We would like to express our most sincere appreciations to Dr Keith Beven for his 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.

The authors are correct in their starting point that it would indeed be good to have a method of going from measured streamflows to spatial patterns of runoff generation using an inversion methodology. They also recognise that this is fraught with problems and needs to be constrained. They provide a procedure that gives spatial maps of runoff but, to my mind, they oversell what information is added. This is partly because they do not closely at the sub-catchment patterns produced – most of the pattern in their maps comes from the differences between stream gauge observations of subcatchment specific discharges. It is also partly because they assume that uncertainty in their input data is negligible (as this makes the inversion much more difficult).

First of all, 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 and caused some misunderstandings here as well. So let us first try to clarify on a few things:

(1) No new information can be created from nothing in the inversion. 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 Referee #2, to study the impact of these above factors, and an extra section will be added to the manuscript to discuss the limitations of this method. Also, to avoid "overselling", all claims lacking support from the actual results will be removed.

(2) The purpose of setting streamflow error variance to zero is by no mean to assume the uncertainty in the streamflow is negligible. The main purpose is to ensure perfect physical consistency of the inverted runoff field (reproduction of input streamflow). Another reason is that we want the inversion to work without initial guess (an important feature we want), and in this case, the only choice is to set streamflow errors to zero. There are, of course, errors in the streamflow, and under our setting, ALL these errors in the streamflow observations will pass onto the inverted runoff field, which allows us to

treat the streamflow errors SEPARATELY from the inversion process itself (not to ignore it).

Setting streamflow errors to non-zeros will not increase the computational complexity of the inversion, but it does make it “much more difficult” in the sense that you have to find the most appropriate relative error numbers for the streamflow measurements and initial guess – this is a huge task because data assimilation is essentially a game of finding a good trade-off between them. Also, inversion with non-zero streamflow errors will break the water balance by adding to or subtracting from the observed amount, greatly complicating the error analysis. The error propagation during the inversion process is already quite complicated, and including observations errors here will make the problem even messier, if not intractable. So instead of swallowing two hot potatoes in one bite, we prefer to isolate one from the other and investigate one at a time.

In essence this is not so different to many geophysical inversions which produce nice images but which cannot be checked. Here the check is on reproducing the gauged discharges, but that is also the input that produces the patterns.

Yes, our inversion is just another geophysical inversion. The key thing is what underlying geophysical model you use. So far the runoff interpolation methods we have encountered in literature are mostly based on assumptions like the lag correlation and water balance. We use a full-scale routing model and apply much more physical knowledge we have on how the water may move in time and space to the estimation problem. As a result, we achieve a physical consistency in time and space (reproduction the gauge discharge) that is not seen elsewhere to our knowledge (the referee may a better knowledge on this though?). We think this “not so different” inversion does make a significant difference hydrologically.

About the check on the inversion: In order to check the consistency of the inversion method itself, it should be checked against the inputs first. We absolutely agree that we need to check it against other independent data sources, like precipitation, as well. In fact, our check is not against the inputs only – we compared the inverted runoff to the estimates from the LSM (the “synthetic truth” in Figures 4, 5, 6). These “synthetic truth” runoff fields are derived from NLDAS rainfall data by the LSM and they are not just the “input” to the inversion. And as suggested later, we will add comparisons against rainfall patterns in the revisions.

For this paper to be publishable I would like to see the authors add an investigation of the inferred sub-catchment runoff patterns in relation to rainfall patterns. The runoff should show some time delay banding that might be quite unrelated to rainfall patterns (in some cases the inferred bands could apparently match elevation related precipitation patterns, but that should not always be the case).

Yes, we will compare the inverted runoff to the rainfall patterns in the revisions. There is time delay effects in the inverted runoff but not very pronounced within the sub-catchments because most of the spatial patterns and contrasts are created by small sub-catchments that have a very short travel time. We will provide a zoom-in image with finer color scales to show the stratification caused by travel time progressions.

I would also like to see an investigation of the impact of uncertainty on the inversions. The assumption that the gauged discharges are known without error is clearly a useful constraint, but not a very realistic one. We know very well that even USGS gauge data can be subject to some uncertainty, and for both high and low flows that might be significant. It seems unlikely that they may well have looked at this in the Kalman filter framework, but they have chosen not to present the results. Why is that? Some negative values were inferred already (p17). Does the inversion become unreasonably or unstable in the face of realistic uncertainties?

We are now carrying out a series of new synthetic experiments, as also suggested by Referee #2, to study the impact of a number of factors including streamflow measurement errors, gauge density, missing measurements, and routing model errors. The purpose of these extra experiments is to study how the inversion degrades and fails due to these factors and an extra section will be added to the manuscript to discuss the limitations of this method.

In the new experiments, random perturbations will be added to the streamflow measurements, but the inversion will still use the "zero error" setting. Again, as we said in the response to an earlier comment, we do not assume the gauge measurements are error-free, and the zero error setting is merely a tool to enforce physical consistency and a way to help limit our uncertainty analysis to the inversion process itself.

We did do some limited Kalman filter experiments with non-zeros streamflow errors but found it hard to interpret two types of errors simultaneously. The negative runoff issue is no worse than the zero error setting in these experiments. The negative runoff arises not from the error setting but mostly from a particular condition where the downstream measurement is smaller than or too close to the sum of all contributing upstream sub-catchments. In our synthetic experiments, the occurrence of negative runoff is extremely rare (<0.1%) and negligible. In real USGS data, this unbalanced condition happens slightly more but only when the downstream gauge basin is very large. It happens because the water gets extracted or diverted in the middle (for irrigation or other purposes) and the inversion has to put the "extra" upstream water back onto the land. It also happens on other places we encountered, for example, the downstream gauges on Nile river often report smaller flow than upstream gauges in part due to the re-infiltration as it flows across the dry land. We don't think the "negative" runoff is a big problem as long as it could be reasonably interpreted or attributed, e.g. flow regulation, re-infiltration, evaporation, and so on.

Some specific comments

P3 When precipitation is measured by rain gauges, radars, or satellite sensors, the measured value is validated at the same time and location where it rains or snows. So is evapotranspiration by towers/satellites and soil moisture by probes/microwave sensors. it is not fully equivalent to runoff. – altimetry is not even equivalent to discharge at the same point without an (uncertain) rating curve.

Yes, the water stage measurement needs to be translated to flow through a rating curve and this involves uncertainty. This happens to both USGS river gauges and satellite altimetry and the difference is probably that USGS may be able to do a better curve calibration. The text is changed to reflect this.

p5 Equation 2 is NOT the St. Venant equations but a Diffusive Wave approximation to it.

Yes, exactly. The text is changed. Thanks so much for pointing this out! (The same thanks to Referee #2!)

Figures 4/5. The authors are overselling their technique somewhat here. The scale of the figures disguises the fact that the routing depends only on the distances from the outlet + the routing model parameters – which are assumed known and fixed. That means that points at similar routing distances from the catchment outlet will have similar routing characteristics.

Yes, the inversion relies entirely on the information from streamflow measurements and routing parameters. There is no way to distinguish the response from pixels with similar routing distances or travel times unless extra gauges can be installed. We fully recognize this limitation, and this limitation is due to the availability of information.

Thus the structure apparent in the figures is coming directly from the discharges at the gauge sites. Differences between pixels at similar routing times from the subcatchment outlets cannot be distinguished (this is part of the ill-posedness of the inverse problem. Thus very different visual patterns would result from the inversion if, for example, the number of gauges used was halved. Thus the authors suggestion of robustness here is not being defined with respect to the identification of actual patterns of runoff (and reproduction of the actual gauge discharges is being enforced without uncertainty). As in any inverse problems, yes you get an answer, but the authors are not giving any indication of the realism of that answer over and above being consistent with the gauged discharges that they assume known.

We agree and the inversion problem is indeed highly under-constrained. As a result, the effectiveness or robustness of the proposed method should be studied together with data availability. To provide better "realism" check on the answers, we have compared it to the LSM derived runoff field and are comparing to rainfall pattern as well. The effect of cutting down gauge density is also being investigated.

In fact, for some of these patterns, simply plotting an explicit fixed delay specific discharge in each subcatchment would probably not look so much different (without all the matrix inversions!!). I would suggest that the authors should add such a comparison to demonstrate how much added value there is in their method.

The sub-catchments overlap a lot (downstream ones contain upstream ones), and we can't figure out a really "simple" way to make a map of sub-catchment specific discharge without at least resolving the water balance relationship from the drainage network. One can also use those small sub-catchments that are mutually disjoint in space, but there aren't many left then. And should the "fixed delay" be the same for both large and small sub-catchments? Or not? If the referee can provide more details on how to construct such a daily specific discharge map in a simple way without matrix inversions, we are happy to compare it to our complicated approach.

Referee #2 suggested using the water balance approach (according to catchment-subcatchment summation relationship) as the baseline to calculate the added value. 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.

P20 For example, runoff fields inverted from the future SWOT mission can be used to identify and correct missing or overestimated precipitation estimates from the Global Precipitation Measurement (GPM) mission.

Surely not????? 1. SWOT will only give discharges with uncertainty. 2. Inversion gives only runoff generation, a further (highly nonlinear) inversion would be necessary to get back to rainfall. A Kirchner type inversion requires single valued storage-discharge functions. 3. Inversion adds uncertainty – will this be sufficiently constrained to be used for corrections? The authors should leave this out until it is proven useful.

We surely need to constrain this statement very carefully. Also, we failed to make this more clear in the manuscript – we did not mean Kirchner type of inversion and instead we mean to constrain satellite rainfall errors through the mass balance of water at surface. The surface water balance relationship is: $ds/dt = p - e - r$, where s is the total soil water storage, p is the precipitation, e is the evapotranspiration, and r is the runoff. This is a linear function and there is no need for the inversion of the storage-discharge relationship.

There will be errors (presumably large) in satellite altimetry derived streamflow. Whether the altimetry would contribute positively to the whole water cycle estimation will depend on the relative magnitude of errors in all different terms and their independency. We know that satellite rainfall (e.g. TRMM) can have very large errors on stormy days (easily exceeding 100%), which is far larger than the ground river gauge errors. The satellite altimetry sensor will be much worse than ground stations but SWOT design goals (a few centimeters RMSE in height measurements) seem to give a better prospect than rainfall sensors because the uncertainty in the stage-to-flow relationship (slope will also be measured for estimation through Manning's equation) still looks better than the microwave emissivity-to-rain rate relationship. On a wet stormy day where the ds/dt and e terms are minor to p and r, we see a reasonable potential for the altimetry measurement to make some positive contributions.

Abstract and Conclusions: such studies are limited to scales where the spatial and temporal difference between the two can be ignored.

But as noted above, this is really only a way of distributing the specific discharge at a gauge according to some transformed distance scale / time delays. It does not properly distribute in space because it cannot differentiate between effective different time delays. It would be really interesting to see what effect this has at much smaller scales, when the time delay "bands" inferred by the method should be apparent.

We are not sure if we are reading the comment that "it does not distribute in space because it cannot differentiate between effective different time delays" correctly. Bands of different time delays within the same sub-catchment do exist in the inverted runoff, but they are not pronounced due to the stronger contrast between different sub-catchments. Another reason is that we use a large time step (daily) while a lot of the sub-catchments are small, and it takes no more than 2-3 days for the water to flow through them. So there won't be many daily bands within these small sub-catchments. We will provide a zoom-in image with finer color scales to show how that happens. Also, even if the time banding effects are not well shown on one image, it doesn't mean the travel time is not important or can be replaced by a "fixed" value – see our response to the earlier comment about creating the specific discharge map using a "fixed delay".