Evaluating Performance of Watershed Hydrologic Simulations at Daily Time Step Using Autoregression

M.D. Tomer[1], P.C. Beeson[2], D.W. Meek[1], D.N Moriasi[3], C.G. Rossi[4], and A.M. Sadeghi[2]

USDA/ARS locations: 1-Ames IA; 2-Beltsville MD; 3-El Reno OK; 4-Temple TX

National Laboratory for Agriculture and the Environment

What's this about? Watershed models that simulate the hydrology of large rivers, including the Soil and Water Assessment Tool (SWAT) are typically run on a daily time step. However, accuracy of simulations (i.e., calibration/validation results) are most frequently reported on aggregate time periods (i.e., monthly and/or annually; see Gassman et al., 2007; Douglass Mankin et al., 2010). More consistent evaluation of hydrologic performance at the daily time step should enable river basin models to better represent dynamics that drive contaminant transport, improving their utility for water quality management. But there are statistical challenges to evaluating daily hydrologic simulations, because daily discharge from rivers varies across orders of magnitude.

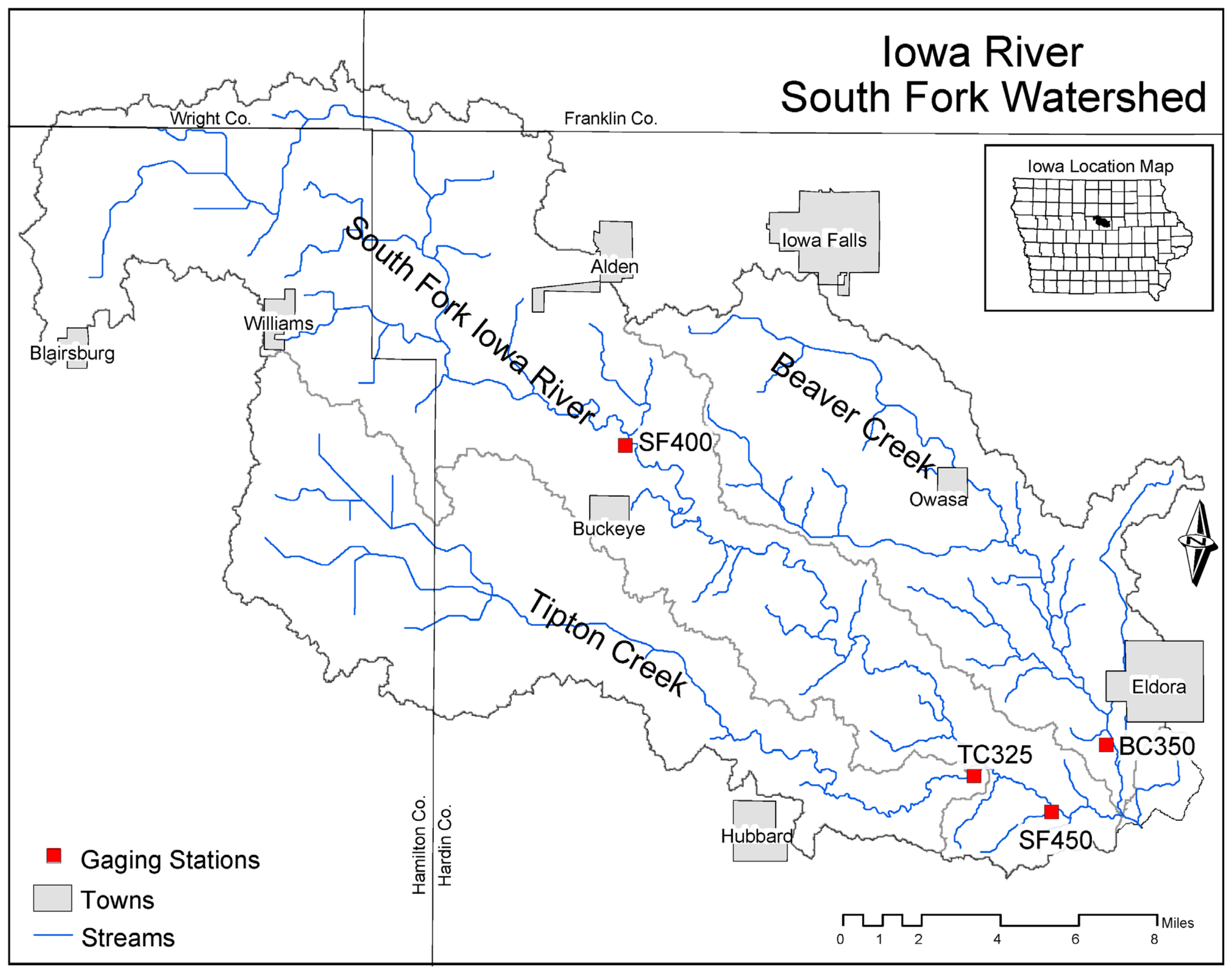
Goal: To evaluate autoregression as a tool to evaluate simulation of daily discharge from rivers.

Bottom line result: Autoregression can generate a calibration target to assess accuracy of daily simulations from measured data alone. Caveats exist, but they are explained by stream flashiness.

What is autoregression? In the context of large-basin hydrology, autoregression provides a statistical representation of the measured discharge record, based on long term trends, cycles of seasonal variability, and the inherent co-dependence of a given day's discharge on antecedent flows. It is expressed as an equation that generates an estimate of the measured daily discharge (Q) record. The form applied for this study is shown below. The coefficients a_0 through a_8 were fitted through iteration, using SAS proc NLIN. Note Q data are $\log_e$ transformed.

$$\overbrace{y_i}^{\text{ln Q}} = \underbrace{a_0}_{\text{intercept}} + \overbrace{a_1\sqrt{t_i} + a_2 t_i + a_3 t_i^{3/2}}^{\text{long term trend}} + \underbrace{a_4 \cos\left[\frac{2\pi}{365}(t_i - a_5)\right]}_{\text{annual (seasonal) cycle}} + \overbrace{a_6(y_{i-1} - \hat{y}_{i-1}) + a_7(y_{i-2} - \hat{y}_{i-2}) + a_8(y_{i-3} - \hat{y}_{i-3})}^{\text{dependence on antecedent flows}} + \underbrace{e}_{\text{error term}}$$

Context: The study watershed was the South Fork of the Iowa River (Tomer et al., 2008). Daily flow data at four gage stations were available from 2001 through 2009. Beeson et al. (2011) evaluated several sources of input precipitation data and presented monthly validation results. Here we took the two best precipitation input results using rain gage and radar rainfall data (verbatim, no recalibration) and calculated daily performance. There were separate periods for calibration (2001-04 & 2008-09) and validation (2005-07) but we only present results for the full 9-yr period here. We also fit autoregressive equations to data from the four gages and calculated daily performance.

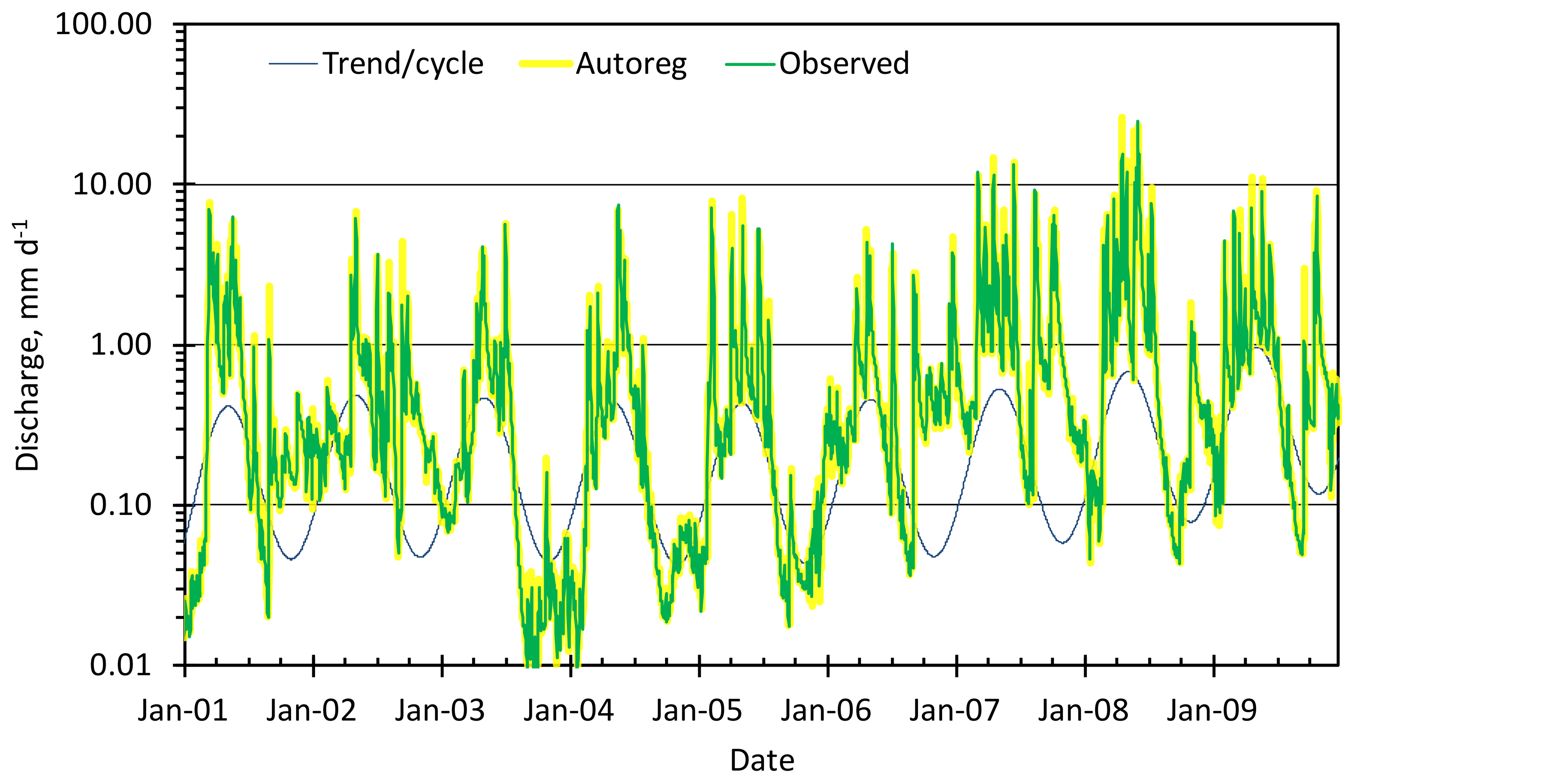


Gage	Area (km^2)	Ave Q (mm yr^{-1})
SF450	580	305
SF400	256	322
TC325	198	281
BC350	182	307

Results: Autoregressive equation coefficients (with standard errors). Differences among coefficients (which are significant for a_6, a_7) are consistent with differences in basin size and stream flashiness.

Gage Site (R^2)	Intercept	Trend			Annual cycle		Autoregressive terms		
	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
SF450	-6.38	0.39	-0.011	0.00010	1.16	132.7	1.38	-0.56	0.15
(0.97)	(1.60)	(0.16)	(0.005)	(4.5×10^{-5})	(0.17)	(8.3)	(0.02)	(0.03)	(0.02)
SF400	-1.73	ns	ns	1.2×10^{-7}	1.27	115.4	1.31	-0.50	0.15
(0.95)	(0.27)			(5.5×10^{-8})	(0.24)	(11.0)	(0.02)	(0.03)	(0.02)
TC325	-4.39	0.32	-0.012	0.00013	1.25	130.0	1.29	-0.47	0.14
(0.97)	(1.92)	(0.19)	(0.006)	(5.9×10^{-6})	(0.19)	(9.0)	(0.02)	(0.03)	(0.02)
BC350	-1.00	-0.0023	1.7×10^{-6}	3.0×10^{-10}	0.99	125.3	1.13	-0.34	0.14
(0.93)	(0.37)	(0.001)	(6.8×10^{-7})	(1.4×10^{-10})	(0.12)	(7.3)	(0.02)	(0.03)	(0.02)

Example plot of autoregressive equation results, shown with observed data for the gage at SF450.



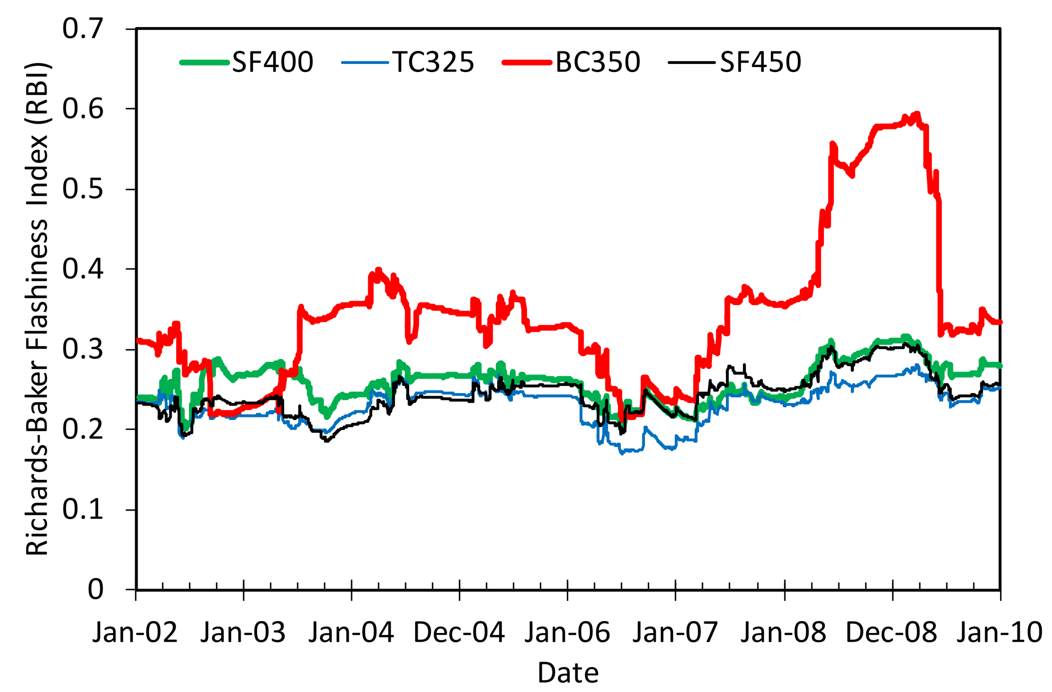
Model performance statistics for daily SWAT output using two precipitation inputs, and for data estimated from autoregression are shown below. Note autoregressive equations were fit with a constraint to minimize bias, and generated a 95% confidence interval of [ln Q] for each day. The frequencies at which SWAT generated daily flows outside these intervals are also given.

Note: NSE=Nash-Sutcliffe Efficiency, PBIAS=% Bias, RSR=root mean square error : std. deviation ratio.

Gage site	Rainfall input to SWAT model/ AR eqn.	Natural log transform [ln(mm/d)]					Original variate [mm/d]		
		NSE	PBIAS	RSR	%<0.95 CI	%>0.95 CI	NSE	PBIAS	RSR
SF450	Rain gage	0.53	-29.3	0.69	42.0	17.8	0.74	15.1	0.51
	Radar	0.55	-28.5	0.68	41.7	17.3	0.74	11.1	0.51
	Autoregressive eqn.	0.97	0.0	0.13	0.9	3.3	0.82	1.2	0.38
SF400	Rain gage	0.41	-41.2	0.77	43.0	13.5	0.61	17.9	0.63
	Radar	0.45	-38.1	0.74	41.4	13.5	0.61	13.1	0.62
	Autoregressive eqn.	0.96	0.0	0.17	1.6	3.8	0.77	2.6	0.41
TC325	Rain gage	0.50	-7.2	0.71	32.9	24.0	0.58	10.7	0.65
	Radar	0.54	-3.3	0.68	30.6	24.9	0.56	4.6	0.66
	Autoregressive eqn.	0.97	0.0	0.14	1.1	3.5	0.81	2.5	0.37
BC350	Rain gage	0.60	-7.4	0.63	23.3	16.4	0.56	-2.8	0.67
	Radar	0.60	-21.0	0.63	29.2	11.2	0.54	6.4	0.67
	Autoregressive eqn.	0.93	0.0	0.20	0.8	3.8	0.29	4.7	0.77

Interpretation: Poorer fit of the autoregressive equation at BC350 (R^2=0.93) was inherently due to greater flashiness of this stream (see right). Stream flashiness may be a good performance metric to evaluate model simulations at a daily time step. Here, daily SWAT output matched flashiness to within 20% at the three SF and TC stations where observed stream flashiness was <0.3. The autoregressive-estimated discharge, when transformed from log to original units, gave better estimates of Q than the SWAT model at the same three stations.

Stream flashiness (plotted as a 365-d moving average) is a ratio of summed changes in daily Q to summed daily Q (Baker et al., 2004).



Conclusions

❶For these four gages, if simulations could show NSE >0.9 and RSR<0.2 (in transformed scale), that would be comparable to a re-estimation of the original data series using autoregression.

❷Autoregression can generate daily 95% confidence intervals for daily data but these intervals expand where stream flashiness increases and fit of the autoregressive equation decreases.

❸Therefore, autoregression can provide targets for simulation model performance, based on measured data alone, if stream flashiness is not too high. In this study, where the observed daily Richards-Baker Flashiness Index was less than about 0.3, autoregressive equations provided a consistent target to evaluate SWAT performance.

❹Stream flashiness is a useful metric to evaluate simulations of daily river discharge.

❺Autoregression can help evaluate simulations of daily discharge, and warrants further testing.