

ABSTRACT

Tropical Pacific sea surface temperatures (SSTs) and the accompanying El Niño–Southern Oscillation phenomenon are recognized as significant components of climate behavior. The atmospheric and oceanic processes involved display highly complicated variability over both space and time. Researchers have applied both physically derived modeling and statistical approaches to develop long-lead predictions of tropical Pacific SSTs. The comparative successes of these two approaches are a subject of substantial inquiry and some controversy. Presented in this article is a new procedure for long-lead forecasting of tropical Pacific SST fields that expresses qualitative aspects of scientific paradigms for SST dynamics in a statistical manner. Through this combining of substantial physical understanding and statistical modeling and learning, this procedure acquires considerable predictive skill. Specifically, a Markov model, applied to a low-order (empirical orthogonal function–based) dynamical system of tropical Pacific SST, with stochastic regime transition, is considered. The approach accounts explicitly for uncertainty in the formulation of the model, which leads to realistic error bounds on forecasts. The methodology that makes this possible is hierarchical Bayesian dynamical modeling.

1. Introduction

The interannual variation of tropical Pacific sea surface temperature (SST) is an important factor in the variability of the global climate system. The dominant feature of this field is the episodic warming and cooling of ocean waters with periods of approximately 3–5 yr, namely, the El Niño–Southern Oscillation (ENSO) phenomenon. In recent years, long-lead predictions of tropical Pacific SSTs have improved greatly in light of better observational networks, analysis schemes, and understanding of the processes that govern the interaction of the atmosphere and ocean. Although statistical methods for SST prediction have performed as well as or better than deterministic, physically derived dynamical methods (Barnston et al. 1999), there is a common perception in the climate community that much of the potential of statistical models has been exhausted. Of course, such

suggestions refer to a collection of particular forms of statistical models, some of which are reviewed in section 3. Our intention in this article is to introduce alternative, Bayesian statistical formulations and prediction procedures that express a variety of *qualitative* notions and principles present in the literature regarding SSTs and their evolution. The Bayesian approach also allows for the quantification of uncertainty related to our physical understanding and for its stochastic representation.

Our model is “statistical” in that, although it is guided by a qualitative expression of the physics, no formal physical model is included. This is not a limitation of the Bayesian approach. See Royle et al. (1998) and Wikle et al. (1999, manuscript submitted to *J. Amer. Stat. Assoc.*) for examples of Bayesian analyses relying on physical models. Indeed, the Bayesian statistical view does not recognize a strict dichotomy between statistical and deterministic approaches. Though beyond the scope of this article, the Bayesian viewpoint offers opportunities to incorporate deterministic models into statistical, long-lead SST forecasting in a fashion that accounts for a variety of uncertainties. Furthermore, Bayesian prediction results and associated uncertainties are efficient inputs to decision making and forecasting impacts of

Corresponding author address: L. Mark Berliner, Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH 43210-1247.
E-mail: mb@stat.ohio-state.edu

SST behavior. (See Berger 1985 and Bernardo and Smith 1994 for general discussion.)

Many of the long-lead statistical prediction schemes for SSTs that have been used to date have focused on relatively simple linear models (Barnston et al. 1999). The statistical sciences have undergone a dramatic change in recent years as computationally based nonlinear methodologies have been developed and applied in complex settings. In particular, the hierarchical Bayesian statistical paradigm has benefited from this revolution. The tropical Pacific SST prediction problem provides an exceptional context in which to demonstrate the long-lead predictive power of, and associated quantification of uncertainties possible with, hierarchical Bayesian dynamical modeling.

Our suggestion to use the Bayesian approach in statistical modeling and prediction is not new for the climate sciences (e.g., Epstein 1985; Tarantola 1987). Also see Hasselmann (1998), Leroy (1998), and Berliner et al. (2000) for applications of Bayesian analysis in climate change studies. Indeed, many familiar statistical procedures actually are Bayesian procedures, or nearly so. Examples include optimal interpolation or kriging (e.g., Kitanidis 1986; Omre 1987; Cressie 1993, section 3.4.4), and some data assimilation procedures such as Kalman filtering (e.g., Meinhold and Singpurwalla 1983; Lorenc 1986; Courtier 1997). It is our intention in this article to establish a more general Bayesian statistical paradigm that we believe will enrich statistical practice in the climate sciences.

To illustrate the Bayesian approach and some of the flexibility associated with it, we develop a Bayesian space–time model for forecasting monthly tropical Pacific SSTs at a 7-month lead time. This lead time was chosen to demonstrate how the methodology could be applied to produce operational forecasts at least 6 months in advance, with the consideration that time is required to acquire new data for the new forecast. The methodology can be readily adapted to different lead times. Keys to the model include incorporation of the following features.

- For each time (month) we consider a spectral model for the data, focusing on a reduced empirical orthogonal function (EOF) basis set.
- We assume that the spectral coefficients of the model are stochastic and time-varying. That is, they are assumed to follow a multivariate time series model.
- The parameters of that time series model are themselves time-varying, yielding a methodology that is inherently nonlinear. Models, reflecting warm, cold, and normal regimes, are considered.
- Prognostic variables that indicate possible future transitions among regimes are modeled as random, with probabilities that depend upon the behavior of surface-wind anomalies in the western Pacific, which is a qualitative expression of physical processes associated with tropical Pacific SSTs.

We include uncertainties in the development of the model and present SST predictions in light of this uncertainty.

Section 2 gives a brief discussion of Bayesian analysis in science. General notions related to stochastic modeling of SST anomalies are presented in section 3. The Bayesian model for SSTs is described in section 4. The presentation focuses on our statistical modeling rather than the mathematics of Bayesian formalism and computation. Some of the more technical aspects of the model are described in the appendix. Prediction results are given in section 5 and a brief discussion concludes the paper in section 6.

2. Bayesian analysis

a. Basic notions

Our presentation is brief and pragmatic. For more extensive discussion, see Bernardo and Smith (1994) and Berger (1985). For discussions targeted to the climate sciences, see Epstein (1985), Berliner et al. (1998), and Wikle et al. (1998, 1999, manuscript submitted to *J. Amer. Stat. Assoc.*).

One way to explain the Bayesian viewpoint is to present the approach as a natural solution to prediction problems: the goal is to produce a probability distribution for future values of some process conditional on past observational data and our knowledge of the process. Implicit in this view is a strong reliance on statistical modeling and probability calculus. While probability calculus provides mechanisms for computation, the basis of the Bayesian statistical approach is that all unknown quantities of interest are modeled as random variables. That is, probability models are used to represent and quantify uncertainty, even regarding unknown constants or deterministic quantities.

Let Z represent observable data and Y represent unknown quantities. We first formulate a statistical probability model $p(z|y)$ for the data; this is the conditional distribution of Z given Y . Note that this step is common to both traditional and Bayesian formulations. Next, we formulate a *prior* (to observing the data) probability distribution for Y , denoted by $\pi(y)$. The key to Bayesian learning-from-data is Bayes' theorem, which provides a formula for computing the conditional probability distribution of unknowns given the observed data. Specifically, the *posterior* (after observing that $Z = z$) distribution for Y is given by

$$\pi(y|z) \propto p(z|y)\pi(y), \quad (1)$$

where \propto denotes "is proportional to." The proportionality constant in this formula is the constant (function of z , but not y) that ensures that $\pi(y|z)$ integrates to 1.

Bayesian inference relies on the posterior distribution obtained in (1). Typically, selected summaries (e.g., percentiles, means, and variances) of the posterior are reported. This is especially necessary when the target of

our inference, Y , is very high-dimensional, as in the SST problem.

b. Hierarchical models

From (1), we can develop highly complex, space–time dynamic models based on various results from probability theory (e.g., Wikle et al. 1998; Wikle and Cressie 1999). A particularly valuable modeling strategy is the hierarchical framework. In this approach, a complete joint probability model [the right-hand side of (1)] for all data and unknowns is written as the product of a series of conditional models. The probability theory is easy to describe. Suppose that instead of a single variable Y in (1), we wish to model three random variables: Y_1 , Y_2 , and Y_3 . Their joint prior probability distribution, denoted by $\pi(y_1, y_2, y_3)$, may be complicated and difficult to formulate directly. However, probability theory says that the joint prior can be written as a product:

$$\pi(y_1, y_2, y_3) = \pi_{1|2,3}(y_1|y_2, y_3)\pi_{2|3}(y_2|y_3)\pi_3(y_3), \quad (2)$$

where, for example, $\pi_{1|2,3}$ is a conditional density for Y_1 , given Y_2 and Y_3 . (Analogous expressions exist for an arbitrary number of variables.) The essence of this method is that specification of the components of the product in (2) may be simpler and more readily guided by our science and experience, compared to the direct formulation of the joint prior.

The essence of hierarchical thinking in the context of space–time processes is contained in the following representation. First, we envision three basic collections of variables to be modeled:

- data: our observations,
- process: those physical, state variables of interest, and
- parameters: unknown physical constants and parameters introduced in the statistical components of the model.

To correspond to these three collections of variables, hierarchical thinking suggests a model with three primary components (e.g., Berliner 1996):

- data model: $p(\text{data}|\text{process}, \text{parameters})$,
- prior process model: $\pi(\text{process}|\text{parameters})$, and
- prior on parameters: $\pi(\text{parameters})$.

As indicated in (2), the product of these three components is a joint probability distribution. Once the modeling process is completed, Bayes' theorem yields the posterior distribution, say $\pi(\text{process}, \text{parameters}|\text{data})$. The distribution, $p(\text{data}|\text{process}, \text{parameters})$, is typically a measurement error model. Formulation of this component is usually critical in both Bayesian and non-Bayesian statistical approaches. Formulations of $\pi(\text{process}|\text{parameters})$ and $\pi(\text{parameters})$ are cornerstones of the Bayesian approach. Here, the modeler uses scientific knowledge of the process whenever possible.

Uncertainties in one's knowledge are quantified probabilistically.

Of course, construction of the component models above may be challenging. However, substantial flexibility in modeling is achieved, particularly in very high dimensions. Often, the component models are quite simple, but simplicity in the components does not mean that the final model is simple. Indeed, the models that arise can capture very complex, nonlinear, nonstationary space–time structure.

A second issue arising in the use of Bayesian hierarchical models is that historically they have been too complicated to implement (i.e., to compute the posterior) without major simplification. However, large-scale hierarchical modeling has recently become operationally feasible due to the popularization of numerically intensive simulation approaches, such as Markov chain Monte Carlo (MCMC); see, for example, Gilks et al. (1996). Recent examples related to atmospheric science, where MCMC was used, include Royle et al. (1998) and Wikle et al. (1998; 1999, manuscript submitted to *J. Amer. Stat. Assoc.*).

3. Pacific SSTs and ENSO

The primary dataset is a compilation of monthly averaged SST analysis fields obtained from the International Research Institute for Climate Prediction, Lamont-Doherty Earth Observatory data library Web site (<http://ingrid.ldeo.columbia.edu/SOURCES/.CAC/.sst/>). Specifically, we used the Climate Analysis Center (CAC) monthly SST anomalies over the tropical Pacific region 29°S–29°N latitude and 124°E–70°W longitude, at a 2° × 2° spatial resolution for the period January 1970–July 1999. This dataset combines the 2° × 2° CAC dataset from January 1970 through October 1981 with a regrided (2° × 2°) version of the current 1° × 1° interpolation product offered by the Climate Prediction Center for the period November 1981–July 1999. Anomalies were based on the climatology from the period January 1970–December 1985. Let \mathbf{Z} , be an m -dimensional vector denoting the vectorized spatial field *anomaly* at $m = 2261$ oceanic locations for month t ; $t = 1, \dots, T$, where $T = 343$.

We will also make use of zonal wind components at a height of 10 m above the surface. Specifically, we considered zonal wind anomalies from the National Centers for Environmental Prediction–National Center for Atmospheric Research reanalysis dataset (Kálnay et al. 1996). The data were regrided to the same spatial domain as described for the SST dataset. Similarly, the temporal coverage and anomaly period for the wind data is the same as that for the SST dataset described above.

a. Markovian modeling

Many authors have provided evidence that ENSO can be considered as a system forced by relatively high-

frequency atmospheric transients (e.g., Lau 1985; Lau and Chan 1986, 1988; Vallis 1988; Zebiak 1989; Kleeman and Moore 1997; Blanke et al. 1997; Eckert and Latif 1997; Moore and Kleeman 1999). In particular, Penland and Magorian (1993), Penland and Sardeshmukh (1995), and Penland (1996) make a strong case that the projection of tropical Pacific SST anomalies onto their leading EOFs can be modeled reasonably as a first-order Markov process with Gaussian spatial noise. The justification for this linear model is that since the nonlinear processes and forcing associated with SST have much shorter timescales than the SST evolution we are interested in, the nonlinear terms can be absorbed into the spatial noise term. Penland and collaborators also demonstrated that moderately successful SST anomaly forecasts are possible with this Markovian approach. However, as is often the case with climatological implementation of such stochastic methods, estimation issues have only rarely been considered in the climate literature, and little attempt has been made to quantify all sources of uncertainty. The hierarchical Bayesian approach presented here addresses some of these issues. Furthermore, as stated previously, it also allows for the inclusion of time-varying dynamics, through an inhomogeneous Markovian process model.

The general idea for forecasting Pacific SST anomalies using these notions is via stochastic evolution models of the form

$$\mathbf{A}_{t+\tau} = \mathbf{H}\mathbf{A}_t + \boldsymbol{\eta}_{t+\tau}, \quad (3)$$

where $\mathbf{A}_t \equiv (A_t^{(1)}, \dots, A_t^{(K)})'$ is the vector of length K , that results from projecting the SST anomalies onto the first K EOFs (i.e., $\mathbf{Z}_t \approx \boldsymbol{\Phi}\mathbf{A}_t$, where $\boldsymbol{\Phi}$ is the $m \times K$ EOF matrix), \mathbf{H} is a $K \times K$ propagator matrix of regression parameters, and $\boldsymbol{\eta}_{t+\tau}$ is a K vector of innovations. For the results presented here, we choose $K = 10$; the first 10 EOFs explain approximately 74% of the variance in the SST anomalies in our dataset. The innovation vectors $\{\boldsymbol{\eta}_t\}$ are assumed to be mutually uncorrelated over time and typically are specified to have Gaussian distributions,

$$\boldsymbol{\eta}_{t+\tau} \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \quad \text{for all } t = 1, \dots, T,$$

where \sim is to be read as “is distributed as,” and $\boldsymbol{\Sigma}_\eta$ is a $K \times K$ covariance matrix. Notice the Markovian nature of (3) in that the model for $\mathbf{A}_{t+\tau}$ depends on the present value \mathbf{A}_t and the past values \mathbf{A}_{t-1} , \mathbf{A}_{t-2} , \dots , only through the present, \mathbf{A}_t .

Such models can be motivated by linearizations of underlying dynamics. In that view, linearization of a particular physical model would lead to propagator matrices \mathbf{H} that might depend on τ and \mathbf{A}_t . Alternatively, we could suggest that the model is “statistical” and replace \mathbf{H} in (3) with \mathbf{H}_t . To proceed in this fashion, the \mathbf{H}_t must be parameterized or restricted in some way, since otherwise there are insufficient data to estimate unrestricted $\{\mathbf{H}_t\}$. Specifically, for each time t there

would be a single data vector of length K available to estimate the K^2 elements of \mathbf{H}_t .

b. Wind bursts

As stated previously, one dominant paradigm suggests that the stochastic forcing of SSTs at ENSO timescales is, through the coupling of the atmosphere and ocean, related to the spatial distribution of wind stress. Furthermore, so-called wind bursts, associated with the intraseasonal oscillation, are increasingly thought to be an important factor in the onset and development of ENSO events (e.g., McPhaden and Yu 1999; Moore and Kleeman 1999; Nakazawa 1998). Based on a linear regression between 7-month-lagged 10-m wind anomalies (low-pass filtered to include periods greater than about 24 months) and the similarly filtered Southern Oscillation index (SOI; the standardized sea level pressure anomalies at Tahiti minus those at Darwin), the spatial region with the highest coefficient of determination (R^2 value) is centered around 5°N and 157°E . Thus, we will make use of a wind index at time t , based on the spatial average of 10-m wind anomalies at 5°N and 156° – 158°E , to help predict tropical Pacific SSTs at time $t + \tau$. Although it has a relatively strong lagged association with the filtered SOI ($R^2 \approx 0.7$, with negative association), we do not claim that this statistic is directly representative of wind bursts. However, our search for such a predictor was motivated by these phenomena. Furthermore, it is plausible that if wind bursts in the western Pacific at the intraseasonal timescale are associated with ENSO behavior τ months in the future, then a low-pass filtered measure of this activity would help predict the future ENSO regime. A more detailed discussion is given in section 4d below.

c. Regime switching: Preliminary data analysis

We develop a relatively parsimonious parameterization for time-varying propagator matrices, \mathbf{H}_t . As a simple motivation, consider the plot of the leading spectral coefficients $A_{t+\tau}^{(1)}$ and $A_t^{(1)}$ shown in Fig. 1 as it relates to a non time-varying evolution model like that in Eq. (3). Though such a model includes more dimensions than depicted in this plot, rough intuition suggests that it corresponds to a single regression-through-the-origin model. Such a least squares line for regressing $A_{t+\tau}^{(1)}$ on $A_t^{(1)}$ is also shown in Fig. 1. Our motivation is to expand this sort of regression model by 1) focusing on the current state’s regime, as suggested by $A_t^{(1)}$ (or, nearly equivalently, by the SOI at time t) being in a warm, normal, or cold regime, and 2) anticipating the future regime. In particular, Fig. 2 presents a plot for the same data, but we have stratified the points by both the current and future regime, as well as the wind statistic discussed above. We also included nine regression models (these are no longer required to be regressions through the origin). The idea is that we may gain significant pre-

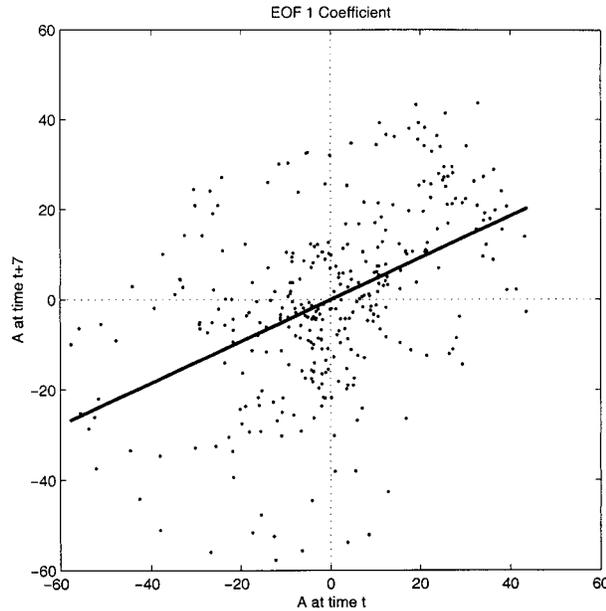


Fig. 1. Scatterplot of the first SST anomaly EOF coefficient, $A_{t+7}^{(1)}$ vs $A_t^{(1)}$. The ordinary least squares fitted regression line is also shown.

dictive power through such extra stratification. [See Zwiers and von Storch (1990) for related discussion in the context of the Southern Oscillation.] Much of the intense statistical modeling in section 4 is a probabilistic formalization of this intuition. How we achieve all of this may not be clear yet, but our goal is easy to illustrate. If we know that the current regime is “normal,” a propagator that takes the current vector of spectral coefficients to another normal vector in τ months should have a different structure than one that takes the current vector into a “warm” (or “cold”) one. Note that Fig. 2 suggests that there is not much linear association between the first spectral component at time t and $t + \tau$ ($\tau = 7$ months) when the current regime is warm. Our model, described below, considers additional spectral components that will allow us to make reasonable predictions from this regime.

4. A Bayesian model for SST prediction

a. Overview of the model

For each t ; $t = 1, \dots, T$, we will model \mathbf{Z}_t as

$$\mathbf{Z}_t = \Phi \mathbf{a}_t + \boldsymbol{\nu}_t, \quad (4)$$

where Φ is an $m \times K$ matrix of EOFs and \mathbf{a}_t is a $K \times 1$ vector of coefficients. We use lower case notation to emphasize that the modeled \mathbf{a}_t 's need not coincide with the original \mathbf{A}_t 's determined by the EOF decomposition. We describe statistical models for the errors $\{\boldsymbol{\nu}_t\}$ below.

Next, we shall formulate a statistical model for the dynamics of $\{\mathbf{a}_t\}$. This model is tailored for making predictions with a τ -month lead and is motivated by the plots shown in Fig. 2. It is a time-varying linear model,

$$\mathbf{a}_{t+\tau} = \boldsymbol{\mu}_t + \mathbf{H}_t \mathbf{a}_t + \boldsymbol{\eta}_{t+\tau}, \quad (5)$$

where $\boldsymbol{\mu}_t$ are K -vector model intercepts, \mathbf{H}_t are $K \times K$ propagator matrices, and $\boldsymbol{\eta}_{t+\tau}$ is statistically independent of \mathbf{a}_t . Models for the innovations $\{\boldsymbol{\eta}_{t+\tau}\}$ will be discussed below.

The $\boldsymbol{\mu}_t$ and \mathbf{H}_t are modeled as random quantities whose distributions are based on qualitative features of SST evolution. Specifically, these models will condition on observable quantities that serve as indicators of regime and prognosticators of regime switching (e.g., a wind statistic). Similar models were used in Lu and Berliner (1999) in a different context. We consider three regimes corresponding to normal (“normal”), warmer-than-normal (“warm”), and cooler-than-normal (“cold”) SST anomalies. Note that, while motivated by ENSO, we do not claim that these regimes are to be interpreted as “warm is equivalent to El Niño” or “cold is equivalent to La Niña.”

Our modeling introduces various other variables, model parameters, and their associated distributions. These components lead to a legitimate statistical model for all random quantities. Once the modeling work is complete, we apply Bayes’ theorem (numerically, using MCMC) to update all distributions based on the data and provide a predictive distribution for $\mathbf{Z}_{T+\tau}$, conditional on the data $\mathbf{Z}_1, \dots, \mathbf{Z}_T$ and the prognostic variables.

b. Data model

The data model is given in (4). The error vectors $\{\boldsymbol{\nu}_t\}$ are included to account for losses of information due to the dimension reduction and other model errors. Such a model is similar to that proposed by Wikle and Cressie (1999). We assume that these error vectors are independent of each other in time and have a common multivariate Gaussian distribution with mean $\mathbf{0}$ and $m \times m$ covariance matrix $\boldsymbol{\Sigma}_\nu$. Symbolically, we write

$$\boldsymbol{\nu}_t \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_\nu),$$

for each month t .

Equivalently, the data model is that the \mathbf{Z}_t are conditionally independent over time and, for each t ,

$$\mathbf{Z}_t | \mathbf{a}_t, \boldsymbol{\Sigma}_\nu \sim \text{Gau}(\mathbf{0} \mathbf{a}_t, \boldsymbol{\Sigma}_\nu). \quad (6)$$

The conditional independence of the $\{\mathbf{Z}_t\}$ does not mean that they are unconditionally independent; if that were the case, there would be no value in studying past $\{\mathbf{Z}_t\}$ to predict future ones. It does mean that in computing the density of the entire dataset, conditional on all unknowns, we simply take the product of the Gaussian density functions dictated by (6).

c. Process model: Stochastic dynamics of the \mathbf{a}_t

We model the $\boldsymbol{\mu}_t$ and \mathbf{H}_t in Eq. (5) as “regime dependent,” using definitions of regimes that include their

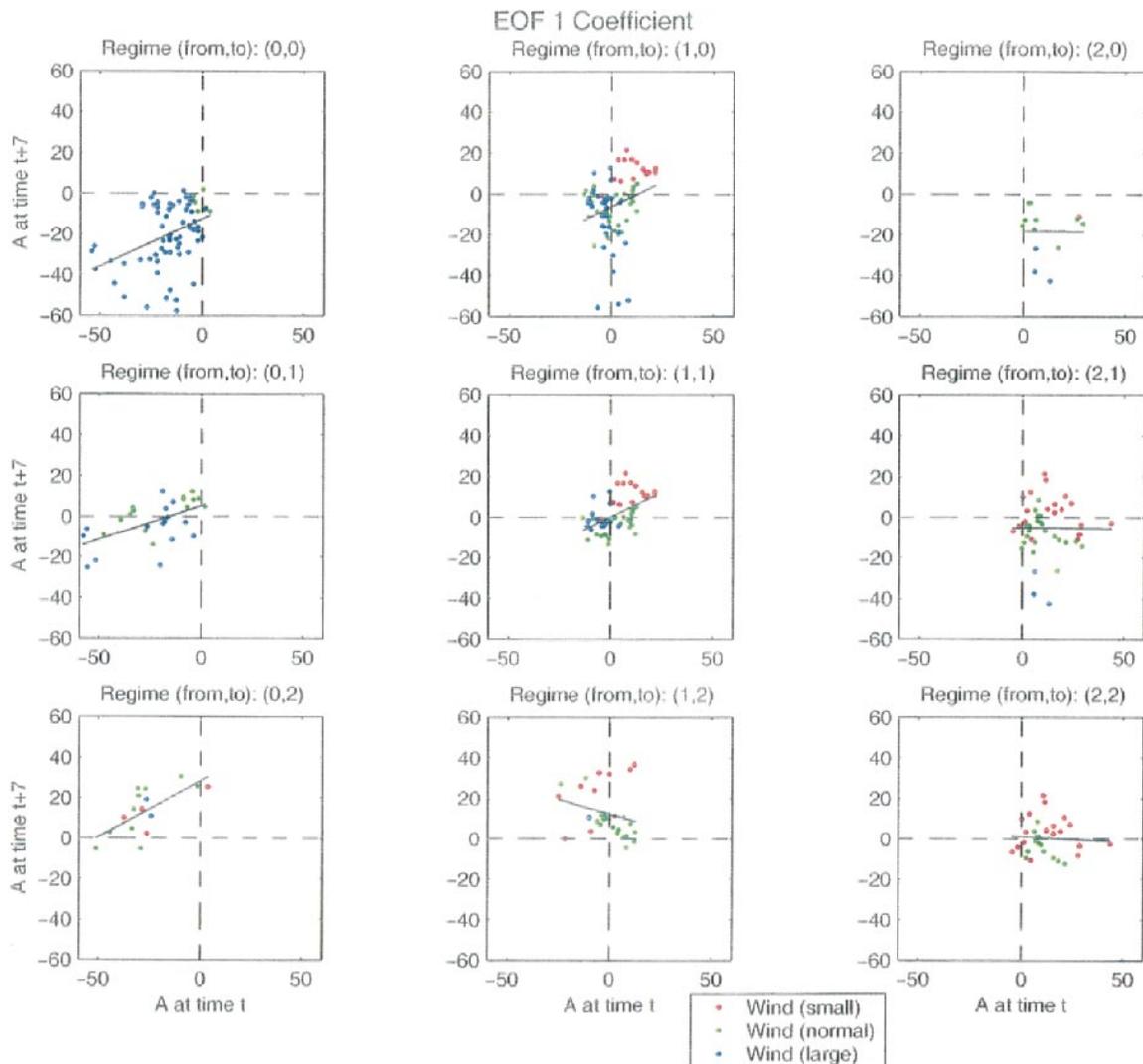


FIG. 2. Same as Fig. 1 except that nine separate regressions are shown, depending on current regime at time t and future regime at time $t + 7$. Note that the regimes are defined such that regime 2 (warm) occurs when the filtered SOI at time t is less than the 33.3d percentile of all filtered SOI values; regime 0 (cold) occurs when the SOI values are greater than the 66.7th percentile; and regime 1 (normal) occurs otherwise. The points are color coded according to the value of the wind statistic at time t , described in section 3b. Classification of wind as small, normal, or large is also relative to the 33.3d and 66.7th percentiles.

evolution in time. We construct two sequences of regime indicators. The first, denoted by I_t , classifies the current state of the system as “warm,” “normal,” or “cold.” The second, denoted by J_t , summarizes selected current information in a fashion that is meant to suggest or anticipate a transition to one of the three regimes. Once these objects are defined, the propagator matrices are formulated as follows:

$$\mathbf{H}_t = \mathbf{H}(I_t, J_t). \quad (7)$$

That is, we will have nine models, as suggested in section 3c. It is worth reiterating that such modeling is precisely one of the strategies enabled by the Bayesian framework. In our case, the reader might object to the

model since one would not know J_t . But the point is that we can develop a model for the evolution of vectors $\{\mathbf{a}_t\}$ that is conditional on the I_t and J_t , and then develop separate models for the I_t and J_t . Once accomplished, probability theory dictates the appropriate predictive analysis.

Note that, in developing a regime-based model, we have increased the number of parameters that must be estimated. Using the time-invariant model (3), one must estimate K^2 elements of \mathbf{H} . To use our nine-model approach, each one being of the sort given by (5), we would need to estimate $9 \times (K + K^2)$ (e.g., =990 for $K = 10$) regression parameters; this is formidable. To develop a comparatively parsimonious evolution model, we break up the vector \mathbf{a}_t , as follows:

$$\mathbf{a}_t = \begin{pmatrix} \mathbf{a}_t^{(l)} \\ \mathbf{a}_t^{(s)} \end{pmatrix},$$

where $\mathbf{a}_t^{(l)}$ is an n_l vector of the large-scale components (corresponding to the first n_l EOFs) and $\mathbf{a}_t^{(s)}$ comprises the remaining $n_s = K - n_l$ smaller-scale components. We consider a stochastic evolution model represented as

$$\begin{pmatrix} \mathbf{a}_{t+\tau}^{(l)} \\ \mathbf{a}_{t+\tau}^{(s)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\delta}(I_t, J_t) \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{G}(I_t, J_t) & \mathbf{G}^{ls} \\ \mathbf{G}^{sl} & \mathbf{G}^{ss} \end{pmatrix} \begin{pmatrix} \mathbf{a}_t^{(l)} \\ \mathbf{a}_t^{(s)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta}_{t+\tau}^{(l)} \\ \boldsymbol{\eta}_{t+\tau}^{(s)} \end{pmatrix}, \quad (8)$$

where the $\boldsymbol{\eta}_t$ are mutually independent, Gaussian random variables:

$$\begin{pmatrix} \boldsymbol{\eta}_{t+\tau}^{(l)} \\ \boldsymbol{\eta}_{t+\tau}^{(s)} \end{pmatrix} \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \quad (9)$$

and we write

$$\boldsymbol{\Sigma}_\eta = \begin{pmatrix} \boldsymbol{\Sigma}_\eta^{ll} & \boldsymbol{\Sigma}_\eta^{ls} \\ (\boldsymbol{\Sigma}_\eta^{ls})' & \boldsymbol{\Sigma}_\eta^{ss} \end{pmatrix}. \quad (10)$$

We form models for these covariance matrices later. Also, in (8), define $\boldsymbol{\mu}(I_t, J_t)$ and $\mathbf{H}(I_t, J_t)$ as

$$\boldsymbol{\mu}(I_t, J_t) = \begin{pmatrix} \boldsymbol{\delta}(I_t, J_t) \\ \mathbf{0} \end{pmatrix} \quad \text{and} \quad (11)$$

$$\mathbf{H}(I_t, J_t) = \begin{pmatrix} \mathbf{G}(I_t, J_t) & \mathbf{G}^s \\ \mathbf{G}^l & \mathbf{G}^{ss} \end{pmatrix}. \quad (12)$$

The critical point is that the only time-varying parameters are those at large scales [$\boldsymbol{\delta}(I_t, J_t)$ and $\mathbf{G}(I_t, J_t)$]. With this reduction, we have [$9 \times (n_l + n_s^2)$] + [$2 \times (n_l \times n_s)$] + n_s^2 regression parameters to estimate. For $n_l = 4$ and $n_s = 6$, there are 264 regression parameters, a large reduction from the 990 parameters that must be estimated if we allow all scales to be regime dependent.

Note that (8) provides distributions for $\mathbf{a}_{\tau+1}$, given \mathbf{a}_1 , $\mathbf{a}_{\tau+2}$ given \mathbf{a}_2 , etc. To complete the stochastic model, we must initialize it by specifying prior distributions on $\mathbf{a}_1, \dots, \mathbf{a}_\tau$. A discussion of these prior distributions can be found in the appendix.

The formulations in (8)–(12) are not unique, nor do we suggest them to be the “best.” The restriction to nine models is not critical; we made that choice for convenience and by analogy to ENSO. One of the advantages of the hierarchical Bayesian framework is that the specific components of the model are subject to debate and refinement. Other model developers might prefer more refined classification of states and/or alternative parameterizations of propagators. While we believe the simple model here is actually quite good, our primary message is the strategy.

d. Regime evolution

Next, we specify models for the two collections of regime indicators:

$$\{I_t : 1 \leq t \leq T\} \quad \text{and} \quad \{J_t : 1 \leq t \leq T\}.$$

The I_t are to classify the *current* state (2 = warm, 1 = normal, and 0 = cold). In general, the model for I_t could be a stochastic one; this would be particularly important if the modeler intended the I_t to be physically meaningful indicators of ENSO. A simpler strategy is adopted here. We use the SOI, after application of a low-pass filter (a recursive Butterworth filter with cutoff at around 24 months; see Rabiner and Gold 1975), to serve as a simple summary of the current state. (We let SOI_t denote this filtered SOI.) We formulate a threshold model (Tong 1990); that is, we define I_t as follows:

$$I_t = \begin{cases} 0, & \text{if } \text{SOI}_t > s_t^+, \\ 1, & \text{if } s_t^- \leq \text{SOI}_t \leq s_t^+, \quad \text{and} \\ 2, & \text{if } \text{SOI}_t < s_t^-, \end{cases} \quad (13)$$

$$I_t = \begin{cases} 0, & \text{if } \text{SOI}_t > s_t^+, \\ 1, & \text{if } s_t^- \leq \text{SOI}_t \leq s_t^+, \quad \text{and} \\ 2, & \text{if } \text{SOI}_t < s_t^-, \end{cases} \quad (14)$$

$$I_t = \begin{cases} 0, & \text{if } \text{SOI}_t > s_t^+, \\ 1, & \text{if } s_t^- \leq \text{SOI}_t \leq s_t^+, \quad \text{and} \\ 2, & \text{if } \text{SOI}_t < s_t^-, \end{cases} \quad (15)$$

where the threshold values $s_t^- \leq s_t^+$ are fixed, but may be chosen to depend upon time. In particular, one might specify that the thresholds vary with month or season (e.g., Trenberth 1997). In this article we did not vary these values in time, though we leave that as an interesting possible enhancement. We discuss our choice for these parameters in the appendix.

The J_t classify the current prognosticator state (2 = warm, 1 = normal, and 0 = cold), based on 1) W_t , where W_t is the current value of the wind statistic described in section 3b, 2) time-by-wind interaction (more explicitly, we consider the sine and cosine of a month indicator multiplied by W_t), and 3) a quadratic term in the wind statistic, W_t^2 . The time-by-wind interactions are included to account partially for seasonal phase locking of ENSO regimes. This information is certainly incomplete. The state of the system in τ months will be determined by a variety of factors. Hence, we use a stochastic model for the J_t .

Note that we do not claim that $J_t = I_{t+\tau}$. Rather, J_t is simply a summary of information suggesting that conditions may be right for a particular transition. The correct view is that our predictive distribution depends on the current state I_t and is a mixture of three distributions, where the mixing probabilities are those corresponding to J_t . We will clarify this more when results are discussed in section 5.

Next, we specify the joint distribution,

$$\pi(\{J_t : 1 \leq t \leq T\} | \{W_t : 1 \leq t \leq T\}, \theta_J), \quad (16)$$

where θ_J denotes a collection of modeling parameters defined below. We assume that the J_t are mutually conditionally independent. Our model is that the joint distribution $\pi(\{J_t : 1 \leq t \leq T\} | \{W_t : 1 \leq t \leq T\}, \theta_J)$ is the product

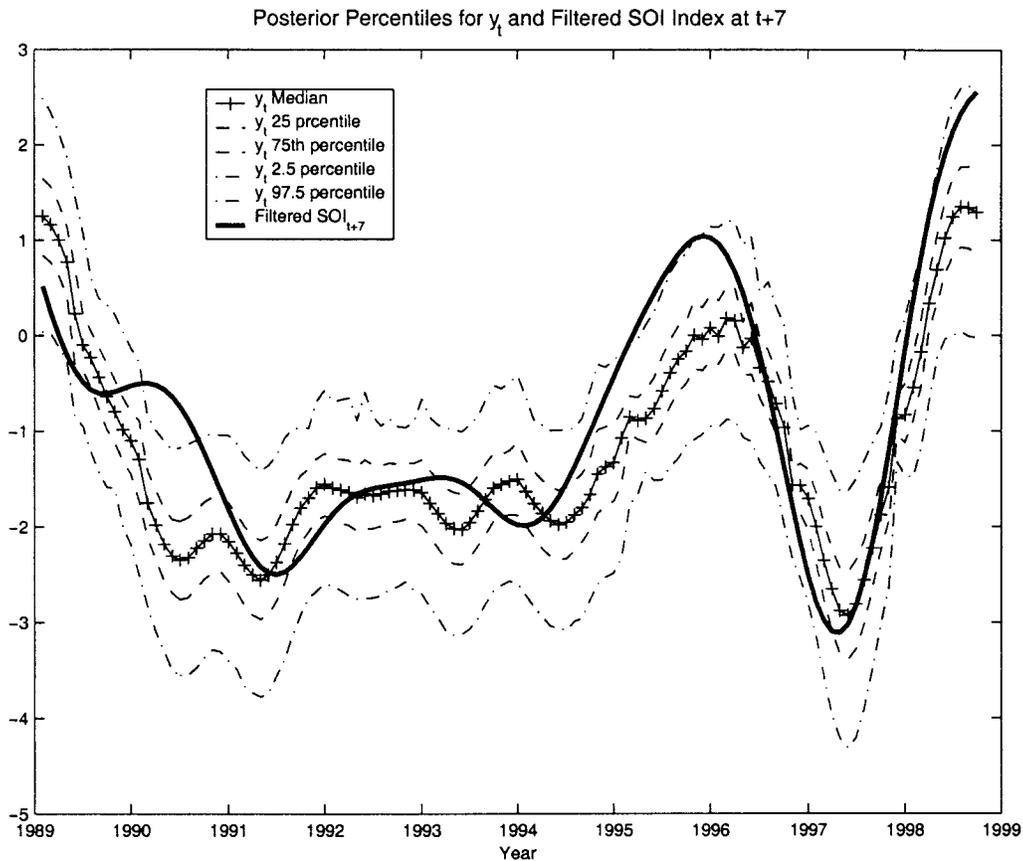


FIG. 3. Posterior percentiles of the hidden process y_t . The SOI at time $t + 7$ is plotted at time t for comparison.

$$\prod_{i=1}^T \pi(J_i | W_i, \theta_j). \quad (17)$$

We could condition on previous lags of W in (17), but preliminary data analyses suggested that these are not beneficial.

The task of building the distributions (17) requires the construction of three probabilities (for J_i being 0, 1, or 2) depending on W_i . We use a hidden or latent variable model; this variable is hidden in the sense that it is never actually observed. For an example of hidden Markov models in atmospheric science, see Hughes and Guttrop (1994). The idea is founded in the statistical notion of generalized linear models (e.g., McCullagh and Nelder 1989); implementation in Bayesian models (involving MCMC as a computational tool) was developed in Albert and Chib (1993).

The development of our hidden variable model is based on the following intuition. Though we do not insist that $J_i = I_{t+\tau}$ exactly, reasonable predictive power would result if this were typically the case. To that end, we consider a hidden random variable y_t , which determines the value of J_t as follows:

$$J_t = \begin{cases} 0, & \text{if } y_t > \gamma_t^+, \\ 1, & \text{if } \gamma_t^- \leq y_t \leq \gamma_t^+, \text{ and} \\ 2, & \text{if } y_t < \gamma_t^-, \end{cases} \quad (18)$$

$$J_t = \begin{cases} 0, & \text{if } y_t > \gamma_t^+, \\ 1, & \text{if } \gamma_t^- \leq y_t \leq \gamma_t^+, \text{ and} \\ 2, & \text{if } y_t < \gamma_t^-, \end{cases} \quad (19)$$

where $\gamma_t^- < 0$ and $\gamma_t^+ > 0$ are to be specified. The y_t and resulting J_t would be particularly useful if the y_t roughly predicted $\text{SOI}_{t+\tau}$, since the latter quantity serves as the classifier of regime $I_{t+\tau}$. With this intuition, options for the selection of the $\gamma_t^- < 0$ and $\gamma_t^+ > 0$ are similar to those regarding the thresholding values (s_t^-, s_t^+). See the appendix for a brief discussion of our choices for the examples presented here. An interesting extension is to model the (γ_t^-, γ_t^+) values as unknown (and, hence, random in the Bayesian paradigm). We will investigate this elsewhere.

Our model for the y_t is that they are conditionally independent, with distributions as follows:

$$y_t \sim \text{Gau}(\mathbf{x}'_t \beta_y, \sigma_y^2), \quad (21)$$

where \mathbf{x}_t is the vector of conditioning variables containing a 1 (for the intercept), W_t , $W_t \sin(\text{mo}_t)$, $W_t \cos(\text{mo}_t)$, and W_t^2 , where mo_t is a month indicator (i.e., $\text{mo}_t = (2\pi M_t/12)$, where M_t are month codes 1, ..., 12). We included $W_t \sin(\text{mo}_t)$ and $W_t \cos(\text{mo}_t)$, so that

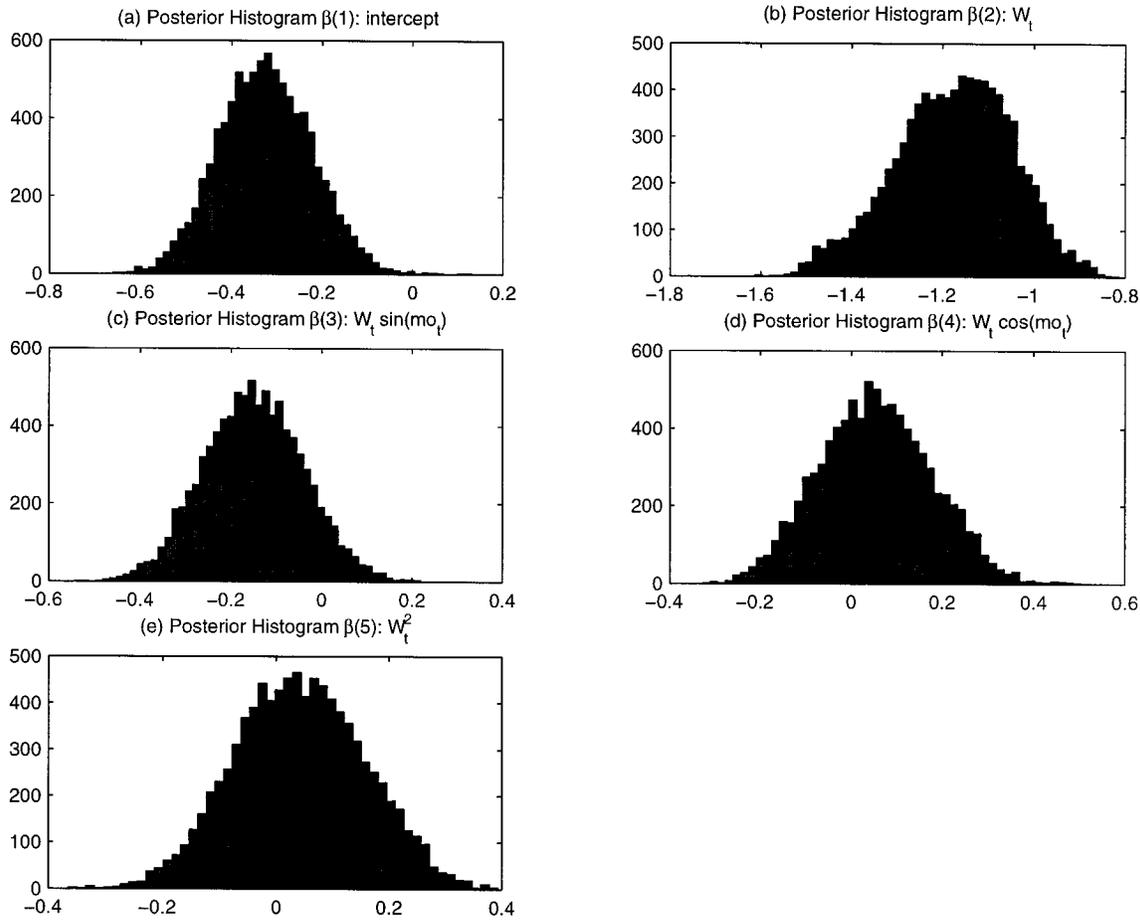


FIG. 4. Histograms of the posterior samples for the hidden-process regression parameters β_j .

the model could respond to potential seasonal dependences on predictability (e.g., Cane 1992, p. 607); and W_t^2 models potential nonlinear dependence. To clarify notation, note that $\theta_j = (\beta_j, \sigma_y^2, \{\gamma_t^-, \gamma_t^+\})$. With this definition, we have

$$\Pr(J_t = 0 | W_t, \theta_j) = \Pr(y_t > \gamma_t^+), \tag{22}$$

$$\Pr(J_t = 1 | W_t, \theta_j) = \Pr(\gamma_t^- \leq y_t \leq \gamma_t^+), \text{ and } \tag{23}$$

$$\Pr(J_t = 2 | W_t, \theta_j) = \Pr(y_t < \gamma_t^-), \tag{24}$$

where these probabilities are computed using the Gaussian distribution in Eq. (21).

e. Prior on model parameters

Completion of the Bayesian model requires specification of a prior distribution for the various parameters introduced earlier. These include the data-model covariance matrix Σ_v , all the regression coefficients defined in (8), as well as the error covariance matrix Σ_η , and the parameters θ_j . This prior specification is a critical and challenging aspect of a Bayesian analysis, both

generally and in our example. Further, our specifications in this article are by no means exhaustive, but rather serve as reasonable illustrations. Hence, to communicate best the overall modeling strategy and our results, we defer details of our prior development to the appendix. We performed some analyses (as indicated in the appendix) to gauge the sensitivity of results with respect to these prior specifications.

5. Results

a. Implementation

Our Bayesian prediction results were obtained numerically using an MCMC procedure known as the Gibbs sampler (Geman and Geman 1984; Gelfand and Smith 1990). A general overview of these methods can be found in Gilks et al. (1996). Applications of these methods to atmospheric science problems can be found in Wikle et al. (1998; 1999, manuscript submitted to *J. Amer. Stat. Assoc.*) and Royle et al. (1998). The essence of the MCMC approach is to learn about a complex joint probability distribution by simulating realizations

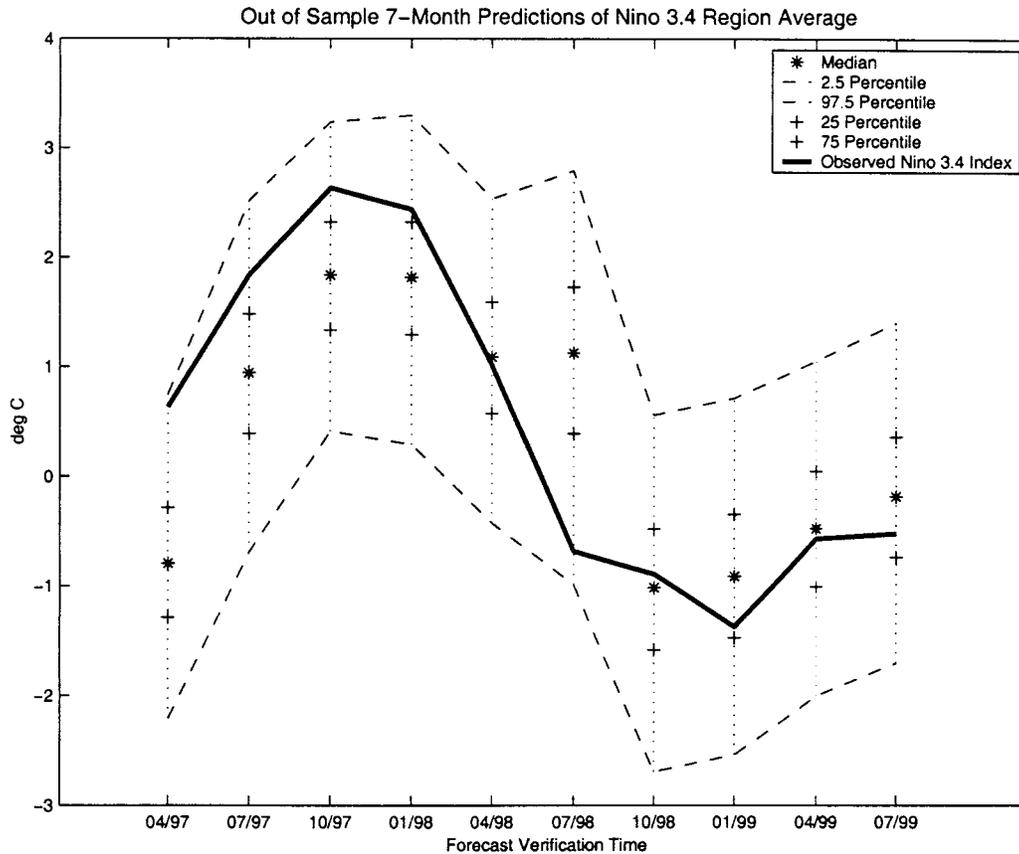


FIG. 5. Percentiles of 7-month Niño-3.4 region averages of SST-anomaly predictions. The observed value is shown for comparison. Note that each of these predictions are out-of-sample, in that only data up to seven months before the given verification time were used in developing the model and generating the prediction.

from it. This is accomplished by specifying a Markov chain with a stationary, ergodic distribution that coincides with the distribution of interest (in our context, the posterior distribution of all model parameters, given the data). Details of the specific Gibbs sampler implementation used for this article can be found at the Web site <http://www.stat.ohio-state.edu/~sses/papers.html>.

For each forecast period discussed below, we obtained 10 000 iterates from the Gibbs sampler, discarding the first 1000 to account for the Markov chain "burn-in." All reported statistics were then based on Monte Carlo estimates from these Gibbs-sampler iterations. Note that 10 000 iterates from this model can be obtained in about 11 hours on a 1999 vintage desktop workstation running MATLAB.

b. Model output

1) HIDDEN PROCESS

A key feature in the Bayesian model is the hidden regime process $\{J_t\}$, or, equivalently, $\{y_t\}$. Figure 3 shows a summary of the posterior of y_t for the last 10 years from an implementation of our model for a

7-month (i.e., $\tau = 7$) forecast based on data through December 1998. The plot shows several percentiles of the posterior distribution as functions of time. For comparison, the filtered SOI at time $t + 7$ is also plotted at time t ; recall from section 4d that we intend that y_t should roughly correspond with $\text{SOI}_{t+\tau}$. Although the posterior medians do not track the SOI exactly, central 95% prediction intervals obtained from the posterior of the y_t 's typically cover $\text{SOI}_{t+\tau}$, suggesting that the hidden-process model is reasonable. Indeed, instead of predicting SST directly, one could use y_t itself as a predictor of ENSO activity. Such a predictor would have prediction bounds similar to those shown in Fig. 3. However, for this figure, note that all values except the last seven months are actually hindcasts. Proper, out-of-sample forecasting is considered below.

It is of interest to know if the explanatory variables that were included in the model for y_t are useful. Figure 4 shows histograms for the posterior distribution of the parameters β_y . The histograms for the sine and cosine interaction components and the quadratic component of the model cover zero with nontrivial frequency, suggesting that these terms may not be critical in determining the future climate state.

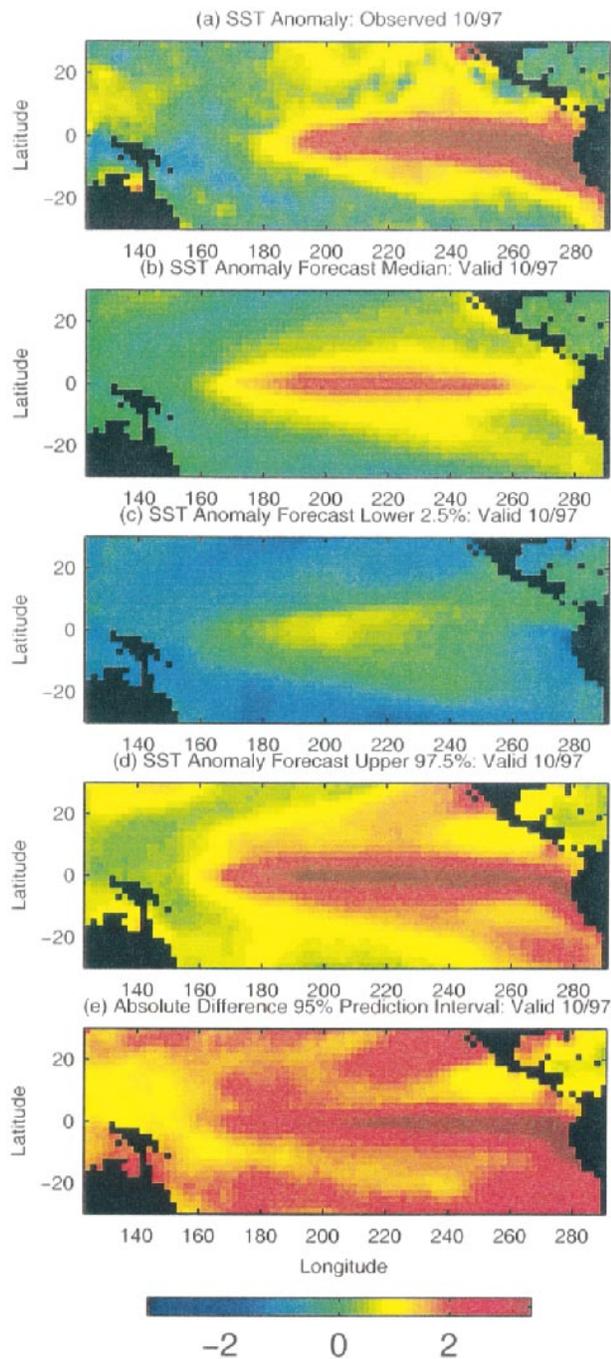


FIG. 6. SST anomalies for Oct 1997. (a) Observed anomalies, (b) posterior median 7-month out-of-sample prediction of smoothed SST anomalies from the Bayesian hierarchical model, (c) lower 2.5th percentile from the posterior distribution of the 7-month prediction, (d) upper 97.5th percentile from the posterior distribution of the 7-month prediction, and (e) length of 95% prediction interval, which has endpoints equal to the 2.5th and 97.5th percentiles.

2) LONG-LEAD PREDICTION OF NIÑO-3.4 AVERAGE

Alternative indices of ENSO activity based on SSTs can be defined as averages of SST anomalies over selected regions, such as the Niño-3.4 region (5°S – 5°N , 120° – 170°W). Figure 5 shows out-of-sample 7-month predictions of the Niño-3.4 index every three months, verifying from April 1997 through July 1999, respectively. By “out-of-sample,” we mean the results are fair in the sense that they are based on separate MCMC analyses that use only data up to seven months prior to the indicated verification time. Figure 5 presents percentiles of the posterior distribution of these predictions. For comparison, the time series of the observed Niño-3.4 index is shown for the same periods. One feature that stands out is that the observed index is within the 95% prediction intervals for each forecast period. Note that the intervals shown here are true prediction intervals in that they incorporate the variability in the predicted SST anomalies that arise from the ν process given in (4). Like virtually all predictions of the 1997–98 warm event onset, the bulk of the predicted distribution is lower in magnitude than the observed Niño-3.4 index. However, comparison to other methods, presented in Fig. 1 of Barnston et al. (1999), shows that even our predicted medians are as good or better than current operational and experimental methods, dynamical or statistical, at 6–7-month leads.

3) LONG-LEAD PREDICTION OF ANOMALY FIELDS

A key strength of the present Bayesian approach is that it accounts for uncertainty and provides distributional results reflecting that uncertainty. This is not only the case for summaries such as those presented in Fig. 5, but for the full spatial anomaly field. Figure 6a shows the observed anomaly field for October 1997. The field of sitewise medians of our out-of-sample 7-month forecast of the smoothed anomaly field, $\Phi_{a,T+\tau}$, for this period (based on data through March 1997) is shown in Fig. 6b. Note that we would have forecasted a strong warm event, although we did not capture the unusual magnitude of this early warm event. This figure also shows the sitewise lower 2.5th and upper 97.5th percentiles, as well as the absolute difference between the two. Note that our prediction is most uncertain in the region of strongest warming and that a very strong warm event is within the ensemble of possible predictions.

4) LONG-LEAD PREDICTION: MIXTURES OF REGIMES

In addition to the distributional information presented above, the forecaster also has information available about the potential regimes. Recall that our forecasting distribution is actually a mixture of three forecasting distributions. For example, Fig. 7 shows the out-of-sample 7-month prediction of the smooth anomaly field for October 1998, based on data through March 1998. In

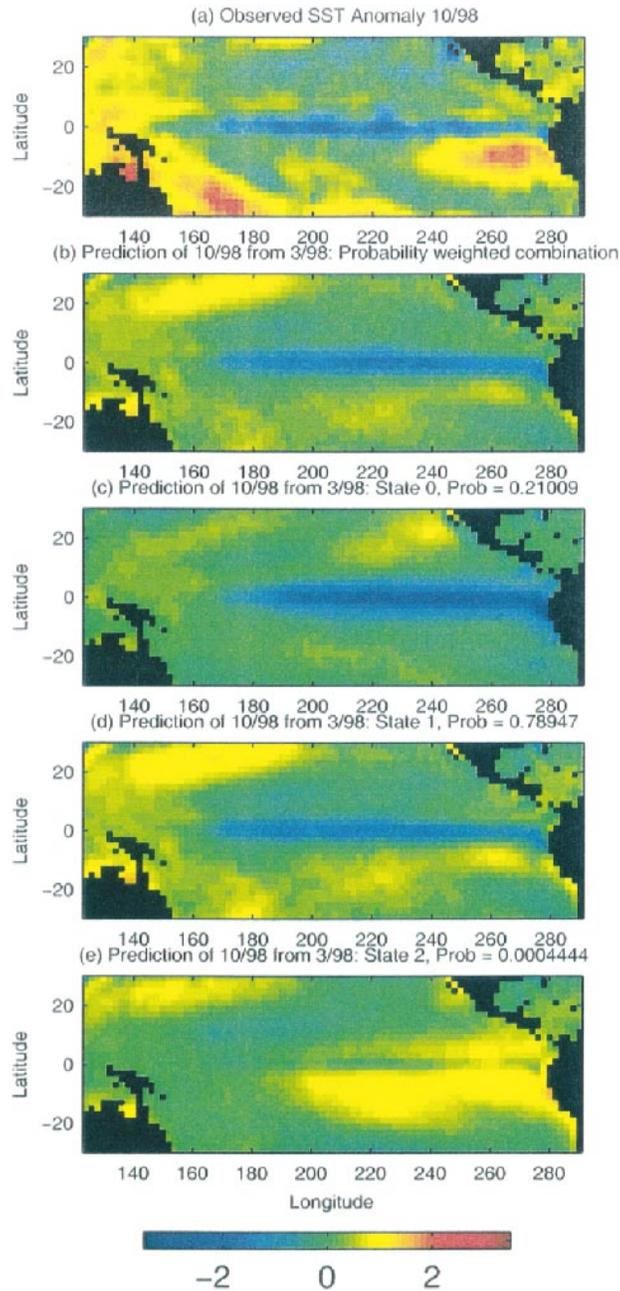


FIG. 7. SST anomalies for Oct 1998. (a) Observed SST anomalies; (b) out-of-sample 7-month probability-weighted mixture of the maps shown in (c), (d), and (e) below, where the mixture weights are based on the model-generated posterior probabilities of moving to the future regime; (c) posterior mean 7-month prediction of smoothed SST anomalies from the Bayesian hierarchical model assuming that the hidden climate regime in Oct 1998 will be cold (regime 0); (d) same as (c) except for future normal regime (regime 1); and (e) same as (c) except for future warm regime (regime 2).

this case, we present three regime-dependent predictions, that is, the predictions (posterior means) if the future regime is cold (regime 0; Fig. 7c), normal (regime 1; Fig. 7d), or warm (regime 2; Fig. 7e). In addition,

we give the mixing probabilities, based on the Bayesian posterior distribution, for each possible future regime. The probability is about 0.79 that the climate regime will be normal, 0.21 that it will be cold, and nearly zero

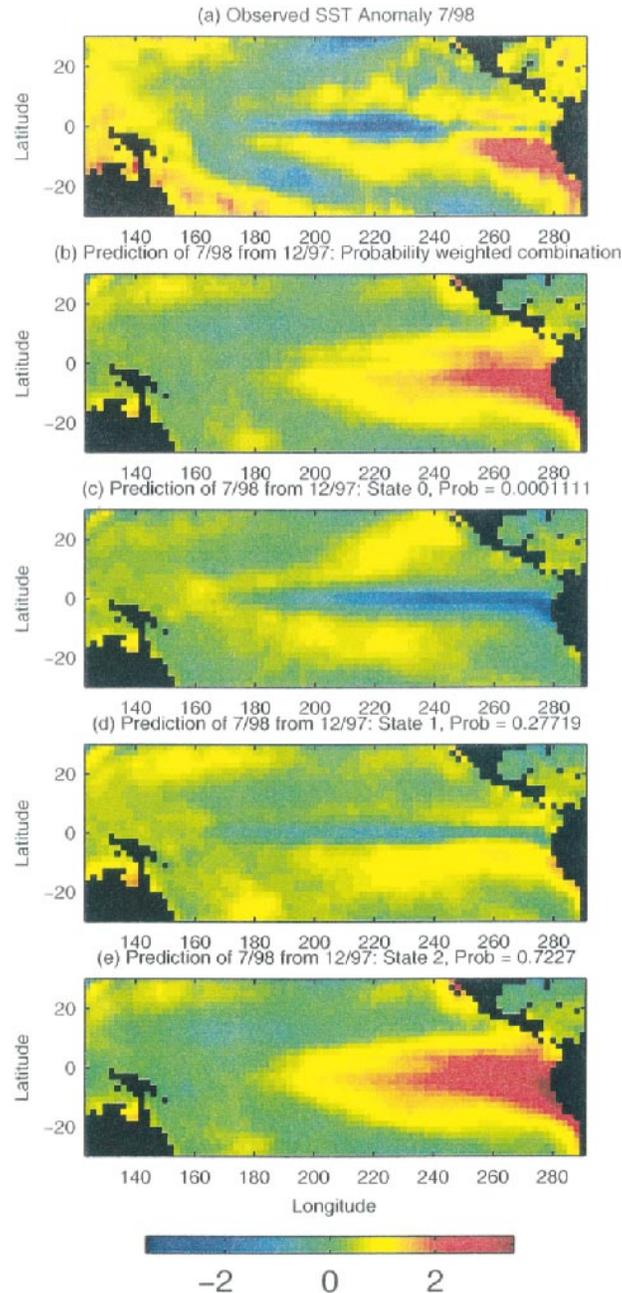


FIG. 8. Same as Fig. 7 except that the out-of-sample 7-month predictions are for Jul 1998.

that it will be warm, in October 1998, based on data through the previous March. The overall posterior mean prediction is the probability weighted sum of these three maps, as shown in Fig. 7b. For comparison, the observed anomalies in October 1998 are also shown in Fig. 7a. The model has done a good job of predicting the continuing cold event (although it recognizes that there is likely to be a deviation from the classic cold-event pat-

tern). Note that forecasters could reweight these maps based on their experience or other information.

Similar mixture prediction results are shown in Fig. 8 for the 7-month prediction of July 1998, based on data through December 1997. Note from Fig. 5 that this was a period during which the model did not forecast well (at least in terms of the median forecast). In this case, the model assigned more weight to the regime 2 (warm)

prediction than it should have in retrospect. But one can see that the cold pool in the central Pacific is suggested by the regime 0 and regime 1 predictions. This ability to recognize several potential forecasts is one of the strengths of the hierarchical mixture approach.

6. Discussion

Recent results of long-lead-prediction model comparisons for the 1997–98 ENSO event suggest that statistical models performed well when compared to deterministic and hybrid models (Barnston et al. 1999; Anderson et al. 1999). Nevertheless, there is often a perception that statistical models for long-lead prediction have exhausted their potential and that dramatic future improvements will come only from deterministic models. In this paper, we present an alternative statistical long-lead-prediction model that makes use of advancements that have revolutionized statistical modeling over the last decade. We do not claim that this particular Bayesian hierarchical space–time mixture model is dramatically superior to existing methodologies. Rather, we present it to illustrate that models such as these open the door to serious exploration of complicated statistical models for long-lead prediction. In this sense, the potential for statistical models in long-lead prediction is every bit as great as that for deterministic models. Furthermore, we note that since this is a desktop workstation model, even comparable results to large deterministic models is a significant advance relative to computational effort.

We have illustrated how one might incorporate physical understanding (e.g., about winds and the SST dynamical system) into this model and account for the uncertainty of using such information (by specifying prior distributions). The output from this model then gives reasonable predictions in a distributional context, having factored in this uncertainty. It could be argued that this is all we can ask of a forecasting procedure, namely, to give the best forecast *and* a realistic assessment of how good it is likely to be.

There is much more that could be done to the present model to improve it. We have noted many of these throughout the exposition above. It was not our intention at this time to find the optimal model, but rather to present a viable alternative methodology. We are developing an operational version of this model, the details of which will be presented elsewhere.

Acknowledgments. This research was supported by the U.S. Environmental Protection Agency under Assistance Agreement R827257-01-0. C. K. Wikle’s research was also partially funded by the Geophysical Statistics Project at the National Center for Atmospheric Research under NSF grant DMS93-12686. The authors thank D. Shea for his help in data acquisition for this project. We are grateful to K. Trenberth, two referees,

and the editor, Francis Zwiers, for their valuable comments and suggestions.

APPENDIX

Priors on Parameters

a. Prior distribution for Σ_ν

The critical portion of the statistical model (4) focuses on use of the first K spatial EOFs. We model Σ_ν based on some of the remaining EOFs, denoted as Φ^c . That is, the error terms $\{\nu_i\}$ are intended to account for spatial structure lost in dimension reduction. The intuition is to define Σ_ν so that it models spatial variation represented by the $m - K$ remaining SST anomaly EOFs. For the results presented here, $m = 2261$ and, in the notation of section 4c, we chose $K = 10$ with $n_l = 6$ and $n_s = 4$. The first 10 EOFs explain approximately 74% of the variance in the SST anomalies. However, since $m - K$ is very large, we did a second dimension reduction for simplicity; namely, we only used EOFs $K + 1$ through $K + k$ (we set $k = 20$, which accounts for an additional 14% of the variance of the SST anomaly field). Of course, k EOFs do not specify an $m \times m$ positive-definite covariance matrix. Hence, to proceed with the intuition given above, we write

$$\Sigma_\nu = \alpha \left(c_\nu \mathbf{I} + \sum_{j=1}^k \lambda_j \Phi_j^c \Phi_j^{c'} \right), \quad (\text{A1})$$

where c_ν was chosen to account for the remaining variance that corresponds to the last $(m - K - k)$ EOFs. Finally, the quantity α is assumed to have an inverse Gamma distribution, $\alpha \sim IG(7, 0.17)$, where $IG(q, r)$ represents an inverse gamma distribution with shape parameter q and scale parameter r . The choices of $q = 7$ and $r = 0.17$ imply that α has prior mean 1 and prior variance 0.04. The results presented in this paper are not overly sensitive to these choices.

b. Prior distributions for $\mathbf{a}_1, \dots, \mathbf{a}_\tau$

To complete the dynamical model for $\{\mathbf{a}_\tau\}$, we specified the following “initializing” prior distributions:

$$\mathbf{a}_t \sim \text{Gau}(\mathbf{A}_t, \Sigma_{A,0}); \quad t = 1, \dots, \tau, \quad (\text{A2})$$

where the \mathbf{A}_t are the least squares fitted spectral coefficients from the EOF decomposition of $\{\mathbf{Z}_t\}$; we specified $\Sigma_{A,0}$ to be the estimated marginal covariance matrix of \mathbf{A}_t multiplied by a constant to “inflate” the prior covariance to account for the fact that these estimates were based on a rather limited set of realizations. We chose the value of 4 for this constant and noted that the results are not sensitive to this choice.

c. Prior distribution for process-model parameters

Consider the dynamic model parameters δ_{ll} , \mathbf{G}_{ll} , \mathbf{G}^s , \mathbf{G}^l , \mathbf{G}^{ss} , and Σ_η . The critical issue for developing prior

models for the intercept and regression parameters is that we do not actually know the hidden process $\{J_t\}$. For purposes of prior development, we assumed that J_t is determined by $\text{SOI}_{t+\tau}$. Then, we simply estimated the parameter matrices by ordinary least squares. These estimates were then used as the prior means. The corresponding prior variances were obtained by inflating the ordinary least squares estimated parameter variances (we arbitrarily used a constant inflation factor of 4 for each and noted that the results are not overly sensitive to this choice). Note that we applied the “vec” operator to all the \mathbf{G} matrices to convert them to vectors. We assumed that these vectors have multivariate normal distributions. The inverse covariance matrix Σ_η^{-1} was assumed to have a Wishart distribution with prior mean given by the corresponding ordinary least squares estimate based on the $\{\mathbf{A}_t\}$ from the EOF decomposition of $\{\mathbf{Z}_t\}$.

d. Prior distribution for θ_j

The key task here is the specification of prior distributions for the parameters β_y and σ_y^2 introduced in (21). This is difficult in principle since the $\{y_t\}$ are hidden variables; we do not even specify what physical variables they represent. One option is to use comparatively uninformed prior distributions for the parameters. A second suggestion, used here, is ad hoc, but intuitive. Recall from section 4d that the y_t would be particularly useful if they essentially predicted $\text{SOI}_{t+\tau}$ since this quantity serves as a primary classifier of regime $I_{t+\tau}$. We actually performed an ordinary least squares linear regression analysis fitting the model

$$\text{SOI}_{t+\tau} = \mathbf{x}'_t \beta_{\text{SOI}} + \varepsilon_{\text{SOI}}(t + \tau), \quad (\text{A3})$$

where the errors $\{\varepsilon_{\text{SOI}}(t + \tau) : t = 1, 2, \dots\}$ were assumed to have a constant variance σ_ε^2 . [cf. (A3) to (21).] The fitted parameters β_{SOI} were assigned to be our prior mean vector for β_y . The estimated covariance matrix, \hat{C} for β_{SOI} , was inflated by a constant c_β and the result was used as our prior covariance for β_y . Finally, we specified $\beta_y \sim \text{Gau}(\beta_{\text{SOI}}, c_\beta \hat{C})$, where c_β was chosen to be 4. The results are not overly sensitive to this choice.

Similarly, the estimated value $\hat{\sigma}_\varepsilon^2$ was used as the prior mean for the distribution on σ_y^2 . Specifically, we assigned an inverse gamma prior to σ_y^2 with mean given by $\hat{\sigma}_\varepsilon^2$ ($=0.65$) and a variance of 0.25 and noted that the results are not sensitive to these choices. We assumed that β_y and σ_y^2 are a priori independent. Although this assumption is highly questionable given the prior construction, it is not suspected to be critical to the final results.

e. Threshold parameters

Recall that, to determine the regime at time t , we use the filtered SOI and require threshold values s_t^- and

s_t^+ . For the results presented here, for all times t , we let s_t^- and s_t^+ be the 33.3rd and 66.7th percentiles from the filtered SOI time series $\{\text{SOI}_t : t = 1, \dots, t = T\}$. Similarly, since we have, in a sense, calibrated the y_t process to $\text{SOI}_{t+\tau}$ by our choice of prior distribution given above, we also set γ_t^- and γ_t^+ to be the same as s_t^- and s_t^+ , respectively.

REFERENCES

- Albert, J. H., and S. Chib, 1993: Bayesian analysis of binary and polychotomous response data. *J. Amer. Stat. Assoc.*, **88**, 669–679.
- Anderson, J., H. van den Dool, A. Barnston, W. Chen, W. Stern, and J. Ploshay, 1999: Present-day capabilities of numerical and statistical models for atmospheric extratropical seasonal simulation and prediction. *Bull. Amer. Meteor. Soc.*, **80**, 1349–1361.
- Barnston, A. G., M. H. Glantz, and Y. He, 1999: Predictive skill of statistical and dynamical climate models of SST forecasts during the 1997–98 El Niño episode and the 1998 La Niña onset. *Bull. Amer. Meteor. Soc.*, **80**, 217–243.
- Berger, J. O., 1985: *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, 617 pp.
- Berliner, L. M., 1996: Hierarchical Bayesian time series models. *Maximum Entropy and Bayesian Methods*, K. Hanson and R. Silver, Eds., Kluwer Academic, 15–22.
- , J. A. Royle, C. K. Wikle, and R. F. Milliff, 1998: Bayesian methods in the atmospheric sciences. *Bayesian Statistics 6*, J. M. Bernardo et al., Eds., Oxford University Press, 83–100.
- , R. A. Levine, and D. J. Shea, 2000: Bayesian climate change assessment. *J. Climate*, **13**, 3805–3820.
- Bernardo, J. M., and A. F. M. Smith, 1994: *Bayesian Theory*. Wiley and Sons, 586 pp.
- Blanke, B., J. D. Neelin, and D. Gutzler, 1997: Estimating the effect of stochastic wind stress forcing on ENSO irregularity. *J. Climate*, **10**, 1473–1486.
- Cane, M. A., 1992: Tropical Pacific ENSO models: ENSO as a mode of the coupled system. *Climate System Modeling*, K. E. Trenberth, Ed., Cambridge University Press, 583–614.
- Courtier, P., 1997: Dual formulation of four-dimensional variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **123**, 2449–2461.
- Cressie, N. A. C., 1993: *Statistics for Spatial Data, Revised Edition*. Wiley and Sons, 900 pp.
- Eckert, C., and M. Latif, 1997: Predictability of a stochastically forced hybrid coupled model of El Niño. *J. Climate*, **10**, 1488–1504.
- Epstein, E. S., 1985: *Statistical Inference and Prediction in Climatology: A Bayesian Approach*. Amer. Meteor. Soc., 199 pp.
- Gelfand, A. E., and A. F. M. Smith, 1990: Sampling-based approaches to calculating marginal densities. *J. Amer. Stat. Assoc.*, **85**, 398–409.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 1995: *Bayesian Data Analysis*. Chapman and Hall, 526 pp.
- Geman, S., and D. Geman, 1984: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, **6**, 721–741.
- Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, Eds., 1996: *Markov Chain Monte Carlo in Practice*. Chapman and Hall, 486 pp.
- Hasselmann, K., 1998: Conventional and Bayesian approach to climate-change detection and attribution. *Quart. J. Roy. Meteor. Soc.*, **124**, 2541–2565.
- Hughes, J. P., and P. Guttorp, 1994: A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resour. Res.*, **30**, 1535–46.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kitanidis, P. K., 1986: Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resour. Res.*, **22**, 499–507.

- Kleeman, R., and A. M. Moore, 1997: A theory for the limitations of ENSO predictability due to stochastic atmospheric transients. *J. Atmos. Sci.*, **54**, 753–767.
- Lau, K.-M., 1985: Elements of a stochastic dynamical theory of the long-term variability of the El Niño–Southern Oscillation. *J. Atmos. Sci.*, **42**, 1552–1558.
- , and P. H. Chan, 1986: The 40–50 day oscillation and the El Niño–Southern Oscillation: A new perspective. *Bull. Amer. Meteor. Soc.*, **67**, 533–534.
- , and —, 1988: Intraseasonal and interannual variations of tropical convection: A possible link between 40–50-day oscillation and ENSO? *J. Atmos. Sci.*, **45**, 506–521.
- Leroy, S. S., 1998: Detecting climate signals: Some Bayesian aspects. *J. Climate*, **11**, 640–651.
- Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194.
- Lu, Z.-Q., and L. M. Berliner, 1999: Markov switching time series models with application to a daily runoff series. *Water Resour. Res.*, **35**, 523–534.
- McCullagh, P., and J. A. Nelder, 1989: *Generalized Linear Models*. 3d ed. Chapman and Hall, 511 pp.
- McPhaden, M. J., and X. Yu, 1999: Genesis and evolution of the 1997–1998 El Niño. *Science*, **283**, 950–954.
- Meinhold, J., and N. D. Singpurwalla, 1983: Understanding the Kalman filter. *Amer. Stat.*, **37**, 123–127.
- Moore, A. M., and R. Kleeman, 1999: Stochastic forcing of ENSO by the intraseasonal oscillation. *J. Climate*, **12**, 1199–1220.
- Nakazawa, T., 1998: MJO—A key component in the atmosphere for triggering ENSO. *Proc. Conf. on the TOGA Coupled Ocean–Atmosphere Response Experiment (COARE)*, Boulder, CO, World Meteorological Organization, World Climate Research Program, 165–166.
- Omre, H., 1987: Bayesian kriging—Merging observations and qualified guesses in kriging. *Math. Geol.*, **19**, 25–39.
- Penland, C., 1996: A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D*, **98**, 534–558.
- , and T. Magorian, 1993: Prediction of Niño 3 sea surface temperatures using linear inverse modeling. *J. Climate*, **6**, 1067–1076.
- , and P. D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. *J. Climate*, **8**, 1999–2024.
- Rabiner, L. R., and B. Gold, 1975: *Theory and Application of Digital Signal Processing*. Prentice Hall, 762 pp.
- Royle, J. A., L. M. Berliner, C. K. Wikle, and R. Milliff, 1998: A hierarchical spatial model for constructing wind fields from scatterometer data in the Labrador Sea. *Case Studies in Bayesian Statistics*, C. Gatsonis et al., Eds., Springer-Verlag, 367–382.
- Tarantola, A., 1987: *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier, 613 pp.
- Tong, H., 1990: *Non-linear Time Series: A Dynamical Systems Approach*. Oxford University Press, 564 pp.
- Trenberth, K. E., 1997: The definition of El Niño. *Bull. Amer. Meteor. Soc.*, **78**, 2771–2777.
- Vallis, G. K., 1988: Conceptual models of El Niño and the Southern Oscillation. *J. Geophys. Res.*, **93**, 13 979–13 991.
- Wikle, C. K., and N. Cressie, 1999: A dimension reduced approach to space–time Kalman filtering. *Biometrika*, **86**, 815–829.
- , L. M. Berliner, and N. Cressie, 1998: Hierarchical Bayesian space–time models. *J. Environ. Ecol. Stat.*, **5**, 117–154.
- Zebiak, S. E., 1989: On the 30–60 day oscillation and the prediction of El Niño. *J. Climate*, **2**, 1381–1387.
- Zwiers, F., and H. von Storch, 1990: Regime-dependent autoregressive time series modeling of the Southern Oscillation. *J. Climate*, **3**, 1347–1363.