Statistical Evaluation of WATFLOOD

By: Angela MacLean, Dept. of Civil & Environmental Engineering, University of Waterloo, On.

October, 2005

The statistics program associated with WATFLOOD uses spl.csv file that is produced with the WATFLOOD output. For each station the reliability of the model is assessed using eight different criteria. The criteria used include deviation of runoff volumes *(D_v)*, the absolute percent bias *(APB)*, the root mean squared error *(RMSE)*, the mean absolute error *(MAE)*, the bias *(b)*, the Nash-Sutcliffe coefficient *(N_r)*, the correlation coefficient *(r^2)*, and the modified Garrick coefficient *(G_r)*. The first six criteria test the differences between models and the last three evaluate how well the simulated time series fits the observed data. Each of the criteria is briefly described below.

Deviation of runoff volumes (D_v)

The deviation of runoff volumes D_v, also known as the percentage bias, is perhaps the simplest goodness-fit criterion. Its value is calculated using equation 1:

$$D_v(\%) = \frac{\sum_{i=1}^{N}(S_i - O_i)}{\sum_{i=1}^{N} O_i} * 100 \tag{1}$$

where S_i is the simulated discharge for each time step and O_i is the observed value. N is the total number of values within the period of analysis. For a perfect model, D_v is equal to zero. The smaller the D_v value, the better the performance of the model.

Absolute Percent bias (APB)

The absolute percent bias is a measure of the timing difference between the streamflow observations and the model simulations. It is usually used in conjunction with the D_v criterion. Given an observed and simulated series where the D_v value is small and the *APB* is large, one could conclude that both series share similar volumes but that their timing is not as close. Thus, a good agreement in timing and volume requires D_v and *APB* to be small. *APB* is always greater than D_v, and its value is determined using equation 2:

$$APB(\%) = \frac{\sum_{i=1}^{N} | S_i - O_i |}{\sum_{i=1}^{N} O_i} * 100 \qquad (2)$$

where all values have the same meaning as in equation 1

Root mean square error (RMSE), mean absolute error (MAE), and bias (b)

These three indicators provide a quantitative estimate of the differences between models in units of discharge (m^3/s). The values of these three criteria are used in this study to establish a relative comparison of the three WAT* models. *RMSE, MAE*, and *b* are calculated using equations 3 to 5:

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^{N} (S_i - O_i)^2 \right]^{1/2} \qquad (3)$$

$$MAE = \frac{\sum_{i=1}^{N} \left| S_i - O_i \right|}{N} \tag{4}$$

$$b = \frac{\sum_{i=1}^{N} (S_i - O_i)}{N} \tag{5}$$

where all the terms have the same meaning as above.

Nash-Sutcliffe Coefficient (N_r)

Along with the coefficient of correlation, the Nash-Sutcliffe coefficient (N_r) is a measure of statistical association, which indicates the percentage of the observed variance that is explained by the predicted data. The Nash-Sutcliffe coefficient, also known as the efficiency criterion, is perhaps the most common measurement mentioned in the hydrological literature for evaluating the performance of a model. Initially proposed by Nash and Sutcliffe (1970), N_r is estimated using equation 6:

$$N_r = 1 - \frac{\sum_{i=1}^{N} (S_i - O_i)^2}{\sum_{i=1}^{N} (O_i - O_i^*)^2} \tag{6}$$

where O_i^* is the average measured discharge and all the other variables have the same meaning as above. The second term in equation 6 represents the ratio between the mean square error (*MSE*) and the variance of the observed data. Thus, a value of N_r equal to zero indicates that the model output is not better than that obtained using the simple averaged observed streamflow for the entire period of analysis.

A shortcoming of the Nash-Sutcliffe statistic is that, because of its definition, it puts more emphasis on extreme events than on average flows. Additionally, the timing of the predicted series greatly influences the value of the coefficient.

Garrick Coefficient (Gr)

The Garrick coefficient is a modified form of the Nash-Sutcliffe coefficient. The Garrick coefficient uses daily averages as apposed to one overall average for the whole time step. This in turn emphasizes the average flows and not extreme flows or the seasonal variation. The Garrick coefficient was initially purposed by Garrick, Cunnane and Nash (1978). *Gr* is estimated using equation 7.

$$N_r = 1 - \frac{\sum_{i=1}^{N}(S_i - O_i)^2}{\sum_{i=1}^{N}(O_i - O_d)^2} \qquad (7)$$

where O_d is the average flow measured on date d and all the other variables have the same meaning as above.

Correlation Coefficient (R)

The *R* statistic describes the degree of colinearity between the observed and predicted time series. *R* is determined as indicated in equation 8. A perfect model has a correlation coefficient equal to 1.0. High values of the *R* coefficient indicate better agreement between observations and simulations.

$$R = \frac{\frac{1}{N} * \sum_{i=1}^{N} (O_i - O_i^*) * (S_i - O_i^*)}{\sqrt{\left(\frac{N * \sum_{i=1}^{N} O_i^2 - \left(\sum_{i=1}^{N} O_i \right)^2}{N * (N-1)} \right)} * \sqrt{\left(\frac{N * \sum_{i=1}^{N} S_i^2 - \left(\sum_{i=1}^{N} S_i \right)^2}{N * (N-1)} \right)}} \qquad (8)$$

where all symbols have the same meanings as above.

As with the Nash-Sutcliffe statistic, the correlation coefficient is more sensitive to outliers than to values near the observed mean.