Statistical Science
2003, Vol. 18, No. 4, 422–429

Statistics in Atmospheric Science

Andrew R. Solow

Abstract. This paper reviews the use of statistical methods in atmospheric science. The applications covered include the development, assessment and use of numerical physical models of the atmosphere and more empirical analysis unconnected to physical models.

Key words and phrases: Data assimilation, general circulation model, model assessment, parameterization of physical processes, spatial time series.

1. INTRODUCTION

In broad terms, the ultimate goal of atmospheric science is prediction. The atmosphere (and the ocean, to which it is inextricably coupled) form a classical, if extremely complex, dynamical system. In principle, given complete knowledge of the current state of this system, the physical laws that govern it, the topography of the domain over which these laws operate and external forces acting upon it, it would be possible to predict its future state without error. Needless to say, the complete knowledge needed for perfect prediction is unattainable and the bread-and-butter of atmospheric science is the development of various kinds of approximations that make useful, if imperfect, prediction possible and that can provide a measure of the imperfection of the prediction. Statistics has a role to play in virtually every aspect of this enterprise and the goal of this article is to review some of the ways in which statistical methods have been used in atmospheric science.

In the United States, the application of statistical methods to problems in atmospheric science has been advanced greatly by the establishment in 1994 of the Geophysical Statistics Project (GSP) at the National Center for Atmospheric Research (NCAR). This project has had two leaders: Mark Berliner, who served from 1995 until 1997, and Douglas Nychka, who has served since 1997. The project supports collaboration between atmospheric scientists and statisticians. Its centerpiece is a program that brings doctoral students and postdoctoral researchers together with more senior scientists for collaboration. As a glance at the references of this article will testify, the GSP has made a signal contribution to the advancement of statistics in atmospheric and related science. Some of this work is presented in the volume edited by Berliner, Nychka and Hoar (2000). The monograph by von Storch and Zwiers (1999), although aimed at atmospheric scientists and oceanographers, provides an excellent description of some statistical problems in these fields and also reviews some of the statistical and quasistatistical methods in current use.

The remainder of this review is organized in the following way. In Section 2, some statistical issues arising in the use of physical models of the atmosphere are discussed. In Section 3, some statistical work on atmospheric data that is less closely tied to physical models is reviewed. Section 4 contains some concluding remarks.

2. STATISTICS IN ATMOSPHERIC MODELING

This section reviews some of the ways in which statistics has been used in the development, assessment and use of numerical models of atmospheric dynamics.

2.1 Parameterizing Subgrid-Scale Processes

The dynamics of the atmosphere can be described by a coupled system of partial differential equations (Salby, 1996). These are called the primitive equations. Unfortunately, the primitive equations are too complicated to solve explicitly and atmospheric scientists have turned to large-scale numerical models called general circulation models (GCMs). In a GCM, the primitive equations are discretized and numerically integrated in time on a three-dimensional spatial

Andrew R. Solow is Associate Scientist and Director, Marine Policy Center, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543 (e-mail: asolow@whoi.edu).

grid. Limitations on computing power require that this grid be relatively coarse. For example, the Community Climate Model maintained at the National Center for Atmospheric Research divides the atmosphere into $\sim 8{,}000$ grid boxes with horizontal dimensions of $\sim$300-by-300 km at the equator at each of $\sim$20 atmospheric layers and uses a time step of 20 minutes for integrations of up to 100 years.

A consequence of the relatively coarse spatial grid used in GCMs is that atmospheric processes that operate at smaller scales are not resolved. Although these processes operate at relatively small spatial scales, they can play an important role in atmospheric dynamics. To incorporate subgrid-scale processes into GCMs, atmospheric scientists use parameterizations. A parameterization is essentially an empirical representation of an unmodeled process that can be incorporated into a model (Trenberth, 1992). The following general formulation is based on Berliner, Royle, Wikle and Milliff (1999). Consider a vector-valued dynamical system z_t that can be decomposed as $z_t = (x_t, y_t)$. In general, both components x_t and y_t are also vector-valued. The dynamics of this system are governed by

$$\begin{aligned} x_{t+1} &= h(x_t, y_{t+1}), \\ y_{t+1} &= g(x_t, y_t). \end{aligned} \tag{1}$$

Suppose that the components of x_t are resolved by a numerical model, but the components of y_t are not. The goal is to construct a model for x_t that still incorporates the effect of y_t. Toward this end, construct a parameterization of the unmodeled component:

$$\hat{y}_{t+1} \cong f(x_t; \theta), \tag{2}$$

where $\hat{y}_t$ is a simplified representation of the unmodeled process and θ is a collection of parameters. Finally, substitute this parameterization into the model:

$$x_{t+1} = \tilde{h}\big(x_t, f(x_t; \theta)\big), \tag{3}$$

where $\tilde{h}$ is the function relating x_{t+1} to x_t and the simplified representation $\hat{y}_{t+1}$. As an example of a simplified representation, in the case of cloud parameterizations discussed below, it is common to parameterize the fraction of a GCM grid box that is covered by cloud and not individual clouds themselves. The statistical side of this problem is to fit the parameterization (2) as a regression problem.

From the perspective of modeling the long-term effects on climate of increasing atmospheric concentrations of greenhouse gases, the most important subgrid-scale atmospheric process is cloud formation. Clouds have a substantial effect on the Earth's radiation budget. This effect is complicated. Briefly, thick, low-lying clouds act to cool the surface by reflecting solar radiation back into space, while thin, high clouds act to warm the surface by trapping reflected longwave radiation through the greenhouse effect. General circulation models generally treat the amount of clouds and their radiative effects separately. Parameterizations of cloud amount typically involve two steps. In the first step, each grid box is determined to be convective or nonconvective. Convection refers to the mechanism by which water vapor evaporated at the ocean surface (and elsewhere on the Earth's surface) is transported vertically into the atmosphere to form clouds. This determination is based on the vertical profile of the modeled atmosphere. In the second step, the amount of cloud in each convective grid box is estimated from an empirical relationship between cloud amount and modeled variables. Because detailed observational data on clouds are scarce, this relationship is usually estimated from simulations from a cloud-resolving model with overall domain around the size of a single GCM grid box. As an example, Xu and Randall (1996) proposed the following parameterization for the stratiform cloud fraction C_S of the sky in a GCM grid box:

$$C_S = r_H^p\big(1 - \exp(-\alpha\bar{q}_e)\big), \tag{4}$$

where r_H is relative humidity, $\bar{q}_e$ is the large-scale condensate mixing ratio and

$$\alpha = \alpha_0\big((1 - r_H)q^*\big)^{-\gamma}, \tag{5}$$

where q^* is the water vapor mixing ratio. The model provides values of $r_H, \bar{q}_e$ and q^* and the parameters p, α_0 and γ are estimated statistically from model simulations.

Bailey et al. (2000) described a more complicated cloud parameterization. Briefly, in their parameterization, the cloud amount at time $t + 1$ in a grid box depends nonlinearly on the cloud amount at time t and the values of three atmospheric variables—relative humidity, Richardson's number (a measure of shear in the atmosphere) and the time rate of change of generalized convective available potential energy (a predictor of convection)—in the grid box itself and in each of its four nearest neighbors. The resulting nonlinear regression model, which involves a total of 16 regressors, was fitted to data from a cloud-resolving model using a feed-forward neural network and was found to perform reasonably well.

The practice of fitting a parameterization to the output of a deterministic model by regression raises a fundamental question. In the usual regression setting, the response is treated as a random variable whose mean is a function of some regressors. Even if this function is known, the response will be different from the mean as a result of some kind of stochastic variation. In practice, some part of the deviation between observed and fitted values of the response may reflect model error but, provided the deviations show little systematic structure, this is ignored. In the case of model-based parameterization, the regression machinery based on a stochastic model is used for deterministic functional approximation. In this case, the deviations between the "observed" and fitted values of the response can only be due to model error in the parameterization and do not reflect uncertainty in the traditional sense. The statistical treatment of the output of deterministic models was reviewed by Sacks, Welch, Mitchell and Wynn (1989) with particular emphasis on the design of computer experiments. Kennedy and O'Hagan (2001) outlined a comprehensive Bayesian formalism for the analysis of model output that incorporates all sources of error, including model error, in a statistical framework. Berliner (2003) discussed this kind of approach in the context of climate modeling.

2.2 Model Assessment

Once an atmospheric model is constructed, it is necessary to assess its ability to simulate atmospheric dynamics before using it for prediction. One important use of GCMs is forecasting the long-term response of the climate system to actual or hypothetical changes in the atmospheric concentrations of carbon dioxide and other radiatively active gases. As this is an issue of considerable interest to society, there is a need to know whether GCMs provide an accurate picture of climate. In broad terms, the assessment of GCMs is based on comparing model reconstructions of historical climate to historical climate records. This is clearly a statistical problem and it has been addressed in a number of ways with varying degrees of formality. A particularly lucid exposition from a statistical perspective was provided by Levine and Berliner (1999); for a Bayesian perspective on related issues, see Berliner, Levine and Shea (2000). In their notation, the basic statistical model is

$$\Psi = \Psi^s + \tilde{\Psi}, \tag{6}$$

where Ψ is a vector of dimension n of climate observations, Ψ^s is the climate signal and $\tilde{\Psi}$ is internal climate variability with mean 0 and dispersion matrix C. The elements of these vectors correspond to different climate variables (e.g., temperature, pressure, humidity) at different locations in time and space. Typically, both Ψ and Ψ^s are expressed as deviations or anomalies from some baseline. Let the n-dimensional vectors $g_1, g_2, \ldots, g_p$ represent the modeled responses of the climate signal to prescribed radiative forcings and let G be the n-by-p matrix with columns given by these vectors. For example, g_1 could represent the modeled response to the radiative effect of historical increases in greenhouse gas concentrations, g_2 could represent the modeled response to the radiative effect of historical solar variability and so on. Under the assumption that these responses are additive, the statistical model can be written as the regression

$$\Psi = Ga + \tilde{\Psi}, \tag{7}$$

where a is a vector of unknown regression parameters of dimension p.

Under this formulation, Levine and Berliner (1999) proposed as a method of model assessment a formal test of what they referred to as geoequivalency, in analogy to bioequivalency in drug testing. The basic idea is to test the compound null hypothesis $H_0 : a \neq a_M$ against the simple alternative hypothesis $H_1 : a = a_M$, where a_M is the model-based value of a. The following test is due to Brown, Casella and Hwang (1995). Let $\theta = a - a_M$ and $\hat{\theta} = \hat{a}_{\text{obs}} - \hat{a}_M$, where $\hat{a}_{\text{obs}} = (G'C^{-1}G)^{-1}G'C^{-1}\Psi$ is the generalized least squares estimate of a based on the observations and $\hat{a}_M$ is the estimate of a_M based on model simulations. Let $\Sigma = C_{\text{obs}} + C_M$, where $C_{\text{obs}} = (G'C^{-1}G)^{-1}$ and C_M are the dispersion matrices of $\hat{a}_{\text{obs}}$ and $\hat{a}_M$, respectively. Under the assumption that $\hat{\theta}$ has a multivariate normal distribution with mean θ and dispersion matrix Σ, H_0 can be rejected at significance level α if the confidence region

$$R = \{\theta : (\theta'\Sigma^{-1}\theta)^{1/2} \le z_\alpha + (\theta'\Sigma^{-1}\theta)^{-1/2} \cdot (\hat{\theta}'\Sigma^{-1}\theta)\} \tag{8}$$

does not cover the origin.

Several issues arise in the implementation of this approach. Chief among these is the need to estimate the dispersion matrix C and the effect of using this estimate in place of C on the properties of the test. Levine and Berliner (1999) outlined two proposals for estimating C. One is based on repeated control runs of the model—that is, simulations in which none of the radiative forcings that give rise to the responses comprising G is applied to the model. This places the additional burden on the GCM of correctly simulating unforced variability in climate. The second proposal is to estimate C from the observations themselves. The difficulty here is that these observations reflect variations that are not due solely to true natural variability—for example, measurement error and data processing. Finally, there seems to be a problem with interpreting the result of this test. As every GCM is based on some approximations to reality, the null hypothesis is certainly true and there is no logical reason to test it. To put it another way, as the only possible reason for failing to reject the null hypothesis is low power, provided that additional observations are collected through time, this test is destined to reject the null hypothesis.

2.3 Data Assimilation

The previous discussion has dealt with the use of statistical methods in developing an atmospheric model and in assessing its predictive ability. Here, the use of statistical methods in the prediction process itself is discussed. The term data assimilation refers generally to the blending of observations with a numerical physical model for improved prediction. One use of data assimilation is simply to convert measurements of one variable into measurements of another. For example, satellite measurements of energy need to be converted into measurements of temperature. A second use of data assimilation is in numerical weather prediction. Numerical weather prediction is based on integrating a numerical model forward from observed initial conditions. Such predictions can be highly sensitive to errors in these initial conditions and there is a need to incorporate observations on a continuous basis. In both of these applications, the use of a physical model is intended to ensure that the results obey basic physical laws. Data assimilation and related issues are discussed in Daley (1997) and Sneddon (2000).

To begin with, consider the data assimilation problem at a fixed point in time. This would correspond, for example, to the problem of using observations of one atmospheric variable to predict the values of another. The basic model is

$$y = h(x) + \varepsilon, \tag{9}$$

where y is the vector of observations, x is the vector of values to be predicted and ε is a vector of observation errors. The function h is an expression of the physical law relating x and y. To proceed, this model is linearized in some way so that

$$y \cong Hx + \varepsilon, \tag{10}$$

where H is a known matrix. Let x_0 be an initial guess at x. For example, this guess can be based on an earlier assimilation. The prediction $\hat{x}$ is found by minimizing

$$\begin{aligned} J(x) = {} & (y - Hx)'P^{-1}(y - Hx) \\ & + (x - x_0)'R^{-1}(x - x_0), \end{aligned} \tag{11}$$

where P is the dispersion matrix of the observation errors ε and R is the dispersion matrix of the so-called forecast errors $(x - x_0)$. The minimization of $J(x)$ is called a three-dimensional variational problem. The term three-dimensional emphasizes that assimilation is at a set of locations in three spatial dimensions at a fixed point in time. The solution to the three-dimensional variational problem is given by

$$\hat{x} = (H'P^{-1}H + R^{-1})^{-1}(H'P^{-1}y + R^{-1}x_0). \tag{12}$$

The variational problem can be interpreted as a penalized least squares problem where the penalty is the second term on the right-hand side of (11). This problem also has a straightforward Bayesian interpretation. Specifically, $\hat{x}$ corresponds to the posterior mode of the random variable X with multivariate normal prior distribution with mean x_0 and dispersion matrix R and where the distribution of ε is also multivariate normal with mean 0 and dispersion matrix P. The Bayesian formulation leads naturally to results about prediction accuracy and about the design of observations for optimal prediction (Lu, Berliner and Snyder, 2000).

While the solution to the three-dimensional variational problem is straightforward in principle, there are significant problems in implementation. One problem is that the dimension of x can be very large—on the order of 10^5–10^6—so that the inversion of R is challenging (Sneddon, 2000). To circumvent this problem, it is common to make assumptions about the spatial dependence of the forecast errors that result in some simplifying structure in R. Another problem is that the quality of the linear approximation to h in the neighborhood of x_0 may not always be good. A third problem can arise if the assumption of normal error is violated. An example involving the assimilation of precipitation data, for which the assumption of normal error is not appropriate, is discussed in Errico, Fillion, Nychka and Lu (2000).

The second common use of data assimilation is in numerical weather prediction. Here, the goal is to predict the state of the atmosphere at time t based on information available at time $t - 1$. This problem, which is referred to as four-dimensional data assimilation, can be formulated through a state-space model. The observation equation

$$y_t = H_t x_t + \varepsilon_t \tag{13}$$

is the same as (10) with the addition of the time index. The state equation is

$$x_t = M_t x_{t-1} + \eta_t, \tag{14}$$

where M_t is a linearization of the model governing atmospheric dynamics and η_t represents stochastic variability in the atmosphere. The combined state-space model is solved using the Kalman filter. For example, the prediction of x_t is given by

$$\hat{x}_t = M_t \hat{x}_{t-1} + G_t(y_t - H_t M_t \hat{x}_{t-1}), \tag{15}$$

where G_t is the Kalman gain matrix that depends on H_t, the dispersion matrix P_t of ε_t, and the dispersion matrix of the error of predicting x_t based on data available at time $t-1$. Details can be found in Cohn (1997). A Bayesian interpretation is given in West and Harrison (1997).

The same kinds of problems that arise in the implementation of three-dimensional data assimilation also arise in the four-dimensional version. An acute problem arises from the need to invert large dispersion matrices at each time step. Various methods have been used to simplify this calculation, including reducing dimensionality by focusing on directions in which prediction errors grow particularly quickly. The technical report by Fisher and Courtier (1995) describes some of these approaches.

The equations governing the atmospheric dynamics are nonlinear. Briefly, to accommodate this, in the so-called extended Kalman filter, M_t in (14) can be continuously updated by taking a linear approximation in the neighborhood of $\hat{x}_{t-1}$. Other approaches to extending the Kalman filter to the nonlinear case are discussed in Kitagawa (1996) and Evensen and van Leeuwen (2000). The nonlinearity of atmospheric dynamics famously gives rise to extreme sensitivity to initial conditions (Lorenz, 1963). For this reason, to map out prediction uncertainty it is now common to initialize weather predictions from a suitably selected ensemble of initial conditions. Ensemble forecasting is discussed in Sivillo, Ahlquist and Toth (1997) and, from a statistical perspective, in Berliner (2001).

3. EMPIRICAL STATISTICAL ANALYSIS

The previous section reviewed the use of statistical methods in atmospheric modeling. Statistics has also been used to analyze atmospheric data outside the context of physical modeling. This section gives two examples of this kind of work: one concerned with the development of statistical methods for decomposing spatial time series and the other with an application of well-established methods to inference based on a partially incomplete climate record.

3.1 Multivariate Methods for Spatial Time Series

Atmospheric and oceanographic data commonly take the form of spatial time series, that is, time series of the same variable measured at a collection of locations. Often atmospheric scientists are interested in empirically decomposing such data into uncorrelated modes of variation and, specifically, in isolating low-frequency modes for further analysis. Let $y(t)$ be a discrete-time, 0-mean, vector time series of dimension p. For example, $y(t)$ could represent detrended mean annual surface temperature at a set of p measurement sites. The overall goal of this kind of analysis is to effect a decomposition of the form

$$y(t) = \sum_{j=1}^{k} w_j x_j(t) + \varepsilon(t) \tag{16}$$

with $k < p$, where $x_1(t), x_2(t), \ldots, x_k(t)$ is a set of orthogonal univariate index series or temporal modes of variation, w_j is a vector of dimension p that describes the spatial structure of this mode and $\varepsilon(t)$ represents residual variability not captured by the k modes. There are various motivations for this kind of decomposition. One is to attempt to identify physically meaningful modes of variation in atmospheric dynamics such as the El Niño–Southern Oscillation or the North Atlantic Oscillation. Another is to identify modes of variability that are temporally smooth and, therefore, predictable from their past values.

The decomposition in (16) is not unique and the basic problem is to choose one. By a wide margin, the most common approach in atmospheric science and oceanography is to use principal component analysis (PCA), also called empirical orthogonal function (EOF) analysis (Jolliffe, 1986). In PCA, the first mode is given by the linear combination

$$x_1(t) = \sum_{j=1}^{p} v_j y_j(t) \tag{17}$$

of the elements of $y(t)$, where the weights are chosen to maximize the temporal variance of $x_1(t)$ subject to the side condition $\sum_{j=1}^{p} v_j^2 = 1$. It is straightforward to show that these weights are given by the elements of the unit eigenvector corresponding to the largest eigenvalue of the marginal spatial covariance function Σ_0 of $y(t)$. The vector w_1 in (16) can be found by regressing the original component series $y_j(t)$ on $x_1(t)$. Additional index series are constructed from the remaining eigenvectors of Σ_0.

An unsatisfactory feature of this use of PCA is that, because it is based on the marginal spatial covariance function, it makes no use of the temporal pattern of the original time series. As an alternative, Hasselmann (1988) proposed a method called principal oscillation patterns (POPS) analysis. This approach is based on a first-order vector autoregressive model:

$$y(t) = \Theta y(t-1) + \eta(t), \tag{18}$$

where Θ is a matrix of autoregressive parameters and $\eta(t)$ is multivariate white noise. This can be viewed as a linear approximation to the nonlinear dynamics of the field. The approach parallels principal component analysis, but is based on the eigenvectors of the matrix $B = \Sigma_1 \Sigma_0^{-1}$, where Σ_1 is the lag-one autocovariance matrix. In counterpoint to PCA, POPS analysis focuses on identifying (possibly complex) components with strong temporal structure but that may not be strongly connected to the spatial structure of the original field.

As a compromise, Kooperberg and O'Sullivan (1996) proposed a method called predictive oscillation patterns (PROPS) analysis. Briefly, PROPS analysis extracts spatial structures—the w_j in (16)—that minimize an upper bound on the expected squared one-step-ahead prediction error. A recursive algorithm is used to identify these structures. Minimization of the upper bound on prediction error entails making both the variance of the index series large (as in PCA) and their predictability high (as in POPS).

This area of statistical climatology has suffered from a clear statement of the underlying goal of the analysis. The proposal by Kooperberg and O'Sullivan (1996) to focus on predictability is a substantial step in the right direction. Objectives other than prediction will, of course, lead to other methods, but the important point is to develop the method from an explicit objective. A practical issue that needs to be faced in this kind of work concerns the spatial–temporal pattern of data availability. Atmospheric and oceanographic data are sparse and irregularly located in space and time. Many methods of analysis assume some kind of regularity in spacing. For example, PROPS requires the estimation of spectral densities and Kooperberg and O'Sullivan (1996) assume that regularly spaced time series data are available for this purpose. It is common practice in atmospheric science to construct regular data sets by interpolating or gridding sparse, irregularly spaced observations. It does not appear to be widely recognized among atmospheric scientists that the statistical properties of gridded data are different from those of the underlying variables and that care should be taken in removing artifacts created by gridding.

3.2 Testing for Trend in a Partially Incomplete Hurricane Record

So far, this review has focused on examples of general statistical problems in atmospheric science. This final example considers a specific application of well-established statistical methods to address a substantive question in atmospheric science.

A perennial problem in the empirical analysis of historical data in atmospheric science is the unreliability of the early observational record. This is a serious problem because the temporal scale of variability of most interest to studies of climate change presses the limit of the reliable record. One example is the annual record of North Atlantic hurricane counts. Prior to the advent of regular aircraft reconnaissance in 1946, a hurricane that did not come close to land had a significant probability of going undetected. For this reason, attempts to identify a long-term trend in the mean annual number of hurricanes were restricted to the period of the complete record. Interest in this issue stems from a possible connection between overall warming and hurricane activity. As the North Atlantic basin experienced a general warming trend over the 20th century, the corresponding pattern of hurricane activity would shed some light on this connection.

Although the record of North Atlantic hurricanes is complete only since around 1946, the record of landfalling hurricanes is believed to be complete back to 1900 and probably further. Analysis of the postwar record suggests that the mean proportion of hurricanes that make landfall has been constant since 1946. Under the assumption that this result also holds prior to 1946, the possibility arises of supplementing the partially incomplete record of all hurricanes with the complete record of landfalling hurricanes to extend the analysis of overall hurricane activity back beyond 1946. This idea was pursued by Solow and Moore (2000) to test for trend in overall hurricane activity over the period 1930–1998.

Let the random variable Y_t be the true basinwide number of hurricanes in year t $(t = 1, 2, \ldots, n)$ and suppose that Y_t has a Poisson distribution with mean $\mu_t = \exp(\mu_0 + \mu_1 t)$. Interest centers on testing the null hypothesis $H_0: \mu_1 = 0$ of no change in mean number against the general alternative hypothesis $H_1: \mu_1 \notin 0$. Let the random variable X_t be the number of landfalling hurricanes in year t and assume that, conditional on $Y_t = y_t$, X_t has a binomial distribution with y_t trials and unknown landfalling probability p. Finally, let the random variable Z_t be the observed number of hurricanes in year t that did not make landfall and assume that, conditional on y_t and the observed value x_t of X_t, Z_t has a binomial distribution with $y_t - x_t$ trials and unknown sighting probability $q_t = q$ for $t = 1, 2, \ldots, m$ and 1 for $t = m+1, m+2, \ldots, n$.

Using this model, Solow and Moore (2000) tested for trend over the period 1930–1998 using the likelihood ratio test. The relatively short extension beyond 1946 was chosen to ensure the reasonableness of the assumption that the sighting probability q is constant over the period of incomplete record. The contribution to the likelihood function of the observation (x, z) in year t is

$$f_t(x,z) = \sum_{y=x+z}^{\infty} \binom{y-x}{z} q_t^z (1-q_t)^{y-x-z} \cdot \binom{y}{x} p^x (1-p)^{y-x} \cdot \mu_t^y \exp(-\mu_t) \Big/ y!. \tag{19}$$

The lower bound of the summation reflects the fact that the total number of hurricanes in a given year cannot be less than the observed number. The maximum likelihood estimate of μ_1 is around -0.005 with an approximate significance level of around 0.09. Solow and Moore (2002) extended this analysis back to 1900. In doing so, they checked and retained the assumption of a constant sighting probability, but relaxed the parametric log linear model for μ_t in favor of a nonparametric approach based on kernel estimation. The significance level estimated by a parametric bootstrap was around 0.12. Overall, the analysis suggests little evidence of a trend in hurricane activity.

4. CONCLUDING REMARKS

The involvement of statisticians in atmospheric science has grown markedly over the past 10 years. This article has attempted to review in broad strokes some of the areas in which this involvement has contributed and can continue to contribute. As in other fields of application, there has been an ongoing tension between the level of statistical sophistication favored by statisticians working in this area and the practical requirements of atmospheric scientists for whom statistics is not the main object of interest. A particularly gratifying development of the past few years has been a general relaxation of this tension.

A review of this kind is necessarily selective. One important area of statistical work in atmospheric science not reviewed here is downscaling. Downscaling refers to the use of geographically coarse model output (or other types of large-scale climate information) to predict climate on a geographically finer scale. Downscaling is commonly done by using historical data to construct a regression model relating regional or local climate response variables to selected model output (sometimes called model output statistics, or MOS) as regressors. The fitted model can then be used in the future to predict the local variables from model output. Statistical issues that arise in downscaling include the specification of the regression model and the selection of regressor variables. A sample of recent work in this area includes Vislocky and Fritsch (1995), who used generalized additive modeling in predicting aviation weather; Kuligowski and Barros (1998), who considered the use of neural nets in modeling the relationship between local climate and model output; and Bellone, Hughes and Guttorp (2000), who used a hidden Markov model for downscaling large-scale atmospheric patterns to predict local precipitation amounts.

A second important topic not reviewed here is the development of empirical prediction models that rely heavily on physical understanding. These prediction models occupy a middle ground between deterministic physical models and simple empirical models. Two recent examples are the work of Berliner, Wikle and Cressie (2000), who developed a dynamic Bayesian model for predicting sea surface temperature in the equatorial Pacific, and Wikle, Milliff, Nychka and Berliner (2001), who developed a hierarchical Bayesian spatiotemporal model for constructing high resolution predictions of surface winds.

ACKNOWLEDGMENT

The helpful comments of two anonymous reviewers are acknowledged with gratitude.

REFERENCES

BAILEY, B. A., BERLINER, L. M., COLLINS, W., NYCHKA, D. W. and KIEHL, J. T. (2000). Neural networks: Cloud parameterizations. In *Studies in the Atmospheric Sciences. Lecture Notes in Statist.* **144** 97–116. Springer, New York.

BELLONE, E., HUGHES, J. P. and GUTTORP, P. (2000). A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. *Climate Res.* **15** 1–12.

BERLINER, L. M. (2001). Monte Carlo based ensemble forecasting. *Stat. Comput.* **11** 269–275.

BERLINER, L. M. (2003). Physical–statistical modeling in geophysics. *J. Geophys. Res.—Atmospheres*. To appear.

BERLINER, L. M., LEVINE, R. A. and SHEA, D. J. (2000). Bayesian climate change assessment. *J. Climate* **13** 3805–3820.

BERLINER, L. M., NYCHKA, D. and HOAR, T., eds. (2000). *Studies in the Atmospheric Sciences*. *Lecture Notes in Statist.* **144**. Springer, New York.

BERLINER, L. M., ROYLE, J. A., WIKLE, C. K. and MILLIFF, R. F. (1999). Bayesian methods in the atmospheric sciences. In *Bayesian Statistics 6* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 83–100. Oxford Univ. Press.

BERLINER, L. M., WIKLE, C. K. and CRESSIE, N. (2000). Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *J. Climate* **13** 3953–3968.

BROWN, L., CASELLA, G. and HWANG, J. T. G. (1995). Optimal confidence sets, bioequivalence, and the limaçon of Pascal. *J. Amer. Statist. Assoc.* **90** 880–889.

COHN, S. E. (1997). An introduction to estimation theory. *J. Meteorol. Soc. Japan* **75** 257–288.

DALEY, R. (1997). Atmospheric data assimilation. *J. Meteorol. Soc. Japan* **75** 319–329.

ERRICO, R. M., FILLION, L., NYCHKA, D. and LU, Z.-Q. (2000). Some statistical considerations associated with the data assimilation of precipitation observations. *Quart. J. Roy. Meteorol. Soc. Ser. A* **126** 339–360.

EVENSEN, G. and VAN LEEUWEN, P. J. (2000). An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review* **128** 1852–1867.

FISHER, M. and COURTIER, P. (1995). Estimating the covariance matrices of analysis and forecast error in variational data assimilation. Technical Memorandum 220, European Center for Medium-Range Weather Forecasting, Reading, UK.

HASSELMANN, K. F. (1988). PIPS and POPS: The reduction of complex dynamical systems using Principal Interaction and Oscillation Patterns. *J. Geophys. Res. Atmospheres* **93** 11,015–11,021.

JOLLIFFE, I. T. (1986). *Principal Component Analysis*. Springer, New York.

KENNEDY, M. C. and O'HAGAN, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** 425–464.

KITAGAWA, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.* **5** 1–25.

KOOPERBERG, C. and O'SULLIVAN, F. (1996). Predictive oscillation patterns: A synthesis of methods for spatial-temporal decomposition of random fields. *J. Amer. Statist. Assoc.* **91** 1485–1496.

KULIGOWSKI, R. J. and BARROS, A. P. (1998). Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks. *Weather and Forecasting* **13** 1194–1204.

LEVINE, R. A. and BERLINER, L. M. (1999). Statistical principles for climate change studies. *J. Climate* **12** 564–574.

LORENZ, E. N. (1963). Deterministic nonperiodic flow. *J. Atmospheric Sci.* **20** 130–148.

LU, Z. Q., BERLINER, L. M. and SNYDER, C. (2000). Experimental design for spatial and adaptive observations. In *Studies in the Atmospheric Sciences*. *Lecture Notes in Statist.* **144** 65–78. Springer, New York.

SACKS, J., WELCH, W. J., MITCHELL, T. J. and WYNN, H. P. (1989). Design and analysis of computer experiments (with discussion). *Statist. Sci.* **4** 409–435.

SALBY, M. L. (1996). *Fundamentals of Atmospheric Physics*. Academic Press, San Diego, CA.

SIVILLO, J. K., AHLQUIST, J. E. and TOTH, Z. (1997). An ensemble forecasting primer. *Weather and Forecasting* **12** 809–818.

SNEDDON, G. (2000). A statistical perspective on data assimilation in numerical models. In *Studies in the Atmospheric Sciences*. *Lecture Notes in Statist.* **144** 7–21. Springer, New York.

SOLOW, A. R. and MOORE, L. (2000). Testing for a trend in a partially incomplete hurricane record. *J. Climate* **13** 3696–3699.

SOLOW, A. R. and MOORE, L. (2002). Testing for trend in North Atlantic hurricane activity, 1900–98. *J. Climate* **15** 3111–3114.

TRENBERTH, K. E., ed. (1992). *Climate System Modeling*. Cambridge Univ. Press.

VISLOCKY, R. L. and FRITSCH, J. M. (1995). Generalized additive models versus linear regression in generating probabilistic MOS forecasts of aviation weather parameters. *Weather and Forecasting* **10** 669–680.

VON STORCH, H. and ZWIERS, F. W. (1999). *Statistical Analysis in Climate Research*. Cambridge Univ. Press.

WEST, M. and HARRISON, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer, New York.

WIKLE, C. K., MILLIFF, R. F., NYCHKA, D. and BERLINER, L. M. (2001). Spatio-temporal hierarchical Bayesian modeling: Tropical ocean surface winds. *J. Amer. Statist. Assoc.* **96** 382–397.

XU, K.-M. and RANDALL, D. A. (1996). A semiempirical cloudiness parameterization for use in climate models. *J. Atmospheric Sci.* **53** 3084–3102.