Hydrology and
Earth System
Sciences

Discussions

Interactive comment on "Linear Optimal Runoff Aggregate (LORA): A global gridded synthesis runoff product" *by* Sanaa Hobeichi et al.

L. Gudmundsson (Referee)

lukas.gudmundsson@env.ethz.ch

Received and published: 7 September 2018

I did enjoy reading the paper of Hobeichi and colleagues, which describes an interesting approach to compute a weighted ensemble mean of runoff fields, conditional on observed river flow. Overall the paper is well written. Nevertheless, I do have some comments and requests for clarification as summarized below:

** Overall comments and questions **

(1) Please refer early to Fig. 3 in the methods description.

(2) Maps of the performance of LORA would be very helpful.

(3) Some example time series, including the time-varying uncertainty would be inter-

esting.

(4) Does the fact that the weights are computed based on a discrete set of donor-catchments lead to discontinuities in the runoff fields? How would e.g. the field look like for one individual month?

(5) I really appreciate the authors effort to also include uncertainty estimates in their product. Unfortunately, I did not find any validation of this uncertainty estimate or a full interpretation of what it means. I know the uncertainty estimates are introduced elsewhere, but as this is a relatively new approach it requires extra care. For example:

(5.1) It is not clear to me whether this uncertainty estimate is related to the "confidence interval" (i.e. an estimate for the range of the "true" ensemble mean") or whether it is related to the "prediction interval" (i.e. an estimate of the range in which new observations would fall).

(5.2) if the uncertainty is an "prediction interval", an evaluation of the widths of the "uncertainty bounds" with respect to the distribution of the residuals would be helpful. Especially, compared to the spread of the input ensemble. For this, metrics from ensemble forecasting (e.g. the continuous ranked probability score (CRPS) or "reliability plots/histograms" might be helpful).

(5.3) the uncertainty bounds can produce negative runoff. Should this be the case or is this an artefact?

(6) I would appreciate some more information (figures, tables) on the actual "weights" and "biases"

(7) some of the references have artefacts, e.g. "nan" values instead of page numbers. . .

** Specific comments and questions **

Section 2.1: The authors might also be interested in the following global-scale data source: Do et al (2018, doi:10.5194/essd-10-765-2018) & Gudmundsson et al (2018,

doi: 10.5194/essd-10-787-2018). Sorry for this self-citation.

Section 2.2: What about other comprehensive model ensembles, such as ISIMIP2a (http://dx.doi.org/10.5880/PIK.2017.010)

page 4, line 25: Is there some empirical evidence supporting the assumption of constant bias ratio?

Generall for methods: I found the mixed use of $Ragg\_k\hat{\ }j$ and $q\_k\hat{\ }j$ confusing. Would it be possible to clarify the difference/commonality once and then only use one of them?

page 5, line 2: not really clear what you mean with this sentence. what is the number of records?

page 5, line 4: An alternative approach for dealing with negative values might be to log-transform runoff/stream flow before doing the computations and back transform it in the end. See e.g. Gudmundsson & Seneviratne (2015, doi: 10.5194/hess-19-2859-2015; 2016, doi: 10.5194/essd-8-279-2016). Again, my apologies for self-citation.

page 5, line 10: Some more details on the transformation process would be appreciated. I know it is published in Hobeichi et al. (2018) but it would make the paper easier to understand if it would be outlined in more detail here.

page 5, line 20: transfer of "weights" from the 3 most similar basins; why 3 and not a larger/smaller number? Any empirical motivation for this? Would optimizing this hyper-parameter help to get even better results?

page 6. line 8: Note that "Tundra" and "Subarctic" will not always have permanent ice! Also: The chosen climate zone map is very uncommon, and most readers will not be familiar with it. Therefore, it needs to be presented in a figure. Alternative: why not use a common climate-zone definition (e.g. Köppen-Geiger?)

page 7. l 15: should be fig 4 not fig 3

Fig 4: I found the "relative" improvement difficult to grasp upon first reading. Sugges-

tion: just show the performance for WPin WPout (and omit individual models, as they are shown later)

page 9, line 18:

(1) "mean runoff uncertainty". What was exactly calculated? Note that simply averaging sqrt(sigma) will yield wrong results. Instead the rules for propagation of uncertainty should be used (https://en.wikipedia.org/wiki/Propagation_of_uncertainty). If the authors did already consider this, I apologize for this comment.

(2) Note that there are strong perceptions on what "reliability" means in ensemble forecasting. See e.g. "reliability diagrams" and "reliability histograms"

page 9, line 31 & Fig 8: how did you compute the mean uncertainty? (refer to my comment above)

page 10, line 31: not Fig 3...