Probabilistic streamflow forecasts based on hydrologic persistence and large-scale climate signals in central Texas

Wenge Wei and David W. Watkins, Jr

ABSTRACT

Skillful streamflow forecasts at seasonal lead times may be useful to water managers seeking to provide reliable water supplies and maximize system benefits. In this study, streamflow autocorrelation and large-scale climate information are used to generate probabilistic streamflow forecasts for the Lower Colorado River system in central Texas. A number of potential predictors are evaluated for forecasting flows in various seasons, including large-scale climate indices related to the El Niño/Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), North Atlantic Oscillation (NAO) and others. Results indicate that, of the predictors evaluated, only hydrologic persistence and Pacific Ocean sea surface temperature patterns associated with ENSO and PDO provide forecasts which are statistically better than climatology. An ordinal polytomous logistic regression approach is proposed as a means of incorporating multiple predictor variables into a probabilistic forecast model. Forecast performance is assessed through a cross-validation procedure, using distribution-oriented metrics, and implications for decision making are discussed.

Key words | hydrologic persistence, large-scale climate signals, polytomous logistic regression, probabilistic streamflow forecasts

Wenge Wei
David W. Watkins, Jr (corresponding author)
Department of Civil and
   Environmental Engineering,
Michigan Technological University,
Houghton,
MI 49931,
USA
E-mail: *dwatkins@mtu.edu*

INTRODUCTION

Reliable streamflow forecasts with lead times of even one season can have a significant effect on the performance of reservoir operation policies and operation efficiency (Karamouz & Zahraie 2004; Sun *et al.* 2006). In recent years, much effort has been made to develop mid- to long-term (seasonal to annual) hydroclimatic and streamflow forecasting models for water management in the United States. For example, the NOAA Climate Prediction Center issues seasonal forecasts of temperature, precipitation and soil moisture, as well as a drought outlook, for the entire US. In the western US, the USDA Natural Resources Conservation Service provides streamflow forecasts in the first half of the year based on observed snowpack conditions. At many stream gauge locations throughout the US, the National Weather Service provides probabilistic seasonal flow forecasts through a procedure known as Ensemble Streamflow Prediction, or

ESP (Day 1985; Smith *et al.* 1992). Traditionally, each of the meteorology traces has been assumed to represent an equally likely scenario for the future; more recently, methods have been developed to condition the probabilities of the historical meteorological traces based on seasonal climate forecasts (e.g. Croley 2000; Duan *et al.* 2006). On a global scale, the NOAA/Columbia University International Research Institute for Climate and Society (IRI) is one institution that issues seasonal forecasts of temperature and precipitation.

Significant progress has been made in understanding the influence of large-scale ocean–atmospheric patterns, such as El Niño–Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), the North Atlantic Oscillation (NAO) and Atlantic Multidecadal Oscillation (AMO), on regional climate anomalies around world. Numerous studies have shown that statistical models incorporating large-scale

ocean–atmospheric patterns can improve the ability to forecast streamflow with long lead times (e.g. Hamlet & Lettenmaier 1999; Sharma 2000; Tootle & Piechota 2006). At least three previous studies have identified teleconnections for central Texas. Piechota & Dracup (1996) found strong correlation between the Southern Oscillation Index (SOI) and the Palmer Drought Severity Index (PDSI), indicating the potential of improved hydroclimatic forecasts, with up to one year in lead time, for the region. However, a strong relationship between SOI and streamflow was not found. One possible reason for this is that PDSI is a mathematical function of temperature and precipitation, and provides a general indication of drought, whereas streamflow tends to integrate climatic processes over interseasonal time scales, and this seasonal averaging may limit forecast accuracy. For instance, streamflow is a function of both surface runoff and groundwater discharge, and groundwater recharge and discharge processes often exhibit lag times markedly longer than those of rainfall–runoff processes (Alley 1985). Furthermore, groundwater basins seldom align with surface watersheds, which may confound statistical analyses of climate and streamflow variables measured at specific gauge locations.

In another study, Rajagopalan et al. (2000) found a correlation between summer PDSI and winter Pacific Ocean Sea Surface Temperature (SST) anomalies. However, they also found epochal variations in this correlation, with the period of 1963–1995 showing weaker teleconnections than the period 1895–1962. Of course, without a means of predicting these epochal shifts in teleconnections, such variation tends to confound statistical forecasting methods based on the entire historical record. It is widely hypothesized that interdecadal North Pacific variability modulates ENSO precipitation teleconnections (e.g. Gershunov & Barnett 1998), but Rajagopalan et al. (2000) were not able to conclude that either NAO or PDO has any effect on ENSO precipitation teleconnections in central Texas.

Finally, Tootle et al. (2005) completed a study of the influence of interdecadal, decadal and interannual oceanic–atmospheric influences on streamflow in the United States. Unimpaired streamflow was identified for 639 stations for the period 1951–2002, and the phases (cold/negative or warm/positive) of ENSO, PDO, NAO and AMO were identified for the year prior to the streamflow year (i.e. long lead time). Statistical significance testing of streamflow, based on the

interdecadal, decadal and interannual oceanic–atmospheric phase (warm/positive or cold/negative), indicated no spatially coherent teleconnections for central Texas. Although this study focused on a specific annual period (October–September) and a particular forecast lead time, it provides an indication that long-lead climate forecasts may not be useful to water managers in this region.

Streamflow forecasts may be either deterministic or probabilistic, but probabilistic methods are often preferable for water management because they can provide more information about uncertainty. Categorical streamflow forecasts are common, providing the probabilities of flow being in each of a number of categories (e.g. low, medium or high). Probabilities of each category could be generated directly or indirectly. Piechota et al. (1998) proposed linear discriminant analysis to produce the probabilities of each category of streamflow directly. This method involves nonparametric kernel density estimation of the probability density function for each flow category. Regonda et al. (2006) employed logistic regression to directly predict the probability of streamflow above a given threshold. They applied this approach to categorical forecasts of the spring (April–June) streamflow at six locations in the Gunnison River basin. However, this approach treats the response variable as binary, i.e. equal to 1 if the streamflow value exceeds a given threshold and 0 otherwise. A drawback of this approach for multiple categories is that the logistic regression needs to be repeated to obtain the probability corresponding to each category threshold, and the sum of probabilities is not guaranteed to equal 1.

In this paper, a statistical method called polytomous logistic regression for ordinal response (Kutner et al. 2004) is proposed to generate probabilistic forecasts with seasonal lead times for the Highland Lakes system in central Texas. In the method, the response variable (streamflow) has multiple discrete outcomes rather than binary. Further, the response categories (e.g. below normal, normal, above normal) could be considered as ordered, thus allowing a parsimonious and easily interpreted logistic model, called a proportional odds model, that may be employed to generate a probabilistic (multi-category) forecast using a single model. A number of distribution-oriented metrics, such as the Brier Skill Score and the Ranked Probability Skill Score, may be used to assess model performance (Wilks 1995). This is demonstrated for the Highland Lakes system in Texas for a number of potential

predictor variables, including streamflow autocorrelation (hydrologic persistence), sea surface temperatures and other large-scale climate signals.

CASE STUDY DATA

The Lower Colorado River Authority (LCRA) operates the Highland Lakes system in central Texas, a series of six lakes on the Lower Colorado River. As a water conservation and reclamation district, the LCRA provides the water supply and flood control to a 33-county area, including the City of Austin and several rice irrigation districts along the Texas Gulf Coast (see Figure 1). In addition, the LCRA produces wholesale power for a 53-county service area and provides water resources for lake recreation activities and in-stream flow maintenance. To meet rapidly growing demands, reliable reservoir inflow forecasts with seasonal lead times would potentially be very beneficial; however, hydrologic forecasts are not used by the LCRA for a number of reasons, including high seasonal and annual variability of streamflow, the absence of easily measured hydrologic indicators such as snowpack (Watkins & O'Connell 2005).

To explore the patterns of streamflow and the influence of teleconnections in central Texas, monthly streamflow data are acquired from two sources: (1) aggregate inflows to the Highland Lakes (upstream), based on USGS gauge measurements and adjustments made by LCRA staff to account for runoff from ungauged areas and (2) unregulated tributary flows to the Colorado River downstream of the Highland Lakes, as determined by the Texas Water Availability Model
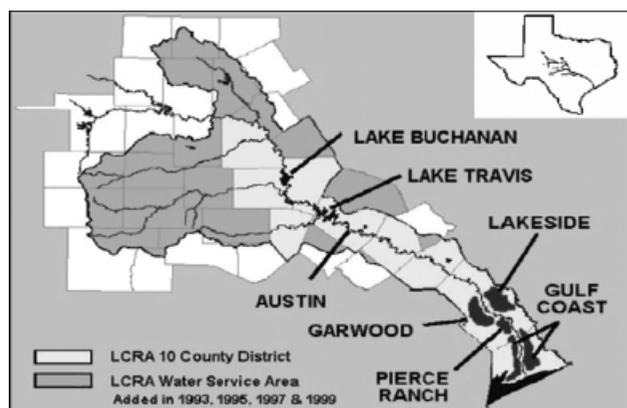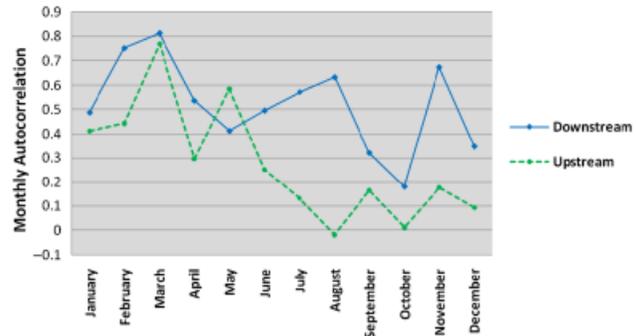

Figure 2 | Monthly autocorrelations for aggregate inflows to the Highland Lakes (upstream) and the Texas Water Availability Model (downstream) data. Correlation coefficients are computed using raw flow data.

(WAM) (Wurbs 2005). The reservoir inflow data spans a total of 57 years, from 1950 to 2006, and the naturalized downstream flow data spans 59 years, from 1940 to 1998. For most of the analyses, the flow data are normalized through a two-step process–first a logarithmic transformation, then conversion to a standardized anomaly by subtraction of the mean (of the log values) and division by the standard deviation (of the log values). While this transforms the data so that the statistical assumption of normality is more valid, it should be noted that the correlation coefficients are generally inflated by the log-transform, and thus an effort is made to illustrate the results in terms of the raw flow data.

Figures 2 and 3 show the monthly and seasonal autocorrelations of these two time series. Monthly autocorrelations of (upstream) reservoir inflows range from a high of nearly 0.8 for February and March flows to a low of essentially 0 for July and August flows. Seasonal correlation coefficients also peak in the winter season, with a value of
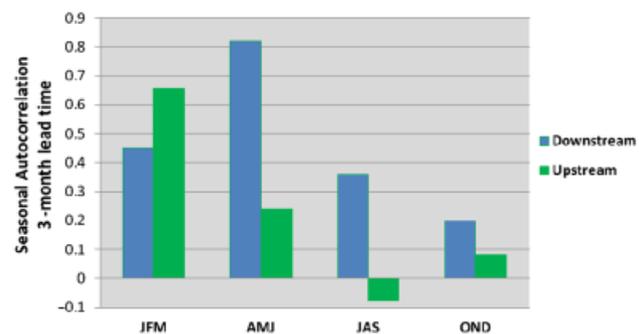

Figure 1 | Lower Colorado River Authority District.

Figure 3 | Seasonal autocorrelations for aggregate inflows to the Highland Lakes (upstream) and the Texas Water Availability Model (downstream) data. Correlation coefficients are computed using raw flow data.

0.66. (The correlation coefficient that is significant at the $p = 0.05$ level is 0.20.) It may be surprising that the seasonal correlation between OND and JFM flows is higher than the average of monthly correlation coefficients during this period. One reason for this may be that averaging over a three-month period reduces the "noise" that results from individual storm events which have a significant effect on monthly flow totals.

The tributary flows to the Colorado River downstream of the Highland Lakes, estimated as the WAM naturalized flows at Mansfield Dam minus the naturalized flows at Bay City, have monthly autocorrelation coefficients that reach a maximum of 0.8 for February and March and have a minimum of nearly 0.2 for September and October. The average monthly autocorrelation for the downstream data is about 0.27 higher than the upstream data. Seasonal autocorrelations peak in the spring season with a value of 0.82, which is lagged by one season in comparison with the upstream data. All seasonal autocorrelation coefficients for the downstream data are significant at $p = 0.10$ or less and the average autocorrelation coefficient is about 0.23 higher than the upstream data.

Based on these autocorrelation coefficients, seasonal streamflow forecasts for certain times of the year may be based solely on hydrologic persistence. We investigate the predictive skill of these forecasts, as well as the potential for large scale ocean–atmosphere interactions to provide additional forecast skill. The oceanic–atmospheric phenomena investigated as potential predictor variables for streamflow in Central Texas are the El Niño–Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO), the North Atlantic Oscillation (NAO), the Atlantic Multidecadal Oscillation (AMO) and the Pacific North American (PNA). The ENSO and NAO generally have a two- to seven-year periodicity (Philander 1990), while PDO and AMO exhibit long-term periodicity (about 25–60 years) (Mantua 1997; Kerr 2000; Gray et al. 2004).

Various indices were selected to quantify the magnitude of these ocean–atmospheric oscillations. The Niño 3.4 index, which characterizes the tropical Pacific Ocean sea surface temperature (SST) anomalies between latitudes 5°S and 5°N and longitudes 170°W and 120°W, was selected as an indicator of ENSO, and monthly index data were obtained from the National Weather Service (NWS) Climate Prediction Center (CPC) (http://www.cpc.ncep.noaa.gov/data/indices). The PNA index values, a measure of atmospheric pressure

anomalies at four locations in the northern hemisphere (Horel & Wallace 1981) were also obtained from the CPC. The PDO index values were obtained from the University of Washington (http://jisao.atmos.washington.edu/pdo). Finally, NAO index values were obtained from the National Center for Atmospheric Research (http://www.cgd.ucar.edu/cas/jhurrell/indices.html) and the AMO index values (Kerr 2000) were obtained from the National Oceanic and Atmospheric Administration (NOAA) Climate Diagnostics Center (CDC) (http://www.cdc.noaa.gov/Climateindices). In all cases, monthly index values for the period 1940–2006 were used for the analysis.

In addition, SST data was analyzed directly for correlations with streamflow. The data used was the extended reconstructed sea surface temperature (ERSST) analysis (Smith et al. 2008), obtained from the National Climatic Data Center (NCDC) through the KNMI Climate Explorer, an online data analysis tool (Oldenborgh & Burgers 2005). Six correlation patterns with high statistical significance ($p < 0.01$) were identified and referenced as ERSST1–ERSST6, corresponding to the following seasonal streamflows: (1) winter reservoir inflows, (2) spring reservoir inflows, (3) winter downstream flows, (4) spring downstream flows, (5) summer downstream flows and (6) fall downstream flows. For each SST pattern, a normalized index was computed based on average seasonal temperatures over a 4° by 4° area, similar to the procedure of Block & Rajagopalan (2007).

STATISTICAL FORECAST MODEL

Multiple logistic regression is most frequently used to model the relationship between a binary response variable and a set of predictor variables, which may be either numerical or categorical. Let $p = Pr(Y = 1)$ denote the probability of success. The ratio of $p/(1-p)$ is called the odds and the function $\log(p/(1-p))$ is called logit(p), which is in fact the logarithm of the ratio of the probability of success to the probability of failure. A multiple logistic regression model can be expressed as follows:

$$\text{logit}(p) = \log(p/(1-p)) = \beta_0 + \beta_1 x_1 + \cdots \beta_k x_k \quad (1)$$

where the parameters β_j are usually estimated using maximum likelihood theory (Menard 1995).

In the binary logistic case, logit(p) compares the probability of a category-1 response (success) to the probability of a category-2 response (failure). If the response variable has more than two levels, logistic regression can still be employed by means of a polytomous or multi-category logistic regression model (Kutner *et al.* 2004). For a response variable with J categories, it is necessary to develop $J-1$ logistic regression models. One category will be chosen as the baseline or reference category, and then all other categories will be compared to it. The choice of reference category is arbitrary. Frequently the last category is chosen.

Using category J to denote the reference category, only $J-1$ logits need to be developed. For a nominal response, the jth logit expression for the ith observation is given as

$$\log\left[\frac{p_{ij}}{p_{iJ}}\right] = X_i^T \beta_j \quad \text{for} \quad j = 1, 2, \dots J-1 \tag{2}$$

where $\beta_j = \begin{bmatrix} \beta_{0j} & \beta_{1j} & \cdots & \beta_{kj} \end{bmatrix}^T$ and $X_i = \begin{bmatrix} 1 & X_{i1} & \cdots X_{ik} \end{bmatrix}^T$. (Note that vectors β_j are different for each category j.) Given the $J-1$ logit expressions, it is possible to obtain the $J-1$ direct expressions for the category probabilities in terms of the $J-1$ linear predictors, $X^T\beta_{jJ}$ (algebra not shown). The resulting expressions are

$$p_{ij} = \frac{\exp(X_i^T \beta_j)}{1 + \sum_{k=1}^{J-1} \exp(X_i^T \beta_k)} \quad \text{for} \quad j = 1, 2, \dots\dots J-1. \tag{3}$$

The estimates of the $J-1$ parameter vectors $\beta_1, \beta_2, \dots, \beta_{J-1}$ can be obtained simultaneously using maximum likelihood estimation. The sum of probabilities of each category for the ith observation is equal to 1. For example, for three response categories, we use category $J=3$ as the baseline category, and there are two comparisons to this reference category. Let p_{ij} denote the probability that category j is selected for the ith response, and then the logit expressions for the two comparisons are

$$\log_e \frac{p_{i1}}{p_{i3}} = X_i' \beta_1, \quad \log_e \frac{p_{i2}}{p_{i3}} = X_i' \beta_2. \tag{4}$$

If we add the constraint $p_{i1} + p_{i2} + p_{i3} = 1$, we can then obtain the probabilities of each category for the ith observa-

tion by solving these three algebraic equations as below:

$$p_{i1} = \frac{\exp(X_i^T \beta_1)}{1 + \exp(X_i^T \beta_1) + \exp(X_i^T \beta_2)} \tag{5}$$

$$p_{i2} = \frac{\exp(X_i^T \beta_2)}{1 + \exp(X_i^T \beta_1) + \exp(X_i^T \beta_2)} \tag{6}$$

$$p_{i3} = \frac{1}{1 + \exp(X_i^T \beta_1) + \exp(X_i^T \beta_2)}. \tag{7}$$

If multiple response categories are treated as ordered, the logistic regression model can be reduced to $J-1$ cumulative logits as follows:

$$\log\left[\frac{p(Y_i \leq j)}{1 - p(Y_i \leq j)}\right] = \alpha_j + X_i^T \beta \quad \text{for} \quad j = 1, 2, \dots, J-1. \tag{8}$$

The difference between the ordinal response logits and the nominal response logits is that each of the $J-1$ parameter vectors β_j is unique for the nominal case; for ordinal response, the slope coefficient vector α is identical for each of the $J-1$ cumulative logits and only the intercepts α_i differ. Finally, the cumulative probabilities $p(Y_i \leq j)$ for the ordinal logistic regression model are given as follows:

$$p(Y_i \leq j) = \frac{\exp(\alpha_i + X_i^T \beta)}{1 + \exp(\alpha_i + X_i^T \beta)} \quad \text{for} \quad j = 1, 2, \dots J-1. \tag{9}$$

The goal of this study is to develop a framework for developing categorical forecasts of streamflows. Since these categories may be treated as ordered, thus the ordinal polytomous logistic regression model, which is also called the proportional odds model, can be used to produce tercile probability forecasts (below-normal, normal and above-normal categories). This may be more effective, yielding a more parsimonious model with easily interpreted results. The software package VGAM developed in R (http://www.r-project.org) was used to derive the ordinal polytomous logistic regression model. This software employs the maximum likelihood method to estimate the model parameters (Yee 2010).

PREDICTOR SELECTION

In this study, a total of seven potential predictor variables are examined for each of the two forecast locations and four

seasonal forecast periods. Therefore, 128 (2^7) alternative models can be constructed with each predictor either included or excluded from each of the eight forecast models. Automatic search procedures are employed to screen the most promising models according to a specified criterion without requiring the fitting of all possible regression models. For logistic regression modeling, two commonly used criteria are Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC), which are defined, respectively, as follows:

$$AIC = -2Ln(L(\boldsymbol{b})) + 2P \tag{10}$$

$$BIC = -2Ln(L(\boldsymbol{b})) + PLn(n) \tag{11}$$

where $\boldsymbol{b}$ denotes the vector of estimated parameters of the logistic regression model (using the maximum likelihood method), L($\boldsymbol{b}$) is the log-likelihood function, P is the number of estimated parameters and n is the total number of observations. Promising models will yield relatively small values for these criteria.

In this study, a forward stepwise search procedure is used to select the best logistic regression model (Seber & Lee 2003). Essentially, this search method develops a sequence of regression models, at each step adding or deleting a predictor variable according to a decision rule. For logistic regression, the decision rule is based on the likelihood-ratio test and its significance (p values), which are obtained from a chi-squared distribution with the associated degree of freedom.

In the forward stepwise procedure, a predictor variable will be added to the model at each step only if the p value is less than a critical value (usually 0.05). Additionally, a predictor variable in the model will be deleted when its p value exceeds a critical value. The procedure will terminate until no further predictor variables can be added with resulting p values less than the critical value, i.e. there are no predictors considered sufficiently helpful to enter the regression model.

The forward stepwise method is applied to select predictor variables from a set of potential predictors, which include streamflows and large-scale climate signals observed in the seasonal period prior to the forecast. For convenience, the seasons are defined as winter (January–March), spring (April–June), summer (July–September) and fall (October–December). Semi-annual streamflow (January–June) is also predicted based on observations from the previous fall. A summary of the logistic regression models selected using the forward stepwise method for reservoir inflows is presented in Table 1. Climate indices and streamflow for the season prior to the predicted seasonal streamflow are designated by (-1). The results show that streamflow persistence is a statistically significant predictor ($p = 0.05$) for winter, spring, fall and Jan–June streamflow forecasts. Amongst the large-scale climate signals, either the ERSST1 pattern (shown in Figure 4) or PDO is a significant predictor in the logistic regression model for winter streamflow forecasts, but not both, likely due to high colinearity between these predictors. For spring streamflow

Table 1 | Best logistic regression models from forward stepwise regression for seasonal streamflow forecasts of aggregate inflows to the Highland Lakes (upstream). Values in brackets are p values for variable entering the model. Percent improvement is relative to climatology (median historical value)

| Model | Selected predictors (p value) | Model fitting criteria | | | | | |
| | | Regression model | | Climatology | | Improvement (%) | |
		AIC	BIC	AIC	BIC	AIC	BIC
Winter-A	Fall(-1) (<0.001) PDO(-1) (0.017)	94.93	107.08	127.01	131.06	25.3	18.3
Winter-B	Fall(-1) (<0.001) ERSST1(-1) (0.010)	92.73	104.88	127.01	131.06	27.0	20.0
Spring	Winter(-1) (0.003) ERSST2(-1) (<0.001) PNA(-1) (0.002)	105.69	117.95	129.24	133.33	18.2	11.5
Fall	Summer(-1) (0.004)	122.41	130.58	129.24	133.33	5.3	2.1
Jan–Jun	Fall(-1) (0.004)	119.83	127.93	127.01	131.06	5.6	2.4
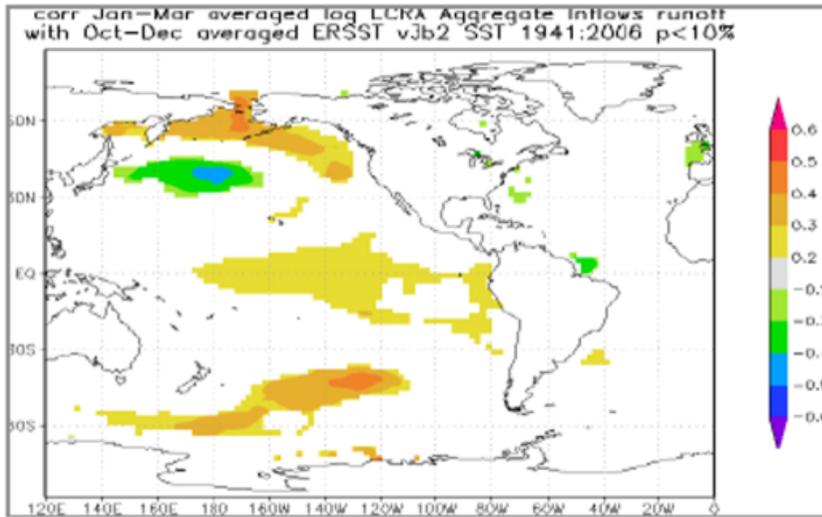
Figure 4 | Correlation map of winter aggregate inflow to the Highland Lakes with fall sea surface temperatures. Circled regions indicate strong positive and negative correlations used to derive the ERSST1 index.

forecasts, however, both ERSST2 (shown in Figure 5) and PNA are significant predictors to be retained in the model. The other large-scale signals (ENSO, NAO and AMO) are not statistically significant in the logistic regression model. Also shown is the relative improvement of forecast models with the selected predictor(s) in terms of model selection criterion AIC and BIC over forecasts based on seasonal climatology (i.e. forecasts equal to the median historical value). The greatest improvement is observed for winter

streamflow forecasts, whereas the stepwise selection procedure indicates there are no significant predictor variables for summer streamflow forecasts.

A summary of logistic regression models selected for seasonal streamflow forecasts of downstream flows is presented in Table 2. Results show that streamflow persistence is a significant predictor for all seasons. Amongst the large-scale climate signals, SST indices (based on the correlation patterns shown in Figures 6–8) have the most significant impact on the
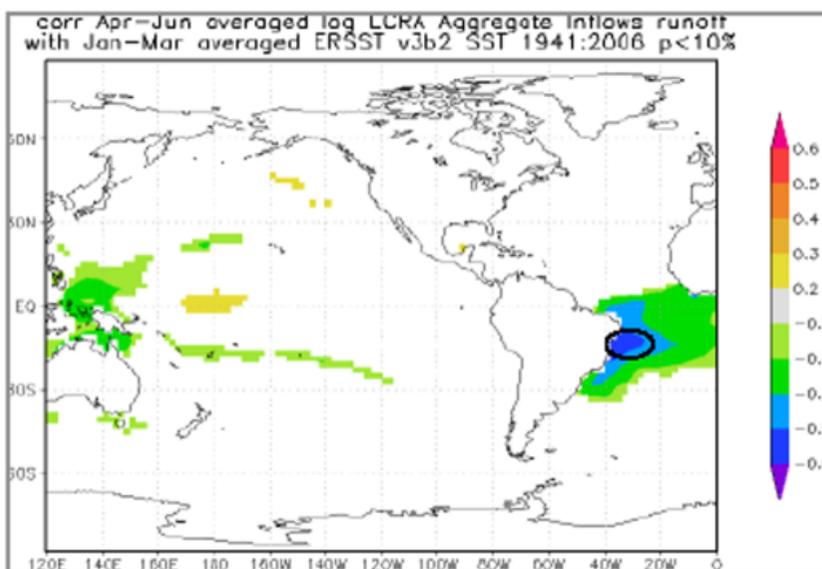

Figure 5 | Correlation map of spring aggregate inflow to the Highland Lakes with winter sea surface temperatures. Circled regions indicate strong negative correlations used to derive the ERSST2 index.

Table 2 | Best logistic regression models from forward stepwise regression for seasonal streamflow forecasts of unregulated tributary flows, from the Texas Water Availability Model (downstream). Values in brackets are *p* values for variable entering the model. Percent improvement is relative to climatology (median historical value)

| Model | Selected predictors (*p* value) | Model fitting criteria | | | | | |
| | | Regression model | | Climatology | | Improvement (%) | |
		AIC	BIC	AIC	BIC	AIC	BIC
Winter	Fall(−1) (<0.001)	88.47	99.70	109.47	113.21	19.2	11.9
	ERSST3(−1) (0.013)						
Spring	Winter(−1) (<0.001)	83.82	91.39	111.62	115.41	24.9	20.8
	ERSST4(−1) (<0.001)						
Summer	Spring(−1) (<0.001)	107.39	119.86	133.60	137.76	19.6	13.0
	ERSST5(−1) (0.007)						
Fall	Summer(−1) (0.002)	100.97	108.54	111.62	115.41	9.5	6.0
	ERSST6(−1) (0.005)						

logistic regression models. The other large-scale signals do not have a statistically significant impact on the unregulated flows downstream of the Highland Lakes; however, it should be noted that each of the SST patterns used to develop the climate indices can be related to one of the named oscillations.

Forecast verification

Forecast verification aims to evaluate the agreement between forecasts and observations (Katz & Murphy 1997; Stephenson 2003). The Brier skill score (BSS) and the ranked probability skill score (RPSS) are common statistics used to measure the

improvement in the accuracy of multicategory probability forecasts over a naive forecasting method such as climatology. The Brier skill score (BSS) (Doggett 1998) is defined as

$$BSS = 1 - BS/BSC \tag{12}$$

where BS is the Brier score, and is defined as

$$BS = (1/n)\Sigma(f_i - l(\text{obs}_i))^2 \tag{13}$$

where f_i is the forecast probability of event i occurring; $l(\text{obs}_i)$ is an indicator variable (1 if event in category i occurs, else 0); n is the number of events; and BSC is the climatologically expected value of BS, equal to $BSC = P_i(1-P_i)$, where P_i is the
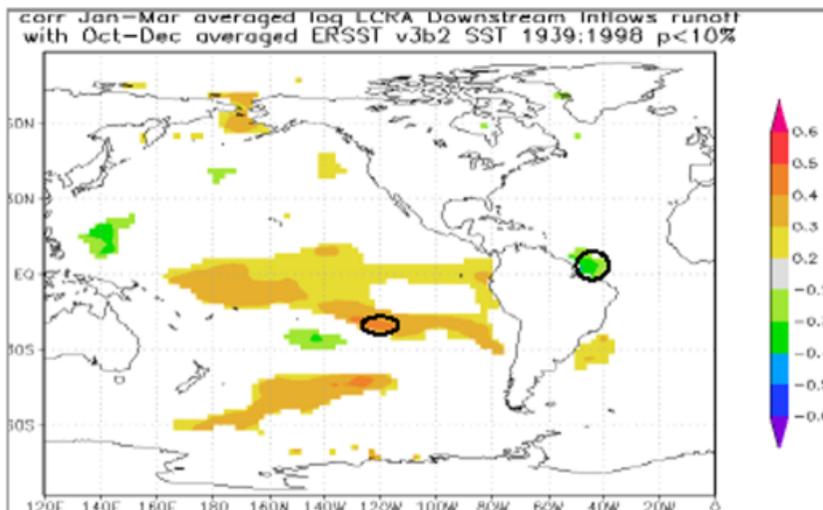

Figure 6 | Correlation map of winter downstream flows with fall sea surface temperatures. Circled regions indicate strong positive and negative correlations used to derive the ERSST3 index.
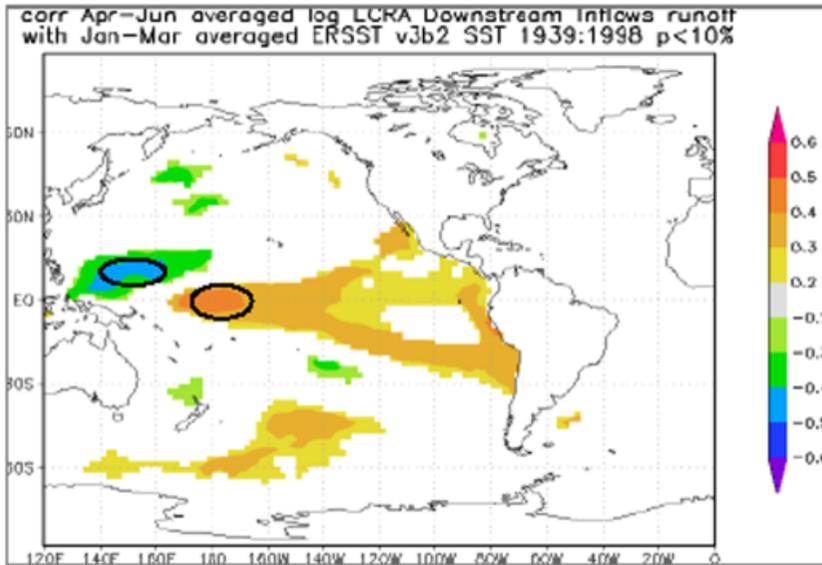
Figure 7 | Correlation map of spring downstream flows with winter sea surface temperatures. Circled regions indicate strong positive and negative correlations used to derive the ERSST4 index.

climatological probability of the event. The Brier score is often applied to events that exceed a given threshold, but it can also be applied to categorical events. In this study, we consider three tercile categories, i.e. below normal, normal and above normal. Accordingly, $BSC = (1/3)(1-1/3) = 0.222$ for each category. A perfect forecast has a value of BSS of 1; positive values between 0 and 1 indicate forecast performance better than climatology and negative values indicate forecast performance worse than climatology.

The ranked probability score (*RPS*) evaluates the sum of the squared differences in the cumulative probability space, so that

$$RPS = \frac{1}{K-1} \sum_{m=1}^{K} \left[\left(\sum_{k=1}^{m} f_k \right) - \left(\sum_{k=1}^{m} o_K \right) \right]^2 \qquad (14)$$

where K is the number of forecast categories (below normal, normal and above normal), f_k is the forecast probability for
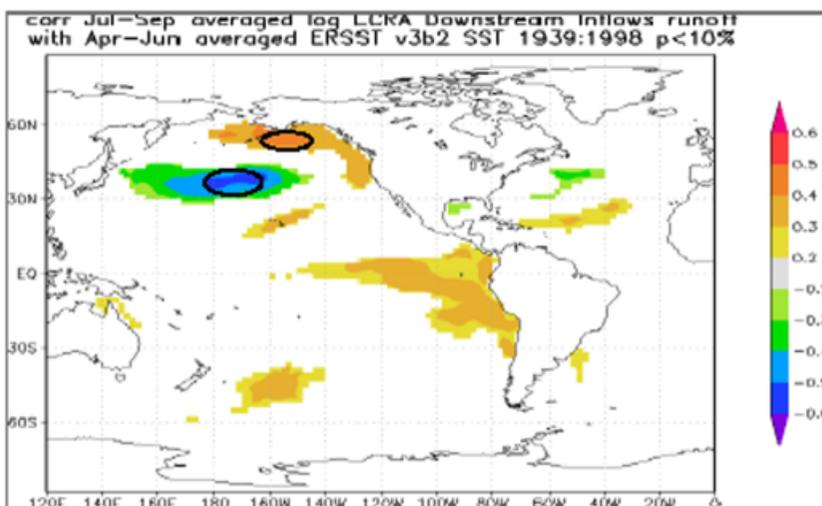

Figure 8 | Correlation map of summer downstream flows with spring sea surface temperatures. Circled regions indicate strong positive and negative correlations used to derive the ERSST5 index.

the kth point and o_k equals 0 or 1 to indicate whether or not the observed flow is in the kth category. The use of *RPS* results in higher penalties for forecasts further away from actual outcomes, rather than scoring based on only hit and miss. The *RPS* can assume a number between 0 and 1, with a perfect forecast scoring 0. The ranked probability skill score (*RPSS*) then measures the relative improvement of using a forecast over climatology alone and is given by

$$RPSS = \frac{\overline{RPS} - \overline{RPS}_{climatology}}{o - \overline{RPS}_{climatology}} = 1 - \frac{\overline{RPS}}{\overline{RPS}_{climatology}} \qquad (15)$$

The probabilities of streamflow in each category in the climatology forecast (i.e. prior probabilities) are the same and equal to 1/3 due to the definition of the flow regimes. Thus, the *RPS* values of the three categories (below normal, normal and above normal) in the climatology forecast are 0.278, 0.111 and 0.278, respectively. A perfect *RPSS* is 1 and negative scores indicate that forecasts performed worse than climatology.

In this study, forecast performance is evaluated using a leave-one-out cross-validation method, in which one observed streamflow value is held out and the remaining data are used to generate a prediction. This process is repeated for each value in the dataset and the cross-validated forecasts are then evaluated using *BSS* and *RPSS*.

RESULTS AND DISCUSSION

The ordinal polytomous logistic regression models with the minimum *AIC* and *BIC* values, indicated in Tables 1 and 2, were applied to generate tercile probability forecasts for flows in the Highland Lakes system. The Brier skill score (BSS) and the ranked probability skill score (RPSS) were used to evaluate the performance of cross-validated (leave-one-out) forecasts. Summaries of the results for both data sets (upstream and downstream of the reservoirs) are presented in Tables 3 and 4.

Results show that (upstream) reservoir inflows for the winter season can be predicted with significant skill with one season lead time based on either persistence and ERSST1 or persistence and PDO. The *BSS* and *RPSS* values indicate that

Table 3 | Brier Skill Score (BSS) and Ranked Probability Skill Score (RPSS) for leave-one-out cross-validated seasonal forecasts of inflows to the Highland Lakes reservoir system

Forecast model	Forecast skill score	
	BSS (%)	RPSS (%)
Winter-A	22.9	19.2
Winter-B	28.2	25.8
Winter*	18.7	12.7
Spring	13.8	8.4
Spring*	0.2	−2.0
Fall	−3.2	−4.7
Jan–Jun	2.7	1.4

*Denotes values for a forecast model based on hydrologic persistence only.

winter streamflow forecasts have an average improvement in skill of 25.6% and 22.5% improvement, respectively, over climatology. Table 3 also shows the *BSS* and *RPSS* for winter streamflow forecasts based only on persistence (18.7% and 12.7%, respectively), which indicates that including ERSST1 or PDO can provide a significant improvement over a forecast based on streamflow persistence only. For spring streamflow forecasts, *BSS* and *RPSS* values indicate that forecast skill can be improved by about 8–13% over climatology based on persistence, ERSST2 and PNA together. Spring streamflow forecasts based only on persistence have no skill. Hydrologic persistence also shows no skill as a predictor for fall streamflows, and the other large-scale climate indices (ENSO, PDO, NAO, AMO and PNA) are not useful predicators for any season.

Table 4 | Brier Skill Score (BSS) and Ranked Probability Skill Score (RPSS) for leave-one-out cross-validated seasonal forecasts of downstream unregulated flows

Forecast model	Forecast skill score	
	BSS (%)	RPSS (%)
Winter	32.3	14.1
Winter*	27.6	8.5
Spring	36.1	16.7
Spring*	33.4	12.6
Summer	16.2	14.8
Summer*	9.2	8.6
Fall	−4.8	−3.9

*Denotes values for a forecast model based on hydrologic persistence only.
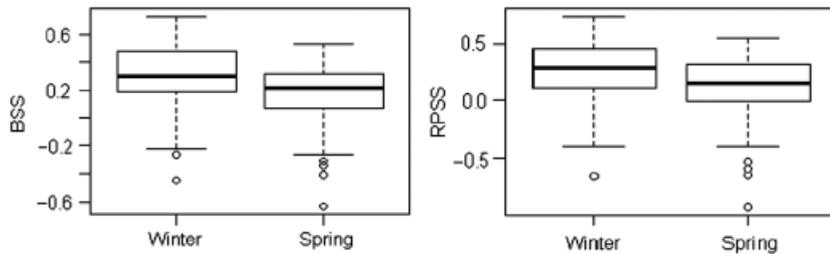
Figure 9 | Box plots of *BSS* (left) and *RPSS* (right) of upstream flow forecasts for winter and spring seasons. Forecasts are based on dropping 10–12% of the observations randomly. The boxes correspond to the interquartile ranges (25th and 75th percentiles), the horizontal line in each box is the median and whiskers extend to the median of lower and upper quartiles (12.5th and 87.5th percentiles). Circles represent the lowest values.

Results in Table 4 show that streamflow persistence (auto-correlation) is a useful predictor for downstream unregulated flows for winter, spring and summer seasons. In addition, the derived SST indices ERSST3, ERSST4 and ERSST5 provide additional forecast skill for these three seasons, respectively. The forecast skill score indicate between 12–33% improvement over climatology-based forecasts for these seasons. For fall forecasts, however, skill scores are very close to 0, indicating no improvement over climatology, despite the correlations found in the regression analysis using all data together (i.e. not holding data out as in cross-validation).

These results for both upstream and downstream flow forecasts are generally similar to the findings of Rajagopalan *et al.* (2000) and Tootle *et al.* (2005), but with some important differences. Both of these previous studies concluded that weak relationships exist between climate indices and stream-flow in central Texas. In this study, only one identified oceanic–atmospheric mode, PDO, was found to provide significant improvement in winter streamflow forecasts (a second, PNA, provided a small improvement in spring fore-casts). However, derived SST indices, each of which have a spatial correlation pattern similar to ENSO or PDO, were

found to provide significant improvements in forecast skill when included in the regression models. Some differences are also attributed to different lead times–this study focuses on a seasonal lead time while the investigation by Tootle *et al.* (2005) was based on an annual lead time.

To further evaluate the forecast models, we tested the forecasts by randomly holding out two observations from each of the three (climatology-based) categories, thus dropping approximately 10–12% of the data at random, and repea-ting this procedure 100 times. The *BSS* and *RPSS* metrics for both upstream and downstream flows during winter and spring are shown as box plots in Figures 9 and 10. The median skill scores are essentially the same as the leave-one-out cross-validation forecasts, although there is considerable variability in the skill scores due to resam-pling. This variability would be important to water managers evaluating forecasts over short observational periods.

Comparisons of observations and leave-one-out cross-validated forecasts are shown in Figures 11 and 12 for winter reservoir inflows and spring downstream flows, respectively. Since the ordinal polytomous logistic regression
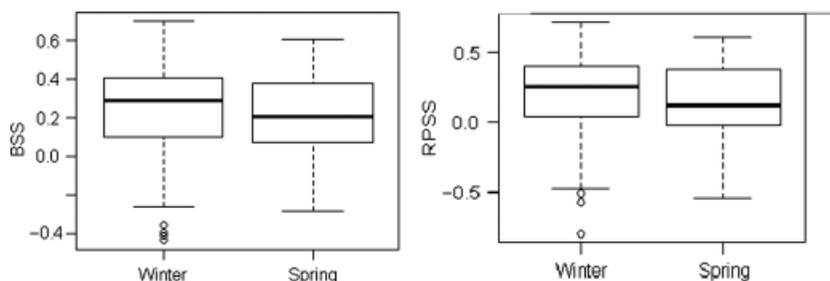

Figure 10 | Box plot of *BSS* (left) and *RPSS* (right) of downstream flows forecasts for winter and spring. Forecasts are based on dropping 10–12% of the observations randomly. The boxes correspond to the interquartile ranges (25th and 75th percentiles), the horizontal line in each box is the median and whiskers extend to the median of lower and upper quartiles (12.5th and 87.5th percentiles). Circles represent the lowest values.
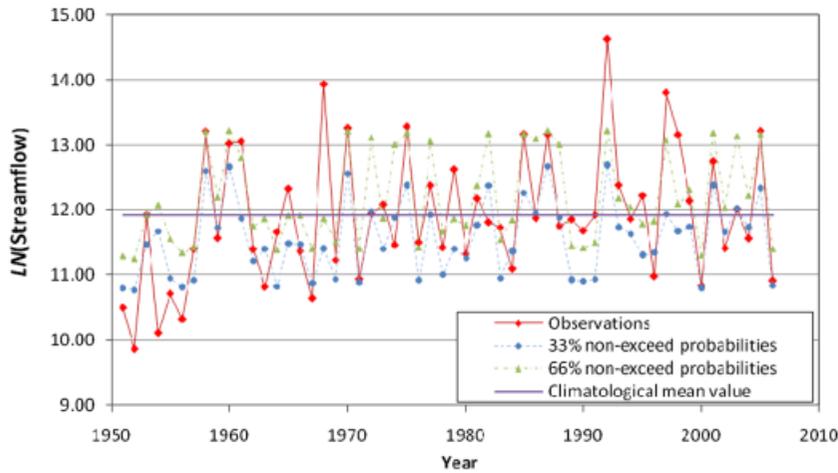
Figure 11 │ Plot of observations and cross-validated forecasts for winter reservoir inflows to the Highland Lakes. Units on the y axis are the natural logarithm of flow in acre-feet per month (1 acre-ft equals approximately 1234 m³).

models provide tercile probability forecasts for flows, the following method is used to obtain these streamflow volume forecasts:

(1)   First we calculate the empirical cumulative probabilities based on climatology forecasts. This assumes that each flow observation is equally likely and the probability of each category is equal to 1/3.

(2)   Given the forecasted probability of each category, we adjust the empirical cumulative probabilities so that the probability of each category matches the forecasted probability instead of 1/3. This is similar to the approach of Croley (2000).

(3)   To obtain the streamflow volumes corresponding to the 33% and 66% non-exceedance probabilities, we interpolate quantiles of a log-normal distribution according to the adjusted cumulative probabilities.

Results show many years in which the climate-based forecasts provide a significant improvement over climatology. For instance, the forecasts accurately predict high reservoir inflows in 10 of the 12 years in which winter inflows exceeded 442 400 acre-ft ($\ln(442\,400) = 13.0$). Perhaps more importantly, the forecasts accurately predict low reservoir inflows ($\ln(\text{streamflow}) < 11.0$) in 8 of 11 cases. Downstream flow forecasts for the spring season have even better skill,
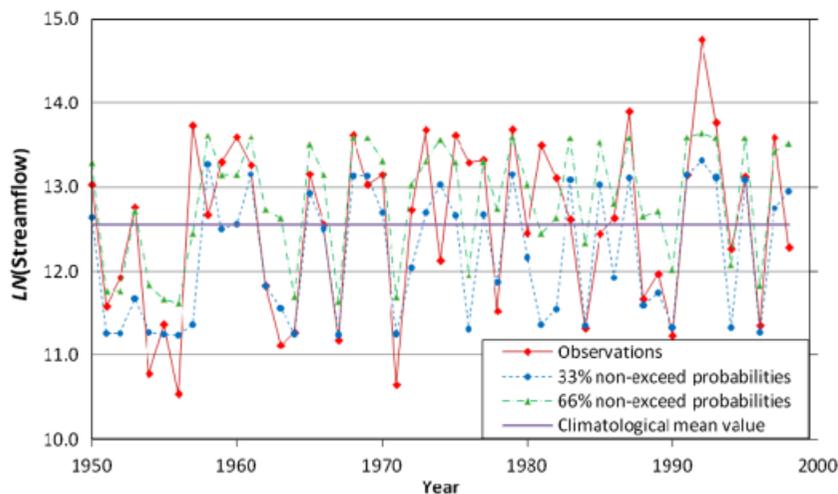

Figure 12 │ Plot of observations and cross-validated forecasts for spring streamflows downstream of the Highland Lakes. Units on the y axis are the natural logarithm of flow in acre-feet per month (1 acre-ft equals approximately 1234 m³).

with accurate predictions of high flows (ln(stream-flow) > 13.5) in 9 of 11 cases and accurate predictions of low flows (ln(streamflow) < 11.5) in all 11 cases.

These results indicate that hydrologic persistence (streamflow autocorrelation) can provide skillful forecasts of reservoir inflows during the winter and spring seasons and downstream flows during the winter, spring and summer seasons. One reason for this is that streamflow in the winter months is closely related to soil moisture, which tends to be higher during fall and winter, with persistence from fall through winter and early spring in central Texas. This is not the case for late spring and summer in the upper watershed, when soils dry and high runoff mainly results from convective storm events. Persistence in soil moisture extends further into the spring, and sometimes early summer, in the more humid portion of the watershed downstream of the reservoirs, which may explain why downstream flows are somewhat predict-able during the summer season.

It is more difficult to explain how large-scale climate signals affect streamflow in central Texas. Either the derived SST index ERSST1 or the PDO index was found to be useful in predicting winter reservoir inflow, and there is strong correlation between the two indices ($r = 0.88$), indicating that ERSST1 is essentially a surrogate for PDO. There is also strong correlation between ERSST2 and PDO ($r = 0.62$). Warm-phase PDO winters correspond to blocking high pres-sure over the northeastern Pacific Ocean, which shifts the jet stream northward, leading to warmer and drier than average conditions, and thus lower soil moisture, in central Texas. Conversely, cool-phase PDO winters tend to be cooler and wetter than average, with higher soil moisture. There is strong persistence in the PDO index from October to March, and so the fall PDO index (October to December) is a good indicator of soil moisture through the winter season. This persistence may extend to early spring, which may explain why ERSST2 is a significant indicator for spring season. PNA has a similar but weaker effect on soil moisture; nonetheless it is a statis-tically significant indicator for spring conditions.

In contrast to the upstream flows, none of the climate indices considered were selected by the stepwise method for inclusion in the logistic regression models for downstream flow forecasts. However, each of the identified SST patterns (Figures 6–8), from which new indices were derived, can be related to typical PDO or ENSO SST patterns in the Pacific

and Atlantic Oceans. (PDO and Nino3.4 indices were likely not selected due to colinearity with these patterns.) Figure 6 shows the temporal correlation map between fall SSTs and winter downstream flows. While ENSO is most well known to affect SSTs in the equatorial Pacific, it also strongly affects the climate of northeastern Brazil (e.g. Souza Filho & Lall 2003). The SST pattern shown in Figure 7, used to forecast spring downstream flows, exhibits a classic ENSO pattern in the western Pacific. Similarly, the SST pattern exhibited in Figure 8, used to forecast summer downstream flows, is typical of PDO patterns in the northern Pacific.

CONCLUSION

This work aimed to develop seasonal streamflow forecast models for the Highland Lakes system in central Texas. Hydrologic persistence (streamflow autocorrelation), five large-scale climate indices (Nino3.4, PDO, NAO, AMO and PNA) and six derived SST indices were screened for inclusion in an ordinal polytomous logistic regression model. Results indicate that hydrologic persistence is a useful predictor of seasonal streamflows both upstream and downstream of the Highland Lakes reservoir system during the winter and spring. Summer downstream flow forecasts based on persis-tence also exhibit significant skill. In addition, winter reser-voir inflow forecasts may be significantly improved by including either a derived SST index or the PDO index, and spring reservoir inflow forecasts may be improved by includ-ing a derived SST index and PNA. Similarly, including derived SST indices, related to ENSO and PDO SST patterns, improves downstream flow forecasts during the winter, spring and summer.

The methods presented here are completely transferable to other regions where significant hydrologic persistence and/or teleconnections between seasonal streamflow and large-scale climate anomalies exist. Stepwise linear regression with selection of predictor variables based on information criteria proved an effective method of screening a large number of potential predictors. Ordinal polytomous logistic regression proved an effective and parsimonious method for producing the probabilistic (categorical) streamflow forecasts. Both of these methods assume linearity, however, while rela-tionships between streamflow and climate anomalies are likely

to be nonlinear and include multivariate interactions due to the complexity of ocean–atmosphere dynamics (Araghinejad *et al.* 2006). Linear regression and logistic regression models are difficult to interpret if nonlinearity and/or interactions are present. To this end, nonlinear statistical methods such as data mining (machine learning) may be considered in future work to improve the predictive skill of seasonal forecasts.

ACKNOWLEDGEMENT

REFERENCES

Alley, W. M. 1985 Water balance models in one-month-ahead stream flow forecasting. *Wat. Res. Res.* **21**(4), 597–606.

Araghinejad, S., Brun, D. H. & Karamouz, M. 2006 Long-lead probabilistic forecasting of streamflow using oceanic-atmospheric and hydrological predictors. *Wat. Res. Res.* **42**, W07411.

Block, P. & Rajagopalan, B. 2007 Interannual variability and ensemble forecast of the Upper Blue Nile Basin Kiremt season forecast. *J. Hydrometeorol.* **8**, 327–342.

Croley, T. E. II 2000 *Using Meteorology Probability Forecasts in Operational Hydrology*. ASCE Press, Reston, VA.

Day, G. N. 1985 Extended streamflow forecasting using NWSRFS. *J. Wat. Res. Plann. Mngmnt.* **111**(2), 157–170.

Doggett, K. 1998 *Glossary of verification terms (revised June, 1998)*. National Oceanic and Atmospheric Administration. Retrieved November 13, 2002 from http://www.sel.noaa.gov/forecast_verification/verif_glossary2.html.

Duan, Q., Ajami, N. K., Gao, X. & Sorooshian, S. 2006 Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Wat. Res.* **30**, 1371–1386.

Gershunov, A. & Barnet, T. P. 1998 ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Observations and model results. *J. Climate* **11**(7), 1575–1586.

Gray, S. T., Graumlich, L. J., Betancourt, J. L. & Pederson, G. T. 2004 A tree-ring based reconstruction of the Atlantic Multidecadal Oscillation sing 1567 A. D. *Geophys. Res. Lett.* **31**, L12 205.

Hamlet, A. F. & Lettenmeier, D. P. 1999 Columbia River streamflow forecasting based on ENSO and PDO climate signals. *J. Wat. Res. Plann. Mngmnt.* **125**(6), 333–334.

Horel, J. D.& Wallace, J. M. 1981 Planetary scale atmospheric phenomena associated with the Southern Oscillation. *Weather Review* **109**, 813–829.

Kerr, R. A. 2000 A North Atlantic climate pacemaker for the centuries. *Science* **228**, 1984–1986.

Karamouz, M. F. & Zahraie, B. 2004 Seasonal streamflow forecasting using snow budget and El Niño-Southern Oscillation climate

signals: application to the Salt River Basin in Arizona. *J. Hydrol. Engng.* **9**(6), 523–533.

Katz, R. W. & Murphy, A. H. 1997 *Economic Value of Weather and Climate Forecasts*, Cambridge University Press, Cambridge, ch 2, pp 19–74.

Kutner, M. H., Nachtsheim, C. J. & Neter, J. 2004 *Applied Linear Regression Models*. McGraw-Hill, New York.

Mantua, N. 1997 *Relationships Between a Naturalized Columbia River Flow Record and Large Scale Climate Variations over the North Pacific*. JISAO Climate Impacts Group, Year 2 Progress Rep., University of Washington. Seattle, WA.

Menard, S. 1995 *Applied Logistic Regression Analysis*. Sage Publishing, Thousand Oaks, CA.

Oldenborgh, G. J. van & Burgers, G. 2005 Searching for decadal variations in ENSO precipitation teleconnections. *Geophys. Res. Lett.* **32**(15), L15 701.

Philander, S. G. 1990 *El Niño, La Niño, and the Southern Oscillation*. Academic Press, San Diego, CA.

Piechota, T. C., Chiew, F. H. S., Dracup, J. A. & McMahon, T. A. 1998 Seasonal streamflow forecasting in eastern Australia and the El Niño–Southern Oscillation. *Wat. Res. Res.* **34**, 3035–3044.

Piechota, T. C. & Dracup, J. A. 1996 Drought and regional hydrologic variation in the United States: associations with the El Niño–Southern Oscillation. *Wat. Res. Res.* **32**(5), 1359–1373.

Rajagopalan, B., Cook, E., Lall, U. & Ray, B. K. 2000 Spatiotemporal variability of ENSO and SST teleconnections to summer drought over the United States during the twentieth century. *J Climate* **13**(24), 4244–4255.

Regonda, S. K., Rajagopalan, B. & Clark, M. 2006 A new method to produce categorical streamflow forecasts. *Wat. Res. Res.* **42**, W09501.

Seber, G. A. F. & Lee, A. S. 2003 *Linear Regression Analysis* 2nd edn. John Wiley & Sons, New York.

Sharma, A. 2000 Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: 1. A strategy for system predictor identification. *J. Hydrol.* **239**, 249–258.

Smith, J., Day, G. N. & Kane, M. D. 1992 Nonparametric framework for long range streamflow forecasting. *J. Wat. Res. Plann. Mngmnt.* **118**(1), 82–92.

Smith, T. M., Reynolds, R. W., Peterson, T. C. & Lawrimore, J. 2008 Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis (1880–2006). *J. Climate* **21**, 2283–2296.

Souza Filho, F. A. & Lall, U. 2003 Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: applications of a multivariate, semi-parametric algorithm. *Wat. Res. Res.* **39**(11), 1307–1325.

Stephenson, D. B. 2003. Glossary. In: *Forecast Verification: A Practitioner's Guide in Atmospheric Science* (Jolliffe, I. T. & Stephenson, D. B. (Eds.)). John Wiley & Sons, Chichester, pp 13–36.

Sun, L., Moncunill, D. F., Li, H., Moura, A. D., Souza Filho, F. A. & Zebiak, S. E. 2006 An operational dynamical downscaling prediction system for Nordeste Brazil and the 2002-04 Real-Time Forecast Evaluation. *J. Climate* **19**, 1990–2007.

Tootle, G. A. & Piechota, T. C. 2006 Relationships between Pacific and Atlantic Ocean sea surface temperatures and U.S. streamflow variability. *Wat. Res. Res.* **42**, W07 411.

Tootle, G. A., Piechota, T. C. & Singh, A. 2005 Coupled oceanic-atmospheric variability and U.S. streamflow. *Wat. Res. Res.* **41**, W12 408.

Watkins, D. W. Jr. & O'Connell, S. M. 2005 Teleconnections and disconnections in central Texas: a guide for water managers. In: *Climate Variations, Climate Change and Water Resources Engi-neering* (Garbrecht, J. & Piechota, T. (Eds.)). ASCE Press, Reston, VA, pp 103–114.

Wilks, D. 1995 *Statistical Methods in Atmospheric Science: An Intro-duction*. Academic Press, San Diego, CA.

Wurbs, R. 2005 Texas water availability modeling system. *J. Wat. Res. Plann. & Mngmnt.* **131**(4), 270–279.

Yee, T. W. 2010 The VGAM package for categorical data analysis. *J. Statist. Software* **32**(10), 1–34.