Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2015-536-RC3, 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

A. Pugliese (Referee)

alessio.pugliese3@unibo.it

Received and published: 18 March 2016

The paper "Ordinary kriging as a tool to estimate historical daily streamflow records" by Farmer W.H. shows a comparative assessment of kriging techniques, exploring the performances obtainable employing Ordinary Kriging, under different model settings, for the prediction of daily streamflow series in ungauged basins. The paper is well written and is rather complete in all its section, the topic is of wide interest in the hydrological field, thus I believe it is suitable for the publication in HESS after some little improvements that in my view the author should consider to take into account.

Printer-friendly version

Discussion paper

Major comments

1. Even if there is a relationship between the covariance function $C(x_1, x_2)$ between two data points x_1, x_2 and the variogram, which is, by definition: $\gamma(h) = 1/2E[(Z(x_2) - Z(x_1))^2]$, where $h = x_2 - x_1$ is the spatial Euclidean difference between the two data points and E is the expected value of the squared increment of Z, relative to the spatial lag h. All the textbooks and papers on geostatistics refer to the variogram, rather than employ the covariance directly, as the major controller of the spatial correlation. C and γ are two sides of the same coin, because $\gamma(h) = C(0) - C(h)$, though the variogram has some more features, which is why it is the main function to look at. For instance, most of the variables that might be referred to as "spatial fields" may (or may not) have a nugget effect, which is a unique feature of the variogram. Moreover, there are some variables that might be "non-stationary". In this case, one can denote non-stationarity as the variogram diverge and never reach the "sill", while non-stationarity might not be seen from the covariance. I think the mathematical notations and equations (1), (2), (3) and (4) (L25 P4) are formally incorrect as they refer to the covariance, rather they should refer to semivariances of the increment $z(x + h) - z(x)$, both theoretical or experimental (see for examples, Cressie, 1993; Journel and Huijbregts, 1978). Although there is a way to employ the covariance matrix too, which derives from the optimization of the prediction variance, the author did not report the correct one. I would recommend the author to rewrite the system of equation (2), (3) and (4) and stick with the variogram. Furthermore the author cites Skøien (2006) as the refernce for solving the kriging system. There are a couple of mistakes with this reference: (1) that paper focuses on solving an "adapted" ordinary kriging linear system to fit with regularized variorgams, so maybe this is not the best choice for someone who wants to discover more about kriging tecniques and (2) that paper never reports covariaces within matrices of the kriging linear system to solve, rather it reports correctly variograms.

2. I think that the comparison with Top-kriging here is not informative as it should and might be even misleading. Firstly, it does not report the best model setting. Even if the author specifies here and there that the comparison with Top-kriging is not definitive, I would strongly recommend to point out that Top-kriging model performances might not be the best obtainable in this study area. Or, in case preliminary analyses have shown that it is instead the best model setting, this should be clearly said throughout the manuscript. Secondly, I think the paper, which is intentionally unbalanced towards the two ordinary kriging methods, does not accomplish the assessment Top-kriging deserves. Indeed, the latter is actually an ordinary kriging too, technically it is a "modified" ordinary kriging, where the modification relies just on the variogram. The author instead groups this method together with DAR and QPPQ, whereas it should be grouped with the ordinary kriging methods. Concluding, does this comparison with Top-kriging reflect what the title says? At the very end, is this informative? Thus, should the Top-kriging be removed from the comparative assessments with other models?

Minor comments

1. L 24-25 P10. The author conclude that kriging techniques are biased and inaccurate in the tails of the distributions, and prove it with Fig. 4. This is somehow understandable and even quite normal. The kriging techniques are weighted average. Predicting streamflows within a leave-one-out cross-validation, when the lowest or highest streamflow is removed, plus it is perhaps orders of magnitude lower or higher than streamflows from donor sites, it is predictable that the outcome shows upward or downward biases, respectively, in those regimes. I think this thought might be taken into account, or at least pointed out clearly, maybe after those lines or elsewhere in the first sections.

Printer-friendly version

Discussion paper

2. In case the author adopted unconstrained kriging methods, that is when kriging weights are both positives and negatives, predicting low flows may lead to negative estimates. Would be interesting to know, if any, how many negatives are produced.
3. By taking the logs of the standardized streamflows, the author implicitly removes zero flows, if any, from the dataset. Is there any catchment with intermittent regime? If so, would be interesting to see how zero flows are treated.
4. L. 33 P9: it is not clear whether or not the author adopted the leave-one-out cross-validation for the DAR and QPPQ method too. I think so, but I would recommend to rephrase and be a little clearer about the cross-validation procedure used for all the methods reported. Would be even informative to know if any other cross-validation methods have been used in the past for DAR and QPPQ.
5. L. 31-32 P6. I think the parentheses might be removed and extending the text for a few lines might improve the reasoning.

Technical notes and misspellings

1. L. 32 P6, I've noticed the author use to put the punctuation mark before the right hand parenthesis, please correct throughout the manuscript with ").”
2. L2 P8. "[. . .] developed form", should be "developed from".
3. L8 P8. "[. . .] between the 5% and 15% non-exceedance probability" should be perhaps "between the 5% and 15% error"?
4. L30-31 P10. There are two "similarly" adverbs very close one to another. Please, consider to substitute one of them.

5. Fig. 4 height might be increased. In general Fig. 4 form factor might be changed to improve the readability of the figure itself.

References

Cressie, N.A.C., 1993. Statistics for spatial data, Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley.

Journel, A., Huijbregts, C., 1978. Mining Geostatistics. Acad. Press.

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