WATER RESOURCES RESEARCH, VOL. 33, NO. 10, PAGES 2423–2424, OCTOBER 1997

Comment on "The use of artificial neural networks for the prediction of water quality parameters" by H. R. Maier and G. C. Dandy

Vincent Fortin, Taha B. M. J. Ouarda, and Bernard Bobée
Institut National de la Recherche Scientifique, Sainte-Foy, Québec, Canada

Nonparametric approaches to prediction and forecasting of complex physical systems, such as artificial neural networks (ANNs), are set to spread with modern computing possibilities. As *Maier and Dandy* [1996] correctly point out, data-driven models make prediction and forecasting possible under less stringent hypotheses. This paper is highly appreciated for introducing ANNs to the water resources community and presenting a practical application. However, as many readers may be unfamiliar with this type of model, we would like to clarify links that exist between ANNs and autoregressive-moving average (ARMA) models, which the authors did not emphasize. We would also like to suggest the use of a different network configuration, which may prove more appropriate for time series forecasting.

To discuss the links that exist between ANNs and ARMA models, a common vocabulary is needed. ANNs, which were originally designed to solve artificial intelligence (AI) problems such as speech and hand writing recognition, are often described using AI terminology. While this may have been appropriate when the purpose of ANNs was to loosely model the human brain, it is now a major source of confusion. For example, the authors name "training" or "learning" the process by which the weights of an ANN are adjusted so as to minimize mean squared error (section 2.2.3). However, in fact, this is absolutely equivalent to the calibration process of any stochastic model: training is equivalent to "calibrating." The algorithm most often used for calibration of an ANN, called "backpropagation" by the AI community (see section 3 for instance) is nothing else but the well-known steepest-descent method of optimization [*Cauchy*, 1847]. Backpropagation also refers in some of the literature (and in this paper) to the type of feed forward network used by the authors, but this terminology is misleading as other equally valid algorithms may be used for calibration, including the conjugate gradient method [*Fletcher and Reeves*, 1964], simulated annealing [*Aarts and Korst*, 1989], genetic algorithms [*Holland*, 1992], and evolutionary programming [*Fogel et al.*, 1989]. At this point in their development, ANNs are simply another type of "black-box" model [*Welstead*, 1994]. An ANN needs to be configured and calibrated like any other type of model, and using similar algorithms, it does not, in any way, "learn" a relationship from the data, unless we define as learning the process of fitting a model to observed data.

Machine learning is best defined by *Mogili and Sunol* [1993, p. 756] as "a process in which a computer program improves its performance, acquires knowledge and solves new problems in a specified domain". These authors also correctly state that ANNs "... are unsuitable for knowledge acquisition purposes as they do not allow for knowledge extraction in the form of rules ... ". By themselves, ANNs are not appropriate for acquiring and organizing knowledge but can be useful as part of a larger automated learning system [*Mogili and Sunol*, 1993]. In short, there is nothing mystical about an artificial neural network, and using AI terminology sustains the hope that ANNs may be almost as efficient as humans in picking up patterns in data, which is still far from the truth.

When properly used and understood, ANNs often prove very efficient: being able to model nonlinear relationships, they have a clear advantage over most commonly used stochastic models. In fact, the type of feed forward ANN used by the author has been shown to be a type of nonlinear autoregressive (NAR) model [*Connor et al.*, 1994]. No matter the number of hidden layers and the number of hidden nodes in each layer, a feed-forward ANN used for time series analysis may always be described by (1), for some finite value of p:

$$x_t = h(x_{t-1}, x_{t-2}, \cdots, x_{t-p}, \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \cdots, \mathbf{y}_{t-p}) + e_t \tag{1}$$

where h can be any function, $\mathbf{y}_t$ represents a vector of related time series, and e_t is a random shock of zero mean and finite variance.

NAR models suffer some of the same shortcomings as linear autoregressive (AR) models. Specifically, they cannot model dependence of x_t on previous random shocks $e_{t-1}, e_{t-2}, \cdots$, and so cannot model moving average (MA) components of a time series. Recurrent ANNs, which use as inputs previous observed random shocks, are a generalization of feed forward ANNs and allow modeling of complex nonlinear autoregressive-moving average systems (NARMA). They can thus be described by (2), for some finite values of p and q:

$$x_t = h(x_{t-1}, x_{t-2}, \cdots, x_{t-p}, e_{t-1}, e_{t-2}, \cdots, e_{t-q}, \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \cdots, \mathbf{y}_{t-p}) + e_t \tag{2}$$

where h can be a nonlinear function. Fully recurrent NARMA networks [*Connor et al.*, 1994] have been shown to perform much better than the NAR model proposed in the paper by *Maier and Dandy* [1996] when moving average components are present in the stochastic process. We thus question the statement by the authors that feed forward network (or backpropagation network, as named by the authors) "... are the type of ANN most suited to forecasting applications" (section 3). Serious thought should be given to recurrent designs for forecasting time series.

Understanding that the feed-forward ANNs used in this paper are simply a class of NAR, we can comment on two other statements made by the authors about feed forward ANNs. It is mentioned that "... when developing ANN models, the statistical distribution of the data does not have to be known ... " (section 3), meaning in particular that ANNs are

Paper number 97WR00969.
0043-1397/97/97WR-00969$09.00

applicable when the random shock component is not normal, contrary to ARMA models. In fact, the objective function used to calibrate both types of models (mean squared error) is justified mainly because it maximizes the likelihood of the model under the hypothesis of normal (or at least symmetrical) random shocks. If their distribution is known to be skewed, then the objective function needs to be adapted for both types of models. It is also mentioned that a disadvantage of model-driven approaches, such as ARMA models, is that "... the model order has to be determined before the unknown model parameters can be estimated" (section 3). We would like to point out that the process of selecting the number of hidden layers and hidden nodes in each layer is also quite complex and must be conducted prior to calibrating the network [*Bebis and Georgiopoulos*, 1994]. Furthermore, for ARMA models, there exist well-established statistical procedures, such as MAICE [*Akaike*, 1972], to help select the order of the model.

We would also have liked the authors to substantiate their claim that ANNs can efficiently model nonstationarities in the data without differencing or otherwise transforming it (section 3), instead of only referring to a research report [*Maier and Dandy*, 1995]. Of course, it is well known that feed-forward ANNs can approximate any function to any degree of accuracy [*Hornik et al.*, 1989], including nonstationary functions. However, in practice modeling nonstationary data with an ANN is quite difficult since the network, once calibrated, will be faced with input patterns the like of which it has never seen before: instead of interpolating between similar previously observed input-output examples, the ANN will be asked to extrapolate. Also, ANNs are usually not very good at extrapolating, given their large number of free parameters. In fact, it can be argued that nonparametric, data-driven models, such as ANNs, are in general not suited for extrapolation, and at best constitute very arbitrary models for this task [*Fortin et al.*, 1997]. Extrapolation only makes sense when based on clearly stated verifiable hypotheses; complex black-box models do not qualify. Not knowing precisely what *Maier and Dandy* [1995] have discovered, we would prefer nonstationary stochastic models [*Young*, 1994] or multivariate adaptive regression splines [*Friedman*, 1991; *Lall et al.*, 1996] for modeling nonstationary data without transforming it prior to the analysis.

All this is not to say that ANNs were not useful in the practical problem studied by *Maier and Dandy* [1996]. However, as suggested by the authors, a thorough comparison with ARMA models would be needed to confirm their value. Furthermore, recurrent architectures should be considered. In fact, since an ANN is at best a nonlinear ARMA model, we suggest fitting a linear ARMA model first and then verifying, by the use of an ANN, if the shortcomings of the ARMA model are caused by the hypothesis of linearity. Finally, we strongly believe that presenting an artificial neural network as what it really is, a type of nonlinear AR or ARMA model and not quite a model of the brain, will help users in the field of water resources understand and apply this powerful tool wisely.

References

Aarts, E., and J. Korst, *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*, John Wiley, New York, 1989.

Akaike, H., Use of an information theoretic quantity for statistical model identification, *Proc. 5th Hawaii Int. Conf. Sys. Sci.*, 249–250, 1972.

Bebis, G., and M. Georgiopoulos Feed-forward neural networks–Why network size is so important, *IEEE Potentials*, 27–31, 1994.

Cauchy, A. L., Méthode générale pour la résolution des systèmes d'équations simultanées. *C. R. Acad. Sci.*, Paris, *25*, 536–538, 1847.

Connor, J. T., R. D. Martin, and L. E. Atlas, Recurrent neural networks and robust time series prediction. *IEEE Trans. Neural Networks*, *5*(2), 240–254, 1994.

Fletcher, R., and C. M. Reeves, Function minimization by conjugate gradient, *Comput. J.*, *7*(2), 149–154, 1964.

Fogel, L. J., A. J. Owens, and M. J. Walsh, *Artificial Intelligence Through Simulated Evolution*, John Wiley, New York, 1989.

Fortin, V., J. Bernier, and B. Bobée, Simulation, Bayes, and bootstrap in statistical hydrology, *Water Resour. Res.*, *33*(3), 439–448, 1997.

Friedman, J. H., Multivariate adaptive regression splines, *Ann. Stat.*, *19*(1), 1–141, 1991.

Holland, J. H., *Adaptation in Natural and Artificial Systems*, 2nd ed., MIT Press, Cambridge, Mass., 1992.

Hornik, K., M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks*, *2*, 359–366, 1989.

Lall, U., T. Sangoyomi, and H. D. I. Abarbanel, Nonlinear dynamics of the Great Salt Lake: Nonparametric short-term forecasting, *Water Resour. Res.*, *32*(4), 975–985, 1996.

Maier, H. R., and G. C. Dandy, Comparison of the Box-Jenkins procedure with artificial neural network methods for univariate time series modelling, *Res. Rep. R127*, Dep. of Civ. and Environ. Eng., Univ. of Adelaide, Adelaide, Australia, 1995.

Maier, H. R., and G. C. Dandy, The use of artificial neural network for the prediction of water quality parameters, *Water Resour. Res.*, *32*(4), 1013–1022, 1996.

Mogili, P. K., and A. K. Sunol, Machine learning approach to design of complex distillation columns, in *Applications of Artificial Intelligence in Engineering VIII*, Vol 2, *Applications and techniques*, edited by G. Rzevski, J. Pastor, and R. A. Adey, pp. 755–770, Comput. Mech. Publ., Southampton, England, 1994.

Welstead, S. T., *Neural Network and Fuzzy Logic Applications in C/C++*, John Wiley, New York, 1994.

Young, P. C., Time-variable parameter and trend estimation in nonstationary economic time series, *J. Forecasting*, *13*, 179–210, 1994.

B. Bobée, V. Fortin, and T. B. M. J. Ouarda, Institut National de la Recherche Scientifique, Université du Québec, P. O. Box 7500, Sainte-Foy, Quebec, Canada G1V 4C7. (e-mail: bernard_bobee@inrs-eau.uquebec.ca)

(Received June 18, 1996; accepted March 28, 1997.)