Hydrol. Earth Syst. Sci. Discuss., 10, C5780–C5785, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C5780/2013/

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Interpolation of extensive routine water pollution monitoring datasets: methodology and discussion of implications for aquifer management" *by* Yuval et al.

Yuval et al.

lavuy@tx.technion.ac.il

Received and published: 22 October 2013

We thank the Reviewer for the efforts of reading our manuscript. Below are our responses.

Response to the General comment

The Reviewer does not agree with our use of the IDW interpolation method. There is no reference to our exhaustive assessment of the interpolation results and the quality of the interpolation which they revealed. Such exhaustive assessment is not always possible to carry out objectively using the alternative methods which the Reviewer

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

suggests later, especially not for a work dealing with dozens of pollutants. Nor does the Reviewer refer to the inherent trade-off between the interpolation accuracy and spatial coverage which our work demonstrated in a quantitative way. This trade-off may seem trivial but many times practitioners tend to overlook it. The issue is valid for any type of interpolation used and our manuscript provides admittedly simple but quantitative tools to assess it in an objective way. We are disappointed by the general rejection of the work by the Reviewer. We are sure that her/his experience could have greatly benefited the manuscript had she/he commented on the specific points that the manuscript make.

Responses to the specific comments The comments were written in one paragraph. Our responses break it down so that the different points can be answered separately.

C1. IDW is a non-statistical method for spatial interpolation.

R1. Statistics is a science of collecting and analysing numerical data for the purpose of inferring and gaining understanding of the phenomenon that the sample of numerical data observed. In that sense any method, whatever it is, which sheds light on a phenomenon using the relevant observations is statistical. And, our work indeed shed such light on a true practical problem. We used a very simple interpolation method, and we will explain in our responses to other points below why we resorted to using it. We are well aware and familiar with the available spectrum of geostatistical tools but to our opinion, for the circumstances of our work and the points which we wanted to make, IDW is sufficient, if not the best method.

C2. The authors claim that given the variability structure of their data and it not following well-behaved statistical distributions that this eliminates a statistical approach like those developed in the ïňĄeld of geostatistics.

R1. Not fully correct. We indeed mentioned the non-standard distributions of our data sets as a problem but the main point which the Reviewer omits is our claim that methods like Kriging variants or Thin Plate Smoothing smear some values (whatever

HESSD

10, C5780–C5785, 2013

transformation they have gone through) to areas clean of a pollutant. A quote from the manuscript: "Clearly, interpolation between high localized pollution values and zero, or very low values, results in an unrealistic smearing of the pollution beyond the true polluted area and leads to inefficient use of groundwater resources.". This is an issue of prime importance to water resource management which sophisticated interpolation methods are not good at without making many additional subjective decisions (e.g., some arbitrary cut-offs based on the Kriging uncertainty).

C3. Data transformations and developing methods based on statistical mixture distributions (to handle the speciïĭňĄc distributions described) are two immediate ideas for applying geostatistical methods to this problem.

R3. We agree with the Reviewer that data transformations might have helped handling the large dynamic range of data values. However, they would not have helped in denoting with zero interpolation values areas believed to be clean of the pollutant. This is crucial as aquifer managers would not accept water production from areas suggested by an interpolation as having some level of pollutant even if the values are low and below the health standards.

C4. This in combination with a universal kriging approach that permits the inclusion of covariates to possibly help discern areas of differing distributions would seem (to me) to be the more scientiïĭňĄc approach.

R4. We believe the Reviewers meant to refer to Co-Kriging, not Universal Kriging saying that it would "permits the inclusion of covariates to possibly help discern areas of differing distributions". Land use data which are available for the entire Coastal aquifer might have been used for such purpose. However, inaccuracies and shortcomings encountered in the land use data (and we carried out much work with these data) may result in very bad interpolations. Moreover, this still will not solve the issue discussed above of the desire to interpolate zero values to areas clean of pollutant.

C5. Rather the authors decide to promote an overly simplistic approach.

HESSD

10, C5780–C5785, 2013

R5. Occam's razor principle, extended from hypothesis testing to the use of models, asserts that among competing models the simplest one should be used. Indeed it should not be overly simple for the task at hand. We claim that in our case our interpolation model is sufficient and serves the task better. Beyond general rejection, the Reviewer did not explain why our approach in this case is "overly simplistic".

C6. This leads to an approach and results that are not very scientiïňĄcally defensible.

R6. We presented results of some computations. We carried out an exhaustive assessment of their worth. A valid comment would be that the results of this assessment seem insufficient for the purpose for which our interpolation was designed and applied, and, that other methods could have achieved better ones. We do not think this is true. Nor did the Reviewer claimed that, beyond a general rejection of the IDW method.

C7. The quality and quantity of the data are involved in all inferential steps of geostatistical modeling, as they would be in any statistical approach, and this crucial concept is missing in IDW.

R7. This comment is not clear.

C8. The authors also seem to imply that geostatistical methods are not well suited for sparsely monitored (sampled) scenarios.

R8. No such claim was made in the manuscript. We said that "In sparsely monitored areas, and most aquifer areas can be considered as such, this results in grid points at which interpolation is not carried out." This was referring to the inability to interpolate pollutant concentrations to the whole aquifer area with adequate accuracy level. This claim is true for any method used.

C9. Process like IDW that are far more amenable to automation, the pushing of a few buttons after making subjective choices on the size and shape of areas of inclusion zones, are not easily able to tell the difference between sparsely and more intense monitored scenarios.

HESSD

10, C5780–C5785, 2013

R9. Pushing a few buttons without any understanding of the algorithm and software that implement it is indeed a very sad practice that the authors rue and frequently complain about. Let it be clear that every piece of code used for preparing our manuscript was written by the authors. It's part of a large software package which we authored and use. Indeed, IDW was initially added to this package as a simple benchmark method to compare to others, some of which the Reviewer suggests. However, in this case we found IDW the most suitable. And, automation was indeed an important consideration. We deal with dozens of pollutants and are interested in carrying out statistical analyses of their spatial features. We thus need a process that can be automated based on a few defensible general decisions (e.g., the distance which pollutants may move during certain period of time, which dictated our inclusion zones dimensions). To compare to Kriging for example, creating for each pollutant the empirical semi-variogram, selecting nugget and sill and other variables, and fitting a theoretical function to the semi-variogram is not practical in this case. Nor is it a good scientific practice given the impact that those many subjective decision may have on the final joint analyses.

C10. I have sat on several panels and reviewed enough articles seeing a similar concept, this idea of making IDW better somehow or even making it statistical. My answer is always the same that this has already been done before and the results of that work are the current and developing field of geostatistics.

R10. We have not found in the literature a study of a whole aquifer's water quality using routine monitoring data of dozens of pollutants. The Reviewer has not pointed out any such work either. We have noticed a few interesting issues regarding using effectively and efficiently such data set in our circumstances. Given the lack of literature on that specific subject, we thought that practitioners in the field may benefit from our understanding, whatever the interpolation method which is used. Even if it's not the best, we believe that IDW is not wrong to use in our case and it demonstrates a few points which have not (to our best knowledge) appeared in this context in the literature.

HESSD

10, C5780–C5785, 2013

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 9363, 2013.