in Proceedings of the Second Federal Interagency Hydrologic Modeling Conference, Las Vegas, Nevada, July 29 – August 1, 2002: Subcommittee on Hydrology of the Interagency Advisory Committee on Water Information.

DEVELOPMENT OF A NEURAL NETWORK MODEL FOR DISSOLVED OXYGEN IN THE TUALATIN RIVER, OREGON

By Stewart A. Rounds, Hydrologist, U.S. Geological Survey, Portland, Oregon, USA

Abstract

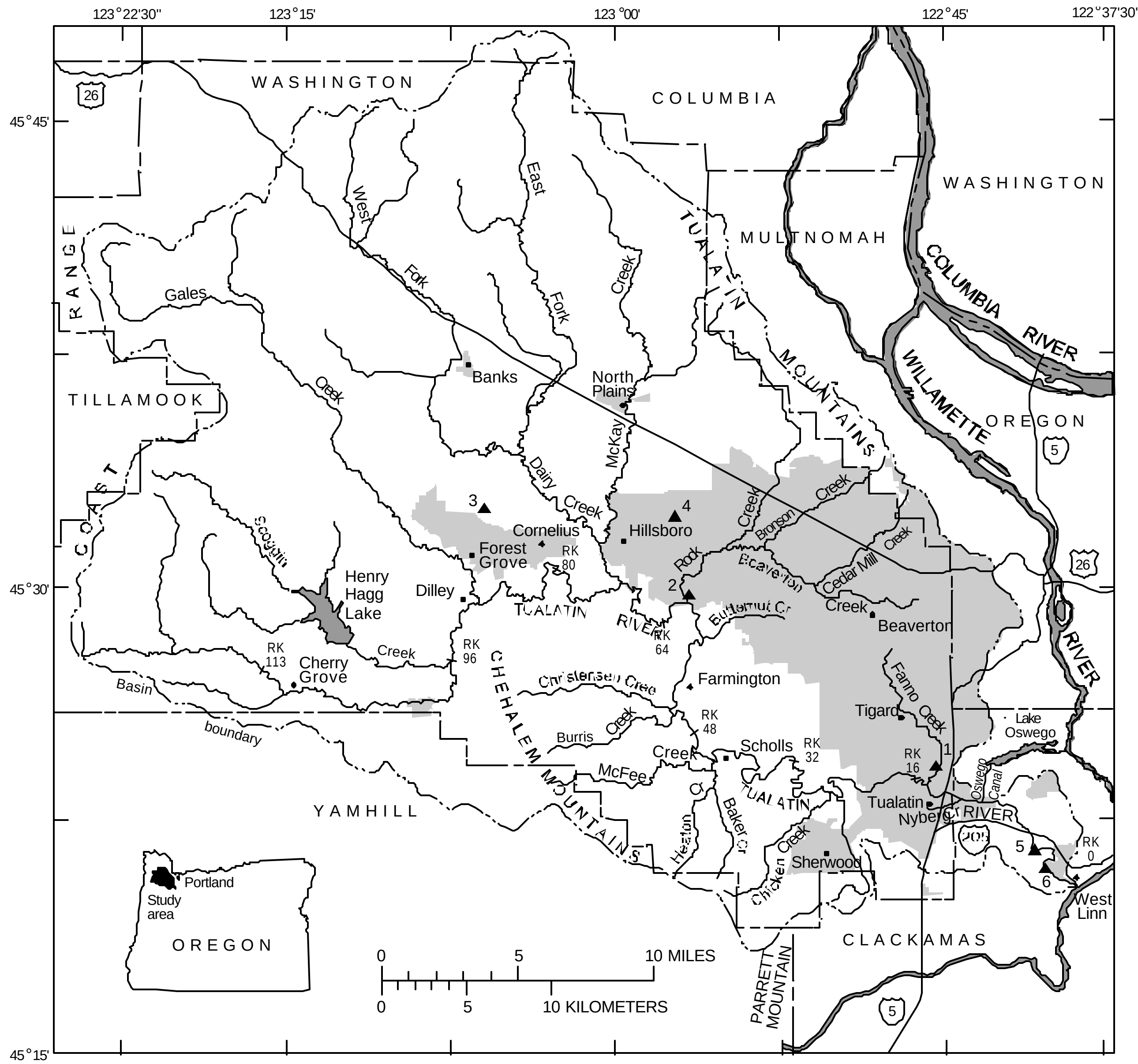
Dissolved oxygen concentrations in the lower reaches of the Tualatin River in northwest Oregon are the result of many processes. Temperature imposes a seasonal signal through the solubility of oxygen in water. Streamflow determines the travel time through the system and affects the amount of oxygen consumed via processes such as ammonia nitrification and the decomposition of organic material in the sediment and water column. Streamflow also affects the rate of oxygen exchange across the air/water interface. The available solar energy limits the photosynthetic production of oxygen by phytoplankton.

Many of the processes that affect dissolved oxygen concentrations in the Tualatin River – solubility, sediment oxygen demand, photosynthesis, respiration, biochemical oxygen demand, and reaeration – are controlled to some extent by physical and meteorological factors such as streamflow, air temperature, and solar radiation. To test the extent of that control, an artificial neural network model was constructed to predict dissolved oxygen concentrations in the Tualatin River at the Oswego Dam using only air temperature, solar radiation, rainfall, and streamflow as inputs. The Oswego Dam is a low-head structure located on a bedrock sill 5.5 kilometers upstream from the river's mouth. Hourly dissolved oxygen concentrations have been collected there since 1991; the available dataset spans more than 10 years.

Feedforward neural network modeling techniques, the most widely used type, were applied to this dataset. Data were segregated into calibration, verification, and test subsets. Two neural network models were constructed in series: the first model simulated daily mean dissolved oxygen concentrations, while the second superimposed any daily periodic signals. The final calibrated neural network models predicted the dissolved oxygen concentration with acceptable accuracy, producing high correlations between measured and predicted values (correlation coefficient of 0.83, mean absolute error less than 0.9 milligrams per liter). By some measures, neural network model performance was better than that of a calibrated, mechanistic model of dissolved oxygen in the Tualatin River. As expected, however, dissolved oxygen concentrations affected by factors other than the physical and meteorological factors used as model inputs, such as large point-source ammonia releases, were not predicted well by the neural network model. Nevertheless, the neural network model demonstrated potential for use as a river management and forecasting tool to predict the effects of flow augmentation and near-term weather conditions on Tualatin River dissolved oxygen concentrations.

INTRODUCTION

The Tualatin River drains a 1,840 km^2 (square kilometer) catchment on the west side of the Portland metropolitan area in northwest Oregon (fig. 1). Approximately 450,000 people live in the basin, mainly within a well-defined urban area. The population relies on the Tualatin River as a source of domestic, industrial, and irrigation water; habitat for fish and other wildlife; and a place to recreate. The river also receives highly treated municipal and industrial wastewater from the urban population.



Base modified from U.S. Geological Survey 1:100,000, topographic quadrangles, 1978-84

EXPLANATION

▲ Reference location

Map number	Site identification number	Site name and river kilometer location
1	452359122454500	Durham Wastewater Treatment Plant (RK 15.0)
2	452938122565500	Rock Creek Wastewater Treatment Plant (RK 61.3)
3		Agrimet meteorological station at Verboort, Oregon
4		Portland-Hillsboro Airport meteorological station
5	14207200	Tualatin River at Oswego Dam (RK 5.5)
6	14207500	Tualatin River at West Linn, Oregon (RK 2.9)

Designated urban growth area—From Metro, 1998

RK 16 River kilometer

Figure 1. Map showing the Tualatin River and selected data-collection sites.

The Tualatin River begins in the forested Coast Range mountains on the west side of the basin, where it flows for 24 kilometers (km) on a sometimes-steep bedrock substrate before reaching the valley bottom. Once on the valley bottom, the river meanders through sedimentary deposits in a predominantly agricultural region. In this 54-km meander reach, the river widens to approximately 15 meters with a mean depth of about 3 meters, but remains relatively shaded. The river next enters a 43-km backwater reach, where the water is slowed by the presence of a bedrock sill at river kilometer (RK) 15 and by a low-head dam at RK 5.5. The river further widens and deepens in the backwater reach, becoming wider than 50 meters with typical depths of 4.5 meters. The river's width in this reach prevents efficient shading; as a result, solar energy inputs often are sufficient to promote large algal populations during the summer. Downstream of the small dam at RK 5.5, the river again follows a bedrock channel with many riffles and intervening pools before joining the Willamette River at West Linn, Oregon.

Streamflow in the Tualatin River reflects the seasonal patterns in precipitation typical of the Pacific maritime climate in western Oregon. The highest flows occur in the winter rainy season between November and April, while the lowest flows normally occur in the late part of the dry summer period. The Tualatin is not a large river, with typical wintertime flows of 30-90 cubic meters per second (m^3/s) and summertime flows of only 4-6 m^3/s. Low flow can cause residence times in the river's backwater reach to be as long as 14-17 days.

Water-Quality Problems

Historically, the backwater reach of the Tualatin River exhibited many water-quality problems during the low-flow summer period. Low streamflow, coupled with plentiful nutrients (nitrogen and phosphorus), warm water, and ample light energy, provided sufficient time for large populations of phytoplankton to flourish before being transported downstream and out of that reach of the river. Algal blooms often degraded the aquatic health of the river by driving the pH above 8.5 and causing large variations in the dissolved oxygen (DO) concentration (3-5 mg/L [milligrams per liter]). After a bloom, respiring algae and decomposing organic material from the bloom often decreased the DO concentration to less than minimum acceptable levels (6 mg/L). Instream nitrification of large loads of ammonia discharged from wastewater treatment plants (WWTPs) contributed to low DO concentrations. Even after standard treatment controls were adopted at the WWTPs in the 1970s and 1980s, the river continued to have problems associated with high pH and low DO.

In response to these water-quality problems and in accordance with the Federal Clean Water Act, Total Maximum Daily Loads were adopted in 1988 for the Tualatin River and its major tributaries. The WWTPs were upgraded to state-of-the-art facilities in the early 1990s to remove phosphorus and ammonia. The object of phosphorus removal was to limit the growth of phytoplankton in the river. Ammonia discharges were decreased to reduce the instream consumption of DO by ammonia nitrification.

In 1990, the U.S. Geological Survey (USGS) began a long-term water-quality assessment of the Tualatin River with goals of (a) identifying the sources of nitrogen and phosphorus to the river, (b) assessing the transport and fate of those nutrients in the river, (c) identifying and quantifying the processes affecting DO in the river, and (d) constructing and using a process-based model of nutrients, phytoplankton, and DO in the Tualatin River. Results of the study have been published in a series of reports (for example: Kelly, 1997; Rounds and Doyle, 1997; Kelly and others, 1999; Rounds and others, 1999; Rounds and Wood, 2001).

Factors Affecting Dissolved Oxygen in the Backwater Reach

The USGS assessment revealed that the DO concentration in the Tualatin River, in the absence of large loads of ammonia from the WWTPs, is largely determined by several simple physical and meteorological factors: streamflow, air temperature, and solar radiation. DO is affected by many biological processes such as respiration, photosynthesis, and decomposition. Although biological processes directly influence DO, physical and meteorological factors control and limit the effects of those biological processes (Rounds and Wood, 2001).

Seasonal trends in DO are constrained by its solubility in water, which is a strong function of temperature. Consumption of DO by decomposition processes occurring in the water column and the sediments also is a function of water temperature and streamflow. Rates of the biologically mediated decomposition reactions are influenced by water temperature, and the DO consumed by those reactions is limited by the time that a given water parcel resides in a particular reach of the Tualatin River. Algal respiration and photosynthesis only affect the river's DO when large populations of phytoplankton are present, and such populations are possible only when sufficient light energy is available and when streamflow is low enough (< 8.5 m^3/s) to allow sufficient time for the phytoplankton to grow before being transported out of the backwater reach (fig. 2). Low concentrations of phosphorus can limit algal growth, but only during large algal blooms and near the surface of the river, where sunlight for photosynthesis is in ample supply.

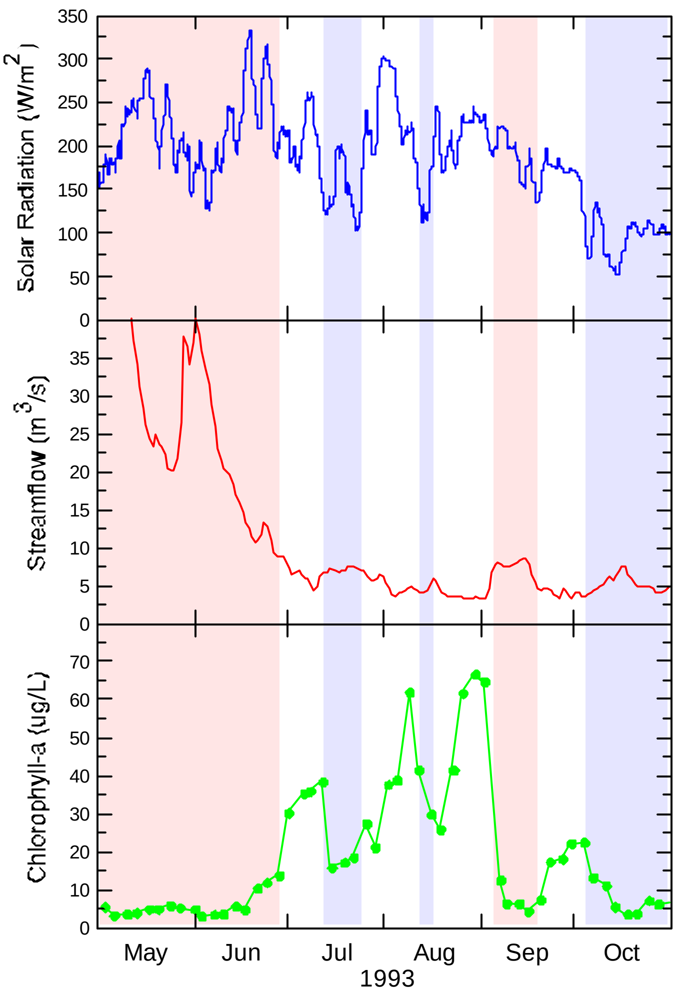


Figure 2. Favorable streamflow and light conditions are necessary before sizable algal blooms can occur in the Tualatin River. Shaded periods are unfavorable for algal growth due to high flow (red) or low light (blue) conditions. Data from 1993 (Doyle and Caldwell, 1996).

Using these findings, the USGS process-oriented model of DO in the Tualatin River was successful in simulating patterns in measured DO concentrations that result from seasonal temperature variations, periodic blooms of phytoplankton, and point-source discharges of oxygen-consuming substances such as ammonia. The model simulated the river's DO concentrations with good accuracy, producing a mean absolute error of 0.9, 1.0, and 1.6 mg/L at three important locations in the river's backwater reach over a 42-month period spanning seven hydrologically distinct summers between 1991 and 1997 (Rounds and Wood, 2001). Process-oriented models, however, typically require copious data on meteorological conditions and the quantity, temperature, and chemical characteristics of all inflows, outflows, and instream sites. Many of these data, particularly the chemical data, are not available in real time. Therefore, process-oriented water-quality models cannot be run in real time to provide feedback to river managers who may need model results to set an appropriate level of flow augmentation.

Objectives and Approach

If the DO concentration in the Tualatin River could be predicted only from data that are collected in real time, then river managers would be better able to manage the river's water quality. The river's DO is influenced greatly by physical and meteorological factors, but whether the DO concentration can be predicted from such factors with any accuracy was unknown.

The purpose of this study was to determine the extent to which the DO concentration in the Tualatin River at the Oswego Dam (fig. 1) can be predicted solely from physical and meteorological measurements such as streamflow, air temperature, solar radiation, and rainfall, using multiple linear regression and artificial neural network modeling techniques. Other real-time water-quality measurements (water temperature, specific conductance, etc.) are available and could be included in this analysis, but the primary goal was to find out whether the information present in the streamflow and meteorological measurements is sufficient to predict DO with an acceptable level of error. The extent to which other measurements of water quality (water temperature and specific conductance) might improve the predictions also was tested. Future work may incorporate these techniques into a real-time water-quality forecasting tool.

ARTIFICIAL NEURAL NETWORKS

An artificial neural network (ANN) is a mathematical structure designed to mimic the information processing functions of a network of neurons in the brain (Hinton, 1992; Jensen, 1994). ANNs are highly parallel systems that process information through many interconnected units that respond to inputs through modifiable weights, thresholds, and mathematical transfer functions. Each unit processes the pattern of activity it receives from other units, then broadcasts its response to still other units. ANNs are particularly well suited for problems in which large datasets contain complicated nonlinear relations among many different inputs. ANNs are able to find and identify complex patterns in datasets that may not be well described by a set of known processes or simple mathematical formulae.

In this application, simply suspecting that streamflow, air temperature, solar radiation, and rainfall influence instream DO concentrations is sufficient to apply an ANN. Unlike a process-based model, it is not necessary to know exactly how those variables interact, the nature of the physical/chemical/biological processes that cause those patterns, or any mathematical representation of those processes before applying an ANN. As a result, ANN models can be developed more quickly and with less expense than typical process-based models. Because ANNs contain no internal "knowledge" of the processes behind the data patterns, however, they are less able to provide additional insight into those processes (Conrads and Roehl, 1999). Nevertheless, ANNs can be useful tools for finding and predicting patterns in water-quality data.

Hundreds of different types of ANNs exist. The most commonly used type of ANN is a type of feedforward network termed the multilayer perceptron, an example of which is illustrated in figure 3. In this type of network, the artificial neurons, or processing units, are arranged in a layered configuration containing an input layer, usually one "hidden" layer, and an output layer. Units in the input layer introduce normalized or filtered values of each input into the network (ANNs work best if the inputs are scaled to the same range of values). Units in the hidden and output layers are connected to all of the units in the preceding layer. Each connection carries a weighting factor. The weighted sum of all inputs to a processing unit is calculated and compared to a threshold value. That activation signal then is passed through a mathematical transfer function to create an output signal that is sent to processing units in the next layer. Training an ANN is a mathematical exercise that optimizes all of the ANN's weights and threshold values, using some fraction of the available data. Optimization routines can be used to determine the ideal number of units in the hidden layer and the nature of their transfer functions. ANNs "learn" by example; as long as the input dataset contains a wide range of the types of patterns that the ANN will be asked to predict, the model is likely to find those patterns and successfully use them in its predictions.

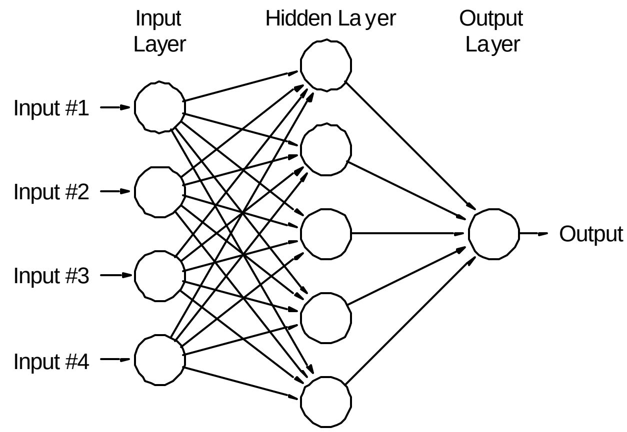


Figure 3. A representation of a simple 3-layer feedforward artificial neural network with four inputs, 5 hidden nodes, and one output.

In this study, Statistica Neural Networks software (StatSoft, 2000) was used to create and train the ANN models. Simple 3-layer feedforward networks were used, where the number of units in the hidden layer was optimized by the software and by manual testing. Standard training methods (back-propagation and conjugate gradient descent) were used for initial network identification and selection of the best set of inputs. Final ANN models were trained using Levenberg-Marquardt optimization, which is the fastest and most reliable ANN training method for relatively small networks with a single output (DO).

DATA PREPARATION AND DECORRELATION

This investigation focuses on predicting DO concentrations in the Tualatin River at the Oswego Dam (USGS station 14207200) for the May-October periods of 1991 through 2000. May through October is the general time frame for the low-flow summer period when most DO problems are likely to occur. Hourly DO data are available for the Oswego Dam station during that time period. To capture the effects of physical and meteorological forcings on the river's DO concentration, measured values of streamflow, solar radiation, air temperature, and rainfall were available for the same time period (table 1).

To maximize the number of useful records in the dataset, secondary sources of data were sometimes used to fill gaps in the data from the primary source. For example, rainfall data from the Portland-Hillsboro Airport were used to fill gaps in the rainfall record from the Verboort Agrimet station. Some gaps remained in the data, but most gaps were small, resulting in more than 40,000 useful records for modeling. The time period August 23 - September 3 of 1996 was excluded from the model dataset due to a rare release of a large quantity of ammonia from one of the WWTPs that affected the DO at the Oswego Dam; physical and meteorological factors are unrelated to that point-source problem and inclusion of that time period would only serve to distort other patterns in the data.

Table 1. Sources of data. The first station listed is the primary source for each parameter.
[USGS = U.S. Geological Survey; BOR = Bureau of Reclamation; NWS = National Weather Service]

Parameter	Data Frequency	Data Source	Station (ID Number)	Map Number (fig. 1)
Dissolved Oxygen	Hourly	USGS	Tualatin River at Oswego Dam (14207200)	5
Streamflow	Hourly	USGS	Tualatin River at West Linn, OR (14207500)	6
Solar Radiation	Hourly	USGS	Durham WWTP (452359122454500)	1
Air Temperature	Hourly	USGS	Tualatin River at Oswego Dam (14207200)	5
		USGS	Rock Creek WWTP (452938122565500)	2
		BOR	Agrimet meteorological station at Verboort, OR	3
Rainfall	Daily	BOR	Agrimet meteorological station at Verboort, OR	3
		NWS	Portland-Hillsboro Airport meteorological station	4

Because ANN models have no underlying knowledge of the processes affecting the input and output variables, it is critical to examine the data for periodicity, cross-correlations, and important time lags (Roehl and Conrads, 2000). Results from such analyses can be used to maximize the signals in the input data that will help to predict the output.

Periodicity

Each parameter's data were analyzed by Fourier transform to determine the presence of periodic signals. Solar radiation, air temperature, and DO all had strong periodic signals at daily time scales; periods of 24 and 12 hours characterized the most important signals. Streamflow appeared to have useful signals at time scales longer than a day or two, but only weak patterns at shorter time scales. Figure 4 illustrates typical power spectrums from these data. Strong signals at daily time scales can obscure important correlations and time lags in the data (Risley and others, 2002); therefore, the short and long time scale signals in the data were separated. A low-pass filter was used to remove the 24-hour and shorter periodic signals from each time series, preserving any periodic signals at time scales longer than one day; the resulting time series were equivalent to the 24-hour running average of each input. Long-term patterns and short-term periodicity in the data were simulated with separate models.

Cross-Correlations and Time Lags

Multiple linear regression and ANN techniques work best if the data inputs are as independent as possible. To test for interdependence, the data were correlated against one another using linear regression techniques. The analysis was extended to identify important time lags by analyzing the correlation coefficients with an imposed time lag. Figure 5 illustrates how the low-pass filtered DO data correlate with other time-lagged and low-pass filtered input variables. Each input variable appears to offer some information that might be helpful in predicting DO; solar radiation and air temperature appear to offer the best linear correlations.

Figure 5 also illustrates several important time lags that are present in the data. The signal in the solar data is maximized when the solar data are lagged in time by 1.75 days; in other words, DO

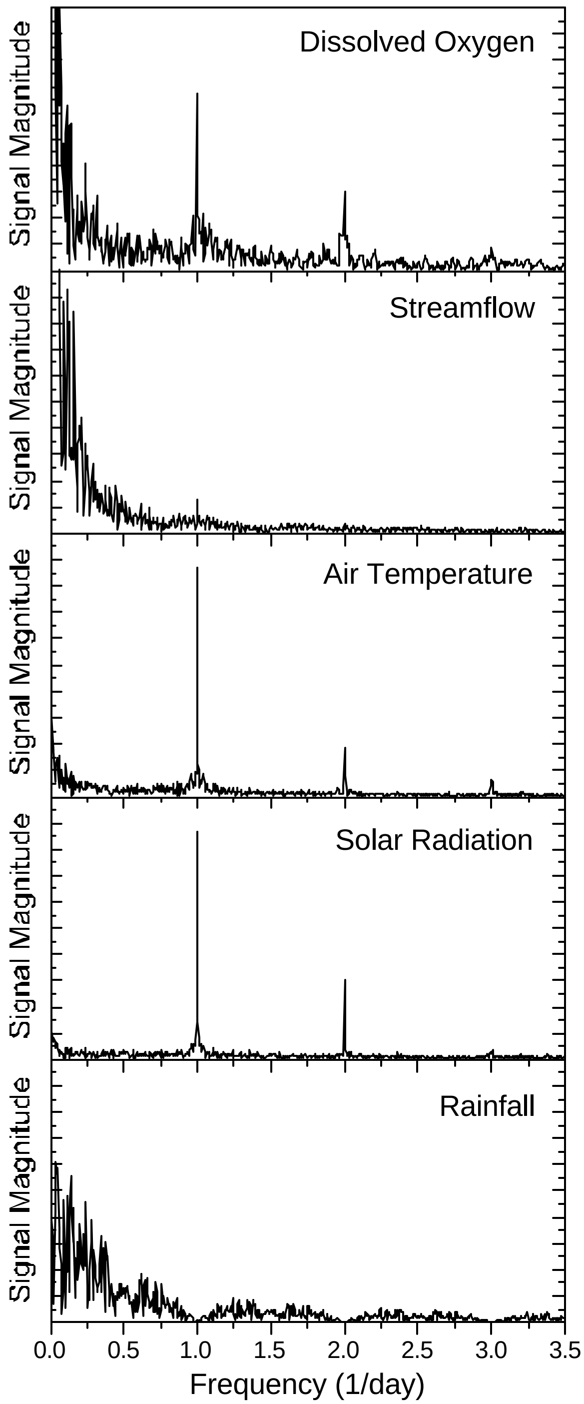


Figure 4. Typical power spectrums for the dissolved oxygen, streamflow, air temperature, solar radiation, and rainfall data.

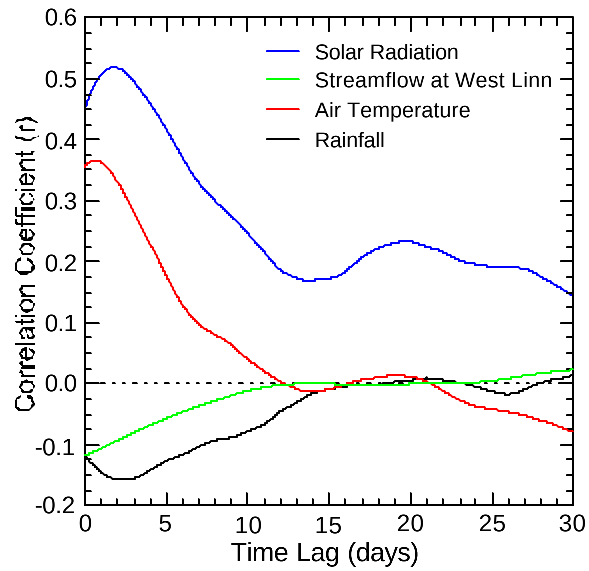


Figure 5. Correlations and time-lags between low-pass filtered (daily-mean) dissolved oxygen and other low-pass filtered (daily-mean) inputs.

has its highest correlation with the solar insolation rate that occurred about 2 days previous. That time lag has a physical basis because the available solar energy affects the amount of DO produced by photosynthesis, and the effects of very sunny or very cloudy days on algal growth are not immediate. The best air temperature signal is lagged by 2/3 of a day, and the greatest correlation with rainfall is lagged by 2.3 days.

Many of the DO cross-correlations are minimized at a time lag on the order of 12 days, which makes sense because the typical summer residence time in the backwater reach of the Tualatin River is on the order of 10-14 days, depending on streamflow. Autocorrelation of the DO data also shows a minimum at a time lag of about 12 days. Streamflow data from several main-stem Tualatin River gages were analyzed; all were found to be highly cross-correlated (not shown). Data from just one gage, therefore, were sufficient to capture the signal in the streamflow data.

LOW-FREQUENCY (DAILY MEAN) MODELS

The cross-correlation and time-lag results indicate that the information in the input data is maximized by using the following manipulations of those data as model input, where "lp" denotes the low-pass filter, "lag" means time-lagged, and the symbols Q, S, AT, and R stand for streamflow, solar radiation, air temperature, and rainfall, respectively:

- low-pass filtered data: lp-Q, lp-S, lp-AT, R (raw rainfall data were daily)
- low-pass filtered data from 12 days previous: lp-Q-12, lp-S-12, lp-AT-12, R-12
- time-lagged filtered data: lp-AT-lag (2/3 day), lp-S-lag (1.75 days), R-lag (2.3 days)
- miscellaneous: year, day-of-year, fraction-of-day

The data from 12 days previous provide long-term slope information. To avoid correlations between unmodified and time-lagged data, the time-lagged inputs were calculated as differences between the original and lagged data. Using differences is a good way to decorrelate inputs.

Initial searches for the best ANN to predict the low-pass (daily mean) DO revealed that some inputs were more important than others. Indeed, some inputs seemed to convey little useful, independent information. To create the most efficient model, the model inputs were culled to leave only these eight: lp-Q, lp-AT, lp-S, lp-Q-12, lp-AT-12, lp-S-lag, day-of-year, and year. Note that rainfall data were eliminated. Any signal in the rainfall data apparently was redundant with information in the air temperature and solar data; the presence of many zero values on dry days also may have decreased the utility of the rainfall data.

In all of the models tested, half the data were randomly selected for model training (calibration). Half the remaining data were used for verification, and the rest were used as an independent test dataset. Statistica Neural Networks uses the training data for training the model – optimizing the model's weights and threshold values. During training, the verification data are used as feedback to ensure that the model does not become overtrained; overtraining is a condition in which the model finds patterns to decrease the error in the training dataset that are not reflected in the larger dataset. Because the verification data are used to prevent overtraining and create a more robust model, they are not a true, independent test of the model. For that reason, it is useful to reserve a third portion of the dataset for an independent test of the trained model. In all cases, model performance was almost identical for each of the training, verification, and test datasets.

Multiple Linear Regression

Multiple linear regression may be viewed as a special case ANN model that uses linear transfer functions and no hidden layers. If the linear model performs as well as a more complex ANN, then using the nonlinear ANN may not be justified; thus, linear models are useful as a basis for comparison. Multiple linear regression analysis of the low-frequency data (8 inputs as listed above) revealed that the patterns in the data must be highly nonlinear, as the linear model failed to capture the important patterns in the measured DO data (table 2).

Artificial Neural Network

Optimization revealed that the best ANN for these eight inputs included one hidden layer with seven processing units. Logistic transfer functions ($1/(1+e^{-x})$) were used for all hidden-layer units. Other popular transfer functions, including the hyperbolic tangent function, were tried but none produced better results. ANN predictions were markedly better than the linear model and in many cases better than the results from the USGS process-based model, with a mean absolute error of only 0.83 mg/L and a correlation coefficient of 0.837 (fig. 6, table 2). The ANN model captured most of the important patterns in the data, producing remarkable fits to the measured DO considering that the predictions were based only on streamflow, air temperature, solar radiation, year, and day-of-year. Sensitivity analyses showed that the most important predictor variables were lp-Q, day-of-year, lp-S, and lp-S-lag, respectively.

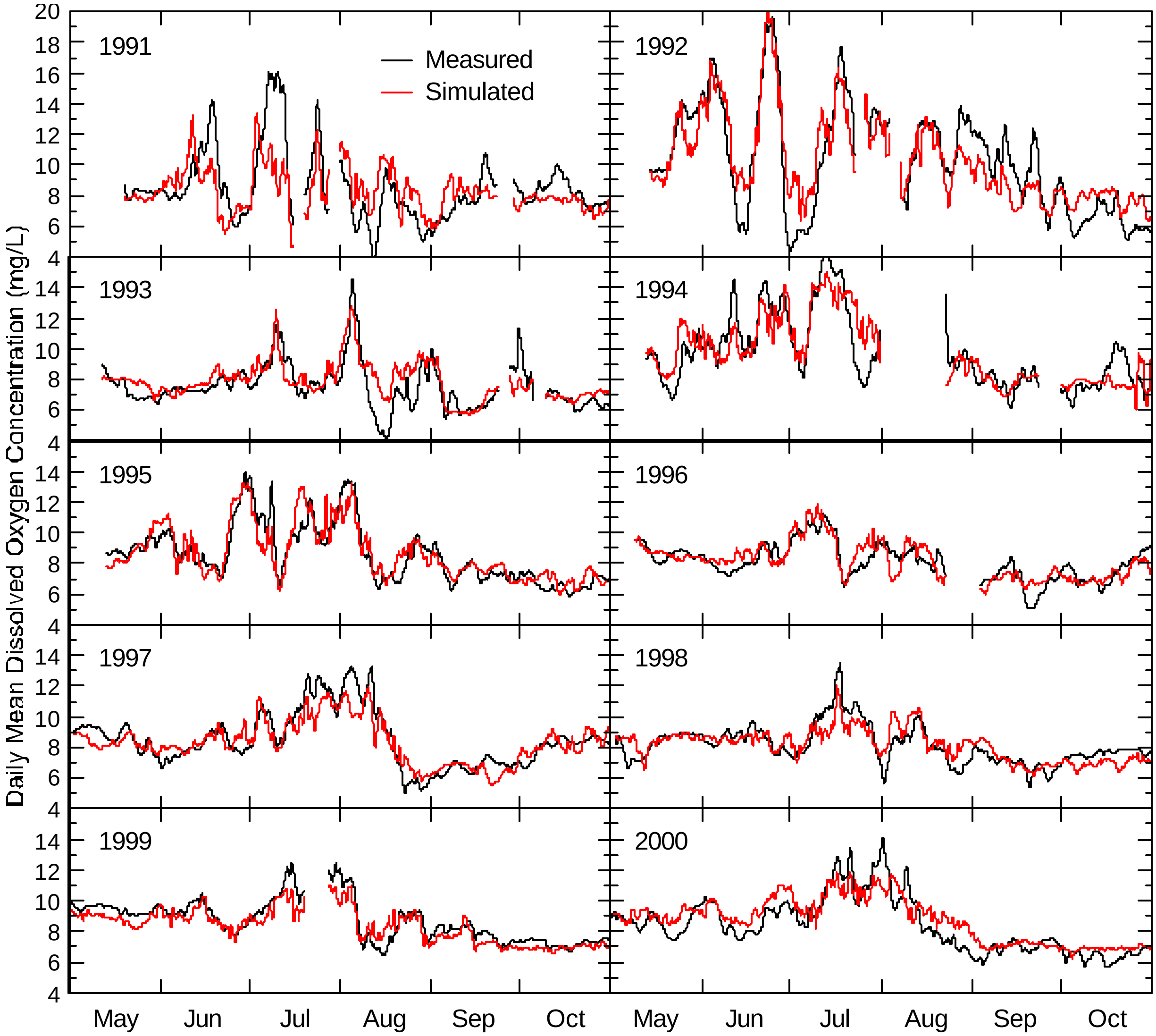


Figure 6. Measured and simulated daily-mean (low-pass) dissolved oxygen concentrations for the Tualatin River at Oswego Dam (station 14207200). Simulated values were calculated by the low-frequency ANN model (8 inputs, 1 hidden layer with 7 processing units, 1 output).

ANN models built with additional inputs of water temperature and specific conductance from the Tualatin River at Oswego Dam produced slightly better results (mean absolute error of 0.75 mg/L), but additional models to predict those inputs would be required if this more complex ANN were used for DO forecasting. For the purpose of forecasting, future values of inputs must be known or estimated in a reliable manner. That exercise is left as a subject for future study.

FINAL HOURLY MODEL

High-frequency signals in the data were separated from low-frequency signals by subtracting the low-pass filtered data from the original data. The "high-pass" (hp) solar signal, for example, was determined as: hp-S = S – lp-S. High-pass data have the long-term trends removed and reflect only daily and shorter variations. These high-pass data, as well as the output of the low-frequency ANN model (call that "lp-DO_{ANN}"), were used as inputs to a new ANN to predict the measured hourly DO. The high-pass air temperature and solar inputs were included at several time lags to capture the 12-hour and 24-hour signals in their data; the power spectra in figure 4 show that the streamflow and rainfall data have no useful information at these short time scales.

Testing of various ANNs with time-lagged, high-pass air temperature and solar inputs revealed that the information in those inputs was captured adequately with time lags of 2, 23, and 56 hours for air temperature and 4, 25, and 58 hours for solar radiation. Further testing showed that the best model needed only nine inputs: lp-DO_{ANN}, hp-AT-lag (3 lags), hp-S-lag (3 lags), day-of-year, and year. Final training and optimization yielded an ANN with one hidden layer containing 10 processing units. As with the low-frequency ANN, logistic transfer functions were used for the hidden-layer units. The final hourly ANN model captured the long-term and daily patterns in the measured DO data, fitting the data with a mean absolute error of 0.86 mg/L and a correlation coefficient of 0.831 (table 2). Figure 7 illustrates the daily variations that the model produced in the final DO predictions for a subset of the data. These final predictions appear to be accurate enough to be useful. Future work will focus on incorporating these and other ANN models into real-time water-quality forecasting tools.

Table 2. Goodness-of-fit statistics for the models predicting Tualatin River DO at Oswego Dam.

Model Type	Time Scale	Number of Points	Mean Absolute Error	Root Mean Square Error	Correlation Coefficient
Multiple Linear Regression	low-frequency	40,388	1.29 mg/L	1.69 mg/L	0.589
ANN	low-frequency	40,388	0.83 mg/L	1.14 mg/L	0.837
	final hourly	40,372	0.86 mg/L	1.21 mg/L	0.831

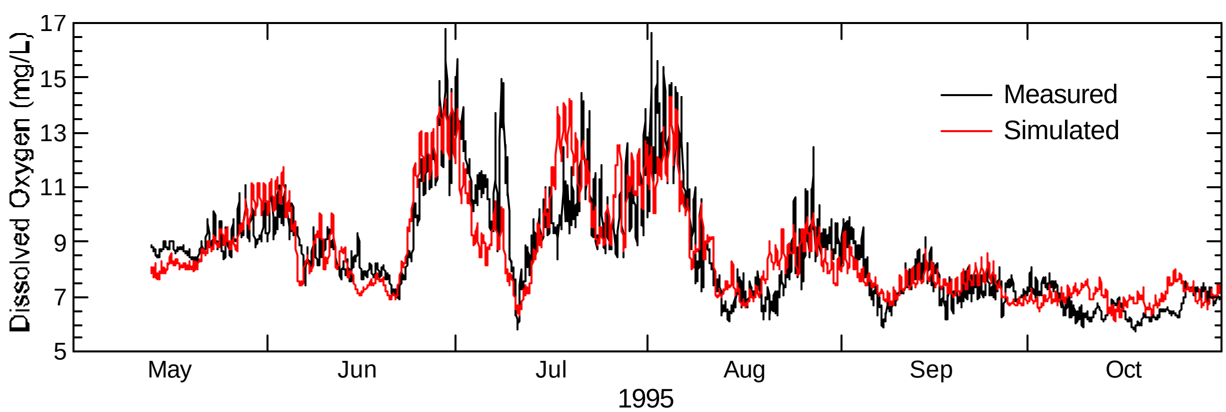


Figure 7. Measured and simulated hourly dissolved oxygen concentrations for the summer of 1995 in the Tualatin River at Oswego Dam (station 14207200). Simulated values were calculated by the final hourly ANN model (9 inputs, 1 hidden layer with 10 processing units, 1 output).

CONCLUSIONS

Artificial neural network models were developed to simulate daily mean and hourly DO concentrations in the Tualatin River at the Oswego Dam. The DO at that site is affected by its solubility as well as biological processes such as algal photosynthesis and respiration, sediment oxygen demand, biochemical oxygen demand, and ammonia nitrification. The effects of these biological processes were hypothesized to be constrained by a small set of physical and meteorological factors: streamflow, air temperature, solar radiation, and rainfall. Neural network and regression models were built to test this hypothesis, using data from May-October of 1991-2000.

Multiple linear regression models failed to capture the long-term patterns in the DO data, producing poorly correlated results. Neural network models, however, were successful in predicting patterns in the DO data on daily, weekly, and seasonal time scales. ANN model performance was good, with mean absolute errors less than 0.9 mg/L. The ANN predictions often were better than those from a USGS process-based model of the Tualatin River. As applied to the Tualatin River, however, ANN and process-based models have different purposes. The process-based model is most useful for providing insight into how the river works, identifying important processes, and testing the effects of point-sources and management strategies. The ANN model has tremendous potential as a forecasting tool, but yields less insight into the specifics of riverine processes.

Now that it has been demonstrated that DO in the Tualatin River can be predicted with acceptable accuracy from a small set of physical and meteorological measurements, future work will concentrate on the development of a real-time DO forecasting tool using these ANN techniques. Refinements may include the prediction of water temperature as a first step, so that DO solubility estimates may be included as model input.

REFERENCES

Conrads, P.A. and Roehl, E.A., 1999, Comparing physics-based and neural network models for simulating salinity, temperature, and dissolved-oxygen in a complex, tidally affected river basin, *in* Proceedings of the 1999 South Carolina Environmental Conference, Myrtle Beach, South Carolina: http://smig.usgs.gov/features_0399/nnm1.html, accessed 4/15/2002.

Doyle, M.C. and Caldwell, J.M., 1996, Water-quality, streamflow, and meteorological data for the Tualatin River Basin, Oregon, 1991-93: U.S. Geological Survey Open-File Report 96-173, 49 p., CD-ROM

Hinton, G.E., 1992, How neural networks learn from experience, *Sci. American*, **267**, 144-151.

Jensen, B.A., 1994, Expert systems - neural networks, *Instrument Engineers' Handbook*, 3rd ed.: Radnor, Pennsylvania, Chilton, p. 48-54.

Kelly, V.J., 1997, Dissolved oxygen in the Tualatin River, Oregon, during winter flow conditions, 1991 and 1992: U.S. Geological Survey Water-Supply Paper 2465-A, 68 p.

Kelly, V.J., Lynch, D.D., and Rounds, S.A., 1999, Sources and transport of phosphorus and nitrogen during low-flow conditions in the Tualatin River, Oregon, 1991-1993: U.S. Geological Survey Water-Supply Paper 2465-C, 94 p.

Risley, J.C., Roehl, E.A., and Conrads, P.A., 2002, Estimating water temperatures in small streams in western Oregon using neural network models: U.S. Geological Survey Water-Resources Investigations Report 02-4218.

Roehl, E.A. and Conrads, P.A., 2000, Real-time control of the salt front in a complex, tidally affected river basin, *in* Proceedings of the ANNIE 2000 Conference, Nov. 5-8, 2000, St. Louis, MO: ASME Press: http://smig.usgs.gov/features_0900/annie.html, accessed 4/15/2002.

Rounds, S.A. and Doyle, M.C., 1997, Sediment oxygen demand in the Tualatin River Basin, Oregon, 1992-1996: U.S. Geological Survey Water-Resources Investigations Report 97-4103, 19 p.

Rounds, S.A., Wood, T.M., and Lynch, D.D., 1999, Modeling discharge, temperature, and water quality in the Tualatin River, Oregon: U.S. Geological Survey Water-Supply Paper 2465-B, 121 p.

Rounds, S.A. and Wood, T.M., 2001, Modeling water quality in the Tualatin River, Oregon, 1991-1997: U.S. Geological Survey Water-Resources Investigations Report 01-4041, 53 p.

StatSoft, Inc., 2000, Statistica Neural Networks, version 4.0E: Tulsa, Oklahoma.