Reply to Reviewer Comments to the Manuscript

CLIMFILL v0.9: A Framework for Intelligently Gap-filling Earth Observations

Verena Bessenbacher, Sonia I. Seneviratne, Lukas Gudmundsson
Geophysical Model Development, `doi:10.5194/gmd-2021-164`

RC: ***Reviewer Comment***, AR: *Authors response*, ☐ Manuscript text

1. Letter to the editor

Dear Prof. Tomomichi Kato,

Please find enclosed the second revised version of the article entitled “CLIMFILL v0.9: A Framework for Intelligently Gap filling Earth Observations” (Paper GMD-2021-164). In the revised version, we have

- implemented the technical corrections provided by Anonymous Referee #1.
- revised the manuscript according the revisions suggested by Anonymous Referee #2 in all the cases where the comments of Anonymous Referee #2 were still applicable to the current version of the manuscript, as we have previously agreed via Email correspondence.
- updated the results and Figures taking into account the fact that we were able to run CLIMFILL on a larger time period whilst the revisions took place. One major drawback of the new methodology introduced in the revised version was that the computational expensiveness grew such that we had to restrict ourselves to gapfilling only a year (2003) of data. Now, since we had some time in between revisions, we were able to run CLIMFILL for the time period 2003-2020, and included the results. This also leads to adding a new Figure (Figure 12) that showcases the interannual variability, which was previously impossible since results were only available for one year.
- thoroughly revised the text to improve readability and comprehension. We replaced the RMSE with pearson correlation in Figure 11 and 12 to improve intuitive understanding of CLIMFILLs advantages and shortcomings.

These minor changes can be viewed in the accompanyed difference-document. We are confident that the above-mentioned revisions of the paper together with the updated supporting information have increased the value of the submitted manuscript. We are looking forward to your reply and the reply of the Referees, if applicable.

Yours Sincerely,

Verena Bessenbacher (on behalf of all co-authors)

2. Anonymous Referee #1

RC: ***The authors have substantially revised their manuscript and entirely revisited the methodology. I congratulate them for the thorough work and I believe the manuscript has improved and in now publishable.***

I only have a few minor comments that can be addressed without a further round of revisions:

AR: *We thank Anonymous Referee #1 for this positive feedback*

RC: ***Section 2.1, step 1: I agree an approximation might be needed for computational reasons, but it would be good to mention that a consequence is that uncertainty in the interpolation cannot be estimated.***

AR: *We agree that with this the multistep process in step 1 the uncertainty of the interpolation cannot be estimated. Firstly, thin-plate-spline interpolation does not offer uncertainty estimates, and secondly it is unclear how the uncertainties of the different iterations could be aggregated. Further research would be needed to answer this question. We have added the following sentence to this section:*

> The interpolation of the daily anomalies follows Das et al (2018), who suggest reducing complexity of kriging/Gaussian Process regression by repeated interpolations on random sub-samples of all available data points and averaging the resulting estimates. In particular, the missing values in the anomalies are estimated by randomly selecting 1000 observed points per month over which the interpolation is calculated. This is repeated five times and the mean of all interpolations for each missing point is taken as the gap-fill estimate. As a consequence of these adaptations, the interpolation step becomes computationally feasible, but the uncertainty of the interpolation cannot be estimated. Finally, monthly maps and anomalies are summed up to form the inital gap-fill estimate from step 1.

RC: ***l. 302: There is a small mistake here: In the JS divergence, zero means that both distributions are identical.***

AR: *Thanks for spotting this error. We have corrected the sentence, it reads now:*

> This measure compares the distance between two multivariate distributions, where a value of ~~one means that the two samples are from the same distribution~~zero means that both distributions are identical, and one indicates that the distributions are not overlapping.

RC: ***Caption of figure 11: Please clarify that the top panel indicates RMSE while the bottom panel indicates physical values.***

AR: *Figure 11 now shows pearson correlation and not RMSE. To clarify this and reply to this comment, We have changed the caption to:*

> ~~RMSE~~ Pearson correlation of regionally averaged mean seasonal cycles between CLIMFILL and the original ERA-5 data over IPCC reference regions (AR6 regions, Iturbide et al. 2020) (top panel maps)~~and regional averages~~. Regionally averaged mean seasonal cycle of the physical values over selected regions in original ERA-5 data, satellite-observed ERA-5 data and ~~gap-filled~~ gap filled CLIMFILL data (bottom panels). The selected regions ~~are in areas with the largest fractions of missing values globally or~~ show exemplary advantages and problems of the framework, see text. For all other AR6 regions see ~~Supplementary~~ Appendix Fig. A2.

RC: ***Reference to Allard (2013): This is a citation to the review of a book. Please check whether you prefer to refer the book itself.***

AR: *We have replaced the reference to Allard (2013) and Chilès and Delfiner (1999) with the reference to the most recent edition of the respective book:*

> In the geoscientific literature, among the most commonly used approaches for estimating unobserved points are spatial and temporal interpolation methods, including nearest neighbour regression as well as kriging and derivatives thereof (Liu et al., 2018; Cowtan and Way, 2014; Haylock et al., 2008; Cressie et al., 2006, for an overview see ~~Cressie and Wikle 2015; Allard et al. 2013; Chilès and Delfiner 1999)~~ Cressie and Wikle 2015; Chiles and Delfiner 2012)).

where the full reference is: Chiles, J.-P. and Delfiner, P.: Geostatistics: modeling spatial uncertainty, Wiley series in probability and statistics, Wiley, Hoboken, N.J, 2nd ed edn., 2012.

3. Anonymous Referee #3

RC: ***The authors proposed a method to fill missing values in gridded hydrometeorological data. The key idea is to estimate missing data for one variable (e.g. soil moisture) by using some other independent variables (e.g. precipitation, ground temperature). The authors applied their method and claimed that it performed better than simple interpolation.***

Geoscientists are always bothered by missing data. Efficient and reliable methods of filling in missing data are always eagerly expected. Although the method by authors could be potentially promising, I am not totally convinced so in the current form of manuscript due to a simple single reason. I guess (because I found only very short qualitative description for this in lines 474-482) that the performance of data reconstruction is quite condition-dependent (e.g. the parameters, quality of original data, period and location of validation, etc). I would like to see discussion how sensitive the performance is to such conditions. I totally understand this is a demanding request, but at least I wish to see at least a relevant and detailed discussion in the manuscript.

AR: *We thank Referee #3 for the feedback and the detailed comments below. We note, however, that the comments seem to refer to the originally submitted version of the manuscript, and not the resubmitted one which was under discussion. In discussion with the Topical Editor it was decided that we reply to the comments wherever they are still relevant to the current version of the manuscript.*

Specific comments

RC: ***Line 216 "soil moisture values at this point in a 3-month backward window (s=90 days) from the current date (l = 0days), corresponding to previous work indicating the soil moisture memory effect acts on "monthly to subseasonal time scales": If I understood correctly, the parameter 90 and 0 was more or less "subjectively" decided by the authors. Should the users of your method need to conduct sensitivity tests for these parameters? If yes, is it implementable (i.e. parameter combinations can be easily exploded). If not, what is the rationale?***

AR: *The chosen parameters were decided by taking into account previous work with embedded features (Ghiggi et al, 2019, Gudmundsson et al, 2015) and are intended to reflect the timescales in which soil moisture memory effects are predominantly taking place (see e. g. Nicolai-Shaw et al. 2016, full citation in paper). Ideally,* *any parameter that is set or needs to be set within the CLIMFILL framework would be determined using the Cross-Validation procedure described in Section 2.4. However, we discuss in Section 2.4 that some of the parameters are set such that computational expense still is within manageble limits. Similarly, we do currently not recommend sensitivity tests for the parameters mentioned here, since as already noted parameter combinations can easily explode and running many instances of CLIMFILL to test different parameter combinations is too expensive to be feasible.*

RC: ***Line 237 "the algorithm repeatedly iterates over the variables until convergence is reached": The term "convergence" needs definition. Convergence toward what?***

AR: *In the current version of the manuscript, we have defined convergence in line 232-234:*

> The algorithm is stopped (stopping criterion) once the change in the estimates for the missing values is small between iterations (convergence) or a maximum number of iterations is reached (early stopping).

RC: ***Lines 301-327 In this part, numerous parameters appear without showing concrete background (e.g 3 d running mean, 5-pixel side length, embedded features for (s=7,l=0), (s=23, l=7), (s=150, l=30), (S=7, l=-7)). First, how these variables were determined? Second, how sensitive are these settings to the results of gap filling? Third, the authors claim that these are from the domain knowledge. Then, if one intendedly set counterfactual parameters (parameters against the domain knowledge), will the performance deteriorate? I am asking the last question because sometimes machine learning is so powerful that any data can be "predicted" by problematic assumption or data.***

AR: *The mentioned parameters all relate to the first step of the framework (interpolation step), which has been replaced by a Gaussian Process - based gapfilling in the revised version of the manuscript, such that none of the mentioned parameters are used anymore.*

RC: ***Line 404 "multivariate property of land climate interactions that is currently underestimated in satellite data": What does it mean by underestimation in property? I found more detailed explanation in lines 426-428, but anyway it is a bit hard to read.***

AR: *This paragraph and Figure 12 which it relates to is removed in the revised version of the manuscript.*

RC: ***Line 411 Figure 11: If I understood correctly, the difference between the "Original ERA5 data" and "Satellite-observable ERA 5 data" is mainly sourced from the grid cells to include for areal mean calculation. First, is this correct? If this is the case, the discrepancy in lines make sense because these are looking at different grid cells in each region. Then I am wondering why terrestrial water storage (TWS) is always perfectly matches between two. Is this because TWS has wide satellite-observation coverage hence the two lines are virtually the same? Is this more or less the case for ground temperature except snow free period? If all above is the case, I would say that displaying the results for TWS and ground temperature is misleading and confusing because they must well overlap.***

AR: *This is correct: the difference between the two is that unobserved grid cells are missing in one, and not missing in the other. In Figure 11 in both versions of the mansucript, TWS does not always perfectly match between original and satellite observed, there are slight differences (see Figures 11 and 12 and Appendix Figures A1 and A2). These differences are small because only a small amount of values (15 %, see Figure 5) are missing for TWS.*

RC: ***Line 419 "precipitation and terrestrial water storage estimates show little change": I think precipitation shows change.***

AR: *This part is formulated differently in the new version:*

> The most difficult case is precipitation. Precipitation estimates are only slightly improved with CLIMFILL compared to initial interpolation. Precipitation is influenced by several processes that are not captured within the four selected variables. For example, frontal rain patterns are mostly not explained by land surface properties but are governed by large scale circulation. This is a challenging case and could still furthermore be improved, for example by adding wind patterns to capture more synoptic features.

RC: ***Figure A1: The figure caption says that "all combinations of variables", but the combination of soil moisture and ground temperature seems missing. Is there any reason for this?***

AR: *We agree that the caption is misleading. We have changed it to:*

> Improvement of multivariate distribution with CLIMFILL gap filling: 2D-histogram ~~of~~ for all other combinations of variables ~~for not satellite-observable values~~ (apart from the one already shown in Figure 7) for the original ERA-5 data, ~~interpolation~~ satellite-observable ERA-5 data, the Interpolation gapfill and CLIMFILL gapfill.