# Managing Water Temperature TMDLs
# Under Economic and Environmental Uncertainty

Jonathan D. Kaplan
Department of Economics
California State University, Sacramento

Richard E. Howitt
Department of Agricultural and Resource Economics
University of California, Davis

Michael L. Johnson
John Muir Institute and Ecotoxicology Program
University of California, Davis

Joshua H. Viers
Information Center for the Environment
University of California, Davis

**Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Denver, Colorado, August 1-4, 2004**

Managing Water Temperature TMDLs
Under Economic and Environmental Uncertainty

## I. Introduction

The US Environmental Protection Agency (EPA) is in the process of drafting Total Maximum Daily Load (TMDL) standards for waters designated as impaired under the Clean Water Act. As of 1998 the EPA listed over 300,000 miles of rivers and streams, and 5 million acres of lakes throughout the United States that are impaired by a number of pollutants (EPA 2003). A TMDL for a particular watershed defines the sum of the allowable load of a single pollutant from all contributing sources of pollution including a margin of safety to account for seasonal variation in water quality.

The specificity of any strategy to manage a TMDL will be limited by the degree of uncertainty about the environmental and economic activity within the impaired watershed and how they are interrelated. Further complicating the problem is the spatial heterogeneity among environmental and economic activity within the watershed. If the activities throughout the watershed are highly dispersed, a single uniformly adopted strategy for meeting TMDL objectives may be rendered ineffective. Needless to say, environmental agencies at the local, state, and federal levels face a daunting task of designing strategies to achieve TMDL objectives given uncertainty about the environmental processes within a watershed and the co-existing patchwork of environmental and economic activity within a watershed, which may be intricately interrelated.

If an environmental agent acquires information (thus reducing the uncertainty) about the environmental and economic characteristics of a watershed, then the agent can adapt the TMDL strategy to exploit the newly acquired information and perform more efficiently in achieving water quality goals than a strategy based on pre-existing knowledge. In addition, given the spatially-diverse activity within a watershed, the agent who designs and implements the strategy is often faced with a choice of how spatially specific a strategy to undertake. At one extreme the agent may specify alternative actions for each infinitely small area within the watershed, thereby exploiting the heterogeneity of the landscape, but requiring considerable information to implement. Conversely, the agent may design a strategy specifying a single set of actions for all areas within the watershed, ignoring the heterogeneity of the landscape, thus requiring little or no new information. With information acquisition the agent learns about the environmental processes and the heterogeneous landscape, and therefore may adapt the optimal strategy to exploit the estimated relationship between water quality and the spatially-diverse, environmental and economic activities.

Understanding how an environmental agency responds to uncertainty when making decisions to improve or protect environmental quality provides relevant analysis that will contribute to a growing literature on environmental agency decision-making under uncertainty. Kaplan, Howitt and Farzin (2003) were the first to empirically analyze information acquisition and adaptive management in an environmental agency problem. Other studies (Baerenklau; Horan, Shortle and Abler; Johansson) evaluating the design of environmental programs have not explicitly considered the role of information acquisition and learning that may allow for more effective targeting of agency resources and thus greater environmental protection or improvement. This paper contributes by adapting a state of the art statistical methodology for estimating water quality processes, agency uncertainty and information acquisition to a unique data set consisting of environmental and economic relationships within the Navarro River watershed that will serve as a basis for evaluating optimal policy design when information is incomplete and costly.

Estimating pollution relationships in complex environmental systems invariably faces obstacles due to limited data and ad hoc or arbitrary distributional assumptions about the model parameters. The sequential entropy filter (SEF) first presented by Kaplan and Howitt (2002) estimates distributions for the parameters of a watershed model when the data comes from a small sample with minimal assumptions about the distributional structure. To alleviate the small sample estimation problem and the imposition of arbitrary distributional assumptions about the estimated coefficients and errors we employ an adaptation of the SEF to estimate in-stream water temperature dynamics. The SEF also captures changes in model parameters over time and space, thereby facilitating the empirical analysis into whether the environmental and economic costs of less-informed strategies can be reduced through information acquisition and adaptive management.

This paper explores this information acquisition and adaptive management problem with an empirical application drawn from data collected from the Navarro River watershed, located in Mendocino County, California. The current TMDL for the Navarro River watershed limits in-stream water temperature and sediment loading, both of which impair critical salmon habitat needed to maintain its population (EPA 2000). The goal of the analysis is to advance our understanding of optimal TMDL design by expanding the empirical tools needed to analyze environmental adaptive management problems in general, and the case of in-stream water temperature for the Navarro River watershed, in particular.

The paper proceeds as follows. Section II provides some background on the Navarro River watershed. The following section describes the empirical methodology. Section IV details the empirical application where we estimate the in-stream water temperature daily cycle within the Navarro River watershed as a function of temporal and spatial environmental and economic activities. Section V presents the results from policy simulation. Section VI concludes.

**II. Background**

The Navarro River watershed is located in southern Mendocino County of California, USA. It is unique in that it is a moderately sized watershed (~800km$^2$) that is both hydrologically contained (i.e., it flows directly into the Pacific Ocean) and heterogeneous in its land uses, which include timber production, animal grazing, and viticulture operations. Although the current human population is only 3500, Euro-Americans have inhabited the watershed for 150 years. Recent changes in land use such as expanding residential development and increasing viticulture is affecting aquatic resources through degrading water quality. This is most notable in the application of the federal Clean Water Act (Section 303(d)) to the degraded beneficial use of cold-water fisheries by elevated stream temperatures and excess stream sediment.

Riparian forests of the Navarro River watershed include redwood (*Sequoia sempervirens*), Douglas fir (*Pseudotsuga menziesii*), and intermingled hardwoods (*Acer spp.*, *Lithocarpus sp.*, *Quercus spp.*, etc.). Riparian vegetation in this watershed is a heterogeneous mix of both upland forests and true "riparian" forests. These riparian forests consist largely of willow (*Salix spp.*) and alder (*Alnus spp.*), as well as diagnostic understory species such as wild grape (*Vitis californica*) and berry (*Rubus spp.*). Although annual grasslands are found throughout the watershed, their proximity to streams and rivers is limited. Ecologically, it is the montane forests that drive terrestrial ecosystem productivity and provide allochthonous material, large woody debris, and shading to aquatic ecosystems. Thus intact riparian forests provide significant microhabitat benefits to freshwater ecosystems (Malanson 1995, Naiman et al. 2000) and it is their diminishing cover within the watershed that may compromise water quality. Previous research in this watershed has shown that typical “riparian” species, such as willows (*Salix spp.*), contribute comparatively little to stream shading; although some localized shading does occur on smaller pools and riffles (Viers 2003). In the Navarro River watershed, it is primarily the canopy structure of large upland tree species (*Sequoia sp.* and *Pseudotsuga sp.*) within the riparian zone that influences the overall shading conditions of stream segments consisting of a series of pools, riffles, and runs (Viers et al. 2004). This has led to the development of various policies regarding the removal of large trees within a riparian buffer and is an active area of economic analysis.

## III. Empirical Methodology

Modeling temperature control policies presents several interesting challenges when compared to many environmental policy models. First, the temperature in the river changes constantly by time of day, however, the fish only incur environmental damages when the temperature exceeds certain thresholds for given lengths of time. Thus the damage is based on the definite integral of a daily temperature cycle once it exceeds a certain level. The critical levels for salmon are at $18^{o}$c due to heat stress protein production, and above $24^{o}$c salmon experience acute respiration problems.

This model is based on a single river, one watershed, and three reaches in the watershed. The basic temperature data is collected at 6-minute intervals, presenting a massive (50,000) observation data set for the 147 summer and fall days considered. Given that a critical exposure time for salmon is 2 hours above $24^{o}$c, we aggregated the 6-minute observations to 7,000, 30-

minute observations for purposes of estimation and policy simulation. One advantage of temperature data is that the period of the daily cycle is fixed at 48 half hour intervals. However both the mean temperature and the amplitude of the daily fluctuations changes between the three reaches of the river and with the daily progression in any given reach. The daily temperature cycles are not stationary, and the estimates need to be able to systematically change as a function of time and exogenous variables. We use an adaptation of the SEF, a sequential cross entropy Bayes estimator, to explain the in-stream water temperature daily cycles as a function of daily stream flow in each reach, daily air temperature, and an annual reach specific shade index. The temperature control policy variables are flow and shade cross section variables that can only be changed slowly but influence the mean and amplitude of the rapid half hour temperature variables. It is this combination of fast time series dependent variables and slow cross section policy variables that create challenges for a conventional estimation approach.

The fundamental equation that explains the half hour temperature in each reach of the river is:

$$
\begin{aligned}
(1)\quad Temp_{tr} = {} & pi + dpimr + dpiur + sipi * s_r + (1 - sishr * s_r - pdur * dur_r - pdmr * dmr_r) \\
& * \left\{ \begin{array}{l} \left[ \sum_j par1_{jt} * val_j \right] * \sin\left(phival_t\right) + \left[ \sum_j par2_{jt} * val_j \right] * \cos\left(phival_t\right) \\ + \left[ \sum_j par3_{jt} * val_j \right] * \sin\left(phival_t\right)^2 + \left[ \sum_j par4_{jt} * val_j \right] * \cos\left(phival_t\right)^2 \end{array} \right\} \\
& + err_{tr}
\end{aligned}
$$

Equation (1) shows that the daily temperature cycle is defined by the combination of linear and quadratic sine and cosine functions. The dependent variable $Temp_{tr}$ is the in-stream water temperature for half-hour $t$ for reach r for a given day. We suppress the day index since equation (1) is sequentially updated with each day. The daily mean water temperature for the lower reach is denoted as pi. The dummy variables *dpimr* and *dpiur* shift the mean temperature to captures differences between the lower reach and the middle and upper reaches respectively. The shade index ($s_r$) for a given reach also shifts the mean temperature through the *sipi* coefficient. The amplitude of the cycle in each reach is also influenced by the annual shade index in that reach ($s_r$) through the amplitude coefficient parameter *sishr*. We also include two reach specific dummy variables *pdur* and *pdmr* to capture differences in amplitudes across reaches. The daily temperature cycle is a function of $phival_t$, which is the time of day expressed in radians and evaluated at the mid point of each half hour interval.

The time-varying coefficients $par1_{jt} \ldots par4_{jt}$ define the amplitude of the daily cycle as functions of the explanatory variables $val_j$ where $j$ denotes maximum, minimum and mean daily air temperature and daily in-stream flow. The dynamic nature of the estimator is defined by an equation of motion for the intercept coefficient (*pi*) and each of the sine and cosine coefficients, $par1_{jt} \ldots par4_{jt}$.

Each of these coefficients is estimated with SEF, a generalized maximum entropy (GME) (Golan et al 1996) estimator, in which the objective function minimizes the probability distance, often termed the cross entropy, between the current probability distribution for each parameter and the prior probabilities. The sequential estimation procedure uses to derive the coefficient distributions in each half-hour are the prior distributions for estimating the coefficients in the following estimation period. Kaplan and Howitt (2002) show that the resulting estimator is an optimal Bayes estimate.

Given two sets of support values $zvals_{jp}$ and $zvals4_{jp}$ spanning the discrete distribution for the various estimated coefficients, the $j$th equation of motion for the first linear sine function coefficient $par1_{jt}$ is written as:

$$(2) \quad \sum_p pshr1_{jp} * zvals_{jp} = \left( \sum_p qp1_{jp} * zvals_{jp} \right) * \left( \sum_p pday1_{jp} * zvals4_{jp} \right) + \left( \sum_p perrst1_{jp} * zvals4_{jp} \right)$$

Where $qp1_{jp}$ are the fixed prior probabilities for the parameter in question, the equation of motion coefficient is $\sum_p pday1_{jp} * zvals4_{jp}$ and the error term on the equation of motion is $\sum_p perrst1_{jp} * zvals4_{jp}$. Note that the equation of motion defines the evolution of the time varying parameters in equation (1) since:

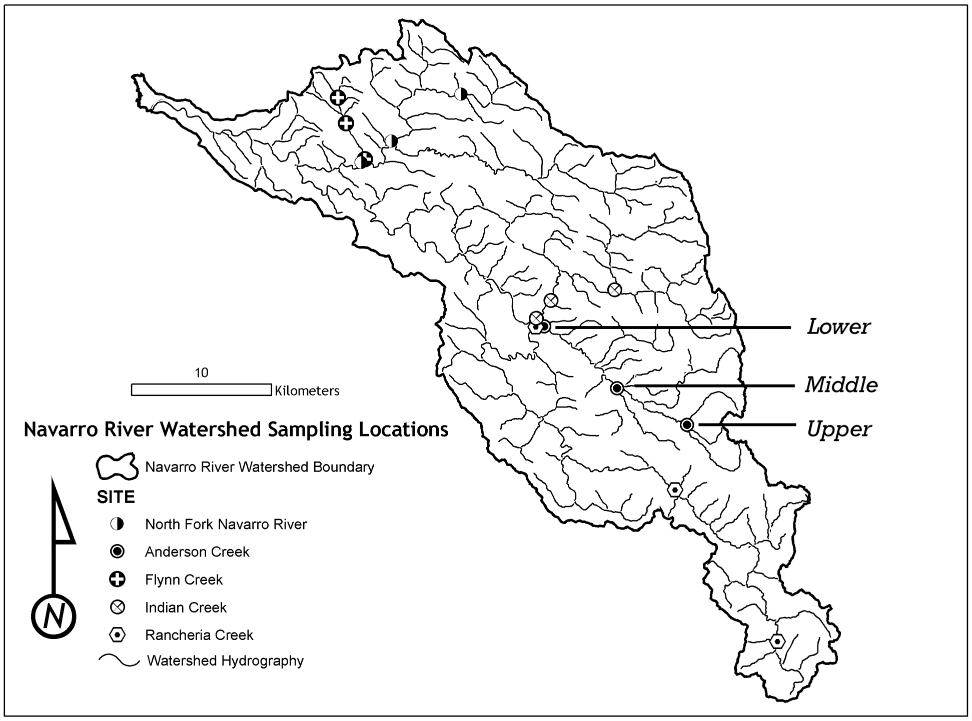
$$(3) \qquad par1_{jt} = \sum_p pshr1_{jp} * zvals_{jp}$$

## III. Empirical Application

The empirical model developed in this analysis focuses on the relationship between in-stream water temperature and the environmental and economic activities within one sub-watershed of the Navarro River watershed. In the analysis, daily water temperature cycles in three reaches of Anderson Creek (lower, middle, and upper), located with the Navarro River watershed (Figure 1), are sequentially estimated for the period spanning June through October when in-stream water temperatures approach or exceed 18$^{o}$c may exceed 24$^{o}$c. Among the many environmental and economic factors contributing to the water temperature, the model captures previously identified factors such as air temperature, in-stream flow, and riparian shade. Data from 2000 on water temperature and these identified factors are used to estimate the mean and amplitude of the in-stream water temperature daily cycle.

The data provides measures of maximum, minimum and average daily air temperature measured in Ukiah, CA, daily average in-stream flow, and a seasonal stream shade measure derived from a spatially explicit Geographic Information System model developed by Viers, Quinn, and Johnson (2004) to account for riparian and topographic induced shade. As mentioned, the estimated amplitude and intercept for each day then serve as prior distributions for estimating subsequent amplitude and intercept of the daily temperature cycle.

Figure 1. Map of Anderson Creek sampling locations in the Navarro watershed.




The results from the empirical estimation are then used to evaluate alternative TMDL strategies. First we consider the affect on in-stream water temperature if the riparian shade is increased by 50 percent. Second we impose a flow restriction on withdrawals, presumably withdrawals taken by vineyards and rural residential dwellings within the Anderson Creek catchment area. Lastly we implement an integrated strategy that employs both a shade and flow strategy for reducing the duration of time at which the daily in-stream water temperature cycle exceeds 18$^{o}$c and 24$^{o}$c.

The sequential learning of the model over the first ten days of the estimation database is shown by the fit to the nonstationary dynamic temperature cycle. Figures 2-4 show the in sample fit of the model for all three reaches, and the differences of the temperature process over time and space.

Figure 2. Lower Reach Daily Temperature Cycle and Estimated Cycle (First Ten Days)

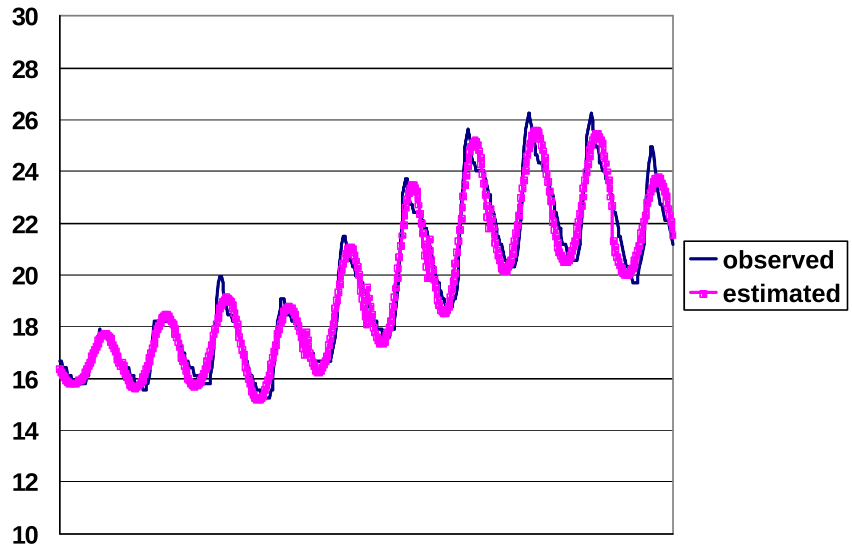



Figure 3. Middle Reach Daily Temperature Cycle and Estimated Cycle(First Ten Days)

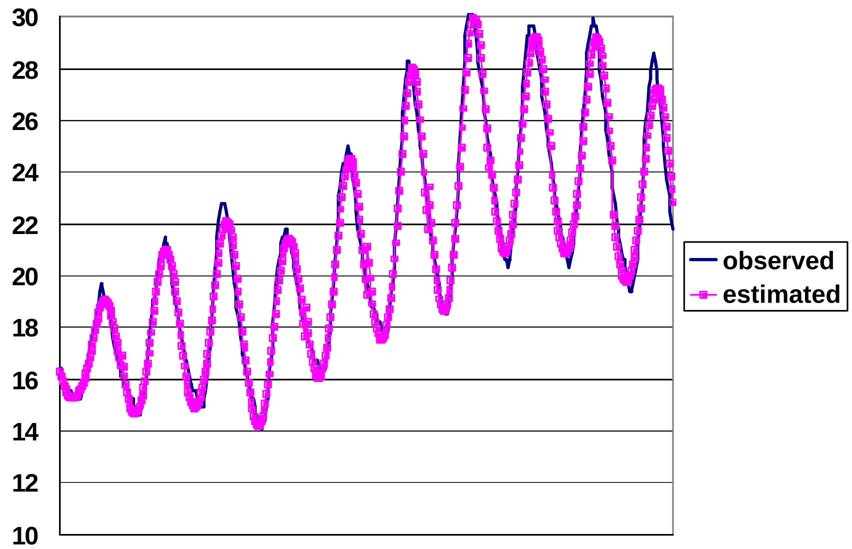
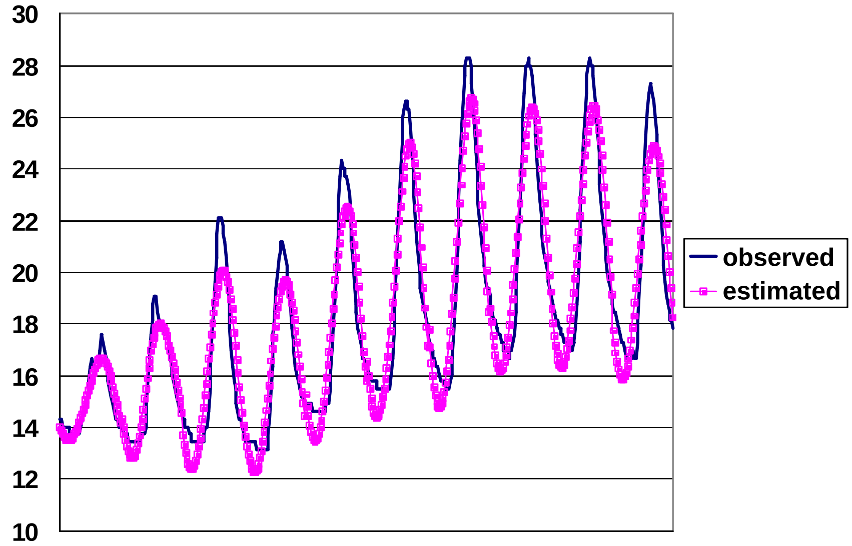


Figure 4. Upper Reach Daily Temperature Cycle (First Ten Days)




Figures 2 – 4 show that the in-sample model fit for a ten day estimation sample is good, despite the changes in amplitude and mean, and the differences between the river reaches. The mean squared errors for the in-sample estimates are lower reach 0.447, middle reach 0.407, upper reach 1.508.

Out of sample forecasts are obtained by fixing the coefficient values at their tenth day value, and suppressing the equations of motion predictions for the ten days after the estimation sample. Essentially this forecast adopts an open loop approach to the problem, with the only change in variables coming from the j daily variables that measure the maximum, minimum, and mean daily air temperature and the daily flow in the river. A closed loop forecasting approach will be assessed later using the information implicit in daily shifts over the season. Despite the open loop specification, the out of sample predictions shown in figures 5 - 7 are quite good, with the mean and amplitude of the temperature cycle responding to changed air temperature and flow levels. The mean squared errors for the out-of-sample estimates are: lower reach 1.668, middle reach 1.674, upper reach 2.739.[1]

Figure 5. Lower Reach Daily Temperature Cycle and Out of Sample Predictions

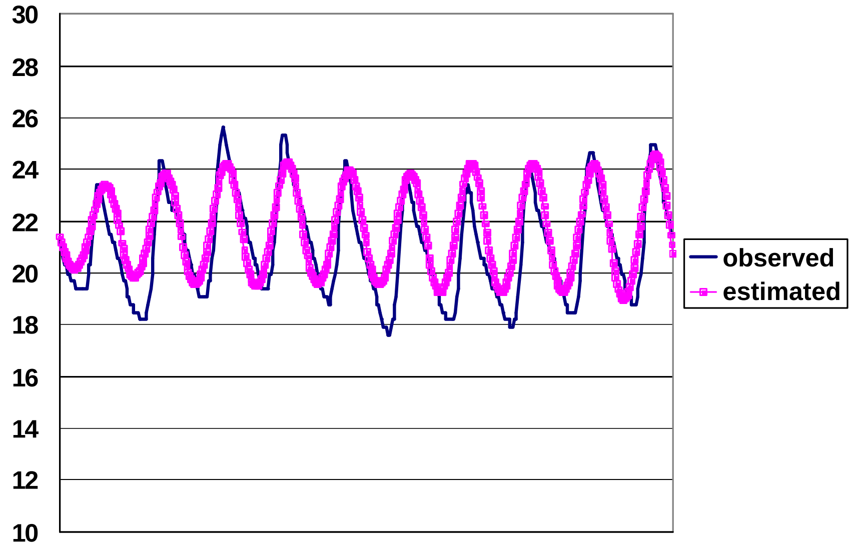
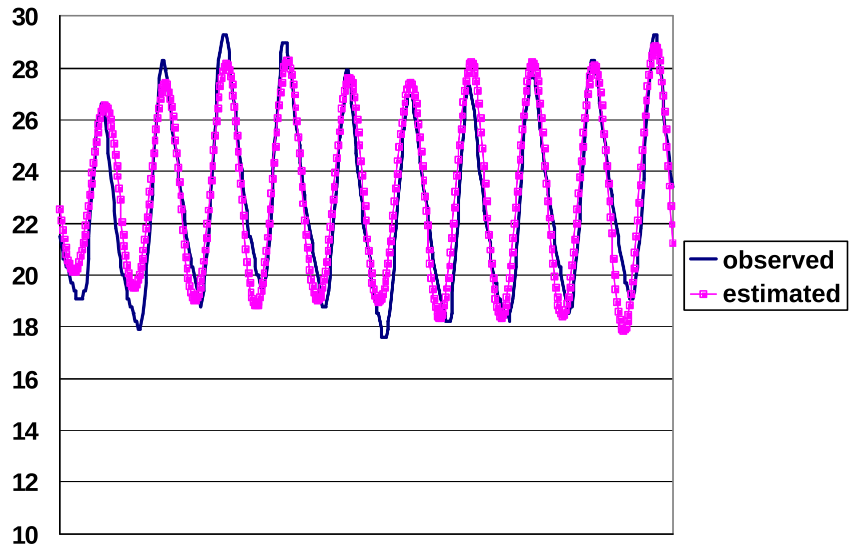
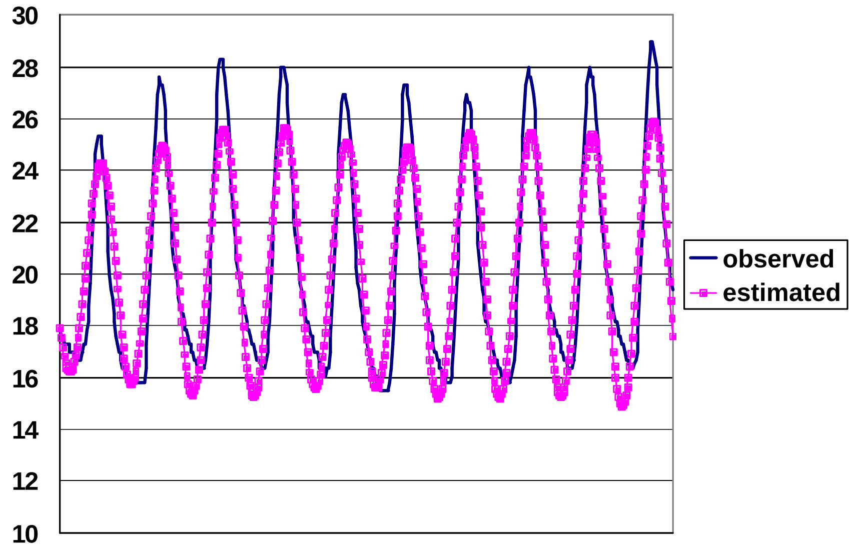



Figure 6. Middle Reach Daily Temperature Cycle and Out of Sample Predictions




[1] We also ran a 50-day in-sample sequential estimation of daily temperature cycles with a ten-day out-of-sample open-loop prediction. The corresponding in-sample mean squared errors are 0.505, 0.420, and 1.873 for the lower middle and upper reaches, respectively. The out-of-sample mean squared errors are 1.288, 1.457, and 2.422 for the lower middle and upper reaches, respectively.

Figure 7. Upper Reach Daily Temperature Cycle and Out of Sample Predictions




## IV. Policy Results

The effect of changes in the two policy variables on the daily temperature cycle is examined by simulating the out of sample forecasts with changes in the variables. The two policy variables are: first, a 50% increase in riparian shade index, and second a 50% increase in flows in the river due to restrictions on groundwater pumping in the watershed. The payoff to the policy variables comes from reduction of the time that fish face temperatures above the stress level of 18$^o$c or the critical exposure level of 24$^o$c. Note that given the structural model used to estimate in-stream water temperature, the policy goal of reducing in-stream water temperature can be achieved by either a reduction in the mean of the daily cycle or a reduction in its amplitude, or a combination of both effects.

The shade increasing runs shown in figure 8 show that a 50% increase in shade decreases temperature by an average of approximately 2$^o$c. This reduction in the mean temperature results in the critical 24$^o$c threshold being avoided in the days simulated however there is still a six-hour period in which the stress temperature is exceeded.

A 50% increase in the daily flow level causes a slight shift in the phase and a noticeable reduction in the amplitude of the temperature cycle. Note however, that unlike the shade policy, the mean does not shift. The shift in amplitude is sufficient to significantly reduce the hours of critical exposure above 24$^{o}$c, but not eliminate exposure from the simulated sample.

Figure 8. Out Of Sample Predicted Water Temperature Before and After Shade Policy Implementation

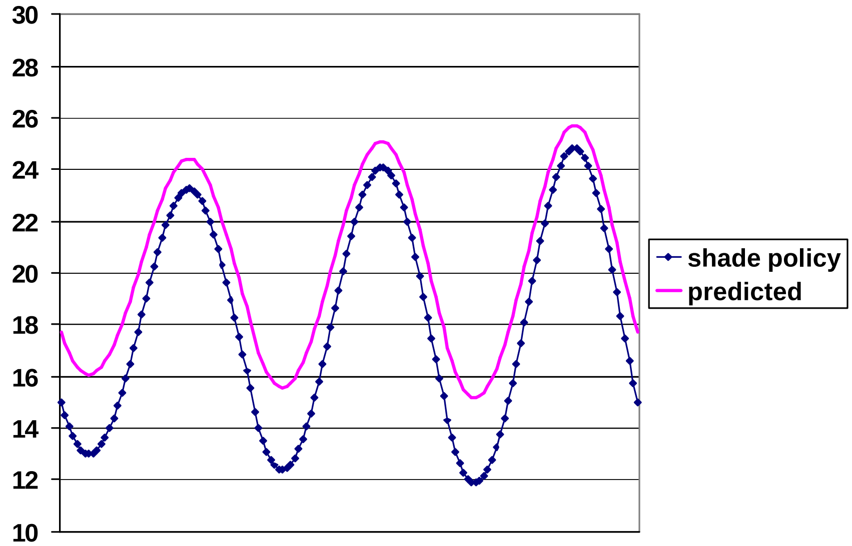



Figure 9. Out Of Sample Predicted Water Temperature Before and After Flow Policy Implementation

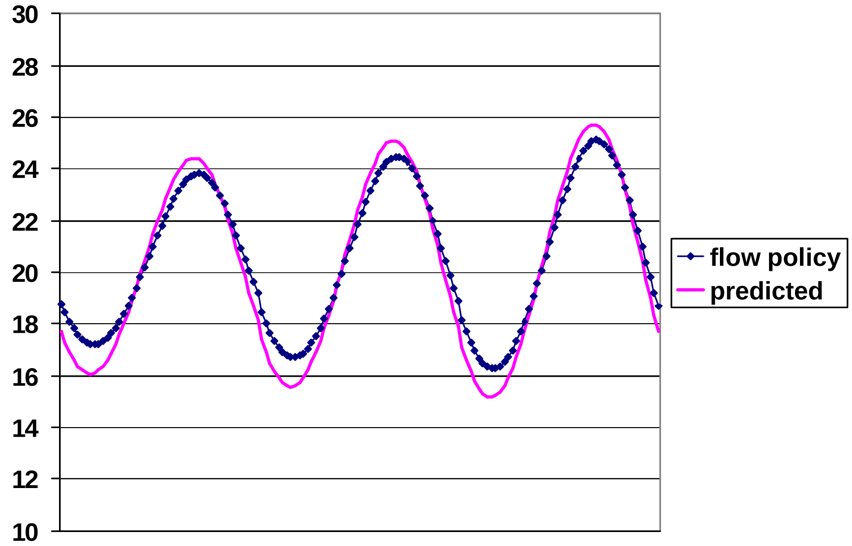



Figure 10. Out Of Sample Predicted Water Temperature Before and After Combined Shade and Flow Policy Implementation

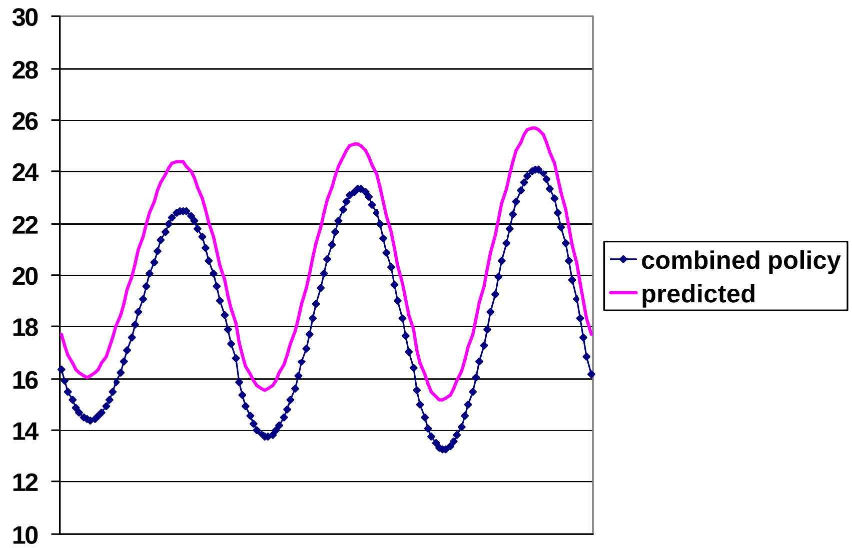



Figure 10 shows a combined increased shade and flow policy that decreases the mean of the temperature cycle and reduces its amplitude. Under the combined policy scenario the maximum temperature never exceeds 24$^o$c on the three days examined, and time when the temperature exceeds18$^o$c is considerably reduced.

## V. Conclusion

The non stationary SEF model developed in this paper shows that fast moving time series variables can be combined with slowly changing cross section variables to produce a model that fits very well in sample, and produces reliable out of sample estimates. Preliminary results show that both the mean and amplitude of the temperature cycles respond to changes in the shade and flow parameters to the extent that a 50% change in both parameters modifies the cycle to avoid the critical 24$^{o}$c threshold, and reduce exposure to temperatures above 18$^{o}$c. Preliminary tests using a fifty-day, 2400 observation estimation period, and a similar out of sample period for forecasting show that the mean squared error measures for the estimation increased and for the forecasts decreased slightly. Further modeling work on extending the parame ter equations of motion to reflect daily time trends during the year will enable out of sample forecasts to use this information in an open loop estimate of future responses.

## VII. References


Baerenklau, K.A. "Green Payment Programs for Nonpoint Source Pollution Control: How Important Is Targeting for Cost-Effectiveness?" *Journal of Agricultural and Resource Economics* 27 (2002): 406-19

Golan.A, G. Judge & D.Miller 1996, *Maximum Entropy Econometrics* John Wiley, Chichester, & New York.

Horan, R.D., Shortle, J.S., and D.G. Abler. "Ambient Taxes Under m-Dimensional Choice Sets, Heterogeneous Expectations, and Risk-Aversion." *Environmental and Resource Economics* 21 (2002): 189-202.

Johansson, R.C. "Watershed Nutrient Trading under Asymmetric Information." *Agricultural and Resource Economics Review* 31(2002): 221-232.

Kaplan, J.D., Howitt, R.E., and Y.H. Farzin, "An Information-Theoretical Approach to Budget Constrained Nonpoint Source Pollution Control." *Journal of Environmental Economics and Management* **46** (2003): 106-130.

Kaplan, J.D., and R.E. Howitt. "Estimating Nonpoint Source Pollution: An Application of a Sequential Entropy Filter." *Water Resources Research* 38(3) (2002): 10.1029/2000WR000088.

Malanson, G. P. 1995. Riparian landscapes. New York, Cambridge.

Naiman, R. J., R. E. Bilby, and P. A. Bisson. 2000. Riparian ecology and management in the Pacific Coastal Rain Forest. Bioscience 50:996-1011.

United States Environmental Protection Agency. "Overview of Current Total Maximum Daily Load - TMDL - Program and Regulations," Online Report, available online at http://www.epa.gov/owow/tmdl/overviewfs.html. 2003.

____. *Navarro River TMDLs for Temperature and Sediment (December 2000 Final)*, online at http://www.epa.gov/owow/tmdl/examples/sediment/ca_navarro.pdf. 2000.

Viers, J.H. 2003. Remote Methodologies of Watershed Assessment: Applications toward Conservation and Restoration in the Navarro River Watershed. Doctoral Dissertation. University of California, Davis, Davis.

Viers, J.H., J. F. Quinn, and M. L. Johnson. 2004. RipTopo: a riparian-topographic stream shading model applied to the Navarro River Watershed, California, USA. Environmental Management Submitted.