Decision-Support System for Long-Range Stream Flow Forecasting

By Michael Bender,[1] and Slobodan P. Simonovic,[2] Member, ASCE

Abstract: A sensitivity analysis of model forecasts provided possible insight into development of a hybrid technique to improve forecast quality using engineering knowledge and experience. A rule-based decision support system (DSS) is developed to demonstrate this technique for selecting long-range water-supply forecast models. The knowledge-based system is also used to document experience gained in developing time series models for the purpose of tutoring inexperienced time series modelers. The DSS allows automation of data manipulation, modeling tasks, and as a facility for coding descriptions of expert knowledge. It also acts as an aid for learning time-series concepts through modules which assist users in producing representative time series models. Use of Unix workstations for decision support is explored through the integration of various software tools with the expert system development tool Nexpert Object. Nexpert Object provides the platform for managing coded experience, as well as controlling external tools for analysis and presentation of results.

INTRODUCTION

The development of a DSS for time-series modeling is part of a case study for evaluating the effectiveness of long-range stream flow forecasting at Manitoba Hydro, a hydroelectric utility in the province of Manitoba, Canada. Fig. 1 summarizes the project. A review of current practices was made for several utilities, including Ontario Hydro (OH), Quebec Hydro (QH), Manitoba Hydro (MH), and British Columbia Hydro (BCH). The specific forecast requirements and scenario for producing long-range forecasts at Manitoba Hydro provided the basis for comparing currently used models with statistical time-series modeling approaches such as Deseasonalized AutoRegressive Moving Average (ARMA) models, and Seasonal Auto-Regressive Integrated Moving Average (SARIMA) models. Forecasts were produced for a range of possible flow conditions by a number of statistically representative models. Evaluation of these forecasts suggested a technique for improving forecasting quality through the selection of a single model for a given system state (Bender and Simonovic 1993). This differs from previous forecast approaches, which may use weighted combinations of models when there exists a set of choices (McLeod et al. 1987; Newbold and Granger 1974). Implementation of the findings within a computer-based application of the case study served to document the procedure and allow efficient use of the knowledge and experience gained by the researchers. Presentation of results using workstation tools facilitated efficient information transferral. Developing the application using expert systems technology and a decision-support development tool facilitated potential improvement of forecasts by documenting the knowledge base that governs the procedure.

[1]Res. Asst., Dept. of Civ. Engrg., Univ. of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2.

[2]Prof., Dept. of Civ. Engrg., Univ. of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2.

Note. Discussion open until June 1, 1994. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on November 2, 1992. This paper is part of the ***Journal of Computing in Civil Engineering***, Vol. 8, No. 1, January, 1994. ©ASCE, ISSN 0887-3801/94/0001-0020/$1.00 + $.15 per page. Paper No. 5029.

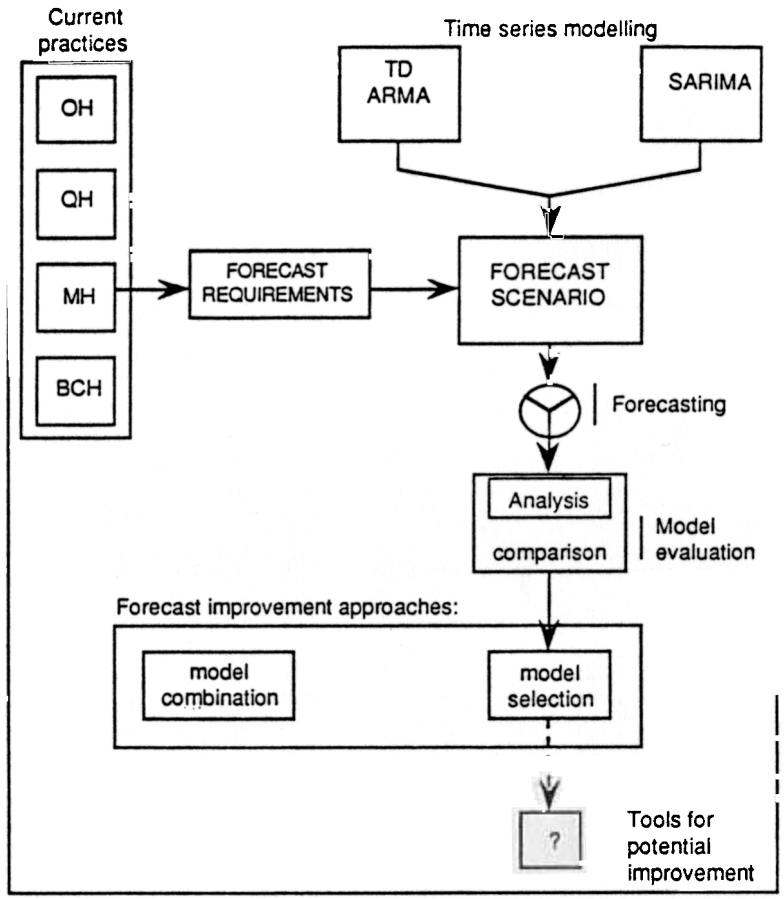

FIG. 1 Project Overview

DECISION-SUPPORT SYSTEMS

In the past, computers have been used in the field of water resources as a means of providing information to solve management problems. This use has expanded into techniques such as decision support systems (DSS), which position the user at the center of the decision-making process (Thierauf 1988; Guariso and Werthner 1989; Loucks and daCosta 1991). By amplifying the capabilities of the user and by eliminating impediments to rational functioning, DSS improves the decision-making process by placing information in the hands of the user at the proper time and place, and by providing complete flexibility in the choice and sequence of analysis, and in the presentation of results.

The approach used in the development of a DSS for long-range stream flow forecasting has the ability to interactively respond to users' ad hoc

requests in "real time." The dynamic forecasting environment compels the user to ask for information that often cannot be supplied in advance. The user has to be reactive, or proactive, by making short-term and long-term decisions based on the information and tools at hand.

Because judgment and insight are critical in forecasting, the DSS is designed to support different user needs, from model search and identification to producing forecasts. To make this DSS an effective tool, two main processes are addressed in the development: (1) Problem representation; and (2) problem solving. Problem representation involves searching for available mathematical forms for forecasting process representation. All the opportunities related to these representations can be identified and implemented. In addition, the DSS centers on problem solving—identifying current forecasting needs and responding to them.

An integral part of problem representation is the "perception" principle. This principle states that a forecast must be able to perceive future trends that have impacts on the operation of water resource projects before they actually occur. The great value of the DSS is that users can perform "what if" sensitivity analyses, so they can try different decisions without having to deal with the consequences. In this way a DSS for long-term stream flow forecasting can guide the user through most optimistic, pessimistic, or expected forecasts.

A characteristic property that makes this DSS approach original is an emphasis on the use of expert-system technology, which has attracted the attention of many analysts in the water resources field. Expert, or knowledge-based, systems have two main components: the knowledge base itself, and a method, the inference engine, for applying that knowledge to the solution of a specific problem. Combining a knowledge-based system with an interactive decision support system provides what is called an Intelligent decision support system (IDSS) (Simonovic and Savic 1989). The emphasis is still on support, however, whether a system is "intelligent" or not.

In general, several roles are envisioned for expert systems in long-range flow forecasting. An expert system can be developed to serve as a stand-alone tool. It may also act as an alternative form of documentation in place of textbooks or manuals. The form accepted in the presented research is as an interface between a complicated computer modeling process and a less experienced user. Through the introduction of an expert system component, a DSS role has been expanded from a tool for interactive forecast analysis to a vehicle for communication, experimentation, training, and learning.

DECISION-SUPPORT DOMAIN

Long-range forecasting for reservoir management using time-series modeling techniques can define two general forms of model structuring. Both correlate previous time lags, expressed as a combination of AutoRegressive (AR) and moving average (MA) parameters (Box and Jenkins 1970). One form of time series modeling uses deseasonalization techniques to produce stationary data series. The resulting model is an ARMA model. SARIMA models are the other form of time series modeling, either in multiplicative or nonmultiplicative form [(1)]. For a detailed discussion of these modeling techniques [see Abraham and Ledolter (1983), Bowerman and O'Connell (1987), and or Salas (1980)].

$$\text{SARIMA } (p, d, q) \times (P, D, Q)_L \quad \text{or } \phi(B)\Phi(B^L)(1 - B)^d(1 - B^L)^D x_t = \theta(B)\Theta(B^L)a_t \quad \ldots\ldots (1)$$

Examination of the autocorrelation function (ACF) and partial autocorrelation function (PACF) provides a thorough basis for analysis of the system behavior under time dependence, and will suggest the appropriate parameters to include in the model. The identification of a stationary series requires consideration of the extent in which the correlation functions are significant. This can become difficult for a data set with complicated seasonal patterns or trends. Experience in modeling time-series data, and familiarity with identifying stationary processes, is a great asset.

Models adequacy can be evaluated by three tests. A t-test is used to check the significance of individual parameters. A parameter may be removed from the model structure if it is statistically insignificant, and the remaining parameters should be recalculated. A chi-square test is used to evaluate the validity of the model in general. The probability values associated with the residual autocorrelation chi-square test should be greater than 5%. If one or more of the probabilities are not, yet the model passes the other tests, the model structure may be completely inadequate and a new differencing scheme should be tested. Finally, residual ACF and PACF values can be checked for additional significant lags. Review of these correlation functions will give an indication of whether all of the appropriate model parameters have been included.

More than one differencing scheme may produce a reasonably stationary series, so more than one acceptable SARIMA model structure may be found. These models may be immediately compared by observing any differences in chi-square probabilities. If one model exhibits considerably greater probabilities, it may be chosen above the others. Backcasts, values already used to define the model, can also be made and residuals compared. There are occasions, though, where one model may not be recognizably superior to other models.

When there is a set of possible models to produce forecasts for stream flow series, some type of combination may reduce forecast errors, but there may be situations where a single model is superior, or more accurate, than any other model structure or forecast combination. This research investigated model performances under a variety of flow conditions to suggest model choices that are dependent on current system conditions. Comparison of forecast errors and visual observations of forecast series of several available models, suggested a system of model ranks for each of the flow scenarios to generate simple guidelines for selecting a specific model. A comprehensive forecast comparison plan was envisioned after acknowledging that any given model structure represents a certain portion of the system. Depending on the focus of the system needs, and system characteristics, model performance may be sensitive to the forecast scenario. Since flow data was supplied for several Manitoba river basins, the forecast scenario remained consistent with Manitoba Hydro system needs.

A sensitivity analysis produces a large number of forecasts to compare, allowing evaluation of the sensitivity of the models to various flow conditions, and providing an in-depth understanding of model performance. The mean squared error (MSE) is a statistical parameter that provides a convenient measure of performance in the same units and magnitude as variance. Model selection is based on minimum MSE, and observations on general forecast curve characteristics. Visual examination of actual forecasts is beneficial in making observations on model behavior concerning the appropriateness of the model behavior characteristics, or ability to handle large system fluxes such as spring runoff. Special attention is focused on

tendencies to overestimate or underestimate the flow. If it is possible to determine which model produces the best forecast under a general array of flow possibilities, then that model should be chosen.

Intricate observations on performance are often made as model behavior may be subtle. An understanding of the consequences of strengths, and especially weaknesses, of model choices is vital to the selection of a model. Spurious observations in the real-time operational demand for forecasts may be costly. This risk can be minimized by documenting the knowledge of an expert within a decision support system that acts to confirm the expert, or to act on the behalf of the expert.

Generation of rules for context-sensitive model selection uses knowledge of modeling techniques and engineering experience to automatically produce optimal forecasts for previously studied river basins. The potential of this forecast tool is to remove inconsistencies in forecasting decision making, reduce time requirements for producing updated forecasts, and aid both experienced engineers and novice users in clarifying important decision issues in forecasting.

DESCRIPTION OF DECISION-SUPPORT SYSTEM

The model selection process, and observations concerning the quality of forecasts, suggested the use of a set of if-then rules to document the knowledge. Nexpert Object provided a straightforward means by which the rules could be built, and the inference engine to execute the DSS. Nexpert Object provides the means by which rules can be written for: (1) Control of the analysis process; (2) execution of external programs and packages to perform various tasks; and (3) control of the display of text and the form of additional queries to the user. These properties are used to arrange the system of rule premises, and concluding actions within the long-range stream flow forecasting DSS: (1) To control access to the knowledge base; (2) to assign properties describing the system; (3) to govern the selection of a model; and (4) to check statistical tests for the model.

A Unix workstation, SUN Sparc Station 1+, was chosen to develop the system. It enabled the simultaneous execution of multiple programs and graphical displays. Networking between computers provided the ability to transfer data or knowledge to and from numerous sources. The Unix operating system provided a powerful and flexible environment.

The means by which users view the application is an important aspect of development. Information and knowledge must be passed in a concise form that promotes user comprehension of the related issues while improving efficiency by automating many of the menial or straightforward tasks. Numerous textual displays are required to describe necessary guidelines and modeling options. Automation reduces the necessity of a novice user to fully comprehend the details of time-series modeling. The emphasis of system advice is focused on questions such as "why?" or "what if . . .?" User focus may then be maintained for decision making on much higher level problems such as general system behavior and model structure.

Nexpert Object controls the display from a session control window that asks questions, and provides answer options. Flexible text windowing features supplied by Nexpert, control the display of messages and advice to the user. Development time for the DSS using Nexpert Object was low because the developer does not require extensive programming abilities, merely the ability to describe the system structure and various pieces of knowledge within the modules. Execution of the decision-support tool for

testing or establishing the validity of a hypothesis or system state is handled entirely by Nexpert Object once the structure and rules are in place.

The scope of the system demands the utilization of numerous specialized software and programming tools. Integration of these tools in an efficient, and transparent, manner will not only prevent cluttering of the application but actually enhance the aesthetic value and tutorial potential of the DSS. Tools used to develop the system include the statistical analysis system (SAS), Xgraph, Unix scripts, ASCII text files, and FORTRAN programs. SAS is the primary tool used to perform statistical analysis of the data including calculation of correlation functions, estimation of model parameters, testing model adequacy, and producing forecasts. Xgraph is a plotting tool that graphs a specified set of data within the *X*-window graphical environment. It is useful in plotting forecasts or historical series of data. Unix scripts are analogous to DOS batch files. A script is a text file consisting of a series of operating system commands. Several of these files are used to control the formation of input files and execution of both SAS and Xgraph applications. Numerous FORTRAN programs control data manipulation such as transformation, deseasonalization, and transfer of data to formatted input for SAS, or form of output from SAS. Arrangement of these tools is shown in Fig. 1, where the small black box represents the entire DSS where the various components may be invisible to the user. Although Nexpert Object is used to control the application, unix scripts are also used to control timely execution of the other development tools, or execution tools. These control tools and execution tools are the significant types of components that operate together to provide the user with the system task option to produce forecasts (forecasting), or develop a new model (model development).

The DSS developed for this case study contains two system modes or tasks (Fig. 2). The first, model development, assists in the initial development of time series models. Advice is given to help guide decisions by the user, and the DSS automates many of the necessary functions including data manipulation, and execution of the statistical software. The model development mode provides the user with relevant knowledge and instruction to produce an adequate model using one of a number of time series

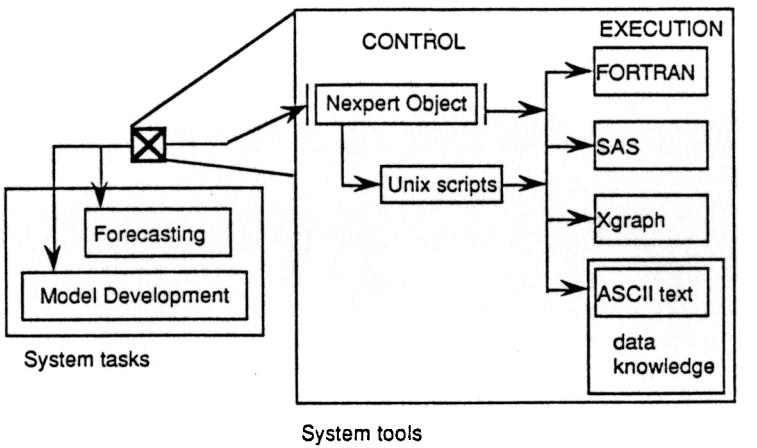

FIG. 2. DSA Tasks and Tools

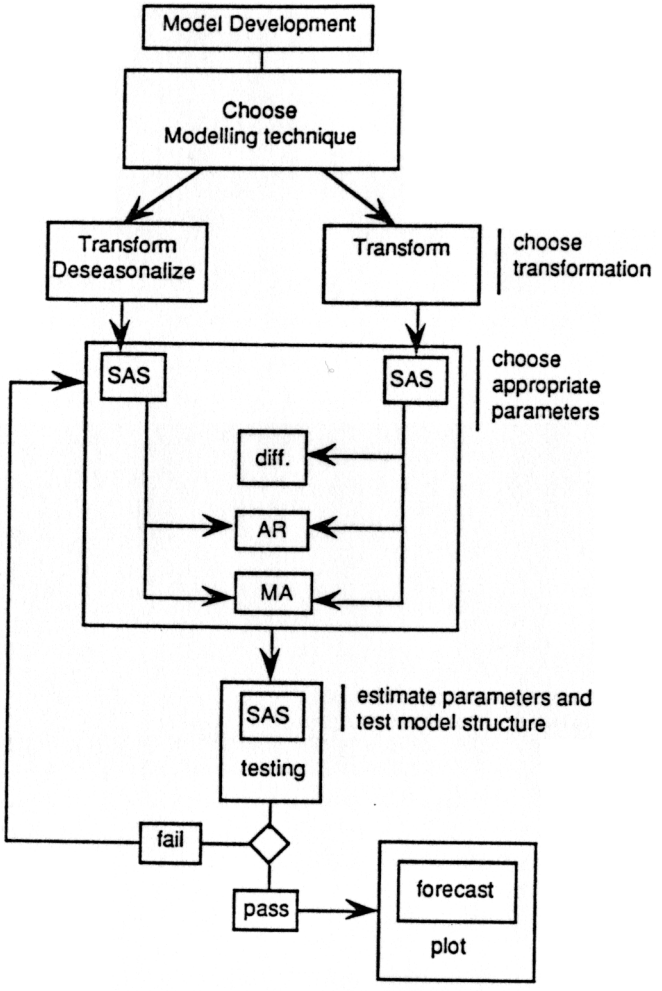

FIG. 3. Model Development User Mode

analysis procedures (Fig. 3). Advice instructs the user in choosing an appropriate modeling technique and type of data transformation or manipulation. The application automates all aspects of data transformation, preparation of input files for SAS, execution of SAS, and transformation of SAS output for display to the user. The user is directed to choose appropriate parameters for the model. Automation allows efficient, consistent testing of the model structure. This allows the user to produce advanced time series models and forecasts of data without having to be experienced in this field of study.

The second user mode, forecasting, enables the user to produce forecasts based on recent flow behavior, using time-series model structures that have already been developed and tested. Model-selection rules are accessed to choose the most appropriate available model for forecasting. Forecasts are automatically produced for the chosen model. Fig. 4 represents the process

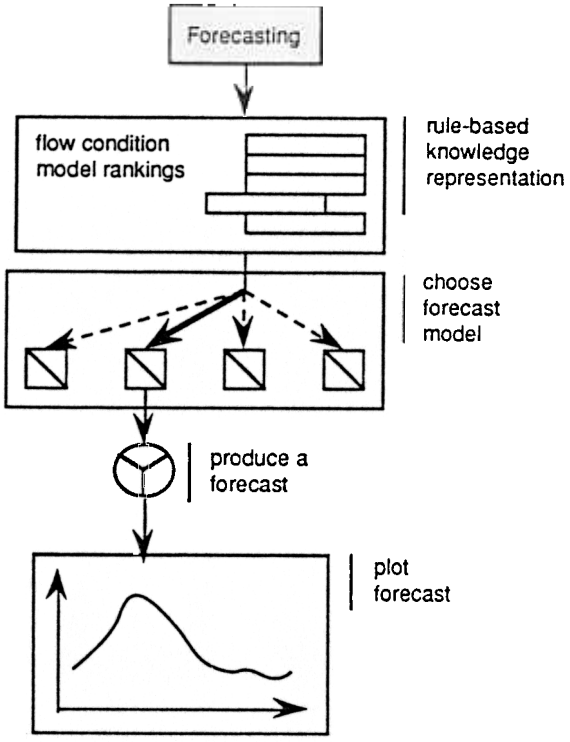

FIG. 4. Forecasting User Mode

TABLE 1. DSS Goals

Hypothesis (1)	Description (2)
(*a*) Forecasting	
Set_model	Find most appropriate model to produce forecasts
Get_flow	Get description of current flow conditions
(*b*) Model Development	
Make_model	Execute model development procedure
Do_first	Identify modeling procedure and transformation type
Do_second	Choose modeling structure and estimate parameters
Do_third	Produce forecast
Reject_model	Test model: find flaws, suggest procedure to rehabilitate model

of producing a forecast within the application. The user is queried for information regarding the present system state. Rule-based knowledge accesses the model rankings for the specified conditions. Producing and plotting of forecasts is also automated.

Consultation of the time-series DSS, as previously described, directs the

TABLE 2. DSS Input Information: Forecasting

Forecasting	
Parameters (1)	Options (2)
Set	Numerous data sets
—	—
Flow	Low
—	Median
—	High
—	Uncertain
Peak	Low
—	Median
—	High
—	Uncertain
—	—
—	—
—	—
—	—
—	—
—	—

TABLE 3. DSS Input Information: Model Development

Model Development	
Parameters (1)	Options (2)
Procedure	Seasonal
—	Deseasonalized
Transform	Log
—	None
—	—
—	—
P_lags/q_lags/var	Options dependent on procedure
—	—
—	—
—	Transformation
Parameters	Significant
—	Insignificant
Chi_square	Adequate
—	Inadequate
Residuals	Significant
—	Insignificant

user to produce either a forecast using the performance rules that govern model choice, or develop a representative model. In order to produce monthly forecasts for a specified stream, the rules corresponding to generalizations from the performance sensitivity analysis must be available within the knowledge base. If they are not available, the best the DSS can suggest is a model structure based on general river basin characteristics. Performance rules are contingent on the expert performing the sensitivity analysis with the best

available time-series models, drawing conclusions, and updating the knowledge base. Adding groups of performance rules is straightforward with Nexpert Object. From a window environment, the premise and conditions for each rule are entered. Nexpert will attempt to execute one of the rules whenever the user selects a river basin for forecasting. When the available model structures are specified in the knowledge base, execution of SAS will estimate parameters based on the most recent update of flow records. Fig. 4 represents the process of accessing rules describing performance rankings and choosing a particular model structure. The dashed lines represent paths to possible model structures while the path to the chosen model is shown by the bold arrow. Table 1 describes the hypotheses used to suggest or control the decision-making process for producing forecasts. Tables 2 and 3 provides a short summary of variables used to describe the necessary knowledge. Automation of the time series forecasting procedures and model selection rules results in plots of the desired monthly forecasts against historical monthly averages.

The conditions for passing the model selection rules will generate a series of questions for the user whenever required information is not available. Specifically, these questions are related to the present state of the system, in qualitative terms. Fig. 5 shows interaction with a historical data set when the user is uncertain of the system state. Visual inspection determines whether flow tendencies are high, median, or low. Statistical analysis of the data will provide a reasonable, or concrete, assessment of system state but automation of interpretation of the system state may remove benefits derived from user intuition and experience with the system. The user's ability to grasp system tendencies and effects from various sources is the advantage of allowing user input.

Development of model structures by the user poses entirely different problems for DSS development. In this case, the DSS acts to improve model development efficiency for an experienced modeler by automating many of the menial calculations, manipulations, and commands. However, the DSS also acts as a tutor for inexperienced modelers. Advice is available throughout the consultation, and must be thorough and concise. Messages and advice is provided through the use of text windows which remain visible for consultation, or can be discarded by the user.

Advice is required for a novice user when choosing the most appropriate modeling technique (see Fig. 3). One of the observations from work on this project is that ARMA models are better suited in Canadian conditions for river basins with a relatively large amount of upstream storage such as a series of small lakes that act to filter out large stream flow fluctuation. SARIMA models tended to outperform ARMA models for streams that exhibit a great dependency on precipitation events where large fluctuations are possible. Appropriate data manipulation must also be chosen. For ARMA models, choices are restricted by the decision to deseasonalize the data.

Knowledge, or advice, is required to specify the proper structure for the model. This has traditionally been a trial and error process, but expert advice by the DSS can reduce the development time. At this stage, the DSS acts as liaison and interpreter for SAS (Fig. 6). SAS input files are created, SAS is executed, and output is graphically formatted for the user. Advice is provided to explain model requirements, correlation functions, and tests being used to evaluate model adequacy. However, the most important advice is directed toward choosing a model structure that will likely represent the behavior of the flow series. Fig. 6 shows SAS output on the lower right.

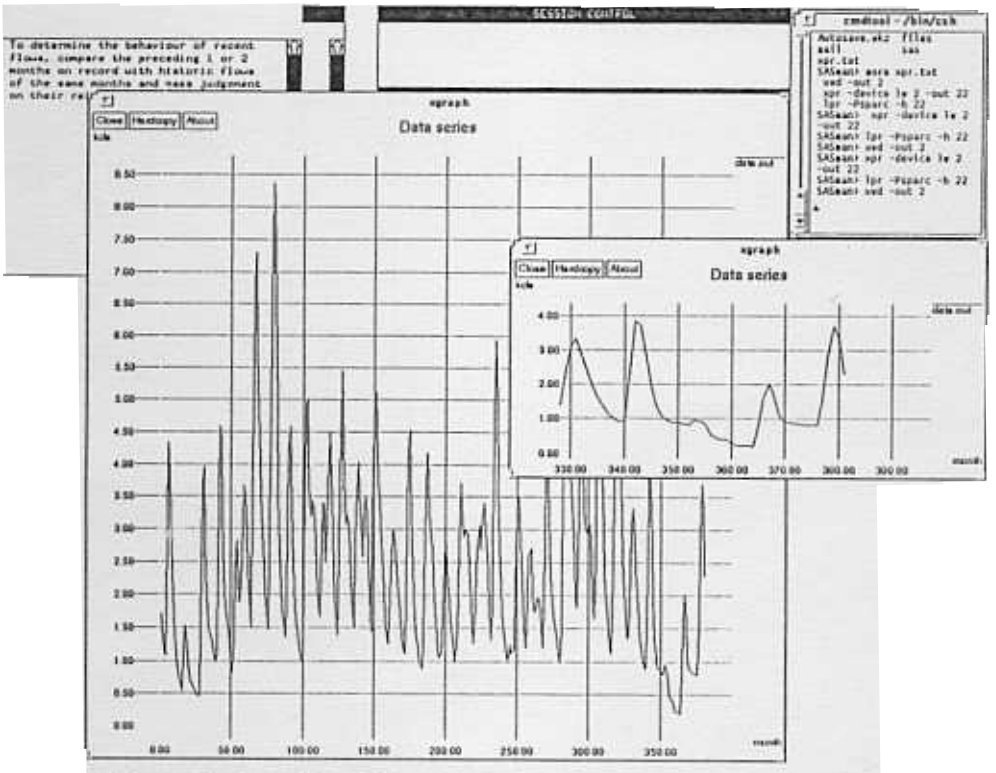

FIG. 5. Display of Data Series for Qualitative Description of System State

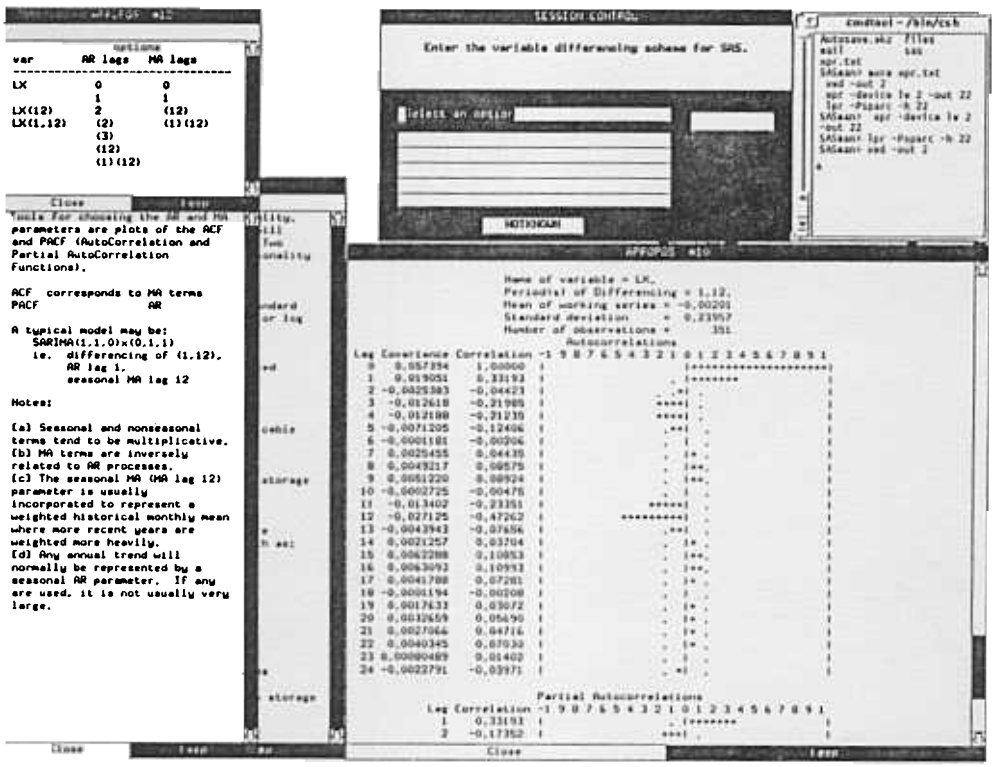

FIG. 6. Construction of Model Structures

Recommendations and advice are given on the right. This screen display attempts to establish a differencing scheme, entered in the session control window at the top of the screen, to use for a SARIMA model structure. The DSS will suggest possible structures and instruct the user in the interpretation of the ACF and PACF. Most importantly, the DSS will restrict the user according to the chosen modeling technique. For example, TD ARMA models do not require differencing because deseasonalization was already performed to ensure a stationary correlation structure. Suggested AR and MA parameters will not include seasonal AR or seasonal MA processes since these seasonal properties are usually drowned out by deseasonalization.

Parameters are estimated through SAS, and the model structure is tested by the three model tests previously described. If the model structure is inadequate, the user is returned to the point in the parameter selection process where changes are most likely needed. For example, if the t-test suggests a particular parameter is insignificant, the user will need to define the model without that parameter. If the chi-square test fails, but the t-test is satisfactory, and no additional significant lags are apparent from examination of the residual ACF and PACF, the user may be required to select another differencing scheme or transformation altogether. In other words, the user is looped back only as far as necessary to save repetition of satisfactory steps in the model development process. Table 1 describes the hypotheses used to suggest or control the decision-making process for producing adequate model structures. Table 3 summarizes parameters used to describe knowledge and input information for model development. When a satisfactory model is found, the user can produce a plot of forecasts.

Through careful guidance, and concise advice, a novice user will be able to produce an adequate model. For experienced time-series modelers, development time may be greatly reduced by automation of many of the steps in the modeling process. The DSS exhibits a great deal of power in being flexible enough to handle both types of users without being cumbersome for users with either extreme of experience and knowledge.

CONCLUSIONS

An application of decision-support techniques has been demonstrated for water-supply forecasting. Experience in modeling time series for long-range forecasting has been acquired and preserved within a DSS. A rule-based knowledge system has been used to represent knowledge governing the selection of a suitable forecast model for improvement of forecasting accuracy, based on observations of model performance under various flow conditions.

The DSS presented in this paper has been built within Nexpert Object, a knowledge system development tool. It allows future expansion of advice, messages, and management of data and forecast models. The database can be expanded to accommodate any number of data sets and available models. The statistical tools used to build the SARIMA models can also be used to build advanced variations such as transfer function noise (TFN) models by cross-correlating multiple data sets. Such models may reveal more information concerning trends in the data.

Use of flexible modeling techniques, guided by observations of model performance under varied flow conditions, may present users with potential forecast improvements. Many large hydroelectric utilities demand forecasts of natural inflows for long-range planning of reservoir releases. Intelligent automation of model selection, forecasting, or model development may streamline forecasting operations. Benefits include improved time management for modelers and forecasters, improved consistency of forecasting procedures, and improved efficiency in planning long-range reservoir releases.

ACKNOWLEDGMENTS

This research project was made possible through a graduate scholarship from the research and development committee at Manitoba Hydro, and in part by an operating grant from the Canadian National Science and Engineering Research Council (NSERC). We would also like to thank Dave Cormie at Manitoba Hydro for his hours of consultation time.

APPENDIX I. REFERENCES

Abraham, B., and Ledolter, J. (1983). *Statistical methods for forecasting*. Wiley and Sons, New York, N.Y.

Bender, M. J., and Simonovic, S. P. (1993). "Time series modelling for long-range stream flow forecasting." *J. Water Res. Plng. and Mgmt.*, ASCE.

Bowerman, B. L., and O'Connell, R. T. (1987). *Time series forecasting*. Duxbury Press, Boston, Mass.

Box, G. E. P., and Jenkins, G. (1970). *Time series analysis, forecasting and control*. Holden-Day, San Francisco, Calif.

Guariso, G., and Werthner, H. (1989). *Environmental decision support systems*. Halsted Press, London, United Kingdom.

Introduction manual, reference manual, user's guide. (1991). Neuron Data Inc., Palo Alto, Calif.

Loucks, D. P., and da Costa, J. R. (1991). *Decision support systems: water resources planning*. Springer-Verlag, New York, N.Y.

McLeod, A. I., Noakes, D. J., Hipel, K. W., and Thompstone, R. M. (1987). "Combining hydrologic forecasts." *J. Water Res. Plng. and Mgmt.*, ASCE, 113(1), 29–41.

Newbold, P., and Granger, C. W. J. (1974). "Experience with forecasting univariate time series and the combination of forecasts." *J. R. Statistics Society*, 137, 131–165.

Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L. (1980). *Applied modelling of hydrologic time series*. Water Resources Publications, Littleton, Colo.

Simonovic, S. P., and Savic, D. A. (1989). "Intelligent decision support and reservoir management and operations." *J. Comp. in Civ. Engrg.*, ASCE, 3(4), 367–385.

Thierauf, R. J. (1988). *User-oriented decision support systems: Accent on problem finding*. Prentice Hall, N.J.

APPENDIX II. NOTATION

The following symbols are used in this paper:

AR = autoregressive time-series expression;
$ARMA$ = autoregressive moving average time-series model;
a_t = white-noise series;
B = backshift operator;
d, D = degree of differencing (nonseasonal, and seasonal)
L = seasonal lag;
MA = moving average time-series expression;
p, P = number of AR terms (nonseasonal, and seasonal);
q, Q = number of MA terms (nonseasonal, and seasonal);

SARIM seasonal autoregressi tegrated age mode
structure
position time
dat ries
ϕ coefficien and
MA coefficien