INVITED COMMENTARY
HYDROLOGICAL PROCESSES
Hydrol. Process. **22**, 2723–2725 (2008)
Published online 28 May 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hyp.7072

Do Nash values have value? Discussion and alternate proposals

Robert E. Criss* and William E. Winston

Department of Earth and Planetary Sciences, Washington University in St. Louis, MO, USA

*Correspondence to: Robert E. Criss, Department of Earth and Planetary Sciences, Washington University in St. Louis, MO, USA. E-mail: criss@wustl.edu

Received 8 April 2008
Accepted 21 April 2008

Schaefli and Gupta (2007) point out the need for useful metrics to evaluate and compare the performance of hydrologic models, make helpful suggestions and invite discussion. They discuss the shortcomings of the 'Nash–Sutcliffe Efficiency', hereafter NSE, originally proposed by Nash and Sutcliffe (1970):

$$NSE = 1 - \frac{\sum(Q_{\mathrm{calc}} - Q_{\mathrm{obs}})^2}{\sum(\overline{Q}_{\mathrm{obs}} - Q_{\mathrm{obs}})^2} \tag{1}$$

where Q_{obs} is the observed streamflow, Q_{calc} is the streamflow predicted by some models, and $\overline{Q}_{\mathrm{obs}}$ is the mean observed flow over the interval of interest. Schaefli and Gupta (2007) point out that few modellers in other environmental fields know what an NSE value is. However, they follow Legates and McCabe (1999), Seibert (2001) and several others who propose to modify and improve this metric by the substitution of a suitable objective benchmark Q_{bench} for the mean flow $\overline{Q}_{\mathrm{obs}}$ in Equation (1).

Schaefli and Gupta (2007) do not mention the one major point made by Legates and McCabe (1999) and Krause *et al.* (2005), which is that the NSE overemphasizes large flows relative to other measurements because the various deviations are squared. This problem is shared by the familiar correlation coefficients, R and R^2, and while this weighting scheme is statistically questionable, it could be argued that large flows have the greatest importance to humans. However, this defect is serious because the very flows that are overrepresented in such 'goodness-of-fit' calculations are for many reasons the least accurately measured. To offset this problem, Legates and McCabe (1999) propose that Equation (1) be replaced with a more general formulation of efficiency 'E' based on absolute values:

$$E_j = 1 - \frac{\sum |Q_{\mathrm{calc}} - Q_{\mathrm{obs}}|^j}{\sum |Q_{\mathrm{bench}} - Q_{\mathrm{obs}}|^j} \tag{2}$$

where j is a power, and concluded that $j = 1$ was the best choice.

We question whether the NSE or any such modification has the optimal value to the hydrologic community. While both NSE and E_j are conveniently dimensionless and easy to calculate, neither is bounded as they range from a value of 1 indicating perfect agreement between model and measurement to a value of minus infinity. Oddly, large negative values returned by Equations (1) or (2) do not necessarily indicate that some model is abysmally poor, but only that the observed streamflow is very steady or otherwise closely matches the 'benchmark' values!

We propose that a 'volumetric' efficiency (VE) circumvents several of these problems and presents some new advantages:

$$VE = 1 - \frac{\sum |Q_{\mathrm{calc}} - Q_{\mathrm{obs}}|}{\sum Q_{\mathrm{obs}}} \tag{3}$$

On a standard hydrograph, i.e. a plot of discharge Q versus time, the areal deviation between the predicted and actual hydrographs represents a volume of water. For an unbiased model, scaled so that the predicted total volume of water delivered over a given time interval matches the actual volume delivered, the VE thus ranges from 0 to 1 and represents the fraction of water delivered at the proper time; its compliment represents the fractional volumetric mismatch. For detailed time series, this calculated volumetric fraction is accurately indicated by Equation (3), although if a few data points are available other algorithms would be more accurate. Unlike the usual correlation coefficients or the NSE, E_j and other related metrics, the VE has obvious physical significance to hydrologists. The VE would be particularly helpful in comparing the performance of similarly scaled, rainfall-runoff transfer functions. It has the additional advantage of being easy to calculate, and of treating every cubic metre of water the same as any other cubic metre, whether it be delivered during slow recession or during peak flow.

Figure 1 shows how the correlation coefficient (R^2), the linear regression slope m, the NSE and the VE vary when a typical, complex, actual hydrograph, representing Q_{obs}, is compared to a 'calculated' hydrograph, 'Q_{calc}'. In Figure 1(a) the 'calculated' hydrograph is determined by progressively offsetting the time axis of the real hydrograph, and in Figure 1(b) it is determined by introducing bias by multiplying all actual discharge values by a constant factor. The various metrics for the mathematical comparisons of the 'calculated' and real hydrographs are then determined by standard linear regression (slope m, correlation coefficient R^2) or from Equations (1) and (3). Note that the VE decreases linearly as model agreement degrades, unlike the other metrics. Interestingly, the NSE is seen to be similar to R^2 in Figure 1(a), but very different in Figure 1(b). Figure 1(b) also shows that R^2 is independent of model bias, a serious defect pointed out by Legates and McCabe (1999) and many others.

Alternatively, given that hydrologists need to be understood when communicating their important predictions of flood and flow to others, they might just use the widely understood correlation coefficient. Its defect regarding bias for hydrologic applications (Figure 1(b)) would be significantly offset by simply reporting the regression line for a Q_{calc} versus Q_{obs} graph. Moreover, as this measure is attributed to Gauss, it has two centuries of precedence.

The above suggestions do not eliminate the need for useful benchmarks mentioned in earlier papers. For any given situation, the metrics calculated from Equations (1) to (3) are only useful when directly compared to values for other models, or benchmarks applied to the same problem. However, we question whether it is useful to directly incorporate such benchmarks into the formula for the particular metric. This is not only because of the above problems, but because the optimal benchmark will differ for different applications, which indeed is the reason why so many benchmarks have been proposed. In our view, comparisons of stream flow made by the NSE and allied algorithms are more distorted and less amenable to understanding than the VE.

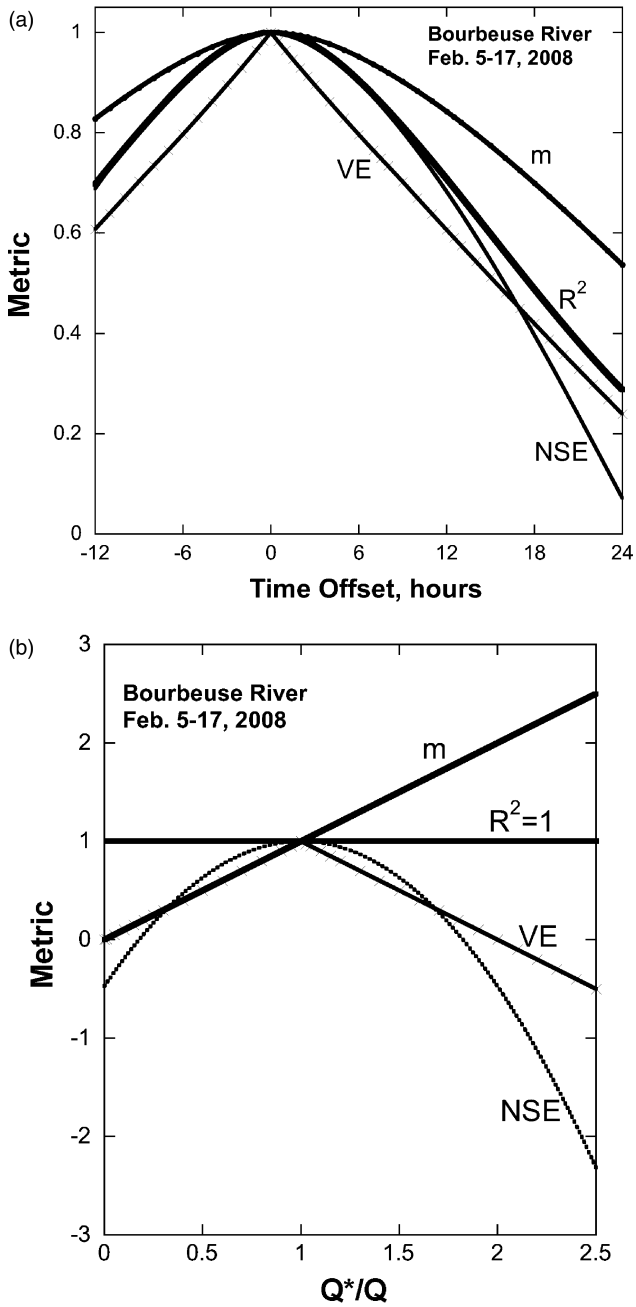

Figure 1. (a) Comparison based on various metrics of the flood hydrograph of 5–17, February 2008 for the Bourbeuse River, Missouri, relative to the same hydrograph that is offset temporally by the indicted amount. (b) Similar diagram for the Bourbeuse River hydrograph, compared to itself after introducing systematic bias by multiplying all measured discharge values by a constant factor, indicated by the ratio Q^*/Q. Analogous results would be returned for other choices of the real hydrograph

Hydrol. Process. 22, 2723–2725 (2008)
DOI: 10.1002/hyp

References

Krause P, Boyle DP, Base F. 2005. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* 5: 89–97.

Legates DR, McCabe GJ. 1999. Evaluating the use of "goodness of fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research* 35: 233–241.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models: Part 1- A discussion of principles. *Journal of Hydrology* 10: 282–290.

Schaefli B, Gupta HV. 2007. Do Nash values have value? *Hydrological Processes* 21: 1075–1080.

Seibert J. 2001. On the need for benchmarks in hydrologic modeling. *Hydrological Processes* 15: 1063–1064.

Hydrol. Process. **22**, 2723–2725 (2008)
DOI: 10.1002/hyp