PART 4. STATISTICAL METHODS AND MODELS

Statistical methods and models play a key role in the interpretation and synthesis of observed climate data and the predictions of numerical climate models. Important advances have been made in the development and application of both frequentist and Bayesian statistical approaches and, as noted previously, the methods yield similar results when either an uninformed prior is used for the Bayesian analysis or a very large dataset is available for estimation. Recent reviews of statistical methods for climate assessment are summarized, including procedures for trend detection, assessing model fit, downscaling, and data-model assimilation. Methods for hypothesis testing and model selection are presented, and emerging issues in statistical methods development are considered.

Levine and Berliner (1999) review statistical methods for detecting and attributing climate change signals in the face of high natural variations in the weather and climate, focusing on "fingerprint" methods designed to maximize the signal-to-noise ratio in an observed climatic dataset (Hasselmann, 1979; 1993). The climate change detection problem is framed in terms of statistical hypothesis testing and the fingerprint method is shown to be analogous to stepwise regression of the observed data (*e.g.*, temperature) against the hypothesized input signals (carbon dioxide concentrations, aerosols, *etc.*). Explanatory variables are added to the model until their coefficients are no longer statistically significant. The formulation and interpretation of the hypothesis test is complicated considerably by the complex spatial and temporal correlation structure of the dependent and explanatory variables, and Levine and Berliner discuss various approaches for addressing these concerns. The selection of the best filter for isolating a climate change signal within the natural climate record is shown to be equivalent to the determination of an optimal (most powerful) statistical test of hypothesis.

Solow (2003) reviews various statistical models used in atmospheric and climate science, including methods for:

- fitting multivariate spatial-time series models, using methods such as principal component analysis (PCA) to consider spatial covariance, and predictive oscillation patterns (PROPS) analysis and maximum covariance analysis (MCA) for addressing both spatial and temporal variations (Kooperberg and O'Sullivan, 1996; Salim *et al.*, 2005);
- identifying trends in the rate of occurrence of extreme events given only a partially observed historical record (Solow and Moore, 2000, 2002);
- downscaling GCM model predictions to estimate climate variables at finer temporal and spatial resolution (Berliner *et al.,* 1999; Berliner, 2003);
- assessing the goodness of fit of GCMs to observed data (McAvaney *et al.*, 2001), where goodness-of-fit is often measured by the ability of the model to reproduce the observed climate variability (Levine and Berliner, 1999; Bell *et al.,* 2000); and
- data assimilation methods that combine model projections with the observed data for improved overall prediction (Daley, 1997), including multi-model assimilation methods (Stephenson *et al.,* 2005) and extended Kalman filter procedures that also provide for model parameter estimation (Evensen and van Leeuwen, 2000; Annan, 2005; Annan *et al.*, 2005).

Zwiers and von Storch (2004) also review the role of statistics in climate research, focusing on statistical methods for identifying the dynamics of the climate system and implications for data collection, forecasting, and climate change detection. The authors argue that empirical models for the spatiotemporal features of the climate record should be associated with plausible physical models and interpretations for the system dynamics. Statistical assessments of data homogeneity are noted as essential when evaluating long-term records where measurement methods, local processes, and other non-climate influences are liable to result in gradual or abrupt changes in the data record (Vincent, 1998; Lund and Reeves, 2002). Statistical procedures are reviewed for assessing the potential predictability and accuracy of future weather and climate forecasts, including those based on the data-model assimilation methods described above. Zwiers and Storch offer that for the critical tasks of determining the inherent (irreducible) uncertainty in climate predictions vs. the potential value of learning from better data and models, Bayesian statistical methods are often better suited than are frequentist approaches.

Methods for Hypothesis and Model Testing

A well-established measure in classical statistics for comparing competing models (or hypotheses) is the likelihood ratio (LR), which follows from the common use of the maximum likelihood estimate for parameter estimation. For two competing models M_1 and M_2, the LR is the ratio of the likelihood or maximum probability of the observed data under M_1 divided by the likelihood of the observed data under M_2, with large values of the likelihood ratio indicating support for M_1. Solow and Moore (2000) applied the LR test to look for evidence of a trend in a partially incomplete hurricane record, using a Poisson distribution for the number of hurricanes in a year with a constant sighting probability over the incomplete record period. The existence of such a trend could indicate warming in the North Atlantic Basin, but based on their analysis, little evidence was apparent. In cases such as that above in which the LR tests models with the same parameterization and simple hypotheses are of interest, the LR is equivalent to the Bayes Factor, which is the ratio of the posterior odds of M1 to the prior odds of M1. That is, the Bayes Factor represents the odds of favoring M1 over M2 based solely on the data, and thus the magnitude of the Bayes Factor is often used as a measure of evidence in favor of M1.

An approximation to the log of the Bayes Factor for large sample sizes, Schwarz's Bayesian Information Criterion or BIC, is often used as a model-fitting criterion when selecting among all possible subset models. The BIC allows models to be evaluated in terms of a lack of fit component (a function of the sample size and mean squared error) and a penalty term for the number of parameters in a model. The BIC differs from the well-known Akaike's Information Criterion (AIC) only in the penalty for the number of included model terms. Another related model selection statistic is Mallow's Cp (Laud and Ibrahim, 1995). Karl *et al.* (1996) utilize the BIC to select among ARMA models for climate change, finding that the Climate Extremes Index (CEI) and the United States Greenhouse Climate Response Index (GCRI) increased abruptly during the 1970s.

Model uncertainty can also be addressed by aggregating the results of competing models into a single analysis. For instance, in the next section we report an estimate of climate sensitivity (Andronova and Schlesinger, 2001) made by simulating the observed hemispheric-mean near-surface temperature changes since 1856 with a simple climate/ocean model forced radiatively by greenhouse gases, sulfate aerosols and solar-irradiance variations. A number of other investigators have used models together with historical climate data and other evidence to develop probability distributions for climate sensitivity or bound estimates of climate sensitivity or other variables. Several additional efforts of this sort are discussed below in Section 5. An increasing number of these studies have begun to employ Bayesian statistical methods (*e.g.*, Epstein, 1985; Berliner *et al.*, 2000; Katz, 2002; Tebaldi *et al.*, 2004, 2005).

As noted in Katz (2002) and Goldstein (2006), Bayesian methods bring a number of conceptual and computational advantages when characterizing uncertainty for complex systems such as those encountered in climate assessment. Bayesian methods are particularly well suited for problems where experts differ in their scientific assessment of critical processes and parameter values in ways that cannot, as yet, be resolved by the observational record. Comparisons across experts not only help to characterize current uncertainty, but help to identify the type and amount of further data collection likely to lead to resolution of these differences. Bayesian methods also adapt well to situations where hierarchical modeling is needed, such as where model parameters for particular regions, locations, or times can be viewed as being sampled from a more-general (*e.g.*, global) distribution of parameter values (Wilke *et al.*, 1998). Bayesian methods are also used for uncertainty analysis of large computational models, where statistical models that emulate the complex, multidimensional model input-output relationship are learned and updated as more numerical experiments are conducted (Kennedy and O'Hagan, 2001; Fuentes *et al.*, 2003; Kennedy *et al.*, 2006; Goldstein and Rougier, 2006). In addition, Bayesian formulations allow the predictions from multiple models to be averaged or weighted in accordance with their consistency with the historical climate data (Wintle *et al.*, 2003; Tebaldi *et al.*, 2004, 2005; Raftery *et al.*, 2005; Katz and Ehrendorfer, 2006; Min and Hense, 2006).

Regardless of whether frequentist or Bayesian statistical methods are used, the presence of uncertainty in model parameters and the models themselves calls for extensive sensitivity analysis of results to model assumptions. In the Bayesian context, Berger (1994) reviews developments in the study of the sensitivity of Bayesian answers to uncertain inputs, known as robust Bayesian analysis. Results from Bayesian modeling with informed priors should be compared to results generated from priors incorporating more uncertainty, such as flat-tailed distributions, non-informative and partially informative priors. Sensitivity analysis on the likelihood function and the prior by consideration of both non-parametric and parametric classes is often called for when experts differ in their interpretation of an experiment or a measured indicator. For example, Berliner *et al.* (2000) employ Bayesian robustness techniques in the context of a Bayesian fingerprinting methodology for assessment of anthropogenic impacts on climate by examining the range of posterior inference as prior inputs are varied. Of note, Berliner *et al.* also compare their results to those from a classical hypothesis testing approach, emphasizing the conservatism of the Bayesian method that results through more attention to the broader role and impact of uncertainty.

Emerging Methods and Applications

While the suite of tools for statistical evaluation of climate data and models has grown considerably in the last two decades, new applications, hypotheses, and datasets continue to expand the need for new approaches. For example, more sophisticated tests of hypothesis can be made by testing probability distributions for uncertain parameters, rather than single nominal values (Kheshgi and White, 2001). While much of the methods development to date has focused on atmospheric-oceanic applications, statistical methods are also being developed to address the special features of downstream datasets, such as streamflow (Allen and Ingram, 2002; Koutsoyiannis, 2003; Kallache *et al.*, 2005) and species abundance (Austin, 2002; Parmesan and Yohe, 2003).

As models become increasingly sophisticated, requiring more spatial and temporal inputs and parameters, new methods will be needed to allow our limited datasets to keep up with the requirements of these models. Two recent examples are of note. Edwards and Marsh (2005) present a "simplified climate model" with a "fully 3-D, frictional geostrophic ocean component, an Energy and Moisture Balance atmosphere, and a dynamic and thermodynamic sea-ice model. . . representing a first attempt at tuning a 3-D climate model by a strictly defined procedure." While estimates of overturning and ocean heat transport are "well reproduced", "*m*odel parameters *we*re only weakly constrained by the data." Jones *et al.* (2006) present an integrated climate-carbon cycle model to assess the implications of carbon cycle feedback considering parameter and model structure uncertainty. While the authors find that the observational record significantly constrains permissible emissions, the observed data (in this case also) "proves to be insufficient to tightly constrain carbon cycle processes or future feedback strength with implication for climate-carbon cycle model evaluation." Improved data collection, modeling capabilities, and statistical methods must clearly all be developed concomitantly to allow uncertainties to be addressed effectively.

Box 4.1: Predicting Rainfall: An Illustration of Frequentist and Bayesian Approaches

Consider how we use probability theory in weather prediction. We have a vast storehouse of observations of temperature, humidity, cloud cover, wind speed and direction, and atmospheric pressure for a given location. These allow the construction of a classic or frequentist table of probabilities showing the observed probability of rainfall, given particular conditions. This underscores the fact that observations of a stable system permit the construction of powerful predictive models, even if underlying physical processes are not known fully.

So long as the same underlying conditions prevail, the predictive model based on historical weather will remain powerful. However, if an underlying factor does change, the predictive power of the model will fall and the missing explanatory variables will have to be discovered. For example, if an underlying condition for cloud stability and formation of rainfall change because of reduced air pollution that cause the concentration of cloud condensation nuclei (CCN) to decline, the historic observations will not provide as powerful a prediction of rainfall as before. Under such conditions it is useful to consider a Bayesian approach in which cloud condensation nuclei are considered a potential additional explanatory variable. We can start with the old model, then modify its probability of rainfall, given different concentrations of cloud condensation nuclei. With each observation, our prior estimates of rainfall will be modified eventually leading to a new more powerful model, this time inclusive of the new explanatory variable.

Ideally, we want the full distribution of rainfall in a location. This has proven difficult to do, using the frequentist method, especially when we focus on high impact events such as extreme droughts and floods. These occur too infrequently for us to use a large body of observations so we must "assume" a probability distribution for such events in order to predict their probability of occurrence. While it may be informed by basic science, there is no objective method defining the appropriate probability distribution function. What we choose to use is subjective. Furthermore, the determinants of rainfall have been more numerous than once believed, often varying dramatically even on a decadal scale. For example, in the mid twentieth century, it was thought possible to characterize the rainfall in any location from thirty years of observations. This approach used the meteorological data for the period: 1931 to 1960 to *define the climate norm* around the earth. By the mid-80s however, it was clear that that thirty-year period did not provide an adequate basis for predicting rainfall in the subsequent years. In short, we learned that there is no "representative" sample of data in the classical sense. What we have is an evolving condition where tele-connections such as El Nino Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO), as well as air pollution and other factors determine cloud formation, stability and rainfall.

As we gain experience with the complex of processes leading to precipitation, we also develop a sense of humility about the incomplete state of our knowledge. This is where the subjectivity in Bayesian statistics comes to the fore. It states explicitly that our predictions are contingent on our current state of knowledge and that knowledge will be evolving with new observations.

PART 4 REFERENCES

Allen, M.R., and W.J. Ingram, 2002: Constraints on the future changes in climate and the hydrological cycle. *Nature*, **419**, 224–232.

Andronova, N. and M.E. Schlesinger, 2001: Objective estimation of the probability distribution for climate sensitivity. *Journal of Geophysical Research*, **106(D19)**, 22605-22612.

Annan, J.D., 2005: Parameter estimation using chaotic time series. *Tellus A*, **57(5)**, 709-714.

Annan, J.D., J.C. Hargreaves, N.R. Edwards, R. Marsh, 2005: Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Modelling*, **8(1-2)**, 135-154.

Austin, M.P, 2002: Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecological Modelling*, **157(2)**, 101–118.

Bell, J., P.B. Duffy, C. Covey, L. Sloan and the CMIP investigators, 2000: Comparison of temperature variability in observations and sixteen climate model simulations. *Geophysical Research Letters*, **27(2)**, 261-264.

Berger, J.O, 1994: An overview of robust Bayesian analysis (with discussion). *Test*, **3(1)**: 5-124.

Berliner, L.M., 2003: Physical-statistical modeling in geophysics. *Journal of Geophysical Research*, **108(D24)**.

Berliner, L.M., J.A. Royle, C.K. Wilke, and R.F. Milliff, 1999: Bayesian methods in the atmospheric sciences. In: *Bayesian Statistics 6* [Bernardo, J.M., J. O. Berger, A. P. Dawid and A. F. M. Smith (eds.)]. Oxford University Press, 83-100.

Berliner, L.M., R.A. Levine, D.J. Shea, 2000: Bayesian climate change assessment. *Journal of Climate*, **13(21)**, 3805-3820.

Daley, R., 1997: Atmospheric data assimilation. *Journal Meteorological Society Japan*, **75(1B)**, 319-329.

Edwards, N.R. and R. Marsh, 2005: Uncertainties due to transport-parameter sensitivity in an efficient 3D ocean-climate model. *Climate Dynamics*, **24(4)**, 415-433.

Epstein, E.S., 1985: Statistical inference and prediction in climatology: a Bayesian approach. *Meteorological Monographs*, **20(42)**, American Meteorological Society, Boston, Massachusetts.

Evensen, G. and P.J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, **128(6)**, 1852-1867.

Fuentes, M., P. Guttorp, and P. Challenor, 2003: Statistical assessment of numerical models. *International Statistical Review*, **71(2)**, 201–221.

Goldstein, M., 2006: Subjective Bayesian analysis: Principles and practice. *Bayesian Analysis*, **1(3)**, 403-420.

Goldstein, M. and J. Rougier, 2006: Bayes linear calibrated prediction for complex systems. *Journal American Statistical Association*, **101**, 1132-1143.

Hasselmann, K., 1979: On the signal-to-noise problem in atmospheric response studies. In: *Meteorology of Tropical Oceans* [D.B. Shaw (ed.)]. Royal Meteorological Society, pp. 251-259.

Hasselmann, K., 1993: Optimal fingerprints for the detection of time-dependent climate change. *Journal of Climate*, **6(10)**, 1957-1971.

Jones, C.D., P.M. Cox, and C. Huntingford, 2006: Climate-carbon cycle feedbacks under stabilization: Uncertainty and observational constraints. *Tellus B*, **58(5)**, 603-613.

Kallache, M., H.W. Rust, and J. Kropp, 2005: Trend assessment: Applications for hydrology and climate research. *Nonlinear Processes in Geophysics*, **12**, 201-210.

Karl, T.R., R.W. Knight, D.R. Easterling, and R.G. Quayle, 1996: Indices of climate change for the United States. *Bulletin American Meteorological Society,* **77(2),** 279-292.

Katz, R.W, 2002: Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate Research*, **20**, 167-185.

Katz, R.W. and M. Ehrendorfer, 2006: Bayesian approach to decision making using ensemble weather forecasts. *Weather and Forecasting*, **21**, 220–231.

Kennedy, M.C. and A. O'Hagan, 2001: Bayesian calibration of computer models. *Journal of the Royal Statistical Society Series B*, **63(3)**, 425–464.

Kennedy, M.C., C.W. Anderson, S. Conti, and A. O'Hagan, 2006: Case studies in Gaussian process modelling of computer codes. *Reliability Engineering & System Safety*, **91(10-11)**, 1301-1309**.**

Kheshgi, H.S. and B.S. White, 2001: Testing distributed parameter hypotheses for the detection of climate change. *Journal Climate*, **14**, 3464-3481.

Kooperberg, C. and F. O'Sullivan, 1996: Predictive oscillation patterns: A synthesis of methods for spatial-temporal decomposition of random fields. *Journal American Statistical Association*, **91**, 1485-1496.

Koutsoyiannis, D., 2003: Climate change, the Hurst phenomenon, and hydrological statistics. *Hydrological Sciences Journal*, **48(1)**, 3-24.

Laud, P.W. and J.G. Ibrahim, 1995: Predictive model selection. *Journal of the Royal Statistical Society Series B*, **57(1)**, 247-262.

Levine, R.A. and L.M. Berliner, 1999: Statistical principles for climate change studies. *Journal of Climate*, **12**, 564-574.

Lund, R. and J. Reeves, 2002: Detection of undocumented changepoints: A revision of the two-phase regression model. *Journal of Climate*, **15(17)**, 2547-2554.

McAvaney, B.J., C. Covey, S. Joussaume, V. Kattsov, A. Kitoh, W. Ogana, A.J. Pitman, A.J. Weaver, R.A. Wood, and Z.C. Zhao, 2001: Model evaluation. In: *Climate Change 2001: The Scientific Basis*. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, C.A. Johnson (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881 pp.

Min, S.K. and S. Hense, 2006: A Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models. *Geophysical Research Letters*, **8(3)**.

Parmesan, C. and G. Yohe, 2003: A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Raftery, A.E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, **133(5)**, 1155–1174.

Salim, A., Y. Pawitan, and K. Bond, 2005: Modelling association between two irregularly observed spatiotemporal processes by using maximum covariance analysis. *Journal of the Royal Statistical Society, Series C*, **54(3)**, 555-573.

Solow, A.R. and L. Moore, 2000: Testing for a trend in a partially incomplete hurricane record. *Journal Climate*, **13**, 3696-3699.

Solow, A.R. and L. Moore, 2002: Testing for a trend in North Atlantic hurricane activity, 1900-98. *Journal Climate*, **15**, 3111-3114.

Solow, A.R., 2003: Statistics in atmospheric science. *Statistical Science*, **18(4)**, 422-429.

Stephensen, D.B., C.A.S. Coelho, F.J. Doblas-Reyes, and M. Balmaseda, 2005: Forecast assimilation: a unified framework for the combination of multi-model weather and climate predictions. *Tellus A*, **57(3)**, 253-264.

Tebaldi, C., L.O. Mearns, D. Nychka, and R.L. Smith, 2004: Regional probabilities of precipitation change: A Bayesian analysis of multimodel simulations. *Geophysical Research Letters*, **31**, L24213.

Tebaldi, C., R.L. Smith, D. Nychka, and L.O. Mearns, 2005: Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles. *Journal Climat*e, **18(10)**, 1524–1540.

Vincent, L.A., 1998: A technique for the identification of in homogeneities in Canadian temperature series. *Journal of Climate*, **11(5)**, 1094-1104.

Wilke, C.K., L.M. Berliner, and N. Cressie, 1998: Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, **5(2)**, 117-154.

Wintle, B.A., M.A. McCarthy, C.T. Volinsky, and R.P. Kavanagh, 2003: The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology*, **17**, 1579-1590.

Zwiers, F.W. and H. von Storch, 2004: On the role of statistics in climate research. *International Journal of Climatology*, **24(6)** 665-680.