CREST-VEC: A framework towards more accurate and realistic flood simulation across scales

Zhi Li[1], Shang Gao[1], Mengye Chen[1], Jonathan J. Gourley[2,] Naoki Mizukami[3], Yang Hong[1]

[1]School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, OK, 70372, USA
5 [2]NOAA/National Severe Storms Laboratory, Norman, OK, 73072, USA
[3]National Centers for Atmospheric Research, Boulder, CO, 80307, USA

Correspondence to: Yang Hong (yanghong@ou.edu)

Abstract. Large-scale (i.e., continental and global) hydrologic simulation is an appealing yet challenging topic for the
10 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 resource
applications. Towards this goal, we craft two objectives for improving 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. First, in contrast to a
15 conventional fully gridded model, this framework can significantly reduce the computational cost of river routing by at least
ten times, based on experiments at regional (0.07 sec/step vs. 0.002 sec/step) and continental scales (0.35 sec/step vs. 7.2
sec/step). This provides adequate time efficiency for generating operational ensemble streamflow forecasts and even
probabilistic estimates across scales. Second, the performance using the new vector-based routing is improved, with the
median-aggregated NSE (Nash-Sutcliffe Efficiency) score increasing from -0.06 to 0.18 over the CONUS. Third, with the lake
20 module incorporated, the NSE score is further improved by 56.2%, and the systematic bias is reduced by 17%. Lastly, over
20% of the false alarms on two-year floods in the US can be mitigated with the lake module enabled, at the expense of only
missing 2.3% more events. This study demonstrated the advantages of the proposed hydrological modeling framework that
could provide a solid basis for continental and global scale water modeling at fine resolution. Furthermore, the use of ensemble
forecasts can be incorporated into this framework; and thus, optimized streamflow prediction with quantified uncertainty
25 information can be achieved in an operational fashion for stakeholders and decision-makers.

1. Introduction

Flooding all over the world has affected millions of people, especially those who reside in floodplains (Tellman et al., 2021).
In the US, flooding, as the primary cause of billion-dollar weather disasters, costs $3.9 billion monetary losses and 15 deaths
per year over the past four decades according to the NOAA National Centers for Environmental Information U.S. Billion-
30 Dollar Weather and Climate Disasters (2021). In light of frequent flooding in the US, several public agencies have been

operating real-time flood forecasting systems, such as the NOAA NSSL (National Severe Storms Laboratory) FLASH project (https://flash.ou.edu) and NOAA Office of Water Prediction (OWP). However, flood warnings are still either missed or unverified due to uncertainties ranging from precipitation forcing, hydrologic model structure, model parameterization, and/or hydrologic routing. As revealed by Martinaitis et al. (2017), 12.8% of flash floods in the US go completely unwarned per year, let alone falsely warned. Apart from pursuing accurate weather forecasts, improving hydrologic simulation is the key to issuing flood warnings properly.

Flow routing in hydrology is the lateral transport of water on the land surface, subsurface, and in waterways (namely hydrologic compartments). It is an inseparable component in hydrologic simulation to redistribute and exchange water between compartments and is also relatively time-consuming. In a lumped hydrologic model (watershed as an integrated unit), routing can be simplified to convolution in time, such as the Unit Hydrograph (UH) or referred to Impulse Response Function (IRF) (Chow, 1988). However, variable velocities over the land surface and in waterways are difficult to be physically considered. Parameterization is a pragmatic way, but too many parameters could lead to equifinality (Beven, 2006). In addition, only outlet streamflow can be simulated in a lumped model. A semi-distributed model was then born to resolve flow pathways using Digital Elevation Models (Quinn et al., 1991). Owing to ever-increasing computing power, gridded hydrologic models with spatially distributed routing become feasible over large domains (Shaad, 2018). Terrain (or hillslope) routing and river channel routing at grid scales can be explicitly represented in model settings with distributed solvers such as linear reservoirs (Liston et al., 1994; Wang et al., 2011; Shen et al., 2017), kinematic wave model (Vergara et al., 2016), and diffusive wave model (Lighthill & Whitham, 1955; Ponce et al., 1978; de Almeida & Bates, 2013). More recently, vector-based routing has attracted more attention instead of raster-based routing for large-scale (i.e., continental and global extent) simulation. In theory, vector-based routing and raster-based routing differ in defining unit catchments and river networks. For instance, a raster-based routing model discretizes both catchments and river networks on Cartesian coordinates, while a vector-based routing model builds upon the irregular shape of unit catchments (i.e., polygon) and river networks (i.e., polyline).

The pioneering experiment of vector-based routing can be dated back to the early 2000s, in which river network models were incorporated in emerging Geographic Information System (GIS) software (Wang et al., 2000). With the burgeoning availability of global-scale hydrography datasets (e.g., HydroSHEDS and NHDPlus), vector-based routing models have been gaining considerable interest in recent years (David et al., 2011; Lehner & Grill, 2013; Mizukami et al., 2016; Paiva et al., 2011; Yamazaki et al., 2011). Among those developments, three frameworks have become popular and stand out in the hydrologic model community. First, David et al. (2011) introduced the RAPID routing framework that is based on the Muskingum method. The RAPID has been coupled with the National Water Model operated by the NOAA OWP (Office of Water Prediction) (Lin et al., 2018) and the Global Flood Awareness System (GLoFAS) developed by the ECWMF (European Center for Median-Range Weather Forecasts). Second, Yamazaki et al. (2011) developed the CaMa-Flood framework which generates flood inundation at a large scale by solving the 1D diffusive equation and spilling water over floodplains. Third, the recent development of the mizuRoute framework by Mizukami et al. (2016) offers terrain routing and multiple channel routing schemes (e.g., IRF and kinematic wave), making it more physically based compared to RAPID which ignores terrain routing.

65 The mizuRoute has been used together with the hydrologic framework SUMMA (Structure for Unifying Multiple Model Alternatives) (Knoben et al., 2021) and is planned to be implemented in the CESM (Community Earth System Model). These vector-based routing models overcome several challenges for large-scale hydrologic simulations faced by raster-based routing models. First, higher model resolution in raster-based models comes at the expense of higher computational cost, which prohibits global hydrological simulations at tens or hundreds of meters. However, the vector-based routing model is much

70 more scalable and computationally efficient, irrespective of increasing resolution. Second, river networks can be more realistically represented in a vector form. In conventional hydrologic models, the river network in a raster form has to be delineated based on a DEM as a preprocessing step. River networks generated in such a way do not always align well with natural river centerlines. For studies investigating hydrologic connectivity in particular, river grid cells in a raster form can easily become discontinuous without considering river topology. Alternatively, river networks in popular hydrography data

75 are digitalized based on satellite optical imagery and manual inspection (Lin et al., 2021). Another weakness of raster-based routing stems from the traditional D8 flow strategy which means water in the central grid can only be permitted to flow through one of its neighboring grid cells (Tarboton, 1997). On the contrary, vector-based routing offers a more flexible approach from vector representation of river networks.

To date, modern vector-based routing models such as RAPID and mizuRoute have neglected the subsurface routing, which is

80 either assumed to be minimum (Mizukami et al., 2020) or treated 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 have 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 a local scale,

85 but also downstream rivers. One of the functions of lakes and reservoirs in the US is for flood control, so simulation without incorporating such a process is likely to result in falsely issued flood warnings.

 In light of the advantages of vector-based routing, this study introduces a coupled modeling framework CREST-VEC (Coupled Routing and Excess STorage with VECtor routing), which strives to facilitate real-time flood forecasting across scales. This framework seamlessly integrates the current operational flash flood forecast model structure – CREST model and

90 the vector-based routing framework – mizuRoute. We utilize a case study to demonstrate the advantages of this coupled framework and to investigate some updates we made to improve the existing routing scheme. Four questions are posed in this regional case study: (1) What are the performance gains for CREST-VEC compared to the 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? In the second part,

95 we apply this framework to the continental US for a comprehensive evaluation. We ask one additional question: How many floods are falsely alarmed without considering reservoir operations? It is anticipated that findings from this work could motivate the future development of large-scale hydrologic models and raise awareness on whether and how much flood forecasts by model simulations should be trusted without proper representation of lakes.

2. Data and methods

2.1 Hydrography data

In this study, we use the vectorized river network and Hydrologic Response Unit (HRU) dataset derived from the high-accuracy Multi-Error-Removed-Improved-Terrain (MERIT) Hydro hydrography dataset (Yamazaki et al., 2017, 2019). The flowlines were created from the 90-meter DEM data (MERIT DEM), covering the full global land surface (60°S-90°N). A minimum channelization threshold of 25 km^2 (upstream area) was applied to restrict river channel grid cells in the MERIT Hydro dataset. The HRUs were processed along with flowlines by the TauDEM software and trimmed with the HydroBASINS level-II boundaries. Detailed processing of the hydrography data is listed in Lin et al. (2019). This set of hydrography data has been validated against 30-year Landsat imagery (Lin et al., 2021) and empowered the global reconstruction of historical streamflow (Lin et al., 2019; Yang et al., 2021). Over the CONUS, we have obtained 341,921 river reaches and the same amount of unit catchments for the routing component.

Lakes and reservoirs in the U.S. play a significant role in regulating streamflow (Tavakoly et al., 2021). Major river basins (e.g., Mississippi and Columbia River Basins) are highly regulated, as shown in Fig.1a and 1b, results obtained from Lehner et al. (2011). The HydroLAKES dataset provides a global catalog of lake polygons and pour points that can be easily integrated into hydrologic models (Messager et al., 2016). Over one million natural lakes and constructed reservoirs were identified globally, with a minimum surface area larger than 10 ha. Over the U.S., there are 96,874 lakes recorded in the HydroLAKES data, of which 94,865 are natural lakes without human intervention, and 1,992 (17) lakes are reservoirs (regulated lakes), as shown in Fig.1c. Of regulated lakes or reservoirs, 20.0% are primarily used for irrigation, 19.9% for hydroelectricity, 17.6% for water supply, 17.2% for flood control, 14.1% for recreation, 1.9% for navigation, 0.7 for fisheries, and 8.6% for others (Fig.1d). The total lake volume, estimated from the lake bathymetry, is a required field in our modeling framework to approximate outflow.

[INSERT FIGURE 1 HERE]

Figure 1. Maps of (a) percent of regulated river and (b) regulated lake volume; (c) bar plot of lake classifications; (d) pie plot of US regulated lake or reservoir function purposes.

2.2 Forcing data

Forcing data is required as model inputs to drive the hydrologic model. Hourly precipitation rates are obtained from the Multi-Radar Multi-Sensor (MRMS) data, operated at the NOAA NSSL (Zhang et al., 2016). The MRMS is a state-of-the-art radar-gauge merged product, providing instantaneous rates at a 1-km spatial resolution over the CONUS and parts of southern Canada and northern Mexico. We used the one-hour accumulated and gauge-corrected precipitation product in this study for streamflow simulation. The performance and hydrologic utility of MRMS data has been corroborated in previous studies (Li et al., 2020, 2021). The daily temperature from the PRISM (Parameter-elevation Relationships on Independent Slopes Model) is used to simulate snow accumulation and melt (PRISM Climate Group, 2014). The PRISM team routinely collects meteorological data from meteorological stations over the U.S. and interpolates them into 4-km gridded data based on the

elevation dependence (Daly et al., 2008). The potential evapotranspiration (PET) data is obtained from the USGS FEWS data port (https://earlywarning.usgs.gov/fews) at daily and 1° spatial resolution (Allen et al., 1998). Forcing data at different spatial resolutions is re-gridded to a 1-km model resolution. All of these data are collected from the simulation period in complete calendar days from 2015 to 2019.

2.3 CREST model

As jointly developed by the University of Oklahoma (OU) and NASA, the CREST model has been released for a decade (Wang et al., 2011). It is a distributed hydrologic model whose primary purposes are (1) flood simulation and forecasting, (2) evaluating the hydrologic utility of satellite precipitation datasets, and (3) water resources management (Xue et al., 2013; Tang et al., 2016; Gourley et al., 2017; Gao et al., 2021; Li et al., 2021; Chen et al., 2022). Owing to its relatively simple structure and computationally efficient simulation, the CREST model has been promoted by the NOAA NSSL for real-time flash flood forecasting over the continental U.S. and its territories (Gourley et al., 2017; Flamig et al., 2020). As shown in Fig. 2, the effective rain (deficit of rainfall rates and evaporation rates) reaches the land surface and is partitioned into fast runoff from urban impervious area ratio and infiltration into the soils. A Variable Infiltration Curve (VIC) model is incorporated to determine the infiltration rate (Liang et al., 1994). Surface runoff is generated when infiltration rates become higher than the maximum infiltration capacity. In the meantime, slow-flowing interflow is produced while soil water content is depleted. In the CREST model, flow routing is handled in two ways. Terrain routing and in-channel river routing are done by the kinematic wave model which simplifies the Saint-Venant equation by ignoring the acceleration and forcing terms (Vergara et al., 2017). The interflow is routed by a conceptual linear reservoir with parameterized velocity (Shen et al., 2017). We refer to the CREST model hereafter as a standalone package that couples the water balance model with gridded terrain and channel routing. The original code is written in C++.

To account for snowmelt, we coupled the original CREST model with the Snow-17 model, which is part of the National Weather Service River Forecast System in the U.S (Franz et al., 2008). The Snow-17 model is a conceptual snowmelt scheme that simulates snow accumulation and ablation based on temperature and precipitation as inputs (Anderson, 2006). Although the physics behind it is not as comprehensive as the energy balance model, Snow-17 is advantageous for having less required input data and performing "at least as good as" energy-based models (Ohmura, 2001).

2.4 mizuRoute

The mizuRoute river routing model, developed at the NCAR (National Center for Atmospheric Research), is a vector-based routing framework that incorporates both terrain and channel routing for large-domain river routing applications (Mizukami et al., 2016, 2021). For the terrain routing, the IRF or UH is used with parameters associated with gamma distribution to adjust the shape and scale. For the channel routing, user-defined options are IRF, kinematic wave with Lagrangian solution, and kinematic wave with Euler solution. A recent version of mizuRoute (Version 2.0.1) includes two lake routing schemes (Gharari et al., 2022; Vanderkelen et al., 2022) – one based on Döll et al. (2003) with a simple level-pool equation for natural lakes and

the other more complicated one based on Hanasaki et al. (2006) which includes reservoir operation rules. These two schemes have been applied to the other global hydrologic models (e.g., WaterGAP, VIC, and CWatM) to account for regulated streamflow. The original code is written in Fortran.

The current version of mizuRoute does not explicitly account for subsurface runoff routing over terrain, which is critical in the Great Plains and regions where streams are intermittent across a year (Salas et al., 2017). In this study, we enable an option to turn on or off subsurface routing as defined in the model configuration file. Similar to surface runoff routing, the subsurface flow is routed using the IRF scheme but with a much slower velocity and reduced magnitude. We use a two-parameter Gamma distribution function to materialize the IRF method as shown in eq. 1.

$$y(t) = \frac{1}{\Gamma(a)\theta^a} t^{a-1} e^{-\frac{t}{\theta}} \tag{1}$$

Where t is the time variable, a is a shape parameter, and θ is a time-scale parameter. Both a and θ determine the flood peaking time and flashiness. After calculating instantaneous rates based on the gamma function, we use convolution to compute flow rates Q at time t. $R(t-s)$ is the (sub)surface runoff at the time $(t-s)$, and s is an increment of time from 0 to $tmax$ (also denoted as the time window). The default values of a and θ for hillslope surface routing are set to 2.5 and 8000. For subsurface flow routing, the a and θ are 10 and 86400, respectively.

$$Q(t) = \int_0^{tmax} y(t) \times R(t-s) ds \tag{2}$$

2.5 CREST-VEC

The framework, CREST-VEC, and the difference compared to its precedent CREST model are shown in Fig. 2. The main difference comes from the routing process, where the original CREST model routes surface flow and interflow via a kinematic wave routing model and a conceptual linear reservoir model in a gridded manner. However, the CREST-VEC model requires area-averaged time series of surface and subsurface flow at each river reach to be separately routed downstream. The gridded outputs from the CREST model (i.e., surface runoff and subsurface runoff) are extracted and averaged over each unit catchments or HRU using the newly developed Python package EASYMORE (EArth SYstem Modeling REmapper), publicly available from https://github.com/ShervanGharari/EASYMORE.

The framework is loosely coupled with two models written in different programming languages. A bash file calls three executables after model compilation subsequently (CREST-EASYMORE-mizuRoute). The input files for this model chain include forcing data (gridded precipitation, potential evaporation, and temperature), topography data (gridded digital elevation model, flow direction, flow accumulation, river network topology, and hydrologic response unit), and configuration files. The topography data can be accessed from the HydroSHEDS website which consists of grid-based and vector-based topography data.

We use the IRF scheme in this study for both terrain routing and channel routing in this study and activate the lake model with the Döll et al. (2003) lake model. The parameters for lake parameters such as the outflow coefficient a and exponent b of eq.3, are based on suggested values in Döll et al. (2003) and Gharari et al. (2022). For lakes that have monitored storage provided

by the US Geological Survey (USGS), we directly insert storage time series into the model. As reservoir operation is not considered in this study, we exclude observed streamflow that is regulated by reservoirs and regulated lakes, as shown in Fig. 1c. So, only results from natural lakes, which account for 98% of US lakes or reservoirs, are considered valid for statistical comparison. To initialize model states, especially for initial lake volumes, we warm up the CREST-VEC model from 1948 to 2014 using the GLDAS forcing (Global Land Data Assimilation System) at a daily time step.

$$Q_{out} = a \times S_f \times (S_f/S_{f,max})^b, \tag{3}$$

where a and b are the outflow coefficient (1/day) and exponent, respectively; S_f is the actual lake storage (m^3); $S_{f,max}$ is the maximum lake storage (m^3).

[INSERT FIGURE 2 HERE]

Figure 2. Schematic view of the *CREST-VEC* framework. The red arrow highlights the newly added subsurface routing option to the original mizuRoute framework.

3. Results

3.1 Case study: Houston region

As mentioned in the objectives of this study, we first conduct a case study analysis to assess the relative contributions of subsurface flow routing and lake routing to streamflow simulation based on the CREST-VEC framework. The original CREST model is used as a benchmark. We chose the Houston region (Fig. 3a) because there are two large natural lakes - Lake Barker and Lake Addicks that impact hydrologic simulations (Fig. 1a). For the CREST model with gridded routing, we calibrate the model using the DREAM (Differentiable Evolution Adaptive Metropolis) optimizer (Vrugt et a., 2009) from 2016-06-01 to 2017-06-01 at an hourly time step and performed evaluation from 2017-06-01 to 2020-01-01. The NSCE is used as the objective function for calibration, and the model is warmed up for one year from 2016-06-01 to 2017-06-01. We run the CREST model at three spatial resolutions: 1 km, 250 m, and 90 m. To be comparable with CREST-VEC simulations, whose hydrography data is built upon a 90-meter resolution DEM, we only use CREST model results at 90 meters for statistical comparisons and use the results at 1 km and 250 meters to assess computational efficiency. The evaluation metrics shown in Fig. 3c are based on the evaluation period. The river flows from 22 stream gauges are curated from the USGS.

3.1.1 Model speedup

Figure 3b shows the computational cost (elapsed time at seconds per step) for a series of model configurations for the routing process. All the tests were run on a single core Intel i7-6700K CPU (4.00 GHz). The grid-based CREST model costs 0.01, 0.08, and 0.12 seconds per step at 1-km, 250-m, and 90-m resolutions, respectively. However, the CREST-VEC model can reduce this to approximately 0.002 seconds per step, regardless of grid resolutions from forcing data. There is little difference among the three scenarios (i.e., CREST-VEC, CREST-VEC+subq: CREST-VEC plus subsurface routing, and CREST-VEC+subq+lake: CREST-VEC with subsurface routing and lake routing). Relatively speaking, CREST-VEC can speed up the current operational CREST model at 1-km by 10x, let alone at finer resolutions.

3.1.2 Performance improvement

Regarding model skills, the CREST model and CREST-VEC achieve similar median NSE (Fig. 3c) based on observations

230 from 22 stream gauges, even though the CREST model takes advantage of automatic calibration. CREST-VEC and CREST-VEC+subq overestimate flows downstream of two natural lakes, resulting in poor scores. But after incorporating lake routing schemes, the CREST-VEC+subq+lake model achieves not only better median scores but also less spread (quantified by the interquartile range). Notably, both CREST-VEC+subq and CREST-VEC+subq+lake have positive NSE values and smaller uncertainty ranges, primarily owing to included subsurface routing. The time series in Fig. 4 highlights the model performance

235 at three stream gauges affected by upstream lakes. The CREST-VEC overestimates streamflow by a considerable amount (i.e., three times higher than observation in Hurricane Harvey), resulting in low NSE scores: 0.11, 0.16, and 0.18, respectively. With lake routing considered in the CREST-VEC+subq+lake, the simulated streamflow aligns well with observations, achieving NSE scores of 0.61, 0.65, and 0.64, respectively. Although the CREST model captures streamflow magnitude after calibration with the NSE scores – 0.37, 0.52, and 0.54, the peak timing is at least one-day delayed for Hurricane Harvey. In summary, the

240 advantages for the general CREST-VEC framework against the gridded CREST model are threefold: (1) improve computational efficiency by at least ten times, (2) improve overall model skill, (3) reduce uncertainty ranges.

[INSERT FIGURE 3 HERE]

Figure 3. (a) Map of the study area (Houston region) showing river networks and water bodies; (b) Computation time per step for CREST at three resolutions and CREST-VEC model at four configurations on the x-axis; (c) Nash-Sutcliffe efficiency values for
245 **CREST and CREST-VEC model.**

[INSERT FIGURE 4 HERE]

Figure 4. Performance of models downstream of two lakes. The Nash- Sutcliffe Efficiency coefficients are obtained from the CREST-VEC model with lake routing and subsurface routing. Three plots of time series of stream gauges (from upstream to downstream: 08073500, 08073600, 08074000) are pointed aside by the map, and the Hurricane Harvey event is highlighted in red box and insets.

250 ### 3.2 CONUS simulation

Moving towards continental-scale hydrologic simulation, the CREST-VEC model excels at reducing computational costs, leaving room for quantifying uncertainties from forcing, model structure, and parameters in real-time. The following question is whether and how much the new lake routing improves a continental simulation. To answer this question, we simulate CREST-VEC with and without lake routing over the CONUS from 2017-06-01 to 2020-01-01 at an hourly time step. Notably,

255 subsurface routing is activated for both models with and without lake routing, so we expect the difference in results to be primarily due to lake simulation. Streamflow data from 5,350 stream gauges in the same period are collected and used for model verification. For this case, the CREST-VEC model parameters are based on the pre-configured CONUS-wide parameters, the same as the ones used in Flamig et al. (2020).

3.2.1 Model speedup

260 Table 1 lists model performance with respect to total computational costs and evaluation scores of streamflow simulation. CREST-VEC certainly improves streamflow simulation not only via a higher resolution (from 1-km to 90-m) but with faster computational speed (149.2 hours to 29.9 hours in total; 7.2 sec/step to 0.37 sec/step for routing step only). Considering all

preprocessing steps altogether, CREST-VEC model is still at least four times faster than the original framework. To be noted, a considerable amount of time is spent on mapping gridded runoff data to a vector form (>50% of the time). Future attention should be drawn on how to optimize the efficiency while preserving certain degrees of accuracy for this process.

3.2.2 Performance improvement

The median NSE score has increased from -0.06 (gridded) to 0.12 (no lake) and 0.18 (lake). The fraction of gauges with positive NSE scores has been improved from 41.8% (gridded CREST) to 50.6% (CREST-VEC without lake) and to 56.2% (CREST-VEC with lake). However, the CREST-VEC results are more biased than gridded CREST, partly due to the systematic overestimation of streamflow by the IRF routing scheme in the CREST-VEC. The difference would be primarily attributed to the different routing processes, as CREST permits leakage in the interflow reservoir, thereby leading to lower positive bias. Results with lake simulation have reduced BIAS from 27% to 17%, as part of the water is being held in the lake. The CC (Correlation Coefficient), however, does not vary much between scenarios with and without lake simulation, as shown in Fig. 5. One of the reasons is that the CREST-VEC model does not simulate regulated lakes or reservoirs which have strong control of streamflow time shifts. Notably, the IQRs (interquartile ranges) of NSE and BIAS for lake simulation are lower than without lakes, meaning that this method particularly boosts scores at gauge locations that had poor performance previously.

Figure 6 depicts the spatial map of model skill (with lake) and its difference between scenarios with and without lake simulation. CREST-VEC with lake module in regions like the West Coast and Upper Mississippi River Basin have relatively good performance (NSE>0.4), yet over the Great Plains and East Coast, the model bias is high (BIAS>1), yielding low NSE scores. Similar issues are found in the literature with other models (Clark et al., 2008; Newman et al., 2015; Mizukami et al., 2017; Salas et al., 2017; Lin et al., 2019; Konben et al., 2020; Knoben et al., 2020; Yang et al., 2021; Tijerina et al., 2021). Taking the Great Plains as an example (highlighted box in Fig. 6c), the model physics of CREST-VEC does not correctly represent the real hydrologic processes by two means. First, the surface runoff (before routing) simulated by CREST-VEC is biased. We compare the annual surface runoff by CREST-VEC to the public community dataset GRFR (Global Reach-level Flood Reanalysis) in Fig. 7. The runoff in GRFR is simulated by the VIC model and undergoes stringent bias correction against observations via the discrete quantile mapping technique (Yang et al., 2021; Lin et al., 2019). There is a 116.3% higher surface runoff by the CREST-VEC than theirs, partly explaining the high BIAS and low NSE scores in such region. We suspect the singular bulk soil layer represented in the CREST model yields such systematic differences. Second, the missing representation of playas, small and rain-fed lakes that are prominent in the Great Plains, leads to falsely produced runoff (Hay et al., 2016; Solvik et al., 2021). However, even when accounting for multiple hydrologic model structures, performance in this region is still ranked as one of the poorest (Clark et al., 2008; Knoben et al., 2020). For example, Knoben et al. (2020) analyzed 36 hydrologic models over the U.S., in which the maximum KGE scores out of those models are lower than 0.5 over the Great Plains.

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%

[INSERT FIGURE 5 HERE]

Figure 5. Boxplot of model performance comparing results with lake routing and without it.

[INSERT FIGURE 6 HERE]

Figure 6. Spatial map of model performance with the lake (left column) and the difference between with and without lake simulation (right column). (a): NSE scores; (b) NSE differences (results with lake minus results without lake); (c) BIAS; (d) BIAS difference; (c) Correlation Coefficient (CC); (f) CC difference. The blue box in (c) highlights the region where high positive BIAS is present.

[INSERT FIGURE 7 HERE]

Figure 7. (a) density plot of CREST-VEC simulated annual surface runoff against GRFR (Global Reach-level Flood Reanalysis) in the Great Plains; (b) cumulative density function (CDF) of CREST-VEC and GRFR simulated annual runoff.

3.3 How likely are floods falsely detected?

In this section, we shift gears to explore how likely US floods are falsely alarmed if no lake simulations are included. We selected 283 gauges that are downstream of natural lakes (Fig. 8), with most of them located in the middle and eastern US. The hourly time series of streamflow of those gauges are compared against advised flood thresholds (2-year flooding) provided by the US Geological Survey. They fit a log-Pearson III type distribution to the annual maxima streamflow from long-term records and extract values with given flood frequency. Following a similar approach as in Yang et al. (2021), consecutive yet independent events have to be two days apart from one another. From there, we calculated the Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI) based on the contingency table.

As expected, median FAR is reduced from 0.63 (without lake simulation) to 0.50 (with lake), resulting in a slightly higher CSI of 0.36 than that of 0.31 for no lake simulation (Fig. 8a). Additionally, previous research reported that simulation results with the lake module mitigate the seasonal variability of the river discharge (Tokuda et al., 2021). The decrease in FAR values implies five instances: (1) decrease in false alarms while hits remain the same; (2) increase in hits while false alarms remain

the same; (3) decrease in false alarms while increase in hits; (4) decrease in both false alarms and hits; and (5) increase in both false alarms and hits. We find that however POD values decrease from 0.87 without lakes to 0.85 with lakes, from which we

320 can infer that both hits and false alarms are decreasing, but false alarms decrease at a higher rate. That is a fact of reducing simulated flood peak, which results in fewer hits in flood forecasts but meanwhile less falsely alarmed floods. As most studies focus on flood detection, they inevitably arrive at more falsely detected floods. Too many false alarms could make people disregard the warnings, despite a real threat, causing the "cry wolf" effect.

Maps in Fig. 8b display the distributions of flood detectability with lake simulation and its improvements compared to results

325 without lake simulation. High POD and FAR values co-exist in the Great Plains, where the model simulates considerably higher streamflow values than observations. Moderate FAR values are found near the Florida panhandle and parts of Georgia. Lower FAR values are found in the Midwest and West Coast. Compared to results without the lake, FAR values are reduced reasonably over the East Coast, Midwest, Gulf Coast, and West Coast, although POD values remain relatively unchanged or even decreased.

330 Five local cases are shown in Fig. 9, which depicts the river topology and time series of hourly streamflow. One can infer that these lakes are not heavily regulated from recorded streamflow time series, therefore showing the effectiveness of our model. In Fig. 9a, the simulated streamflow without lakes is heavily overestimated, peaking at 1200 cms in the year 2017, whereas the actual flow rate is around 400 cms. The scenario with lake simulation, however, produces a magnitude much closer to the observation. Due to decreased systematic bias, the lake scenario boosts the NSE score from -0.2 to 0.5. There is also an 8%

335 less chance of issuing false alarms than the model without lake simulation. Figure 9b shows a case where FAR is reduced from 0.70 to 0.17, a reduction rate of 75.7%. The flood detectability, i.e., CSI, is greatly improved from 0.29 to 0.57. Figure 9c exemplifies a case with all improved metrics (i.e., NSE, POD, FAR, and CSI). All these three cases in Figs. 9a-c are located along the St. Johns River, in which we expect a systematic improvement along this river after incorporating the lake simulation. Figure 9d displays more common cases where a reduction of FAR comes at the expense of reducing POD (i.e., flood detection),

340 almost at the same pace. Figure 9e shows that although the model with lakes produces better baseflow, it underestimates flood peaks, resulting in lower NSE values (0.3) than results without lakes (0.4). It implies that parameters governing the lake outflow need to be improved.

[INSERT FIGURE 8 HERE]

Figure 8. Flood detection performance comparing lake and no lake simulation. (a) Similar to Figure 5, but for flood detectability;

345 **(b) Similar to Figure 6, but for flood detectability.**

[INSERT FIGURE 9 HERE]

Figure 9. Five case examples of streamflow time series at gauges downstream of lakes: (a) St. Johns River near Sanford, FL; (b) St. Johns River near Cocoa, FL; (c) St. Johns River near De Land, FL; (d) Big Muddy River at Plumfield, IL; (e) Mississippi River at Clinton, IA.

350

4. Discussion

4.1 Vector vs. Raster routing

In this study, we compare the advantages of vector-based routing with respect to conventional raster-based routing in two aspects: (1) model efficiency and (2) model accuracy. Overall, the vector-based routing shows great promise, as it speeds up the routing process by at least ten times, compared to grid-based routing, for both the regional simulation (0.07 sec/step vs. 0.002 sec/step) and the CONUS simulation (0.35 sec/step vs. 7.2 sec/step). In terms of results against observations, the CONUS-wide performance is improved regarding NSE values. However, the variable river reach lengths (from hundreds of meters to tens of kilometers) in large-scale simulation pose challenges for estimating routing parameters such as the time and shape parameters in a unit hydrograph. Second, most land surface models are still grid-based, making a type mismatch (grid-based land surface model vs. vector-based routing model) (Lehner & Grill, 2013). To integrate the two, we need a processing step by mapping surface and subsurface runoff onto representative HRUs. Different aggregation strategies are present and subject to the primary purpose of interest. At present, there is an ongoing effort to seamlessly integrate these two processes together (Gharari et al., 2020). However, it is yet to be efficient and draws further attention to improving this mapping scheme. Third, the many-to-one river network is established but not for one-to-many, meaning river bifurcation is challenging to represent and tackle (Yamazaki et al., 2014).

Raster-based routing comes at the resolution of the input DEM data, albeit at a slower computational speed. Having matured over the years, most raster-based routing models are seamlessly integrated with water balance models so that model can be set up at minimum effort by a modeler. Mostly derived from hydrodynamic models, the concept of "raster" can be extended to "grid" because of the emerging unstructured grids such as triangles, curvilinear, hexagons, etc. These flexible grid types mimic the real flow directions and reduce computational costs. However, they are yet to be accepted/applied by the hydrologic model community.

As one objective of this study, we want to examine the potential improvement from the with-lake configuration on streamflow simulation over a wide range of hydrometrical and geographical settings in the CONUS, rather than provide some optimal model setup and parameterization at the CONUS scale, which we believe is way beyond our scope and several steps forward from the current CREST-VEC or any existing CONUS models. As far as what qualifies 'an adequate base simulation', there may be some room for debate but should be some bottom-line principles: first, one should be clearly aware of the sources of uncertainties, including forcing, model structure, parameterization, streamflow observation as the reference, etc. Optimization, though effective in improving the model performance, compensates for uncertainties from the other sources simply via adjusting model parameters. This has been acceptable for operational purposes but is not appropriate for this study where a modification of the model structure is introduced. Instead, we use an a-priori parameter set that was developed based on remote sensing datasets and also evaluated at the CONUS scale (Vergara et al., 2016). The physical base of these a-priori parameters set a solid foundation for examining the new with-lake configuration, thus should not be compromised via parameter tunning.

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 modeling, snow melting, hillslope routing, in-channel river routing, and reservoir operations are simplified, and some processes such as vegetation and groundwater modeling are missing from the current version. 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. Since it is a compromise between model complexity and efficiency, we hope to continuously push the envelope on this front to optimize the real-time flood forecast system.

4.2 Room for improving large-scale hydrologic simulation

Large-scale hydrologic simulation is still a long-standing challenge for the hydrologic community, especially with debates on developing a "one-model-fits-all" structure or a "malleable" structure (Burek et al., 2020; Clark et al., 2015a; Fenicia et al., 2011; Savenije, 2008). The CREST model, in our study, systematically overestimates surface runoff over the Great Plains and Southeast, a result of some misrepresented or missing processes, yet excels in flash flood simulation. Diverse hydrologic model structures, on the other hand, hope to overcome individual limitations and offer joint benefits (Horton et al., 2021). We, therefore, promote the "malleable" model structure from the efficiency point of view - a flexible structure disables redundant hydrologic processes. Then, the central question becomes: How do we adapt the model to variable catchment processes? In such a context, intercomparisons and discussions of different hydrologic models in varying catchment processes become particularly valuable (Clark et al., 2015a; Knoben et al., 2020; Tijerina et al., 2021). Notably, simply relying on the NSE or KGE score to assess the model performance can be misguiding (Clark et al., 2021).

Hydrologic calibration is powerful in boosting 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 scales such as sub-daily (Mizukami et al., 2021). Our current study used an IRF scheme in which the impulse response function is derived from a diffusive wave equation (see Lohman et al., 1996; Mizukami et al., 2016) and includes two parameters – diffusivity and celerity. These parameters need to be exposed to calibration in addition to the hydrologic model parameters. Furthermore, to fully understand the routing model's impact on streamflow simulations, it is necessary to consider other routing schemes including a diffusive wave model as well as a kinematic wave model, which may be suited for flood forecasting.

Lastly, the computational costs for large-scale simulation can be optimized from accelerated hardware (multi-core CPUs and GPUs) once codes are parallelized and scalable. Advances in Reduced Order Modeling (ROM), a surrogate model which develops a parsimonious solution to replace the computationally intensive part, hold promise to reduce costs (Clark et al., 2015b). For instance, to integrate reservoir simulation into the CREST-VEC system, we can build an offline ML model which is promising in mimicking human decisions (Yang et al., 2021) and plug it into the system.

4.3 Towards improved flood forecasting with lake routing

Flood forecasts are difficult because of their rarity, and their hits and misses are typically low while false alarms are high (Bartholmes et al., 2009; Cloke & Pappenberger, 2009). Results in this study demonstrate a dilemma in which the model with a lake module reduces false alarms but at the cost of more missed flood events compared to the one without a lake module. Although the combined metric CSI has a certain degree of improvement, this leaves a question – should we reduce a large number of false alarms at the expense of missing a small number of real events? Before discussing this point, we acknowledge that the current lake routing process is simple and imperfect, and improvement in this process possibly leads to an optimal situation where both false alarms and misses can be improved. However, in most situations, tradeoffs exist in hydrologic predictions. A good strategy in our case would be running both simulations with and without the lake module concurrently and making the "without lake" results the worst-case scenario. Since the CREST-VEC model has the advantage of efficiency, running two scenarios is totally feasible. A decision-maker can be trained to assess the situation – results from two scenarios disagree – from the perspective of flood severity and consequences.

5. Conclusion

This study compares a conventional raster-based routing scheme with the emerging vector-based routing approach in hydrologic models for regional case and continental simulations. From the continental run, we demonstrate the improvement in streamflow simulation after incorporating the lake storage and release module. Last but not least, flood-related false alarms can be greatly reduced by including the lake module. The following points summarize the primary findings of the study:

1. Vector-based routing can accelerate continental-scale river routing by up to ten times, compared to a grid-based routing, for both a regional case (0.07 sec/step) and a continental case (0.002 sec/step). This leaves adequate room for generating ensemble predictions with variable forcing, parameters, and/or model structures. Furthermore, it improves streamflow simulation from -0.06 to 0.18, according to the aggregated median NSE values.

2. A newly developed lake model increases the NSE score by 56.2% and reduces systematic BIAS by 17% for the continental simulation.

3. Flood false alarm ratios can be mitigated by 20.6% after enabling the lake module at the expense of missing 2.3% more floods on a continental scale.

We recommend the use of ensemble simulations stemming from different model structures to overcome and adapt to varying catchment processes. Optimized streamflow prediction with quantified uncertainty information can be achieved in an operational manner for stakeholders and decision-makers. Future studies can fully investigate the limitation and uncertainty of different forcing, parameters, and/or model structures to catchment signatures such as climatology, dominant hydrologic processes, lithology, etc. Vector-based routing, in such a context, can enable fair comparison by excluding the effect of different routing schemes while focusing on discrepancies in water balance models alone. For future work, we hope to have

the best possible model-simulated streamflow product in the US, fused with multi-model structures and observations. Another direction is to improve current lake and reservoir outflow simulation with a hybrid model – process-based and ML-based.

Acknowledgement

The first author is sponsored by the University of Oklahoma Hydrology and Water Security (HWS) program
450 (https://www.ouhydrologyonline.com/) and Graduate College Hoving Fellowship.

Code and data availability

The CREST-VEC model code is publicly available in Zenodo: https://doi.org/10.5281/zenodo.6305817 (Li, 2022) (last access: February 28, 2022). The MRMS radar-based precipitation data is archived at the Iowa State University https://mesonet.agron.iastate.edu/archive/ (last access: July 3, 2021). The daily temperature data PRISM is accessed from the
455 Oregon State University: https://prism.oregonstate.edu (last access: September 16, 2020). The hydrography MERIT-Hydro river network data (ESRI shapefile) available at https://www.reachhydro.org/home/params/merit-basins (last access: August 10, 2021). The HydroLAKES Version 1.0 global lake data is downloaded at https://www.hydrosheds.org/page/hydrolakes (last access: August 11, 2021)

Author contributions

460 Z.L., Y.H., and J.J.G. conceived this study; Z.L. and N.K. implemented the model code and simulated the model in this study; Z.L., S.G., and M.C. designed the experiments and analyzed the data; Z.L. wrote the original draft; All co-authors contributed to reviewing and revising the manuscript.

References

Allen, R.G., Pereira, L., Raes, D., and Smith, M.: Crop Evapotranspiration, Food and Agriculture Organization of the United
465 Nations, Rome, Italy. FAO publication 56. ISBN 92-5-104219-5. 290p, 1998.

Anderson, E. A.: Snow accumulation and ablation model–SNOW-17. US National Weather Service, Silver Spring, MD, 61, 2006.

Bartholmes, J., Thielen, J., Ramos, M., and Gentilini, S.: The European flood alert system EFAS – Part 2: statistical skill assessment of probabilistic and deterministic operational forecasts. Hydrol. Earth Syst. Sci., 13 (2) 141–153, 2009.

470 Beven, K., and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol., 249(1-4), 11-29, 2001.

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320, 18-36, https://doi.org/10.1016/j.hydrol.2006.07.007, 2006.

Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., Guillaumot, L., Zhao, F., and Wada, Y.: Development of the Community Water Model (CWatM v1.04) – a high-resolution hydrological model for global and regional assessment of integrated water resources management, Geosci. Model Dev., 13, 3267–3298, https://doi.org/10.5194/gmd-13-3267-2020, 2020.

Chen, M., Li, Z. and Gao, S.: Multisensor Remote Sensing and the Multidimensional Modeling of Extreme Flood Events. In Remote Sensing of Water-Related Hazards (eds K. Zhang, Y. Hong and A. AghaKouchak). https://doi.org/10.1002/9781119159131.ch5, 2022.

Chow, V. T., Maidment, D. R., and Mays, L. W.: Applied Hydrology: McGraw-Hill Series in Water Resources and Environmental Engineering, McGraw-Hill, Inc., New York, 1988.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44, W00B02, https://doi.org/10.1029/2007WR006735, 2008.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski D., Rupp, D. E., Woods, R. A., Freeer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger G. N., and Marks, D. G.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resour. Res., 51, 2498 – 2514, https://doi.org/10.1002/2015WR017198, 2015a.

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P., Kumar, M., Leung, L. R., Mackay, D. S., Maxwell, R. M., Shen, C., Swenson, S. C., and Zeng, X.: Improving the representation of hydrologic processes in Earth System Models, Water Resour. Res., 51, 5929– 5956, https://doi.org/10.1002/2015WR017096, 2015b.

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., Shervan, G., Freer J. E., Whitfield, P. H., Shook, K. R., and Papalexiou, S. M.: The abuse of popular performance metrics in hydrologic modeling. Water Resour. Res., 57, e2020WR029001. https://doi.org/10.1029/2020WR029001, 2021.

Cloke, E. and Pappenberger, F.: Ensemble flood forecasting: A review. J. Hydrol., 375, 613-626. https://doi.org/10.1016/j.jhydrol.2009.06.005, 2009.

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., and Pasteris, P. A.: Physiographically-sensitive mapping of temperature and precipitation across the conterminous United States. Int. J. Climatol., 28, 2031-2064, 2008.

David, C. H., Maidment, D. R., Niu, G., Yang, Z., Habets, F., & Eijkhout, V.: River Network Routing on the NHDPlus Dataset, J. Hydrometeorol., 12(5), 913-934, 2011.

de Almeida, G. A. M., and Bates, P.: Applicability of the local inertial approximation of the shallow water equations to flood modeling, Water Resour. Res., 49, 4833– 4844, https://doi.org/10.1002/wrcr.20366, 2013.

Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation. J. Hydrol., 270, 105-134. https://doi.org/10.1016/S0022-1694(02)00283-4, 2003.

Fenicia, F., Kavetski, D., and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47, W11510, https://doi.org/10.1029/2010WR010174, 2011.

Flamig, Z. L., Vergara, H., and Gourley, J. J.: The Ensemble Framework For Flash Flood Forecasting (EF5) v1.2: description and case study, Geosci. Model Dev., 13, 4943–4958, https://doi.org/10.5194/gmd-13-4943-2020, 2020.

510 Franz, K. J., Hogue, T. S., and Sorooshian, S.: Operational snow modeling: Addressing the challenges of an energy balance model for National Weather Service forecasts. J. Hydrol., 360(1-4), 48-66, 2008.

Freeze, R. A., Role of subsurface flow in generating surface runoff: 2. Upstream source areas, Water Resour. Res., 8, 1272–1283, doi:10.1029/WR008i005p01272, 1972.

Gao, S., Chen, M., Li, Z., Cook, S., Allen, D., Neeson, T., Yang, T., Yami, T., and Hong, Y.: Mapping dynamic non-perennial
515 stream networks using high-resolution distributed hydrologic simulation: A case study in the upper blue river basin. J. Hydrol., 600, 126522, 2021.

Getirana, A. C. V., Boone, A., Yamazaki, D., Decharme, B., Papa, F., and Mognard, N.: The Hydrological Modeling and Analysis Platform (HyMAP): Evaluation in the Amazon Basin, J. Hydrometeorol., 13(6), 1641-1665, 2012.

Gharari, S., Clark, M. P., Mizukami, N., Knoben, W. J. M., Wong, J. S., and Pietroniro, A.: Flexible vector-based spatial
520 configurations in land models, Hydrol. Earth Syst. Sci., 24, 5953–5971, https://doi.org/10.5194/hess-24-5953-2020, 2020.

Gharari, S., Vanderkelen, I., Tefs, A., Mizukami, N., Stadnyk, T. A., Lawrence, D., and Clark, M. P.: A Flexible Multi-Scale Framework to Simulate Lakes and Reservoirs in Earth System Models, Earth and Space Science Open Archive, 2022.

Gourley, J. J., Flamig, Z. L., Vergara, H., Kirstetter, P., Clark, R. A., III, Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J. M., Hong, Y., and Howard, K. W.: The FLASH Project: Improving the Tools for Flash Flood Monitoring and
525 Prediction across the United States. Bull. Amer. Meteor. Soc., 98, 361–372. https://doi.org/10.1175/BAMS-D-15-00247.1, 2017.

Hanasaki, N, Kanae, S., and Oki, T.: A reservoir operation scheme for global river routing models, J. Hydrol., 327, 22-41, 2006.

Hay, L, Norton, P, Viger, R, Markstrom, S, Regan, RS, and Vanderhoof, M: Modelling surface-water depression storage in a
530 Prairie Pothole Region. Hydrol. Process., 32, 462– 479. https://doi.org/10.1002/hyp.11416, 2018.

Horton, P., Schaefli, B., and Kauzlaric, M.: Why do we have so many different hydrological models? A review based on the case of Switzerland. Wiley Interdiscip. Rev.: Water, e1574. https://doi.org/10.1002/wat2.1574, 2021.

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., and Woods, R. A.: A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments. Water Resour. Res., 56, e2019WR025975.
535 https://doi.org/10.1029/2019WR025975, 2020.

Knoben, W. J. M., Clark, M., Bates, J., Bennet, A., Gharari, S., Marsh, C., Nijssen, B., Pietroniro, A., Spiteri, R., Tarboton, D. J., and Wood, A. J.: Community Workflows to Advance Reproducibility in Hydrologic Modeling: Separating model-agnostic and model-specific configuration steps in applications of large-domain hydrologic models, Earth and Space Science Open Archive, https://doi.org/10.1002/essoar.10509195.1, 2021.

540   Lehner, B., Liermann, C.R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J.C., Rödel, R., Sindorf, N. and Wisser, D.: High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. Front. Ecol. Environ., 9, 494-502. https://doi.org/10.1890/100125, 2011.

Lehner, B. and Grill, G.: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. Hydrol. Process., 27: 2171-2186. https://doi.org/10.1002/hyp.9740, 2013.

545   Li, H., Wigmosta, M. S., Wu, H., Huang, M., Ke, Y., Coleman, A. M., and Leung, L. R.: A Physically Based Runoff Routing Model for Land Surface and Earth System Models, J. Hydrometeorol., 14(3), 808-828, 2013.

Li, Z., Chen, M., Gao, S., Hong, Z., Tang, G., Wen, Y., Gourley, J.J., and Hong, Y.: Cross-examination of similarity, difference and deficiency of gauge, radar and satellite precipitation measuring uncertainties for extreme events using conventional metrics and multiplicative triple collocation. Remote Sensing, 12(8), 1258, 2020.

550   Li, Z., Chen, M., Gao, S., Luo, X., Gourley, J. J., Kirstetter, P., Yang, T., Kolar, R., McGovern, A., Wen, Y., Rao, B., Yami, T., and Hong, Y.: CREST-iMAP v1. 0: A fully coupled hydrologic-hydraulic modeling framework dedicated to flood inundation mapping and prediction. Environ. Model. Soft., 141, 105051, 2021.

Li, Z.: CREST-VEC: A framework towards more accurate and realistic flood simulation across scales (v1.0). Zenodo. https://doi.org/10.5281/zenodo.6305817, 2022.

555   Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res., 99(D7), 14415– 14428, https://doi.org/10.1029/94JD00483, 1994.

Lohman, D., Nolte-holube, R. and Raschke, E.: A large-scale horizontal routing model to be coupled to land surface parametrization schemes. Tellus A, 48: 708-721. https://doi.org/10.1034/j.1600-0870.1996.t01-3-00009.x, 1996.

560   Lighthill M. J. and Whitham G. B.: On kinematic waves I. Flood movement in long rivers. Proc. R. Soc. Lond. A229281–316. http://doi.org/10.1098/rspa.1955.0088, 1955.

Lin, P., Rajib, M. A., Yang, Z., Somos-Valenzuela, M., Merwade, V., Maidment, D. R., Wang, Y., and Chen, L.: Spatiotemporal Evaluation of Simulated Evapotranspiration and Streamflo over Texas Using the WRF-Hydro-RAPID Modeling Framework. J. Am. Water Resour. Assoc., 54(1), 40-54. https://doi.org/10.1111/1752-1688.12585, 2018.

565   Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. G., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J., and Wood, E. F.: Global reconstruction of naturalized river flows at 2.94 million reaches. Water Resour. Res., 55, 6499– 6516. https://doi.org/10.1029/2019WR025287, 2019.

Lin, P., Pan, M., Wood, E.F. Yamazaki, D., Allen, G. H.: A new vector-based global river network dataset accounting for variable drainage density. Sci Data 8, 28, https://doi.org/10.1038/s41597-021-00819-9, 2021.

570   Liston, G. E., Sud, Y. C. and Wood, E. F.: Evaluating GCM land surface hydrology parameterizations by computing river discharges using a runoff routing model. J. Appl. Met., 33: 394–405, 1994.

Martinaitis, S. M., Gourley, J. J., Flamig, Z. L., Argyle, E. M., Clark, R. A., III, Arthur, A., Smith, B. R., Erlingis, J. M., Perfater, S., and Albright, B.: The HMT Multi-Radar Multi-Sensor Hydro Experiment, Bull. Amer. Meteor. Soc., 98(2), 347-359, 2017.

575   Messager, M.L., Lehner, B., Grill, G., Nedeva, I., and Schmitt, O.: Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nat. Comm.: 13603. https://doi.org/10.1038/ncomms13603, 2016.

Messager, M.L., Lehner, B., Cockburn, C. Lamouroux, N., Pella, H., Snelder, T., Tockner K., Trautmann, T., Watt, C. and Datry, T.: Global prevalence of non-perennial rivers and streams. Nature 594, 391–397 (2021). https://doi.org/10.1038/s41586-021-03565-5, 2021.

580   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,
585   https://doi.org/10.1002/2017WR020401, 2017.

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.

Montanari, A., Young, G., Savenije, H.H.G., Hughes, D., Wagener, T., Ren, L.L., Koutsoyiannis, D., Cudennec, C., Toth, E.,
590   Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S.J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z., and Belyaev, V.: "Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022. Hydrol. Sci. J. 58 (6) 1256–1275, 2013.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J.
595   R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrol. Earth Syst. Sci., 19, 209–223, https://doi.org/10.5194/hess-19-209-2015, 2015.

NOAA National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters. https://www.ncei.noaa.gov/access/billions/, DOI: 10.25921/stkw-7w73, 2022.

600   Ohmura, A.: Physical basis for the temperature-based melt-index method. J. Appl. Meteorol., 40, pp. 753-761, 2001.

Paiva, R.C.D., Collischonn, W., and Tucci, C.E.M.: Large scale hydrologic and hydrodynamic modeling using limited data and a GIS based approach. J. Hydrol., 406 (3), 170–181. https://doi/org/10.1016/j.jhydrol.2011.06.007, 2011.

Paniconi, C., and Putti, M.: Physically based modeling in catchment hydrology at 50: Survey and outlook, Water Resour. Res., 51, 7090– 7129, https://doi.org/10.1002/2015WR017780, 2015.

605 Ponce, V. M., R.-M. Li, and D. B. Simons: Applicability of kinematic and diffusion models, J. Hydraul. Div., 104(HY3), 353–360, 1978.

PRISM Climate Group, Oregon State University, https://prism.oregonstate.edu, data created 4 Feb 2014, accessed 16 Sep 2020.

Quinn, P., Beven, K., Chevallier, P. and Planchon, O.: The prediction of hillslope flow paths for distributed hydrological
610 modelling using digital terrain models. Hydrol. Process., 5: 59-79. https://doi.org/10.1002/hyp.3360050106, 1991.

Rameshwaran, P., Bell, V. A., Brown, M. J., Davies, H. N., Kay, A. L., Rudd, A. C., and Sefton, C.: Use of abstraction and discharge data to improve the performance of a national-scale hydrological model. Water Resour. Res., 58, e2021WR029787, https://doi.org/10.1029/2021WR029787, 2022.

Salas, F. R., Somos-Valenzuela, Marcelo A., Dugger, A., Maidment, D. R., Gochis, D. J., David, C. H., Yu, W., Ding, D.,
615 Clark, E. P., and Noman, N.: Towards Real-Time Continental Scale Streamflow Simulation in Continuous and Discrete Space. J. Am. Water Resour. Assoc. 54, 7- 27, https://doi.org/10.1111/1752-1688.12586, 2018

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resour. Res., 46, W05523, https://doi.org/10.1029/2008WR007327, 2010.

Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., Eisner, S., Müller Schmied, H., Sutanudjaja, E.
620 H., Warrach-Sagi, K., and Attinger, S.: Toward seamless hydrologic predictions across spatial scales, Hydrol. Earth Syst. Sci., 21, 4323–4346, https://doi.org/10.5194/hess-21-4323-2017, 2017.

Savenije, H. H.: HESS Opinions" The art of hydrology". Hydrol. Earth Syst. Sci., 13(2), 157-161, 2009.

Shaad, K: Evolution of river-routing schemes in macro-scale models and their potentials for watershed management, Hydrol. Sci. J., 63, 1062-1077. https://doi.org/10.1080/02626667.2018.1473871, 2018.

625 Shen, X., Hong, Y., Zhang, K., and Hao, Z.: Refining a distributed linear reservoir routing method to improve performance of the CREST model. J. Hydrol. Eng., 22(3), 04016061, 2017.

Solvik, K., Bartuszevige, A. M., Bogaerts, M., and Joseph, M. B.: Predicting playa inundation using a long short-term memory neural network. Water Resour. Res., 57, e2020WR029009. https://doi.org/10.1029/2020WR029009, 2021.

Tang, G., Zeng, Z., Long, D., Guo, X., Yong, B., Zhang, W., and Hong, Y.: Statistical and Hydrological Comparisons between
630 TRMM and GPM Level-3 Products over a Midlatitude Basin: Is Day-1 IMERG a Good Successor for TMPA 3B42V7?, J. Hydrometeorol., 17(1), 121-137, 2016.

Tarboton, D. G.: A new method for the determination of flow directions and upslope areas in grid digital elevation models, Water Resour. Res., 33(2), 309– 319, https:/doi.org/10.1029/96WR03137, 1997.

Tavakoly, A. A., Gutenson, J. L., Lewis, J. W., Follum, M. L., Rajib, A., LaHatte, W. C., and Hamilton, C. O.: Direct
635 integration of numerous dams and reservoirs outflow in continental scale hydrologic modeling. Water Resour. Res., 57, e2020WR029544. https://doi.org/10.1029/2020WR029544, 2021.

Tellman, B., Sullivan, J.A., Kuhn, C. Kettner, A. J., Doyle, C. S., Brakenridge, G. R., Erickson, T. A. and Slayback, D. A.: Satellite imaging reveals increased proportion of population exposed to floods. Nature 596, 80–86, https://doi.org/10.1038/s41586-021-03695-w, 2021.

640   Tijerina, D., Condon, L., FitzGerald, K., Dugger, A., O'Neill, M. M., Sampson, K., Gochis, D., Maxwell, R.: Continental hydrologic intercomparison project, phase 1: A large-scale hydrologic model comparison over the continental United States. Water Resour. Res., 57, e2020WR028931. https://doi.org/10.1029/2020WR028931, 2021.

Tokuda, D., Kim, H., Yamazaki, D., and Oki, T.: Development of a coupled simulation framework representing the lake and river continuum of mass and energy (TCHOIR v1.0), Geosci. Model Dev., 14, 5669–5693, https://doi.org/10.5194/gmd-14-

645   5669-2021, 2021.

Vanderkelen, I., Gharari, S., Mizukami, N., Clark, M. P., Lawrence, D. M., Swenson, S., Pokhrel, Y., Hanasaki, N., van Griensven, A., and Thiery, W.: Evaluating a reservoir parametrisation in the vector-based global routing model mizuRoute (v2.0.1) for Earth System Model coupling, Geosci. Model Dev. Discuss. [preprint], https://doi.org/10.5194/gmd-2022-16, in review, 2022.

650   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. https://doi.org/10.1016/j.jhydrol.2016.06.011, 2016.

Vrugt, J. A., ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., and Higdon, D.: Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling, Int. J. Nonlinear Sci. Numer.

655   Simul., 10, 273–290, https://doi.org/10.1515/ijnsns.2009.10.3.273, 2009

Wang, X., White-Hull, C., Dyer, S., and Yang, Y.: GIS-ROUT: A River Model for Watershed Planning. Environ. Plann. B Plann. Des., 27(2), 231–246, https://doi.org/10.1068/b2624, 2000

Wang, J., Hong, Y., Li, L., Gourley, J.J., Khan, S.I., Yilmaz, K.K., Adler, R.F., Policelli, F.S., Habib, S., Irwn, D., Limaye, A.S., Korme, T. and Okello L., The coupled routing and excess storage (CREST) distributed hydrological model, Hydrol. Sci.

660   J., 56 (1), 84-98, 2011.

Xue, X., Hong, Y., Limaye, A. S., Gourley, J. J., Huffman, G. J., Khan, S. I., Doriji, C., Chen, S.: Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins? J. Hydrol., 499, 91-99. https://doi.org/10.1016/j.hydrol.2013.06.042, 2013.

665   Yamazaki D, Kanae S, Kim H, and Oki T.: A physically based description of floodplain inundation dynamics in a global river routing model. Water Resour. Res. 47: 1– 21. https://doi.org/10.1029/2010WR009726, 2011.

Yamazaki, D., Sato, T., Kanae, S., Hirabayashi, Y., and Bates, P. D.: Regional flood dynamics in a bifurcating mega delta simulated in a global river model, Geophys. Res. Lett., 41, 3127– 3135, https://doi.org/10.1002/2014GL059744, 2014.

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.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., and Pavelsky, T. M.: MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset, Water Resour. Res., 55, 5053–5073, https://doi.org/10.1029/2019WR024873, 2019.

Yang, Y., Pan, M., Lin, P., Beck, H. E., Zeng, Z., Yamazaki, D., David, C. H., Lu, H., Yang, K., Hong, Y., and Wood, E. F.: Global Reach-Level 3-Hourly River Flood Reanalysis (1980–2019), Bull. Amer. Meteorol. Soc., 102(11), E2086-E2105, 2021.

Yang, T., Zhang, L., Kim, T., Hong, Y., Zhang, D., amd Peng, Q.: A large-scale comparison of Artificial Intelligence and Data Mining (AI&DM) techniques in simulating reservoir releases over the Upper Colorado Region, J. Hydrol., 602, 126723, 2021.

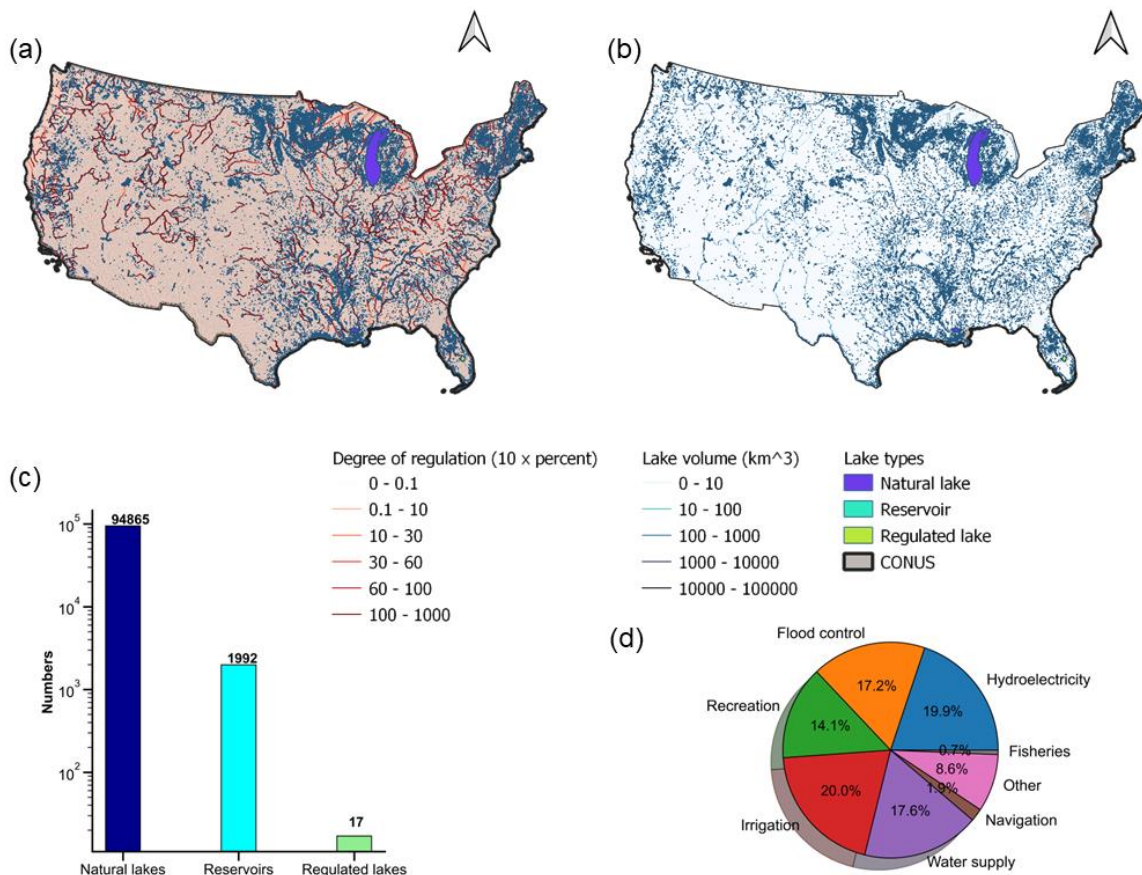
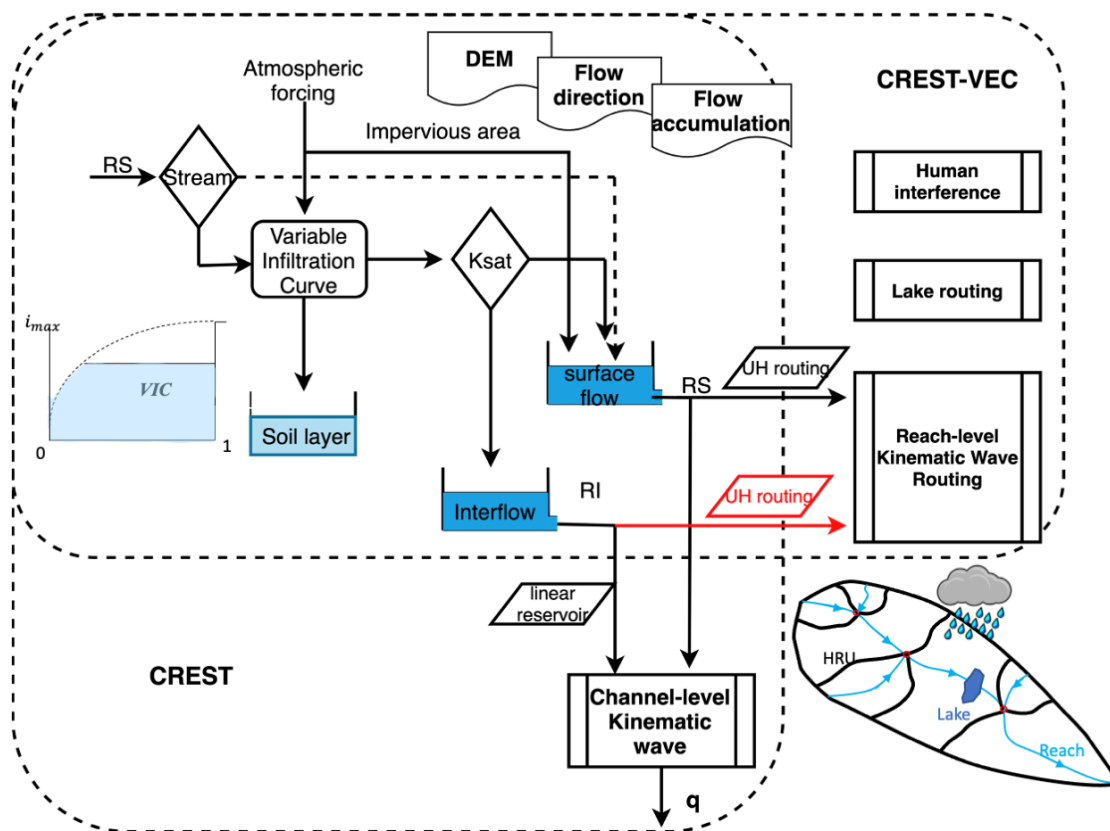



Figure 1. Maps of (a) percent of regulated river and (b) regulated lake volume; (c) bar plot of lake classifications; (d) pie plot of US regulated lake or reservoir function purposes.

685   **Figure 2. Schematic view of the CREST-VEC framework. The red arrow highlights the newly added subsurface routing option to the original mizuRoute framework.**

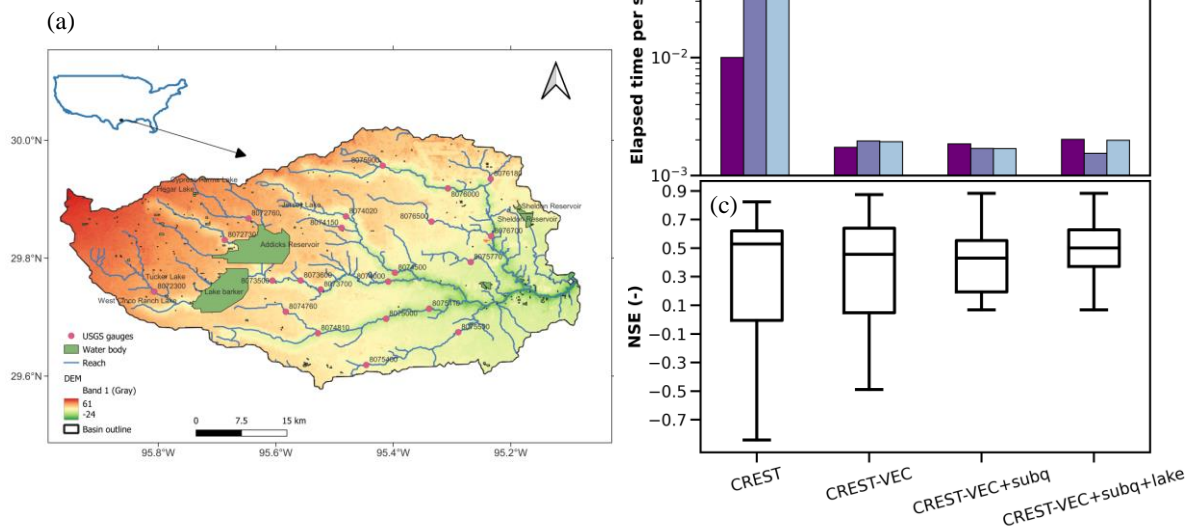



Figure 3. (a) Map of the study area (Houston region) showing river networks and water bodies; (b) Computation time per step for CREST at three resolutions and CREST-VEC model at four configurations on the x-axis; (c) Nash-Sutcliffe efficiency values for CREST and CREST-VEC model.

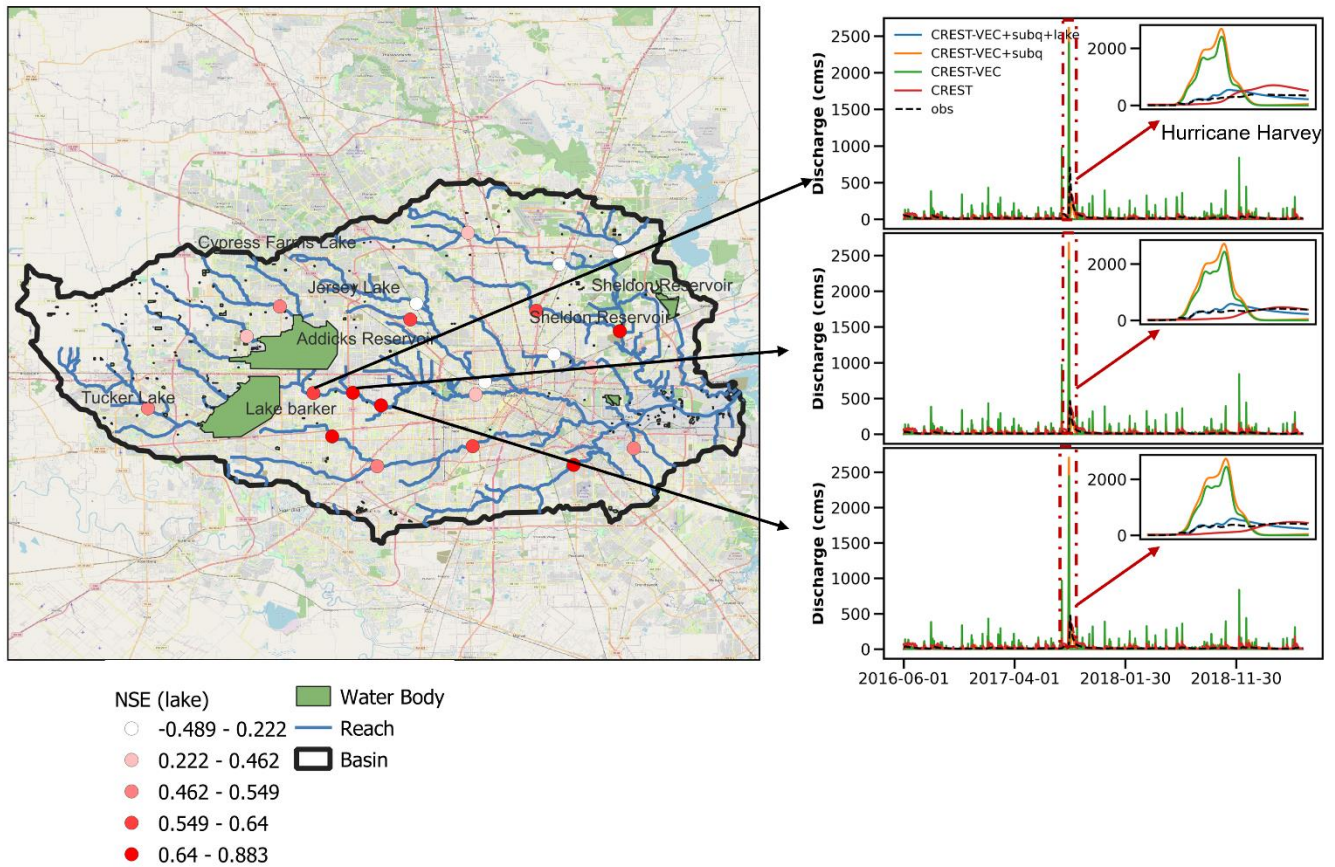



Figure 4. Performance of models downstream of two lakes. The Nash-Sutcliffe Efficiency coefficients are obtained from the CREST-VEC model with lake routing and subsurface routing. Three plots of time series of stream gauges (from upstream to downstream: 08073500, 08073600, 08074000) are pointed aside by the map, and the Hurricane Harvey event is highlighted in red box and insets.

695

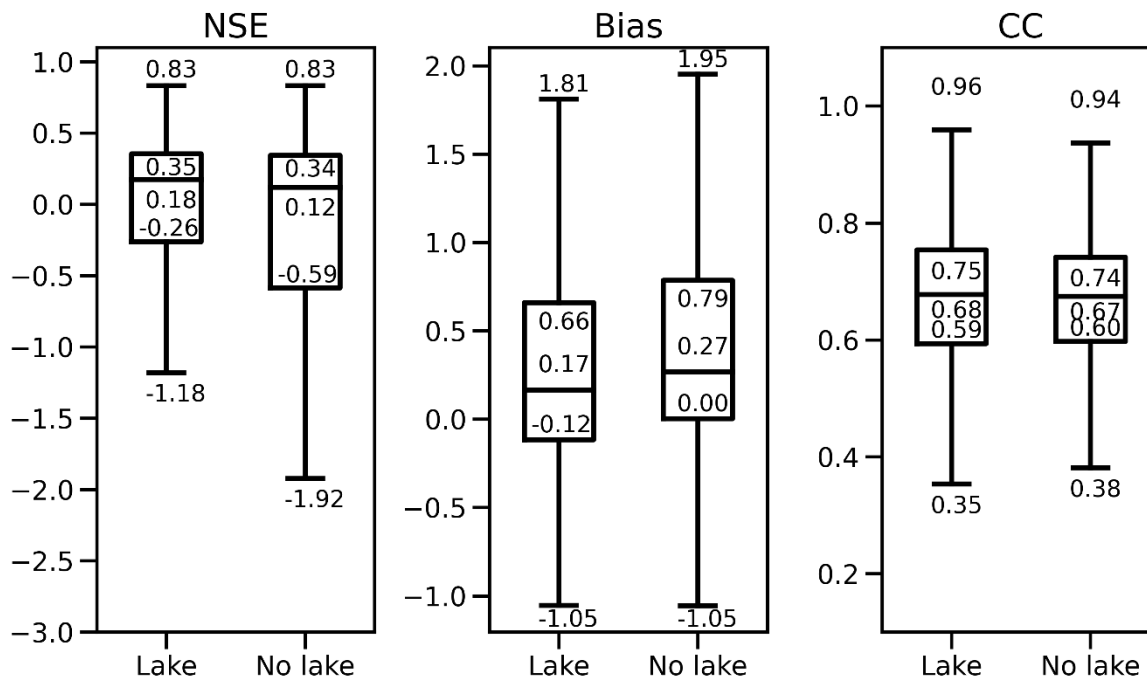



Figure 5. Boxplot of model performance (NSE, Bias, and CC) comparing results with lake routing and without it. The numbers for each box represent first quantile minus 1.5 times interquartile range, first quantile, median, third quantile, and third quantile plus 1.5 times interquartile range. **Figure 5. Boxplot of model performance comparing results with lake routing and without it.**

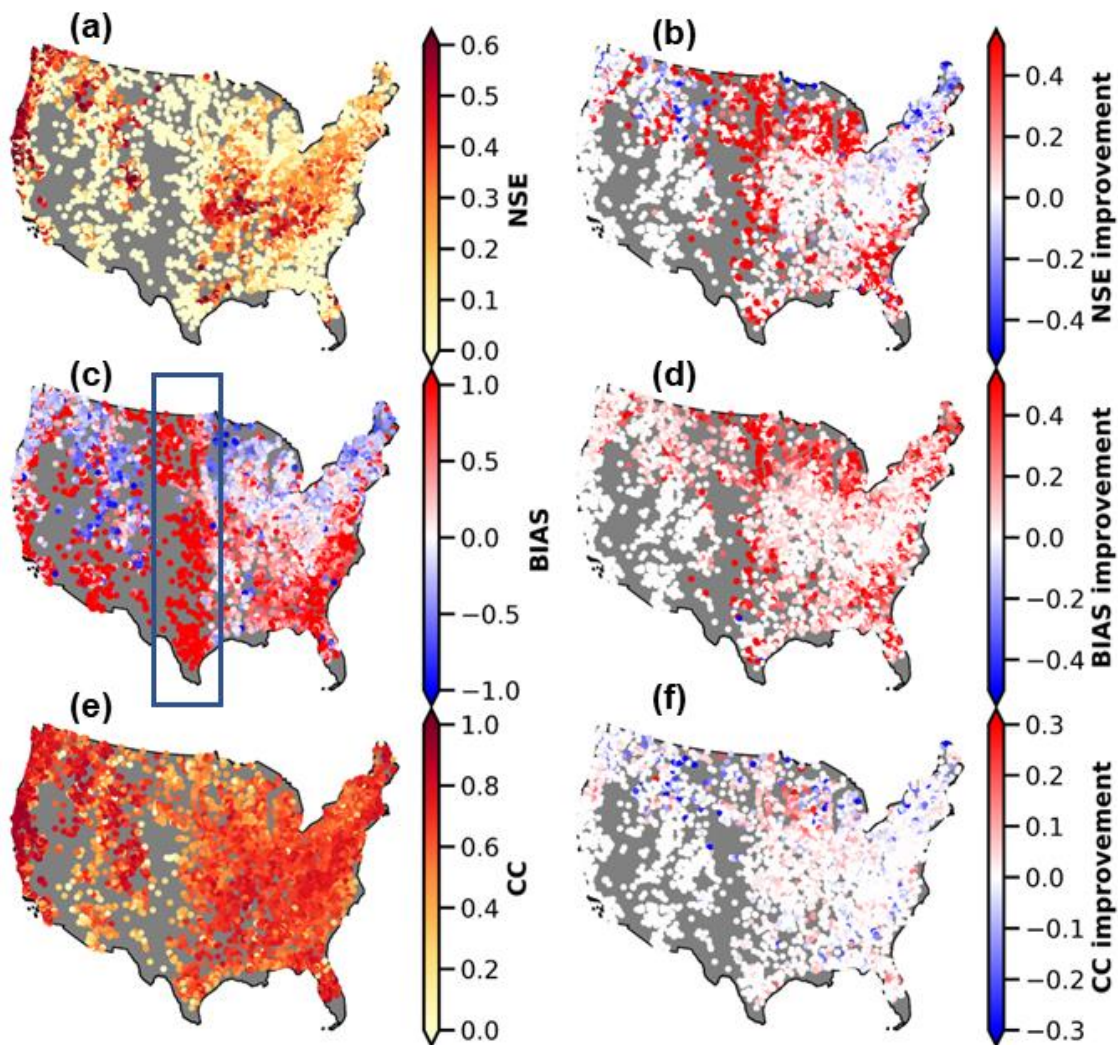



Figure 6. Spatial map of model performance with the lake (left column) and the difference between with and without lake simulation (right column). (a): NSE scores; (b) NSE differences (results with lake minus results without lake); (c) BIAS; (d) BIAS difference; (c) Correlation Coefficient (CC); (f) CC difference. The blue box in (c) highlights the region where high positive BIAS is present.

Figure 7. (a) density plot of CREST-VEC simulated annual surface runoff against GRFR (Global Reach-level Flood Reanalysis) in
705 **the Great Plains; (b) cumulative density function (CDF) of CREST-VEC and GRFR simulated annual runoff.**

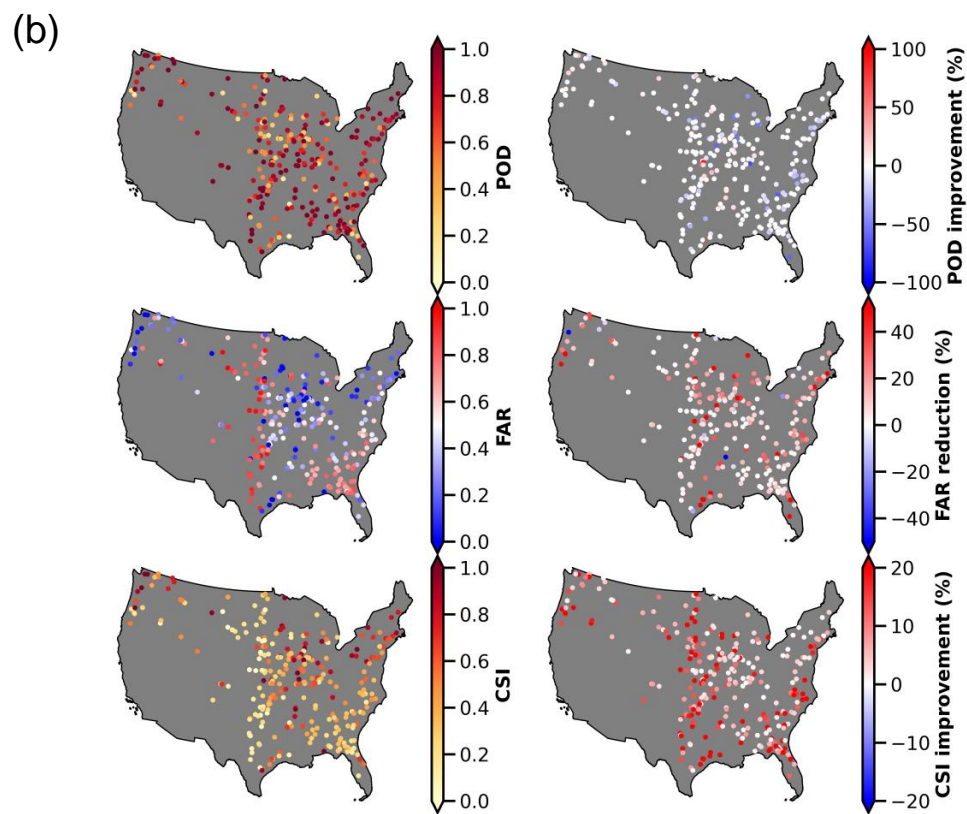
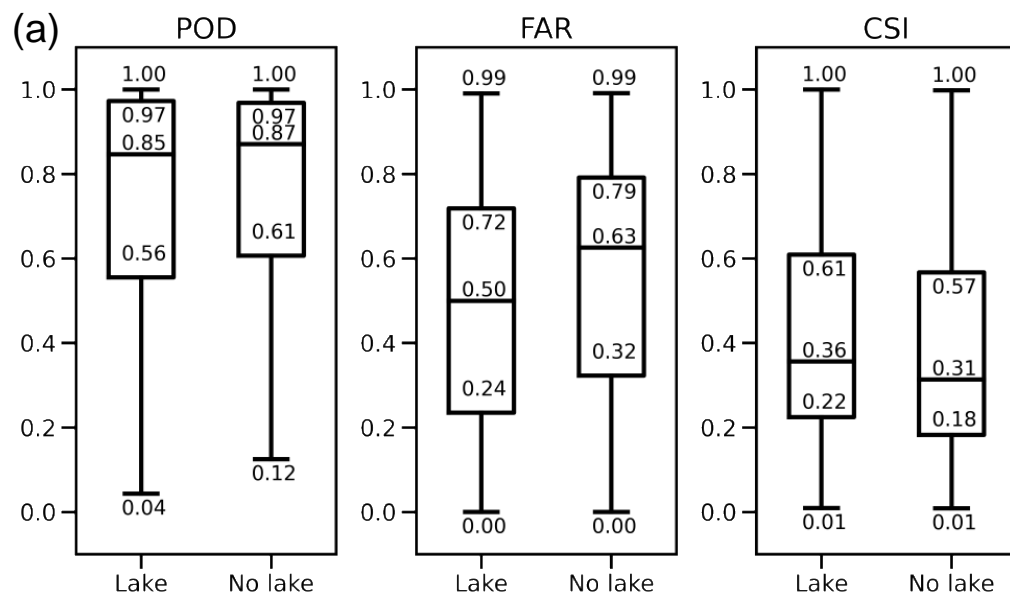
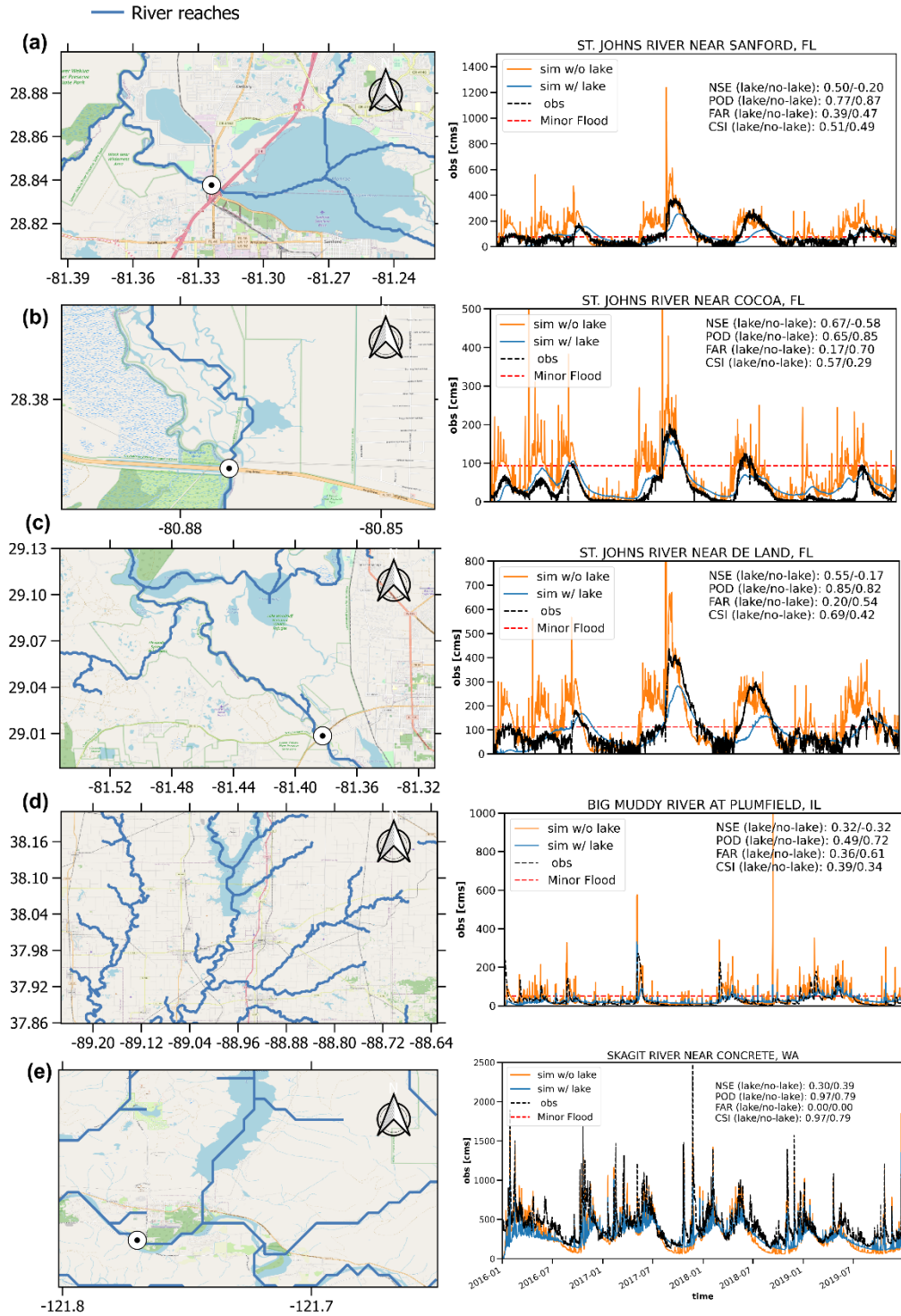



Figure 8. Flood detection performance comparing lake and no lake simulation. (a) Similar to Figure 5, but for flood detectability; (b) Similar to Figure 6, but for flood detectability.

710 **Figure 9. Five case examples of streamflow time series at gauges downstream of lakes: (a) St. Johns River near Sanford, FL; (b) St. Johns River near Cocoa, FL; (c) St. Johns River near De Land, FL; (d) Big Muddy River at Plumfield, IL; (e) Mississippi River at Clinton, IA. Image~~s~~ copyright © Google Map~~courtesy of Google Map~~.**