Review of 2024-3480

The manuscript proposed by Beraud et al. presents a new methodology to filter outliers and estimate glacier surface elevation from dense digital elevation model time series.
Their study relies on the same data as the Hugonnet et al (2021) study.
Instead of Gaussian Process Regression, Beraud et al implement an additional LOWESS filtering step, before feeding the fitlered time series to a localized b-splines scheme in order to interpolate surface elevation measurements.

While the work and the methods shows promise, I strongly recommend that the authors take the necessary steps to strengthen the manuscript.
The paper appears to be in an early stage of development and would benefit from significant revisions before it is considered for publication: sections 4 and 5 are particularly challenging to understand. With the necessary improvements, I am confident the paper has the potential to make a valuable contribution to the field.

I recognize, from first hand experience, the amount of effort that went into this work. This is why I want to restate my strong support for the manuscript as I think it brings an important contribution to the field. I strongly encourage the authors to address the points I have highlighted, as doing so will enhance the clarity, rigor, and overall impact of the work.

Greg Guillet, University of Oslo

General Comments

- From this version of the manuscript it is not clear at all to me how uncertainties are calculated, nor what do they actually represent. More details on this throughout the manuscript are severely needed.
- Section 4 needs a vocabulary overhaul.
  The authors use very unspecific language which makes it harder to grasp what is the point they are trying to make. I have addressed some of these in my specific comments.
- As it stands, the writing in Section 5 makes it very hard to understand. I had to re-read some sentences multiple time to make sure understood the statements correctly. I think significant efforts are needed to
- There are some problematic statements which show some confusion between GPs and splines. I think these can and should be addressed with minimal changes by removing the statements I have highlighted in my specific comments.
  In addition, the authors need to reframe comparisons between methods in the manuscript as such : comparisons between the method proposed by the authors and the GP from Hugonnet et al. (2021).
- A lot of the sentences start with unspecific pronouns such as "this" or "it" - which required me to backtrack a few times. This severely hampers the readability of the manuscript.

- Acronyms are not always defined on their first use - please check this.
- Finally, there are quite a lot of frenchisms and typically french sentence constructions - it's not a major problem, but the manuscript would gain in readability if the authors addressed this.

Specific comments

Abstract

> We compare the produced dataset to previous studies [...]

If I am not mistaken, you only compared it to the Hugonnet et al. (2021) results.
I would hence keep it singular here.

Introduction

I generally agree with all the statements made in the introduction and they are accurate.
However, as it stands, I find the structure of this introductory section quite confusing.
The authors often switch between very broad statements to nichely precise descriptions of methods/problems without clear guidance of the reader into why this matter, which ultimately leads to significant repetition.
This is especially true between lines 30 and 50, which is followed by a lengthy statement on glacier surface velocity inversion methods which seems a bit out of place - I would either make the point clearer as to why this is important here or remove it.

Within these lines, I suggest a little restructuring :

- What are surges and why are they important (essentially your lines 16 to 23)
- Why are elevation changes particularly important when studying surges (which is basically your lines 35 to 34) but I would focus on 1) visual interpretation (i.e. mostly inventory efforts) and 2) quantitative interpretation - i.e. computation of shear stress, and even more, ice thickness/bed geometry inversions. I would remove the mentions to velocity products
- Problems with the data and then the problems with the methods used up to know and how you're solving them.

> [...] are not yet fully understood and this subject continues to be the object of developments and theories (Benn
> et al., 2023; Terleth et al., 2021; Thøgersen et al., 2024; Crompton et al., 2018)

I suggest to keep citations in chronological order, and to use e.g. in front of the citation array here, as these are but a subset of very relevant publications addressing the topic.

> L37: benefits

Benefit.

> L39: The use of suitable digital elevation models (DEMs)

You already defined used DEM on line 38, define it there.

> L40: suitable

I am not sure I grasp what this means in this context, is it DEMs that have high enough accuracy ? Please clarify.

> L49: They need techniques [...]

In general, I suggest avoiding sentences starting with "they" and would try to rephrase.

Paragraph starting at line 52

May I suggest adding my own work - Guillet and Bolch, 2023 - where we develop a Bayesian outlier filtering and uncertainty quantification framework to compute thickness changes from DEMs, specifically for surge-type contexts.
While we do not specifically tackle dense time series, the methods would be the same for denser data arrays and akin to data assimilation.
This is barely a suggestion, and I leave the decision to add this reference to the authors.

- Guillet, Gregoire, and Tobias Bolch. "Bayesian estimation of glacier surface elevation changes from DEMs." Frontiers in Earth Science 11 (2023): 1076732.

> L65: "To accurately estimate the parameters"

What does "parameters" refer to here ? I assume this is thickness change/dynamical thickening but be precise.

Paragraph starting at line 68

I would suggest to bolster this paragraph and be a bit less succinct - you have done a lot of work here and this is a good place to showcase a short summary.

> We algorithms from the literature to filter outliers

This reads awkward. I suggest changing for "We use established algorithms [...]"

Section 2

Paragraph starting at line 80

There are several sentences starting with "they" in a rapid succession, which hampers the readability of the whole paragraph. Please rephrase.

> L92 : The temporal sampling is heterogeneous in time and space.

Rephrase.

> L92-95

It took me 2 reads of these sentences to understand what was clearly meant. First, "below 50 days apart" should be "less than 50 days apart", but a more general question is whether both sentences are needed. I feel they give redundant information, that is illustrated in Figure 2.
I would rephrase to avoid this repetition and clarify the statement.

> L 96-97: (European Space Agency and Airbus, 2022)

Reference should be after the mention to the DEM.

Section 3

Section 3.1

> 1. Spatial filter: we filter out pixels with a difference of more than 400 m between the ASTER DEM and GLO-90 reference
> DEM.
> 2. Merging of strips: we merge the DEM strips on the same day by keeping, at the pixel level, the elevation with the highest
> correlation score at overlaps.

1. Is this 400m threshold arbitrary ? It's not a problem if it is - I am wondering if there is a reference to back this up or if its from the author's personal experience with the data.
   Also, the notion of "spatial" filter is a bit confusing as you are filtering pixels that show and absolute difference in elevation (z coordinate) between the ASTER and GLO-90 DEMs, correct ? Or are you operating in the x, y plane ?
2. This will probably reveal my total lack of knowledge on DEM generation methods but, is the correlation score a sufficient enough metric to discard data ?
   Or is there value in computing a median DEM from all the available DEMs for this day and using this instead ?

> L111-117

This could be condensed a bit.

Section 3.2

> L125: For our dataset, the output of the regression is to sensitive [...]

too sensitive

> L 125-126 : For our dataset, the output of the regression is to sensitive to noise overall and too smooth over surges to be used directly as an interpolation of the elevation, so we use it for

> filtering only.

This is a pretty important point which I feel is a bit brushed over by the authors here, as it is pretty well documented that LOWESS struggles in non-stationary contexts.

Section 3.3

> L147: "thereby improving the state of the art in this domain"

I suggest to remove this. This is very vague and does not give any useful information.

> L165-170.

I have to say I strongly disagree with this statement and I think there is a bit of confusion that needs to be addressed.
While GPs do require a prior on the functional form to fit a given dataset, the shape of the kernel function can actually be interpreted as physical variations in the latent variable, i.e. the shape of the Linear + Periodic kernel used in Hugonnet et al (2021) reflects the sub-linear and periodic signal one could expect from surface elevation changes.
Framed as a causal inference problem, a GP should not be used with prior belief on the structure of the data, but on the physical process the data represents.

ALPS also puts a "prior" on the functional form to fit, as it relies on a combination of b-splines to adequately approximate local regions of the data.
Even if you were interpolating with linear functions, you would be making an assumption that there is no oscillatory behavior between data points - which, by definition is a prior assumption.
The main difference here is thus that the "prior-like" assumptions are expressed through, for example, the smoothness of the derivative (splines) instead of a covariance function (GP).
In a more general way, any model requires prior assumptions.

As a side note, I want to point out that standard smoothing splines represent a special case of GPs, as shown by Kimeldorf and Wahba (1970).
In a highly-abstract way, you are, implicitly, fitting different GP models to different regions of the dataset - one should not stretch this analogy too far, as it breaks down when considering uncertainty estimates, as GPs are probabilistic.

All in all, I would argue against the generalization of this statement to every GP model, and only focus on the one defined by Hugonnet et al. (2021), otherwise this is not an apple-to-apple comparison.
You are comparing the results of two different methodologies and how they fare at interpolating time series of surface elevation data in the presence of transient physical events, no more, no less.
In addition, and I think this point is an important plus for the author's approach : I would imagine ALPS is more scalable than any GP model over larger sample and dimensions datasets, as GP are knowingly computationally expensive.
This is a strong pro for your method and I think you should address a bit more, as surface elevation time series are likely to get denser in the future.

- Kimeldorf, George S., and Grace Wahba. "A correspondence between Bayesian estimation on stochastic processes and smoothing by splines." The Annals of Mathematical Statistics 41.2 (1970): 495-502.

Section 4

Section 4.1

> L188-189: Modifications of this kernel to allow
> for stronger changes in elevation have not proven to be efficient enough

I would be a bit more careful and specific here, as I know of successful attempts at this with different GP kernels.
I would write something of the like : "Reparametrization of the kernel used by Hugonnet et al (2021) [...]".
It seems nitpicky, but I think it makes your point stronger.
Also be careful when using "efficient" - I get that you mean that the interpolation result is not satisfactory or rather than the median is too far from what you would expect, but efficiency is a very vague term.

> L 190-192: It does conserve nearly all known surge events in our study area and period, with one exception being surge events with strong melt before and after the surge

When you say study area and period, you mean the 4 surges you look at, correct ?
In which case, I would rephrase this as I think in its current form, it is not really reading well : "nearly" is a rather imprecise formulation that could be interpreted as trying to hide the cases where it does not work so well.
Just say 3 out of 4. Then, "events" is singular.

> L193: smooths

Smoothes

> L197-203

This whole section is confusing.
I would remove "unfavorable terrain" and be a bit more specific - directly mention textureless and steep regions.
"unrealistic erratic" is also confusing, I get what you try to mean here, I would use something like "Unstable" or "Oversensitive to Noise".
"The filtered-out areas (data gaps) are more prevalent with our method, mostly over unfavourable terrain": more prevalent with your methods than what ? The GP from Hugonnet et al (2021) I assume ? Be more specific here.
What I get from this section is that your method works better than Hugonnet et al (2021) on-glacier, in parts with that are relatively smooth, but tend to over-filter in areas of low contrast/rough terrain - am I correct ? In any case please make the section easier to understand.

> L205

Again, this is confusing. Please rephrase.

Section 4.2

> L209-212;

An interesting point is that the GP used in Hugonnet et al (2021) shows an increasing trend in surface elevation, completely omitting the actual data.
GPs tend to "fall-back" to the median when there is no data but here, both the median and the uncertainty increase.
Can you plot the uncertainty of each measurement ?

> L211: undulations

Replace with "periodic component"

> L215: "[...] creating wavelet artifacts"

The term wavelet design something different in signal processing and I would refrain from using it here. I would use "spurious high frequency oscillations" or something similar.

> L216: "removes completely the surge
> signal"

filters the surge signal out

> L218: have weakest changes

Weaker. But please consider changing to a more specific term.

> L220-221: Some glaciers are more affected by data gaps than
> others, in agreement with areas with a low number of observations (Fig. 1, e.g. Shisper
> glacier)

Shisper is not highlighted on Figure 1 and is not in the studied glaciers.

Section 4.3

> L227: with a decreasing speed (2009-2012, a1),

This figure does not show the velocity.
Also, the reference to the figure is broken.

> L235: wavelets

Same as before.

> L241: to the end of the surge, it then extends 5-10 more

to its termination, the surge propagates

> L246:

Add uncertainty estimates - at least some part of the discrepancy is in there.
I imagine there is an underestimation in the surface area of the reservoir zone ? Did you account that the surge also simultaneously drains the northern (Yutmaru?) tributary (centerline RGI2000-v7.0-L-14-27499) ?

> L253 :The "build-up front" or kinematic wave

To stay consistent with current terminology I would use surge front, not build-up front or kinematic wave.

> L254-255: representing a regular advance of about 460 m per year, which is approx-
> imately 6 times faster than the surface velocity, according to the NASA MEaSUREs ITS_LIVE
> project repository

I am not sure I get what you mean here - the surface velocity data at Khurdopin clearly shows seasonal behavior with velocities reaching around 400-450 $\mathrm{m} * \mathrm{yr}^{-1}$, starting in 2013 with a quasi-linear increase in velocity up to 2017.
Although I might be wrong, i would expect you to be able to see that the surge front advance rate is slower between 2000-2012 than 2012-2017 when the glacier slowly starts to shift to a velocity weakening regime.

> L275-76: The buildup and emptying of the first surge seems weaker than the second one, and extends less up-glacier of the junction, compared to the second surge

Again, refrain from using weaker as it gives the false idea that the surge did not dissipate as much energy - something we have no idea on.
The peak velocities of both of Yazghil surges are actually pretty similar and both are visible up to the glacier front in the surface velocity record.

> L277-278: This may be related to the effect of the tributary surge, that stopped at the junction
> but could have yet increased mass input by a blocking effect.

I really don't get what you mean by that, please explain.

Section 5

Section 5.1

This whole section is very confusing.
I do not understand what the first sentence is supposed to mean, how can an uncertainty estimate over a quantity reflect the filtering capabilities of a filter ?
What does it mean that the surge of Khurdopin shows that "that a discrepancy of a hundred meters is credible on exceptional events." ?

In addition, to further test the outlier filtering side of your methodology, you could generate false erroneous measurements and further quantify how well your method performs at filtering simulated outliers.

> L281: keep true elevations

Be careful with the use of "true".
All measurements are imperfect representations of the "true" elevation, which is by definition, unattainable.

> L309

Add a full stop before "To test"

Section 5.2

> L332 : appears

Appear

> L337: pre-surge thickening front or kinematic wave

Same as before. The propagation of a thickening front is one of the definitions of a surge. In your case it is still the early stage and has not reached the dramatic proportions it will eventually attain.
Stop using kinematic wave.

> L341: There is

There are

> L343: a bit later than our spring 2016 estimate

Unspecific. If it's spring versus October, then it's around 4/6 months later, just say it.

> L360-370:

Be a bit more specific. Please add uncertainty estimates. Mention that differences you show are between median values

> L376: one of the shortest surge cycles in HMA.

Is this from Bhambri et al. (2017) ? Make sure to add proper reference

> L378: Our data suggest it started 1-2 years later, implying a longer quiescence phase of 11-13 years

Do not make this a general statement on the dynamics of Yazghil glacier - ≈ 8 years of quiescence is not different from ≈ 11 when the number of considered events is 2.

> L385: The order of magnitude of the imbalances corresponds to the order of magnitude of the measurement uncertainty

Can't agree more ! Please add them.

> L383: bulge front

Just use bulge. Also, as mentioned before, it's not pre-surge.

Section 5.4

Typos

> L421: (exponential sine-squared (ESS) kernel)

I doubt being so specific is really needed here.
Just remove it. Also, components of a kernel are traditionally called "terms".

Conclusions

> L449-451

Succession of sentences starting with "it"

Figures

In general, all captions need to be reworked to describe all individual panels of the figures. I provide more specific comments hereafter.

Figure 1

Readers familiar with surges and HMA will know where the glaciers you mention are, I am not sure this is the case for the broader audience - maybe you could zoom in on a bounding box around the selected glaciers.
I am not sure the whole Kararokam region needs to be displayed since you focus on specific glaciers.

Figure 3

Caption : I cannot find any mention to "TS" in the plot - remove this from the caption since you write time series.

It would be beneficial to know which glacier centroid/vertex this surface elevation time series is sampled from.

This is a bit of a nitpick here but I would refrain from using two similar colors for the lines in "Interpolation it. 1" and "Interpolation it. 2" - being colorblind, I can't see the difference between them.

In addition, it would be beneficial if you showed the uncertainty associated to each measurement on the plots to the right.

Finally, I see no mention to any Student-T distribution in the paper (because the methods you rely on make no explicit assumption on the distribution of the data). Rename the "t-interval" into "Confidence Interval".

Figure 6

It's really hard for me to see the individual points between raw elevation and filtered measurements.
I think the symbols and the figure in general are just too small.
Again, just remove t-confidence interval and use confidence interval. i think it's too specific for most readers - if they want to know more, they will read Shekhar et al. (2021).

Figure 7

Increase the size of the figure and individual panels.

Figure 8

Hovmöller diagrams are a widely used plot in the community.
Just use "Interpolated surface elevation time series along the centerline of..." or something of the sort.

Figure 9

Again, just specify which glaciers these time series are sampled from - and where on the glacier.

Figure 12

Please add the red A, B and C regions in the captions. It's a shame to have to go into the text to grasp what the figure shows.

Tables

Please add uncertainty estimates in all tables - as a hunch feeling I would typically assume that this is what drive the reported imbalance (except for Hispar).

Table 2

This table is pretty confusing.
I would suggest replacing Table 2 with a figure showing the distributions for each glacier.
This would avoid having 2 columns as the 90th percentile and show the full distribution.