

PAPER • OPEN ACCESS

Comparison between two interpolation methods: Kriging and EPH

To cite this article: G Berton 2018 *J. Phys.: Conf. Ser.* **1141** 012130

View the [article online](#) for updates and enhancements.

You may also like

- [Testing PSF Interpolation in Weak Lensing with Real Data](#)
Tianhuan Lu, Jun Zhang, Fuyu Dong et al.
- [Optimal sensor configuration for flexible structures with multi-dimensional mode shapes](#)
Minwoo Chang and Shamim N Pakzad
- [Resilient algorithms for reconstructing and simulating gappy flow fields in CFD](#)
Seungjoon Lee, Ioannis G Kevrekidis and George Em Karniadakis



DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Comparison between two interpolation methods: Kriging and EPH

G Berton¹

Société de Calcul Mathématique

111 Faubourg Saint Honoré, 75008 Paris, France

E-mail: gottfried.berton@scmsa.eu

Abstract. The aim of this study is to compare two methods of interpolation, namely Kriging (a standard algorithm), mainly used in geostatistics, and the Experimental Probabilistic Hypersurface (developed by SCM SA). We study several technical points, such as their ability to take uncertainties into account, to return an uncertainty on the interpolation, the quality of the numerical procedures, etc. The Experimental Probabilistic Hypersurface (EPH) is a minimal information model, which only uses the existing data and makes as less artificial hypothesis on the data as possible. The Kriging, on the contrary, relies on an estimation of the variability of the data using a variogram.

1. Introduction

The objective of this paper is to compare two interpolation methods: the Kriging, which is a standard algorithm, and the Experimental Probabilistic Hypersurface, a method developed by the company SCM SA (see the book [PIT]), which is based on a minimal information principle, that is making as less artificial hypothesis on the data as possible. In a previous paper [ICAPP], it was stated that EPH gives better results when the information is poor. Here, we continue the study and present typical situations where Kriging or EPH fail in the interpolation. We also study several technical points, such as their ability to take uncertainties into account, to return an uncertainty on the interpolation and the quality of numerical procedures. We would like to thank Yann Richet from IRSN (French Institute for Radioprotection and Nuclear Safety) for bringing such questions to our attention.

2. Basic information about Kriging and EPH

The Kriging method includes the variability of the data in order to estimate the values at unknown locations. Each observation is interpreted as the outcome realization of a random variable. At each point of the domain is associated a random variable. These random variables are part of a stochastic process. The Kriging algorithm has two steps:

- The estimation of a dependence model between the random variables (variogram). This model is a function of the distance between the variables and not of their position. Indeed, two random variables spatially close will take close values; two random variables situated far from each other take very different values, that is, they are independent. One fundamental

¹ To whom any correspondence should be addressed.



parameter of this model is the range parameter: it is the distance from which two random variables are considered as totally independent;

- The construction for each point to reconstruct is a linear combination of the data. The weights are based on a spatial dependence model between the random variables, mentioned above. The kriging weights are chosen to minimize the variance: for each trajectory of the stochastic process modelled, the linear combination of the values at data locations should be as close as possible to the value at the point to reconstruct, especially for the trajectory that generated the data.

Kriging makes several model hypothesis: the variability depends only on the distance between the data, the process to reconstruct is Gaussian and second-order stationary, there are linear correlations between the random variables. On the contrary, the EPH does not introduce arbitrary information or assumptions about the data. This method allows to propagate the existing information toward unknown locations. It can be used to reconstruct missing data or to make predictions. The propagation of information relies on a general principle of maximal entropy (or minimal information) which is itself an increasing function of the distance to the measurement point.

Each of the measurements gives its own contribution to the final result, under the form of a discrete probability density having maximal variance for the fixed entropy (Gaussian distribution, see [PIT]). Such a density takes the form of a Dirac function at the measurement point location (the value is known precisely), and becomes less and less concentrated when moving away from it. The variance parameter is calculated so as to maintain the information minimal at every point. At the end of the process, the individual laws are recombined in order to get a single one depending on the distance between the target-point and each measurement (inverse distance weighting).

3. Situation with strong variability

When the data vary strongly, the EPH gives a very irregular interpolation. When interpolating an exponentially increasing function, the left part is overestimated because the high values at right have an influence upon the left part of the curve as shown in Figure 1. This can seem strange at first sight, but the reader should bear in mind that the EPH makes no assumption on the variability of the data: the weights of the EPH are the same whatever the variability of the function. The high variability of the function is external information that should be added in the model. On the contrary, the Kriging analyses the variability of the data, and includes it in the interpolation. This is the main conceptual difference between EPH and Kriging.

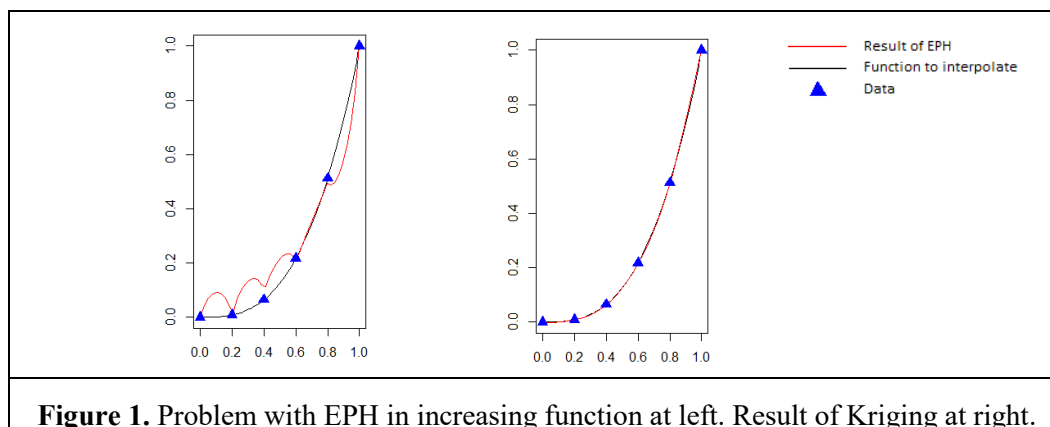
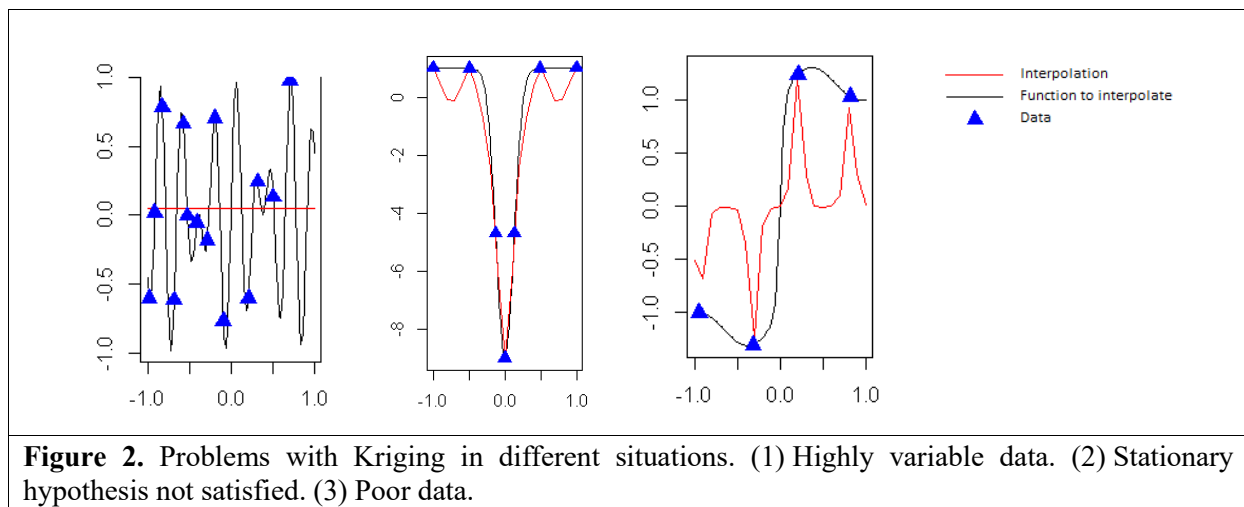


Figure 1. Problem with EPH in increasing function at left. Result of Kriging at right.

In situations of strong variability, the Kriging can perform badly as well. If the data to reconstruct is an oscillating function with sharp peaks of different amplitude, the Kriging can't establish a proper

dependence model between random variables; Maximum Likelihood Estimation fails in the R software (DiceKriging library). The resulting interpolation is then a constant surface which is not correct, see Figure 2.1. The EPH adapts very well to irregular oscillations because it works without the knowledge of the variation of the data.

The Kriging algorithm states that the correlation between two variables solely depends on distance and not on location. This assumption is seldom satisfied in reality, as the example Figure 2.2 shows, where in the middle there is a high variability, whereas at the border, there is no variability. The EPH does not rely on such artificial assumptions.



4. Situation with poor information

In a previous paper [ICAPP], it was clearly exposed that, when working with few data, the EPH gives better results than Kriging. The reason why Kriging fails is the following: when the information is poor, Kriging knows little about the variability of the data. The dependence between the random variables cannot be correctly estimated and most of the time, the range parameter is underestimated, which means that all the variables are assumed to be independent whatever the distance between them is, resulting in a constant interpolation as shown on Figure 2.3. The EPH does not suffer from this problem, since it does not require estimating a parameter.

5. Cluster of measures

The EPH is very sensitive to clusters of data. Each point has a weight depending on its distance to the point to reconstruct, regardless of its position. Hence, all the points of the group have the same weight, and the weight of the group is then too high. For the Kriging algorithm, there is no such problem, it gives the same weight to an isolated point or to a group of points if they are at the same distance. However too close data can lead to numerical problems for Kriging as explained below.

6. Numerical stability of the methods

Kriging requires the inversion of a covariance matrix to calculate the weights of the interpolation. If the conditioning of the matrix is bad, then a small imprecision on the value of the parameters can lead to a completely different calculation of the weights. Bad conditioning occurs (i) when the data are very dense, or two data are very close, as we see in Figure 3.1, (ii) when the range parameter of spatial dependence overestimated, the result of this problem is shown in Figure 3.2, where there was numerical instability due to accumulations of truncation errors.

When using a bad conditioned matrix, we face two problems: in certain cases, the covariance matrix cannot be inverted, and the software R returns an error. The second problem is that a slight change on the location of the data and points to reconstruct can lead to a very different weight and very different interpolation as shown on Figure 3.3: the location of the points were slightly changed (0.03) and the result is very different. The EPH does not present such difficulties because it does not require inverting a matrix.

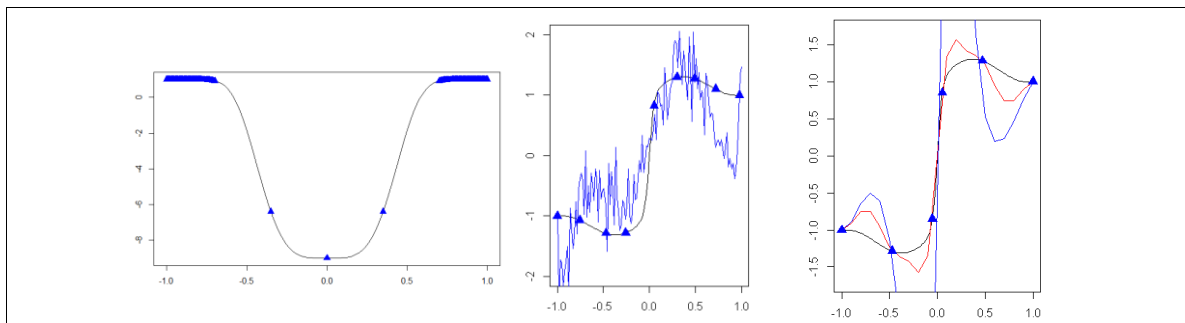


Figure 3. Quality of the numerical procedures in Kriging. (1) Non-invertible matrix due to high density of data. (2) Range parameter overestimated (erroneous result of kriging in blue). (3) Results of kriging before (blue) and after (red) a slight change in location of the data.

7. Taking uncertainties into account

There are different uncertainties to take into account: the location of the data, the position at which the interpolation is made, the values of the data. The EPH is by nature probabilistic and the result returned at each point is a probability law. It is then easy to take the uncertainties into account in the model. The user first sets a number of runs. For each run, one draws at random a value for each parameter supposed to be known with an uncertainty. Then the EPH is computed for these values. At the end, the laws returned at each run are combined according their probability. The law obtained is less concentrated, and the reconstructed surface is smoother when taking into account the uncertainties.

Kriging is based on the use of a variogram to model the dependence between random variables. To take into account the uncertainties on location of the measures, one has to draw at random the positions that can take the data and construct a variogram for each possible repartition of the data. The variograms are averaged, and used to compute the covariance matrix. Taking into account the uncertainties upon the values is more complicated: the usual method is to add a constant to the variogram, but this does not work when the uncertainty is the same over the entire domain.

8. Uncertainty on the interpolation

The result returned by the EPH at an estimation point is a probability law and does not have a Gaussian shape, unlike Kriging. It is possible to calculate not only the expectation but also the median, most probable value, confidence interval, etc. The result of Kriging is deterministic, it returns a Kriging variance; which only reflects the density of measures around, and not their values: the more data there is around the smaller the uncertainty. From this variance, one can construct a confidence interval by assuming that the data are Gaussian: the outcome estimate has always a Gaussian form.

9. Conclusion

A great difference between Kriging and the EPH has been exposed: Kriging depends on the variation of the data to make its interpolation, whereas the EPH uses only the data and does not try to invent other information from them. When information varies strongly, the EPH gives irregular interpolation and Kriging's result is correct. When the data varies irregularly, Kriging cannot get any information on the dependence of the data and fails, the EPH adapts very well to it. When information is poor,

Kriging cannot know about variation. From a general point of view, Kriging perform better when lot of data are available and the dependency between variable is rather simple; EPH will be advantageous in the case of poorly studied phenomena. EPH uses nothing else but the existing data so, if external data are available, it is not the best model to handle them.

10. References

[PIT] Olga Zeydina et Bernard Beuzamy : Probabilistic Information Transfer. Ouvrage édité et commercialisé par la Société de Calcul Mathématique SA. ISBN: 978-2-9521458-6-2, ISSN : 1767-1175. Relié, 208 pages, mai 2013.

[ICAPP] F. Godan (SCM), O. Zeydina (SCM), Y. Richet (IRSN), B. Beuzamy (SCM) : Reactor Safety and Incomplete Information: Comparison of Extrapolation Methods for the Extension of Computational Codes. Proceedings of ICAPP 2015 Nice, France, May 3-6, 2015, Paper 15377.