Statistical Modeling of Daily Water Temperature Attributes on the Sacramento River

Jason Caldwell[1]; Balaji Rajagopalan, A.M.ASCE[2]; and Eric Danner[3]

Abstract: The Sacramento River is the largest river in California, and an important source of water for agricultural, municipal, and industrial users. Input to the Sacramento River comes from Shasta Lake and is controlled by operators of Shasta Dam, who are challenged with meeting the competing needs of these users while also maintaining a cold water habitat for Endangered Species Act (ESA) listed winter-run Chinook salmon. The cold water habitat goals are constrained by the volume of cold water storage in the lake, which operators attempt to selectively deploy throughout the critical late summer/fall window. To make informed decisions about the release of this limited cold water resource, skillful forecasts of downstream water temperature attributes at the seasonal time scale are crucial. To this end, we offer a generalized linear modeling (GLM) framework with a local polynomial method for function estimation, to provide predictions of a range of daily water temperature attributes (maximum daily water temperature, daily temperature range, number of hours of threshold exceedance, and probability of threshold exceedance/nonexceedance). These attributes are varied in nature (i.e., discrete, continuous, categorical, etc.), and the GLM provides a general framework to modeling all of them. A suite of predictors that impact water temperatures are considered, including current and prior day flow, water temperature of upstream releases, air temperature, and precipitation. A two-step model selection is proposed. First, an objective method based on Bayesian Information Criteria (BIC) is used in a global model to select the best set of predictors for each attribute; then the parameters of the local polynomial method for the selected best set of predictors are obtained using generalized cross validation (GCV). Daily weather ensembles from stochastic weather generators are coupled to the GLM models to provide ensembles of water temperature attributes and consequently, the probability distributions to obtain risk estimates. We demonstrate the utility of this approach by modeling water temperature attributes for a temperature compliance point on the Sacramento River below Shasta Dam. Regulations on the dam depress the water temperature forecasting skill; to show this, we present skillful results from applying the approach to an unregulated location in the Pacific Northwest. The proposed method is general, can be ported across sites, and can be used in climate change studies.
DOI: 10.1061/(ASCE)HE.1943-5584.0001023.

Author keywords: Water temperature; Generalized linear model; Stochastic weather generation; Water management; Seasonal forecasting; Climate impacts.

Introduction

The Sacramento River is the largest river in California, and an important source of water for agricultural, municipal, and industrial users (Fig. 1). Input to the Sacramento River comes from Shasta Lake and is controlled by operators of Shasta Dam, who are challenged with meeting the competing needs of these users while also maintaining a cold water habitat for Endangered Species Act (ESA) listed winter-run Chinook salmon (*Oncorhynchus tshawytscha*). In the late summer and fall, strong atmospheric heating causes water temperatures in the Sacramento River to get too warm for winter-run Chinook salmon, and operators at Shasta Dam adjust the water release volume and temperature in attempts to maintain temperatures below downstream compliance targets. However, there is a limited amount of cold water storage in the reservoir and releasing too much cold water during the early summer and midsummer can exhaust the cold water pool, resulting in increased thermal stress during late summer and early fall when fish are particularly vulnerable. Therefore, to enable the most efficient management of the available resources it is essential to have effective modeling and forecasting of downstream temperature attributes. Modeling efforts that effectively incorporate climate forecast information are needed to provide adequate guidance for making decisions at a variety of time scales (i.e., subdaily to decadal). An integrated modeling framework has been developed to provide short-range forecasting for day-to-day operations (Danner et al. 2012; Pike et al. 2013). However, this approach is limited to 96 h forecasts, and water managers need longer-range capabilities to make informed decisions on how to manage the cold water storage over the course of an entire season. This paper presents the development of a seasonal component to the framework that provides managers with monthly to seasonal forecasts of key water temperature attributes and probabilistic estimates of risk of meeting or exceeding predetermined thresholds of each attribute.

[1]Civil Environmental and Architectural Engineering, Univ. of Colorado, Boulder, CO 80309.

[2]Senior Project Engineer, Leonard Rice Engineers, Inc., 1221 Auraria Parkway, Denver, CO 80204; formerly, Civil Environmental and Architectural Engineering, Univ. of Colorado, Boulder, CO 80309 (corresponding author). E-mail: rajagopalan.balaji@colorado.edu

[3]Research Ecologist, Fisheries Ecology Division, National Marine Fisheries Service, NOAA, 110 Shaffer Rd., Santa Cruz, CA 95060.

Note. This manuscript was submitted on June 4, 2013; approved on May 6, 2014; published online on August 25, 2014. Discussion period open until January 25, 2015; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, © ASCE, ISSN 1084-0699/04014065(12)/$25.00.

Water Temperature Models

Recent advances in water temperature monitoring and modeling have facilitated the collection, analysis, and understanding of the complexity of water temperature behavior (Webb et al. 2008; Dunham et al. 2005; Isaak 2011). These concepts and theories are essential to the development of comprehensive models of water temperature. In general, water temperature models fall into two groups: deterministic and empirical. Deterministic models employ an energy budget approach by simulating water temperature through fluid flow and heat transport equations (e.g., Taylor 1998; Carron and Rajaram 2001; Brock and Caupp 1996; Pike et al. 2013). These models capture the physical processes of water temperature dynamics through consideration of unsteady flow, advective-dispersive transport of heat, and heat flux across the air–water and water–sediment interfaces. These require input of detailed data on system geometry, flow, physical attributes of river, and climate in order to robustly calibrate and validate the models. Furthermore, they are computationally intensive, which can be a deterrent for operational agencies with limited resources.

Empirical models, in contrast, generally consist of fitting regression models to daily water temperature as a function of air temperatures and streamflow (e.g., Bogan et al. 2006; Caissie et al. 2001; Neumann et al. 2003). These models rely on the strong correlation between air and water temperature, which is driven by the joint dependence on solar radiation (Benyahya et al. 2007). Such statistical models typically capture variability over large geographic extents (e.g., long stretches of river) and also offer the benefits of computational efficiency and the ability to quantify uncertainty (Benyahya et al. 2007). At daily or weekly time steps, statistical models have the ability to capture water temperature variability; however, at shorter time steps (e.g., hourly), the autocorrelation within the water temperature time series makes the regression increasingly difficult (Mohseni et al. 1998). Empirical modeling techniques, therefore, offer a distinct advantage at longer time scales (i.e., seasonal), while deterministic models are preferred at short (i.e., subdaily) time scales. Neural networks models have also been proposed (e.g., Chenard and Caissie 2008) which can capture nonlinearities in the relationship. However, these models require large amounts of data and subjective inputs, and are computationally intensive.

While regression models (mostly linear) have been the staple of modeling average daily water temperatures (e.g., Neumann et al. 2003), they are not suitable for modeling attributes of daily water temperature that may be skewed (i.e., daily maximum), binary (i.e., probability of threshold exceedance), or discrete (i.e., number of hours of exceedance).

Motivation and Study Site

Most empirical water temperature models focus on estimating average daily water temperatures. However, the management of river temperatures for the protection of coldwater fishes requires understanding and forecasting of a variety of attributes, as these fish are sensitive to acute maximum temperatures and prolonged exposure to higher temperatures (e.g., Myrick and Cech 2000; Van Vleck et al. 1988). Thus, forecasts of these attributes are needed for water resource managers to efficiently plan water releases from reservoirs so as to optimally manage the cold water supply with respect of fish habitat.

To address this need, we propose a generalized linear model (GLM) framework that provides a flexible alternative to modeling a variety of temperature variables. Weather and seasonal climate forecasts can be easily integrated for use in seasonal operational planning (e.g., Neumann et al. 2003, 2006; Towler et al. 2010a, b). Recently, GLMs have been used in stochastic weather generation (Furrer and Katz 2007), waste water quality modeling (Weirich et al. 2011), and in climate applications (Chandler 2005; Chandler and Wheater 2002). Our method uses local polynomials to model and predict a key set of seasonal water temperature attributes: daily temperature maximum (DTX), daily temperature range (DTR), probability of threshold exceedance (POE), and number of hours of exceedance (NHE). This method will provide a simpler and complementary tool to the existing short-term water temperature models for the Sacramento River. We applied the model to the Balls Ferry compliance point approximately 40 km below Shasta Dam on the Sacramento River (Fig. 1), a key management target for meeting temperature objectives on the river (USBR 2004).

Study Site

The construction of Shasta Dam, which was completed in 1945, cut off access to critical spawning and rearing habitat for winter-run Chinook salmon. Prior to the dam's construction, the Sacramento River yielded large volumes of cold water during the winter/spring and smaller volumes of warmer water during the rest of the year (Myrick and Cech 2000). Dam operations now allow cooler water releases during the critical late summer season, which can partially mitigate this loss of habitat (Van Vleck et al. 1988; SFEP 1992; Yates et al. 2008). To protect winter-run salmon eggs and rearing habitat, a temperature target of 13.3°C was established for habitat between Keswick Dam and Red Bluff Diversion Dam (Fig. 1; Bettelheim 2001; Hallock and Fisher 1985).

The installation of a temperature control device (TCD) on Shasta Dam during the mid-1990s improved the capability of maintaining cool downstream water temperatures by enabling

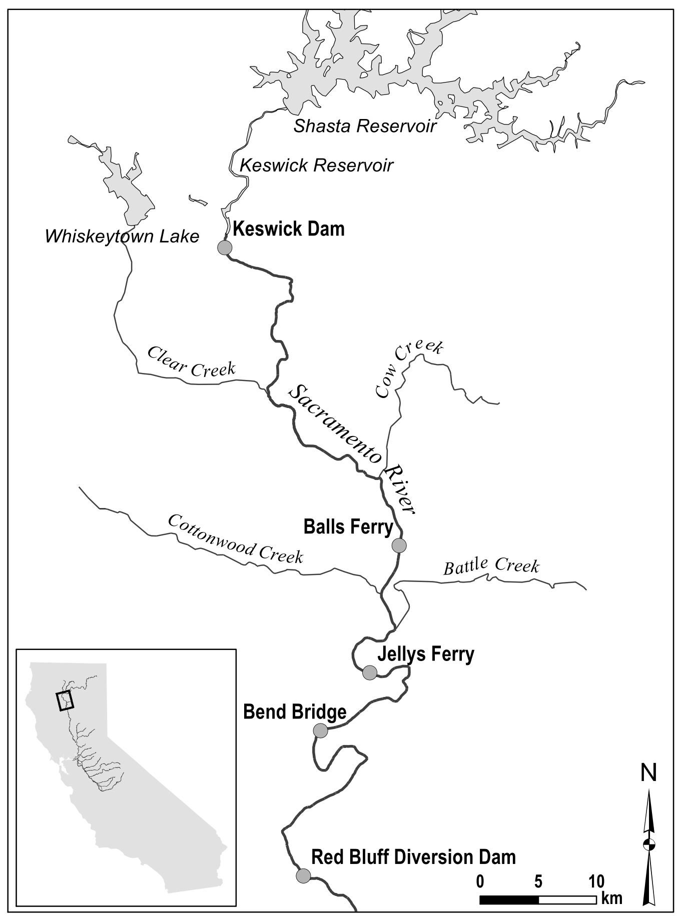

Fig. 1. Map of the study area in the upper Sacramento River basin with compliance points and primary infrastructure (map courtesy of Andrew Pike)

the release of water from different levels within the thermally stratified reservoir. With this level of control, managers attempt to conserve the cold water supply as late in the season as possible, and release it in the late summer/early fall when upper reservoir temperatures are already approaching the compliance limit. As a result, downstream water temperatures are now a combined function of operations and hydrometeorology. Inclusion of the daily flow and water temperature releases in any statistical modeling efforts may partially account for these operations as the volume of water released and its respective temperature ultimately advect downstream and alter the thermal properties.

Methods

Generalized Linear Models

In a GLM, the response or the dependent variable Y can be assumed to be a realization from any distribution in the exponential family with a set of parameters (McCullagh and Nelder 1989). A smooth and invertible link function transforms the conditional expectation of Y to a set of predictors [Eq. (1)]

$$G[E(Y)] = f(\mathbf{X}) + \varepsilon = X\beta^T + \varepsilon \tag{1}$$

$G(.)$ = link function; $\boldsymbol{X}$ = set of predictors or independent variables; $E(Y)$ = expected value of the response variable; T = transpose operator; and ε = error assumed to be normally distributed with variance (σ_ε). In a linear model (the standard linear regression), the function $G(.)$ is identity and Y is assumed to be normally distributed. Depending on the assumed distribution of Y, there exist appropriate link functions (McCullagh and Nelder 1989). The model parameters, β, are estimated using an iterative weighted least squares method that maximizes the likelihood function as opposed to an ordinary least squares method in linear modeling. The GLM can be used to model a variety of response variables—for skewed variables with a lower bound of 0 such as daily maximum water temperature or daily water temperature range, the Gamma distribution assumption of Y and its associated link function is appropriate; for number of hours of temperature exceedance, the Poisson distribution and its associated link functions can be used; for probability of exceedance, a binomial distribution and its link function (i.e., logistic regression) is the approach. McCullagh and Nelder (1989) provide information about a variety of distributions, link functions, and parameter estimation.

To obtain the best set of predictors for the model, there are objective criteria such as the Akaike Information Criteria (AIC) or Bayesian Information Criteria (BIC), both of which penalize the likelihood function based on the number of parameters (Venables and Ripley 2002). Models are fit using all possible subsets of predictors and also link functions; for each, the AIC and BIC are computed and the model with lowest AIC or BIC is selected as the best model. Models can also be tested for significance against a null model or an appropriate subset model using a chi-squared test. BIC is used in this study as it tends to be slightly more parsimonious compared to AIC.

The function f in Eq. (1) is linear and fitted to the entire data (i.e., it is a global fit) and therefore, can miss capturing local nonlinearities (Fig. 2). To address this, a nonparametric approach based on local polynomials (Loader 1999) is used for fitting f. In this, the function is estimated 'locally' for any desired point $\boldsymbol{x}$. A small set of neighbors ($K = \alpha N$; N is the total number of data points and α is a value in the range of 0 to 1) of x are identified and a polynomial of order p is fitted via weighted least squares method, wherein, the nearest neighbors are assigned highest weight and the farthest the least, using a bisquare or tricubic weight function (Loader 1999). The fitted polynomial is used to estimate the response variable Y at the desired point x. This process is repeated for all desired points of estimate. Note that if α and p are set to 1 and the neighbors assigned equal weights, then this collapses to the standard linear regression. This local estimation can be used to fit any distributional form of Y, thus providing an additional degree of flexibility to the GLM framework. The choice of α and p are obtained using a generalized cross validation criteria (GCV) that is similar to AIC or BIC in that it penalizes higher order models and strives for parsimony. The GCV can be used to obtain the local polynomial parameters (α and p) and the best set of predictors (e.g., Regonda et al. 2005); however, here, the local polynomials are fitted to the best predictor set obtained from BIC using the global fit. This hybrid approach is preferred for computational efficiency. It is well known from past studies (see references in the introduction) that stream temperature on any day is strongly influenced by the air temperature, precipitation, streamflow, and the temperature of water released upstream on that day. Furthermore, for long reaches downstream of reservoirs, operations on a day also influence the stream temperature of the following day. Thus in light of the lagged

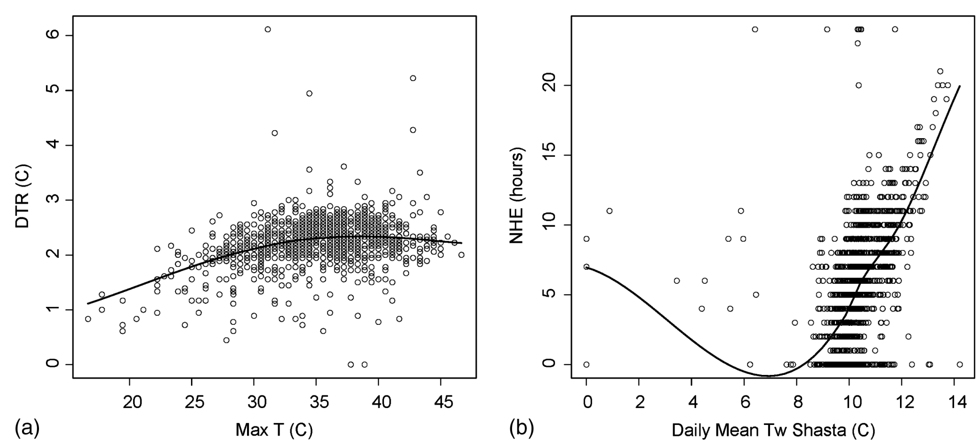

Fig. 2. Scatterplots of (a) DTR and maximum air temperature; (b) NHE and daily mean water temperature release at Shasta for the season of July through September; the local estimation curve is plotted to indicate local non-linear features in the data

effect, we considered a suite of potential predictors that include current and previous day's maximum temperature, minimum temperature, streamflow, stream temperature, and precipitation (see Table 1 for listing).

Local polynomial-based GLMs have been used recently for seasonal streamflow forecasting (e.g., Regonda et al. 2006; Grantz et al. 2005; Bracken et al. 2010), flood frequency estimation (Apipattanavis et al. 2010) and turbidity threshold exceedance modeling (Towler et al. 2010a, b). Here we build on these by modeling the four daily water temperature attributes (DTX, DTR, POE, and NHE) at Balls Ferry. We applied the threshold temperature of 13.3°C, as this is the current compliance target for protecting ESA-listed salmon on the Sacramento River. DTX is a critical indicator for the severity of high water temperatures on any given day; however, depending on the magnitude of DTX, there is opportunity for fish to adapt provided that there is a large diurnal range (DTR) and/or that the hours above that threshold are minimal (NHE). The probability of exceedance (POE) measures whether the mean daily temperature exceeds the threshold, and therefore serves as a measure of compliance. Together they provide a comprehensive prediction of the water temperature conditions and thus help in better planning of reservoir operations.

Incorporating Seasonal Climate Information

When using the GLM models for seasonal water resources planning, daily air temperature and precipitation for the season are required to drive the model. The hydrologic variables such as streamflow can be prescribed as a decision variable or optimized such that the optimal solution of flow and temperature releases result in the fewest number of exceedances and least volume of cold water usage (Neumann et al. 2006). Stochastic weather generators can provide ensembles of daily weather variables. There is a rich literature on traditional weather generators and nonparametric weather generators based on K-nearest neighbor time series bootstrap (Lall et al. 1996; Furrer and Katz 2008) and references therein for a review of traditional weather generators and nonparametric methods. The K-nearest neighbor based stochastic weather generators (Rajagopalan and Lall 1999; Yates et al. 2003) have been enhanced with addition of Markov Chains (Apipattanavis et al. 2007) and labeled the semiparametric weather generator (SWG). In this, a daily weather vector for day t is simulated based on weather vector on day $t-1$ and the precipitation states (i.e., wet, dry) on days $t-1$ and t. K-nearest neighbors of the weather vector on day t are obtained from historical days within a small window centered on day t (a 7-day window as proposed in Apipattanavis et al. 2007 was used) and one of them is resampled using a weight function that gives more weighting to the nearest neighbor and least to the farthest (Lall and Sharma 1996); this approach is described as unconditional generation where in the sequences are generated from historical data. For seasonal planning, weather sequences are required that are based on probabilistic seasonal climate forecasts, which are three category (or tercile) probabilities. Each historical year is assigned one of the three probabilities based on its tercile category. For example, if the forecast probabilities are 25:40:35 for A:N:B categories (A being above normal or upper tercile, B being below normal or lower tercile, and N being normal or middle tercile), then years with seasonal precipitation falling in upper tercile get a weight of 0.25, the middle gets a weight of 0.40, and the lower gets 0.35. These weights are used to modify the weights of the K-nearest neighbors from the unconditional case so as to reflect the seasonal forecast information. Thus generated weather sequences

Table 1. Best Set of Predictors, Local Polynomial Parameters and Fitting R^2 for the Four Temperature Attributes by Month

Month/season	NHE	POE	DTR	DTX
January	Intercept only	Intercept only	pcp,tx,tn,pcp1,tn1,q1	tx,tn,q,tw,tx1,tn1,tw1
	2, 0.50, <0.01, NA	*2, 0.70, <0.01, NA*	*2, 0.07, <0.01, 0.22*	*2, 0.07, <0.01, NA*
February	Intercept only	Intercept only	tx,tn,q,tx1	tx,tn,tw,tn1
	1, 0.50, <0.01, NA	*2, 0.55, <0.01, NA*	*2, 0.05, 0.12, 0.18*	*2, 0.05, <0.01, NA*
March	pcp,tn,tw,pcp1,tx1,tn1,tw1	Intercept only	pcp,tx,tn,q	tx,q,tw,pcp1,tx1,tn1,tw1
	2, 0.50, 0.15, 0.14	*2, 0.70, <0.01, NA*	*1, 0.55, 0.11, 0.66*	*2, 0.06, <0.01, <0.01*
April	pcp,tx,tn,tw,tn1,q1,tw1	tx	pcp,tx,tn,q	tx,tx1,tn1,q1,tw1
	2, 0.50, 1.73, 0.68	*2, 0.55, 0.11, 0.03*	*1, 0.70, 0.12, 0.58*	*2, 0.06, <0.01, <0.01*
May	pcp,tx,tn,q,pcp1,tx1	pcp,tx,tn,tw,pcp1,q1,tw1	pcp,tx,tn,q,tx1	tx,q,tx1
	2, 0.50, 2.84, 0.14	*2, 0.55, <0.01, 0.35*	*2, 0.05, 0.01, 0.17*	*2, 0.95, <0.01, 0.02*
June	tx,tn,q,tx1,tn1,tw1	tw	pcp,tx,tn,pcp1,tx1,q1	tx,tn,q1,tw1
	1, 0.50, 2.73, 0.53	*1, 0.50, 0.05, 0.25*	*2, 0.31, 0.03, 0.07*	*2, 0.05, <0.01, 0.48*
July	tx,q,tw,tn1	Intercept only	pcp,tx,tn,q	q,tw
	2, 0.95, 2.58, 0.36	*2, 0.65, 0.08, 0.23*	*2, 0.40, 0.03, 0.48*	*1, 0.05, <0.01, 0.53*
August	pcp,tn,q,tw,tw1	Intercept only	pcp,tx,tn,tx1	pcp,tx,q,tw,tw1
	2, 0.90, 1.57, 0.48	*2, 0.50, 0.03, 0.05*	*1, 1.00, 0.08, 0.15*	*2, 0.05, <0.01, 0.03*
September	pcp,tx,tn,tw,tx1,q1,tw1	tn,q1,tw1	pcp,tx,q,tx1	tx,tw,tx1,tw1
	2, 0.65, 2.61, 0.55	*2, 0.75, 0.45, 0.54*	*2, 0.55, 0.02, 0.50*	*2, 0.05, <0.01, 0.22*
October	pcp,tn,q,tw,tx1,q1,tw1	tn,tw,tx1	pcp,tx,tn,tw	tx,tn,tw,tw1
	1, 0.50, 4.75, 0.57	*2, 0.50, 0.58, 0.60*	*2, 0.85, 0.03, 0.72*	*2, 0.05, <0.01, 0.21*
November	pcp,tx,tn,tx1,tw1	tx,tn,tw,tx1,tw1	tx,tn	tx,tn,tx1,tw1
	2, 0.50, 3.17, 0.66	*2, 0.75, 0.32, 0.66*	*1, 0.25, 0.09, 0.60*	*2, 0.05, <0.01, 0.07*
December	pcp,tw,pcp1,tn1	pcp,tn1,tw1	pcp,tx,tn,q,tx1	tx,tn,tw,tx1,q1,tw1
	1, 0.50, 0.02, 0.85	*2, 0.85, <0.01, 1.00*	*2, 0.06, <0.01, 0.02*	*2, 0.07, <0.01, 0.99*
JAS	pcp,tx,tn,q,tw,tx1,tw1	tx,q1,tw1	pcp,tx,tn,tx1,tw1	pcp,tx,q,tw,tw1
	2, 0.50, 2.11, 0.61	*2, 0.50, 0.24, 0.44*	*2, 0.97, 0.05, 0.27*	*2, 0.02, <0.01, 0.36*

Note: Variables listed are daily values of precipitation (pcp), maximum air temperature (tx), and minimum air temperature (tn) at Redding, California; and, water temperature (tw) and flow released (q) from Shasta Dam. Variables appended with a "1" indicate prior day values. Water temperature attributes include number hours of exceedance (NHE), probability of exceedance (POE), daily temperature range (DTR), and daily maximum temperature (DTX). Bottom line of values provides the polynomial order, alpha, GCV, and R^2 separated by commas. R^2 was computed only for values above the thresholds provided in Table 2.

are described as conditional generation and are consistent with the seasonal climate forecasts. This is demonstrated in Apipattanaivis et al. 2007 and also in a recent modified version of this approach in Caraway et al. (2014). This weather generator has been applied to crop modeling and agriculture planning in Argentina (Podesta et al. 2009) and has been implemented for the study region in this research (Caldwell 2013). Using the SWG, daily weather sequences are generated which are then incorporated in the GLM models to obtain ensembles of various water temperature attributes, and ultimately, cumulative distribution functions (CDFs). Using the CDFs, it is possible to compute the probability of exceeding threshold values for each water temperature attribute over the seasonal planning horizon by calculating the area under each CDF curve relative to any appropriate threshold value(s). Water managers will then be informed of the relative change in risk for the upcoming season relative to climatology. This study used the observed mean values from the window of July to September for the period 1994–2007 and from the same window in 2008, predicted and observed to be both warmer and drier than normal, to indicate the utility of seasonal forecasts information in operations as hypothetical planning scenarios.

Model Evaluation

The model framework using the daily water temperature data was applied at the Balls Ferry compliance point and utilized daily weather data from the Redding Airport, Redding, California (Fig. 1), for the period 1994–2007. Daily and hourly streamflow data, and release volume and release temperature from Shasta Dam are also available for this period (CDEC 2011). From the hourly water temperature data, DTX, DTR, POE, and NHE were computed for each day. Daily meteorological values from Redding (NCDC, http://www.ncdc.noaa.gov) and Shasta Dam release temperature and flow served as predictors for each variable for each month (Table 1). Then the predictive skill of the fitted GLM models was evaluated using a cross validation model. For this procedure, ten percent of the data was randomly excluded, then the model was fitted using the remaining data, and the excluded values were predicted. The standard root mean square error (RMSE) for this prediction was computed, with the process repeated 250 times. This is a robust method for skill evaluation as it stresses the model more than the traditional approach where in the model is fitted on a period of data and validated on another. Besides, this provides a distribution of the model skill and not a single value.

Ensembles of daily weather were generated from the SWG; while ensembles of the water temperature attributes were generated using the GLM models. The observed daily streamflow and release temperatures on the selected days were used as surrogates for standard operating procedures. Thus, generated attributes are displayed as boxplots along with the corresponding mean values from the observations. This exercise is designed to demonstrate the ability of capturing the historic variability of water temperature attributes. A total of 100 simulations of daily weather were generated, each 14 years in length.

The utility of seasonal climate forecast in generating conditional weather sequences and consequently, conditional stream temperature attributes was then demonstrated. Four representative seasonal climate forecasts for conditional simulations for the period July through September—dry (D), hot (H), very dry (VD), and very hot (VH) conditional scenarios were used. The A:N:B ratios associated with each scenario are as follows: for temperature—hot (40:35:25) and very hot (60:35:05); and for precipitation—dry (25:35:40) and very dry (05:35:60). Stream temperature attributes from these conditional scenarios are compared with climatology (CL) in their probability distribution functions to indicate changes relative to climate forecast input. In particular, this study used the seasonal climate forecast issued by the International Research Institute for Climate and Society in June 2008 for the summer of 2008 (July through September) for the dry and hot scenarios (Fig. 3) to conditionally generate daily weather sequences for the summer season and, consequently, the distribution functions of water temperature attributes.

Results

The developed methods were applied for the entire year; however, the primary months of concern for cold water pool management stretch from May through October. Here, the results focus on a portion of the summer months (July–September) when the water temperatures on the Sacramento River are the highest and have the greatest management implications. The best GLM model predictors was selected from a suite of ten using BIC on a global fit and then the local polynomial GLM was estimated using GCV for each water temperature attribute by month, shown in Table 1. For a number of months and attributes, the parameters α and p deviate from 1, indicating local nonlinear features. The best predictors for NHE and POE indicate the intercept only fits during the months of January and February when no or very few days exceed 13.3°C, thus obviating the need for predictors. In addition, POE has intercept-only fits during the months of July and August, where the complement is true and most days exceed 13.3°C. In general, for NHE more predictors are included compared to other water temperature attributes, due to the fact that this is somewhat of a noisy variable relative to others. In many cases, prior day values are often selected as one of the variables to account for lagged and persistence that is present in the system as described previously.

During the period of July–September, scatterplots of maximum air temperature and DTR, with a local polynomial smoother [Fig. 2(a)] indicate that DTR is linearly proportional to maximum air temperature up to 30°C, after which it reaches an asymptote at DTR ~2.5°C. There is an upper limit to the diurnal range when maximum air temperature reaches 30°C. NHE is poorly correlated to daily mean water temperature releases at Shasta for temperatures less than 8°C [Fig. 2(b)], but are linearly proportional for values above 8°C. Such nonlinear features also exist for other stream temperature variables (figures not shown), which underscores the utility of the application of a local polynomial GLM.

The model is generally able to predict the observations of the four daily temperature attributes during the July–September period quite well (Fig. 4) with fitting R^2 values ranging from 0.27 (for DTR) to 0.61 (for NHE and DTX). For POE, this is not calculated as R^2 is not appropriate. The NHE is well modeled, with the exception of an underestimation for higher observed hours of exceedance, especially 15 h and beyond [Fig. 4(a)]. The probability of exceedance [Fig. 4(b)] too is under estimated (i.e., lower probability values when there is exceedance and vice versa). The predictions of DTR values greater than 2.5°C is under estimated [Fig. 4(c)] and somewhat less so in the case of DTX [Fig. 4(d)] for values beyond 16°C. DTR and DTX are influenced by the clustering of values in the observed data between 1–3°C and 13–15°C, respectively.

To test the predictive skill of the models, the RMSE was computed in a cross validation model by dropping 10% of the observations at random repeated for 250 times. As mentioned in a preceding section this is a robust approach to evaluating predictive skill and has been used in water quality applications

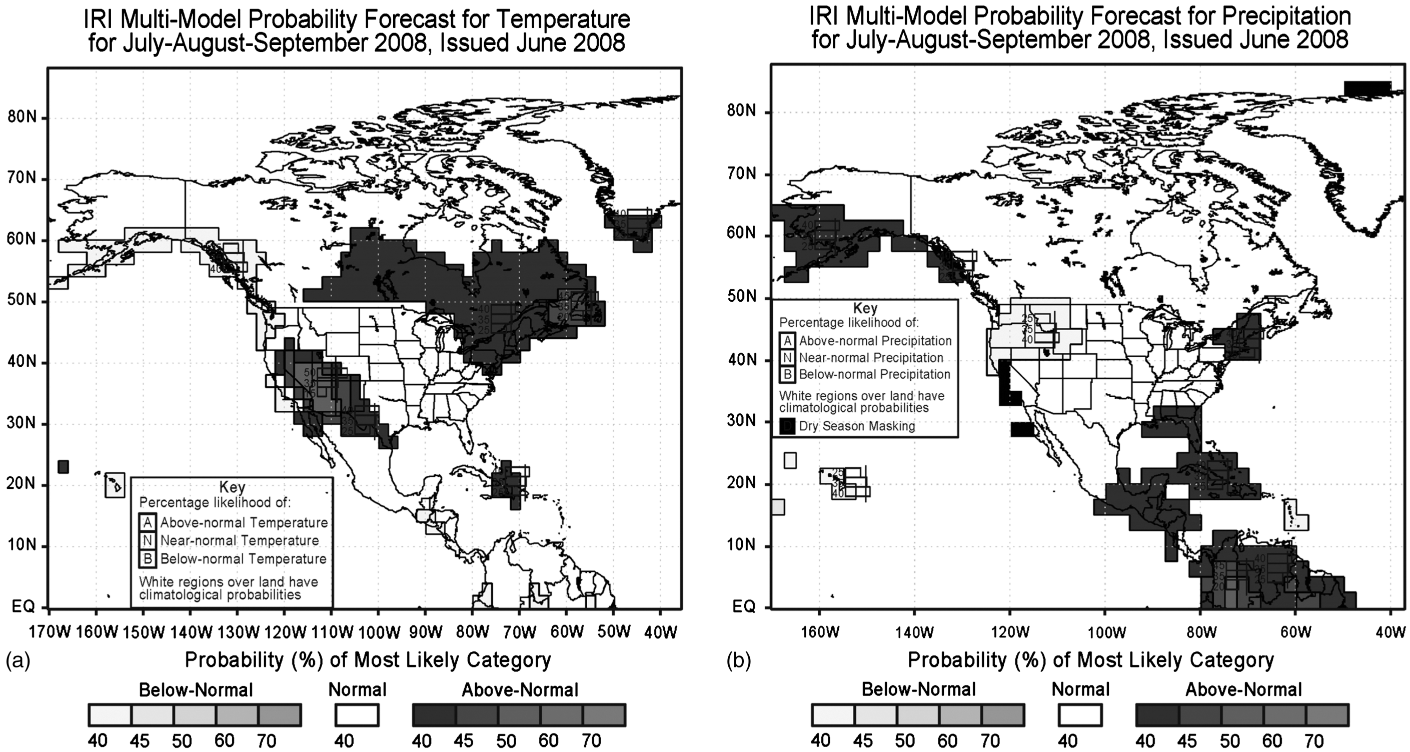

Fig. 3. Probabilistic seasonal climate forecasts of temperature (a) and precipitation; (b) for the period July–September 2008 as applied in the dry and hot conditional scenarios for the weather generator (reprinted from IRI; figure converted to greyscale from IRI website at http://iri.columbia.edu/our-expertise/climate/forecasts/seasonal-climate-forecasts/)

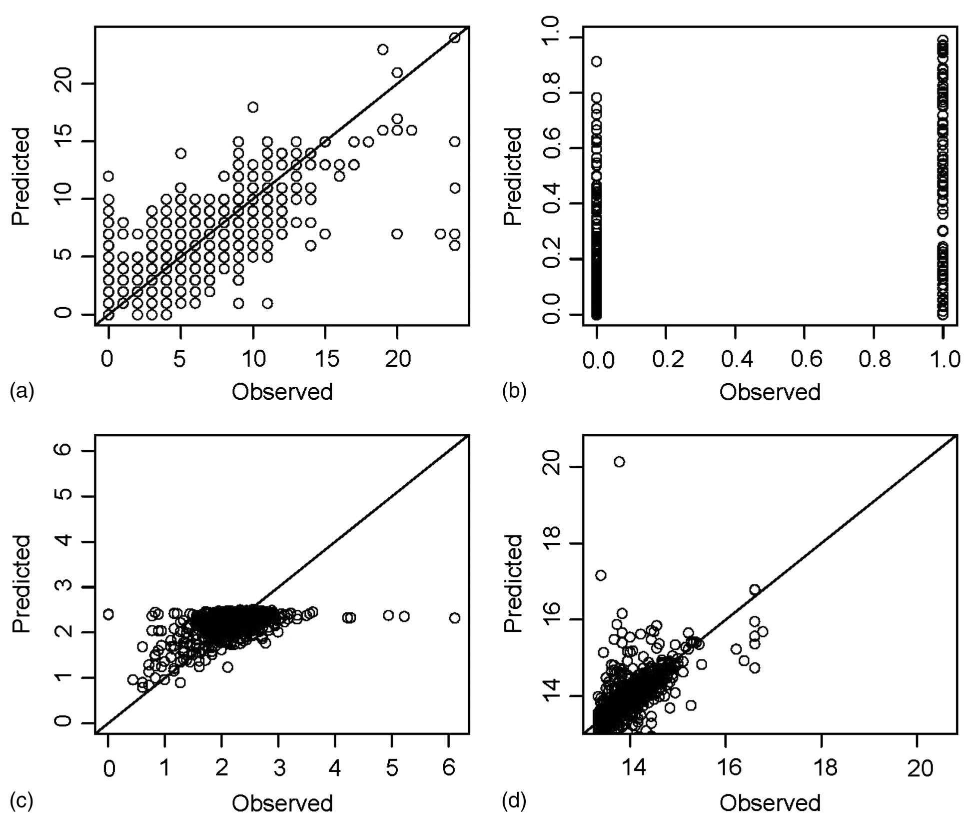

Fig. 4. Comparison of observed and predicted values from the GLM for (a) NHE [hours]; (b) POE [probability]; (c) DTR [°C]; (d) DTX [°C] for the season of July through September

(Towler et al. 2009; Zachman et al. 2007). For NHE and POE, the skills are generally best with lower RMSE values during January to June, with higher RMSE values during the summer and fall seasons [Figs. 5(a and b)]. Although, the RMSE values are higher in summer and fall, for NHE the actual magnitude difference between winter and spring versus summer and fall is 1. POE had poor skill in the fitting as shown in Fig. 4 and this carries to the prediction skill. For DTR and DTX, mean RMSE values are highest during the winter and spring when flood control and snowmelt runoff are more dominant drivers of water temperatures than meteorology or standard operations [Figs. 5(c and d)]. Variability in the RMSE, however, is low during the summer months for DTR and DTX, except for August for DTX, indicating a greater level of confidence in the predicted values. The mean RMSE values consistently range between 0 and 1°C for DTR and DTX, which may be a tolerable threshold for seasonal planning efforts.

Simulating Water Temperature Attributes from Unconditional Daily Weather Ensembles

The SWG was used to generate ensembles of daily weather sequences based just on the historical data—i.e., unconditional simulations—and, consequently generated ensembles of daily water temperature attributes from the GLM models and displayed them as boxplots for each month along with the observed monthly mean values. The weather generator ensembles provide a rich variety in the water temperature attributes and capture the historical mean very well in the boxes (Fig. 6). Exceptions include: (1) underestimation of NHE during the months of April, May and November, and overestimation in June [Fig. 6(a)]; and (2) underestimation of POE during the months of September and November [Fig. 6(b)]. Any biases in the DTR and DTX are not discernible [Figs. 6(c and d)]. These results indicate that the coupling of a SWG to the local polynomial GLM approach to predict and simulate water temperature attributes is quite robust.

Water Temperature Attributes Conditional on Seasonal Forecast

Conditional daily weather sequences were generated for the July–September period for the four climate scenarios (D, VD, H, and VH) described earlier. For the D and H scenarios, forecasts for 2008 (seen in Fig. 3) were used. The generated daily weather ensembles are provided with the local polynomial GLM to provide ensembles of water temperature attributes and the respective CDFs from the ensembles and climatology (Fig. 7). The steepness of the slope of a CDF curve can be interpreted as a larger contribution to the cumulative probability across a given range of values; the probability of falling between the values is determined by the difference in cumulative probability values on the *y*-axis. For NHE less than 3 h and greater than 12 h, all of the CDFs of NHE during warmer and drier conditions (dashed) indicate enhanced risk (i.e., increased probability of exceedance) compared to climatology (black) [Fig. 7(a)]. The largest risk of NHE > 12 h occur in the very hot conditional simulations with a probability of ~0.40, compared to the other conditional simulations (~0.20) and climatology (<0.10). Similarly, the conditional runs indicate an increased risk of POE compared to climatology [Fig. 7(b)]. The very hot simulation estimated the greatest shift in POE with the steepest slope at POE values >0.60 compared to other curves. There is an increased probability of decreased diurnal range (DTR) in the conditional simulations, as expected in a hotter regime—this can be seen by

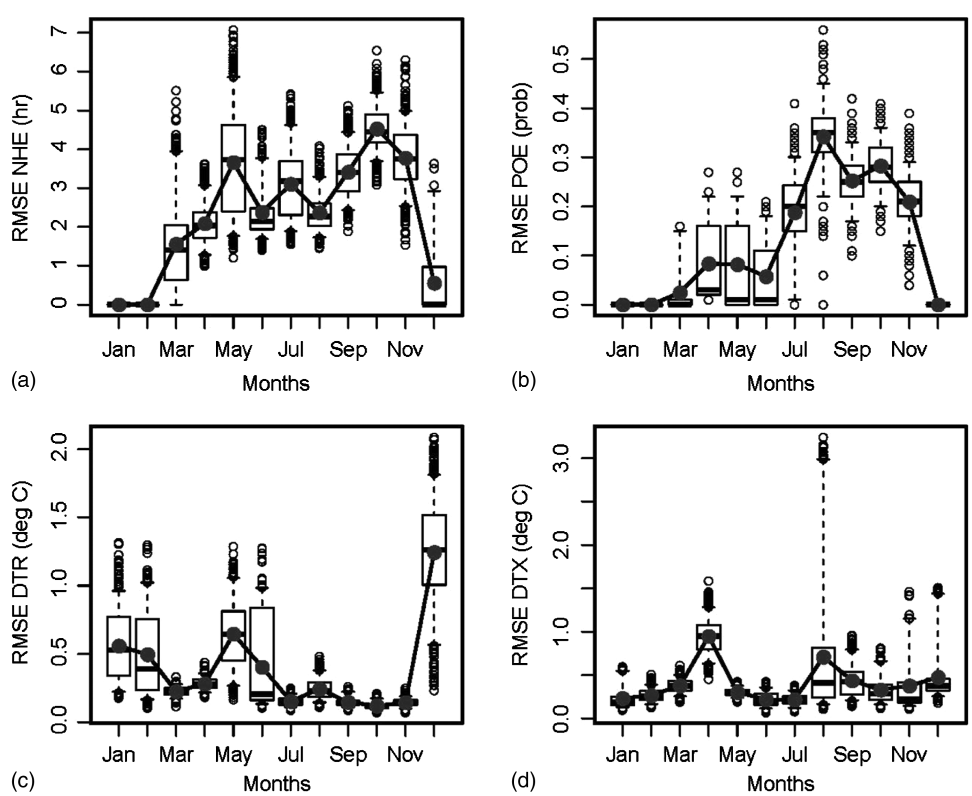

Fig. 5. Cross validation model results using RMSE to examine skill for (a) NHE; (b) POE; (c) DTR; (d) DTX by month; observed climatological means of each variable (points connected with line) are shown; boxplots provide the median as the horizontal line; box height is the interquartile range, whiskers indicate 5th and 95th percentiles, and hollow points indicate values outside the whiskers

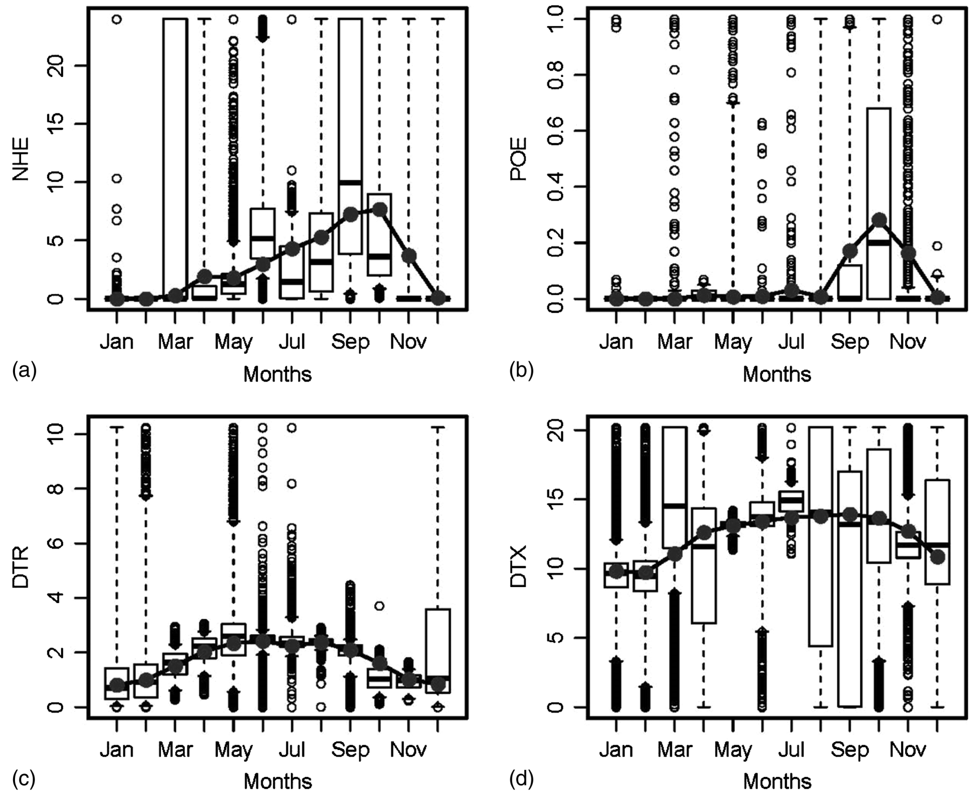

Fig. 6. Comparison of observed and unconditional simulations of (a) NHE; (b) POE; (c) DTR; (d) DTX using the GLM coupled with the SWG by month; observed climatological means of each variable (points connected with line) are shown; boxplots of unconditional simulated values provide median as horizontal line; box height is the interquartile range, whiskers indicate 5th and 95th percentiles, and hollow points indicate values outside the whiskers

the CDFs being shifted to lower DTR values relative to climatology [Fig. 7(c)]. Likewise, there is an increased risk of higher DTX [Fig. 7(d)] compared to climatology.

As described in the comparison to climatology, these exceedance probabilities can provide an indication of risk with respect to thresholds relevant for management. For example, using 2008 mean values of the temperature attributes as a reference, the exceedance probabilities for the four climate scenarios are computed from the CDFs and listed in Table 2, and their relative changes in risk from climatological risk are shown as a barplot in Fig. 8. All of the conditional scenarios show increased risk of NHE greater than 13.35 h, indicating a tendency for more days with mean water temperature above 13.3°C (Table 2). In addition, there is a decreased risk of POE above 0.79, except for the very hot scenario (Table 2 and Fig. 8). Increased risk is noted for DTR < 2.33°C, which would suggest that recovery time for fish will be diminished during hotter or drier than normal conditions if mean water temperatures are high, regardless of the magnitude of the climate shift. Despite the fact that the GLM model generally underestimates higher values of DTX (Fig. 4), increases in risk of DTX > 14.78°C are evident for all conditional scenarios (Table 2 and Fig. 8).

Summary and Discussion

This study reports the development of a complementary statistical modeling tool using local polynomial based GLMs that provides monthly to seasonal forecasts of key water temperature attributes and probabilistic estimates of risk of meeting or exceeding predetermined thresholds of each attribute. The GLM framework can model a variety of variables such as discrete, binary, and continuous, among others. For example, models were fitted to predict a variety of water temperature attributes such as number of hours of exceedance (discrete), probability of temperature exceeding a threshold (binary), and daily maximum water temperature and daily water temperature range (both continuous). Models were fitted for each month and for each variable separately using a large pool of predictor variables based on atmospheric variables (e.g., temperature, precipitation from current and previous day) and water variables (e.g., flow, temperature from previous day). Based on cross validated skill scores, the models performed well, especially during the summer months of interest. A stochastic weather generator (SWG) was applied to generate ensembles of daily weather sequences in an unconditional manner (i.e., assuming all of the historical years are equally likely) and conditional manner (based on probabilistic seasonal climate forecast). Local polynomial GLM ensembles of water temperature attributes were also generated. These ensembles were consistent with the seasonal forecast, demonstrating the ability of the proposed methodology to provide projections of water temperature attributes before the start of the season. The ensembles of water temperatures provide the estimates of risk of exceeding various compliance thresholds. These risk estimates can be of immense help to water managers in making plans for additional water or changes in operations before the start of the season to help mitigate water temperature risk in a sustainable manner.

This integration of the GLM and SWG allows investigation of the relative change in risk of meeting temperature criteria at Balls Ferry by performing both unconditional and conditional simulations. As a proof of concept, the threshold values from 2008 were used to indicate relative changes of risk captured by the conditional

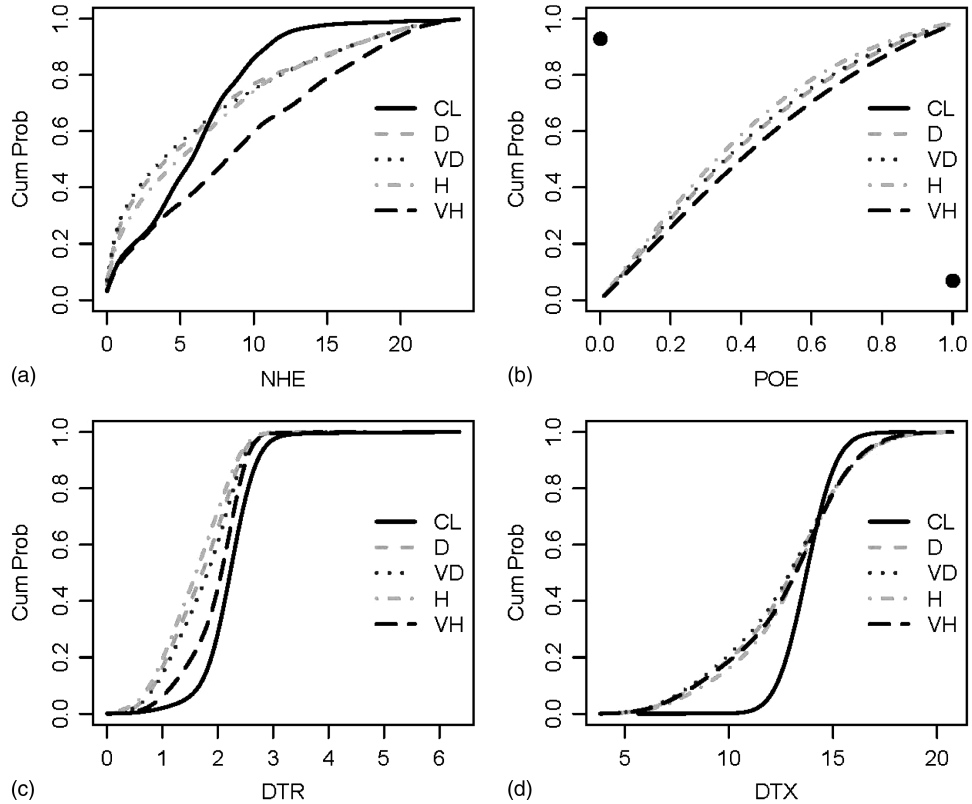

Fig. 7. Cumulative distribution functions for the season of July through September for (a) NHE; (b) POE; (c) DTR; (d) DTX for the observed climatology (CL), dry (D), very dry (VD), hot (H), and very hot (VH) conditional simulations

scenarios (Fig. 8). For NHE, DTR, and DTX, the forecasts issued in June 2008 would have been sufficient to convey an increased risk of violating the thresholds from Table 2; however, a highly skewed forecast—like that offered in the very hot scenario—would have been required to suggest a POE value of 0.79 for the three-month period. Though not applied in the current study, the flows and water temperature releases associated with each conditional scenario could be adjusted from the observed values used in the GLM to modify the predicted values of each water temperature attribute.

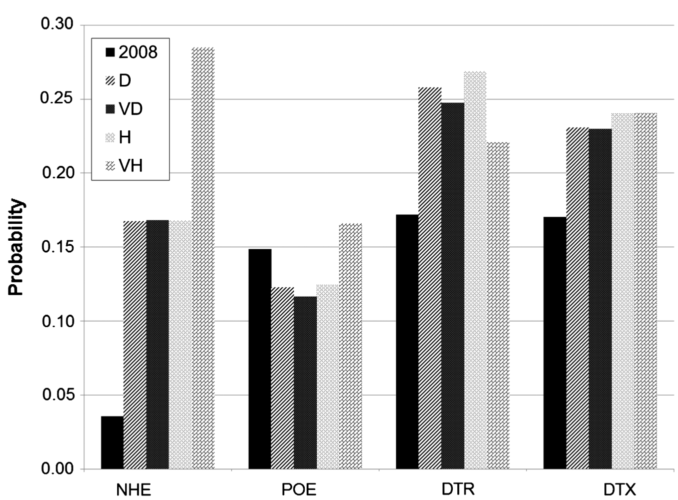

Fig. 8. Probabilities of threshold criteria (from Table 2) for the 2008 observed, dry, very dry, hot, and very hot conditional scenarios plotted in that order for each threshold criteria

For example, operators could apply a designated 90-day flow and temperature regime derived from historical data (i.e., a prior extremely hot or dry year) to adjust the predicted values from the GLM until the relative risk is reduced to a climatological value or other acceptable level. In essence, multiple flow and temperature regimes could be applied through the GLM to determine an optimal solution for reducing cold water usage and maintaining temperatures downstream.

Atmospheric variables are typically well-correlated with water temperatures, particularly during the summer months when solar radiation is at a maximum. In addition, water temperature is generally inversely proportional to flow as larger volumes of water take longer to warm and cool. The skill of the GLM is directly proportional to the strength of these correlations between the hydrometeorological predictors and the water temperature at Balls Ferry. During the summer months, the water temperatures on the Sacramento River are also strongly influenced by the temperature and volume of water released from Shasta Dam. The interactions between the environment and operations are highly nonlinear and, therefore, the local polynomial based GLM is capable of predicting

Table 2. Probabilities of Threshold Criteria Computed from the CDFs in Fig. 7

Variable/threshold	CL	D	VD	H	VH
NHE > 13.4	0.04	0.17	0.17	0.17	0.28
POE > 0.79	0.15	0.12	0.12	0.12	0.17
DTR < 2.33 C	0.17	0.26	0.25	0.27	0.22
DTX > 14.78 C	0.17	0.23	0.23	0.24	0.24

Note: Threshold values correspond to July–September seasonal means from 2008.

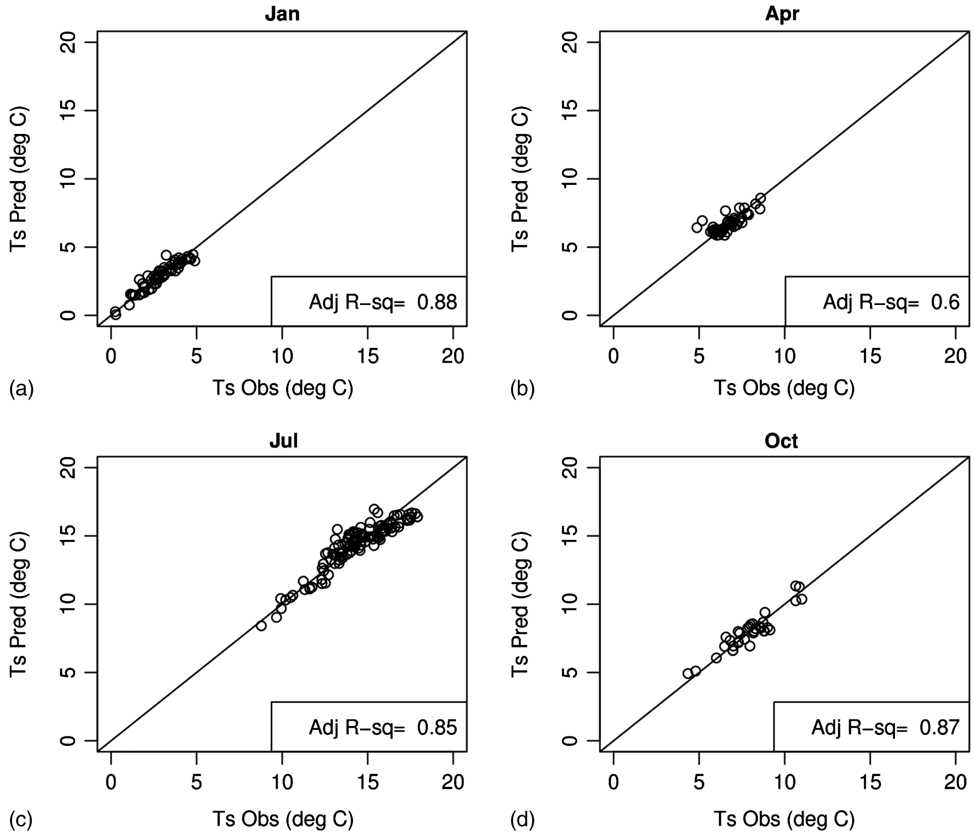

Fig. 9. Scatterplots of observed versus predicted values for January, April, July, and October from the GLM model fits in the unregulated Methow River (reprinted from "Statistical modeling of daily and subdaily stream temperatures: Application to the Methow River Basin, Washington" by R. J. Caldwell et al., *Water Resources Research*, 49, ©2013 American Geophysical Union. Reproduced with permission.)

the response in temperature attributes. Predictor variables from the prior day state of the hydrologic system improve the model fits by including residence time of water release impacts and persistence into the model (i.e., if the prior day water temperatures are cool, there is a natural tendency for today to also be cool).

Since water management (e.g., power generation) may involve subdaily management of releases (Carron and Rajaram 2001), the GLM might be even more effective during the summer months if subdaily dam operations were included in the model fitting process, along with specific information on releases from the temperature control devices on Shasta Dam. Unfortunately, these data are either unavailable, discontinuous, or require reconstruction using detailed hydraulic modeling efforts. To some degree, this relationship was included in the GLM through the mean daily flow and water temperature released at Shasta. This framework was applied to modeling stream temperatures in the Methow River basin, an unregulated system in the State of Washington (Caldwell et al. 2013), to assess the impact of climate change on fish habitat. Results from this application (Fig. 9, reproduced from Caldwell et al. 2013) indicate that model predictions are very good in comparison to the current application (Fig. 4). This indicates that the methodology is portable to other watersheds and can provide improved skill when water management impacts are minimal.

Water projects in the western United States have fundamentally altered temperature regimes in major rivers, particularly downstream of large dams. While dams such as Shasta can selectively release colder water to meet downstream temperature criteria, the current operations approach does not have sufficient forecasting capabilities. High resolution water temperature models have been developed to improve forecasting; but, these models are limited to forecasts of hours to several days, and cannot provide seasonal-scale planning guidance in a timely manner, unless coupled with input from statistical models. This modeling approach provides skillful management options in a decision support system (Caldwell 2013), thus, offering potential for efficient management of river systems to mitigate river temperature impacts on fish habitat.

Protection of the coldwater fish habitat in the Sacramento River is a challenge in this highly altered river system. Careful and innovative management strategies are needed, as any additional changes in water temperature in response to climate could result in conditions that favor nonnative species (May and Brown 2002). Yates et al. (2008) suggest that future warming in air temperatures of 2–4°C could lead to additional threshold temperature exceedances, particularly in August and September of drought years. In addition, maintaining the cold pool in Shasta Lake would be difficult through the summer and fall (Yates et al. 2008). As such, additional research is needed to improve seasonal forecasts of water temperatures during the critical summer and late fall period. Integration of the GLM with hydraulic models of Shasta Lake would also be beneficial by ensuring upstream criteria within the reservoir are met and by providing an input for flow and temperature release information to the GLM, as opposed to the use of a surrogate such as the simulated flow and temperature values from the SWG. Optimization techniques could then be applied directly to monitor cold water storage and both in-lake and downstream habitat.

Acknowledgments

This research is funded by the National Aeronautics and Science Administration's Earth-Sum Science Applied Sciences Program, Grant # NNX08AK72G. Andrew Pike NMFS (Southwest Fisheries Science Center) and Russ Yaworsky (U.S. Bureau of Reclamation) provided helpful guidance on thermal criteria and reservoir operations, respectively. We thank the three anonymous reviewers, editor and associate editor for their valuable comments and suggestions which helped improve the manuscript.

References

Apipattanavis, S., Podestá, G., Rajagopalan, B., and Katz, R. W. (2007). "A semiparametric multivariate and multisite weather generator." *Water Resour. Res.*, 43(11), W11401.

Apipattanavis, S., Rajagopalan, B., and Lall, U. (2010). "Local polynomial–based flood frequency estimator for mixed population." *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0000242, 680–691.

Benyahya, L., Caissie, D., St-Hilaire, A., Ouarda, T. B. M. J., and Bobee, B. (2007). "A review of statistical water temperature models." *Can. Water Resour. J.*, 32(3), 179–192.

Bettelheim, M. (2001). *Temperature and flow regulation in the Sacramento River and its effect on the Sacramento pikeminnow: A literature review*, State of Colorado, Dept. of Fish and Game.

Bogan, T., Othmer, J., Mohseni, O., and Stefan, H. (2006). "Estimating extreme stream temperatures by the standard deviate method." *J. Hydrol.*, 317(3–4), 173–189.

Bracken, C., Rajagopalan, B., and Prairie, J. (2010). "A multisite seasonal ensemble streamflow forecasting technique." *Water Resour. Res.*, 46(3), W03532.

Brock, J. T., and Caupp, C. L. (1996). "Application of a DSSAMt water quality model—Truckee River, Nevada for Truckee River Operating Agreement (TROA) DEIS/DEIR: Simulated river temperatures for TROA." *Technical Rep. No. RCR96-7.0*, Rapid Creek Research, Boise, ID.

Caissie, D., El-Jabi, N., and Satish, M. G. (2001). "Modeling of maximum daily water temperatures in a small stream using air temperatures." *J. Hydrol.*, 251(1–2), 14–28.

Caldwell, R. J. (2013). "An integrated framework for modeling and mitigating water temperature impacts in the Sacramento River." Ph.D. dissertation, Univ. of Colorado, Boulder, CO, 208.

Caldwell, R. J., Gangopadhyay, S., Bountry, J., Lai, Y., and Elsner, M. M. (2013). "Statistical modeling of daily and subdaily stream temperatures: Application to the Methow River basin, Washington." *Water Resour. Res.*, 49(7), 4346–4361.

California Data Exchange Center (CDEC). (2011). *Hydrologic data download*, California Dept. of Water Resources, ⟨http://cdec.water.ca.gov/⟩ (Jan. 9, 2011).

Caraway, N. M., McCreight, J. L., and Rajagopalan, B. (2014). "Multisite stochastic weather generation using cluster analysis and k-nearest neighbor time series resampling." *J. Hydrol.*, 508, 197–213.

Carron, J. C., and Rajaram, H. (2001). "Impact of variable reservoir releases on management of downstream water temperatures." *Water Resour. Res.*, 37(6), 1733–1743.

Chandler, R. E. (2005). "On the use of generalized linear models for interpreting climate variability." *Environmetrics*, 16(7), 699–715.

Chandler, R. E., and Wheater, H. S. (2002). "Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland." *Water Resour. Res.*, 38(10), 1–11.

Chenard, J.-F., and Caissie, D. (2008). "Stream temperature modelling using artificial neural networks: Application on Catamaran Brook, New Brunswick, Canada." *Hydrol. Processes*, 22(17), 3361–3372.

Danner, E. M., et al. (2012). "River temperature forecasting: A coupled-modeling framework for management of river habitat." *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 5(6), 1752–1760.

Dunham, J. B., Chandler, G., Rieman, B. E., and Martin, D. (2005). "Measuring stream temperature with digital dataloggers: A user's guide." *GTR-RMRS-150*, USDA Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

Furrer, E. M., and Katz, R. W. (2007). "Generalized linear modeling approach to stochastic weather generators." *Clim. Res.*, 34, 129–144.

Furrer, E. M., and Katz, R. W. (2008). "Improving the simulation of extreme precipitation events by stochastic weather generators." *Water Resour. Res.*, 44(12), W12439.

Grantz, K., Rajagopalan, B., Clark, M., and Zagona, E. (2005). "A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts." *Water Resour. Res.*, 41, W10410.

Hallock, R. J., and Fisher, F. W. (1985). *Status of winter-run Chinook salmon in the Sacramento River*, Anadromous Fisheries Branch, California Dept. of Fish and Game, 28.

Isaak, D. J. (2011). "Stream temperature monitoring and modeling: Recent advances and new tools for managers." *Stream notes*, Stream Systems Technology Center, Rocky Mountain Research Station, USDA, Fort Collins, CO, 1–7.

Lall, U., Rajagopalan, B., and Tarboton, D. G. (1996). "A Nonparametric wet/dry spell model for daily precipitation." *Water Resour. Res.*, 32(9), 2803–2823.

Lall, U., and Sharma, A. (1996). "A nearest neighbor bootstrap for time series resampling." *Water Resour. Res.*, 32(3), 679–693.

Loader, C. (1999). *Local regression and likelihood, statistics and computing*, Springer, New York, 290.

May, J. T., and Brown, L. R. (2002). "Fish communities of the Sacramento River basin: Implications for conservation of native fishes in the Central Valley, California." *Environ. Biol. Fishes*, 63(4), 373–388.

McCullagh, P., and Nelder, J. A. (1989). *Generalized linear models*, Chapman and Hall, London.

Mohseni, O., Stefan, H. G., and Erickson, T. R. (1998). "A nonlinear regression model for weekly stream temperatures." *Water Resour. Res.* 34(10), 2685–2692.

Myrick, C. A., and Cech, J. J., Jr. (2000). "Swimming performances of four California stream fishes: Temperature effects." *Environ. Biol. Fishes*, 58(3), 289–295.

Neumann, D. W., Rajagopalan, B., and Zagona, E. A. (2003). "Regression model for daily maximum stream temperature." *J. Environ. Eng.*, 10.1061/(ASCE)0733-9372(2003)129:7(667), 667–674.

Neumann, D. W., Zagona, E. A., and Rajagopalan, B. (2006). "A decision support system to manage summer stream temperatures." *J. Am. Water Resour. Assoc.*, 42(5), 1275–1284.

Pike, A., et al. (2013). "River temperature prediction: Revisiting a stochastic dynamics approach." *Water Resour. Res.*, 49(9), 5168–5182.

Podesta, G., et al. (2009). "Decadal climate variability in the Argentine Pampas: Regional impacts of plausible climate scenarios on agricultural systems." *Clim. Res.*, 40, 199–210.

Rajagopalan, B., and Lall, U. (1999). "A k-nearest neighbor simulator for daily precipitation and other weather variables." *Water Resour. Res.*, 35(10), 3089–3101.

Regonda, S. K., Rajagopalan, B., Clark, M., and Pitlick, J. (2005). "Seasonal cycle shifts in hydroclimatology over the western United States." *J. Clim.*, 18(2), 372–384.

Regonda, S. K., Rajagopalan, B., Clark, M., and Zagona, E. (2006). "A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River basin." *Water Resour. Res.*, 42(9), W09404.

San Francisco Estuary Project (SFEP). (1992). *State of the estuary: A report on conditions and problems in the San Francisco Bay/ Sacramento-San Joaquin Delta Estuary*, SFEP, Oakland, CA.

Taylor, R. L. (1998). "Simulation of hourly stream temperature and daily dissolved solids for the Truckee River, California and Nevada." *Water Resources Investigations Rep. No. WRI 98-4064*, U.S. Geological Survey, Reston, VA.

Towler, E., Rajagopalan, B., Gilleland, E., Summers, R. S., Yates, D., and Katz, R. W. (2010a). "Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory." *Water Resour. Res.*, 46(11), W11504.

Towler, E., Rajagopalan, B., and Summers, R. S. (2009). "Using parametric and nonparametric methods to model total organic carbon, alkalinity, and pH after conventional surface water treatment." *Environ. Eng. Sci.*, 26(8), 1299–1308.

Towler, E., Rajagopalan, B., Summers, R. S., and Yates, D. (2010b). "An approach for probabilistic forecasting of seasonal turbidity threshold exceedance." *Water Resour. Res.*, 46(6), W06511.

United States Bureau of Reclamation (USBR). (2004). *Long-term central valley project operations criteria and plan*, U.S. Dept. of Interior, Bureau of Reclamation, Mid-Pacific Region, Sacramento, CA, 238.

Van Vleck, G. K., Deukmejian, G., and Kennedy, D. N. (1988). *Water temperature effects on Chinook salmon with emphasis on the Sacramento River—A literature review*, Dept. of Water Resources, Northern District, CA. http://www.water.ca.gov/pubs/environment/fish/water_temperature_effects_on_chinook_salmon-emphasis_on_the_sacramento_river__a_literature_review/watertempeffecchinooksalmon.pdf.

Venables, W. N., and Ripley, B. D. (2002). *Modern applied statistics with S*, Springer, New York.

Webb, B. W., Hannah, D. M., Moore, R. D., Brown, L. E., and Nobilis, F. (2008). "Recent advances in stream and river temperature research." *Hydrol. Processes*, 22(7), 902–918.

Weirich, S. R., Silverstein, J., and Rajagopalan, B. (2011). "Effect of average flow and capacity utilization on effluent water quality from US municipal wastewater treatment facilities." *Water Res.*, 45(14), 4279–4286.

Yates, D., et al. (2008). "Climate warming, water storage, and Chinook salmon in California's Sacramento Valley." *Clim. Change*, 91(3–4), 335–350.

Yates, D., Gangopadhyay, S., Rajagopalan, B., and Strzepek, K. (2003). "A technique for generating regional climate scenarios using a nearest neighbor algorithm." *Water Resour. Res.*, 39(7), 1199.

Zachman, B., Rajagopalan, B., and Summers, R. S. (2007). "Modeling NOM breakthrough GAC adsorbers using nonprametric regression techniques." *Environ. Eng. Sci.*, 24(9), 1280–1296.