General Comments

The authors provide a description of a new hydrologic model framework, CREST-VEC, which couples CREST and vectorbased routing framework, mizuRoute, and sub-surface runoff and lake schemes are also incorporated. Compared with the raster-based routing scheme implemented in the original CREST, the new framework shows a significant improvement in computational efficiency and reproducibility in river discharge. Therefore, the proposed model has an essential potential for a flood forecast and alert system, as the authors suggested, and the manuscript is informative for the readers of this journal. However, since two contexts (speed-up of the model and introduction of new physical processes) are mixed in the manuscript, it requires some revision of the manuscript and experimental configurations for a well-structured article.

Response:

Thank you for your comments and suggestions. We made changes to restructure our manuscript as suggested. Specific changes are listed below.

Specific Comments

For the abstract and corresponding sections:

The authors are highly requested to discuss 1) vectorization of the routing scheme and the improvement in computational efficiency (and the ensemble simulation, as they suggested) and 2) the introduction of new physical processes, the change in reproducibility of river discharge, and the false alert problem separately. Therefore, the article should be divided into two parts for the sake of readability. However, if they are still to be published as one paper, some parts need to be reorganized. For example, the introduction should describe the significance of the subsurface flow and lake routing for the flood forecast.

Response:

Thanks for your suggestions. We made following changes to our manuscript.

In the **Abstract**, we specifically mentioned the objective of this study:

“Large-scale (i.e., continental and global) hydrologic simulation is an appealing yet challenging topic for the hydrologic community. First and foremost, model efficiency and scalability (flexibility in resolution and discretization) have to be prioritized. Then, sufficient model accuracy and precision are required to provide useful information for water resources applications. Towards this goal, we craft two objectives to improve US current operational hydrological model: (1) vectorized routing and (2) improved hydrological processes. This study presents a hydrologic modeling framework CREST-VEC that combines a gridded water balance model and a newly developed vector-based routing scheme.”

In the **Introduction**, we added one paragraph to introduce the importance of subsurface routing and lake routing specifically to flood prediction.

L.90-97: “To date, modern vector-based routing models such as RAPID and mizuRoute have neglected the subsurface routing, which is either assumed to be minimum (Mizukami et al., 2020) or treated in the way same as surface routing (Lin et al., 2019; Yang et al., 2021). However, subsurface routing is an important hydrologic process and dominates over regions that feature intermittent flow behaviors (Freeze, 1972). For flood simulation, ignoring subsurface routing could underestimate the peak flow and miscalculate the flood timing, both of which directly affect decision-making processes. An equally important research thrust is the representation of lakes and reservoirs in vector formats, since they markedly alter flow response not only at local scale, but also downstream rivers. One significant function of lakes and reservoirs in the US is for flood control, so simulation without incorporating such process is likely to result in falsely issued flood warnings."

The authors used computation time per step as a measure of computational efficiency. However, the number of computation steps varies depending on time-step width and constraints such as CFL conditions in some models (not sure if this is the case with your model). Therefore, it would be more appropriate to use the time it takes to compute a given period on a given spatial resolution (e.g., one year) rather than the computation time per step. The number of parallels and parallel efficiency are also important indicators for comparing computation algorithms

Response:

Thanks for your suggestions. In this study, we used average time taken per step for the whole simulation period – 5 years. Since the legacy model CREST does not have parallel capacity, we only compared single-core performance in this study. Parallel efficiency has been heavily discussed in Mizukami et al. (2021). In addition to comparing routing efficiency, we also take into account the total computational cost listed in Table 1, from which the new scheme is about six times faster than the operational model, even at higher resolution.

Table 1. Statistical comparison of model performance over the continental U.S. Bolded numbers indicate the best metrics among the three model configurations. The computational speed is calculated as an average speed over a whole simulation period.

Metrics	**Gridded CREST (Flamig et al., 2020)**	**CREST-VEC (w/o lake)**	**CREST-VEC (w/ lake)**
Simulation resolution	1 km	**90 m**	**90 m**
Total computational cost (hours)	149.2	**29.9**	32.96
Computational Speed for routing (sec/step)	7.2	**0.35**	0.37
Max NSE	0.71	**0.87**	**0.87**
Median NSE	-0.06	0.12	**0.18**
% gauges NSE>0	41.8 %	50.6 %	**56.2 %**
Max CC	**1.0**	0.96	0.96
Median CC	0.40	**0.67**	**0.67**
Median bias	**9%**	27%	17%

In my understanding, the routing scheme calculates the time lag from runoff from the land surface to downstream. So, why does the bias change between CREST and CREST-VEC, as shown in Figure 4 and Table 1? Or is this bias calculated for peak flows? (One possible factor is the use of externally derived reservoir storage. However, such modifications should not be applied even if they improve the model accuracy since it is difficult to discuss the impacts of the model update when the water budget in the overall model is changed.)

Response:

Thanks for your comments. The bias differences of streamflow shown between CREST and CREST-VEC originate from three parts. First, CREST and CREST-VEC use different routing scheme. CREST adopts a simplified kinematic wave routing (Vergara et al., 2016), while CREST-VEC uses a Unit Hydrograph routing (Mizukami et al., 2016). These two routing schemes in principle have different physical representations. Second, in Figure 4 and Table 1, we listed results both with and without lake module, and certainly results with lake module have less bias with considerable water being held in lakes. Third, Specifically for Figure 4, the CREST model is heavily calibrated utilizing the DREAM optimizer. So, the peak flow is calibrated against observed streamflow. However, the results with CREST-VEC are not calibrated because of large-scale simulation, which is considered as a future work. We included more calibration information in our main text:

L.233-236: "For the CREST model with gridded routing, we calibrate the model using the DREAM (Differentiable Evolution Adaptive Metropolis) optimization (Vrugt et a., 2009) from 2016-06-01 to 2017-06-01 at an hourly time step and perform evaluation from 2017-06-01 to 2020-01-01. The NSCE is used as objective function for calibration, and model is warmed up for one year from 2016-06-01 to 2017-06-01."

1 Introduction

Mainly for line 69: Even if a numerical model can be run at a fine spatial resolution in a realistic amount of time, it does not necessarily mean that the physics assumed in the model hold at the resolution. For example, a 1-D river model with a 100-m resolution is difficult to apply directly to a wide river such as the Amazon River, even if it can be run. The authors' work is technically excellent, but the validity of the model physics should be discussed in the discussion section.

Response:

Thanks for your comments. We admit current limitations in model physics in the discussion section and pasted here.

L. 699-706: "The CREST-VEC model, by no means, represent all physical hydrological processes. Instead, it is a conceptual flood forecast model that aims to deliver timely flood information to stakeholders, decision-makers, and broader users. We admit that some processes such as vadose zone modelling, snow melting, hillslope routing, in-channel river routing, and reservoir operations are simplified, and some processes such as vegetation and groundwater modelling are missing for the current version. Since there is always the trade-off between model complexity and efficiency, we hope to continuously push the envelope of this front to optimize real-time flood forecast system."

Line 84-85: Raster-based and vector-based models are reviewed in detail in the introduction section. However, information relevant to the research question listed here should be added. If previous studies on the impact of subsurface processes and lake considerations on a river model are presented, the reader will better understand the motivation for the study.

Response:

Thanks for your suggestions. We tailored four research questions to be reflective of the overall information we want to convey. Because previously the model performance has not been evaluated with and without subsurface routing and lake routing, especially regarding the linkage to flood warnings, we emphasize the importance of the two in our research questions.

L. 103-105: “Four questions are posed in this regional case study: (1) What is the performance gains for CREST-VEC compared to CREST model? (2) Does the included subsurface routing improve model performance? (3) Can a simple natural lake simulation improve model performance in a downstream urban area? and (4) How does CREST-VEC model adopt to flood warnings?”

3.2 CONUS simulation

Line 263-289: This discussion using the basin attributes suggests a significant uncertainty in the land surface process, not the routing schemes. Remove it, or explicitly describe the contribution of the CREST-VEC (+subq+lake) to the discussion compared with the results when the same analysis is applied to the original CREST.

Response:

Thanks for your suggestions. As it is also brought up by another reviewer, we deleted this section for brevity.

3.3 How likely are floods falsely detected?

Line 315-317: Previous research also reported the incorporation of a lake scheme mitigates the seasonal variability in the river discharge (Tokuda et al., GMD, 2021).

Response:

Thanks for your suggestions. The reference has been added.

L. 565-566: “Additionally, previous research reported that simulation results with lake module mitigate the seasonal variability of the river discharge (Tokuda et al., 2021).”

Technical Corrections

Line 25: Add "." after "1"

Response: Corrected.

Line 38: "the most time-consuming". The expression is too strong and should be loosened. It is because atmospheric models and (even within the framework of a land surface model) land surface models that account for 2-D groundwater flow are computationally expensive.

Response:

Thanks for your suggestions. We modified this sentence to:

L. 41-42:“It is an inseparable component in hydrologic simulation to redistribute and exchange water between compartments and is also relatively time-consuming”

Line 57: 2011; Yamazaki et al., 2011;) remove the last ";".

Response: Corrected.

Line 92: Remove "MERIT-Hydro" before Lin et al. (2019), which corresponds to Yamazaki et al. (2019)

Response: Corrected.

Line 94: Add the reference for the MERIT DEM (Yamazaki et al., 2017).

Response: Thanks for your suggestions. This reference is added.

Line 118: "rainfall" implies snowfall is excluded?

Response: Thanks for your comments. All types of precipitation are included. We changed “rainfall” to “precipitation” throughout the text.

Line 174: The reservoir operation is incorporated in the mizuRoute but does not the CREST-VEC consider it?

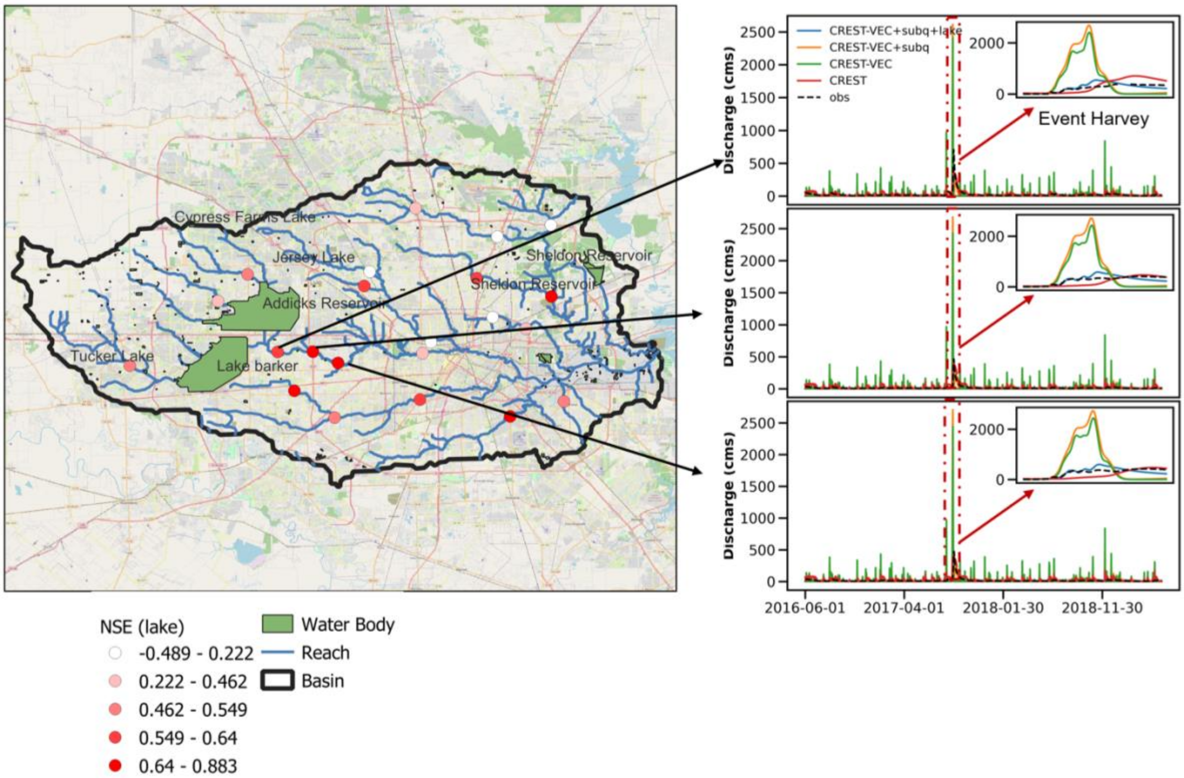
Response: It is included as an option in CREST-VEC, but we did not turn it on in this study for performance analysis because it complicates model configuration.

Line 210: Related to the major comment, why does the CREST-VEC overestimate the peak compared to the original CREST?

Response: Thanks for your comments. We have mentioned three reasons as mentioned in your major comment.

Line 219 (Figure 3): Make the orange lines visible.

Response: Thanks for your suggestions. We replotted this Figure to make it visible.



Line 321: The contribution of the routing scheme to the flood alert system is an essential topic for our society. However, additional analysis is needed since the current analysis is highly affected by the bias of the runoff, and we can not detect the impacts of the new routing and lake schemes from the results.

Response:

Thanks for your comments. We agree that high bias exists in current model settings and such errors can mislead our flood alert system. This work focuses more on the coupled framework which has a potential to replace current operational flood forecast framework. We acknowledge that there are more efforts to be done in the future to correct such bias from atmospheric forcing or model physics. As mentioned previously, we have mentioned our model physics is not perfect – some processes are simplified/missing. Beyond model physics, we need large-scale model calibration efforts to account for streamflow adjustment, as outlined in Discussion Section. But opportunistically, we see an overall improvement in flood prediction as compared to original model.

Line 353-355: Which result (or previous research?) suggests this?

Response:

Thanks for your comments. Our previous sentence is a bit overstating. We have now modified this sentence to:

"In terms of results against observations, the CONUS-wide performance is improved regarding NSE values"

Line 359-361: What is the relationship between the one-to-many river network and the following sentences? What does "on the other hand" mean?

Response:

Thanks for your comments. We intend to show both sides of vector and raster-based routing. On one hand, vector-based routing has xxx deficiencies. One the other hand, raster-based routing has xxx deficiencies. Since this "on the other hand" is confusing, we have deleted it from our main text.

Line 363-368: What is the difference in the sub-grid routing scheme between raster- and vector-based approaches? The section title is "vector vs. raster", but the sub-grid scheme represents the hydrodynamics within each grid, and it has nothing to do with the vector or raster approaches.

Response:

Thanks for your comments. We have deleted this sentence for clarity.

Line 370-396: The paragraphs are too long to discuss the uncertainty caused by the IRF scheme. Make them shorter. (Or, do you have any plan for the calibration with the latest, faster CREST-VEC model?)

Response:

We have shortened this paragraph to only tie to our model results. For future calibration, a feasible approach is through parameter regionalization. It requires careful calibration on a set of unique catchments (by catchment signatures) such as CAMELS dataset. Then, we can transfer calibrated parameters to other regions by feature proximity. Similar work can be found in Mizukami et al. (2017).

L.719-744: "Hydrologic calibration is powerful to boost model accuracy, yet large-scale models oftentimes suffer from the complexity that impedes credible model calibration. River routing schemes and their parameters can affect streamflow simulations especially at fine time scale such as sub-daily (Mizukami et al., 2021). Our current study used IRF scheme in which impulse response function is derived from diffusive wave equation (see Lohman et al., 1996; Mizukami et al., 2016) and includes two parameters – diffusivity and celerity. These parameters need to be included in calibration in addition to the hydrologic model parameters. Furthermore, to fully understand routing model impact on streamflow simulations, it is necessary to consider other routing schemes including diffusive wave as well as kinematic wave."

Line 408 (The section title): Since the reservoir operation is not considered in this experiment, “with a lake scheme” is more appropriate instead of “with regulated flow”

Response:

Thanks for your suggestions. We agree that “with a lake scheme” is more appropriate in this context. We have changed this title to “Towards improved flood forecasting with lake routing”

Line 411-412: Which process should be improved in the current lake scheme? The improvement plan should be discussed in the previous section (4.1).

Response:

The representation of lake routing should and can be improved because it is now simple. A sophisticated lake routing module including more human control should be updated. We have discussed this in Section 4.1.

L. 704-705: “For the lake module, we expect to include more sophisticated multi-layer decision processes instead of a level-pool process. Lake evaporation is another important factor to be considered for improved water balance.”

Line 415-419: Does not this approach cause the “cry wolf” effect?

Response:

This approach still causes “cry wolf” effect, but at a less rate than original flood forecast framework. A large contribution factor is the lake module which by its purpose is to control floods. We see that including lake routing can significantly reduce such “cry wolf” effect.

A list of references mentioned in this response letter:

Mizukami, N., Clark, M. P., Gharari, S., Kluzek, E., Pan, M., Lin, P., Beck, H. E., Yamazaki, D.: A vector-based river routing model for Earth System Models: Parallelization and global applications. J. Adv. Model. Earth Syst., 13, e2020MS002434, https://doi.org/10.1029/2020MS002434, 2021.

Vergara, H., Kirstetter, P., Gourley, J.J., Flamig, Z.L., Hong, Y., Arthur, A., and Kolar, R: Estimating a-priori kinematic wave model parameters based on regionalization for flash flood forecasting in the Conterminous United States, 541, 421-433, doi: 10.1016/j.jhydrol.2016.06.011, 2016.

Mizukami, N., Clark, M. P., Sampson, K., Nijssen, B., Mao, Y., McMillan, H., Viger, R. J., Markstrom, S. L., Hay, L. E., Woods, R., Arnold, J. R., and Brekke, L. D.: mizuRoute version 1: a river network routing tool for a continental domain water resources applications, Geosci. Model Dev., 9, 2223–2238, https://doi.org/10.5194/gmd-9-2223-2016, 2016.

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., Rakovec, O., and Samaniego, L.: Towards seamless large-domain parameter estimation for hydrologic models, Water Resour. Res., 53, 8020– 8040, https://doi.org/10.1002/2017WR020401, 2017.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., and Bates, P. D.: A high-accuracy map of global terrain elevations, Geophys. Res. Lett., 44, 5844–5853, https://doi.org/10.1002/2017GL072874, 2017.