RESULTS FROM INTERNATIONAL INTERCOMPARISONS OF SNOWMELT RUNOFF MODEL PERFORMANCE

Albert Rango

USDA-ARS Hydrology Laboratory
Beltsville, Maryland

Jaroslav Martinec

Federal Institute for Snow and Avalanche Research
Davos, Switzerland

ABSTRACT

The World Meteorological Organization has sponsored several projects to intercompare hydrological models in forecast situations. To facilitate intercomparison, numerical performance evaluation criteria have been used. A combination of a limited number of these criteria has been found useful for intercomparing model results. The use of updating has also been found useful for improving simulation accuracies. Simple updating seems to be best, especially if done in an objective way. A project that more realistically simulates operational forecasts with snowmelt-runoff models is now needed.

INTRODUCTION

Because snowmelt has been recognized as a significant hydrologic process in certain regions of the world, many snowmelt runoff models have been developed to simulate the flow produced by a melting snowpack. As a result of the increase in the number of snowmelt runoff models available, there is a need to be able to intercompare the performance of various models in different snowpack conditions and regions. Such a means for comparing model performance would be especially valuable to operational users who may be trying to choose an appropriate model for use in their particular snow basin.

The World Meteorological Organization (WMO) recognized the need for providing documentation on various snowmelt runoff models and for testing the models on common data sets. WMO sponsored a project entitled the "Intercomparison of Models of Snowmelt Runoff" and also provided documentation on the results (World Meteorological Organization, 1986). This intercomparison project included ten separate snowmelt runoff models and six common data sets. Table 1 lists the ten models used in the project along with their countries and the basins each model used in the testing.

WMO PROJECT DESIGN

Each data set listed in Table 1 was distributed to the model operators in two sections. A six year calibration period was defined in which all the input and output data were made available. This period was particularly for the so-called calibration models to define their model parameters. Of the ten models, only one model (SRM) was not a calibration model and did not require calibration data. The second section was a four year verification period. Only input data were supplied for this time and all output (streamflow) data were withheld from the model operators. Simulations of snowmelt runoff were produced for both the calibration period and the verifiction period. WMO verified the simulations produced by the modelers in the second period (four years).

Table 1. Models participating in WMO project, model countries, and data sets used by each model.

Model name (and abbreviation)	Country	River basin data sets tested in the project: Durance River (France) 2170 km^2	W3-Watershed (USA) 8.7 km^2	Dunajec River (Poland) 680 km^2	Dischma Basin (Switzerland) 43.3 km^2	Illecillewaet River (Canada) 1155 km^2	Kultsjon (Sweden) 1110 km^2
CEQUEAU Model (CEQUEAU)	Canada	X	X	X	X	X	X
Empirical-Regressive Model (ERM)	Czechoslovakia	X		X	X	X	X
HBV-Model (HBV)	Sweden	X	X	X	X	X	X
Rainfall-Runoff Model, Version II (NAM-II)	Denmark	X	X	X	X		
National Weather Service River Forecast System Model (NWSRFS)	USA	X	X	X			
Precipitation-Runoff Modelling System (PRMS)	USA	X	X	X	X		
Snowmelt Runoff Model (SRM)	Switzerland USA	X	X	X	X		
Streamflow Synthesis and Reservoir Regulation Model (SSARR)	USA	X	X	X	X	X	X
Tank Model with Snow Model (TANK)	Japan	X	X	X	X	X	X
U.B.C. Watershed Model (UBC)	Canada	X	X	X	X	X	X

One major advantage of this intercomparison project was that WMO applied quantitative performance evaluation criteria to each model simulations. Typically, when model results are presented in the literature or at presentations, no numerical criteria are presented to assist the audience in evaluating how well a model performed. A major disadvantage of the project was that too many criteria were available (11) and WMO (1986) did not assist readers by selectively comparing criteria and models.

SELECTIVE COMBINATION OF EVALUATION CRITERIA

In this paper we advocate the use of some sort of performance evaluation whenever model results are presented, preferably numerical criteria similar to those presented in the WMO (1986) document. Our recommendation is to use only a selected few criteria in order to avoid confusing a user. Many times the use of one statistical criterion will be sufficient for a specific purpose. At other times the use of multiple criteria may be necessary. As an example of multiple numerical criteria, Figure 1 shows the plotting of the deviation of the runoff volumes criterion, D_V, against the so-called Nash-Sutcliffe coefficient, R^2 (Nash and Sutcliffe, 1970). The values are plotted for all models for both the snowmelt season and the entire hydrological year. The letters are used so that we can refer to models anonomously and are not related to the list given in Table 1.

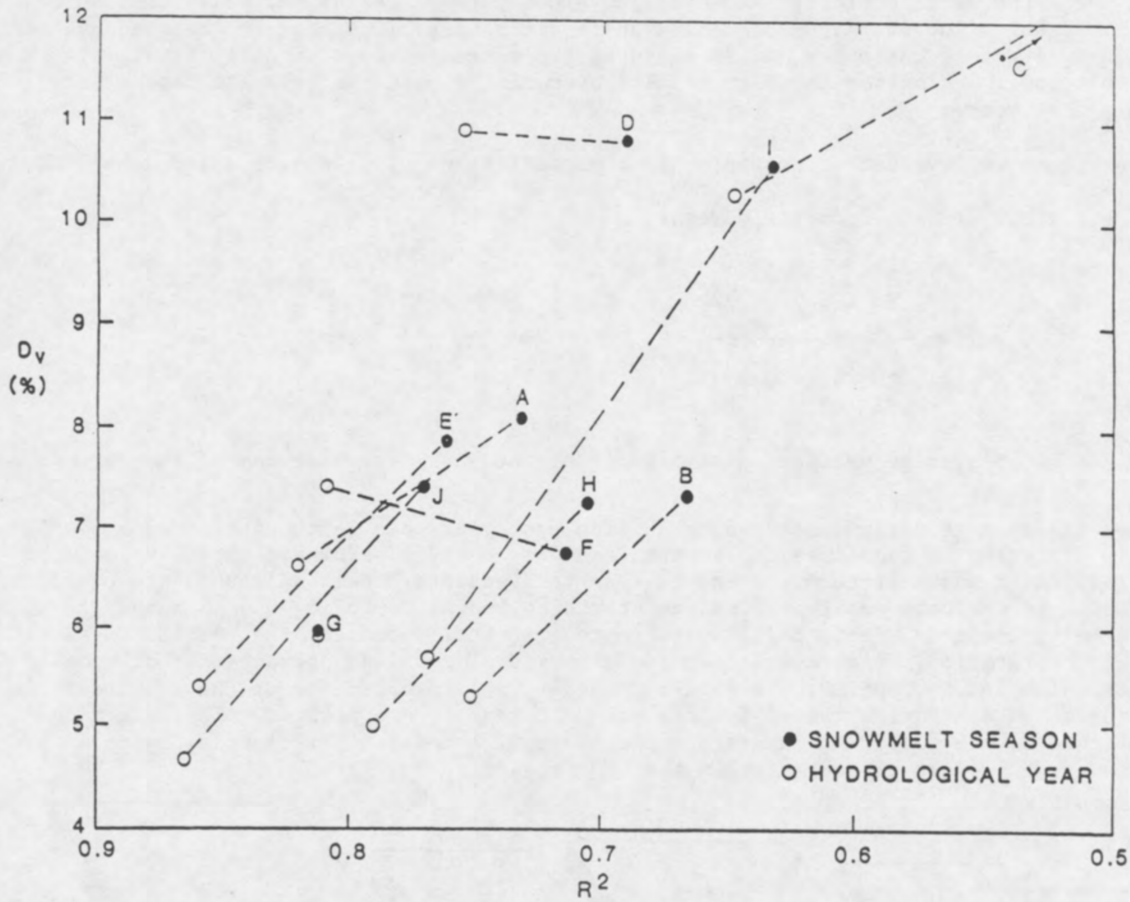

Figure 1. R^2 plotted against D_V for comparison of average results for hydrological years and snowmelt seasons for each model.

The values D_V and R^2 can be defined as follows:

1. Deviation of the Runoff Volumes, D_V

$$D_V(\%) = \frac{V - V'}{V} \cdot 100$$

V = measured runoff volume

V' = computed runoff volume

2. Nash-Sutcliffe coefficient, R^2

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Q_i - Q'_i)^2}{\sum_{i=1}^{n} (Q_i - \overline{Q})^2}$$

Q_i = measured daily discharge

Q'_i = computed daily discharge

$\overline{Q}$ = average measured discharge (for the time period in question)

In Figure 1 the results plotted are average values for all basins and years tested for each model. Results improve as D_V decreases and R^2 increases. D_V measures the goodness of fit for seasonal volume estimates and R^2 measures the correspondence of daily flows. It is apparent that model G provides the best results overall for both the snowmelt season and the hydrological year.

Another important evaluation criterion is the coefficient of gain from daily means, DG.

3. Coefficient of Gain from Daily Means, DG

$$DG = 1 - \frac{\sum_{i=1}^{n} (Q_i - Q'_i)^2}{\sum_{i=1}^{n} (Q_i - \overline{Q}_i)^2}$$

$\overline{Q}_i$ = average measured discharge from past years for each day of the period.

Because the average daily discharge taken from past years can be thought of as a simple model, this criterion is sometimes called the Peasants Model. The higher the DG value, the better a particular snowmelt-runoff model is than the Peasants Model. Although the use of three criteria to evaluate model performance starts to become confusing when a number of models are being compared, it is still possible to graphically present the results to assist in easier interpretation. Figure 2 is a three-dimensional combined presentation of model performance using R^2, D_V, and DG. In Figure 2 the average inaccuracies of the ten models in the WMO project on all basins tested for the snowmelt seasons are plotted. The larger the prism in Figure 2, the poorer the average performance of a model. For example, model C seems to have the largest inaccuracies in the WMO project.

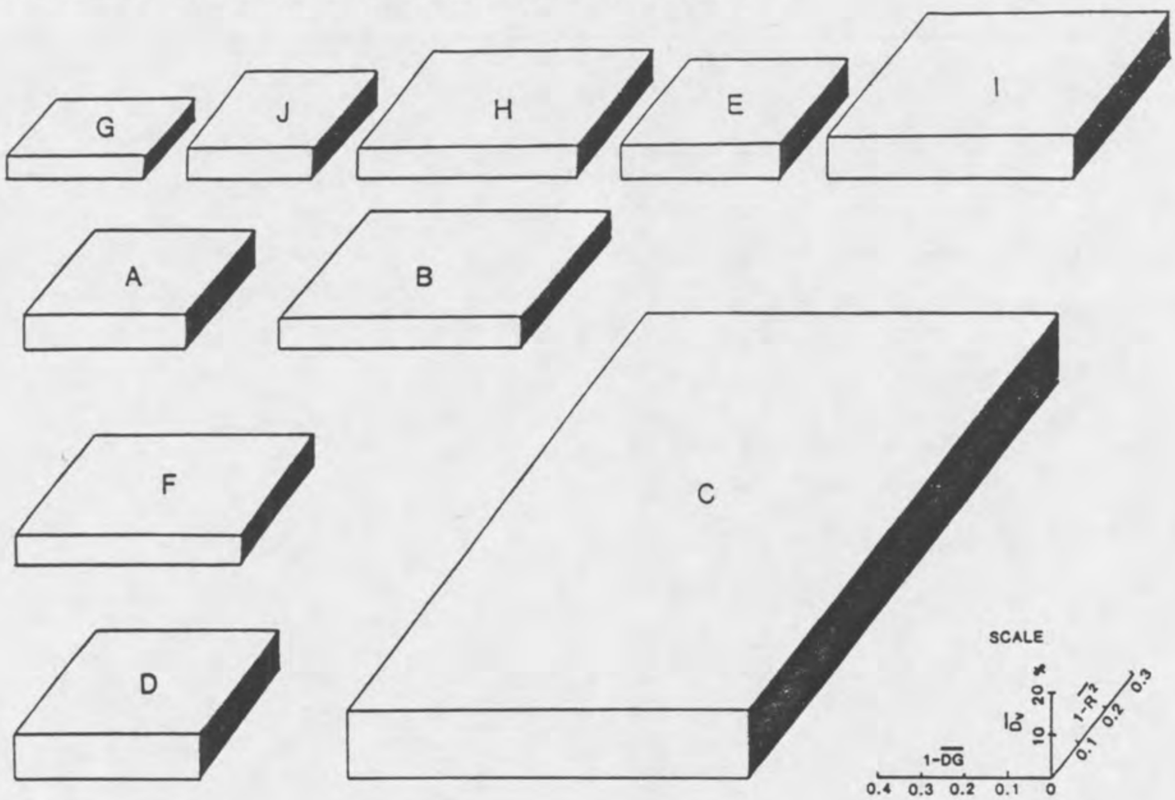

Figure 2. Combined representation of model performance using three criteria: D_V, R^2, and DG. The volumes of the prisms indicate the average inaccuracies of the various models from all results for snowmelt seasons reported in the WMO project.

It is doubtful that there will be many situations where more than three criteria will be necessary for evaluating model performance. When such evaluations are needed, the model owner or user should always use numerical criteria but strive to selectively limit the number of criteria used.

SIMULATED CONDITIONS OF OPERATIONAL RUNOFF FORECASTS

One problem with the WMO intercomparison project was that it was not a true operational forecasting situation where day-to-day modifications could be made in the forecast as new information became available. As a result, WMO has attempted to create a real-time forecasting situation as a follow-up to the intercomparison project. This follow-up project is entitled "Simulated Real-Time Intercomparison of Hydrological Models." In this project all model owners assembled with their models at a common test site. Calibration data had been released earlier for preliminary testing and optimization of model parameters. In contrast to the earlier WMO project, no verification period data were released to the modelers, and the forecast periods were more numerous and shorter than before. The input data for the verification period were released to the modelers just before they were to make their forecast in order to approximate a real-time situation. In this project the input data are the actual observed data for the forecast period.

As in the prior project, all the input data were perfect so that no forecasting of input variables was necessary. The only differences then between this project and the earlier one were that the modelers did not get input data to process or examine until the operational phase of the project was to begin, the input data were only for short time periods corresponding to short forecast periods (although all data were eventually released), and updates could be performed according to each individual model's updating methods. The fact that all input data were known perfectly meant that this project was not a true simulation of operational conditions. The forecasting of input variables can be a major task with a significant effect on model results. The capability for updating, however, is realistic and should produce better results than the earlier project where updating was not possible. The results of the project are currently being compiled and will be published in the future.

SRM UPDATING RESULTS

To illustrate the potential of updating, SRM was implemented on one of the basins used in both projects, namely, the Illecillewaet River basin in Canada. Figure 3 is an outline of the Illecillewaet basin showing the elevation range and division into four elevation zones.
These elevation zones are the basin subunits used for mapping of snow covered area using satellite data.

Figure 4 is an example of snow cover depletion curves for each of the four elevation zones for 1984 in the Illecillawaet basin. The depletion curves were derived from Landsat MSS data. Individual daily values were extracted from these curves and used as input to SRM in the WMO simulated real-time project.

SRM has an option that performs only a very simple form of automatic updating, namely, every 7th day updating of the computed flow with the actual measured flow. The frequency of updating is arbitrary and could be changed to match forecast intervals, e.g., 3-5 days. The first time the updating option for SRM was used was in the second WMO project because simulations were requested for the entire snowmelt season with no updating and normal model updating. Figure 5 shows the 1984 nonupdated hydrograph using snow cover taken from Figure 4 and temperature and precipitation input data from WMO. Figure 6 shows the same year 1984, but the hydrograph has been updated every 7 days with measured flow values.

It is obvious that, even with this simple updating algorithm, the use of updating in simulations or forecasts can have positive benefits for improving the correspondence between computed and measured flow. It is recommended that updating be used when in the forecast situation. The updating, however, should be simple, objective, and easy to duplicate by different users. More research and documentation is needed on the various kinds of updating algorithms available and how they should be applied effectively.

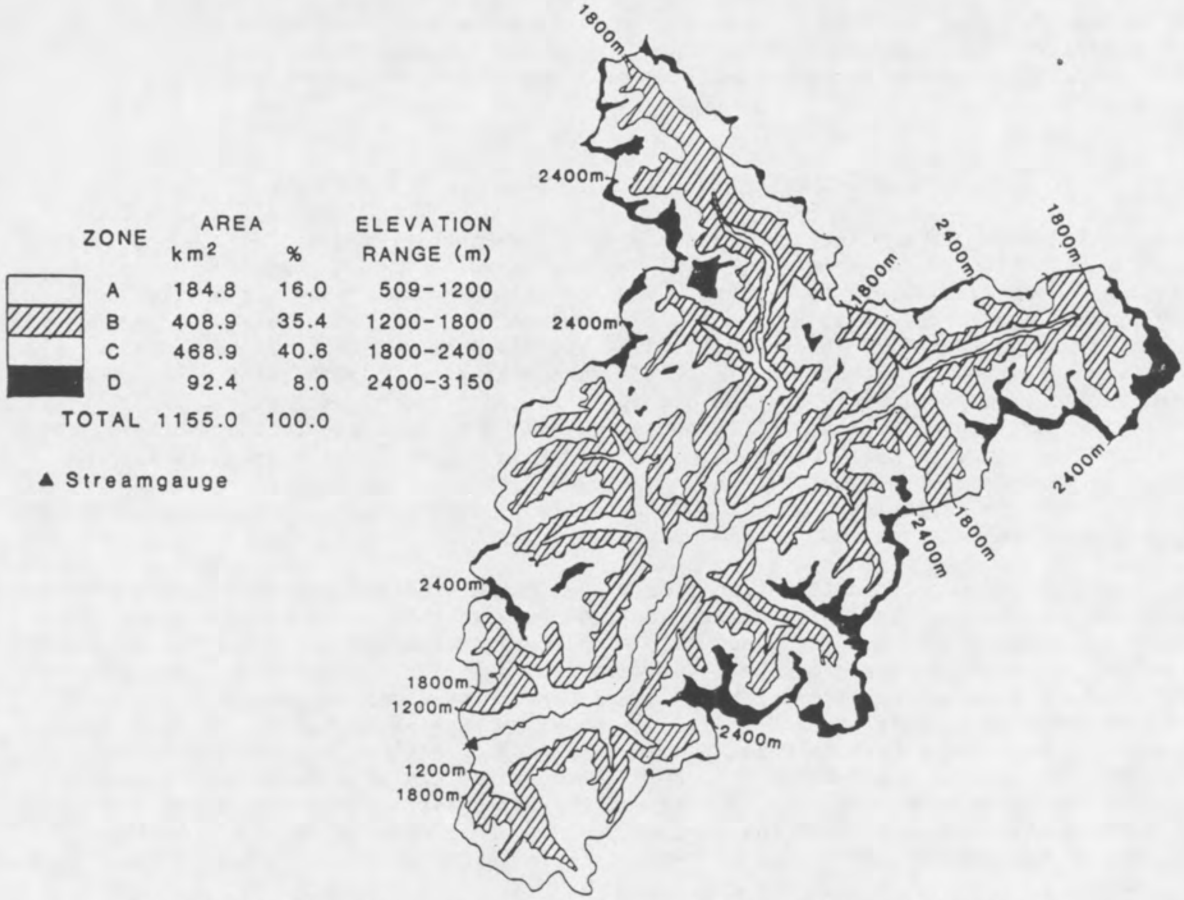

Figure 3. Elevation zones and areas of the Illecillewaet River basin at Greeley, British Columbia, Canada.

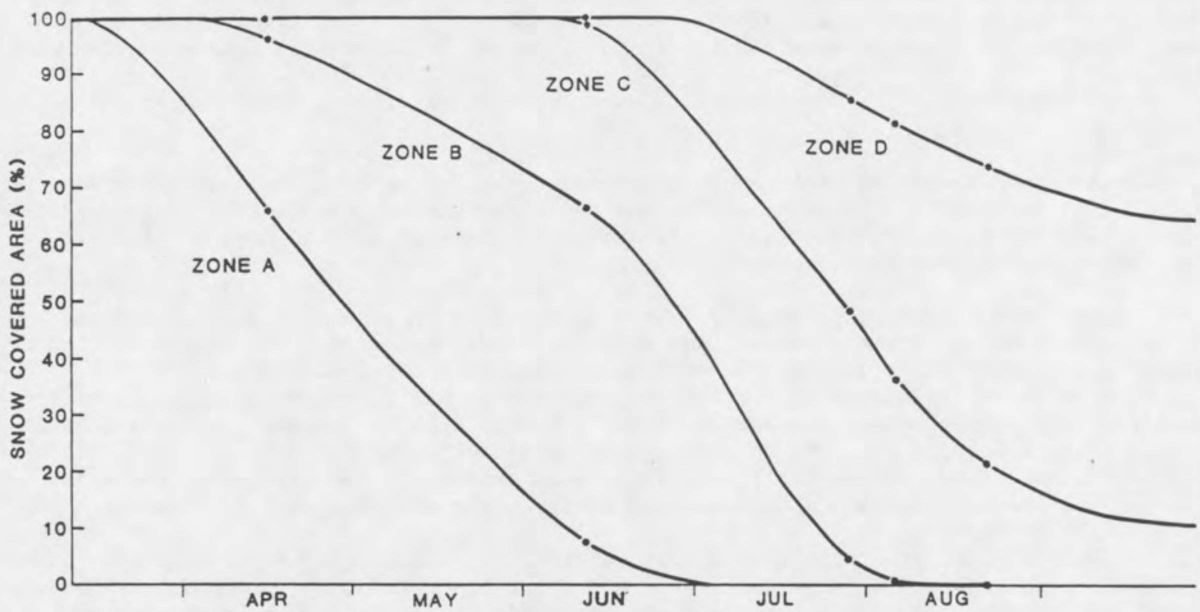

Figure 4. Snow cover depletion curves for various elevation zones of the Illecillewaet River basin, 1984 as obtained from Landsat MSS observations.

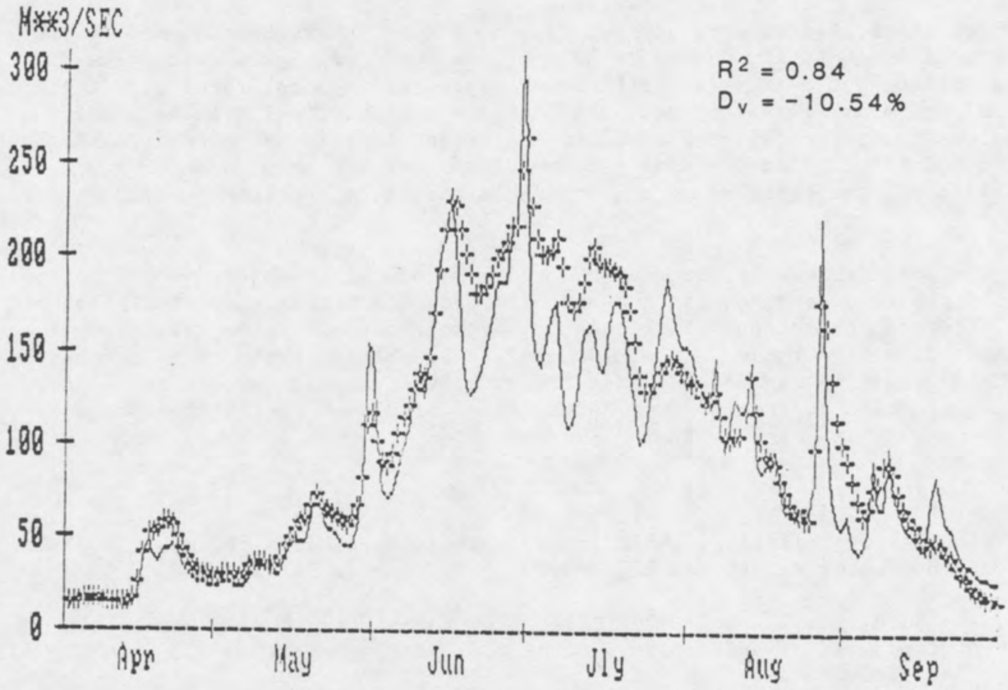

Figure 5. SRM estimated (+) versus measured (-) streamflow for the Illecillewaet basin (1155 km^2) for the 1984 snowmelt season with no updates.

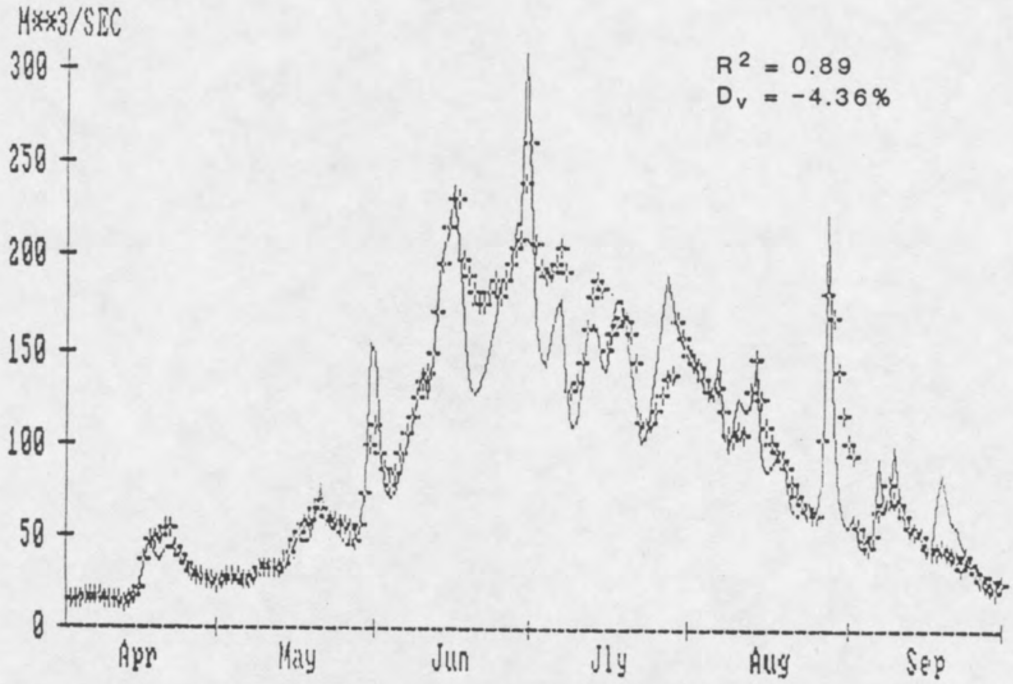

Figure 6. SRM estimated (+) versus measured (-) streamflow for the Illecillewaet basin (1155 km^2) for the 1984 snowmelt season with every 7th day actual discharge updates.

CONCLUSIONS

The WMO has instituted several worthwhile projects on intercomparison of hydrological models for runoff estimation. One major benefit has been highlighting the need for using quantitative performance evaluation criteria when presenting model results. It is suggested that numerical criteria always be used, but that the number of criteria be limited so as to avoid confusion in interpretation of results. A second benefit has been focussing attention on the use of updating to improve model accuracy. A case has been made for simple to use and objective updating algorithms so that they can easily be applied by a variety of model users.

Despite the positive results of the WMO projects, one major objective has yet to be obtained. The projects have not yet successfully provided a realistic simulation of operational forecast conditions. Particularly, a project needs to be conducted using imperfect input data like those that would normally be obtained when using forecasted temperature and precipitation data. Using the excellent experience obtained so far, WMO is the likely organization to sponsor such a realistic, simulated operational project.

BIBLIOGRAPHY

World Meteorological Organization, 1986: Intercomparison of models of snowmelt runoff, Operational Hydrology Report No. 23, WMO-NO. 646, Geneva, Switzerland.

Nash, J. E., and Sutcliffe, J. V., 1970: River flow forecasting through conceptual models, Part 1 - A discussion of principles, Journal of Hydrology Vol. 10(3), pp. 282-290.