Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2015-536-AC2, 2016

Hydrology and Earth System Sciences
Discussions

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: *This manuscript analyses ordinary kriging for estimation of historical daily streamflow. The paper is well written, and includes interesting analyses. Some revisions are still necessary before it can be published. Below are some suggestions for improvement.*

Author Response 1: Thank you for a thorough and valuable review of my work. I greatly appreciate the effort you have put into improving the impact and communication of my work. I hope that, with your improvements, this work will motivate future research.

Reviewer Comment 2: *There is some overlap between results and discussion, where some results are discussed in the results section, and then this is partly repeated in the discussion, which is then more like a summary. I think this part of the manuscript*

could be better organized.

Author Response 2: I will revise the sections so that the 'Results' section is re-named as a 'Results and Discussion'. The 'Discussion' section will be revised and re-named 'General Discussion' as a subsection of 'Results and Discussion'.

Reviewer Comment 3: *Some work is based on the author's PhD (Farmer, 2015). I am not sure how easily accessible this PhD is? I think it is ok to include results from the PhD in this manuscript as previously unpublished, as long as they have not been presented in other peer reviewed journals/conference proceedings. The paragraph on P6 L12-17 should anyway be rewritten. I don't see why it is not intuitive to model each day independently? What is the most extreme? What is meant by stationarity of variogram parameters here, that they are temporally constant?*

Author Response 3: My Ph.D. dissertation is available through my alma mater, but is not freely-available online. Several chapters have been published elsewhere, but the chapter pertaining to this material has not been published.

I will revise the noted paragraph for clarity. Modeling days independently is non-intuitive because our basic understanding of hydrology shows strong temporal dependence across days. I called the averaging of parameters the most extreme case because it exists on a continuum of estimating every day and taking an average of all days. As discussed with a previous reviewer, you could consider any range of averaging (e.g., a 31-day moving average). I will revise to make this point more clearly. Here I am referring to temporal stationarity and will add that adjective.

Reviewer Comment 4: *The brief summary of kriging is not brief; it pretty much includes everything in Skøien et al. (2006), which the author refers to, just in a different way. What is missing is actually a variogram model and the sample variogram, which is of more interest for the analyses than the equations for finding the weights.*

Author Response 4: The reviewer is correct, even though the spherical variogram was

used (line 20, page 5), I could more clearly describe the variogram model considered. Per recommendations from another reviewer, I will revise the kriging system presented to be based on semivariance. I will also add the following description of the spherical variogram model:

$$\gamma(h) = \frac{1}{2}E[(Z(\mathbf{x}+h) - Z(\mathbf{x}))^2] \tag{1}$$

Where $\mathbf{x}$ is a geospatial location and h is a separation distance. To improve stability of the system, the semivariance $\gamma(h)$ is approximated as

$$\hat{\gamma}(h) = \begin{cases} (\sigma^2 - \tau^2)(\frac{3h}{2\phi} - \frac{h^3}{2\phi^3}) + \tau^2 & if \quad h \leq \phi \\ \sigma^2 & if \quad h > \phi \end{cases} \tag{2}$$

Where σ^2 is the sill, ϕ is the range and τ^2 is the nugget variance.

Reviewer Comment 5: *I think it is necessary to discuss in more detail the advantages/disadvantages of OK/TK and pooled/daily variograms beyond the cross-validation results. First of all, OK can be seen as the most extreme case of TK with only one regularization point, and comparing with the results of Skøien et al (2015), it is not surprising that also OK can perform reasonably well for most catchments. However, this is likely to depend on the configuration of observation and prediction locations. Table 1 indicates that the NSE of TK is considerably higher for the 10th percentile. Then the author barely mentions the prediction uncertainty, where I would expect daily variograms to perform better than pooled variograms, and TK should perform better than OK.*

Author Response 5: I agree that this is far from a thorough comparison of ordinary and top kriging. The intention of this paper is to introduce ordinary kriging as a competitor to standard methods. A more robust comparison, as was discussed with another reviewer, may be more appropriate in future research. There are several unanswered questions, including the configuration of observation and prediction locations. Why the 10th percentile of TK is better is not currently known. I will add text to point out this.

In line with the recommendations of another reviewer, I will add some prose to discuss the similarity of ordinary and top kriging. It is especially important to note that one is a modification of the other.

One of the main advantages of kriging methods is the availability of prediction uncertainty. However, the possibility for developing confidence intervals or uncertainty estimates on daily estimates is a subject that merits its own thorough research. As there is no standard method for single-index techniques, there are no grounds for comparison. Further research is needed, but I will note this in the manuscript.

Reviewer Comment 6: *Regarding biases of upper/lower extremes, particularly P8 L3-10 and P12 11-14 – I think it is not surprising that kriging tends to over/underestimate the extremes. Kriging is generally unbiased around the 50th percentile, and a range around this percentile. For the conclusions, kriging is a stochastic approach, not deterministic, and using simulations would only create a range of values which is still centred around the under/overestimated prediction. Higher nugget effects will increase the tendency to smooth the extremes in the interpolated field.*

Author Response 6: I referred to kriging as a deterministic approach because it produces a single estimate of daily streamflow from a single, deterministic input. I will clarify in the text. Of course, as you suggest and as I mention in the manuscript, the estimates can be coupled with prediction uncertainties to be used in a deterministic manner. However, the simulations will remain clustered around biased averages. What is needed is a bias correction factor. The derivation of such is beyond the scope of this work but is of immense interest and I look forward to future research on the topic.

Reviewer Comment 7: *The average underestimation of 40% for the largest streamflows is still somewhat large, it would be useful to further analyze and discuss the causes for such large deviations, and see if there are particular cases where they are larger. On the other hand, the prediction difference might be smaller, as this refers to exceedance probability, something which could be better explained. It is not easy to*

understand from the text what is meant by exceedance probability in P8. The sentence "For low streamflows, below . . ." is not clear either.

Author Response 7: I will clarify the text to refer to non-exceedance probabilities rather than percentiles. The 40% error is in high streamflows, making it all the more concerning. Why such large deviations occur is not well understood. Kriging, in of itself, is a smoothing estimator and the temporal pooling further smooths results. This smoothing may mask the effects of rare, extreme events and dampen their amplitudes. I will note this in the manuscript.

Reviewer Comment 8: *Although not frequently in use as far as I know, there are some methods for unbiased kriging of extreme values. Two of them are IWQSEL (Craigmile et al., 2006) and Modified Ordinary Kriging (Skøien et al, 2008), both methods implemented in the Rpackage intamap. I don't think this would be feasible to include for the analyses in this manuscript, but could be a possibility for future work.*

Author Response 8: I am excited to consider these methods in future work.

Reviewer Comment 9: *The results from the analyses of autocorrelation on P8 should probably not focus so much at what I would refer to as relative differences in percent (as is currently done, although referred to as absolute), rather the absolute errors. As mentioned, relative differences when autocorrelation is below 0.1 can be large, but still negligible. A figure could present the autocorrelation as lines.*

Author Response 9: I will consider revising this figure, as the reviewer suggests. I agree that is could be improved by showing the deviations in some way that minimizes the impact of inconsequential differences.

Reviewer Comment 10: *The discussion about temporal variations of variogram parameters at the end of 3.2 should also include some more thoughts about the reason for the temporal changes, and check the description of the current relationships. I would assume that shorter ranges in Summer is an indication of more heterogeneity*

(more convective precipitation) and that long ranges is an indication of homogeneity. I would also assume that sill is decreasing in Winter because runoff is decreasing, so the ratio sill/mean runoff could be of interest, the same with the nugget/sill ratio. Long ranges in Winter/spring could be related to snow/snow melt. "beyond the range" is confusing.

Author Response 10: As the reviewer suggests, an understanding of the impacts of any parameter changing in isolation is not easily described. I chose to plot the raw parameters because they are constant across all sites when the whole network is used for calibration. Standardizing the sill by mean values thus becomes difficult. Though, I do agree that it would be useful for additional analysis. The ratio of nugget to sill is also useful.

As described on line 15 of page 9, I agree that the decreased range in Summer is emblematic of less homogeneity and more heterogeneity. I'll make this point more clearly. I will also add some of the reviewers' comments to the manuscript after exploring the ratio of sill to nugget. The attached figure shows the 31-day average ratio of nugget to sill. January through April, the nugget accounts for 20-30% of the sill, it dips to only 5% of the sill in mid-May and then steadies to about 15% of the sill through the rest of the year. The pooled parameter sets the nugget at 15% of the sill. Lower Winter sills may be the result of smaller streamflows. The measurement uncertainty in smaller streamflows may be greater, increasing the nugget. For high streamflows in mid-May, the measurement error represented by the nugget may be minimal. Average flows throughout the remainder of the year show a standard degree of measurement error, agreeing with the pooled parameter. I will consider adding this figure and discussion to the manuscript.

Reviewer Comment 11: *P2 L12 "It is postulated" – Who postulated what?*

Author Response 11: This sentence is meant to state my hypothesis that "predictions of daily streamflow time series can be improved by incorporating regional information

beyond the information available at single index streamgages and that, building on previous hydrologic time series analysis, this can be achieved by utilizing the geostatistical method known as kriging." I will revise to more clearly state my hypothesis by stating "It is hypothesized here that..."

Reviewer Comment 12: *P2 L27-28 There are also kriging methods where the predictions are based on external variables in addition to geolocation.*

Author Response 12: There are; I will revise the text to say that co-kriging is a variation of kriging that incorporates the influence of other variables.

Reviewer Comment 13: *P3 L9 "Deterministic rainfall-runoff models" is more commonly used than mechanistic.*

Author Response 13: I will change the term throughout.

Reviewer Comment 14: *P3 L9-10 I would say that the comparison is missing from their work, they did not emphasize the need.*

Author Response 14: I meant to imply that the lack of a comparison emphasizes the need, not that they emphasized the need. I will revise to "Because it has not been previously considered, it is important to explore and contrast the potential of ordinary kriging and top-kriging to predict streamflow time series in ungauged basins."

Reviewer Comment 15: *P3 L25 logarithms of UNIT runoff?*

Author Response 15: Yes, runoff is considered as a depth here, whereas discharge is considered a volume. I will add "unit" for clarity.

Reviewer Comment 16: *P4 L1 The sentence is a bit clumsy (Because . . . because), consider rewriting, such as ". . . reference-quality in the designation (. . .) or in previous flood-frequency studies (. . .)." Remove brackets from next sentence.*

Author Response 16: Will be revised as "...reference quality according to their designation ...". The next sentence will be removed from parentheses.

Reviewer Comment 17: *P4 L10 One out of 33 days on average or some periods of 1-33 days? These were filled by the author or has previously been filled in by USGS?*

Author Response 17: These were previous filled by Farmer et al. (2014) and contained periods ranging in length from one to 33 days. "on the order of one to 33 days" will be revised to "for periods of one to 33 days long".

Reviewer Comment 18: *P4 L16-17 This sentence seems unnecessary complicated, and I am not sure if it is completely correct.*

Author Response 18: Revised to "The inter-site semivariances of data from a measured network can be used to create a system of linear equations predicting the semivariance at an unmeasured site to be a linear sum of the semivariance between all observed sites."

Reviewer Comment 19: *P5 L17 Which previous hydrologic geostatistics is the OK in this manuscript an extension to? And depending on the answer, is OK really an extension or does the manuscript include analyses which are useful as an addition to other* methods?

Author Response 19: The application, not OK itself, is an extension of previous applications. The extension is that, here, time series are being considered rather than streamflow statistics. As the reviewer points out, this method is really an addition rather than an extension. I will revise to "Ordinary kriging of streamflow time series builds off of previous hydrologic applications to predict streamflow statistics to produce a method for handling temporal variation along with spatial variation."

Reviewer Comment 20: *P5 L30-31 I think "the temporal considerations" can be deleted.*

Author Response 20: Will revise to "focused on kriging time series and the temporal behavior of kriging parameters."

Reviewer Comment 21: *P6 L6 I don't think it is the covariance, but it could be the*

variance, due to short scale variability or measurement errors.

Author Response 21: Per the recommendation of another reviewer, I will revise to speak only in terms of semivariance. The nugget, therefore, is the semivariance of co-located points, which can be non-zero due to short-scale variability or measurement errors. Revisions will reflect this understanding.

Reviewer Comment 22: *P6 L8 "the dependent variable" can be deleted, together with "of" (structure of which)?*

Author Response 22: Will be revised to "In some previous hydrologic applications of kriging, the semivariance, which is modeled by the semivariogram, has been assumed to be temporally constant..."

Reviewer Comment 23: *P6 L21 What is stability of the spatial covariance structure?*

Author Response 23: Stability is meant to imply that the parameters of the semivariogram are constant. Will revise to "If the parameters of the semivariogram can be reasonably assumed to be constant, then the computational ..."

Reviewer Comment 24: *P6 L23 Move first sentence to L25 (after daily variogram sentence).*

Author Response 24: Revised.

Reviewer Comment 25: *P6 L31-32 Here it is a bit unclear what is meant by "average model". In L15-16 it is referred to averaging of model parameters, which is definitely different than a variogram model fitted to a temporally pooled empirical variogram. If the difference between average and pooled refers to the difference between treating each daily empirical variogram as equal, or giving them weights according to the number of pairs in each bin, then this should be described more explicit.*

Author Response 25: The reviewer's understanding is correct. Per additional advice from other reviewers, I will revise to more clearly state the difference between the

average and the pooled model. "The average model treats each empirical variogram equally, while the pooled model weights each bin by the number of pairs in each bin."

Reviewer Comment 26: *P7 L11 The usage of top-kriging in this manuscript is not exactly the same as the one described in Skøien and Blöschl (2007). The previous paper uses spatio-temporal regularization, whereas the implementation in rtop only uses spatial regularization. However, it can be assumed that the difference between these is small, and not likely to affect the quality of predictions. A formal comparison has not been done, but the current version of rtop uses a similar method as the one in the manuscript for predicting time series of runoff.*

Author Response 26: Thank you for this information. I will add the statement "Using the same metrics, ordinary kriging was contrasted with an application of top-kriging similar to that defined by Skøien and Blöschl (2007). Top-kriging was applied using the rtop package (Skøien, 2015), which uses spatial regularization rather than the spatio-temporal regularization presented by Skøien and Blöschl (2007). The differences can be assumed to be negligible for this application. Regardless, here, top-kriging was applied..."

Reviewer Comment 27: *P7 L27 Is this poor performance in the recession period typical for all catchments, or is this example worse than many others? I would expect a bias in the extreme, as presented later in the paper, but also that some catchments will even be underestimated during May and June in the figure.*

Author Response 27: As can be seen in the figure, some recessions are reproduced well, while others are misrepresented. See, for example, the recessions from January through March and compare with the recession May through July. This variability in recession reproduction is typical; there is no categorical reproduction. Future research may explore the reproduction of specific streamflow regimes or signatures.

Reviewer Comment 28: *P7 L31 - P8 L2 I find this sentence somewhat unclear, in addition, is the difference significant? For me there is barely any difference between*

the curves. Regarding the figure, I usually don't like grids in figures, but this could be an exception where it might add some information.

Author Response 28: I will revise to note the insignificance of these differences. It may read "Though the differences between the curves from the pooled and daily variograms are not significant, the pooled variogram produces estimates with slightly fewer large errors." I will add grids to the figure.

Reviewer Comment 29: *P8 L17 and -> a*

Author Response 29: Revised.

Reviewer Comment 30: *P11 L15-17 I did not understand this sentence. What is meant by hours and days per site?*

Author Response 30: I will revise to "... (depending on processor speeds, top-kriging required just less than three days of computation time for each site predicted, while ordinary kriging required only hours of computation time per site predicted)."

Reviewer Comment 31: *P11 L18 This was correct in the past, but pooled variogram estimation is now included in the package, together with time series interpolation.*

Author Response 31: Will revise to "At the time of application, there package by Skøien (2015) did not contain a method to estimate pooled variograms directly. More recent versions do contain this functionality."

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

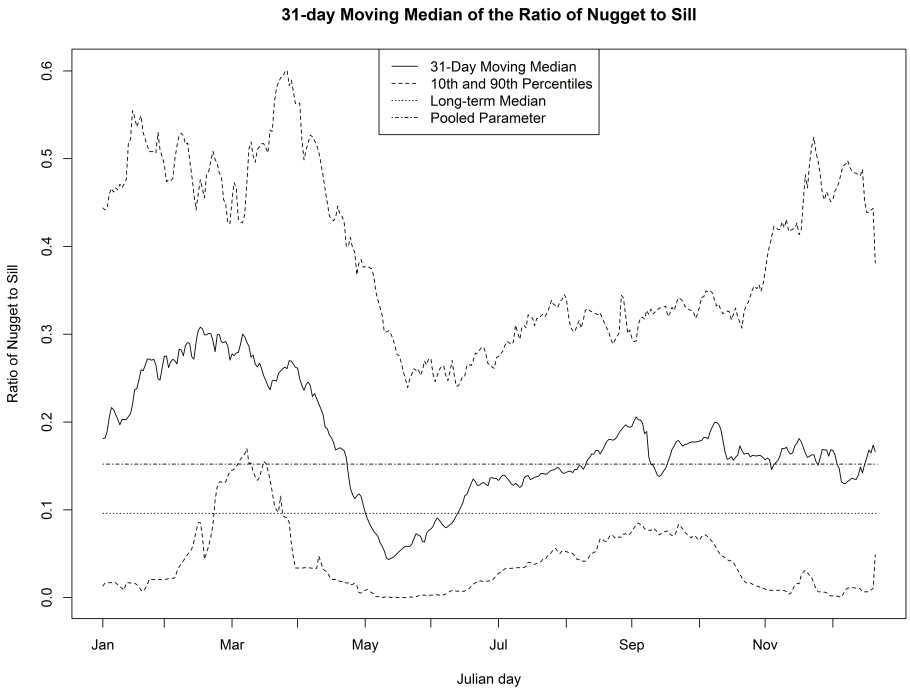

Fig. 1. 31-day moving median of the ratio of nugget to sill.