We would like to express our most sincere appreciations to Referee #2 for his/her great efforts in examining the manuscript and invaluable critics/corrections/suggestions. We are now conducting additional investigations and revising the manuscript according to these suggestions and critics.

Ming Pan and Eric F. Wood

Reviewer comments (in blue color) and responses start here.

Manuscript: hessd-10-6897-2013: Inverse streamflow routing

Major remarks

The authors present an interesting approach to estimate runoff at the grid scale from streamflow measurements at river gauges. This approach may help to improve the data availability for specific catchments. Unfortunately the paper is currently written in a way that the reader might think that the authors find the golden solution for all problems related to the availability of gridded runoff and gaps in streamflow time series. Thereby, the authors largely neglect the limitations of their method and associate characteristics with their method that the method cannot hold. Here, the reader might think that the their estimated runoff data are almost as good as observations, which would be misleading for future studies that might compare simulated runoff data to runoff yielded by the inverse streamflow routing.

Consequently a large part of the conclusions section has to be rewritten, as there are a lot of statements that are not justified. These comprise, e.g., page 6915, lines 21, 25++, page 6916, lines 10.

Yes, and we failed to stress a number of fundamental points about what an inversion can do and cannot do, and this has made the article misleading to readers. We are making a number of clarifications in the revisions. A key point about the inversion method proposed is that absolutely no new information can be created from nothing in the inversion process. That means the inverted runoff fields won't contain any more information (derived or direct) than what is already in the inputs, i.e. the streamflow observations (all those within the same contributing sub-catchment and maximum travel time) and the physical knowledge (routing equations/parameters) we have on how the water may flow. That said, the inversion will degrade and fail as all these inputs degrade and fail, for example, gauges too sparse, observations too sporadic, large errors in river measurements, errors in routing parameters, etc.

We are now carrying out a series of new synthetic experiments, as also suggested by another referee, Dr Keith Beven, to study the impact of these above factors, including streamflow measurement errors, gauge density, missing measurements, and routing model errors. An extra section will be added to the manuscript to discuss the limitations of this method and when/where it starts to fail.

The conclusion section, including all the 3 sentences mentioned here will be re-written to carefully constrain the all the remarks within what the actual results are able to support. All claims lacking support from the actual results will be removed.

The test catchment used for the application of the new method contains a large number (75) of stream gauges. Thus, it is not completely surprising that the inverse method yields some reasonable runoff distribution. But it is likely that this large number of gauges is necessary to yield those results. As such a dense gauge network is not present for many rivers of the globe; this certainly will limit the applicability of the method. Consequently this should be tested, which can be easily done by reducing the numbers of gauges used for the inverse routing.

Yes we agree. The gauge density is a big factor for how much runoff information one can infer from the gauge measurements. We are investigating impact of gauge density as mentioned earlier. The Ohio basin is very heavily gauged, as what happens to most other regions in United States. Actually, there are 790 active gauging stations operating in this region, and the 75 we selected are only a small portion (<10%) of all available. We picked the 75 subset for two reasons: (1) many stations are too close to each other on the same river and they contribute too much redundant information because their contributing sub-catchments are vastly overlapped; and (2) we tried to test on a gauge density that covers the maximum capacity of a satellite altimetry sensor. (SWOT's goal is to measure rivers >50m wide, but this is a moving target.) The 75 gauges/490600 km^2 density won't be reached in most parts of the world, and our ongoing sensitivity tests on gauge density will show how low density would degrade the inversion.

Here, gauges should be randomly selected except for the station at the outlet of the catchment that may always be part of the subsamples of gauges. For lower number of stations, certainly the complexity of the topography might play an important role.

Yes, and we are testing on this in the reduced gauge density experiments.

Given the large number of stations within the catchment, it is necessary to evaluate the added value of the inverse routing technique in comparison to the more simple approach of distributing every measured streamflow over the associated sub-catchment (For example, if a sub-catchment has one inflow Q1 and an outflow Q2, then the discharge Q2 is equally distributed over the area A2 solely belonging to the gauge station 2, while Q1 is distributed equally to the catchment area A1 of station 1. Note that the full catchment area of station 2 is A1+A2.). Will similar pattern arise as for the inverse routing method? If yes, then the inverse routing is likely not adding much value to the distribution and its quality just originates from the large number of stations. (Note that a hint on this is given on page 6911, where it is written that 'the shape of patches follows the boundaries of sub-basins that drain to the input gauge locations.') If a more reasonable pattern is estimated, then this would show an added value of the method.

Another reviewer, Dr Keith Beven, also made a similar suggestion of comparing the inverse routing to a simpler approach. What is described here is a sub-catchment water balance approach.

We are fully aware of this approach as it's already being used by other researchers (through personal communications). The experience from those who use the water balance method shows that the main difficulty of this approach is how to handle the time delay (especially for larger sub-catchments). As a result, the water balance method is only used in relatively small regions (shorter travel time) to estimate monthly runoff so that the travel time difference doesn't quite matter. It is really hard to get daily runoff well resolved in the Ohio basin (15-day maximum travel time) by water balance because it may take days for the downstream gauge to receive a flood peak while upstream gauges are flooded already. If the travel time mismatch is not resolved, a lot of negative runoff estimates will show up in downstream areas simply because the downstream gauge hasn't received it yet. In other words, the water balance method fails at fine time scale. This is one important motivation for us to propose this new approach with a more sophisticated accounting for the travel time such that the runoff can be estimated at a finer time scale.

In summary, I suggest that the authors should focus on what their method can do, and clearly point out what it cannot. Then, the paper may be accepted for publication after major revisions have been made.

Minor Comments

In the following suggestions for editorial corrections are marked in Italic.

p. 6898 – line 17

It is written: "Now inverse routing bridges the gap and provides a best, if not only, mean to estimate runoff field at any spatial or temporal scales from observations."

This a good example for the major problem of the paper addressed above. First, for long-term annual averages, observed streamflow measurements are sufficient to bridge the gap. Second, for specific catchment with a high number of stations (or simple topography), the method may bridge the gap on finer temporal scales, but this will likely not be the case for catchments with only a few gauges and/or a complex topography.

Agree. We are conducting sensitivity experiments with respect to gauge density and the sentence will be carefully re-written according to the new results.

p. 6901 – line 17

Eq. 1 is not the St. Venant equation. It is a diffusive wave equation!

Yes. Thanks so much for pointing this out! (The same thanks to Dr. Keith Beven!)

p. 6902 – line 4

… prescribed and *are* independent …

Yes. The text is corrected.

p. 6904 – Sect. 2.2

If you only account for travel times, then retention characteristics within the catchment are neglected. This is part of the simplifications of your method (which is ok.), and should be stated.

Yes, this is the case. This is added to the method description. The retention behavior and similar effects from lakes and reservoirs make the routing model more nonlinear and the inversion much harder. So, we are still far from having a "golden solution" :)

p. 6908 – line 24

If D is set to 0 all over the basin, then no diffusion takes place, Then Eq. (1) becomes a simple advection equation. If this is part of the method, this is ok. But then I don't understand all the effort you make by using Eq. (1) in sect. 2.2 and all their related derivatives. It seems that you initially pretend to use a sophisticated advection-diffusion approach, which later on turns out not to be the case.

Actually we didn't expect to see D = 0 when we started formulating the method. The D = 0 came purely out of the routing model calibration process. The routing model is designed to work under a wide range of scales so it includes both the advection and diffusion. We chose to run everything at daily time step to control the computational cost, and an unexpected side effect of this large time step (relative to the 12 km pixel size) is that the diffusion effect basically disappeared when everything is aggregated up to daily. As a result, the calibration suggests D = 0. The diffusion effect will show up when a finer time step and/or larger pixel size is used.

p. 6910 – line 8

… usefulness *of* inverse …

The text is corrected now.

p. 6911 – line 15

Here follows another example for statements that suggest more than the method can probably do, as 'the very strong capability' of the method is very likely caused by the high number of gauges (see major remarks above).

Yes. We are conducting sensitivity experiments with respect to gauge density and will revise such statements accordingly.

p. 6912 – line 2

… case *in* Fig. 4 …

The text is corrected now.

p. 6912 – line 21-27

Why innovation is defined as the difference between the synthetic truth (thick green line) and initial guess (blue line) of streamflow?

I would expect that innovation, which is caused by the inverted method, is somehow related to the streamflow constructed from the inverted method. Please clarify.

In dynamic system theories and data assimilation literature, the term “innovation” is particularly defined as the difference between the actual observations and what the model would initially guess. This is purely a tradition/convention, see Anderson and Moore 1979.

The difference between the initial guess and inverted runoff is usually called the filter/smoother “increment”. We haven’t seen a particular term widely used for the difference between reconstructed and initial guess of streamflow.

p. 6913 – line 9-11

It is written:

‘As the inverted runoff fields can perfectly reconstruct the streamflow time series at 10 input gauge stations in Fig. 8, they can also reconstruct the streamflow at any point on the river network’.

Actually I strongly doubt this. The perfect match is just by definition of the method. At least you may show this by reducing the number of gauges, and then by simulating streamflow at gauges that were not used for the inverse routing. To show real positive characteristics of your method, you have to significantly reduce the number of stations (see also major remarks above).

Yes, we agree. This statement is lacking substantial support from the existing results and will be removed. We are conducting sensitivity experiments with respect to gauge density too.

p. 6913 – line 15-17

Again I doubt the generality of this statement. It will work in catchments with good river gauge density. But it will likely fail in large catchments with a low number of stations.

This statement is lacking substantial support from existing results and will be removed.

p. 6913 – line 25

It is written: …not the same well as…

Grammar seems wrong – please correct!

The sentence is rephrased now.

p. 6914 – line 1

… consistently *lower* basin …

Yes, this is a much better way to say it. The text is fixed.

p. 6914 – line 6.

… but *is* often …

The text is fixed.

p. 6914 – line 8-11

It is written:

'This suggests it is more difficult to make significant improvement to the initial guess using real gauge measurements, especially when the initial guess is already very reasonable. Large biases can be easily corrected but small spatial details are much more difficult to recover.'

This is directly related to the effect that runoff is estimated by the method, not reconstructed, and that the station density plays a role. More spatial details could obviously only be recovered if the station density would be higher! (see also major remarks)

Yes, we completely agree. We are conducting sensitivity experiments with respect to gauge density now.

p. 6914 – line 18

… streamflow *is* not the natural …

The text is corrected now.

Figs. 3, 7, 8, 10

Axis descriptions, legends and panel titles are too small. Please increase their size.

Yes, the font size is raised.