

Response to Reviewer 3 Anne Springer

Dear Authors,

I enjoyed reading your manuscript and appreciate this thorough description of the G3P product, which is a timely add-on to the already published data set. I share the impression of the two previous reviewers from the first round that the manuscript is well structured, thoughtfully composed, and easily accessible. I read the whole manuscript and, in particular, also the responses to the two previous reviewers.

I will first provide a general comment on the manuscript, which I think could further improve the usefulness of the manuscript for users of the product without requiring a lot of effort from your side. I will then address the responses to the two other reviewers and finally provide some additional comments on issues that caught my eye. These are very minor issues.

Dear reviewer, dear Anne,
many thanks for your careful and helpful review. We revised the manuscript accordingly, as detailed below.

General remark:

The main aim of the manuscript is to explain which data the G3P product is based on and to inform users about the limitations of the product. I expect the G3P product will become (and I think it is already on its way to becoming) a standard product for the assessment of global groundwater changes. From this perspective, I think it is essential for users to understand, in particular, also the limitations of the product, which will be of great value for any regional studies. I think this is well developed in the discussion section. However, while you can see at first glance where to find information about the origin and contribution of the individual storages, the limitations are less prominently represented. I think users would benefit if this aspect were made more obvious in the manuscript. This does not make the value of the data set smaller, but rather makes it easier to understand what can and cannot be discovered, for example, in regional studies. I think you could emphasize the limitations a little bit more without changing a lot about the manuscript.

Yes, we agree, thanks for this valuable comment. We highlight the parts of the paper that deal with uncertainties in the revised version:

My suggestion would be (and feel free to choose whichever of these options seems feasible to you):

- Consider renaming the discussion section, e.g. **Discussion of limitations** or **Discussion and limitations of the G3P product**.

We rename the title of Chapter 4 into **Discussion and limitations of the G3P product**

- In the discussion section, you could make the limitations of the individual compartments more prominent by adding paragraph titles. However, this may also disturb the flow of the text. As a reader, I would appreciate it, though. Alternatively, you could think about having the general discussion first and then a subsection **Limitations of the G3P product**.

Following this suggestion, we added sub-chapters headings to Chapter 4, with each sub-chapter devoted to the limitations of one of the storage compartment data streams that go into G3P.

- Maybe add in line 34 of the abstract something like:
Importantly, we also discuss several limitations of the data set arising mainly from the compartmental storage data sets.

We extend the sentence starting in line 31 of the abstract by alluding to the presentation of uncertainties and limitations of the individual data streams.

- Throughout the manuscript, you could add some links to the discussion section dealing with the limitations. As many people only read parts of a manuscript, I think this would be helpful. When I was reading the manuscript from the beginning, I was frequently wondering why certain limitations were not mentioned and only discovered the detailed discussion of them at the end. E.g. you could add:

- in line 172:
Importantly, the limitations of the final product are discussed in detail in Section XY.
(Maybe also briefly say why this is relevant?)

In the revised version of the manuscript, we add the following sentence in this paragraph: “The uncertainty quantification for each data stream is described in the respective sub-chapters of Chapter 2, and the limitations of each data set and possible consequences for the quality of the final G3P product are discussed in Chapter 4.”

- In line 635, you could think about starting the results section with something like:
Here, we provide an overview of the general characteristics of the product, including amplitudes, trends, and selected case studies. This is not intended as a complete validation, and when interpreting the figures, the limitations of the product discussed in Section XY should be taken into account.

We add a similar sentence as the one suggested by the reviewer at the beginning of the results chapter in line 637.

Responses to previous reviewers:

Reviewer 1: All points are well addressed in the updated version of the manuscript.

Reviewer 2: All points are well addressed in the updated version of the manuscript.

Regarding comment 4 of this reviewer, the authors could explicitly state somewhere below line 810 what the leakage effect means when interpreting the G3P groundwater data in coastal areas.

While there is already an explanation in this paragraph about what leakage means for mass change data including signal loss to oceans, we explicitly add GWSA to this sentence to make clear that this also applies to the final G3P data.

Other minor remarks:

Section 1:

- I. 123: For me, it was not clear if “filtering of GRACE/-FO satellite data” refers to Level 1 processing, to the monthly fields, or to both. If you mean only the monthly fields, you could maybe replace “satellite data” by “monthly gravity fields”.

Modified to “monthly gravity fields”.

Section 2:

- **Figure 1:** This is a nice figure. You could use the term “Subtraction process” instead of “Subtraction approach” to be consistent with the text. I find the box with “Uncertainty propagation” a bit misleading. This does not include uncertainties from GRACE/-FO, and only uncertainties from the filtered data are propagated. Shouldn’t this start already on the left?

Thanks for these observations. We changed the term to “Subtraction process” in the figure and added another box to make it obvious that also TWSA uncertainty is propagated to the final GWSA data set.

- **L. 334:** One other limitation that could be worth adding is the assumed soil depth of 2 m. I may have misunderstood this point, so please correct me if this is not the case, but I understood from the text that a soil depth of 2 m is assumed throughout the product. As one can see in Figure C1 and Table C2 in Müller-Schmied et al. (2021) (<https://gmd.copernicus.org/articles/14/1037/2021/#App1.Ch1.S3.T10>), for instance, the rooting depth assumed in WGHM is about 4 m in the Amazon region and 1 m over Europe. From my understanding, this would mean, for instance, that in the Amazon region, the temporal variability of GW storage could be highly overestimated. This could be one reason for the differences that you see in Fig. I in your answer to the first reviewer for this catchment. But I do not see the need to discuss any examples in the paper. However, this could also be a suggestion to further improve the product and decrease the uncertainty of the soil storage: to assume a variable soil depth.

Thank you for the valuable suggestion of using spatially variable soil depth that can be considered for improved versions of this product in future.

In the revised manuscript, we made it more clear in Chapter 2.2.2 where we describe the RZSM data that this data stream represents soil moisture for a globally uniform soil depth of 2 meters. At the beginning of Chapter 2.2.2 we already allude to the fact that soil depth / rooting depth varies globally, and in the Discussion and limitation Chapter 4.2, this limitation and potential impact on the resulting GWSA data is discussed.

- **L. 342/343:** I was wondering if the two snow data sets are consistent. If not, could this affect GW trend estimates?

Both datasets, GlobSnow v3.0 and Copernicus Land Monitoring Service (CLMS), are produced by FMI with the same underlying baseline retrieval algorithm, deriving SWE from satellite passive microwave radiometer measurements. The datasets have been validated against independent snow course measurements for SWE. With a one-year overlap period, they do

show a consistent continuation in SWE estimates for the Northern hemisphere and no significant change between the two SWE products themselves. Correlation coefficients as well as RMSE are consistent with time. Therefore, we argue that the two SWE datasets can consistently be combined and do not cause spurious trends or affect trends of the final groundwater product.

In the revised manuscript we extended a sentence on SWE evaluation by the statement that both data sets are consistent, with reference to Figure 3a which shows the evaluation results.

- **L. 411 and L. 413:** You mention “geodetic samples” and “geodetic surveys”. Would it be possible to specify what this refers to?

We add a short explanation to the revised manuscript indicating that the geodetic method for estimating glacier mass changes builds on mapping glacier surface and, thus, volume changes over time.

Section 3:

- **L813:** replace comma by full stop

Done.

- **L. 860:** There is also the surface water correction data set by Deggim et al. (2021). But maybe this idea is sufficiently covered by the references you have already included.

Yes, data sources that went into the Deggim et al. (2021) study are covered by the references that we give in the manuscript, and we focus on those that may provide operational SWS products while Deggim et al. (2021) is a singular product that has not been updated since then.

Section 4:

- **L. 808 ff.:** In the paper by Hohensinn et al. (2026), you discuss the impact of GIA on groundwater trends. As GIA could have a significant impact on groundwater trends at high latitudes (and this does not seem to be included in the error estimate), wouldn't it be worth mentioning this as a limitation here?

Yes, this is a good point. We added a sentence on the uncertainties that arise from GIA correction to the TWSA limitations part in Chapter 4: " While mass effects of GIA are corrected for in Level 2B processing of GRACE/-FO data based on one GIA model (see Chapter 2.2.1), there is substantial uncertainty associated with this correction due to the differences among GIA models (Eicker et al., 2024). This uncertainty is not explicitly propagated to the GWSA data in G3P and adds uncertainties on the long-term trend estimates of GWSA in GIA-affected regions such as Fennoscandia or the Hudson Bay area in North America (Hohensinn et al. 2026)."

Section 6:

- I expect that the product will also be used extensively for regional studies, at least for comparison purposes. At the regional level, users will often have more detailed information on individual storage compartments that are used as input data for the G3P product. Since the individual data sets are provided separately as part of the G3P product, users could combine these with more detailed regional data where available, and thereby also assess the contribution of individual compartments to the resulting groundwater estimates. I think this could provide very interesting feedback on the performance of the product at the regional scale. Maybe it would be worth mentioning this potential in the Conclusion or Discussion section?

Yes, thanks for the suggestion. We added a sentence to the conclusions chapter when talking about user evaluation: "In this context, users may also have more detailed information on individual storage compartments at the regional scale. These may be compared and eventually combined with the data streams that go into the G3P processing chain and may help to better assess the contribution and the uncertainties of individual storage compartments for the resulting GWSA estimates."