Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2015-536-RC1, 2016

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Ordinary kriging as a tool to estimate historical daily streamflow records" *by* W. H. Farmer

E. Pebesma (Referee)

edzer.pebesma@uni-muenster.de

Received and published: 29 February 2016

article

Review of: *Ordinary kriging as a tool to estimate historical daily streamflow records*, by W. H. Farmer; reviewed by Edzer Pebesma.

Ordinary kriging is a well established method for spatial interpolation of geostatistical (field) variables. The current manuscript demonstrates its successful application to the interpolation of daily streamflow for ungauged catchments. I would be the last one to criticize the application of a simple, well-known technique to solve a clear problem in hydrology – hydrology has suffered enough of solutionism, papers defending new methods to solve problems where simpler alternatives would have been good enough. The current paper however leaves a number of questions open, related to understand-

ing why the proposed approach works well.

Physically, streamflow records are aggregates over larger regions: rain falls on the entire catchment, and much of it flows out in the stream. This implies that catchment size has an influence on the characteristics of the variable. Also, streamflow records may have been measured on the same river, introducing correlations because one gauge's catchment is contained by that of another. Ordinary kriging assumes intrinsic stationarity, and hence ignores catchment area and containment.

A recent variety of kriging, called top-kriging, has been developed (and is available as R package rtop, on CRAN) to accomodate variables with varying support, and was designed for this particular problem. In this paper, we see that ordinary kriging performs very similar (in terms of average statistics) to top-kriging. It would be good to better understand why the differences are so small: is it because we divide stream flow over catchment area? Is it because catchments don't contain each other? Is it because area is similar in most, or many cases? A graph of for instance the (temporal) variation of z for each gauge against the size of the catchment might reveal a lot.

The author interprets the results as ordinary kriging (with pooled variogram) being favourable over top-kriging. I would consider them identical, as a difference of 1% in R^2 or RMSE is in my opinion meaningless for operational purposes. What interested me (but is not mentioned in the main text) is that the 90-percentile values for top-kriging perform better with a slightly larger margin. Is there an explanation for that?

I find the reporting of both R^2 and RMSE a bit artificial, in particular the fact that they give rise to different conclusions (I would have concluded that performance is similar). In (Pebesma, E.J., P. Switzer, K. Loague, 2005. Error analysis for the evaluation of model performance: Rainfall-runoff event time series data. Hydrological Processes, 19, 1529-1548, http://dx.doi.org/10.1002/hyp.5587) we point out that R^2 is a scaled version of MSE, so if RMSE gives a different ranking of methods than R^2, this is all due to taking the square root. If then, based on that, pooled ordinary kriging turns out to be favoured, a good story why taking the square root of MSE is important would be no luxury!

The handling of zeroes: which fraction of the observation was zero? How sensitive were the results for the arbitrary number assigned to zero streamflows? (In my experience, when taking logs, this decision may pretty much blow away everything else).

Logarithms: is the goal to estimate z, i.e. not back-transform? If back-transformation was applied, how was this done? The results in Table 1 show results for non-log and log-transformed values, but the main text suggests that z, meaning only logarithms, are considered. What is the case?

Which software and software packages did you use to carry out this study? Can you also cite its authors? Since the data are open, can you provide a script that reproduces the findings?

The paper's main contribution seems how it handles time: instead of repeatedly computing and fitting variograms for each time step, a single (pooled) variogram model is fitted to the average of all variograms, and this seems to work better. As referenced in the paper, we (Gräler, Gerharz and Pebesma, 2011) found similar results when interpolating daily PM10 values over Europe. My impression there (and feeling here) is that the problem with daily fitted models is that occasionally the fit looks crazy, due to extreme values or strange conditions. This paper does not confirm nor deny this, as figure 6 only shows moving 31-day medians of variogram parameters. Can you also show (or describe) the time series of the raw daily values?

Given that figure 6 shows clear seasonal signals in the variogram parameters, an alternative to the current approach would be to use the 31-day median parameter values instead of the temporally constant pooled variogram model. This would be a compromise between the (too noisy?) daily fitted model and the (overly smooth?) constant model. Another question that might be discussed is the option to use spatio-temporal (ordinary, or top) kriging: right now, temporal correlation in streamflow records is ig-

nored. In case the prediction would concern incidentally missing values, the observation directly before or after the missing value might be much more informative than the spatial neighbouring values. It might be the case that missingness means longer periods of no observations, in which case temporal correlation will not help much. Explaining the pattern of missing-ness might help the reader understand why this study did not consider temporal correlation; currently such an explanation is missing.

In general, the manuscript confuses semivariance with covariance; I strongly suggest to use only one of the two.

Textual Details:

1. if Figure 1 would show the (main) rivers, or catchments, we could see the degree to which catchments are contained by each other.
2. 4:21 $i = [1, ..., n]$
3. 4:21 Euclidian location: omit Euclidian
4. 5:9 then should be that
5. 5:9 μ is the Lagrange multiplier, not an estimate of the mean of z
6. 5:10 "In practice, the elements of D cannot be calculated explicitly" what you try to say is that individual covariance values cannot be inferred from a single sample (realisation) of z; only with additional stationarity assumptions, covariance can be modelled as a function of separation distance. Rephrase?
7. 5:12 a variogram is not a model of the covariance
8. 5:21 "theoretical variogram model", replace with "variogram model type".
9. 5:22 "in building the empirical variogram, the covariances": again, variogram/semivariance and covariogram/covariance are two different measures.

10. 5:23 for a complete description, the maximum distance up to which semivariances were computed is needed too (as well as whether the ten groups are of equal distance interval width)

11. 6:6 same confusion: the variogram does not have covariance values.

12. 6:15 "stationarity" → "temporal stationarity"

13. 6:20 "computation of as many variograms" → "fitting of as many variogram models"

14. I feel that it should be pointed out somewhere that averaging variogram model parameters does not necessarily lead to the same model as fitting a model to averaged (pooled) sample variograms.

15. 8:17 "and" → "a"