Replies to reviews

“To what extent does river routing matter in hydrological modelling?”

Nicolás Cortés-Salazar, Nicolás Vásquez, Naoki Mizukami, Pablo A. Mendoza, and Ximena Vargas

We provide responses to each individual point below. For clarity, comments are given in italics, and our responses are given in plain blue text.

Anonymous Referee #2

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

While the approach was rigorous in the sense that it compiled data from thousands of model simulations, it suffers from a number of critical methodological issues.

We thank the referee for his/her time on reviewing our manuscript and providing several constructive suggestions. In this document, we provide our detailed responses and also mention how we plan to address the reviewer’s comments in a future version of this manuscript.

I discuss a few of these major issues below:

Routing is most influential on peak magnitude and timing of large events; both of these are poorly captured by integrated hydrograph metrics such as KGE and NSE. Peak flow differences after calibrating the routing models would be a much more useful metric for evaluating routing model performance. By using integrated measures such as KGE, the critical differences between routing algorithms are not discernable (as seen in nearly all of the reported results).

We agree with this reviewer that including additional high-flow related metrics would increase the impact of this work. Therefore, we will add the annual peak flow bias (Mizukami et al., 2019) and the percent bias in the high flow segment of the flow duration curve (Yilmaz et al., 2008) as calibration metrics. The justification for including these metrics and related equations will be included in section 4.2 (“Objective functions”) of the revised manuscript.

In this work, we decide to include KGE, NSE and log-NSE in the analysis, since these integrated metrics are not only of interest for the hydrologic modeling community – especially for parameter calibration and evaluation (e.g., Fowler et al., 2018; Knoben et al., 2019; Clark et al., 2021) –, but also for the river routing community. Indeed, several examples of river routing scheme assessments can be found using the KGE (e.g., Pereira et al., 2017; Hoch et al., 2019; Qiao et al., 2019; Thober et al., 2019; Munier & Decharme, 2022), NSE (e.g., Yamazaki et al., 2011; Ye et al., 2013; Zhao et al., 2017; ElSaadani et al., 2018; Nguyen-Quang et al., 2018; Fleischmann et al., 2019, 2020) and even the NSE using flows in logarithmic space (Paiva et al., 2013). We believe that the joint analysis of traditional streamflow performance metrics and high flow metrics will expand the target audience of this work. We will provide a proper justification for selecting these metrics in the new section 4 (“Experimental setup”) of the revised manuscript.

Each of the figures in the report are reporting ALL of the output from the simulations, regardless of whether it is important or interpretable or worthy of interpretation.

Figure 5 is the only one that contains results from all the parameter samples, though we will modify this to address the following reviewer's comment (see next response).

For instance, figure 5 reports KGE, NSE, and NSE of log transformed flows for all 3500 simulations with multiple timesteps, multiple routing schemes. In addition to the only interpretable result from this figure is that no routing is outperformed by routing, there is little utility in comparing NSE values of 0.2-0.3 (the approximate median of these simulations) -differences in NSE below about 0.5 are nearly arbitrary in that a hydrograph with an NSE of 0.2 may not be visibly preferable to an NSE of 0.05. The only feature of this plot referred to in the text was the maximum metric value. Why not simply report that?

The original motivation of Figure 5 was to illustrate the impact that incorporating river routing modelling may have on streamflow performance metrics across the VIC parameter space. Nevertheless, we fully agree that it would make more sense to simply illustrate the results for a subset of behavioural parameter sets (as we did in Figure 7). Hence, in the revised manuscript we select and report only the best 1% of runs, following the approach of Melsen et al. (2016).

Critically, because the parameters of the VIC model are arbitrary, the comparisons of even the best models are in effect the results of Monte Carlo calibration, the least efficient optimization approach.

We decide to use a Monte Carlo parameter sampling approach because, rather than seeking for an optimal parameter set, our primary goal is to assess the impacts of different river routing configurations on streamflow metrics across the model parameter space. In particular, Latin Hypercube Sampling is a common strategy to sample the parameter space and identify behavioural parameter sets for a specific target metric (e.g., Andréassian et al., 2014; Broderick et al., 2016; Melsen et al., 2016, 2019; Guse et al., 2017; Khatami et al., 2019). We will clarify this in section 4.1 ("Parameter sampling and streamflow simulations") of the revised manuscript, adding the following text:
"Since we aim to examine the impacts of different routing schemes on streamflow performance metrics across the parameter space, rather than seeking for an optimal parameter set, we use the Latin Hypercube Sampling (LHS) method, which is a common strategy to sample the parameter space and identify behavioral sets for specific target metrics (e.g., Andréassian et al., 2014; Broderick et al., 2016; Melsen et al., 2016, 2019; Guse et al., 2017; Khatami et al., 2019)."

Comparing the 'best' models when these are not rigorously determined to be the actual best for each algorithm (rather than a sampling error) is problematic. For this comparison to be rigorous, I don't see how to do this without simultaneous calibration of routing and land surface parameters, an issue the authors acknowledge in section 5.2.

River routing parameters were excluded from the original setup in order to make a clean numerical experiment (recall that the baseline model has no routing module). However, we appreciate the reviewer's concern, and in response to this comment we are conducting new LHS experiments that include river routing parameters (see response to the following comment).

In practice, the routing parameters (such as Manning's n) would be calibrated in conjunction with VIC model parameters, likely further diminishing any incremental performance differences between the routing models.

In order to address this point, we have repeated the parameter sampling experiment, including 13 VIC parameters and the routing parameters: one (the Manning roughness coefficient) for the Kinematic Wave, Diffusive Wave and Muskingum-Cunge algorithms, and two for the Impulse Response Function method (see Table S1). The new results are displayed in Figure S1.

Table S1. Model parameters sampled in this study.

Parameter	**Units**	**Lower value**	**Upper value**	**Description**
Infilt	-	0.01	0.99	Variable infiltration curve parameter
D_s	-	0.1	0.9	Fraction of $D_{s_{max}}$ where non-linear baseflow occurs
$D_{s_{max}}$	mm/d	0.1	300	Maximum velocity of baseflow
W_s	-	0.1	0.9	Fraction of maximum soil moisture where non-linear baseflow occurs
expt	-	3.1	10	Exponent of Campbell's equation for hydraulic conductivity
d_{max} d_1 d_2 d_3	m	0.5 0.05 d_{max} 0.21 d_{max} 0.74 d_{max}	5 0.2 d_{max} 0.7 d_{max} 0.1 d_{max}	Depth of soil layers 1, 2 and 3
K_{sat}	mm/d	1	1000	Saturated hydraulic conductivity
$T_{max,snow}$	(°C)	-10	10	Maximum temperature for snowfall
α_{thaw}	-	0.75	0.90	Decay of albedo
α_{new}	-	0.85	0.95	Maximum albedo for fresh snow
n	$s/m^{1/3}$	0.024	0.075	Roughness coefficient of Manning (Barnes, 1967)
C	m/s	0.25	10	Advection coefficient (Allen et al., 2018)
D	m^2/s	200	4000	Diffusion coefficient (Melsen et al., 2016)

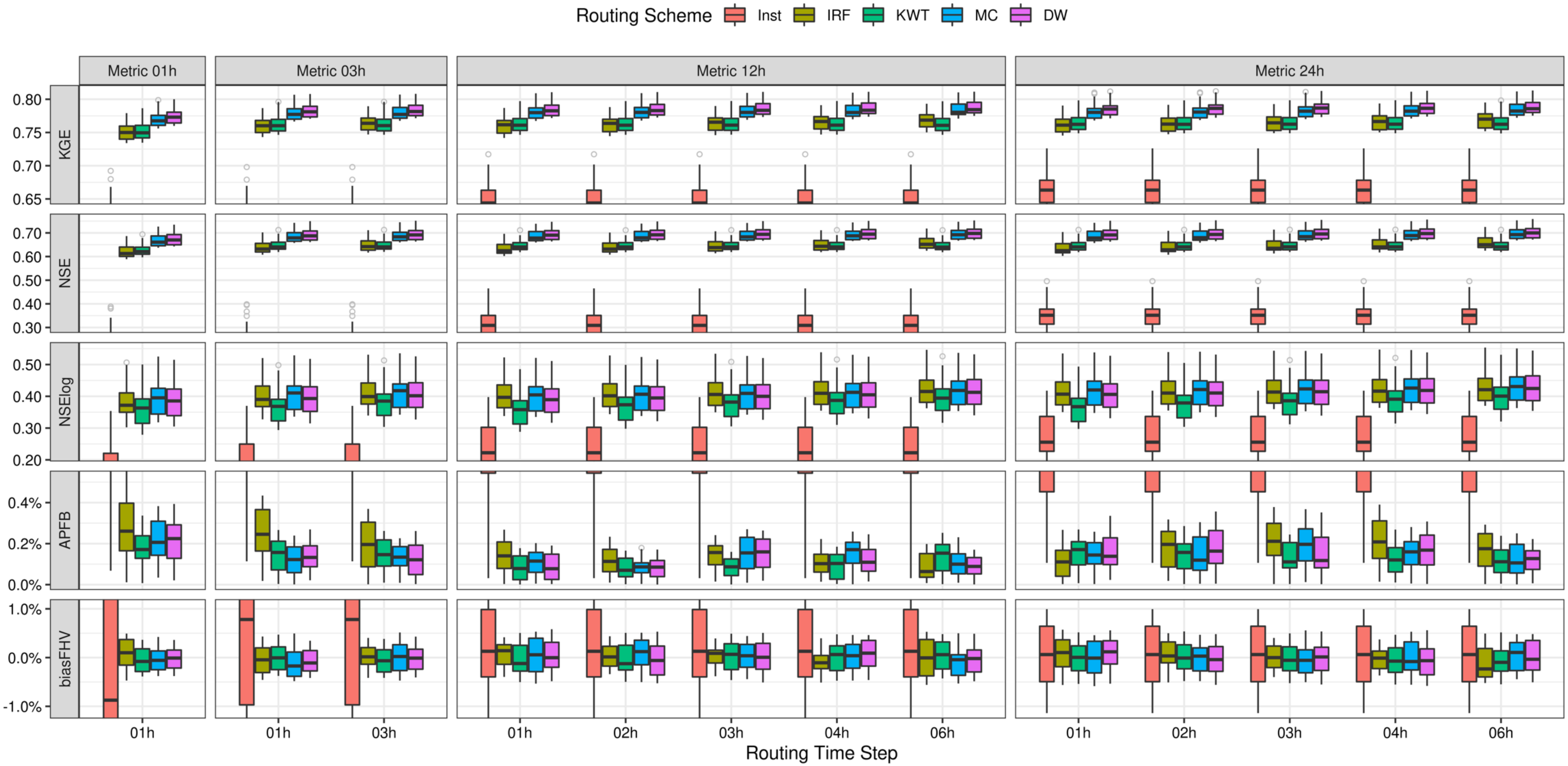

Figure S1. New parameter sampling results for the Cautín at Cajón River basin. Impact of routing scheme and routing time step on performance metrics (rows) computed for the period April/2008-March/2012, using different discharge temporal resolutions for metric calculation (columns) across a large sample of VIC and routing parameter sets obtained through Latin Hypercube Sampling. The results are presented for instantaneous runoff (Inst), Impulse Response Function (IRF), Kinematic Wave Tracking (KWT), Muskingum-Cunge (MC) and Diffusive Wave (DW), and the metrics are the Kling-Gupta Efficiency (KGE), the Nash-Sutcliffe efficiency in both raw (NSE) and logarithmic (NSElog) space, the annual peak flow bias (APFB) and the percent bias in the high flow segment of the flow duration curve (bias FHV). Each boxplot shows, for each configuration, the best 1% from 3500 parameter sets.

As seen in Figure S1, the benefits of representing river routing processes are not only limited to popular performance measures like KGE or NSE, but also to metrics that emphasize low flow simulations (NSE-log), and metrics for high flow applications. We plan to update all the figures of the revised manuscript based on the new experimental setup.

The fundamental results discussed here are obvious without the testing herein.

We provide individual responses to all the reviewer's comments on this matter below:

Routing is better than no routing.

Of course, the incorporation of as many hydrological processes as observed for the natural system of interest – including river routing – is desirable to achieve "fidelius" (Gharari et al., 2021) simulations in hydrology and land surface models. Nevertheless, the degree of improvement for application-specific metric is not obvious, considering that the interplay between model structural uncertainty (here, provided by different routing schemes) and parametric uncertainty is not fully understood and, therefore, it is a topic of active research (e.g., Günther et al., 2020; e.g., Pilz et al., 2020; Spieler et al., 2020; Chlumsky et al., 2021; Zhou et al., 2023). In particular, it is not clear to which degree the perturbation of hydrological model parameters can compensate for the lack of river routing representation – a gap that this work intends to fill and is now highlighted in the introduction.

Low flows are not as impacted by routing differences.

In principle, the effects of a specific model configuration (i.e., routing scheme or routing time step) on certain processes (e.g., high or low flows) are not comparable, since these are assessed with different performance metrics (e.g., Mizukami et al., 2019; Zhou et al., 2023). Additionally, the results presented here show that the implementation and configuration of river routing schemes are also relevant for medium and low flows. For example, including river routing provided higher NSE-log values (Figures 5 and 6 in the original submission) – improving the simulation of low flows – and modified the shape of the mid and low flow segments in the FDC (Figure 10 in the original submission), which are characteristic signatures of 'flashiness' in runoff response and long term baseflow, respectively (Yilmaz et al., 2008). The effects of river routing are also reflected in the partitioning of total runoff between baseflow and surface runoff. In fact, the results presented here show that the parameter search process compensates for the lack of routing by modifying other fluxes and state variables (Khatami et al., 2019) to increase streamflow-oriented performance metrics. In our case, the contribution of baseflow to total runoff increases by >20% when river routing is excluded, which is achieved by modifying soil parameters –especially W_s, one of the most sensitive for baseflow processes (Sepúlveda et al., 2022) – to delay the streamflow response. This result suggests that including routing processes may impact the outcomes from drought-oriented studies, since baseflow is the primary flux sustaining streamflow during water scarcity periods (Karki et al., 2021). All these points are now highlighted in the discussion (Section 6.1).

Parameter compensation occurs in hydrological models.

We agree that this is a well-known issue in the hydrologic modelling community, including compensation of hydrologic model parameters on forcing errors (e.g., Baez-Villanueva et al., 2021; Wang et al., 2023) and hydrologic model structure deficiencies (e.g., Saavedra et al., 2022). To the best of our knowledge, however, this is the first study that characterizes how hydrologic model parameter estimation can compensate for the absence of river routing representation, with focus on performance metrics, parameter values, flux partitioning and signatures used for water resources applications (in our case, flood frequency and flow duration curves). Our results show that, in this case, parameter compensation is not trivial, and depends on the calibration metric and fluxes analysed (Figure 7).

Routing impacts are not visible when averaged over a monthly time step.

The results presented here demonstrate exactly the opposite. Indeed, the second row in Figure 4 (original submission) shows that the difference in monthly flows between routed vs. non-routed flows can be as large as 63.2 m^3/s. We thank the reviewer for this comment, and we will modify the text in order to clarify this point:

"At monthly time steps, the differences between routed and instantaneous runoff reduce considerably, although these still can be as large as 63.2 m^3/s (i.e., a 29.3% difference using routed runoff as the reference)."

Models with instantaneous routing (i.e., no routing) have higher flows.

We agree (and also shown in Figures 9 and 10) that excluding routing yields higher flows. Obvious errors observed in the simulations from a model without routing are actually timing errors (due to the absence of travel time in the channel) and variability error (due to no attenuation accounted for). These types of errors are often assumed to become negligible when temporally aggregating non-routed flows. Nevertheless, this study unveils that the impacts of river routing go beyond the simulation of high flows at sub-daily or daily time steps, affecting streamflow simulations even at the monthly time scale (see previous response). The extent to which these effects propagate across modelling decisions is not trivial and has not been previously documented.

As is, I do not see an additional contribution from this work beyond that which exists in the literature. It is for the above reasons that I must recommend rejection.

We regret that the reviewer failed to agree with us regarding the contribution of the manuscript. We will address all the critiques raised in this review, and we will convey more clearly the relevance of our work in the revised version.

References

Andréassian, V., Bourgin, F., Oudin, L., Mathevet, T., Perrin, C., Lerat, J., et al. (2014). Seeking genericity in the selection of parameter sets: Impact on hydrological model efficiency. *Water Resources Research*, *50*(10), 8356–8366. https://doi.org/10.1002/2013WR014761

Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Mendoza, P. A., McNamara, I., Beck, H. E., Thurner, J., et al. (2021). On the selection of precipitation products for the regionalisation of hydrological model parameters. *Hydrology and Earth System Sciences*, *25*(11), 5805–5837. https://doi.org/10.5194/hess-25-5805-2021

Broderick, C., Matthews, T., Wilby, R. L., Bastola, S., & Murphy, C. (2016). Transferability of hydrological models and ensemble averaging methods between contrasting climatic periods. *Water Resources Research*, *52*(10), 8343–8373. https://doi.org/10.1002/2016WR018850

Chlumsky, R., Mai, J., Craig, J. R., & Tolson, B. A. (2021). Simultaneous Calibration of Hydrologic Model Structure and Parameters Using a Blended Model. *Water Resources Research*, *57*(5), 1–22. https://doi.org/10.1029/2020WR029229

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., et al. (2021). The Abuse of Popular Performance Metrics in Hydrologic Modeling. *Water Resources Research*, *57*(9), 1–16. https://doi.org/10.1029/2020WR029001

ElSaadani, M., Krajewski, W. F. F., Goska, R., & Smith, M. B. B. (2018). An Investigation of Errors in Distributed Models' Stream Discharge Prediction Due to Channel Routing. *Journal of the American Water Resources Association*, *54*(3), 742–751. https://doi.org/10.1111/1752-1688.12627

Fleischmann, A., Paiva, R., & Collischonn, W. (2019). Can regional to continental river hydrodynamic models be locally relevant? A cross-scale comparison. *Journal of Hydrology X*, *3*, 100027. https://doi.org/10.1016/j.hydroa.2019.100027

Fleischmann, A., Paiva, R. C. D., Collischonn, W., Siqueira, V. A., Paris, A., Moreira, D. M., et al. (2020). Trade-Offs Between 1-D and 2-D Regional River Hydrodynamic Models. *Water Resources Research*, *56*(8), 1–30. https://doi.org/10.1029/2019WR026812

Fowler, K., Peel, M., Western, A., & Zhang, L. (2018). Improved Rainfall-Runoff Calibration for Drying Climate: Choice of Objective Function. *Water Resources Research*, *54*(5), 3392–3408. https://doi.org/10.1029/2017WR022466

Gharari, S., Gupta, H. V., Clark, M. P., Hrachowitz, M., Fenicia, F., Matgen, P., & Savenije, H. H. G. (2021). Understanding the Information Content in the Hierarchy of Model Development Decisions: Learning From Data. *Water Resources Research*, *57*(6). https://doi.org/10.1029/2020WR027948

Günther, D., Hanzer, F., Warscher, M., Essery, R., & Strasser, U. (2020). Including Parameter Uncertainty in an Intercomparison of Physically-Based Snow Models. *Frontiers in Earth Science*, *8*(October), 1–15. https://doi.org/10.3389/feart.2020.542599

Guse, B., Pfannerstill, M., Gafurov, A., Kiesel, J., Lehr, C., & Fohrer, N. (2017). Identifying the connective strength between model parameters and performance criteria. *Hydrology and Earth System Sciences*, *21*(11), 5663–5679. https://doi.org/10.5194/hess-21-5663-2017

Hoch, J. M., Eilander, D., Ikeuchi, H., Baart, F., & Winsemius, H. C. (2019). Evaluating the impact of model complexity on flood wave propagation and inundation extent with a hydrologic-hydrodynamic model coupling framework. *Natural Hazards and Earth System Sciences*, *19*(8), 1723–1735. https://doi.org/10.5194/nhess-19-1723-2019

Karki, R., Krienert, J. M., Hong, M., & Steward, D. R. (2021). Evaluating Baseflow Simulation in the National Water Model: A Case Study in the Northern High Plains Region, USA. *Journal of the American Water Resources Association*, *57*(2), 267–280. https://doi.org/10.1111/1752-1688.12911

Khatami, S., Peel, M. C., Peterson, T. J., & Western, A. W. (2019). Equifinality and Flux Mapping: A New Approach to Model Evaluation and Process Representation Under Uncertainty. *Water Resources Research*, *55*(11), 8922–8941. https://doi.org/10.1029/2018WR023750

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M., & Uijlenhoet, R. (2016). Representation of spatial and temporal variability in large-domain hydrological models: Case study for a mesoscale pre-Alpine basin. *Hydrology and Earth System Sciences*, *20*(6), 2207–2226. https://doi.org/10.5194/hess-20-2207-2016

Melsen, L., Teuling, A. J., Torfs, P. J. J. F., Zappa, M., Mizukami, N., Mendoza, P. A., et al. (2019). Subjective modeling decisions can significantly impact the simulation of flood and drought events. *Journal of Hydrology*, *568*(November 2018), 1093–1104. https://doi.org/10.1016/j.jhydrol.2018.11.046

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., & Kumar, R. (2019). On the choice of calibration metrics for “high-flow” estimation using hydrologic models. *Hydrology and Earth System Sciences*, *23*(6), 2601–2614. https://doi.org/10.5194/hess-23-2601-2019

Munier, S., & Decharme, B. (2022). River network and hydro-geomorphological parameters at 1/12 ° resolution for global hydrological and climate studies. *Earth System Science Data*, *14*(5), 2239–2258. https://doi.org/10.5194/essd-14-2239-2022

Nguyen-Quang, T., Polcher, J., Ducharne, A., Arsouze, T., Zhou, X., Schneider, A., & Fita, L. (2018). ORCHIDEE-ROUTING: Revising the river routing scheme using a high-resolution hydrological database. *Geoscientific Model Development*, *11*(12), 4965–4985. https://doi.org/10.5194/gmd-11-4965-2018

Paiva, R. C. D., Collischonn, W., & Buarque, D. C. (2013). Validation of a full hydrodynamic model for large-scale hydrologic modelling in the Amazon. *Hydrological Processes*, *27*(3), 333–346. https://doi.org/10.1002/hyp.8425

Pereira, F. F., Farinosi, F., Arias, M. E., Lee, E., Briscoe, J., & Moorcroft, P. R. (2017). Technical note: A hydrological routing scheme for the Ecosystem Demography model (ED2+R) tested in the Tapajós River basin in the Brazilian Amazon. *Hydrology and Earth System Sciences*, *21*(9), 4629–4648. https://doi.org/10.5194/hess-21-4629-2017

Pilz, T., Francke, T., Baroni, G., & Bronstert, A. (2020). How to Tailor My Process-Based Hydrological Model? Dynamic Identifiability Analysis of Flexible Model Structures. *Water Resources Research*, *56*(8). https://doi.org/10.1029/2020WR028042

Qiao, X., Nelson, E. J., Ames, D. P., Li, Z., David, C. H., Williams, G. P., et al. (2019). A systems approach to routing global gridded runoff through local high-resolution stream networks for flood early warning systems. *Environmental Modelling and Software*, *120*(August). https://doi.org/10.1016/j.envsoft.2019.104501

Saavedra, D., Mendoza, P. A., Addor, N., Llauca, H., & Vargas, X. (2022). A multi-objective approach to select hydrological models and constrain structural uncertainties for climate impact assessments. *Hydrological Processes*, *36*(1). https://doi.org/10.1002/hyp.14446

Sepúlveda, U. M., Mendoza, P. A., Mizukami, N., & Newman, A. J. (2022). Revisiting parameter sensitivities in the variable infiltration capacity model across a hydroclimatic gradient. *Hydrology and Earth System Sciences*, *26*(13), 3419–3445. https://doi.org/10.5194/hess-26-3419-2022

Spieler, D., Mai, J., Craig, J. R., Tolson, B. A., & Schütze, N. (2020). Automatic Model Structure Identification for Conceptual Hydrologic Models. *Water Resources Research*, *56*(9). https://doi.org/10.1029/2019WR027009

Thober, S., Cuntz, M., Kelbling, M., Kumar, R., Mai, J., & Samaniego, L. (2019). The multiscale routing model mRM v1.0: simple river routing at resolutions from 1 to 50 km. *Geoscientific Model Development*, *12*(6), 2501–2521. https://doi.org/10.5194/gmd-12-2501-2019

Wang, J., Zhuo, L., Han, D., Liu, Y., & Rico-Ramirez, M. A. (2023). Hydrological model adaptability to rainfall inputs of varied quality. *Water Resources Research*. https://doi.org/10.1029/2022WR032484

Yamazaki, D., Kanae, S., Kim, H., & Oki, T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*, *47*(4), 1–21. https://doi.org/10.1029/2010WR009726

Ye, A., Duan, Q., Zhan, C., Liu, Z., & Mao, Y. (2013). Improving kinematic wave routing scheme in Community Land Model. *Hydrology Research*, *44*(5), 886–903. https://doi.org/10.2166/nh.2012.145

Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, *44*(9), W09417. https://doi.org/10.1029/2007WR006716

Zhao, F., Veldkamp, T. I. E., Frieler, K., Schewe, J., Ostberg, S., Willner, S., et al. (2017). The critical role of the routing scheme in simulating peak river discharge in global hydrological models. *Environmental Research Letters*, *12*(7). https://doi.org/10.1088/1748-9326/aa7250

Zhou, L., Liu, P., Zhang, X., Cheng, L., Xia, Q., Xie, K., et al. (2023). Improving structure identifiability of hydrological processes by temporal sensitivity with a flexible modeling framework. *Journal of Hydrology*, *616*(June 2022), 128843. https://doi.org/10.1016/j.jhydrol.2022.128843