Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2020-32-AC3, 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: 3 October 2020

[12pt]article

Anonymous Referee #1
Received and published: 3 September 2020

In this paper, the authors developed a new method for river bathymetry and discharge estimation from satellite altimetry data. They firstly estimated river discharge from satellite-observed water elevation data by machine learning. Then, using the estimated discharge and water elevation, they performed the inversion of a hydrodynamic modelto estimate bathymetry and

other parameters by variational data assimilation.General comments: Although the topic of this paper is suitable to NPG, I believe that this paper has some fatal flaws which cannot be fixed in the short period of time. I believe that the current version of the paper cannot be accepted.

First, the design of the authors' synthetic experiment is inappropriate. Their synthetic observations were fully generated by hydrodynamic models with no observation and model errors. They may not consider the real satellite swath, and the temporal resolution of the data (daily) is much higher than the real satellite altimetry. I believe that they have too rich data to examine the potential of SWOT. The richness of the observation data significantly matters when the fully data-driven approach such as neural network is applied but the authors completely ignored this issue. I strongly recommend the authors to perform numerical experiments with more realistic data.

*

On the "inappropriate synthetic experiment". Data are synthetic, that is correct. These data have been generated by calibrated models (Hec-Ras and LisFlood in particular) by a relatively large community. They constitute the reference Pepsi and Pepsi-2 datasets of the SWOT Science Team. They contain 56 000 "examples" (in the machine learning sense) representing 29 heterogeneous rivers; this is a volume of reference data (considered reliable) never reached up to now in the community. These datasets are the most advanced ones in the SWOT community for assessing and benchmarking the different algorithms on a large panel of different rivers presenting different flow regimes, see [Durand et al. 2016, Frasson et al. 2020]. These data are not as realistic as those considered in [Tuozzolo et al. 2019] for example. Indeed, this last reference constitutes one of the first study (among very few today) based on real SWOT like data (airborne ones, NASA AirSWOT mission). (We, the authors of the present manuscript, have performed our previous version of

algorithm for [Tuozzolo et al. 2019]; it was one of the two employed and compared algorithm). Considering real Airborne data or data from the SWOT science simulator (https://swot.jpl.nasa.gov/mission/overview/) does not affect the method-algorithms analysis. It might affect the accuracy of the estimations but not the crucial features of the method.
Considering synthetic data with no noise before considering data with some Gaussian noise is a mandatory step to better understand and evaluate a method capability. This is what is done in the present study following the Discharge Algorithm Working Group research plan of the SWOT Science Team, see [Durand et al., 2016] and Frasson et al. 2020] Phase 1. Moreover some noisy data have been taken into account but the results were not shown in the present study (since the study focuses on Phase 1).
However, following your remark we have added a few comments from the results we have obtained with noise. Please read our more detailed answer to your final comment on "sensitivity analysis".

On the "temporal resolution".
That is correct, the considered data frequency is 1 day only. This corresponds to the important Cal-Val orbit phase of the satellite.

In short, the considered datasets are synthetic, 1-day repeat, covering a very large rivers sets with very different flow characteristics. This responds to an important science issue, at the forefront of the current Discharge Algorithm Working Group (https://swot.jpl.nasa.gov/documents/4050/)

You are right, these points were not sufficiently explained. Now, they are 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.

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 or our next answer to your comment).
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".

On the "neural network" estimations.
Recall that a standard ANN is interpolator but based on a complex multi-resolution model (defined by its architecture) and a large volume of data. After optimization (training stage), the ANN has "identified-learned" invariants, correlations between the four input variables and the output variable Q, see eg. [Mallat (2016) Phil. Trans. R. Soc. A 374: 20150203]. As you know, the results obtained by ANN can be astonishing including in fluid mechanics, see eg. [Brenner, et al. (2019). Physical Review Fluids, 4(10), 100501] , despite no one fully understand how it actually works yet.

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 correlation nor time ones exist between two "examples".
As a consequence, the ANN does not "see" potential space or time correlation between the datasets.
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. Obviously, in this case (eg. with 21 days revisit) and for the reason previously mentioned (identifiability map), the discharge estimations remain valid for a few hours - a day around the observation instant only.
Note that we have performed many other tests demonstrating the robustness of the present ANN estimations and eg. their insensitivity to the test – train river sets.

A remark on these points (non-correlated feature of the examples and robustness of the ANN estimation for less frequent observations) has been added in the dedicated new section entitled “On the sensitivity of the estimations with respect to error measurements and data frequency”.

Again, all the points you mention were not sufficiently clear in the manuscript, or even not mentioned for a few of them. All are now much better highlighted in the new version. Please read the new abstract, the new general introduction, the new conclusion and the sections 2 and 3 in particular.

*

Second, the advantage of the proposed method is unclear for me. In my understanding, there are many methods to infer river discharge from water levels.

Correct; we have tried to present a relatively large bibliography in the general introduction.

*

The authors omitted to compare their neural network with those previous works so that I am

not convinced that machine learning is necessary in this context.

We do not agree; the present study demonstrates that a machine learning approach (the ANN) can help to solve a crucial step in the inversions.
Firstly, note that the mentioned bibliography in our introduction is quite complete; the unsolved issues are clearly explicited.
Secondly, whatever the adopted physically-based inversions method (Kalman Filter or Variational Data Assimilation), one of the most remaining critical challenge is to determine theÂăprior information, in particular the first guess value(s) of these iterative algorithms, see eg. a related discussion in [Frasson et al 2020]. Until recently, this point was not really, or at least not sufficiently, discussed. (Note that this point becomes even more critical if the study relies on a single river only). In particular, the VDA physically-based approach enables to capture space-time variations like almost no other published method does (see the cited bibliography), however a shift remains: it is the bias we address in this article. Please re-read the abstract, the introduction and Section 5.3. in particular.
After optimization, the bias value was depending strongly on the first guess value (and of the method of course too).
Here, this crucial issue seems to be solved for ungauged rivers by a machine learning approach (the ANN) plus the hierarchical flow model for rivers belonging to the learning partition (denoted here by Q-Lset). This is new, robust and absolutely promising. This is evaluated and analyzed for a large number of rivers.
Moreover, as clearly mentioned in the manuscript, the ANN estimation is not a “final product”. The ANN estimation is greatly improved first by the algebraic flow model, second by the VDA process (see this description eg. in the abstract).
As a consequence, the estimation to be compared with the other approaches would be the final estimation and not this intermediate purely-data driven value. Moreover, this purely data-driven estimation is implicitly compared to the “final” estimation (since being the “rough” basic estimation”). And the latter is implicitly compared to others

state-of-the-art methods through the previous articles eg. [Durand et al. 2016, Tuozzolo et al. 2019, Larnier et al. 2020, Frasson et al. 2020].
Recall that we have directly participated to likely the most extensive comparisons published up to now [Durand et al. 2016, Frasson et al. 2020], [Tuozzolo et al. 2019]; benchmarks based on a large number of synthetic rivers plus one of the very few Airborne dataset (AirSWOT mission). These comparisons have been performed with our former algorithm(s), the purely deterministic VDA method presented in [Brisset et al. 2018], [Larnier et al. 2020] (and implemented in the same computational code as the present one, a former version of course). As a consequence, this very solid experience of benchmarking has provided us numerous reference results to compare with. These experiences enable us to claim the results we obtain here (and to implicitly compare them to the mentioned studies).
Obviously, the posed inverse problem is not fully solved; also, other benchmarks between different complimentary approaches should be organized soon.

*

As the authors raised in section 1, there are many methods to perform river bathymetry by assimilating satellite altimetry observations into hydrodynamic models.

Correct but only if one of the other unknown key parameter is provided. More precisely, if one has a good key prior information such as one discharge value or one reference bathymetry value (at a single location is enough), then it has already been demonstrated that the SWOT inverse problem can be solved with a reasonable accuracy, see the references cited in our general introduction. For ungauged rivers, the algorithms have to be able to infer the discharge and the bathymetry. As a consequence, the actual inverse problem is to infer the pair (Q(x,t); b(x)) plus of course a corresponding effective friction parameter K (constant or not). Based on hydrodynamics models, this is an ill-posed inverse problem (see Section 5.3). This is the inverse problem

addressed in the present manuscript, with, to our best knowledge, a capability to solve it never reached up to now.
Note that the present study focuses on the quality of the discharge estimation. Another article in preparation focuses on the quality of the bathymetry estimation obtained by the algorithm.

*

In my understanding, some of them simply applied the flavors of Kalman filter and successfully inferred river bathymetry (and river discharge) using the real satellite data from ENVISAT, ICESAT, and JASON-2 (e.g., Breda et al. 2019 https://doi.org/10.1029/2018WR024010).Âă

Thank you for mentioning this very recent reference. However, the inverse problem addressed in [Breda et al.] is not the same as the present one: the authors infer the bathymetry only (plus an effective friction coefficient), with the discharge given. This inverse problem is mathematically much much easier. Moreover, this is not the encountered inverse problem in ungauged rivers cases. Indeed, see their supporting information TextS4, “the model was forced using in situ observations of discharge at GS 15400000 (upstream boundary condition) and water levels at GS 15940000 (downstream boundary condition)”. The discharge value was imposed at upstream at daily frequency.

Actually, their inverse problem can be solved by others approaches-algorithms too; including the present algorithm, see [Larnier et al 2020]. (Note that this would be interesting to compare the available different methods to solve this inverse problem in this particular case). Note that [Breda et al.] is original not for its classical Kalman Filter approach but for its global optimisation approach based on a genetic algorithm (the SCE-UA algorithm) and the use of multi-satellite like data (which are synthetic too).

*

The authors' method seems to be much more complicated than these previous works and I am not convinced that the complex processes are necessary.Âă

The present method (partly) solves an inverse problem unsolved up to now (considering ungauged river without accurate prior information). The use of the final algorithm implemented into the open-source software DassFlow is not more complex than the use of standard computational hydro-informatics codes. The VDA approach may be qualified as complex in the hydrology community but it is a standard inversion approach in others geosciences community like oceanography for example (including in the SWOT community).
Also, note that a dedicated toolchain based on HiVDI algorithm has been implemented (and validated) to automatically produced discharge estimations from standard datasets used in the SWOT community (datasets from the Pepsi challenges, or as those produced from AirSWOT data or from the SWOT Science Simulator). In other respect, as explained in Section 7, once a one year data assimilation has been processed (ie. after one year of the instrument acquisition), an extremely simple algebraic model can provide in CPU-real-time the discharge estimation.
The critical scientific challenge is to learn ungauged observed rivers (without accurate prior information); this is a complex inverse problem; it seems to require sophisticated mathematical and numerical tools. At the end, our resulting operational system (the calibrated algebraic flow model, Section 7) is very simple.

*

I strongly recommend the authors to perform many sensitivity analyses and to confirm the impact of each process on the performance of their method.

Sensitivity analyses have been performed at each stage of the study and for each

stage of the global algorithm: ANN, the algebraic flow model, VDA-based on the St-Venant equations. Sensitivity analyses (equivalently, inversions robustness) have been previously addressed both for the algebraic flow model and the VDA approach, see [Brisset et al. 2018, Tuozzolo et al. 2019, Larnier et al. 2020, Frasson et al. 2020]. Sensitivity analyses on the ANN step only had been performed but were not shown. As expected, the resulting accuracy is slightly degraded but the robustness remains. Following your comment, a remark on this point (estimations with Gaussian noise with the expected instrument accuracy) have been added in the new subsection 3.4 entitled "On the sensitivity of the estimations with respect to error measurements and data frequency", see also a remark at the end of Section 6.
A thorough analysis of the complete inversion algorithm sensitivity in a context of real like (provided by a simulator or from the AirSWOT campaing aforementioned) should be done during the next benchmarking study. Our past experiments (including those based on the aforementioned real datasets) plus the present ones have convinced us that the presented scientific approach is solid and partly answers to an open problem unsolved up to now.

Following your remarks, we have added numerous clarifications in the new version. This should make the manuscript clearer, in particular by better highlighting the context and the academic feature of the numerical experiments. Moreover, short analyses for noisy measurements have been added, see in particular the end of sections 3 and 6. Finally, recall that finely analyzing an inverse method on perfect data is a mandatory step. The results, obtained for numerous and heterogeneous rivers, show an important improvement of the discharge estimations compared to the previous studies.

We sincerely thank you for your comments which have greatly help to clarify the hydrology problem, the approach capabilities and the limitations of the study, therefore the necessary forthcoming studies to assess estimations for ungauged rivers from eg.

the SWOT simulator.

*

Specific comments:
Major points:

L113: section 2.1.3. should not be “In-situ data”. The authors actually generated synthetic in-situ data by simulation. This is misleading.

Data are synthetic, that is correct. ÂăPlease, see the previous discussion for details.
To be more accurate, we have replaced the term “in-situ data” by “in-situ type data” throughout the text (including in the subsection name of course). Moreover, as already mentioned above, data origins and features have been recalled more clearly (including in the abstract).

*

L118: I believe that daily sampling data cannot be called “SWOT like” observations although it may be accepted in the previous papers.

This point has been better highlighted throughout the text, including in the abstract.
We explicitly refer now to the Cal-Val phase of the instrument and to Phase 1 of the so-called “Pepsi challenge” defined in [Durand et al. 2016, Frasson et al., 2020]. Please, see the previous discussion.

*

L119: As mentioned above, the assumption of perfect observation is problematic.
Please, see the previous related discussion too.

*

L142-145: Why did you calculate Pearson correlation coefficient? The authors did not use this information in this paper.Âă

We compute the R2 correlation criteria because it is a good (and classical) performance criteria to measure the efficiency of an ANN prediction vs the true values. This feature fully applies to the experiments presented in Tab. 1.

*

L148-149: I could not understand why the authors excluded the data whose mean discharge is larger than 10 000 m3. Since machine learning basically interpolates the data, it is generally recommended to make training data cover the wide range of state space. If they cannot have the access to those data, maybe they should not use fully data-driven approaches. Why should the authors choose the inappropriate experiment design?

This is not an "inappropriate experiment".
You are right, as previously mentioned a well trained ANN can accurately represent multi-scale, highly non-linear observed phenomena, in a least-square sense. You are right, a trained ANN constitutes an excellent interpolator but a-priori not an extrapolator (ie. out of the learning range values). Here, one expect that its prediction capabilities hold within the learning partition Q-Lset only.
In our case, the preliminary statistical analysis show that the great majority of “examples” (in the sense of Machine Learning i.e. dataset at one location) presents a mean discharge value lower than 10 000m3/s. The few rivers presenting mean discharge values greater than 10 000m3 are somehow outliers; they represent less than 10% of the examples.

Then we have designed the experiments to show:
1) the capability of an ANN to (roughly) estimate a discharge from the (3+1) input variables only.
2) the (un-)capability of estimations for larger rivers for which one could not acquire data enough to train the ANN.
This is our experiment plan and goals. The obtained conclusions seem clear and robust.

Basically, given a (unmonitored) river presenting the same characteristics as the learning partition, one can expect a more accurate ANN estimation. (Here, the partition Q-Lset contains the rivers with mean discharge lower than 10 000 m3/s). This intuitive feature is verified here, in the present case. We have tested (results not shown here): one obtain a better ANN model (or at worse a similar) for rivers belonging to the training partition, than if training for the whole values range.

However, if one train the ANN from the complete range of discharge values like you suggest it (ie. without excluding rivers with mean discharge values greater than 10 000 m3/s, namely Jamina, Mississipi downstream and Padma), then the ANN predictions for these three rivers are much more accurate. But, recall, that in that case, the same ANN is less accurate or (at best similar) for the other rivers ie. those belonging to Q-Vset-in. This is what show (confirm) the present numerical experiments.

In practice and if one approximatively knowns at what class a river belongs to (this assumption is reasonable for a great majority of rivers in the world eg. from the GRADES database), and if one has data enough to perform a good training process, then it seems to be more efficient to train the ANN with “examples” (datasets) from the corresponding class of rivers. This is what we suggest.

Note that detailing all these numerical experiments and the resulting properties cannot be done in the same article. We believe that the present manuscript demonstrates already a lot of new estimations capabilities and intrinsic properties of the different elaborated algorithms (the ANN, the algebraic flow model and the advanced VDA process).

However, again following your remark, the experiment plan and its goal are now much more detailed, see the new dedicated section 2.4 entitled “On the choice to define two river classes”.

*

L217, Table 2: I recommend the authors to use same metrics for Tables 1 and 2.

As already mentioned, the Pierson correlation coefficient (R2) fully makes sense for the Table 1 experiment since based on the complete set. It is more questionable for small datasets like it is the case for a single river only. That is why it has not been indicated neither in Table 2 nor in Table 3. On the contrary computing the nRMSE or the NSE for the whole dataset is meaningless. These criteria make sense for each river. For Table 1 experiment, the NSE represents a mean value only (like the nRMSE), that is why we initially choose to not indicate it. However, following your remark this criterion is indicated in Table 1 now.

*

L440, Figure 10: How did the authors get the target of bathymetry (red dots)?

The target bathymetry values are those employed in the various calibrated reference flow models (HEC-Ras, LisFlood etc) which have been performed to obtain the synthetic data available in the Pepsi 1 and Pepsi 2 datasets, see [Durand et al. 2016, Frasson et al. 2020] and references therein. Here, the red dots are computed from the effective rectangular values of the unobserved lowest cross-section A0 (W=W0, H0=Z0-b). As indicated in Section 4.3.1, these "true" (= reference model values) are available at the Reference Data Scale only. This point is better detailed now, see Section 4.3.1 and the end of the introduction of Section 6.

*

Minor points:
L23: Maybe the authors can divide this paragraph around this line. This first paragraph is too long and includes several topics.Âă
Corrected.
L61: the estimations accuracy –> the estimation's accuracy
Âă Corrected.
L417: Please fix a typo ("Section ??").
Corrected.

Thank you for your detailed proof reading.