Transactions of the American Fisheries Society 125:622–627, 1996

NOTES

An Empirical Method for the Prediction of Daily Water Temperatures in the Littoral Zone of Temperate Lakes[1]

JOHN E. MATUSZEK AND BRIAN J. SHUTER

Aquatic Ecosystems Research Section, Ontario Ministry of Natural Resources Box 5000, Maple, Ontario L6A 1S9, Canada

Abstract.—We used daily air and water temperatures from 14 lakes in Ontario to develop and test a simple method for constructing lake-specific empirical models for predicting daily littoral water temperatures. Data requirements for prediction are modest (average air temperature, day of the year) and the method is robust and practical, requiring only a few (15–20) well-spaced water temperature observations to construct a single-year model that can generate reasonably accurate predictions for an entire ice-free season. The ability of a singe-lake model to predict several years of temperature data is significantly improved by explicitly including information on ice-out date in the model. Our multiyear model for Lake Opeongo described most (18 of 22) years well. Years that were not well described were usually coincident with El Niño Southern Oscillation events. Used with caution, the method can be an effective tool for supplementing direct monitoring of littoral water temperatures and for generating historical water temperature estimates when direct estimates are lacking. These capabilities should be of particular use to fisheries biologists studying or managing populations of fish species with critical life stages that are affected by littoral water temperatures.

Water temperatures in the littoral zone of temperate lakes strongly influence the reproductive timing, growth, and abundance of many resident fish species (Shuter et al. 1980; Hill and Magnuson 1990; Magnuson et al. 1990; Ridgway et al. 1991; Ridgway and Friessen 1992). Many of these effects involve nonlinear or threshold responses and cannot be accurately predicted from relatively long-term (e.g., monthly) mean temperatures. Therefore, automatic logging of temperature hourly (or daily) is standard practice in current field research programs. However, there are many situations where direct measurements of these data must be supplemented with indirect estimates. For example, data loggers may fail, producing gaps in time series; there may be insufficient resources to equip all study sites with data loggers; and studies of fish population time series may require reconstruction of historical water temperature data.

Attempts at predicting daily water temperatures have been made with complex mechanistic models (Edinger et al. 1968; Robertson and Ragotzkie 1990; Honzo and Stefan 1993) and simpler empirical models (Jacobsen and Bachmann 1974). Air temperature has been highlighted as a significant predictor of water temperature by both mechanistic (Henderson-Sellers 1988; Honzo and Stefan 1992) and empirical modelers (Rawson 1958; McCombie 1959; Jacobsen and Bachmann 1974; Robertson and Ragotzkie 1990).

In this paper, we present a method for constructing lake-specific empirical models that predict daily littoral water temperatures. We substantially extend previous empirical studies (i.e., Jacobsen and Bachmann 1974) by identifying a new model structure and by demonstrating, through extensive interlake and interyear testing, that models based on this structure are more accurate and require less data to develop than those used previously.

Methods

Daily water temperatures were measured during the ice-free season for 14 widely distributed Ontario lakes (44°47″–50°03″N, 76°45″–92°15″W). Lake areas ranged from 36 to 8,270 ha (median, 339 ha), and mean depths ranged from 2.9 to 24.7 m (median, 9.3 m). Most lakes had water temperature data for only one year, 1990 or 1991. Lake Opeongo (5,820 ha; mean depth, 14.8 m) had data for 22 years (1964, 1967–1987).

Daily mean water temperatures were calculated as the mean of 24 hourly measurements made by recording thermographs anchored near shore in 1–2 m of water, with recording sensors placed at a depth of 1 m. Since all our study lakes were dimictic, the measured littoral water temperatures represent epilimnetic temperatures during summer stratification. Ice cover was usually present from early December to late April or early May. Daily

[1] Contribution 95-08 of the Aquatic Ecosystem Research Section, Ontario Ministry of Natural Resources.

air temperatures were calculated as the mean of daily minimum and maximum temperatures measured at nine Environment Canada weather stations in Ontario.

Jacobsen and Bachman (1974) showed that air temperature averaged over a period leading up to day *t* is a useful predictor of surface water temperature on day *t*. In a limited set of interlake (4 lakes) and interyear (16 years) comparisons, they found that the length of the "best" averaging period varied from 10 to 26 d. We carried out a more extensive comparison (14 lakes, 22 years) to determine if there was a length for the averaging period that would work well in most situations. We calculated Pearson correlations between daily water temperatures and mean air temperatures for 5, 10, 15, 20, 25, and 30-d periods, each period extending back in time from the day the water temperature was measured. The air temperature data came from the weather station nearest each lake. Only air temperatures after ice-out were used. We also examined the deterioration of these statistical associations with increases in the distance between air and water temperature recording sites.

For each lake, we used multiple linear regression analysis to construct year-specific predictive equations for water temperature. Independent variables used in these analyses included mean air temperature, day of the year (YDAY) and its transformations (square, cube, inverse, logarithm), and air temperature–YDAY interaction terms. At Lake Opeongo, ice-out varied from YDAY 102 (April 12) to YDAY 135 (May 15). Because air temperatures after ice-out were used, the prediction interval for each model began some time after ice-out. We measured the accuracy and bias of each regression equation by using the square root of the mean squared error (RMSE) and the annual mean residual, respectively. Analyses based on all available water temperature data were compared with analyses based on selected subsets of the available water temperature data to determine the minimum amount of data required to build useable year-specific predictive models for each lake.

For Lake Opeongo, we pooled all 22 years of data to build a single lake-specific multiyear regression model for water temperature from the following set of independent variables: mean air temperature; YDAY adjusted for ice-out (and its transformations: square, cube, inverse, logarithm); YDAY–ice-out interaction terms; and air temperature–YDAY interaction terms. Interannual variation in the predictive capabilities of this model was compared to interannual variation in the following: ice-out date, an index of summer warmth (i.e., maximum summer temperature, as estimated from a fifth order polynomial fit to the daily air temperatures of an ice-free season), and an index of the duration of thermal stratification (i.e., length of the period when the difference between epilimnetic and hypolimnetic water temperatures is large and relatively stable, as estimated from weekly thermal stratification data obtained with bathythermographs).

Results

Single-Year Models

Over a typical ice-free season, water temperatures in our lakes did not relate to air temperatures in a constant manner. Water temperatures followed mean air temperatures most closely in the spring and remained well above mean air temperatures during the summer and fall. This varying relationship prompted us to include a function of time in our predictive equations.

The model structure which generally provided the best annual fits to the daily water temperatures observed at each of our 14 lakes had the form:

$$\text{WTEMP} = C_0 + C_1(\text{ATEMP}) + C_2(\text{YDAY}) + C_3(\text{YDAY})^2;$$

WTEMP = daily water temperature, ATEMP = mean air temperature, YDAY = day of the year, and C_0, C_1, C_2, and C_3 are regression coefficients. Annual RMSEs ranged from 0.69°C to 1.65°C (median, 0.82°C), with all lakes except the shallowest (mean depth, 2.9 m; area, 133 ha) at values of 1.09°C or less. The optimal air temperature averaging period for each lake ranged from 5 to 20 d, however, a 20-d period was near-optimal for all lakes; that is, the 20-d correlations, which ranged from 0.76 to 0.98, were each within 5% of the maximum correlation obtained for each lake. For a given lake, the correlation between daily water temperatures and 20-d mean air temperatures observed at the recording site nearest the lake was only marginally higher than correlations obtained with 20-d mean air temperatures recorded at relatively distant sites. For distances up to 100 km, 100% of all correlations (N = 27) were within 10% of their respective lake-specific maximum values. From 100 to 200 km, 92% of all correlations (N = 20) were within 10% of their respective maxima, and from 200 to 300 km, 90% of all correlations (N = 24) were within 10% of their respective maxima.

Using our basic model structure, we constructed annual models for each of the 22 years of data from Lake Opeongo and obtained RMSE values ranging from 0.76°C to 1.24°C (median, 0.98°C). We used a 10-d averaging period for these models because 10 and 20-d averaging periods were equally effective as predictors in Lake Opeongo and the 10-d averaging period permitted the prediction interval for each model to begin 10 d earlier in spring.

Useable annual models could be constructed from small subsets of the available data by carefully selecting specific days over a season. For example, models based on daily mean water temperatures selected at intervals of 1, 5, 10, and 15 d (N = 177, 35, 18, 11) from the 177 daily values available for Lake Opeongo in 1977 had similar RMSE values (0.90°C, 0.90°C, 0.95°C, and 1.03°C, respectively). Similar results were obtained for all 22 years of data in the Opeongo data set; the median of the RMSE values for models based on all the observations available in each year was only slightly smaller than the median RMSE for models based on observations taken at 10-d intervals (0.98°C versus 1.01°C).

Further data reduction, with little loss of model accuracy, could be achieved by replacing daily mean water temperatures with carefully chosen spot temperatures. At Lake Opeongo, spot temperatures, taken anytime between 1200 and 1500 hours, served as accurate and unbiased estimators of daily mean water temperatures calculated from continuous recording thermographs. Typical regressions of daily mean water temperatures on spot water temperatures yielded slopes of 1, intercepts of 0, R^2 values >0.98, and RMSE values <0.25°C.

Multiyear Models

We pooled the 22 years of data from Lake Opeongo into one large data set and generated the following model:

$$\begin{aligned}\text{WTEMP} = & -32.36 + 0.5069(\text{ATEMP}) \\ & + 0.4090(\text{YDAY}) \\ & - 0.0009455(\text{YDAY})^2,\end{aligned}$$

with ATEMP = the 10-d air temperature mean (R^2 = 0.92, RMSE = 1.33°C, N = 3,761). The addition of the YDAY terms increased the R^2 from 0.83 to 0.92 and decreased the RMSE from 1.89°C to 1.32°C.

The fit of this equation to the data from individual years (Figure 1) illustrates the range in accuracy that might be expected if the equation were used to predict daily water temperatures for years without data. Annual RMSEs from the 22-year equation (range 0.87–2.10°C; median, 1.17°C) averaged 19% higher than comparable RMSEs from the 22 single-year equations (range, 0.76–1.24°C; median, 0.98°C). This result is distorted somewhat by the presence of 4 years (1967, 1972, 1982, 1987) that were poorly predicted by the 22-year equation, having both high RMSE values (1.6–2.1) and high bias measures (range for absolute value of mean residual, 0.96–1.74). With these years excluded, the average increase in RMSE is only 12%.

These results show that interannual variation in the regression coefficients of our Lake Opeongo models is relatively small over most years. Thus, the data from 3 or 4 "typical" years is quite sufficient to generate a relatively accurate, multiyear model. For example, the 3-year equation derived from 1979–1981 data equals the 22-year equation in its ability to predict the entire 22-year data set (i.e., 3-year equation: range and median RMSE 0.86–2.15°C, 1.17°C; 22-year equation: range and median RMSE, 0.87–2.10°C, 1.17°C). However, these results also show that there are "atypical" years, with quite different regression coefficients, that cannot be identified from information on air temperature and date alone.

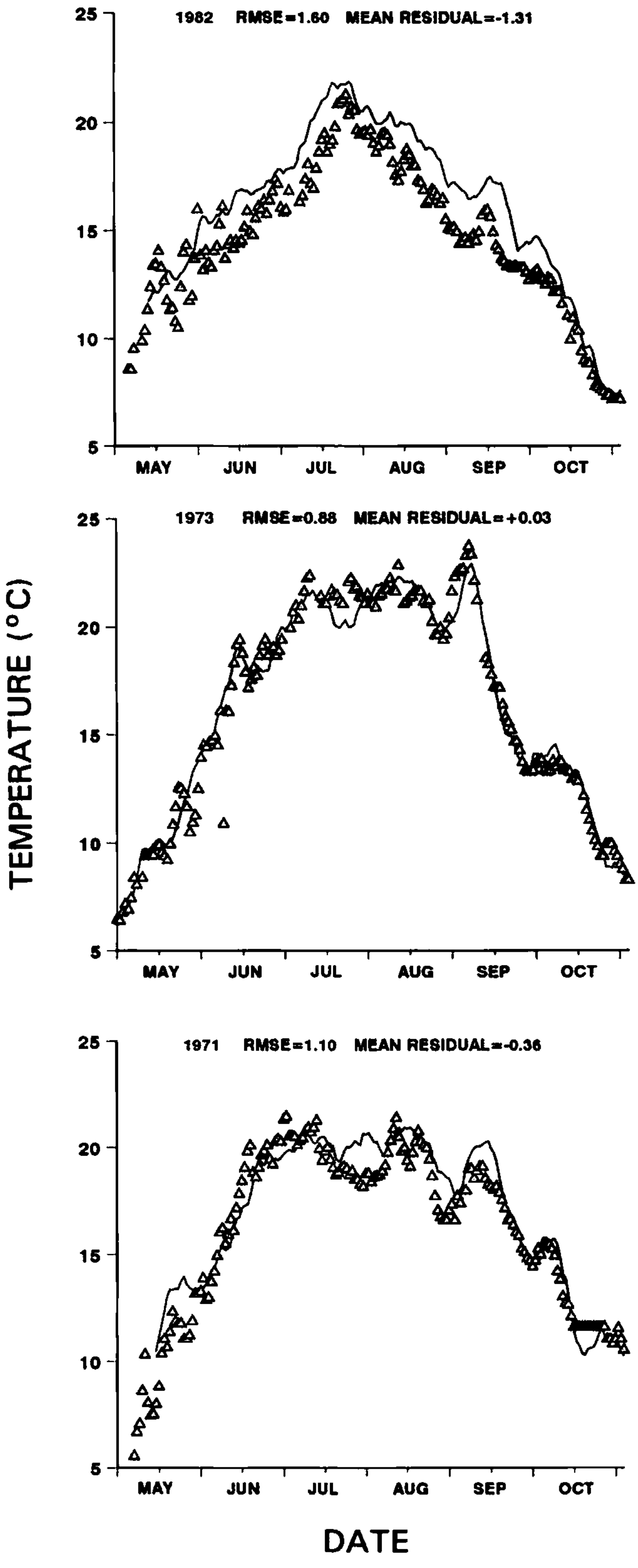
Enhanced Multiyear Models

Our annual indices of stratification duration (range, 7–78 d) and summer warmth (range, 16.5–19.7°C) were not useful in identifying those years that were predicted less accurately by the 22-year basic equation. However, there was a statistically significant (P < 0.01) correlation between prediction bias and ice-out time, with overpredicted years exhibiting late ice out-times and underpredicted years exhibiting early ice-out times.

We reanalyzed the pooled 22-year data set to include the effect of ice-out time. Assuming that

→

FIGURE 1.—Predicted (line) and observed (triangles) daily water temperatures (°C) for Lake Opeongo for selected years. Predictions were made with a 22-year regression equation that used 10-d mean air temperature, day of the year, and day of the year squared. The year 1973 represents a year that was predicted well, 1982 a year that was predicted poorly, and 1971 a year in the middle of the range. Accuracy and biases of each prediction were assessed based on the square root of the mean square error (RMSE) and the annual mean residual, respectively.


1982 RMSE=1.60 MEAN RESIDUAL=-1.31
1973 RMSE=0.88 MEAN RESIDUAL=+0.03
1971 RMSE=1.10 MEAN RESIDUAL=-0.36
TEMPERATURE (°C)
25
20
15
10
5
MAY
JUN
JUL
AUG
SEP
OCT
DATE

the influence of ice-out on littoral water temperatures would be greatest in spring, we introduced a variable that diminished in value as the time after ice-out increased. We also introduced short-term (5-d) and longer term (15- and 20-d) air temperature averages to try to account for water temperature variations caused by very recent air temperatures. Stepwise regression analysis yielded the following model structure (all terms significant at $P < 0.01$):

$$\text{WTEMP} = C_0 + C_1\,(\text{ATEMP1}) + C_2(\text{YDAY}) + C_3(\text{YDAY})^2 + C_4(\text{INVADYD}) + C_5(\text{ATEMP2});$$

ATEMP1 = 15-d mean air temperature, INVADYD = inverse of YDAY adjusted so that ice-out is standardized as YDAY 100, and ATEMP2 = 5-d mean air temperature.

Compared with the basic equation, the enhanced equation, fitted to the 22 years of data, improved the RMSE from 1.32°C to 1.18°C. The 22-year enhanced equation also reduced the RMSE for most years (16 of 22), bringing both its range (0.84–1.88°C) and median value (1.12°C) closer to those obtained by fitting separate equations to each year (range, 0.76–1.24°C; median, 0.98°C).

Discussion

Our results showed that the empirical relationship between littoral water temperature and air temperature is time dependent and that the nature of this time dependence can be simply and effectively expressed in terms of a parabolic seasonal trend, around which daily water temperatures fluctuate in response to changes in mean air temperatures. More complex representations, involving time–temperature interaction terms and higher-order polynomial terms, did not provide better descriptions of this effect.

Our time-dependent empirical models exhibited significantly better predictive capabilities than the time-independent models of Jacobsen and Bachmann (1974). These authors used mean air temperature alone to predict daily water temperatures for four lakes in Iowa, Indiana, and New York. Their 16-year regression for monomictic Clear Lake (area, 1,472 ha; mean depth, 3.7 m), Iowa, produced an RMSE value (1.84°C) considerably larger than the 1.32°C and 1.18°C values obtained from our 22-year regressions for Lake Opeongo. With 2- and 3-year regressions for three other lakes (two monomictic, one dimictic), they obtained RMSE values ranging from 1.17°C to 2.08°C, values again larger than the typical RMSEs achieved by our single-year models for 14 different lakes (all but one ≤1.09°C).

Our results showed that an averaging period of 20 d for air temperature is close to optimal for all the years and lakes we examined, and would be a good averaging period to use for a lake lacking sufficient data for estimation of its optimal period. As noted by Jacobsen and Bachman (1974), the optimal averaging period itself varies from year to year and lake to lake, with no obvious link between it and lake morphometry or climatic conditions. The range in optimal periods found by Jacobsen and Bachmann (10–26 d) was similar to that observed in our lakes (5–20 d).

Useable air temperatures for constructing predictive models could be obtained from recording sites up to 300 km from our lakes. This insensitivity to distance must stem in part from the relatively low variation in physical relief in Ontario. Nevertheless, as long as the air and water temperature recording sites are within the same regional climatic zone, the air temperature recording site need not be right next to the lake in order to supply useful data.

Our results showed that the accuracy of a multiyear model could be improved by including information on year-to-year variation in ice-out date. This finding suggests that it may be simpler to build accurate multiyear models for lakes that do not freeze annually.

Additional year-to-year variation was linked to significant climatic anomalies. Our multiyear analyses of Lake Opeongo identified three extremely atypical years; 1972 and 1982 are the most overpredicted years in the data set, with short stratification periods (13 d, 7 d), low indices of summer warmth (17.5°C, 16.5°C), and late ice-out dates (May 5, May 15); 1987 is the most underpredicted year in the data set, with a long stratification period (56 or more days), high index of summer warmth (19.7°C), and an early ice-out date (April 13). All three of these years are ENSO (El Niño Southern Oscillation) years. A similar association of ENSO years with extreme variation in lake thermal characteristics was noted by Strub et al. (1985) for Castle Lake, California.

Our results demonstrate the robustness of this approach, and therefore the potential utility of the approach in other situations. The models generated produced good fits to daily littoral water temperatures during the ice-free season across a wide range of conditions. The models worked well for small and large lakes, deep and shallow lakes, lakes that experienced the harsh climate of northwestern and northeastern Ontario, and lakes that experienced the more moderate climate of southern Ontario. The models also worked well for most of the year-to-year variations in climate that occurred during two decades at Lake Opeongo.

Acknowledgments

The authors thank C. Sliwinski for his help with data analysis and the reviewers for their comments on the manuscript.

References

Edinger, J. E., D. W. Duttweiler, and J. C. Geyer. 1968. The response of water temperatures to meteorological conditions. Water Resources Research 4:1137–1143.

Henderson-Sellers, B. 1988. Sensitivity of thermal stratification models to changing boundary conditions. Applied Mathematical Modelling 12:31–43.

Hill, D. K., and J. J. Magnuson. 1990. Potential effects of global climate warming on the growth and prey consumption of Great Lakes fish. Transactions of the American Fisheries Society 119:265–275.

Honzo, M., and H. G. Stefan. 1992. Propagation of uncertainty due to variable meteorological forcing in lake temperature models. Water Resources Research 28:2629–2638.

Honzo, M., and H. G. Stefan. 1993. Regional water temperature characteristics of lakes subjected to climate change. Climatic Change 24:187–211.

Jacobsen, T. V., and R. W. Bachmann. 1974. Calculation of water temperatures in Clear Lake, Iowa. Iowa State Journal of Research 48:285–290.

Magnuson, J. J., J. D. Meisner, and D. K. Hill. 1990. Potential changes in the thermal habitat of Great Lakes fish after global climate warming. Transactions of the American Fisheries Society 119:254–264.

McCombie, A. M. 1959. Some relations between air temperature and the surface water temperature of lakes. Limnology and Oceanography 4:252–258.

Rawson, D. S. 1958. Indices to lake productivity and their significance in predicting conditions in reservoirs and lakes with disturbed water levels. Pages 27–42 *in* P. A. Larkin, editor. The investigation of fish-power problems, H.R. MacMillan Lectures in Fisheries, University of British Columbia, Vancouver.

Ridgway, M. S., and T. G. Friessen. 1992. Annual variation in parental care in smallmouth bass *Micropterus dolomieui*. Environmental Biology of Fishes 35:243–255.

Ridgway, M. S., B. J. Shuter, and E. E. Post. 1991. The relative influence of body size and territorial behaviour on nesting asynchrony in male smallmouth bass, *Micropterus dolomieui* (Pisces: Centrarchidae). Journal of Animal Ecology 60:665–681.

Robertson, D. M., and R. A. Ragotzkie. 1990. Changes in the thermal structure of moderate to large sized lakes in response to changes in air temperature. Aquatic Sciences 52:1015–1621.

Shuter, B. J., J. A. MacLean, F. E. J. Fry, and H. A. Regier. 1980. Stochastic simulation of temperature effects on first-year survival of smallmouth bass. Transactions of the American Fisheries Society 109: 1–34.

Strub, P. R., T. Powell, and C. R. Goldman. 1985. Climatic forcing: effects of El Niño on a small, temperate lake. Science 227:55–57.

Received July 19, 1995
Accepted January 22, 1996