Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2020-32-AC6, 2020

Nonlinear Processes in Geophysics
Discussions

Interactive comment on "Hybrid Neural Network – Variational Data Assimilation algorithm to infer river discharges from SWOT-like data" *by* Kevin Larnier and Jerome Monnier

Kevin Larnier and Jerome Monnier

jerome.monnier@insa-toulouse.fr

Received and published: 10 November 2020

Anonymous Referee #2
Received and published: 5 November 2020

This study aims at proposing the hybrid Neural Network (NN) – variational data assimilation algorithm to estimate river discharge from simulated SWOT like data. Such methodological studies are very important and of the scope of the NPG. In addition, investigating the potential benefits of satellites prior to the launches is quite useful to improve satellite missions further. However, I think the present manuscript has some fatal issues that should be solved prior to publication. The authors seemed to investigate the method that would not be applicable to the

real ungauged river basins as elaborate below. I am compelled to suggest this manuscript be rejected.

Major Issues 1.
As described, the SWOT-based estimation of river discharge is useful for ungagged or poorly gauged river basins (P1L14). However, the authors used "too rich" basin information. They used dA (difference in cross section), W (river width), S(slope), and A (cross section) to estimate Q (discharge) by NN (P8L166).

We are so sorry that you have completely misunderstood the addressed inverse problem and the developed method.
The considered information are the measured quantities by SWOT (dA, W, S) (you are right) plus $\mathcal{A}$ the local drainage area (in km^2); obviously (or unfortunately !?) not the river cross-section A (in m^2)...
Undoubtedly, there is a serious misunderstanding; moreover there was a typing error P8L166.
However, this crucial point was indicated in the abstract, in the general introduction (P2L58), in Section 2 (P6L135), in the ANN description (Section 3), in figures titles and in the general conclusion. The only input information in addition to the SWOT like measurements is $\mathcal{A}$ the local drainage area.
However this was not recalled P8L166; now it is done. Moreover, P8L166: obviously, the knowledge of dA does *not* imply the knowledge of A_0... (typing error which have now been corrected).

The employed values of $\mathcal{A}$ are those available in HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales).

Moreover, the Artificial Neural Network (ANN) provides a first rough estimation of Q at reach scale only, see Section 3. The latters being next improved by a (low complexity)

algebraic flow model (Section 4) and finally by the inversion of a complete dynamic flow model inversion (Section 5).
We regret that you did not see the answers to RC1 published sept. 25th; these answers would have brought you some additional clarifications and clues on the addressed challenging scientific problem.

In the revised version published on the journal website Oct. 3rd, see https://npg.copernicus.org/preprints/npg-2020-32/npg-2020-32-AC5-supplement.pdf, we have better highlighted the hypotheses in many locations of the manuscript: in the new abstract, in the general introduction, in the data section, in the conclusion, and in almost each section. Also we have included a flowchart of the complete inversion algorithm with the indication of the unique prior $\mathcal{A}$, the SWOT-like input variables and the output variables, see Fig. 8 P19.

* *The physical based models, which were also used to mimic observation data, simulates Q based on dA, W, S, and A with only one major uncertainty parameter: frictions of river channel. Namely, there is one equation and one uncertain parameter. Solving this problem is too very easy for NN.*

Again, this is obviously not the addressed inverse problem... As indicated throughout the paper (see above), the addressed inverse problem consists to infer : the discharge value $Q(x,t)$ and effective pairs (friction parameter $K(h(x))$, bathymetry $b(x)$ - or equivalently $A_0(x)$).

You are right, if the problem was to solve a single equation with a single parameter, one line of trivial calculation would have been enough.
We recall in Section 5.3 the "Capabilities and limitations of the inversions based on the

flow models only". This section mathematically shows the inversions capabilities from the standard flow models (this includes the basic Manning-Strickler's law of course). To our best knowledge, this basic but very informative analysis is original. Moreover, it nicely explains the obtained bias when inverting physically-informed models if no prior information (eg an accurate mean value of Q) is available, see the cited references or the intercomparisons studies [Durand et al. 2016], [Frasson et al., 2020] (submitted).

* Let us provide some additional point-to-point answers below.

When we refer to the manuscript, we mention either page-lines numbers of the original version (you have received) or the page-line numbers of the version published on the journal website Oct. 3rd (https://npg.copernicus.org/preprints/npg-2020-32/npg-2020-32-AC5-supplement.pdf).

* *Consequently, the present experimental setting of NN was very confusing to me. It is usually impossible to use the cross section A because the cross section under the river surface is unobservable by satellites. The challenge for realistic applications is to estimate Q without using A.*

It is unfortunately a misunderstanding of the considered inverse problem and the developed methods. Please, refer to the previous answer.

* *2. The authors assumed unrealistic dailySWOT observation data while real satellite revisits 1-4 times per 21 days (P1L22). Consequently, I strongly suggest the authors re-consider experimental design that is applicable to real problems.*

You are right, the considered SWOT-like data are synthetic, 1-day repeat. They cover however a very large rivers sets with very different flow characteristics. Moreover this responds to an important science issue, at the forefront of the current Discharge Algorithm Working Group (https://swot.jpl.nasa.gov/documents/4050/).

As mentioned in our RC1 point-to-point answers, the first three months after launch, the instrument will be on a 1-day revisit period; this is the important "fast-sampling" Cal-Val phase, see [Rodriguez, JPL, 2012]. This is the context of the present study.
This point was not sufficiently highlighted. Now, it is much better indicated throughout the manuscript, including in the new abstract, in the general introduction and conclusion, and of course in the data section too,(see the new version https://npg.copernicus.org/preprints/npg-2020-32/npg-2020-32-AC5-supplement.pdf).

Note that if considering the nominal SWOT orbit (which will provide data with 21 days revisit period, depending on the latitude), the scientific challenge which consists to solve the ill-posed inverse problem for ungauged rivers posed by the mission remains the same (see Section 5.3 of the manuscript).
In this case, the time validity of the discharge estimation equals the wave travelling time through the river portion (roughly, a few hours to a day, depending on the case), see eg. [Tourian et al. 2017], [Brisset et al 2018], [Larnier et al. 2020] (with the identifiability map concept in particular). This point is well understood now.
The present remark has been added in the dedicated new section 3.4 entitled “On the sensitivity of the estimations with respect to error measurements or data frequency”.

Moreover let us point out that in the present ANN, the concept of spatial correlation or time correlation between examples does not exist. Indeed, the ANN input variables are dA, W, S and $\mathcal{A}$; one “example” corresponds to a set of $(4+1)$ values which are point-wise, snapshots. No space or time correlations exist between two “examples”. As a consequence, the ANN does not “see” the potential space and time correlations

in the dataset. In our case, if considering less frequent observations (eg. with few days frequency), but of course with similar volume and quality of data, the accuracy of the trained ANN would be similar. We have investigated this assertion for a frequency of 5 days (results not shown here). As expected the obtained accuracy were of same order of magnitude than those presented in Table 2 (new version of manuscript). Obviously, in this case (eg. with 21 days revisit) and for the reason previously mentioned (see the identifiability map concept introduced in [Brisset et al. 2018], [Larnier et al. 2020]), the discharge estimations remain valid for a few hours - a day around the observation instant only.

* *[Other Issues] 1. Experimental design is unclear to me. It is better to add a schematic that shows a flow chart of data used in this algorithm.*

Thank you for your remark. Following this remark and RC1 comment, we have added a flowchart of the complete inversion algorithm with the indication of the prior, the input variables and the output variables. Please, see Fig. 8 p19.

* *2. The paper should add more hydrological papers for reference. For example, I found a data-driven estimation of river width from satellite data (Yamazaki et al. 2014). Comparisons to such existing approach would be beneficial to add values of the manuscript.*

Thank you for mentioning this reference. This reference has already been cited in the new version (Oct. 3rd, see https://npg.copernicus.org/preprints/npg-2020-32/npg-2020-32-AC5-supplement.pdf, see P 5 L40) to mention a potential river width database (the Global With Database). However, this reference does not address at all the present inverse problem.

Moreover, we have added too: [Paiva et al., WRR 2015], [Tarpanelli et al., IEEE 2018], [Lin et al., WRR 2019].

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2020-32, 2020.