

Geosci. Model Dev. Discuss., referee comment RC3
<https://doi.org/10.5194/gmd-2022-61-RC3>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2022-61

Anonymous Referee #3

Referee comment on "CREST-VEC: a framework towards more accurate and realistic flood simulation across scales" by Zhi Li et al., Geosci. Model Dev. Discuss.,
<https://doi.org/10.5194/gmd-2022-61-RC3>, 2022

I have finished my review of the paper "CREST-VEC: A framework towards more accurate and realistic flood simulation across scales", by Li et al., submitted to Geoscientific Model Development. This paper outlines the integration of two existing models – the gridded hydrological model CREST and the vector-based routing tool mizuRoute. The authors examine the (significant) improvement in computational cost associated with using a vector-based routing tool and explore the impact of including or not including lakes on simulations of 5350 streamflow hydrographs across the U.S., with impacts evaluated using standard model fit metrics (e.g., NSE/bias) and using flood detection measures (POD, FAR, and CSI). Lastly, the author examine the correlation of absolute model performance against environmental indices from the CAMELS database.

In general, I found this paper to be a bit disorganized and missing a number of important details. It uses less-than-ideal approaches for comparing populations of model runs and does not seem (to me) to represent a significant contribution either as a model development or in presenting new insights into models more generally. I document below a number of major issues with the paper which should be addressed prior to publication.

- The title of this paper implies it may be a discussion of a fundamentally new model or modelling framework. However, it is the merger of an existing routing model which has already been coupled to multiple existing hydrological models (including gridded models) with yet another hydrological modelling code. The details of this integration are generally missing, so it is difficult to evaluate the challenge in performing this integration, whether it is file-based coupling or a single compiled code, whether anything other than the spatial aggregation of runoff is causing the computational improvements in the shift from CREST to CREST-VEC. It does not appear that this model development effort in and of itself is a contribution. If it is, then the authors need to demonstrate why.
- The authors use results from thousands of model runs in order to assess the benefit of including lakes in vector-based routing. The effort to do this is not trivial, and I applaud

the authors for their ambition in the size of this computational experiment. However, there are some issues with both the magnitude of the contribution and the means of assessing the results. About the question of whether simulating lakes is useful, the inclusion of lakes has already been demonstrated to improve vector-based routing results (Han et al., 2020), though perhaps not in as rigorous of a manner as could potentially be done with this massive model set. However, there are several fundamental issues with the analysis herein. Firstly, the model in general (before and after lakes are included) performs very poorly, with median NSEs on the order of 0.3, and 45% of simulations having NSEs below zero. The authors make no attempt to reduce their analysis to focus on models with that satisfy some adequate base performance. They also spend most of their time discussing improvements to the median, which means that much of the performance can easily be due to changing very poorly performing models to slightly less terrible models. They miss the opportunity to statistically compare the probability distributions of metrics to see whether the distributions are functionally different. At the very least, I would ask that they compare median change in NSE (a more adequate metric for evaluating improvements) rather than the change in median NSE.

- Most of this paper is thematically consistent – examining the benefits of using a vector-based routing model with lakes across the continental U.S.. The second half of section 3.2, however, evaluates the raw performance of the model (with lakes) against CAMELS environmental variables. This would make sense if the authors evaluated the benefit of including lakes (i.e., shift in NSE) against these variables, but as is, this analysis is really an examination of the quality of CREST’s runoff estimates. I suggest that if this is to be retained, this section be re-cast to answer questions such as “under what environmental conditions is the inclusion of lakes more likely to be beneficial”, which can be done by (e.g.) performing the analysis of figure 7 with improvements to NSE rather than the raw score.
- The calibration in the Houston case study is missing important details. How many parameters were calibrated? Which ones? Using what objective function? It also uses an unconventionally small 1 year calibration period and 2.5 year validation/evaluation period, with no reference to a run-up period.
- How were the model parameters determined for the CONUS application?
- The justification for selecting model populations is missing – why/how were the 283 models in section 3.3 selected? Are these all of the gauges downstream of natural lakes in the dataset? What distance threshold defines ‘downstream’? How were the 5 local cases discussed in figure 10 chosen – given the minimum NSE difference is > 0.55 , these are not random samples, but rather seem to be cherry-picked to illustrate the most successful inclusion of lakes. It would be useful to see at least one model where inclusion of lakes degraded model performance, with speculation as to what circumstances might cause this.
- The false detection analysis of section 3.3 also suffers from the inclusion of all models, regardless of quality. Why even analyse the false detection performance with an NSE of 0.1? These models are already not fit for the task, so including them in the analysis could very well skew the results such that they imply performance improvements with lake inclusion, even if this performance improvement is meaningless (a NSE of -0.35 is not functionally better than an NSE of -0.55).
- Supplementary materials should only be provided to corroborate existing evidence in the paper; the authors use Fig S1 to make a completely distinct point. This content should be either removed or incorporated into the main document.
- The authors have multiple discussions of ideas that are very loosely related to the paper and/or not tied to any specific results herein, and should be removed:
 - Ensemble forecasts at line 229
 - Advocating for modular model structure at line 377
 - Support for parameter regionalization line 388
 - Two-way feedback between social systems and catchment signatures (In 398)
 - Future work on machine learning based reservoir operation simulation (line 402)

- The authors use percent change in NSE and BIAS to present results (e.g., in figure 6). However, this metric is very problematic for variables that can be positive or negative, because the denominator can go to zero. Another metric must be used.
- Multiple minor issues
 - Bias is reported both as a percentage 0-100% and as a floating point 0-1 (fig 5)
 - Speedup is usually used to evaluate improvements in computational costs; the authors here only report (less generalizable) differences in raw run times per time step.
 - The reasons provided for changing bias at line 247 are implausible – to influence bias you need to have water leave the domain by means other than streamflow. A more likely scenario is that this would be due to evaporation from the lake surface.
 - D-infinity (line 78) is also a grid-based (not vector-based) algorithm, and should be cited as Tarboton (1997)
 - The CREST model original paper (Wang 2011) should be included upon its first mention. How is this paper not referenced?
 - Many minor text errors not inventoried here given expectation of significant revisions

References

Han, M., J. Mai, B.A. Tolson, J.R. Craig, É. Gaborit, H. Liu, and K. Lee, Subwatershed-based lake and river routing products for hydrologic and land surface models applied over Canada, *Canadian Water Resources Journal*, doi:10.1080/07011784.2020.1772116, 2020

Tarboton, D. G., (1997), "A New Method for the Determination of Flow Directions and Contributing Areas in Grid Digital Elevation Models," *Water Resources Research*, 33(2): 309-319

Jiahu Wang , Yang Hong , Li Li , Jonathan J. Gourley , Sadiq I. Khan , Koray K. Yilmaz , Robert F. Adler , Frederick S. Policelli , Shahid Habib , Daniel Irwn , Ashutosh S. Limaye , Tesfaye Korme & Lawrence Okello (2011) The coupled routing and excess storage (CREST) distributed hydrological model, *Hydrological Sciences Journal*, 56:1, 84-98, DOI: 10.1080/02626667.2010.543087