Hydrol. Earth Syst. Sci. Discuss., 7, C4248–C4254, 2010
www.hydrol-earth-syst-sci-discuss.net/7/C4248/2010/

Hydrology and Earth System Sciences Discussions

Interactive comment on "

Geostatistical regionalization of low-flow indices: PSBI and Top-Kriging" *by* S. Castiglioni et al.

S. Castiglioni et al.

attilio.castellarin@unibo.it

Received and published: 17 December 2010

We thankfully acknowledge E. Pebesma's very useful comments. We find Referee#1's comments and suggestions extremely useful for improving the presentation of the work and enhancing the accuracy and rigorousness of the description of the methodologies we used.

Our reply is structured as follows, we report all referee's comments (indicated by RC) together with our reply (denoted by AR, Authors' Reply).

RC:

The paper gives a fair, open and interesting comparison of two methods for the estimation

(prediction/interpolation) of runoff at ungauged locations / catchments. One

of them, Top-kriging, relies on spatial correlation of the measured values, and takes

into account the differences in support and the nested nature of catchments and subcatchments.

The other, PSBI, obtains predictions by applying a kernel smoother in feature

space, here the first two principle components obtained from a set of 9 variables

related to geomorphology and lithology. The paper is written very well, and deserves

publishing.

AC:

We are pleased by the positive view of the manuscript casted by Referee#1's review. Furthermore, we are particularly thankful also for critical comments raised by Referee#1, as they will help us to significantly improve the presentation of our work.

RC:

The paper concludes that the two approaches perform similar, and that gains can be

expected when the strenghts of both methods were combined. Could the authors also

reflect on how ambitious and complicated this would be? In case such a solution were

not feasable for reasons of constrained resources, in which cases would the authors

expect Top-kriging to beform better, and when would PSBI be the prefered choice?

What about the data requirements of both?

AC:

We are currently working on a possible integration of the two methodologies that aims at combining the advantages of both. A possible solution for the integrated approach would be applying PSBI to the residuals of predictions obtained by Topkriging in cross-validation. Given the discussion of complementary features that characterize the two methodologies, which is already present in the original version of the manuscript and provides some indications on when one methodology is expected to outperform the other (see e.g., from line 26 on page 7247 to line 19 on page 7248 of the original manuscript), we believe that this strategy would combine advantages of the two methodologies. Even though this analysis is clearly out of the scope of the present study, we are persuaded that a reference to it in the discussion section (presenting future developments) is worth including. We also will clarify the discussion of results to make crystal clear in which cases we "expect Top-kriging to perform better, and when would PSBI be the preferred choice".

Data requirements are discussed explicitly in section 4.2.1, but comments on this issue will be reported also in the revised version of the Discussion section.

RC:

On page 7248, line 20: the authors argue that "Top-Kriging and PSBI are both geostatistical

interpolation methods." I would argue against this. In my opinion, explicitly

addressing, quantifying, and exploiting autocorrelations in geographic space sets geostatistical

methods apart from other statistical methods. PSBI is a kernel smoothing

technique that uses a covariance based kernel, and does not explicitly address spatial

autocorrelation in observations or residuals.

AC:

Point taken. Terminology will be revised throughout the text accordingly.

RC:

Using this covariance kernel in the space of two principle components issues a number

of new questions. Was stationarity a fair assumption in this space? Were the principle

components scaled, and if yes, how? If not, was some sort of anisotropy assumed or

fitted? If not, on which ground was isotropy assumed?

AC:

Referee#1 is right, we probably should have provided more details on the implementation of PSBI in this study. We referred to Castiglioni et al., (2009), missing an opportunity to get a more self-contained paper. We presented here only a possible implementation of PSBI, the one that performed the best in Castiglioni et al. (2009) for predicting Q355 in ungauged sited (that is in cross-validation). Therefore, we will include some additional details in the revised version of the manuscript (together with what already reported in the manuscript, see pp. 7239-7240): principal components were not scaled, they were computed for normalized catchment descriptors; preliminary analyses showed that the utilization of anisotropic variograms produced poorer results in cross-validation, therefore we decided to assume isotropy to hold. In general, the suitability of all assumptions was assessed via leave-one-out cross-validation and comparison with standard multiregression techniques.

RC:

And further: why was the kernel

limited to 2 PC's? Why discard 30% of the information, and why not use a positive

definite kernel in all dimensions? (Note that in this case the spherical covariance would

not provide a positive definite model, but the exponential or Matern family would.)

When a dimension reduction is needed, why did the authors not choose the linear combination(s) of features that best explained the variability in the target variable, in Q355? Partial least squares but also ridge regression would do this. How did the PSBI compare e.g. to a direct multiple linear regression? This might reveal the benefits (or disadvantages?) of restricting to two dimensions and doing the smoothing in feature space.

AC:

We applied PSBI as proposed and illustrated in the literature (see e.g., Chockmani and Ouarda, 2004). Since we had only one measure to predict (i.e., Q355), the utilization of Canonical Correlation Analysis could not be used to define a two-dimensional physiographic space. Hence we resorted to PCA to reduce the dimensionality of catchment descriptors (see e.g., Chockmani and Ouarda, 2004). We selected a two-dimensional interpolation, using therefore only the first 2 PCA's, to enable visualization and mapping of Q355 on the physiographic space (see Castiglioni et al., 2009). We did compare PSBI with standard multiregression techniques (see also our reply to previous point), and PSBI significantly outperformed them (see Castiglioni et al., 2009).

We will report these comments in the section dealing with the implementation of PSBI to the study region (Section 4.1.1).

RC:

A number of relatively minor issues follow:

On page 723, I find the use of Q355 and Q95 confusing as the 355 and 95 refer to different properties, number and percentage of days.

AC:

The introduction will be revised by incorporating a better definition of Q355, as also requested by the other referees.

RC:

Page 7237, Line 16: I suggest to replace "most of the information" with "most of the variability". Some critical information might show up in lower order PC's, and the hope of course is that this is not the case, but there is no guarantee as PCA can't tell the difference between information and variability.

AC:

The text will be modified as recommended.

RC:

7239, L 24: The authors should mention whether PCA was done on the covariance matrix or the correlation matrix. As they evaluate eigenvalues larger than 1, it seems to be the correlation matrix, otherwise this criterion does not make sense, but the choice needs to be mentioned explicitly.

AC:

We will explicitly mention that PCA were computed by referring to the correlation matrix

RC:

7240, L 9: R package Gstat should be written gstat. R package Rtop is not on CRAN; could the authors indicate where it can be found?

AC:

We will use the notation indicated by Referee#1.

Rtop will soon be uploaded to CRAN, but the latest version can in the meantime be found from:

http://intamap.geo.uu.nl/~jon/rtop/

and a reference paper is:

Skøien, J., E. J. Pebesma, G. Blöschl, 2009. Rtop – an R package for interpolation along the stream network. In: Proceedings StatGIS 2009, Geoinformatics for Environmental Surveillance (ed. G. Dubois)

Relevant link and reference will be included in the revised manuscript.

RC:

7241, L 13: you did not update the PCA, ..., but ... – to me suggests this could have mattered. Maybe you could further stress here that the PCA computation does not involve the Q355 data at all, so leaving those out does not affect the PCAs.

AC:

We will modify the text as suggested.

RC:

7258: Figure 3 should number sub-figures as a, b and c, and properly refer to those in the figure caption.

AC:

We will modify the figure as suggested.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 7, 7231, 2010.