I have finished my review of the paper “To what extent does river routing matter in hydrological modeling”, by Cortés-Salazar et al., submitted to HESS. This generally well-written paper attempts to examine the influence of routing algorithm and time step on model performance using subsets of 3500 different runoff regimes generated using the VIC hydrological model. Very mild differences were found between algorithm and time step choices, with one exception: where routing was not simulated, the performance was consistently poor relative to models which simulate routing. Unfortunately, this is not a very compelling result.

While the approach was rigorous in the sense that it compiled data from thousands of model simulations, it suffers from a number of critical methodological issues. I discuss a few of these major issues below:

- Routing is most influential on peak magnitude and timing of large events; both of these are poorly captured by integrated hydrograph metrics such as KGE and NSE. Peak flow differences after calibrating the routing models would be a much more useful metric for evaluating routing model performance. By using integrated measures such as KGE, the critical differences between routing algorithms are not discernable (as seen in nearly all of the reported results).
- Each of the figures in the report are reporting ALL of the output from the simulations, regardless of whether it is important or interpretable or worthy of interpretation. For instance, figure 5 reports KGE, NSE, and NSE of log transformed flows for all 3500 simulations with multiple timesteps, multiple routing schemes. In addition to the only interpretable result from this figure is that no routing is outperformed by routing, there is little utility in comparing NSE values of 0.2-0.3 (the approximate median of these simulations) -differences in NSE below about 0.5 are nearly arbitrary in that a hydrograph with an NSE of 0.2 may not be visibly preferable to an NSE of 0.05. The only feature of this plot referred to in the text was the maximum metric value. Why not simply report that?
- Critically, because the parameters of the VIC model are arbitrary, the comparisons of
even the best models are in effect the results of Monte Carlo calibration, the least efficient optimization approach. Comparing the ‘best’ models when these are not rigorously determined to be the actual best for each algorithm (rather than a sampling error) is problematic. For this comparison to be rigorous, I don’t see how to do this without simultaneous calibration of routing and land surface parameters, an issue the authors acknowledge in section 5.2.

- In practice, the routing parameters (such as Manning’s n) would be calibrated in conjunction with VIC model parameters, likely further diminishing any incremental performance differences between the routing models.
- The fundamental results discussed here are obvious without the testing herein. Routing is better than no routing. Low flows are not as impacted by routing differences. Parameter compensation occurs in hydrological models. Routing impacts are not visible when averaged over a monthly time step. Models with instantaneous routing (i.e., no routing) have higher flows. As is, I do not see an additional contribution from this work beyond that which exists in the literature.

It is for the above reasons that I must recommend rejection.