IMPROVED PARAMETER ESTIMATION FOR MULTIPURPOSE HYDROLOGIC MODELS

C. T. Haan
Regents Professor & Sarkeys Distinguished Professor
Agricultural Engineering Department
Oklahoma State University

A-115

University Center for Water Research
Oklahoma State University
Stillwater, Oklahoma

July 1991

IMPROVED PARAMETER ESTIMATION FOR MULTIPURPOSE HYDROLOGIC MODELS

C. T. Haan
Regents Professor & Sarkeys Distinguished Professor
Agricultural Engineering Department
Oklahoma State University

A-115

University Center for Water Research
Oklahoma State University
Stillwater, Oklahoma

July 1991

Acknowledgements

The activities on which this report is based were financed in part by the Department of the Interior, U.S. Geological Survey, through the Oklahoma Water Resources Research Institute. The contents of this publication do not necessarily reflect the views and policies of the Department of the Interior, nor does mention of trade names or commercial products constitute their endorsement by the United States Government. Additional support was provided by the Oklahoma Agricultural Experiment Station as a contribution to Regional Project S-211.

Nature of the Research

Models have become the main tool used for assessing the significance of hydrologic and water quality problems and for proposing and testing solutions to these problems. These models may range in complexity from single, empirical equations to complex, continuous simulation models. Models may be used to estimate a single design quantity such as the 10-year peak flow for the design of a culvert. In this proposal, such a model is termed a single purpose model because the model provides an estimate for a single quantity such as peak flow, runoff volume, or sediment concentration. A model that simultaneously estimates several hydrologic quantities is termed a multipurpose model. A multipurpose model might be used to estimate peak flow rates, runoff volumes, and the total load and peak concentrations of water quality constituents simultaneously.

Regardless of which particular model is used, a set of model parameters must be estimated. Some models require the estimation of a single parameter, while others require the estimation of a large set of parameters. Most modelers attempt to attach physical significance to the parameters in their models. However, it is generally difficult to actually measure in the field these physical parameters. Parameters are generally initially estimated by following guidelines appropriate for the particular model and then adjusted to improve predictions for specific watersheds. The parameter adjustment phase generally relies on some type of observed data. Parameter adjustment is often termed model calibration.

Proper estimation of parameters for hydrologic models represents (i) one of the most important aspects of using hydrologic models, (ii) one of the most difficult aspects of using hydrologic models, and (iii) the aspect of using hydrologic models where misapplication is most prevalent. Users of single purpose hydrologic models have traditionally varied parameters either on a trial and error basis or through some automatic calibration procedure to minimize error sum of squares and maximize the correlation coefficient between observed and predicted results or some other similar criterion.

When multiple model outputs are involved, multiple error sum of squares and multiple correlation coefficients are available. A parameter set that is best with respect to one model output may be suboptimal, in fact very inferior, with respect to a second model output. To this point, user judgment has been the major criterion used to find the parameter set that was the best for the joint estimation of several model outputs. Many times user judgment is very satisfactory and in the case of an experienced user may be superior to any automatic calibration. The results of this research will not be a replacement for user judgment but will provide the user with a single numerical criterion to use as a basis for either manual or automatic parameter optimization in the face of multiple model outputs.

This research provides a framework for parameter estimation and evaluation for hydrologic models, including multipurpose models. This framework should help to improve user skill in parameter estimation for hydrologic models. The results of the research are directly applicable by those currently using hydrologic models in a design, regulatory, or operational mode. The results also provide model developers and evaluators with a consistent and analytic basis for parameter estimation. Thus, given a set of data and a hydrologic model, two users of the model should obtain the same parameter estimates.

Objective

The objective of the research is to develop improved procedures for estimating the parameters of hydrologic models.

Related Research

Parameter estimation and calibration pose many difficulties. They are complicated by the nonlinearity of the models, correlation among the parameters, the presence of threshold values for the parameters, irregular response surfaces with long valleys, lack of suitable estimation criteria, data errors, and the use of rainfall-runoff data for estimating certain parameters when runoff may account for generally no more than 1/3 of a catchment's response to rainfall. It is also common that the response surface may be relatively insensitive to certain model parameters over a wide range of values for these parameters.

Troutman (1985) points out problems in model parameter estimation due to data errors and the difficulty of defining an appropriate objective function to use when selecting the "best" set of parameters. Jackson and Aron (1971) discuss these same problems in a somewhat more descriptive manner.

Parameter estimation techniques may be divided into (1) a priori, (2) curve fitting, and (3) combination methods (Kuczera 1982). The a priori method relies on measurable physical characteristics to provide parameter estimates. The curve fitting procedure relies on finding a parameter set that optimizes some fitting criteria between observed and predicted results. The combination approach relies on both a priori information and curve fitting.

Curve fitting approaches include trial and error calibration (manual adjustment of parameters to obtain the best parameter set in the judgment of the modeler), least squares, method of moments, and maximum likelihood (Kuczera 1982; Troutman 1982, 1985; Sorooshian et al. 1983; Sorooshian 1983; Beck and Arnold, 1977). The most commonly used analytic procedure involves the minimization of the error sum of squares. Liou (1970) pioneered the use of such an approach in an attempt to estimate a subset of the parameters for the Stanford Watershed Model.

Kuczera (1983a) casts hydrologic models in a nonlinear regression framework for parameter estimation. After suitable transformations, using Bayesian procedures and following Box and Tiao (1973), he derives the joint posterior probability density function for the model parameters as

$$p(\underline{\gamma}|\underline{q};\lambda,K) \propto (\underline{a}^T\underline{a})^{-(n-p)/2} \tag{1}$$

where $\underline{\gamma}$ is a vector of model parameters, $\underline{q}$ is a vector of responses, λ and K are parameters of a Box and Cox (1964) transformation, n is the number of data points, p is the order of the autoregressive processes required to remove autocorrelation in the model residuals, and $\underline{a}$ is a vector of random disturbances associated with the autoregressive model. Kuczera (1983a) then takes the mode, $\underline{\gamma}_o$, of this posterior distribution (equation 1) as his estimates for the parameters. The mode is found by minimizing $\underline{a^T a}$.

Kuczera (1983b) extends his work to incorporate additional information on parameters to improve the precision of estimation of poorly defined parameters. He defines compatible data as being data such that differences between their final parameter estimates are not statistically significant. For the model used in his study, he shows that model performance in terms of runoff prediction can be improved when prior model parameter information derived from soil moisture response is incorporated into the estimation procedure.

The approach of Kuczera (1983a, 1983b) is attractive from the standpoint that it recognizes model parameters as random variables and derives their joint probability distribution (equation 1). This is congruous with a growing knowledge base on model parameter uncertainty as reflected in many works published on the topic in the last several years (Sorooshian and Dracup 1978; Wood 1978; Cooley 1983; Yeh 1986; Gupta and Sorooshian 1983; Tung and Mays 1981; Haan 1988).

Even with the advances that have been made in hydrologic model parameter specification, the problem of parameter estimation for models with multiple outputs remains. To date, the vast majority of the analytical work has been aimed at improving the prediction of a single quantity such as peak flow, monthly runoff, sediment concentration or some other water quality parameter. When predicting a single quantity, the specification of an objective function is straightforward, even though satisfying the function may produce computational difficulties.

When a model is used to simultaneously predict more than one hydrologic quantity, the specification of an objective function to use in parameter estimation becomes more difficult. It is complicated by the fact that model parameters are correlated and prediction errors for the various quantities of interest are correlated. Thus, a multivariate objective function that is a weighted sum of univariate objective functions suitably standardized does not necessarily yield good parameter estimates.

Currently, the most common procedures for estimating parameter sets for multiple output models rely on a combination of univariate optimization, manual parameter adjustment, and multiple runs with the model. Often the results are highly dependent on the experience and skill of the modeler. It is not uncommon to actually derive more than one set of "optimal" parameters and to use one set to predict one class of model outputs and other sets to predict other classes of outputs.

If the residuals are heteroscedastic and correlated, transformations must be found to produce homoscedasticity and independence. After transformation, equation (1) becomes

$$p(\underline{\theta}|\underline{y}) \propto |\underline{S}(\underline{\theta})|^{-(n-p)/2} \qquad (2)$$

where $\underline{S(\theta)}$ is analogous to $\underline{V(\theta)}$, except it is computed after applying the required transformations. Equation (2) is seen to be analogous to equation (1). The best estimate of $\underline{\gamma}$ is again taken as the mode $\underline{\gamma}_o$ of the distribution defined by equation (2) and found by minimizing $|\underline{S(\theta)}|$.

Methodology and Findings

This research was carried out in several phases with research papers and graduate student theses and dissertations prepared on each phase. A complete list of publications is contained as an appendix to this report. The study started with some investigations that demonstrated the impact that data uncertainties can have on parameter estimates and how this uncertainty is transferred directly to the hydrologic quantity or the design variable being estimated.

Parameter Uncertainty and Infiltration

Infiltration is commonly estimated using the unsaturated flow equation and a model describing the soil hydraulic properties. Average parameter values are used in the model describing the soil hydraulic properties. In this research, the uncertainty or random variability inherent in soil hydraulic data was described by fitting probability distributions to the parameters of the Van Genuchten model describing soil water characteristic data. Infiltration calculated from the unsaturated flow equation using average parameters in the Van Genuchten model was compared to average infiltration based on the probabilistic behavior of the parameters. Differences were found. A procedure for finding a single set of parameters that yield an infiltration estimate similar to the estimate obtained considering the random variability in the parameters was developed and tested. In general, this set of parameters was not equal to the average parameter values. Three soils and a total of 168 soil cores were used in the analysis.

Parameter Uncertainty and Flood Storage Requirements

A simulation approach that accounted for the random variability of the SCS (Soil Conservation Service) retention parameter, S, and 24-hour rainfall, R, was used to determine the T-year flood storage height for reservoirs on small watersheds. A sample of flood storage heights was simulated and a frequency analysis conducted to determine the T-year flood storage height. This storage was compared to the storage requirement based on the standard approach of using the T-year rainfall and the average value for S. The standard approach was found to underpredict the required storage when compared to the more realistic probabilistic approach. The degree of underprediction was found to be a function of the average curve number with smaller curve numbers having the largest error. Correction factors for the curve numbers were developed that make it possible to find the correct storage requirement based on the T-year rain and the corrected curve number for this application. This work points to the need for considering variability in all of the factors contributing to runoff, not simply rainfall probabilities, if an accurate estimate of a return period design is required.

Data Errors and Parameter Estimation

Parameter uncertainty in hydrologic models is due, in part, to random errors in input data used for calibration of the model. This work investigated the impact of various error distributions associated with input data on the final estimated parameter values using three different estimation criteria - least squares, maximum likelihood, and minimization of the sum of absolute errors.

Errors in precipitation data were found to introduce more uncertainty into parameter estimates than errors in runoff data. Parameter uncertainty increased as the level of error introduced into input data increased. Correlated errors in the input data greatly increased the uncertainty associated with parameter estimates.

Influence of the Number of Years of Data Used in Parameter Estimation

The variability in parameter estimates for a hydrologic model as a function of the number of years of data available for estimating the parameters was investigated. A 25-year record on Spavinaw Creek in Oklahoma and Arkansas was divided into 25 1-year, 12 2-year, and 5 5-year records. Each record was used for parameter estimation. Two objective functions, one based on absolute errors and one based on error sum of squares, were used.

Mean parameter values exhibited greater than expected variability, and variances in estimated parameters did not decrease as expected as the number of years used for parameter estimation increased from 1 to 5 years. One year was found to dominate estimated parameters for any record length that included that particular year. Significant differences between parameter values based on the two objective functions were found only for one parameter based on 1-year optimizations.

Multiobjective Parameter Estimation

A procedure for parameter estimation in multipurpose hydrologic models based on multiobjective programming was developed. The procedure helps overcome a problem that often occurs using traditional parameter estimation techniques--producing parameters that are good with respect to one objective but poor in terms of other objectives. The proposed method was tested using a precipitation runoff modeling system, PRMS (Leavesley, et al. 1983). The multiobjective function used had three objectives for estimating four parameters. It is shown that substantial improvement in parameter estimates can be obtained using the tested method.

Literature Citations

Beck, J.V. and K.J. Arnold. 1977. Parameter Estimation in Engineering and Science. John Wiley, Inc. New York.

Box, G.E.P. and G.C. Tiao. 1973. Bayesian Inference in Statistical Analysis. Addison-Wesley. Reading, Mass.

Box, G.E.P. and O.R. Cox. 1964. The Analysis of Transformations. J.R. Stat. Soc. Ser. B 26(2):211-254.

Cooley, R.C. 1982, 1983. Incorporation of Prior Information of Parameters into Nonlinear Regression Groundwater Flow Models. Part 1, Theory. Water Resources Research 18(4):965-976. Part 2, Application. Water Resources Research 19(3):662-676.

Gupta, V.K, and S. Sorooshian. 1983. Uniqueness and Observability of Conceptional Rainfall-Runoff Model Parameters: The Percolation Process Examined. Water Resources Research 19(1):269-276.

Haan, C.T. 1988. Parametric Uncertainty in Hydrologic Modeling. Invited paper for presentation at 1988 International Symposium on Modeling Agricultural, Forest and Rangeland Hydrology. American Society of Agricultural Engineers, Chicago, Illinois, December, 1988. Accepted for publication in Transactions ASAE.

Jackson, D.R. and G. Aron. 1971. Parameter Estimation in Hydrology: The State of the Art. Water Resources Bulletin 7(3):457-472.

Kuczera, G. 1982. On the Relationship Between the Reliability of Parameter Estimates and Hydrologic Time Series Data Used in Calibration. Water Resources Research 18(1):146-154.

Kuczera, G. 1983a. Improved Parameter Inference in Catchment Models. 1. Evaluating Parameter Uncertainty. Water Resources Research 19(5):1151-1162.

Kuczera, G. 1983b. Improved Parameter Inference in Catchment Models. 2. Combining Different Kinds of Hydrologic Data and Testing Their Compatibility. Water Resources Research 19(5):1163-1172.

Leavesley, et al. 1983. Precipitation - Runoff Modeling System: User's Manual. U.S. Geological Survey Report 83-4238. Denver, Colorado.

Liou, E.Y. 1970. OPSET: Program for Computerized Selection of Watershed Parameter Values for the Stanford Watershed Model. Research Report 34, Water Resources Institute, University of Kentucky, Lexington, Kentucky.

Sorooshian, S. and J.A. Dracup. 1978. Considerations of Stochastic Properties in Parameter Estimation of Hydrologic Rainfall-Runoff Models. Report UCLA-Eng-7852, Engineering Systems Department, University of California, Los Angles, California.

Sorooshian, S. 1983. Surface Water Hydrology: On-Line Estimation. Reviews of Geophysics and Space Physics 21(3):706-721. American Geophysical Union.

Sorooshian, S., V.K. Gupta, and J.L. Fulton. 1983. Evaluation of Maximum Likelihood Parameter Estimation Techniques for Conceptual Rainfall-Runoff Models: Influence of Calibration Data Variability and Length on Model Credibility. Water Resources Research 19(1):251-259.

Troutman, B.M. 1982. An Analysis of Input Errors in Precipitation-Runoff Models Using Regression With Errors in the Independent Variables. Water Resources Research 18(4):947-964.

Troutman, B.M. 1985. Errors and Parameter Estimation in Precipitation-Runoff Modeling. Water Resources Research 21(8):1195-1222.

Tung, Y.K. and L.W. Mays. 1981. Reducing Hydrologic Parameter Uncertainty. Journal of Water Resources Planning and Management Division, Proceedings of American Society of Civil Engineers 107(WR1):245-262.

Yeh, W.G. 1986. Review of Parameter Identification Procedures in Groundwater Hydrology: The Inverse Problem. Water Resources Research 22(2):95-108.

APPENDIX

Theses and Dissertations

Ben Salem, E. 1988. Impact of Parameter Uncertainty in Hydrologic Modeling on the Design of Small Flood Water Retarding Structures. M.S. Thesis, Library, Oklahoma State University, Stillwater, OK.

Ben Jemaa, F. 1988. Impact of Parameter Uncertainty on One-Dimensional Vertical Unsaturated Flow of Water. M.S. Thesis, Library, Oklahoma State University, Stillwater, OK.

Borah, A. 1990. Error Distributions and Parameter Estimation in Rainfall-runoff Modeling. Ph.D. Dissertation, Library, Oklahoma State University, Stillwater, OK.

Yan, J. 1990. Parameter Estimation for Multipurpose Hydrologic Models. Ph.D. Dissertation, Library, Oklahoma State University, Stillwater, OK.

Papers and Journal Articles

Allred, B. and C.T. Haan. Variability of Optimized Parameter Estimates Based on Observed Record Length. ASAE Paper 91-2076 presented at International Summer Meeting, American Society of Agricultural Engineers, Albuquerque, NM, June 23-26, 1991.

Barfield, B.J., C.T. Haan, and D. Storm. Why Model? Proceedings CREAMS-GLEAMS Symposium, Athens, GA, September 27-29, 1989.

Borah, A., and C.T. Haan. Impact of Random Data Errors on Parameter Estimation in Hydrologic Modeling. ASAE Paper 90-2018 presented at International Summer Meeting, American Society of Agricultural Engineers, Columbus, OH, June 24-27, 1990.

Borah, A. and C.T. Haan. Error Distributions and Parameter Estimation in Hydrological Modeling. ASAE paper 89-2079 presented at International Summer Meeting, American Society of Agricultural Engineers, Quebec, Canada, June 25-28, 1989.

Borah, A. and C.T. Haan. 1991. Impact of Random Data Errors on Parameter Estimation in Hydrologic Modeling. Transactions American Society of Agricultural Engineers (in press).

Haan, C.T. Parametric Uncertainty in Hydrologic Modeling. Transactions ASAE 32(1):137-146, 1989.

Haan, C.T. and B.N. Wilson. Uncertainty in Urban Hydrologic Modeling. Proceedings at 18th National Water Resources Conference. American Society of Civil Engineers, New Orleans, LA, May 1991.

Haan, C.T., J.B. Solie and B.N. Wilson. To Tell the Truth - Hydrologic Models in Court. In the Science of Water Resources: 1990 and Beyond. Proceedings of AWRA Conference, Denver, CO, November 4-9, 1990.

Haan, C.T., F. Ben Jemaa, and D.L. Nofziger. 1989. Impact of Parameter Uncertainty on Infiltration. Paper 892164 presented at Joint International Summer Meeting of the American and Canadian Societies of Agricultural Engineering, Quebec, Canada, June 25-28, 1989.

Salem, E.B. and C.T. Haan. 1991. Impact of Parameter Uncertainty on the Design of Small Flood Water Retarding Structures. Applied Engineering in Agriculture 7(1):86-90.

Salem, E.B. and C.T. Haan. Impact of Parameter Uncertainty on the Design of Small Flood Water Retarding Structures. ASAE paper 89-2053 presented at International Summer Meeting, American Society of Agricultural Engineers, Quebec, Canada, June 25-28, 1989.

Wilson, B.N. and C.T. Haan. Simplified Bayesian Procedures for Estimating Soil Erodibility. Paper submitted to Applied Engineering in Agriculture.

Wilson, B.N., C.T. Haan, W.J. Elliot, and J.M. Laflen. Bayesian Estimation of Erosion Parameters, Part II: Application. Transaction ASAE, 1991 (in press).

Wilson, B.N. and C.T. Haan. Bayesian Estimation of Erosion Parameters, Part I: Theoretical Development. Transactions ASAE, 1991 (in press).

Wilson, B.N., C.T. Haan and W.J. Elliot. Bayesian Estimation of Erosion Parameters. ASAE Paper 90-2077 presented at International Summer Meeting, American Society of Agricultural Engineers, Columbus, OH, June 25-27, 1990.

Yan, J. and C.T. Haan. 1991. Multiobjective Parameter estimation of Hydrologic Models - Multiobjective Programming. Transactions of American Society of Agricultural Engineers (in press).

Yan, J. and C.T. Haan. Parameter Estimation for Multipurpose Hydrologic Models. ASAE Paper 90-2016 presented at International Summer Meeting, American Society of Agricultural Engineers, Columbus, OH, June 24-27, 1990.