

Responses to Reviewer Comments

Dear HESS editor:

Here we want to express our sincere appreciation to you, reviewer Dr. Read, and the anonymous reviewer for your great help with our work. The comments, critics, suggestions, insights and guidance provided to us are invaluable to improving our work. Following these comments, we have carefully addressed the issues raised, added additional analysis (e.g. extra performance metrics and sensitivity tests), and revised the manuscript accordingly.

In addition, we also want to thank both you and the two reviewers for your great patience and continued support for our work despite of the long delay in revisions due to personal difficulties.

The reviewer comments are quoted in normal **black** font and the responses are highlighted in **blue** color.

Sincerely,

Colby K. Fisher, Ming Pan, and Eric F. Wood

Reviewer 1:

The purpose of the paper is to demonstrate a methodology for calculating 'continuous' streamflow from a network of point observations, with the targeted application to estimate global discharge in areas with sparse measurements. The topic and methodology are of great interest to the community as we are making progress on increasing access to observations worldwide and growing the remote-sensing based methods for estimating a host of hydrologic variables including river discharge. I have several main concerns regarding the manuscript and then list several minor ones for improvement:

Major: My primary concern is on the generalizability of this approach in other basins with differing characteristics. Of course the method can be applied, but the performance seems highly unknown. For example, given that there are differences in performance of the method based on drainage area, I can imagine other characteristics - mean annual flow, degree of landuse/urbanization, management, etc., also leading to differences in performance. In my opinion since the method was established in a previous work (2013), this paper should include some sensitivity (besides the distribution of gauges, which was a very nice experiment) analysis, either using real data (preferred) or synthetic.

Yes, we agree that sensitivity analysis is necessary to better understand the generalizability of the proposed approach. In addition to gauge distributions, we have added the sensitivity tests against the wave velocity (a parameter that is hard to estimate/calibrate without observations) and gauge density (number of available gauges to remove in the flow reconstruction).

A new figure, Figure 9, has been added to show the results of two sensitivity experiments: 1) selection of a different velocity parameter, and 2) decreasing amounts of available gauge data. Both of these experiments were carried out using the null initial conditions of runoff. For the velocity parameter, the calibrated velocity of 1.4 m/s produces the best performance; however, there is still skill in the discharge reconstruction for velocity values from 50% to 200% of the optimal value (Fig. 9a). With regards to potentially limited availability of observations, the number of gauges assimilated in the ISR model was decreased from 0% to 50% of the full set by randomly removing gauges (Fig. 9b). In all cases, the KGE values indicate adequate model performance; however, these results are dependent on the information contained within each observation, as removing a gauge with a larger contributing area is likely to have a larger impact on the overall model performance than a smaller one.

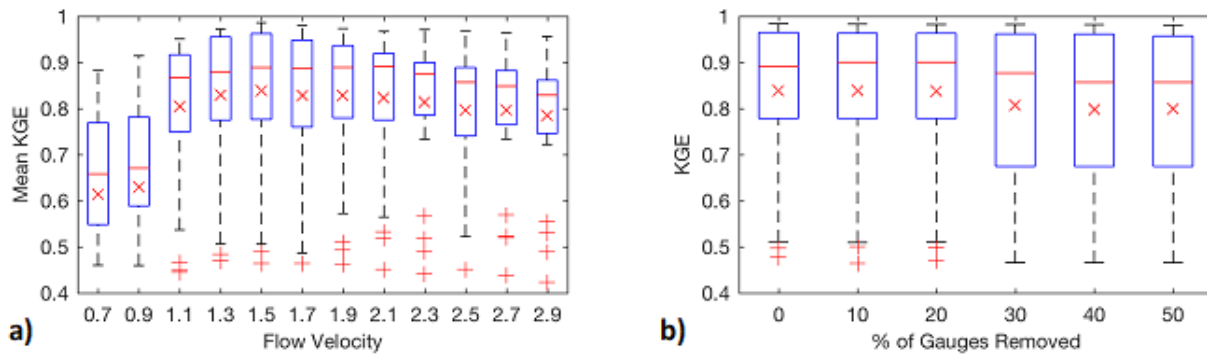


Figure 9: Box plots of KGE values at 25 validation sites for two sensitivity experiments: a) varying wave velocity parameters, b) removing gauges from the observation set. In each of these experiments the null initial condition was used. The mean of each set is denoted with a red X.

Further sensitivity analysis against different flow climatology, geolocations, landcover/landuse, water management, etc. are also necessary. Given that the primary goal of this study is on the introduction of methodology and proof of concept, we think such larger scale experiments should better be carried out in a separate and more application-focused study. So we included such experiments in a second paper that is in review:

Fisher, C.K., M. Pan, and E. F. Wood, 2019: Deriving Continuous Discharge Records from Future SWOT Observations. Water Resources Research, in review.

Further, only one year (2009) is tested - why? Was 2009 particularly wet or dry? These could also impact the results.

The goal of this paper was to perform experiments to illustrate the feasibility of a numerical method to reproduce given discharge records. Year 2009 is an “average” and “typical” year for the study area (Ohio basin) with a normal seasonal cycle and flow values that cover most of the high and low ranges of regional hydrology. There were no dramatic/extended droughts or floods in 2009. Such a choice can minimize the impact of many compounding factors like the deficiencies of simple diffusive wave routing under extreme conditions (high flow and low flow). Also experiments and observations over a “typical” period can be more generalizable.

The second major question is the relative performance of the statistical (kriging) method compared with the one presented. While I agree with the authors that a statistically based approach sacrifices mechanistic understanding, it is important to understand the context of where this work fits in and how it compares (as the authors mention in the introduction).

We agree with the reviewer that it is important for this work to be considered in the context of the previous work done on statistical river kriging (e.g. Paiva et al. (2015) and Yoon et al. (2013)). A careful assessment of the existing statistical approaches shows that it is very difficult to fairly reproduce the kriging results without significant effort from the original authors, given that (1) the study area, study period, training data, underlying parameters etc. are all different, (2) the kriging approach is much more complicated than spatial kriging and its efficacy will depend on training/tuning. Quantitative comparisons against the statistical methods are necessary and we think it is much more fair and convincing to conduct such comparative studies together with all relevant researchers in this area and have all participants working on the same study area/period and input data.

That said, we fully acknowledge that we do not have quantitative metrics to accurately prove the better efficacy of this method. Adjustments to the discussions/conclusions are made to stress this point. Despite this, we believe that the fact that this model accounts for a physical representation of the river system (in the form of a river routing mode, albeit a simple one) is important. We believe that this difference alone provides strong evidence that there should be better physical consistency than with a purely statistical method.

The third major point is on the application itself, where this case was presented for the Ohio river basin, which is a heavily gauged basin (making it great for an initial test, but perhaps poor for demonstrating how the method could be used for global discharge). In my opinion, the likely errors and missing data in the observations are a big concern. In the US we have a relatively low occurrence of missing obs compared with other places in the world where I imagine this could be applied, so the question is: at what point does this method break-down - what are the stress-test results (how many missing obs lead to performance worse than taking the mean or some other metric at the nearest gauge, or applying a statistical method)?

We thank the reviewer for the thoughts on the application of this method to basins with larger errors and missing observations. One way to assess the applicability over other basins and under other conditions is to perform sensitivity tests (as suggested earlier) - see the results from the sensitivity experiments in the response to the earlier comment.

Also, the work presented here was done in the context of the need for global river discharge products for the upcoming NASA Surface Water and Ocean Topography NASA mission. A thorough study of the applicability of this method to observations that are sparse in both space and time has actually been done and is under review in Water Resources Research. In that study the ISR method is applied to synthetic SWOT observations and investigates the performance of the method given the varying discharge estimation errors and intermittent observations of the mission:

Fisher, C.K., M. Pan, and E. F. Wood, 2019: Deriving Continuous Discharge Records from Future SWOT Observations. Water Resources Research, in review.

Minor:

-Only NSE was used as a metric of performance, but there are other statistics that would be interesting depending on whether the user is interested in floods/droughts - 7Q10, bias, etc.

On top of NSE, we have now added more metrics of skill including the Kling-Gupta Efficiency (KGE) score and its three subcomponents: bias ratio (a measure of bias relative to mean), correlation coefficient (a measure of linear correlation), and relative variability (a measure of scaling bias in dynamic range). These new metrics provide a thorough breakdown on how the interpolation may or may not improve the initial guess. Most flood and drought metrics like 7Q10 are very sensitive to extreme values and the values calculated from the data are less robust due to the relatively limited experimental period. So they are not implemented here. Figures 5 and 8 have been updated to include five metrics (NSE, KGE and three KGE breakdowns):

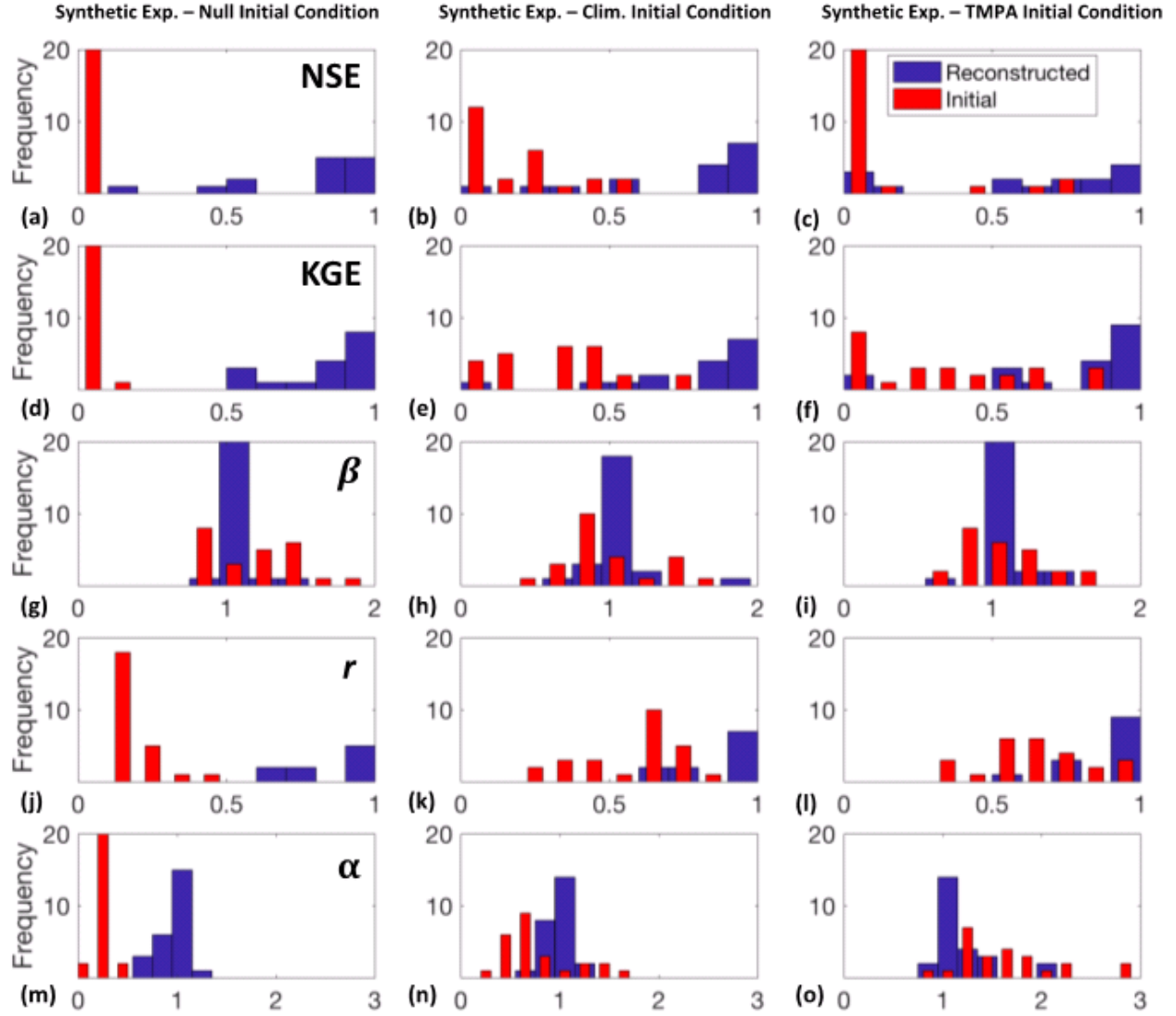


Figure 5: Distributions of NSE values (a,b,c), KGE values (d,e,f) and its component statistics: bias ratio β (g,h,i), correlation coefficient r (j,k,l), and relative variability α (m,n,o), for the three synthetic experiments with varied initial conditions of runoff (shown in three columns). These daily initial conditions are: 1) Null (uniform mean runoff over the entire basin), 2) Climatology (average daily runoff over the entire period from NLDAS), and 3) TMPA (runoff derived from TMPA precipitation and VIC LSM). In each plot, the red bars illustrate the distribution of the statistic values for discharge generated from the initial guess of runoff and the blue bars indicate the same distribution after reconstruction with the inverse routing method.

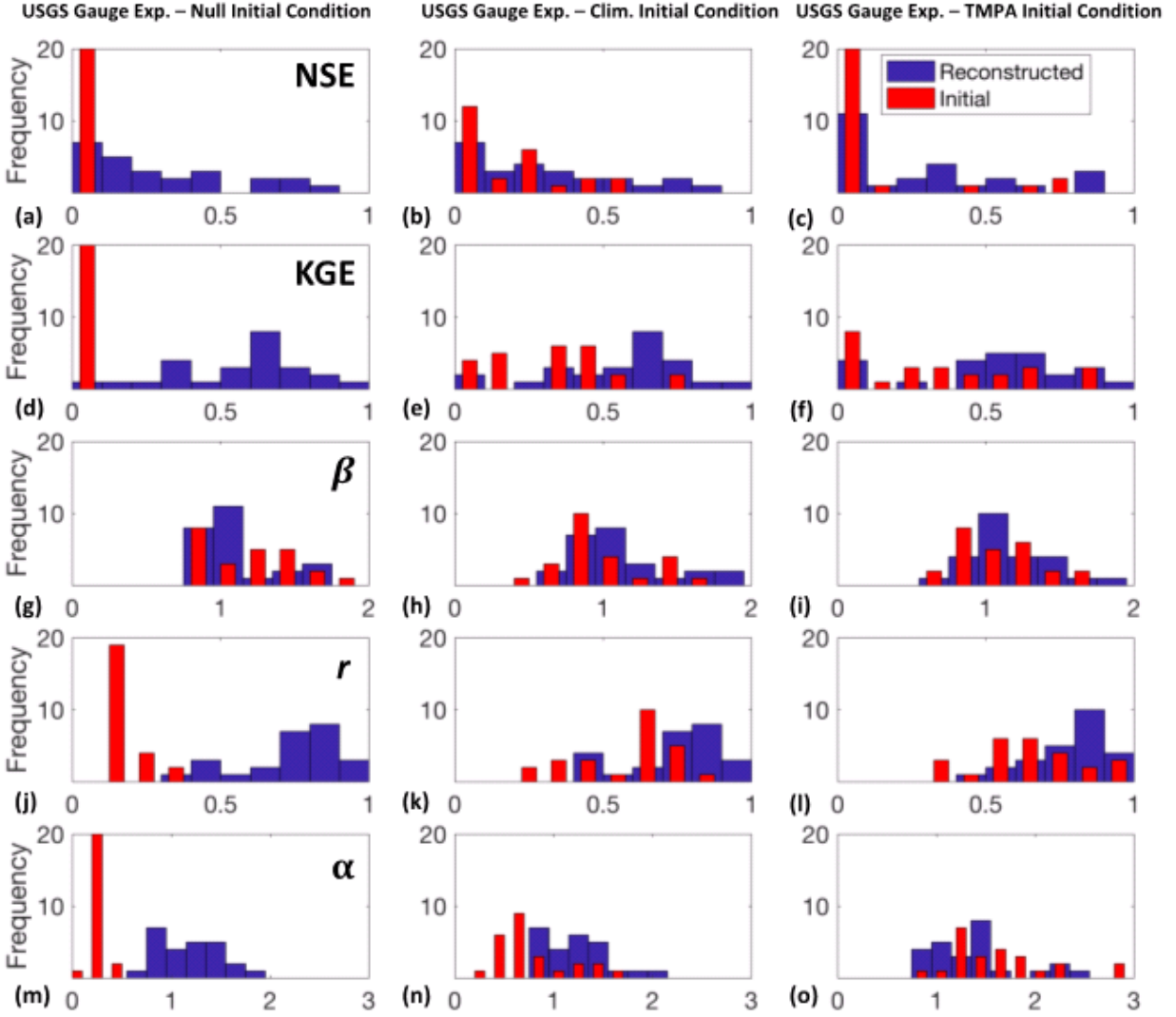


Figure 8: Distributions of NSE and KGE values for the three synthetic experiments with varied initial conditions of runoff and USGS observations as the synthetic truth. The ordering is the same as that in Figure 5.

-Suggest revising the language in the results that often repeats "we can see" (pg 8).

These items have been revised throughout the manuscript to vary the language.

-Avoid saying whether the method did a "good job". Allow the reader to determine that from the results.

Statements such as these have been edited throughout the paper to address this.

-Fig 5: make axes and legend font larger

The figure has been updated accordingly.

-Fig 9: caption has a typo. Dived → Divided?

This has been fixed

Reviewer 2:

My main concern is in regards to the transferability of this method to the remote basins with sparse data, extensively discussed in the conclusion. Given the experiments run, I'm not convinced this method would do any better than the simpler statistical methods typically used for ungauged basins. Line 18-21, page 6, you write "Based on the previous work of Pan and Wood (2013), the wave velocity parameter and the smoothing window for the Ohio River basin were set at 1.4 m/s and 70 days, respectively." In that previous work, Pan and Wood set the values based on a streamflow model calibrated to the gages. How would you get these parameters without having already had a good working model with a lot of data in the area? In my experience, models are very sensitive to wave celerity values, which vary with discharge in non-linear models. So to leave a value constant through time, especially with no model/data to back it up and also considering interannual variability: this all seems quite risky.

We thank the reviewer for the thoughts on the applicability of this model to data poor regions. We fully acknowledge that you need a good forward model for the ISR method to work appropriately. Without this it will be hard to accurately reproduce the discharge in complex basins. Note that the unknown wave celerity (velocity) problem over ungauged basins is not a challenge unique to this ISR based approach but a common challenge to all statistical approaches as well where the lag correlations need to come from somewhere. In Paiva et al. (2015), such lag correlations are based on the analysis of the same diffusive wave model with the need for the same celerity parameter. No method can get a free lunch on the ungauged basin parameter problem simply because we can't make it up, either explicitly in the ISR approach or inexplicitly in statistical approaches, if nothing has been observed. We fully recognize this issue and adjustments have been made to the discussion to stress this caveat. However, poor parameter availability won't be a problem that only the ISR approach will suffer and it is not a reason why it could do less either.

At the same time, modeling communities have proposed different ways to "transfer" parameters from gauged basins to ungauged basins (parameter "transfer" or "regionalizations") according to various similarity measures and physical constraints. Also, the future SWOT satellite mission (a main motivation of this study) will provide river height/width/slope information over many ungauged basins, and the hope is that new data from SWOT may also help infer flow parameters (e.g. lag correlation analysis).

I feel this paper could be greatly improved by backing down the conclusions and instead perhaps alluding to the possibility that such conclusions will be tested in future work somehow (at least comparing the method to other methods in such basins). Or, as the previous reviewer suggested, more sensitivity studies should be presented, in regards to the velocities assumed and the period of record tested on, since the Pan and Wood (2013) paper has already presented the novel method of inverse routing.

Language regarding the performance of the model has been reduced in strength. In addition we believe that the content of the last two paragraphs of the manuscript allude to the future work which will test the applicability of this method to discharge interpolation when discharge observation errors and missing observations are incorporated. Along with this, we believe the focus of this manuscript is different than the work presented by Pan and Wood (2013), where the previous work focused on runoff, versus discharge as presented here. The 2013 paper tried to suggest the potential for reconstructing discharge (see the HESSD discussion threads at), and one of the reviewers believes that this is not warranted at all because no experiments like in this study were performed. See the HESS discussion threads at: <https://www.hydrol-earth-syst-sci.net/17/4577/2013/hess-17-4577-2013-discussion.html>. Consequently, all the related suggestions on flow reconstruction (including the concept of doing so) were removed from the Pan and Wood (2013) paper.

This study puts forward the concept of streamflow interpolation based on inverse streamflow routing for the first time, designs the actual workflow for producing these streamflow estimates, and then conducts experiments to test and validate the interpolation framework. This concept and actual workflow was never framed, tested, or validated in the 2013 paper by Pan and Wood. We think such a step in this paper is critical, even though the main mathematical mechanism for propagating flow information backward in time and space was established in Pan and Wood (2003).