Forecast of Pacific-Indian Ocean SSTs Using Linear Inverse Modeling

contributed by Cecile Penland[1], Klaus Weickmann[2] and Catherine Smith[1]

[1]*CIRES, University of Colorado, Boulder, Colorado*
[2]*Climate Diagnostics Center (CDC), ERL/NOAA, Boulder, Colorado*

Using the methods described in Penland and Magorian (1993) and in previous issues of this Bulletin (particularly the December 1992 and June 1993 issues), the sea surface temperature (SST) anomaly in the Niño 3 region (6°N-6°S, 90°-150°W), as well as the anomaly in the Niño 4 region (6°N-6°S, 150° W-160° E), are predicted. A prediction at lead time t is made by applying a statistically-obtained Green function $\mathbf{G}(t)$ to an observed initial condition consisting of SST anomalies (SSTAs) in the IndoPacific basin. Three-month running means of the temperature anomalies are used, the seasonal cycle has been removed, and the data have been projected onto the 20 leading empirical orthogonal functions (EOFs) explaining 73% of the variance. The Niño 3 region has an RMS temperature anomaly of about 0.7°C; the inverse modeling prediction method has an RMS error of about 0.5°C at a lead time of nine months and approaches the RMS value at lead times of 18 months to two years. The COADS 1950-79 climatological annual cycle has been removed.

The predicted IndoPacific SSTA patterns based on the Dec-Jan-Feb 1995-96 initial condition for the following Mar-Apr-May, Jun-Jul-Aug, and Sep-Oct-Nov 1996, and Dec-Jan-Feb 1996-97, are shown in Fig. 1 (contour interval is 0.2°C). Figure 2a shows the predictions (light solid lines) and verifications (heavy solid lines) of the Niño 3 anomaly for initial conditions Sep-Oct-Nov and Oct-Nov-Dec 1995, and Nov-Dec-Jan and Dec-Jan-Feb 1995-96. The 1-standard deviation expected error for the prediction based on the Sep-Oct-Nov 1995 initial condition is denoted by dashed lines. Figure 2b is the same, but for the Niño 4 region. Verification and prediction do not exactly coincide at zero lead time since SSTAs are projected onto 20 EOFs for the prediction and the truncation error is included in the verification.

Consistent with the forecast published in the December 1995 issue of the Bulletin, this prediction calls for a decay of cold anomalies in the next few months. Warm anomalies are predicted to grow in the southeastern tropical Pacific and extend northward. The rapid decay of the predicted anomalies at lead times greater than six months is an indication of the uncertainty of the prediction at those lead times given current initial conditions.

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

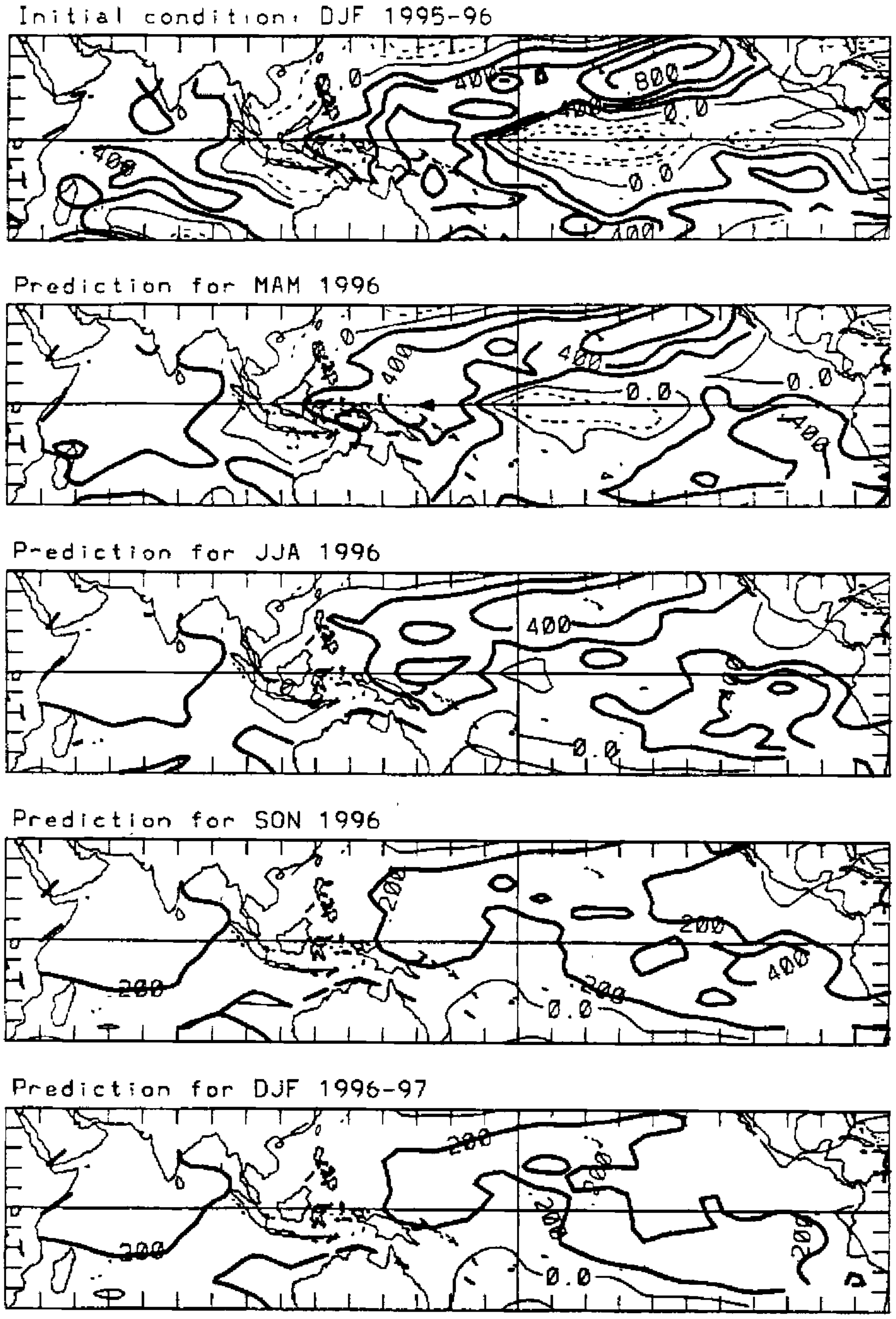
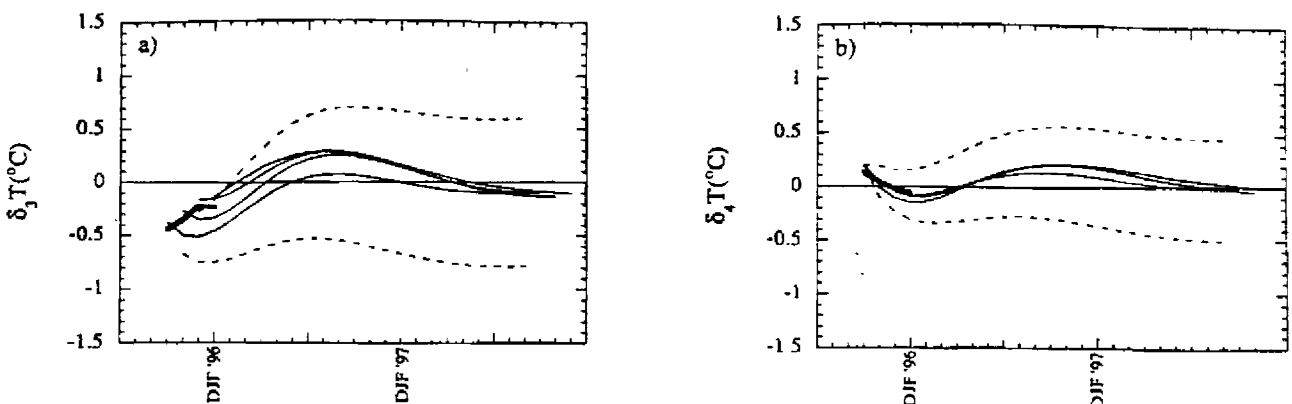


Fig. 1. *Linear inverse modeling forecasts of SST anomalies, relative to the standard 1950–79 COADS climatology both for the training period (1950–84) and for these forecasts. Forecast anomalies are projected onto 20 leading EOFs, based on Dec–Jan–Feb 1995–96 initial conditions (top panel). Contour interval is 0.2°C. Positive anomalies are represented by heavy solid lines, negative anomalies by dashed lines. SST anomaly data have been provided by NCEP, courtesy of R.W. Reynolds. Prediction by linear inverse modeling is described in Penland and Magorian (1993).*

Fig. 2 (below). *(a): Prediction (light solid lines) and verification (heavy solid line) of the Niño 3 SSTA based on initial conditions Sep–Oct–Nov and Oct–Nov–Dec 1995, and Nov–Dec–Jan and Dec–Jan–Feb 1995–96. Dashed lines denote one standard deviation prediction error bars appropriate to a stable linear system driven by stochastic forcing for the Sep–Oct–Nov 1995 initial condition. Verification and prediction do not exactly coincide at zero lead time since SSTAs are projected onto 20 EOFs for the prediction and the truncation error is included in the verification. (b): As in (a) except for Niño 4.*



Forecasts of Tropical Pacific SST Using a Data Assimilating Neural Network Model

contributed by Benyang Tang, William Hsieh and Fred Tangang

Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, B.C., Canada
Web site of the UBC Climate Prediction Group: **http://www.ocgy.ubc.ca**

A neural network model has been developed for forecasting the tropical Pacific SST in the Niño 3 region. Based on our earlier neural network models (Tang et al. 1994; Tangang et al. 1996), this current model has a number of new techniques added to better deal with noisy data.

Normally, when a neural network is trained, only the network weights are adjusted to minimize a cost function which measures only the differences between the network output and the data. In our data assimilating neural network, not only the weights, but also the network input are adjusted. The cost function to be minimized consists of three terms. **The first term** is the cost function of a traditional neural network, measuring the difference between the network output and the data (the output constraint). This is simply the error of the prediction. **The second term** measures the difference between the network input and the raw data (the input constraint). It was proposed by Weigend et al (1996), and was termed "clearning", after the words "learning" and "cleaning", meaning that the neural network learns from the data and cleans the data at the same time. Thus, the data are modified each time a training cycle is performed, based on the assumption that the raw predictor data contain some errors. "Clearning" makes the input data more compatible with the model, alleviating "transient growth" (Blumenthal 1991), somewhat similar to normal mode initialization reducing initial gravity wave propagation with primitive equations in numerical weather prediction. **The third term** measures the difference between the network output and the network input for the next step. It acts as a weak constraint of continuity, forcing the end of one step to be close to the beginning of the next step. This term is usually smaller than the first and second terms, as the first two involve the noisy raw data and the third term contains only the smoothed model input and output. During training, the input for each step is the raw input data (first training cycle) or the cleaned data (from "clearning", for subsequent training cycles). When training is finished, the forecast starts from the network output for the starting month obtained in the training, instead of from the raw data, similar to initialization by adjoint data assimilation. In a forecasting exercise, each step is not a separate entity as in a training cycle--rather, it is a multiple-step application of the trained neural network, with no exposure to the raw or cleaned data between steps.

The data used for training are the Nino 3 SST index and the first 4 EOF coefficients of the FSU monthly wind stress data (Goldenberg and O'Brien 1981). The seasonal cycle, calculated from the 1961-90 data, has been removed from the Niño 3 data. Before the EOF calculation, the wind data were first smoothed with one pass of a 1-2-1 filter in zonal and meridional directions and in time, and detrended and de-seasoned by subtracting from a given month the average of the same calendar months of the previous four years. This pre-EOF processing is the same as that used in Lamont's coupled model (Cane et al. 1986) and in Tang (1995).

The inputs of the neural network for a given month consist of the Niño 3 index and the first 4 wind EOF coefficients of the month and the same 5 numbers for the month that is 3 months earlier, amounting to 10 inputs to the network. These inputs feed into a hidden layer with 4 sigmoidal neurons, which in turn feed into 5 linear output neurons, giving the Niño 3 and the first 4 wind EOF coefficients for the month that is 3 months later. Thus, the time step of the neural network is 3 months. By repeatedly feeding forward the model output as input to the neural network, we can obtain forecasts for longer lead times. The skill of this multiple-step forward feeding is a good check of the predictive power of the neural network.

The neural network has 69 weights to be adjusted: 10 x 4 between input and hidden layer, 4 x 5 between hidden layer and output, and 4 + 5 for the two respective bias vectors. There are 420 training pairs (i.e., sets of predictors and predictands) in the 1961-95 period. (The number of training pairs is smaller for the retroactive real time forecasts described later.) To prevent overfitting, we implemented a termination scheme. For every 5 training iterations, the training is paused and the neural network is fed forward repeatedly to make hindcasts. The average correlation skill of the 3rd step and the 4th step (9 months and 12 months forward, respectively) is calculated. This long-term skill usually increases with training to a maximum (at about 80 to 100 iterations) but then starts to decrease. The training is terminated at this maximum point, even though the one-step error measured by the cost function is still decreasing

To estimate the forecast skill, retroactive real time forecasts for January 1986 to September 1995 were carried out, entailing a total of 118 neural network trainings, one for each month. Figs. 1 and 2 show the correlation skill and the RMS error for the retroactive real time forecast (+) from 1986 to 1995, and the hindcast (x) and persistence forecasts (o) for the whole period (1961-1995). The outputs obtained in the training are used to start the feed forward, so that at the initial time the correlation <1.00 and the RMS error >0.00. The forecast skills are higher than the hindcast skills, largely because the former includes only the more recent years which are less difficult to predict. (Other models also tend to give higher skills in the '80s and the '90s than in the '60s and '70s.) In fact, for identical periods the hindcasts performed here would be expected to outperform the retroactive real-time forecasts, because the hindcasts are based on training that includes the year being forecast--i.e. it is a dependent sample skill estimate that includes some artificial skill. Due to the 1-2-1 filter in time, the initial condition contains information of the next month. Thus, in Figs. 1 and 2, a 3-month lead skill should be interpreted as a 2-month lead skill, and so forth. The skills shown here exceed those realized for the same data using traditional (non-"clearning") neural nets, and for linear regression algorithms.

Fig. 3 shows the latest forecast using a neural network trained with data up to January 1996. Six forecasts of lead times of up to 18 months were initiated from July to December 1995. All 6 initial conditions were obtained from one neural network training. The forecasts starting from July and October 1995 predicted a return to normal conditions by the end of 1996, while the other four forecasts predicted considerable warming in the 96-97 winter.

Blumenthal, M.B., 1991: Predictability of a coupled ocean-atmosphere model. *J. Climate*, **4**, 766-784.

Cane, M.A., S.E. Zebiak and S. Dolan, 1986: Experimental forecasts of El Nino. *Nature*, **321**, 827-832.

Goldenberg, S.B., and J.J. O'Brien, 1981: Time and space variability of tropical Pacific wind stress. *Mon. Wea. Rev.*, **109**, 1190-1207.

Tang, B., 1995: Periods of linear development of the ENSO cycle and POP forecast experiments. *J. Climate*, **8**, 682-691.

Tang, B.,G. Flato and G. Holloway, 1994: A study of Arctic sea ice and sea level pressure using POP and neural network methods. *Atmos.-Ocean*, **32**, 507-529.

Tangang, F.T., W.W. Hsieh and B. Tang, 1996: Forecasting the equatorial Pacific see surface temperatures by neural network models. *Climate Dynamics*, submitted.

Weigend, A.S, H.G. Zimmermann, and R. Neuneier, 1996: Clearning. In *Neural Networks in Financial Engineering*. Refenes, P., Y. Abu-Mostafa, J.E. Moody and A.S. Weigend, Eds. *Proceedings, Neural Networks in the Capital Markets*, October 1995, London, UK. In press.

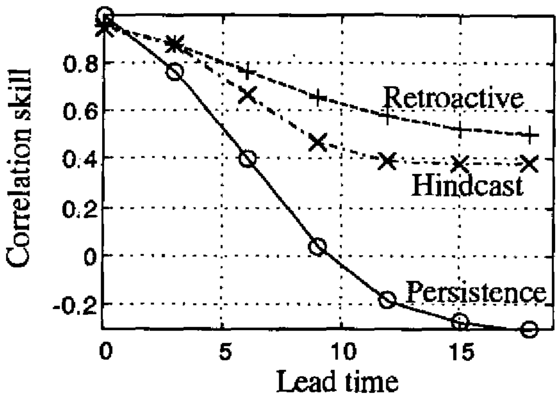


Fig. 1. *Correlation skills for retroactive real time forecasts (+), hindcasts (x), and persistence forecasts (o).*

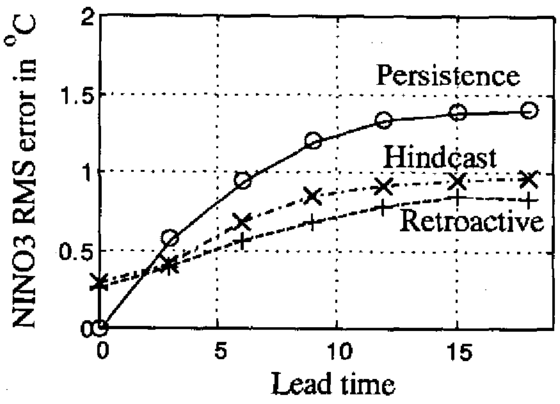


Fig. 2. *As in Fig. 1, except for root–mean–square error.*

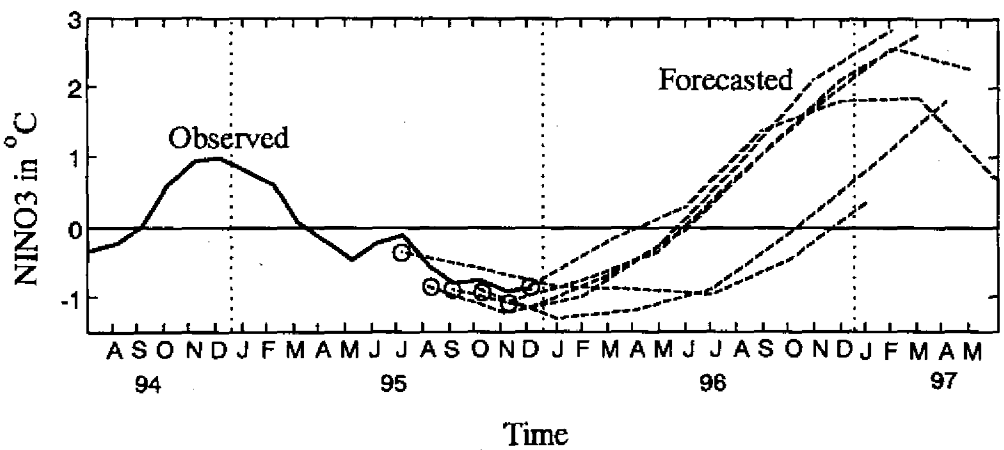


Fig. 3. *Forecasts of Niño 3 SST based on wind stress and SST data through December 1995. The solid line denotes the observed SST, and the 6 dashed lines the forecasts up to lead times of 18 months initiating from July to December 1995.*

Analogue (Non-Linear) Forecasts of the Southern Oscillation Index Time Series

contributed by Wasyl Drosdowsky

Bureau of Meteorology Research Center, Melbourne, Australia

An analogue selection procedure, based on the non-linear time series forecasting technique of Sugihara and May (1990), is applied to the Southern Oscillation Index (Drosdowsky 1994).

The time series to be forecast x_t is "embedded" in an E dimensional space defined by a sequence of lagged coordinates (x_t, x_{t-g}, x_{t-2g},..., $x_{t-(e-1)g}$), where g is the lag interval, usually taken as one time step. The E+1 closest neighbors (analogues) to the current state, defined by the vector x_t, x_{t-g}, x_{t-2g},..., $x_{t-(e-1)g}$, are found and used to construct the smallest simplex containing the current state. Future states of the system are found by projecting each analogue forward nT, where n=1,2,..., time steps and taking a suitably weighted average of the analogues. The optimal embedding dimension E is determined by a trial and error procedure, using the library of patterns formed by the first half of the time series to predict the evolution at each point of the last half of the time series. This effectively determines the window over which the analogue is selected.

The forecast system has been tested on time series with known properties. For the SOI, the optimal embedding dimension is found to be of order 9 to 12. The operational scheme has been used in the monthly Seasonal Climate Outlook issued by the National Climate Centre of the Australian Bureau of Meteorology since mid-1991. Analogues are selected from the entire available SOI time series from 1876 to the present time. An element of persistence is included in the forecast by adjusting the weighted analogue so that the t=0 value agrees with the current observed base value.

The skill of the analogue system has been examined in hindcast experiments (Drosdowsky 1994), and is shown in Fig. 9-1 in the September 1994 issue of this Bulletin. For RMSE the one time step forecasts are approximately equal to persistence while the two or more time step forecasts are more skillful than persistence within the appropriate range of embedding dimension. The spread of the analogues during the forecast period can provide a measure of the confidence level of the forecast.

Beginning with the forecast that appeared in the December 1994 issue, an improved SOI data set has been used. It covers the same Jan. 1876-present period as before, but periods of missing data have been filled. Information on the new data set can be obtained from Rob Allan (rja@dar.csiro.au).

Figure 1 shows the analogue forecast starting from February 1996 and extending through May 1996. The SOI has continued to hover close to zero for the past 3 months. The selected analogues all show similar behavior over the analogue selection period (June 1995 to February 1996) and exhibit similar spread over the forecast period, compared to the forecast issued in December. The analogue forecast shows a weak downward trend from a small positive SOI value in March to weak negative values in April and May. Forecast values for the next three months (in SD units X 10) are:

March 1996	**3.2**
April 1996	**-1.7**
May 1996	**-5.4**

Verification of the forecasts for the previous three months:

December 1995	F= 1.5	V= -5.5
January 1996	F= 2.3	V= 8.4
February 1996	F= 8.4	V= 1.0

Figure 2 shows the analogue forecast starting from January 1996 (one month earlier than for Fig. 1), and Fig. 3 starting from December 1995. While the forecasts from these three start times are not greatly dissimilar (in fact, some of the same years are seen to have been selected for all three starting months), a forecast for a rising SOI has tended to change to one of a falling SOI as the observed starting point has increased from what it was in December. The forecasts beginning from December and February have considerable internal spread. The verification of the two previous months' forecasts (Figs. 2, 3) was good for January and mediocre for December.

Drosdowsky, W., 1994: Analogue (non-linear) forecasts of the Southern Oscillation Index time series. *Wea. Forecasting*, **9**, 78-84.

Sugihara, G. and R.M. May, 1990: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, **344**, 734-741.

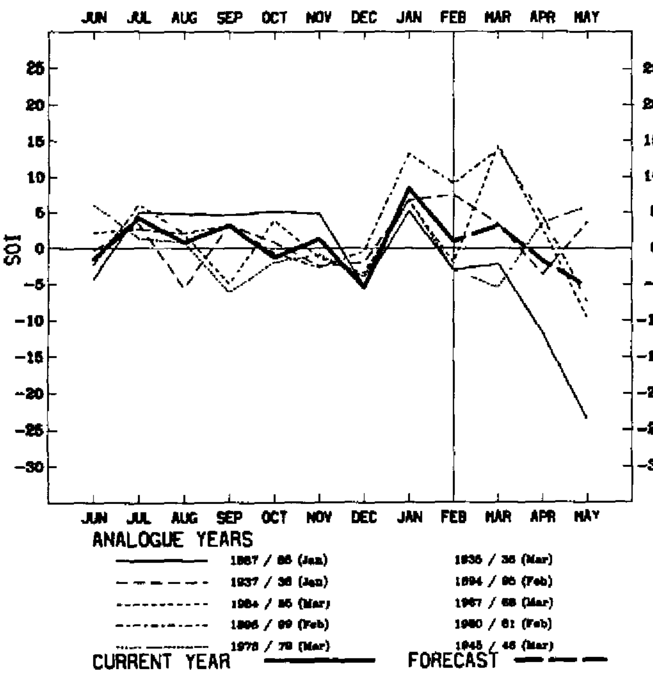


Fig. 1. *Selected analogues and forecasts based on the SOI up to February 1996. Points corresponding to the January, February or March initial condition have been used for selecting possible analogues. For clarity, only the best five analogues are plotted (light dashed or dotted lines), labeled with the year and month corresponding to the current month. (The remaining five analogues are listed to the right.) Heavy solid and dashed curves show the current and forecast values.*

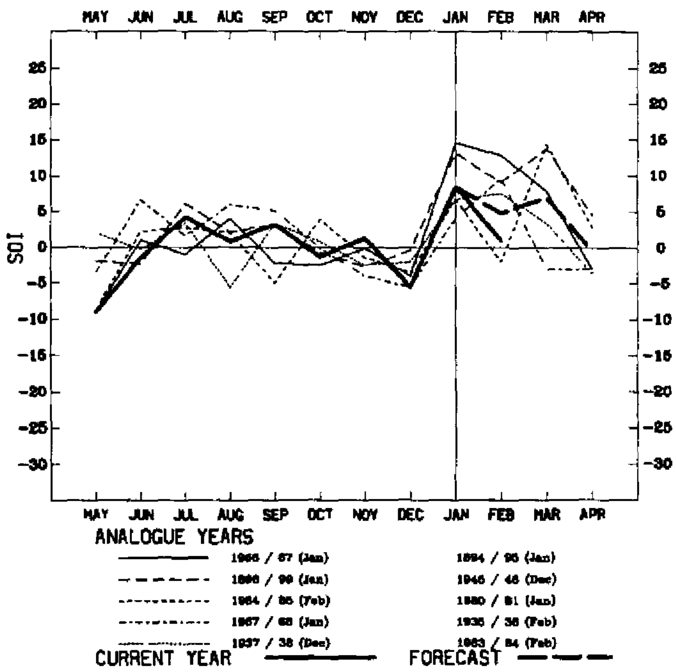


Fig. 2. *As in Fig. 1, except based on the SOI up to January 1996. The verifying value for February is indicated.*

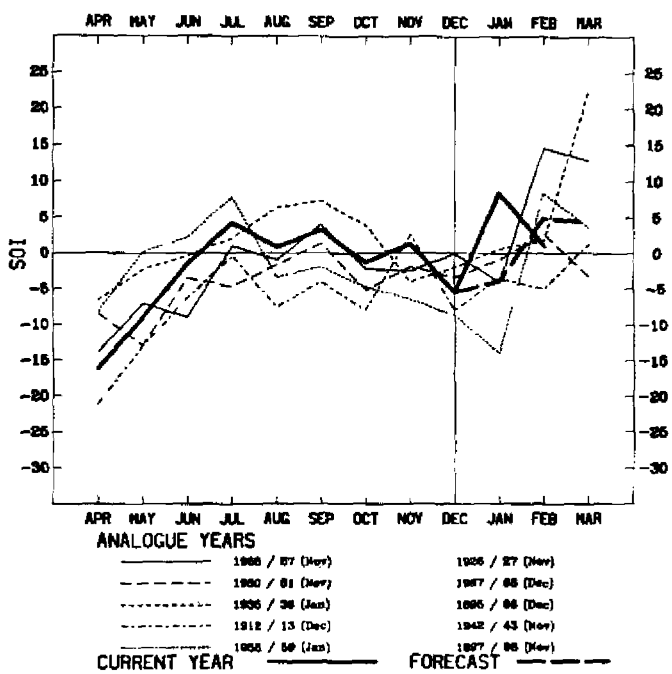


Fig. 3. *As in Fig. 1, except based on the SOI up to December 1995. The verifying values for January and February are indicated.*

A Probabilistic Model of the Number of Intense Atlantic Hurricanes for 1996

contributed by James Elsner

Department of Meteorology, Florida State University, Tallahassee, Florida

A statistical model of seasonal hurricane numbers is obviously an oversimplification of complicated physics and dynamics. Thus it is a mistake to assume that any model is a true representation of the underlying processes. Moreover, often information for making a forecast is available from several different sources.

An approach to probabalistic (or Bayesian) prediction for annual hurricane activity is to first think about and assume some prior information about what to expect for the upcoming season and then calculate the posterior distribution of this expectation. From this follows a predictive distribution for hurricane activity. Here we describe and apply a Bayesian prediction model for seasonal Atlantic-basin intense hurricane activity.

The prediction of seasonal hurricane activity has been given recent attention. The work of Nicholls (1979) in developing regression models of annual hurricane numbers near Australia has been followed by Gray et al. (1992) and Elsner and Schmertmann (1993) for hurricane activity in the Atlantic basin. The historical review of Hess and Elsner (1994) contains many important references to work in this area for Atlantic storms.

Applications of Bayesian statistics to problems in climatology are discussed in Epstein (1985). The problem of detecting climate change from historical time records using a Bayesian approach is presented in Solow (1988).

Let $\Theta = \{\theta_0, \theta_1, \theta_2, \ldots,\}$ be the set of all possible numbers of intense hurricanes in a given year, where θ_0 denotes the occurrence of no intense hurricanes, θ_1, the occurrence of one intense hurricane, etc. Now we have two sources of information. We have estimates $\pi_1(\theta_0), \pi_1(\theta_1), \ldots$ from an expert (Gray et al.) where $\pi_1(\theta_i)$ is the forecast probability of θ_i for a chosen year. Secondly, we have a Poisson regression model (Elsner and Schmertmann 1993) with maximum likelihood criterion. The model can be expressed as

$$\pi_2 = \exp(\gamma_0 + \sum_{i=1}^{5} \gamma_i x_i).$$

The five predictors (x_1, x_2, ..., x_5) include a 10-month forward extrapolation of the 50 mb and 30 mb zonal winds, the magnitude of the vertical shear of these winds, and the average rainfall anomalies (expressed in standard deviations) from the Gulf of Guinea and Sahel regions of west Africa. These are the predictors originally suggested by Gray et al. (1992). The problem is to process these two pieces of information to make the best possible forecast of the number of intense hurricanes (θ).

To construct an overall probabality model for this situation, it makes sense to focus on modeling the performance of the expert. Thus we seek to determine densities $f(p|\theta)$ reflecting the probability p that the expert would be likely to provide under each situation. For example, suppose we review Gray et al.'s past predictions and find that when the annual numbers of intense hurricanes were θ, his predictions p followed a distribution $f(p|\theta)$. Once $f(p|\theta)$ is specified, Bayes's theorem can be applied to the problem to obtain

$$\pi(\theta_0|p) = \frac{f(p|\theta_0)\pi_2(\theta_0)}{\sum_{i=0}^{n} f(p|\theta_i)\pi_2(\theta_i)}$$

similarly

$$\pi(\theta_1|p) = \frac{f(p|\theta_1)\pi_2(\theta_1)}{\sum_{i=0}^{n} f(p|\theta_i)\pi_2(\theta_i)}$$

or

$$\pi(\theta_j|p) = \frac{f(p|\theta_j)\pi_2(\theta_j)}{\sum_{i=0}^{n} f(p|\theta_i)\pi_2(\theta_i)},$$

for $j = 0, 1, 2, \ldots, n$.

The above modeling process is a viable way to proceed. The crucial factor in evaluating Gray et al.'s predictions is the skill of their previous predictions, and anything short of probabilistic modeling of this skill is likely to be inadequate (Berger 1985).

Now we describe a prediction for the upcoming hurricane season. The following table gives the performance of Gray et al.'s forecasts over the past six years (since they began issuing forecasts).

Year	Predicted θ	Observed θ
1995	3	5
1994	2	0
1993	3	1
1992	1	1
1991	1	2
1990	3	1

From this we can sketch a rough prior distribution $f(p|\theta)$ where

$$f(p|\theta) = \begin{cases} 2, & \text{if } \theta=0; \\ 2.33, & \text{if } \theta=1; \\ 1, & \text{if } \theta=2; \\ 2, & \text{if } \theta=3; \\ 3, & \text{if } \theta=4. \end{cases}$$

Input data for the Poisson regression model were obtained over the network from Gray et al.'s site http://typhoon.atmos.colostate.edu /forecasts on December 5, 1995. Regression coefficients for the number of intense hurricanes, estimated from the 1950-95 data, along with the new predictor values are given in Elsner and Schmertmann (1995). The predictor and predictand data sets used to estimate these coefficients are available via our anonymous ftp on metlab1.met.fsu.edu in directory /pub/elsner/Dec1fcst.

The estimated probabilities ($\pi_2(\theta_i)$) for each possible number of intense hurricane are given in the table below.

θ_i	0	1	2	3	4
$\pi_2(\theta_i)$	.259	.350	.236	.107	.036

From this we can apply Bayes's theorem in a straightforward manner.

$$\sum_{i=0}^{4} f(p|\theta_i)\pi_2(\theta_i) =$$

$$2(0.259) + 2.33(0.350) + 1(0.236) + 2(0.107) + 3(0.036) = 1.8915$$

so that,

$$\pi(\theta_0|p) = \frac{f(p|\theta_0)\pi_2(\theta_0)}{\sum_{i=0}^{4} f(p|\theta_i)\pi_2(\theta_i)}$$

so,

$$\pi(\theta_0|p) = \frac{2(0.259)}{1.8915} = 0.274$$

$$\pi(\theta_1|p) = \frac{2.33(0.350)}{1.8915} = 0.431$$

$$\pi(\theta_2|p) = \frac{1(0.236)}{1.8915} = 0.125$$

$$\pi(\theta_3|p) = \frac{2(0.107)}{1.8915} = 0.113$$

$$\pi(\theta_4|p) = \frac{3(0.036)}{1.8915} = 0.057$$

The above represents the posterior distribution of θ. Under very general conditions the mean of the posterior distribution minimizes the Bayes risk when the loss function is quadratic (squared difference). The posterior probabilities are a combination of prior knowledge and statistical evidence (Epstein 1985). In the situation where the prior knowledge is useless, the statistical evidence remains uninfluenced.

Based on the above analysis the probabilistic model estimates a mean of 1.248 intense hurricanes and a probability exceeding 70% of fewer than 2 major storms for 1996. Note this mean is slightly less than the mean of the Poisson distribution (1.352) for 1996 (Elsner and Schmertmann 1995) and is less by nearly one storm from Gray's official forecast, which calls for 2 intense hurricanes for 1996 (Gray 1995).

The underlying philosophy of this work could be extended by noting that π_2 depends solely on the intensity of the Poisson process (that is, the parameter λ). Accordingly, it is consistent to assume that λ is known only in terms of probability statements so that we can perform a Bayesian analysis to determine

π_2 by first assuming some prior distribution for λ. For the Poisson process there exists identical formulas (conjugate distributions) for expressing judgements about λ before and after reviewing the data (Epstein 1985).

Acknowledgments. Partial support for this work came from the Risk Prediction Institute and from NOAA through the Cooperative Institute on Tropical Meteorology (CITM). The suggestion of a Bayesian approach to seasonal hurricane predictions was made by A. Solow. The author is grateful to T. Barnston for his suggestions on improving the presentation.

Berger, J. D., 1985: *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, 617 pp.

Elsner, J. B. and C. P. Schmertmann, 1993: Improving extended-range seasonal predictions of intense Atlantic hurricane activity. *Weather and Forecasting*, **8**, 345–351.

—, 1995: Multiple least-squares regression and Poisson model forecasts of Atlantic tropical storm activity for 1996. *Exp. Long-Lead Fcst. Bull.*, **4**, 4:19–20.

Epstein, E. S., 1985: *Statistical Inference and Prediction in Climatology: A Bayesian Approach.* American Meteorological Society, Boston, 199 pp.

Gray, W. M., 1995: LAD multiple linear regression forecasts of Atlantic tropical storm activity for 1996. *Exp. Long-Lead Fcst. Bull.*, **4**, 4:21–22.

—, C. W. Landsea, P. W. Mielke Jr., and K. J. Berry, 1992: Predicting Atlantic seasonal hurricane activity 6-11 months in advance. *Weather and Forecasting*, **7**, 440–455.

Hess, J. C. and J. B. Elsner, 1994: Historical developments leading to forecasts of annual tropical-cyclone activity. *Bull. Amer. Met. Soc.*, **75**, 1611–1621.

Nicholls, N., 1979: A possible method for predicting seasonal tropical cyclone activity in the Australian region. *Mon. Wea. Rev.*, **107**, 1221–1224.

Solow, A. R., 1988: A Bayesian approach to statistical inference about climate change. *J. Climate*, **1**, 512–521.