Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-415-RC3, 2019

Hydrology and Earth System Sciences
Discussions

Interactive comment on "A geostatistical framework for estimating flow indices by exploiting short records and long-term spatial averages – Application to annual and monthly runoff" *by* Thea Roksvåg et al.

Jon Olav Skøien (Referee)

jon.skoien@gmail.com

Received and published: 22 November 2019

This manuscript describes a methodology for spatial interpolation of time series, which can also take into account short records at the target location. The manuscript is well written and the methodology seems sound. I have some detailed comments below, which I think can all be answered within a minor revision.

Similarly to the other reviewers, and what the authors have written in their answers, I agree that it would be a good idea to replace the mean annual runoff map for Norway

with an assessment of the methods' performance for mean annual runoff.

The manuscript analyses a situation with only one observation for the partially gauged catchment. This is an extreme case, so extreme that an extra observation was introduced for the linear regression to be possible. Whereas this gave a convincing case study, I'd like to see at least some discussions around the effect of more observations also for the PG case. Also, the annual series are based on daily observations, so many observations are actually available for the regression. Could this give another result? Spatio-temporal kriging will most likely not do much better than the spatial interpolation, as it will use the model based covariance rather than the observed covariance also for the PG case, but I think it should be mentioned. It might be a good alternative for a time series with a few (maybe non-consecutive) missing observations.

P3L3 It is referred to how the model is developed for annual datasets, but might be used for indicators with other temporal supports (such as monthly). However, it is never explained why there is a difference between monthly or annual data, except for different correlation lengths etc. I guess this is related to the comments on P3L15, were it is referred to the water balance being "close to preserved . . . with some uncertainty". However, the "almost preserved" is never explained. The same description is used on P11L13. Could the method also be used for daily data? If not, why?

P25 – description of Figs 11-12. It is mentioned that UG has problems predicting large values of runoff, but I'd say it is just as difficult with small values. Additionally, the negative values should be mentioned here, not only in the conclusions. It seems there are no negative values for annual runoff for UG? I think this result is partly related to the fact that the data don't follow the assumption of being normally distributed, which should be discussed. A transformation method such as logtransform could avoid this problem, although log-transformed data on the other hand don't go so well with the linear aggregation assumption, see also Clark (1998).

The reference to Figure 3b comes before Figure 2. Figure 3a is only mentioned in the

reference to Figure 4. Figure 1b is not mentioned at all in Section 2, only in Section 3. Figure 1b is a bit difficult to understand, including the reference to 0-4 catchments. These types of figures are generally difficult to make nice, so I'm only asking the authors to test if other visualizations could work better. The lines of 4 catchments is difficult to see. Maybe it would be possible to add river network on top, to understand the catchment order?

Section 5.2. I think it could be mentioned that the areal model is substantially better than TK when sigma_c » sigma_x. Then later in the section, it says linear regression gives more unstable uncertainty estimates. Is unstable the correct word here?

P27L10 I do not think it is correct that observations from nearby catchments have a larger impact on the target catchment for the centroid model when range_c > range_a. It should be the opposite, large range means that stations further away will also get a weight in the interpolation. Then also the next sentence (L12-15) seems somewhat incorrect. If the values in Table 3 are divided by the similar values in Table 1, maybe the differences could easier be interpreted in light of the range?

The discussion at the end of P31 should also include some thoughts around gauging density. If the density is high, it is more likely that catchments are nested. Can the centroid model be expected to be as good as the areal model for non-nested observations? And when mentioning other environmental variables, it should maybe be stated that these are point values?

Edits

The use of commas should be checked. Several introductory clauses that need commas do not have them, whereas some commas are not needed. Examples:

P1L6 Another property, -> not needed

P4L18 In Western Norway, -> needed

P4L24 The processes explained above, creates -> not needed

P8L30 . . . stationary, Matern . . . -> not needed

Other edits:

P1L5 "The climatic GRF . . ."- I think this could be rephrased. If I understand correct, the GRF learns the spatial pattern from a limited number of years, and can use this information to improve predictions for years without observations.

P2L5 "including data . . . runoff datasets" – I think this part of the sentence can be deleted.

P2L32 Add "spatial" before "correlation"? Maybe the sentence could also be somewhat rephrased, at first reading it appears contradictory when referring to observation locations having both high correlation and high spatial variability.

P3L2 sparse datasetS

P3L10 "the climatic spatial field learns. . ." – I don't think a spatial field can learn, rephrase?

P4L7 "The study is carried out by utilizing" – a bit wordy, shorten?

P4L18 "The large values . . .Norway are mainly . . ." – could be shortened, maybe "This is mainly . . ."

P4L13 "This leaves" – rephrase?

P4L14 The sentence referring to Figures 1a and 1c doesn't really fit in, try to rephrase paragraph.

P4L30-31 The sentence doesn't read well, try to rephrase.

P4L33-34 "The monthly runoff data from Norway" -> "The data from these months"? Caption Fig 1: "FREQUENCY OF annual runoff observations"? And maybe ". . .Figure 1a, subcatchments are plotted on top, and this . . ."

P7L4 "and so on" is not a good expression in a scientific context. "In order" can be

removed.

P7L5-6 I don't think statistical analyses are only because of uncertainties?

P8L6 I'm not sure if it really clear what is meant by point prediction here.

P8L8 rephrase "things"

P9L4 Covariance or correlation?

P9L10 I think this sentence (known covariance structure with unknown parameters) is unclear.

P9L28 linear aggregation of what?

P10L7 The sentence is incomplete.

P10L15 It is not clear what is meant by "uncertainty based on . . ."

P10 Eq 4 I think the second equation needs units.

P10L30 "in Norway" – could maybe be generalized to something like "that models the average runoff over the study area (Norway)"? I think it is only some of the priors that are particular for Norway, the rest of the framework should be general.

P11L19 Isn't it rather that the alternative is for variables that do not require preservation of water balance?

P12L9 should it be 400 mM/year and 4000 mM/year (delete last s)

P12 Eq8 It is not clear where 0.025 comes from, and I also think the value should have a unit.

P16L2 computationalLY feasible

P16L3 remove "of" before matrix operations

P16L13 marginal distributionS?

P17L3 The default covariance function -> this model was also fitted?

P17L10 might influence or will influence?

P18L21 I like the idea of using CRPS for kriging predictions, but as this is (so far) rather uncommon, maybe also clarify here what the predictive cumulative distribution from kriging is in this context?

P20L6 is it the areal or the centroid model that gives larger uncertainty?

P21L19 I would say it's largest for June and for annual data. The sentence starting with "Particularly" is incomplete.

P31L28 I assume that the areal model might outperform the centroid model also for OTHER AREAL variables that are not mass-conserved?

References:

Clark, I.: Geostatistical estimation and the lognormal distribution. Geocongress. Pretoria, RSA., [online] Available from: http://kriging.com/publications/Geocongress1998.pdf, 1998.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-415, 2019.