Forecasting Drought over the Southwest U.S. using Climate Information

Bonnie K. Ray
Department of Mathematical Sciences and
Center for Applied Math and Statistics
New Jersey Institute of Technology
Newark, NJ

Upmanu Lall
Department of Civil & Environmental Engineering and
Utah Water Research Laboratory
Utah State University
Logan, UT

Balaji Rajagopalan, Edward Cook
Lamont-Doherty Earth Observatory of Columbia Univ., Palisades, NY

Abstract

We apply nonlinear forecasting methods to predict summer drought over the southwest region of the U.S. Our methods make use of teleconnections between summer drought indices and the winter El Niño Southern Oscillation (ENSO), as captured by the first few modes of global Sea Surface Temperatures (SST). They also allow for different relationships between drought and climate variables in different climate regimes. Several different nonlinear modeling procedures are presented and their forecast skill assessed over the last two decades. We find that, although methods incorporating the nonlinear relations between climate and drought do not necessarily provide superior forecasts for individual climate regions, these methods provide some improvement in drought pattern forecastability over the broad region of the southwest U.S.

1 Introduction

Manu and Balaji fill in

2 Description of Data

We use the following data to construct forecast models.

1. Seasonal Palmer Drought Severity Index (PDSI), as defined by the National Drought Mitigation Center (NDMC), for the twenty-five climate divisions making up Arizona, New Mexico, and Texas over the period 1895 - 1999.

2. Seasonal global SST anomalies on a 50×50 grid from a "reduced space" optimal smoother algorithm applied to 135 years (1865-1999) of global SST monthly anomalies obtained from the UK Hadley Center archives (Bottomley *et al.*, 1990; Kaplan *et al.*, 1998).

We perform all the analyses for the common period of 1895 - 1999.

The PDSI, a measure of moisture deficit, takes into account both temperature and precipitation, as well as drought information from previous seasons (Palmer, 1965). The PDSI is a standardized index that has comparable meaning at all locations. Negative values of the index imply drought-like (or moisture deficit) conditions and vice-versa. There are variations of the PDSI, such as the modified PDSI, the hydrologic drought index, *etc.*, but all such series are expected to behave similarly with regard to large-scale climate fluctuations.

We are particularly interested in the forecastability of the summer PDSI, defined as the average of the June-August PDSI values, using drought and climate information available only through the previous winter. Good forecasts of drought indicators can provide important information necessary for setting reasonable water management policies. Winter SSTs and PDSIs are defined as the December-February averages of SST at each grid point and at each climate region, respectively.

2.1 Principal Component Analysis

We would like to have some idea about the broad scale forecastability of drought for the entire Southwest US, thus we choose to analyze the first few principal components of the climate region field instead of analyzing each region separately. The method of principal components defines a rotation of variables, given by a particular linear combination of the variables, such that the first derived direction maximizes the variability in the data, the second maximizes the variability among directions uncorrelated with the first, *etc.* (REF). For the summer PDSI field, the first three principal components (PCs) explain more than ?% of the variability in the data. Table 1 gives the derived weights for each region for the first three PCs of summer PDSI over the years 1895-1979 ($n = 85$). We see that PC 1 represents the average summer PDSI over the entire region, whereas PC 2 represents the (FILL IN). PC analysis of the winter PDSI field yields a similar interpretation, with the first three PC's explaining more than ? % of the data variability.

The information in the SST field can also be summarized using PC analysis. For the SSTs, the first three PCs explain more than ? % of the variation in observed SSTs over the entire grid. The first PC of SST can broadly be interpreted as the variation due to ENSO. The second PC (FILL IN).

Since the PCs are orthogonal by definition, we investigate the forecastibility of the first two PCs of summer PDSI by modeling each PC individually as a function of the corresponding PC for winter PDSI and the PCs of the SST field. The next section describes the forecast methods used.

3 Nonparametric Forecasting Methods for Time Series

The analysis of Rajagopolan *et al.* (1999) indicates that teleconnections between summer PDSI and winter ENSO may be nonlinear, requiring different models to describe the relationships during periods of low versus high ENSO activity. We investigate the nature of this relationship

using nonparametric regression methods, which are useful for approximating a general, unspecified functional relationship among variables. For example, let $S_{1,t}$ denote the first PC of summer PDSI in year t, let $W_{1,t}$ denote the first PC of winter PDSI in year t, and let $N_{1,t}$ denote the first PC of the SST field for year t. A general model for $S_{1,t}$ as a function of $W_{1,t}$ and $N_{1,t}$ is

$$S_{1,t} = f(W_{1,t}, N_{1,t}) + \epsilon_t, \tag{1}$$

where ϵ_t represent mean zero, uncorrelated random influences on $S_{1,t}$ that are not captured by the model.

3.1 Kernel Regression

The method of kernel regression has been used extensively for local estimation of regression functions such as that of (1). In general, a kernel estimate of $Y = f(\boldsymbol{x})$ at point $\boldsymbol{x} = (x_1, \ldots, x_p)$ is a weighted average of observations Y_i whose corresponding predictor values X_i are in the neighborhood of $\boldsymbol{x}$, and is defined as

$$\hat{f}(x_1, \ldots, x_p) = \frac{\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{x} - \boldsymbol{X}_i) Y_i}{\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{x} - \boldsymbol{X}_i)}, \tag{2}$$

where $\boldsymbol{H}$ is a $p \times p$ symmetric positive definite matrix of *bandwidths*, which determine the size of the neighborhood, and

$$K_{\boldsymbol{H}}(\boldsymbol{x}) = |\boldsymbol{H}|^{-1/2} K(\boldsymbol{H}^{-1/2}\boldsymbol{x}).$$

Here, K is a p-dimensional kernel function satisfying $\int K(\boldsymbol{x})\, d\boldsymbol{x} = 1$, for example a p-dimensional density function. Kernel regression results in an estimate of f that is locally constant in a neighborhood of $\boldsymbol{x}$. In practice, a product of univariate kernels, such as univariate standard Gaussian density functions, is commonly used for $K(\cdot)$ and $\boldsymbol{H}$ is taken as a diagonal matrix. See, for example, Scott (1996) for an in-depth discussion of nonparametric regression techniques.

The locally constant kernel smoothing method can be extended to allow local polynomial estimation of f. For example, when the degree of the polynomial is one, an estimate of $f(\boldsymbol{x})$ is found using weighted least squares regression of Y_i on $(1, x_{i,1} - x_1, \ldots, , x_{i,p} - x_p)$ with weights determined by the particular kernel and bandwidth chosen. The method of k-nearest neighbors ($k - NN$) is another, closely related, way to obtain local estimates of Y_i, and can be thought of as a kernel regression in which the size of the local neighborhood around x is allowed to vary, thus providing a large window around $\boldsymbol{x}$ when the data are sparse. The method of k-nearest neighbors has been used extensively in hydrological and climatological modeling (REFS).

3.2 MARS

Allowing f to take any possible form using kernel estimation suffers from the "curse-of-dimensionality", *i.e.* the number of data points necessary to obtain a reliable estimate of f grows exponentially with p, the number of predictor variables. If f is constrained in such a way that it still provides a flexible representation of the unknown underlying function yet does not suffer from excessive data requirements, a more stable estimate may be obtained. One such constrained method that has been used successfully in a number of applications is the method of Multivariate Adaptive Regression Splines (MARS)(REFs), in which spline functions are used to approximate $f(\cdot)$ (Friedman, 1991). Splines are piecewise polynomials that maintain continuity of the target function

and selected derivatives across the pieces. The MARS method of fitting regression splines to the data uses a modified recursive partitioning technique to divide the predictor space into an optimal number of overlapping subregions, delineated by threshold points in specific predictors. A truncated, tensor product cubic spline basis function is then used to represent the function $f(\cdot)$. This entails fitting a cubic polynomial in up to p variables for each partition of the predictor space, with the constraint that the function values indicated by the cubic polynomials in different partitions agree at the partition boundaries. The process is easiest to explain in the univariate context. Consider a threshold X_t^* at which the two partitions are formed for the predictor X_t. The cubic polynomial fit (*e.g.*, $Y_t = a_0 + a_1 X_t + a_2 X_t^2 + a_3 X_t^3$) for the regression of Y_t on X_t for $X_t > X_t^*$ and the cubic polynomial fit for the regression of Y_t on X_t for $X_t < X_{t-1}^*$, when evaluated at $X_t = X_t^*$ need to yield the same value. Generally, a certain number of derivatives of each piece need to match at the threshold as well. For the analysis of this paper, we use a version of MARS that considers a truncated cubic spline basis, where only linear terms in each predictor and selected low order interactions (*i.e.*, cross-product across pairs or triplets of predictors) terms are retained. The truncated splines are represented using the notation $(x - t)_+ = (x - t)$ if $x \geq t$ and zero otherwise. If only linear terms are used in the model, the resulting regression surface is composed of planes that are hinged at the thresholds (or spline knots).

Nonlinearity involving variables entering in an additive fashion is consequently approximated through a sequence of approximating linear planes with the knots or thresholds defining these planes chosen by a predictive risk criteria. Non-additive aspects of nonlinearity, *e.g.*, product terms in predictor variables are represented similarly through low order piecewise approximations of these terms across the same series of knots in each variable. Interactions up to order 3, *i.e.* cross products involving up to 3 predictors at a time, are considered in the version of MARS used here. The resulting models can thus have multiple thresholds as well as interactions between predictor variables.

The ability to have interactions is actually of quite some interest, since it provides a mecahnism by which the modulation of a high frequency component, *e.g.* dependence of summer PDSI on winter PDSI can be modulated by a low frequency component such as interannual climate oscillation as represented by ENSO. We use the generalized cross-validation (GCV) criterion to select both the knots and the terms to retain in the MARS models for forecasting summer PDSI. See (REF) for a discussion of model selection criteria for time series models obtained using MARS.

Forecasts are obtained for MARS models by plugging observed values of the predictor variables in year t into the fitted models. Although no theoretical prediction intervals are available, empirical prediction intervals can be obtained using, *e.g.*, bootstrapping.

In the next section, we present models and forecasts for summer PDSI obtained using MARS.

4 Results

We use the first 85 years of data to model to the relationship between summer PDSI and winter PDSI, along with other climatic influences. The remaining 20 years are saved for forecast evaluation.

4.1 Estimated Models

As an example of the types of models obtained using MARS, the following model was estimated for the first PC of summer PDSI.

$$\begin{aligned} S_{1,t} &= 4.589923 - 0.3911(N_{1,t} + 14.083)_+ \\ &\quad 0.1238(W_{1,t} + 5.520)_+(-13.591 - N_{1,t})_+ + 0.0302(W_{1,t} + 5.520)_+(N_{1,t} + 13.591-)_+ \\ &\quad -0.0558(6.610 - W_{1,t})_+(3.931 - N_{1,t})_+ + 0.1332(-2.620 - N_{1,t})_+(-5.283 - N_{2,t})_+ \\ &\quad -0.0167(W_{1,t} - 1.800)_+(-2.620 - N_{1,t})_+(-5.283 - N_{2,t})_+ \end{aligned} \tag{3}$$

Although a direct physical interpretation of the model is difficult since we are working with the PC series, we can see that the summer PDSI is influenced by PC 1 of the SST field, as well as by PC 1 of winter PDSI modulated by PC 1 of SST, with different modulations dependent on the level of the STT PC index. There is also an interaction effect due to the interplay of the first and second SST PCs and an interaction effect due to interplay between all three predictor variables. Of these, an ANOVA decomposition of the fitted model indicates that the SST PC and the interactions between Winter PDSI and SST PC 1 account for most of the variability in PC 1 of Summer PDSI.

The MARS model for PC 2 is

$$\begin{aligned} S_{2,t} &= -0.03203 + 0.8165(W_{2,t} + 0.080)_+ \\ &\quad -0.0289(1.400 - W_{2,t})_+(N_{1,t+12.686})_+, \end{aligned} \tag{4}$$

indicating that the primary influence on summer PDSI 2 is winter PDSI 2, with additional effects due to an interplay between ENSO and winter PDSI conditions.

4.2 Forecast Evaluation

Summer PDSI forecasts are evaluated on the basis of both their in-sample and out-of-sample root mean square prediction errors (RMSEs) and on their forecast skill, as measured by the percentage of forecasts that take values in the same category as was actually observed. Here, the forecast categories are determined using the 33.33 and 66.66 percentiles of the summer PDSI PCs during the model fitting period.

Table 2 presents the forecast results for the MARS models of summer PDSI PC's, along with results obtained using a standard climatological forecast. The climatological forecast predicts summer PDSI using the previous winter PDSI value.

The MARS models provide improved in-sample forecasts of summer PDSI, but only marginally better predictions than the climatological forecasts over the last two decades. In fact, the climatology forecast have greater skill than the MARS model for PC 2. Figure 1 shows the observed and forecast values for PC 2 from 1980-1999. Neither forecast method is able to capture the extremes in the PC values, particularly in 1981, 1997, and 1998. (WHAT MORE CAN WE SAY ABOUT THIS??)

5 Summary and Discussion

Discuss the analysis of individual climate regions or MV models

Table 1: Forecasts Evaluations for Summer PDSI PCs

Component	PC 1				PC 1			
Method	MARS		Climatology		MARS		Climatology	
FC Measure	RMSE	Skill (%)	RMSE	Skill (%)	RMSE	Skill (%)	RMSE	Skill (%)
Hindcast (1895-1979)	1.76	71.76	2.47	67.06	1.43	63.53	1.52	60.00
Forecast (1980-199)	3.84	70.00	2.09	70.00	1.97	50.00	1.36	75.00

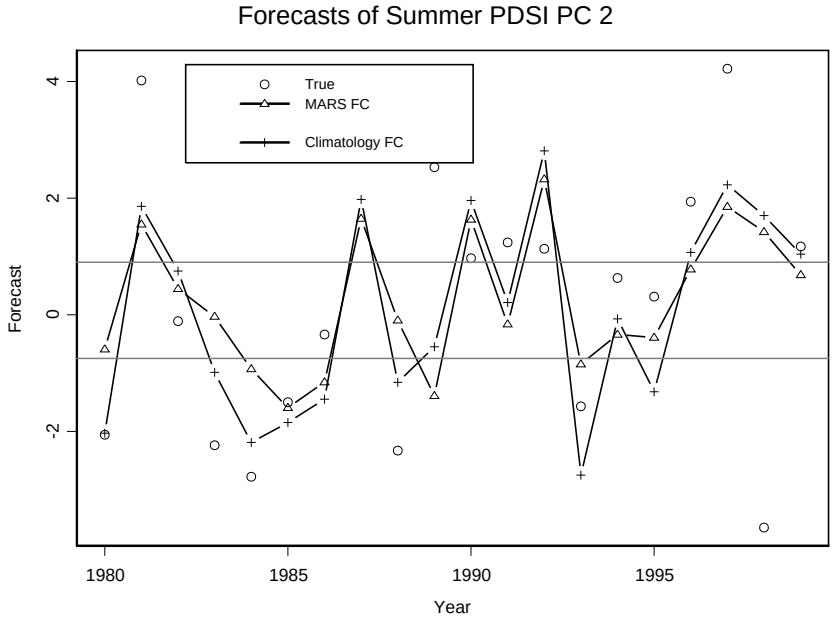

Figure 1: Comparison of forecasts over the last two decades for PC 2 of Summer PDSI in the Southwest U.S.

6 Acknowledgements

The research of all the authors was supported by a grant jointly funded by the National Science Foundation and the National Oceanographic and Atmospheric Association (??).

7 References

Finizio M. and Palmieri S. (1998). "Non-linear modelling of monthly mean vorticity time changes: an application to western Mediterranean" Annales Geophysicae 16, 116-124, Springer Verlag.