Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2015-536-AC1, 2016

Hydrology and Earth System Sciences Discussions
Open Access
EGU

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

W.H. Farmer

wfarmer@usgs.gov

Received and published: 5 April 2016

Reviewer Comment 1: *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 understanding why the proposed approach works well.*

Author Response 1: Thank you for your deep consideration of this work. I greatly appreciate your encouragement and am confident that your comments will greatly im-

Printer-friendly version

Discussion paper

prove this manuscript. I hope that this work has interested you and motivated additional research.

Reviewer Comment 2: *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.

Author Response 2: The reviewer raises two important concerns: the impact of drainage area and the impact of nested basins. Drainage area is incorporated into the estimation by considering depths of streamflow (line 18, page 5). This standardization removes some of the aggregating effects. Nested basins are not explicitly addressed. However, top-kriging does consider this effect. The minimal difference between ordinary and top-kriging suggests that the impact of nested basins is small.

The attached figure shows temporal variation (standard deviation of the logarithm of daily unit runoff) against catchment size. There is some variation with drainage area, but the Pearson correlation is weak (0.05). It appears that taking the logarithms on the unit runoff effectively controls for the effects of drainage area. This may be the reason for minimal improvements provided by top-kriging. This discussion, though not

Printer-friendly version

Discussion paper

the figure, will be added to the manuscript.

Reviewer Comment 3: *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 R2 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?*

Author Response 3: I strongly agree that the methods of kriging, ordinary and top, produce operationally identical performance. I tried to express this fact by noting the inconclusive advantage on lines 11-17 and 25-26 of page 10 and lines 9-21 of page 11 and lines 27-30 of page 12. The advantage of ordinary kriging, as you note earlier, is that it is simpler than top-kriging. If the difference is nearly identical, then the additional complexity and computational load of top-kriging may not be advantageous. I will attempt to make this understanding clearer in the noted lines.

As for the differing distributions of performance, I cannot explain this without a deeper exploration of the weakness of kriging methods, an exploration beyond the scope of this work. However, I will be sure to note this difference in the discussion of results. It could be that the advantages of top-kriging are significant at particular sites or across particular sub-regions, while they are un-impactful elsewhere. This might cause the already-poor sites to continue performing poorly, while another subset might improve significantly. Again, this discussion will be added to the manuscript.

Reviewer Comment 4: *I find the reporting of both R2 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 R2 is a scaled version of MSE, so if RMSE gives a different ranking of methods*

Printer-friendly version

Discussion paper

than R2, 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!

Author Response 4: I agree that the performance metrics presented in Table 1 are not entirely independent of each other. In fact, they were all included for this very reason. The Nash-Sutcliffe model efficiency, as described by your reference and which I understand you to be calling R2, is a function of the mean squared error and is therefore related to the root mean squared error. However, the standardization by variance in the model efficiency can be shown to introduce confounding effects that lead to the differing conclusions of efficiency and root mean squared error. Gupta et al. (2009, DOI: 10.1016/j.jhydrol.2009.08.003 ; 2011, DOI: 10.1029/2011WR010962) show that the model efficiency is an amalgamation of the mean squared error and the Pearson correlation between observed and simulated response, among other things. For this very reason, both the Pearson correlation and root mean squared error are included in Table 1. The intent is to allow the reader to consider several common metrics and to tease out what is driving the Nash-Sutcliffe efficiency. The fact that root mean squared error and model efficiency do not agree on the significance of the differences between ordinary and top-kriging arises from the interplay between Pearson correlation and root mean squared error within the model efficiency. As I will point out in the revised manuscript, the decomposition allows one to better understand the components of the Nash-Sutcliffe model efficiency.

Reviewer Comment 5: *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).*

Author Response 5: A full description of the data set used can be found in Farmer et al. (2014); their data was used without undocumented modification. Across the 182 streamgages considered, there were 1.6 million observations, of which 5,435 were measured as zero. This is an average occurrence of 0.3% at each site. In fact, only 7 out of the 182 streamflow records considered contained days with zero-measured streamflow. Within these, the frequency of zeros is 0.99%, 0.06%, 31.2%, 14.6%, 4.00%, 4.09% and 1.42%. A statement on the prevalence zeros will be added to the manuscript.

The use of a placeholder value can certainly have a significant impact on the interpretation of the results. However, Farmer et al. (2014) found their results relatively insensitive to the selection of the placeholder because of the minimal fraction of zeros and zero-containing sites in the data set. However, if the zeros had been more prevalent or effected more sites, then it would certainly present a significant challenge. This discussion will be added to the manuscript.

Reviewer Comment 6: *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?*

Author Response 6: The kriging system was designed to estimate the natural logarithm of unit runoff. However, to contextualize the results of estimations, some performance metrics were calculated on the back-transformed estimates. This back-transformation was done using simple exponentiation rather than developing a unique bias correction factor. The development of a bias correction factor that could be applied in the case of an ungauged basin was beyond the scope of this manuscript but is surely interesting to future research.

It should also be noted that the skewed distribution of observed daily streamflow may negatively affect the interpretation of the Nash-Sutcliffe model efficiency of streamflow (Gupta et al., 2009, DOI: 10.1016/j.jhydrol.2009.08.003). For this reason, the Nash-Sutcliffe of the logarithms of streamflow is the more reliable metric. This fact further obscures the need for an explicit bias correction in this application. In addition to the previous discussion of performance metrics, this discussion will be added to the manuscript.

Reviewer Comment 7: *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?*

Author Response 7: This work relied on the geoR package developed by Ribeiro and Diggle (2015, geoR: Analysis of Geostatistical Data. R package version 1.7-5.1. http://CRAN.R-project.org/package=geoR). The scripts required to produce this data are not currently available in a publishable format nor do I have the capacity to revise them to a publishable format at this time. However, I do think that a pooled estimation procedure would be a great addition to packages like geoR and gstat. For future work, I would be happy to discuss this development with you further.

Reviewer Comment 8: *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?*

Author Response 8: The daily parameter values do indeed appear crazy. This was why the 31-day moving average was used improve the readability of figure 6. I will add a description of this variability, noting that it was chaotic, to the paragraph in lines 23-28 of page 8. I will also include this in the discussion in lines 27-33 of page 10.

Reviewer Comment 9: *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.*

Author Response 9: It may be that there is a particular averaging procedure that optimizes model performance. This manuscript considered only two end-members of the continuum: daily parameters and time-invariant parameters. Using a moving-average parameter set may dampen daily effects and improve performance but may also reduce the computational advantages of time-invariant models. The exploration of the continuum will be added to the revised manuscript, but is beyond the scope of this work. Certainly, the results herein encourage exploration of alternative time-averaging.

Reviewer Comment 10: *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 ignored. 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.*

Author Response 10: Spatio-temporal kriging presents another opportunity for explicitly handling the temporal variation of streamflows and variogram parameters. As the reviewer suggests, this would be especially important for applications seeking to fill-in sparse records. Here, the question of temporally sparse records was not considered. For purposes of validation, each site was treated as if it were completely ungauged. Furthermore, the work of Skøien and Blöschl (2007, DOI: 10.1029/2006WR005760) showed only limited returns for spatio-temporal considerations in hydrologic time series applications. I will note this in the revised manuscript, but the exploration is left for future work.

Reviewer Comment 11: *In general, the manuscript confuses semivariance with covariance; I strongly suggest to use only one of the two.*

Author Response 11: I will revise the manuscript to more clearly distinguish between covariance and semivariance.

Reviewer Comment 12: *if Figure 1 would show the (main) rivers, or catchments, we could see the degree to which catchments are contained by each other.*

Author Response 12: These will be added to the figure.

Reviewer Comment 13: *4:21 i = [1, ..., n]*

Author Response 13: Revised.

Reviewer Comment 14: *4:21 Euclidian location: omit Euclidian*

Author Response 14: Revised.

Reviewer Comment 15: *5:9 then should be that*

Author Response 15: Revised.

Reviewer Comment 16: *5:9 μ is the Lagrange multiplier, not an estimate of the mean of z*

Author Response 16: Revised to read "...and estimates the LaGrange multiplier, μ, to control for the unknown mean of z."

Reviewer Comment 17: *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?*

Author Response 17: The text in lines 10-14 of page 5 will be revised to read "The single realization of [C] that is produced from the sample observations of z cannot be considered to represent the underlying system. The sample may produce a matrix that is singular or not positive definite, conditions required for solution of the system. Furthermore, the elements of [D], by nature, are unobservable as the value of the dependent variable at the ungauged location is what is being estimated. However, with additional assumptions of stationarity, semivariance can be modeled as a function of separation distance. Several classical models are available to ensure positive definiteness. These models are parameterized by calibration to the empirical variogram of observed semivariance as a function of distance. Once a variogram model is selected, the system becomes . . .”

Reviewer Comment 18: *5:12 a variogram is not a model of the covariance*

Author Response 18: Revised to "using a model of semivariance”. Per recommendations of other reviewers, this portion will be revised to present the kriging system as a function of semivariance alone. This will improve the confusion between covariance and semivariance.

Reviewer Comment 19: *5:21 "theoretical variogram model", replace with "variogram model type".*

Author Response 19: Revised.

Reviewer Comment 20: *5:22 "in building the empirical variogram, the covariances": again, variogram/semivariance and covariogram/covariance are two different measures.*

Author Response 20: "Covariances” will be replaced with "semivariances”.

Reviewer Comment 21: *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)*

Author Response 21: The maximum distance was the maximum inter-site distance in the network, 920 kilometers. The bins were equal-interval bins. The following statement will be added: "... were stratified into ten equal-interval groups based on the

Printer-friendly version

Discussion paper

inter-site distances ranging from zero to the maximum inter-site distance of 920 kilometers, as suggested by ...”

Reviewer Comment 22: *6:6 same confusion: the variogram does not have covariance values.*

Author Response 22: Replaced "covariance" with "semivariance".

Reviewer Comment 23: *6:15 "stationarity" -> "temporal stationarity"*

Author Response 23: Revised.

Reviewer Comment 24: *6:20 "computation of as many variograms" -> "fitting of as many variogram models"*

Author Response 24: Revised.

Reviewer Comment 25: *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.*

Author Response 25: Per the additional recommendations of other reviewers, the discussion in the last paragraph of page 6 will be expanded to further clarify the distinction between the pooled model and the averaged model.

Reviewer Comment 26: *8:17 "and" -> "a"*

Author Response 26: Revised.

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

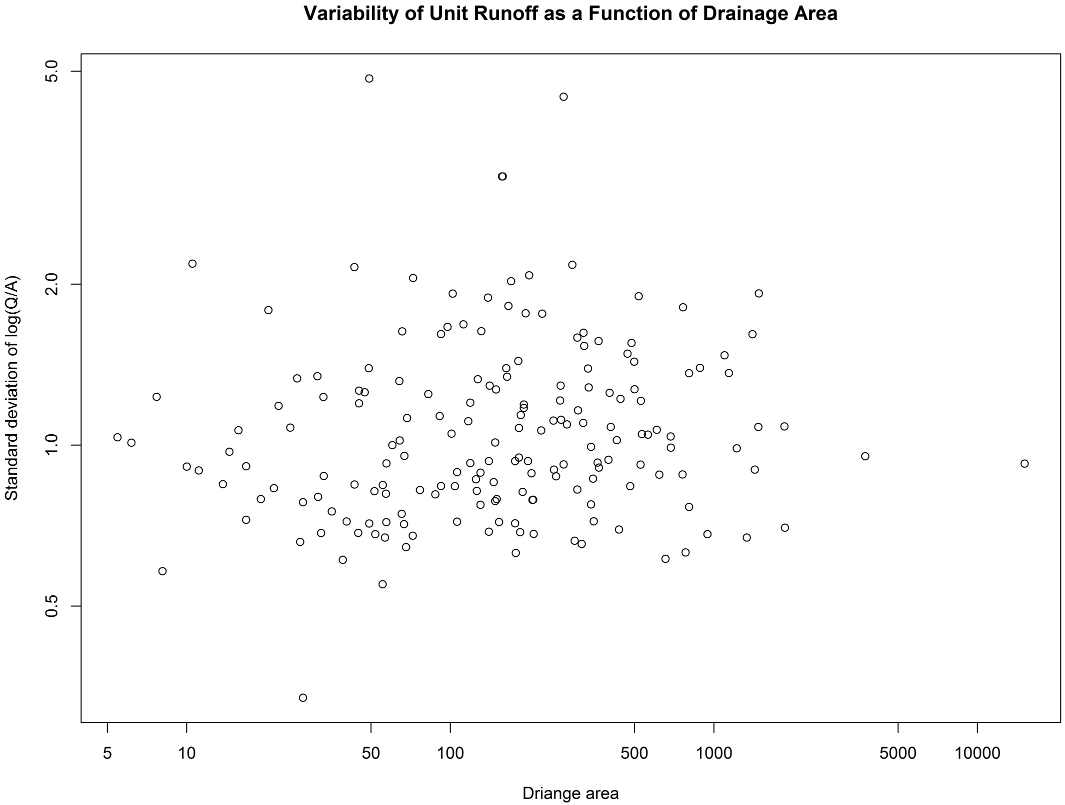

Fig. 1. Temporal variability in streamflow with drainage area.

Printer-friendly version

Discussion paper