Société de Calcul Mathématique, S. A.
Outils d'aide à la décision
depuis 1995

Comparison between two interpolation methods: Kriging and EPH

Gottfried Berton
Société de Calcul Mathématique SA
111 Faubourg Saint Honoré
75008 Paris France

September 2018

Abstract

The aim of this study is to compare two methods of interpolation, namely Kriging, a standard algorithm, 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 (see [PIT]). The Kriging, on the contrary, relies on an estimation of the variability of the data using a variogram and makes several assumptions upon the nature of this variability.

The EPH has the following advantages:

- The assumptions made by the Kriging are rarely satisfied, namely that the dependence between the variables is only a function of the distance between them, the process to reconstruct is Gaussian, there are linear correlations between the variables, etc.

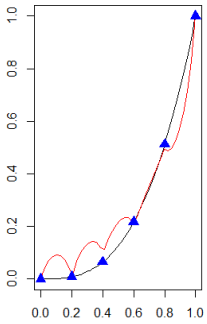
- The variability of the data is not always known, especially when there is poor information. The EPH will give better results in this case.

- The EPH returns an uncertainty under the form of a probability law. Kriging returns a deterministic result under the form of a Kriging variance. Moreover, taking into account the uncertainties upon the values of the measures is easier with EPH.

- The EPH is simpler to carry on than Kriging. The numerical procedures in kriging are more complex and need fitting the parameters of a mathematical model and inverting a matrix. Sometimes this matrix is bad conditioned which gives a bad interpolation, or even no interpolation at all. It can also fail when there is an extreme value in the data.

On the contrary, the EPH has several drawbacks:

- When the data are clustered, EPH gives too much importance to this group of data.
- When data vary strongly, EPH gives a strange result like below:



The concept of the EPH is to add the least possible arbitrary information. In certain specific situations, it can give bad result but this is because there is external information that should be incorporated into the model.

******

I. Introduction

A. The need for a honest approach

After the crash of the Shuttle Challenger in 1986, it appeared that it was due to erosion and failure of the O-ring seals in the right Solid Rocket Booster (SRB). The seals were weakened by the very cold temperature at the time of the liftoff. Richard. P. Feynman, who was a member of the Roger Commission in charge of the investigation of this disaster, wrote the following comment in the Annex of the official report [FEYNMAN]:

"There was no way, without full understanding, that one could have confidence that conditions the next time might not produce erosion three times more severe than the time before. Nevertheless, officials fooled themselves into thinking they had such understanding and confidence, in spite of the peculiar variations from case to case. A mathematical model was made to calculate erosion. This was a model based not on physical understanding but on empirical curve fitting. To be more detailed, it was supposed a stream of hot gas impinged on the O-ring material, and the heat was determined at the point of stagnation (so far, with reasonable physical, thermodynamic laws). But to determine how much rubber eroded it was assumed this depended only on this heat by a formula suggested by data on a similar material. A logarithmic plot suggested a straight line, so it was supposed that the erosion varied as the .58 power of the heat, the .58 being determined by a nearest fit. At any rate, adjusting some other numbers, it was determined that the model agreed with the erosion (to depth of one-third the radius of the ring). There is nothing much so wrong with this as

believing the answer! Uncertainties appear everywhere. How strong the gas stream might be was unpredictable, it depended on holes formed in the putty. Blow-by showed that the ring might fail even though not, or only partially eroded through. *The empirical formula was known to be uncertain, for it did not go directly through the very data points by which it was determined. There were a cloud of points some twice above, and some twice below the fitted curve, so erosions twice predicted were reasonable from that cause alone. Similar uncertainties surrounded the other constants in the formula, etc., etc."*

This is an example of bad-decision making caused by the use of a wrong oversimplified mathematical model. The NASA established a deterministic relationship between the heat of the gas and the erosion to predict the expected erosion in the flight 51-L (the one that failed). The model was not realistic for different reasons:

- It assumed that the erosion depends only on gas temperature. It didn't take into account any variability on operating conditions (temperature of the O-ring itself, etc.).

- It assumed deterministic relationship between erosion and heat of the gas: $erosion = heat^{0.58}$, but the data shown that for one value of heat there are many possible values of erosion.

- It didn't return any uncertainty on the erosion.

The result of this interpolation was trusted by the officials, and used as an argument to accept for flights, seals that had shown important erosion in previous flights. By using a precise value for taking a decision, they ignored the uncertainty on the erosion, as well as the variability due to the initial temperature of the O-ring. This very cold temperature has caused the failure of the primary and secondary O-rings and the crash of the Shuttle.

"Instead of being very concerned that variations of poorly understood conditions might reasonably create a deeper erosion this time, it was asserted, there was "a safety factor of three."

At the end of the day, NASA managers looked for evidence to support mission success rather than evidence indicating possible mission failure. They ignored the contradictory opinion issued by the constructor of the O-ring and the engineers of the NASA and decided at any rate to validate the flight.

This reminds us that having the most honest possible approach is fundamental in science. Nonetheless, thirty years later, this way of working and these methods are still actively used. We are going to present in this article a model that aims at introducing the least arbitrary information and allows to calculate an uncertainty on the result, which is always required in any safety demonstration, including aerospace safety study.

B. Objectives of the article

The objective of this article is to compare two probabilistic 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 discuss 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.

II. Basic information about Kriging and EPH

A. Kriging

1. General presentation

Kriging is a probabilistic interpolation method. This has been developed initially to study the mining deposits in geostatistics. It is not possible to drill the ore field on the entire area, one measure the concentration of ore only from several points of the domain. Kriging allows to estimate the concentration on the entire domain using these observation points.

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 dependence between the variables is a function of the distance between the variables and not on 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.

The Kriging algorithm has two steps:

– The estimation of a dependence model between the random variables;

– 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.

First step: modelling the process

The dependence is modelled by a variogram. This represent the mean difference between the random variables separated from a distance h. The higher the distance, the larger the difference.

In practice, one first constructs an experimental variogram from the observations. Let us assume that one have 25 observations drawn randomly on a domain with dimensions 50 by 50. To each couple of observations $(y_i, y_j),\ i<j,\ (i,j)\in[1,25]$, a point that represents the difference $\left(y_i - y_j\right)^2$ with respect to the distance between the observations is created (black points on

Figure 3). This scatter plot allows to calculate the mean gap in function of the distance (red points).

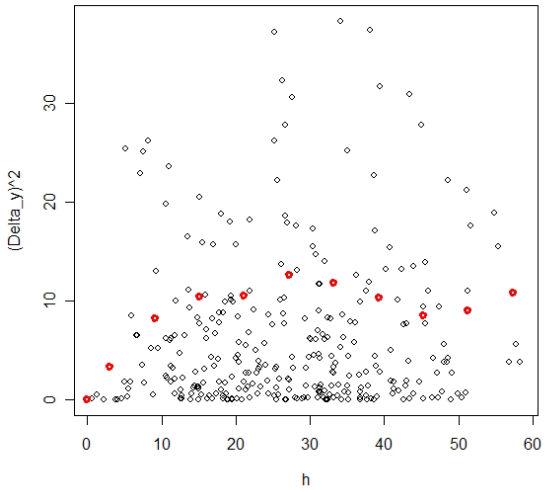


Figure 1: gap between two points in function of the distance

Then, a model is adjusted on this experimental variogram. This allows to obtain an estimation of the mean gap for any distance h.

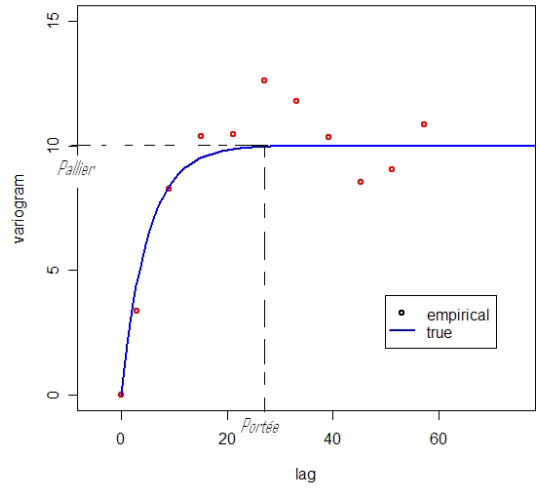


Figure 2: fitting a model on the experimental variogram

This model has an important parameter: the range; it represents a distance. The random variables separated from more than this distance have a minimal dependency. In this example, from a distance of 25, the random variables are considered independent.

Second step: construction of the estimate

The dependency model estimated is used for defining an estimator that is written as a linear combination of the observations. The kriging weights are chosen to minimize the variance of

the error between the random variable at the point to reconstruct and the linear combination of the random variables at each observation. 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.

Let us denote X_0 the location where one want to make the interpolation, and $X_i, i=1,...,N$ the location of the data. Let Y_i be the random variables of the stochastic process associated to locations X_i. The kriging consists in finding the best linear unbiased estimator $Y^* = \sum_{i=1}^{N} \lambda_i Y_i$, such that the expectation of the error is null $E[\varepsilon] = E[Y^* - Y_0]$ and the variance of the error $Var(\varepsilon) = Var(Y^* - Y_0)$ is minimal. A linear estimator gives the minimal variance only if the process modelled is Gaussian: this is an assumption of the kriging method.

The weights minimizing the variance is unique and is equal to $\lambda = \mathrm{K}^{-1}\,\gamma(X_0)$, where

$$\gamma(X_0) = \begin{pmatrix} k(Y_0, Y_1) \\ k(Y_0, Y_2) \\ \vdots \\ k(Y_0, Y_{n-1}) \\ k(Y_0, Y_n) \end{pmatrix}.$$

And,

$$K(p) = \begin{pmatrix} k(Y_1, Y_1) & k(Y_1, Y_2) & \cdots & k(Y_1, Y_n) \\ k(Y_1, Y_2) & k(Y_2, Y_2) & \dots & k(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(Y_1, Y_n) & k(Y_2, Y_n) & \dots & k(Y_n, Y_n) \end{pmatrix}.$$

The weights of this linear combination are different for each point to reconstruct and depend:

- On the distance between this point and each observation;
- On the distance between the observations themselves.

2. Model hypothesis

Kriging makes several model hypothesis:

- the variability depends only on the distance between the data;
- the process to reconstruct is Gaussian;
- the process is second-order stationary;
- there are linear correlations between the random variables.

If the hypothesis are not satisfied, the algorithm may fail. Let us focus on the second-order stationary hypothesis. The kriging assumes that the spatial dependency depends only on the distance between the variables and not on their position. For any couple of points separated from a distance h the covariance between these couple of random variables are the same. This hypothesis is called "stationarity of second order". The main covariance models are:

- Exponential $f(h)=\sigma^2 e^{-\frac{h}{\rho}}$.
- Gaussian $f(h)=\sigma^2 e^{\left(-\frac{h}{\rho}\right)^2}$.
- Matérn: $f(h)=\frac{2^{1-\upsilon}}{\Gamma(\upsilon)}\left(\frac{\sqrt{2}\upsilon h}{\rho}\right)^{\upsilon} K_{\upsilon}\left(\frac{\sqrt{2}\upsilon h}{\rho}\right)$.

Where K_{υ} is a specific function called Bessel function, σ^2 is the variance, common to each random variable to be fixed, h is the distance separating the two variables, and ρ is the range parameter to fix either:

- Manually, after analyzing the variability of the data;
- Using an algorithm: least squares or maximum likelihood.

The estimation of the parameter depends highly on the data. We present below situations where the bad estimation of this parameters leads to bad interpolation. It is important to have a critical point of view upon the estimations obtained and setting manually the parameters in function of the data, which makes the use of kriging difficult.

Each of these functions models a specific dependency. For example, in the exponential model, when the data are sufficiently close, the covariance decreases fast when the distance increases. This increasing is smoother in the Gaussian model. The Matérn model is a compromise because it allows to set this slope manually. In this article we use this model.

This hypothesis of second order stationarity is rarely verified. In the example of a ore field, it assumes an homogeneity of the variability of the concentration of ore on the domain, which is not true in practice. There can be two regions: one having a constant concentration, and the other having a very irregular concentration. We will see on an example that this is impossible to obtain a good interpolation in this situation.

We wish to reconstruct the function $f(x)=1-10e^{-x^2/0.03}$ in black on the plot below. We reconstruct 20 points from the 7 measures represented in blue triangles below. The interpolation is in red below:

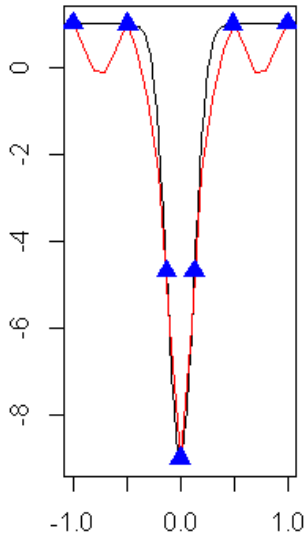


Figure 3: bad estimation of the range parameter using R leading to strange interpolation (red)

In this simulation, the software R estimates the range parameter to 0.13. This coefficient is small, because the algorithm has detected that the data close the one from the other vary strongly in the middle of the domain. However, this coefficient is too small to reproduce properly the constant concentration at the sides.

It is impossible to reconstruct a point situated on the borders when the function is constant on the borders and variable in the middle.

B. EPH

The EPH does not introduce arbitrary information or assumptions about the data, unlike kriging. 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. The EPH model requires two input parameters:

- bounds on each dimension. For example, if the historical data are temporal, one has to set a time min and max on which to perform the reconstruction;

- bounds on the outcome range and discretisation step; the resulting estimates has the form of a discrete probability law on the defined range.

The bounds may come from expert knowledge, physical limits or be defined by a user.

We have N observation points, denoted by A_n, $n=1,...,N$. Let C_n be the outcome value of the n^{th} measure. Let X be the point where we want to obtain an estimate. Let $d_n = d(A_n, X)$ be the distance between the point to reconstruct and the n^{th} measure.

Let t_j be the discretisation of the result range with step τ in υ intervals, and λ be a parameter related to the entropy which is calculated so as to maintain the information minimal at every point (see the book [PIT]).

An entropy is a quantity that is associated to a distribution, and is can be seen roughly as a variance that depends only on the probabilities p_j, and not on the values that the variable can take. The entropy increases linearly with the distance [PIT].

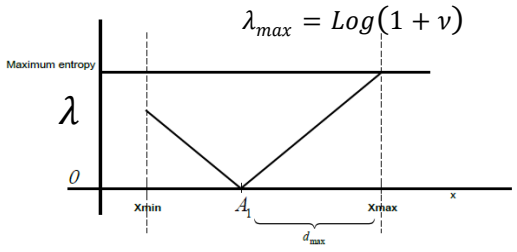


Figure 4: entropy in function of the distance to the point to reconstruct

Each of the measurement n gives its own contribution to the final result, written under the following form:

$$p_{n,j}(X) = \frac{\tau}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t_j - C_n)^2}{2\sigma^2} \right),$$

where $\sigma = \frac{\tau e^{\lambda d_n}}{\sqrt{2\pi e}}$.

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 Gaussian shape comes from the fact that the distribution having maximal variance for a fixed entropy is the Gaussian [PIT].

At the end of the process the individual laws are recombined in order to get a single one depending on the distance of the target-point from each measurement:

$$p_j(X) = \gamma_1 p_{1,j}(X) + ... + \gamma_N p_{N,j}(X),$$

where $\gamma_n = \frac{d_n^{-1}}{\sum_{i=1}^{N} d_i^{-1}}$, $n = 1, ..., N$. This ratio allows to introduce the least information possible (see [PIT] part III p.167).

From this probability law, one can extract a confidence interval. The lower bound of this confidence interval is the quantile 5% of the probability law. The upper bound is the quantile 95%.

III. Situation with strong variability

A. Kriging

In situations of strong variability, the Kriging performs badly. 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, as shown below:

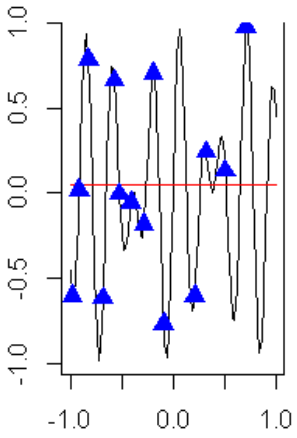


Figure 5: failure of kriging in situation of high variability

The EPH adapts very well to irregular oscillations because it works without the knowledge of the variation of the data. In addition there is no parameter to estimate, which excludes the kind of failure presented above.

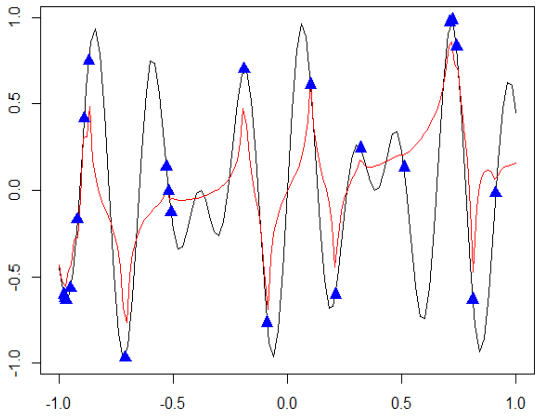


Figure 6: result of EPH in situation of high variability

Kriging can also give a considerably overestimated interpolation when data are variable:

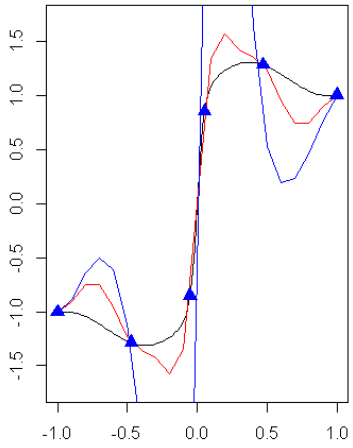


Figure 7: Result of kriging after (red) and before (blue) slightly changing the location of the measures

B. EPH

When the data vary strongly, the EPH can also give 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 the figure below:

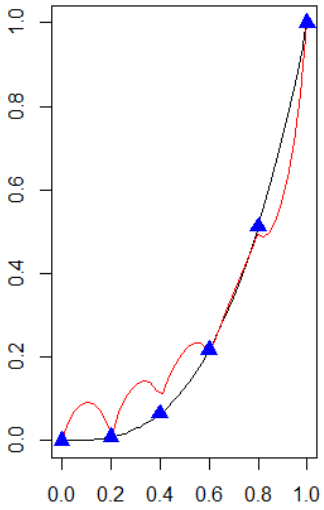


Figure 8: result of EPH (red) when the function is increasing (black curve)

This can seem strange at first sight, but the reader should keep 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 might be. In the EPH, the maximal entropy is reached at the point that is the farthest possible from the measure.

Assuming that the measures sufficiently far add no information is a model hypothesis that is forbidden in EPH. However, if for some reasons, we have got the knowledge that the measures situated at a certain distance bring absolutely no information to the point to reconstruct, that

is they are independent, one could imagine a different model from EPH, where the slope of the entropy would be set manually, like on the figure below:

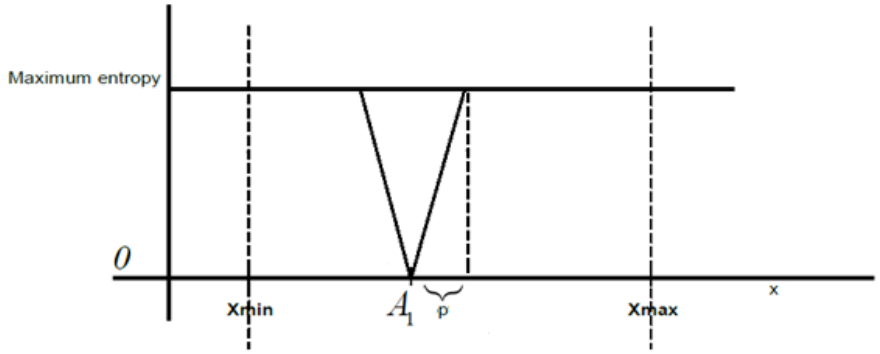


Figure 9: modified growth of entropy with distance

In this model, any measure situated at a distance of more than p brings no information at all. Again, making this hypothesis is forbidden in EPH since we look at making the least possible a priori on the phenomena: this is a minimal information method.

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. The interpolation for kriging is correct in this example:

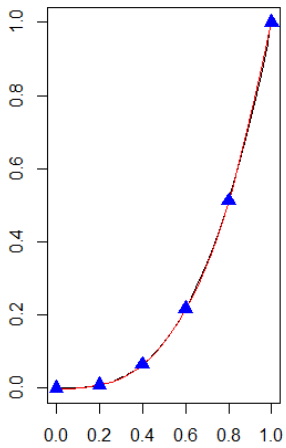


Figure 10: result of kriging when the function is increasing

This is the main conceptual difference between EPH and Kriging. Moreover, the kriging gives the user the responsibility to study the variability of the data and allows the user to add this knowledge in the interpolation.

IV. Situation with poor information

A. Kriging

In a previous paper [ICAPP], it was clearly exposed that, when working with few data, the EPH gives better results than Kriging. When the information is poor, Kriging knows little about the variability of the data and fails estimating the dependence model. 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 below when reconstructing the function $f(x) = 0.2 \times \left(\sin 5x + \sin \sqrt{3x} + \tanh 20x \right)$.

We have reconstructed 20 points on $[-1,1]$ from 4 measures $(-0.94, -0.314)$ and $(0.21, 0.82)$ that are not exactly symmetric.

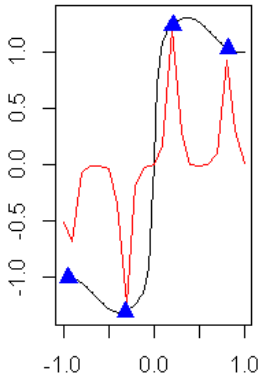


Figure 11: kriging in the case where the points are not symmetric

In a second simulation, we choose two couples of points in a symmetrical disposition $(-0.94, -0.314)$ and $(0.314, 0.94)$. The result obtained is totally different, which is strange:

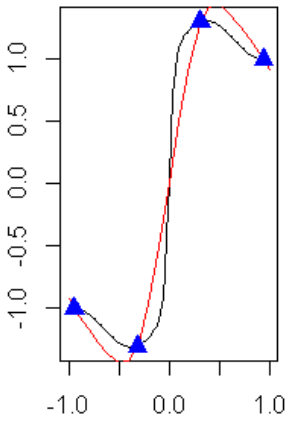


Figure 12: kriging in the case where the points are symmetric

The first simulation gives bad results because the range parameters have been considerably underestimated. The repartition of the data have an impact on this estimation. In the second simulation, the variables far the ones from the others are considered as dependent, because the range is 0.43. In the second simulation, the range is almost null, which means that the data, even the ones close to each other, have no influence between them. In this case, the curve reconstructed is almost constant. The formula of the estimate is $Y_0 = m + K^{-1}k(X_0)(Y-m)$. The covariance between the point to reconstruct and the observation will be null when the distance between them overcome the range, and we will have $k(X_0)=0$ and eventually $Y_0 = m$.

B. EPH

The EPH does not suffer from this problem, since it does not require estimating a parameter. Let us take the same example as above for kriging, the result is more satisfactory for EPH than for kriging, which returned a constant surface.

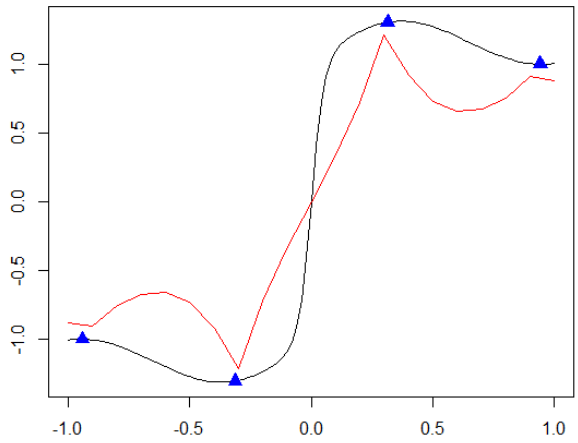


Figure 13: result of EPH with four measures

V. 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 as 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.

Let us illustrate this problem on a simple example. There are 9 measures evenly distributed on the domain. We add a cluster of measures at $(0,0)$ coordinates.

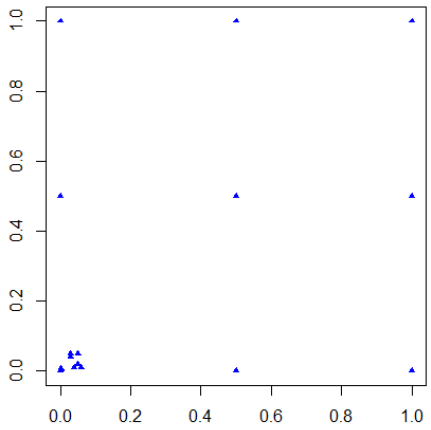


Figure 14: cluster of measures at $(0,0)$

The surface one should obtain is displayed below:

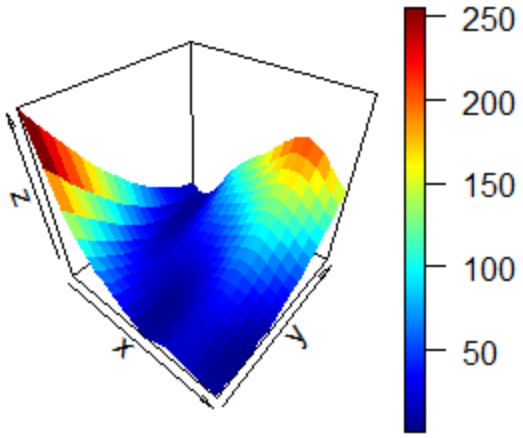


Figure 15: Branin-Hoo surface to be reconstructed

The result obtained with the cluster (at left) is different from the result obtained without the cluster (at right).

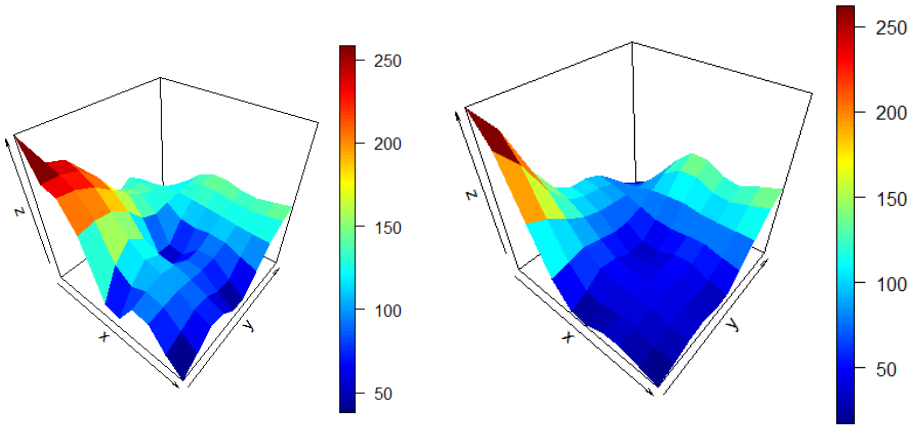


Figure 16: result obtained with the cluster (left), and without the cluster (right)

The surface reconstructed by EPH is very different when there is redundancy in information. The cluster of data in $(0,0)$ has an influence on a large part of the domain. This is not a mistake: there is more information on this corner. The EPH takes into account the distance to the measures, and not the position of these measures: the fact that they are clustered or not doesn't matter. On this example, kriging will give the same interpolation whatever the data are clustered or not.

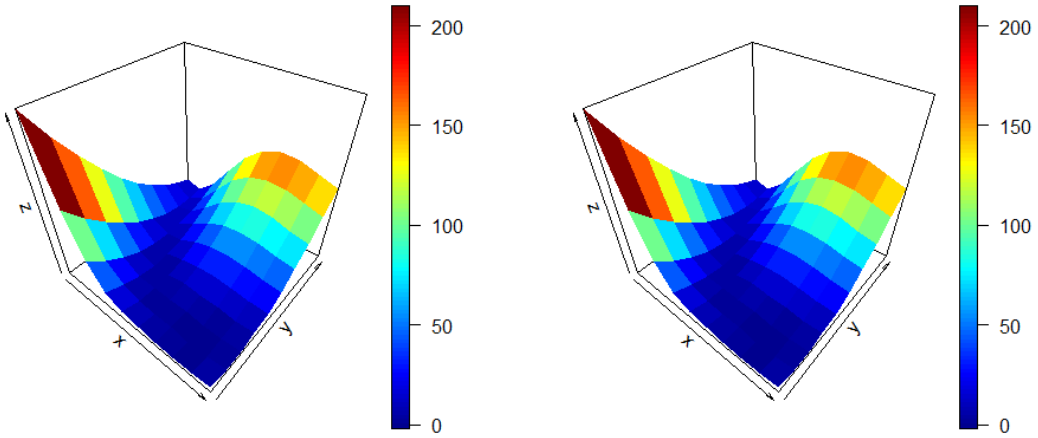


Figure 17: result of the kriging with (left) or without (right) the cluster (the result is the same)

VI. Numerical stability of the methods

In various situations, kriging gives an erroneous result. This can be due to a wrong estimation of the range coefficient. In addition, the matrix used in kriging can be bad conditioned and may cause numerical instabilities. We examine the situations that cause the instabilities and see if EPH suffer from these problems in these specific situations.

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. In certain cases, the covariance matrix cannot be inverted, and the software R returns an error.

The conditioning number allows to measure the dependence of the solution of a numerical system to the data. Let $Ax = b$ a linear system. If A is bad conditioned, then a small variation on b leads to a large variation on the solution x. The condition number is defined by the formula:

$$cond(A) = \|A\|_p \left\|A^{-1}\right\|_p$$

In the case of kriging, the system to be solved is $K\lambda = k(X)$, where K is the covariance matrix, and $k(X)$ the covariance between the observations and the point to reconstruct X.

Bad conditioning occurs:

– When the data are very dense, or two data are very close to each other, as shown below. In this case kriging fails and returns no interpolation at all, because the matrix cannot be inverted.

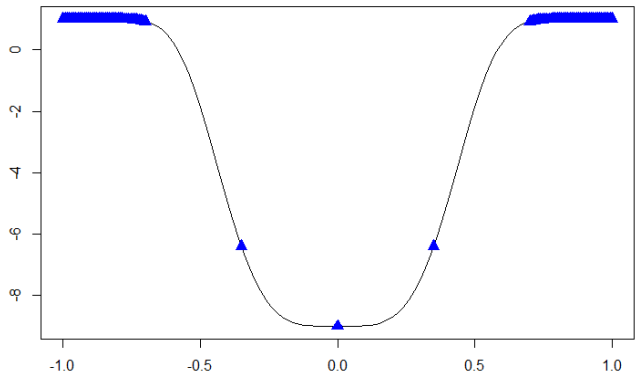


Figure 18: kriging failure in case of high density of data

– When the range parameter of spatial dependence is overestimated. The numerical instability due to accumulations of truncation errors leads to the following interpolation. The conditioning number is 2×10^{17}

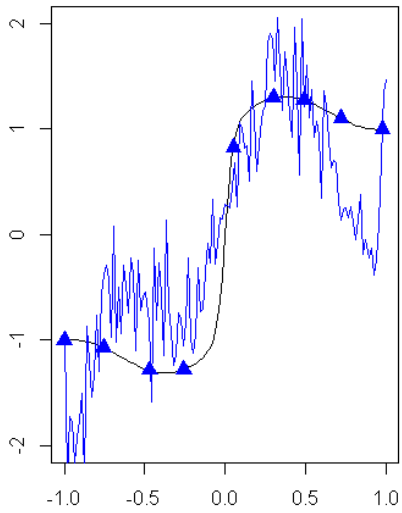


Figure 19: failure of kriging (blue) due to accumulations of truncation errors

In addition, in case of bad conditioned matrix, kriging is not robust to a slight change on the covariance. A small perturbation on the covariance can lead to very different weights and very different interpolation as shown on the example below.

The data are located at $(-1,\ -0.16,\ -0.15,\ 0.97)$. The values of the data are $(-1.005,\ -1.194,\ -1.182,\ 1.000)$. We want to get an estimation at $X = 0.8$.

The matrix is $K = \begin{pmatrix} 1 & 0.69 & 0.68 & 0.13 \\ 0.69 & 1 & 0.99 & 0.51 \\ 0.68 & 0.99 & 1 & 0.52 \\ 0.13 & 0.51 & 0.52 & 1 \end{pmatrix}$. The condition number of this matrix is 2×10^6. The covariance vector is $k(0.8) = \begin{pmatrix} 0.249 \\ 0.830 \\ 0.838 \\ 0.139 \end{pmatrix}$.

Because of the uncertainty on the location of the point to reconstruct, the covariance between the data and the point to reconstruct is known only with uncertainty. We get closer to the fourth point and farther to the three first points, so we modify the covariance vector by adding a vector $\varepsilon = \begin{pmatrix} -0.001 \\ -0.003 \\ -0.007 \\ 0.005 \end{pmatrix}$. The weights change from $(0.14,\ -28.3,\ 28.5,\ 1.02)$ to $(-2.34,\ 242.21,\ -241.06,\ 2.50)$. The value obtained is 0.98 before changing the covariance, and 0.57 after.

The EPH does not have such difficulties because it does not require inverting a matrix.

VII. Ability to take into account the uncertainties

A. EPH

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.

Let us explain this on an example. Let us assume that there is an uncertainty upon the first coordinate of the first measure, that is on the parameter $\xi_1^{(1)}$. Let us assume, for simplicity, that this parameter can only take two values a_1 with probability q_1 and a_2 with probability q_2, and $q_1 + q_2 = 1$.

We first assume that $\xi_1^{(1)} = a_1$ and make the complete construction of the EPH in this case. At a point X, we obtain the law $p_1(j, X)$. Then, we assume $\xi_1^{(1)} = a_2$; at the same point X we obtain the law $p_2(j, X)$. Then, is we take the uncertainty into account, the value given by the EPH at the point X is:

$$p(j, X) = q_1 p_1(j, X) + q_2 p_2(j, X)$$

This procedure extends easily to the case of the uncertainty upon the coordinates of the measure points, the uncertainties upon the value of the measure, and on the position of the test point X. One make the list of the measures with their probabilities, and realize the

construction of the EPH separately in each case. At the end, we construct an "averaged EPH" by combining the different constructions with their respective probabilities.

Simple example

Let us assume that the second point, denoted B, takes only 4 positions, as shown below:

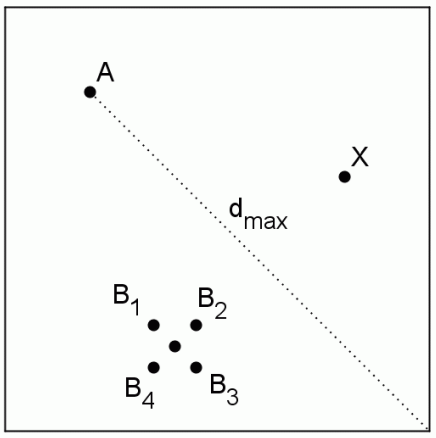


Figure 20: example when B takes 4 positions

The size of the square is 1; the point A has coordinates $(1/5,\ 4/5)$ and point B $(2/5,\ 1/5)$. The points B_i have coordinates $\left(\frac{2\pm 0.25}{5}, \frac{1\pm 0.25}{5}\right)$. We are interested in the point X with coordinates $(4/5,\ 3/5)$.

First of all, we construct the EPH at point X, by considering separately the four situations A,B_1, A,B_2 A,B_3 A,B_4. Then we calculate the mean of the four laws obtained. The dashed curve corresponds to the EPH realized with the point B alone. The solid line correspond to the mean of the curves obtained from the four points B_i. The second curve is slightly less concentrated: this conclusion is a general one; when taking the uncertainties into account the distribution are less concentrated.

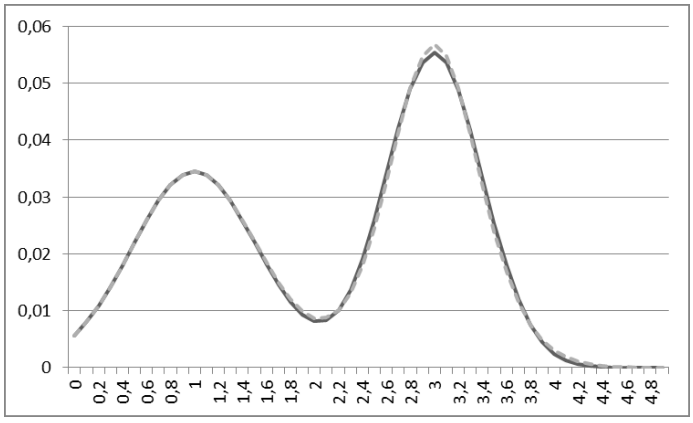


Figure 21: curve obtained when taking into account the uncertainties (solid line)

It is important that the construction is repeated since the very beginning, by taking into account all the possibilities separately.

In practice, there can be an uncertainty upon any data:

- Coordinate of the data;
- Coordinate of the point to reconstruct;
- Values observed for the measure.

A Monte-Carlo method is appropriate in this case, since we are looking for a mean value. We proceed as follows:

First fixing a total number of runs $M = 10^6$:

- For each run $m = 1, ..., M$, 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 averaged with weighting according their probability.

The law obtained is less concentrated, and the reconstructed surface is smoother when taking into account the uncertainties as shown on the example below. We have reconstructed the Branin-Hoo function from 16 measures. The value and position of each measure follow a uniform distribution with support 20 and 0.1 respectively. The number of runs is 200. The figure below represents the true surface:

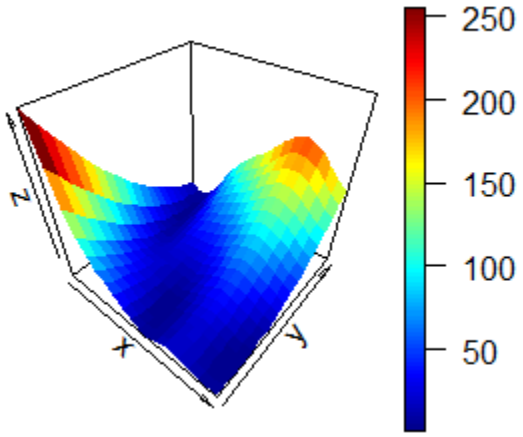


Figure 22: true Branin function to be reconstructed

The figure below at left is without taking into account the uncertainties, and the surface at right is when taking into account the uncertainties. The surface obtained when taking the uncertainties into account at right is smoother.

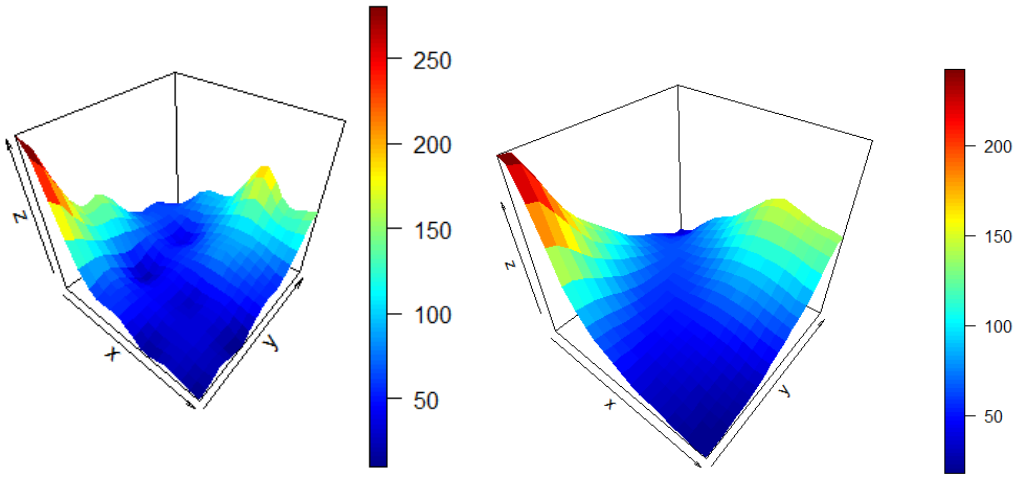


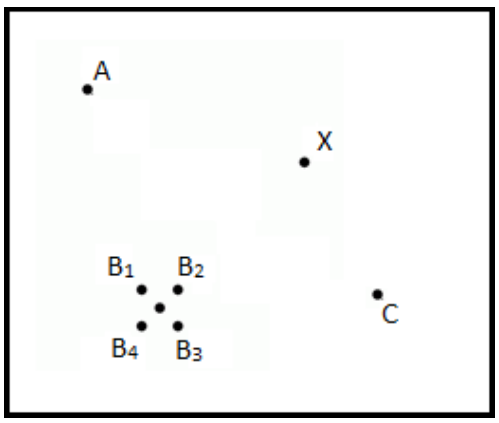
Figure 23: surface obtained without taking into account the uncertainties (left) and taking into account the uncertainties (right)

B. Kriging

1. Uncertainties upon the location of the data

To take into account the uncertainties on the 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.

Let us take again the following example:



We will first calculate the dependence model, that is a function f of the distance. The covariance between A and B will now be equal to:

$$f(A,B) = \frac{f(A,B_1) + f(A,B_2) + f(A,B_3) + f(A,B_4)}{4}$$

We will calculate the new covariance matrix $\tilde{K} = \begin{pmatrix} f(A,A) & f(A,B) & f(A,C) \\ f(B,A) & f(B,B) & f(B,C) \\ f(C,A) & f(C,B) & f(C,C) \end{pmatrix}$ and vector $k = \begin{pmatrix} f(X,A) \\ f(X,B) \\ f(X,C) \end{pmatrix}$.

2. Taking into account the uncertainties upon the values

Taking into account the uncertainties upon the values is more complicated: the common method to take into account the uncertainties is to add a variance upon each random variables, and to assume that this variance is the same for each observation, which is seldom satisfied hypothesis. This is equivalent to adding the variance to the variogram γ so that it satisfies the following equation:

$$\lim_{h \to 0} \gamma(h) = \varepsilon$$

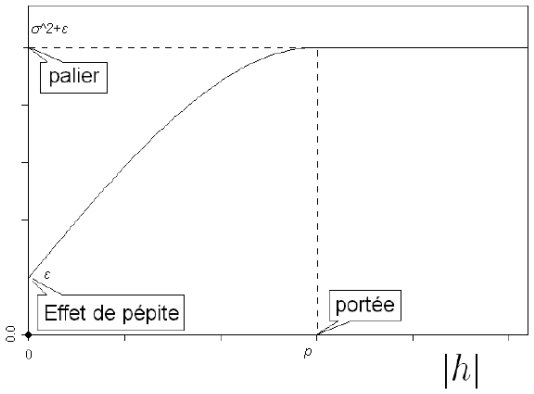


Figure 24: variogram with added variance ε at $h = 0$

Let us consider that that observations are the realization of a stochastic process whose random variables are written as $Y_i + E_i$, where E_i is the random variable that models the measure error associated to observation i, such that $Var(E_i) = \varepsilon$, $\text{cov}(E_i, Y_x) = 0$, and $\text{cov}(E_i, E_j) = 0$. The principle of kriging is to minimize:

$$Var\left(Y_x - \sum_{i=1}^{n} \lambda_i \left(Y_i + E_i \right) \right).$$

The weights that minimize this expression are:

$$\lambda = (K + \varepsilon Id)^{-1} k .$$

Therefore, writing the random variables as $Y_i + E_i$ is the same as adding a variance on the diagonal of the matrix K.

In the simulation below, the following function is reconstructed:

$$f(x) = 0.2 \times \left(\sin 5x + \sin \sqrt{3}x + \tanh 20x \right)$$

The function is represented in black below, and the reconstruction obtained when taking into account the uncertainties is displayed is blue.

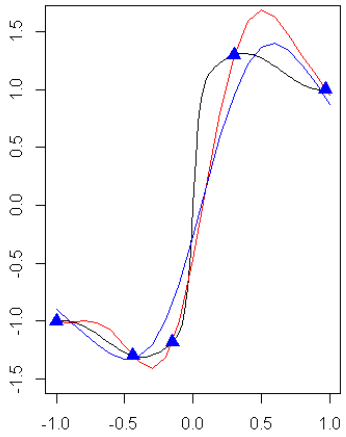


Figure 25: result when taking into account the uncertainties (blue)

The interpolation made when adding a variance is smoother and doesn't necessary pass through the observations.

VIII. Ability to return an uncertainty on the interpolation

A. Kriging

For kriging, the result returned is deterministic, and it will always return a Kriging variance. This variance only reflects the density of measures around, and not their values: the more data there are 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.

B. EPH

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 confidence interval calculated with EPH is often much larger than the one obtained with kriging. This is because EPH doesn't make any hypothesis.

IX. Robustness to extreme values

One aim at reconstructing the Branin-Hoo function introduced above, when adding an extreme value. With this extra point, the kriging produces a constant surface, which result of a failure of the algorithm to estimate a dependence model. One has to modify manually the range parameter to obtain a correct interpolation.

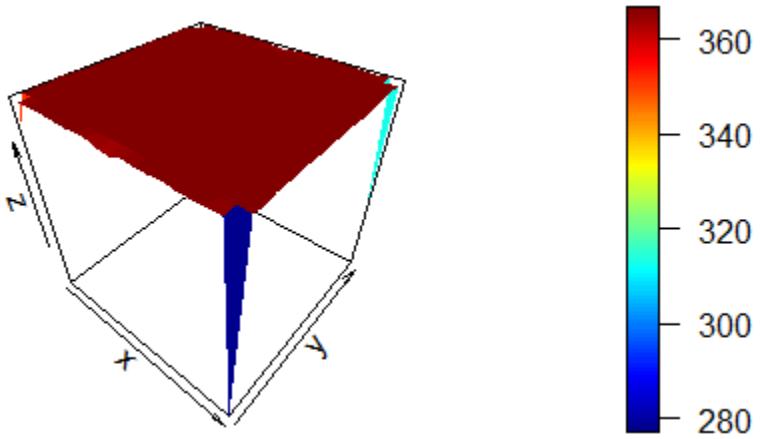


Figure 26: kriging in presence of an extreme value

For EPH, there is no such failure. The extreme value is part of the data and it is taken into account like if it was any other data.

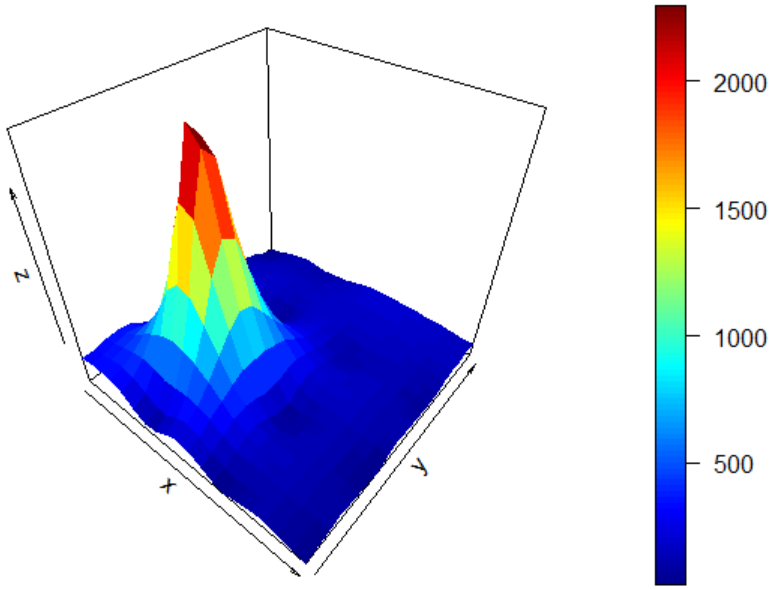


Figure 27: EPH in presence of an extreme value

X. 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, while the EPH adapts very well to it. Basically Kriging fails in the following cases:

- When information is poor, Kriging cannot know about variation.
- When data vary strongly.

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.

XI. References

[PIT] Olga Zeydina et Bernard Beauzamy : 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. Beauzamy (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.

[FEYNMAN] Report of the PRESIDENTIAL COMMISSION on the Space Shuttle Challenger Accident. Volume 2: Appendix F - Personal Observations on Reliability of Shuttle
by R. P. Feynman <https://history.nasa.gov/rogersrep/v2appf.htm> consulted on 9 August 2018