Interpolation

Maria Antonia Brovelli
Outline

- Classification of interpolation methods
- Inverse Distance Weighting (IDW)
- Local Polynomial (LP)
- Radial basis function (RBF)
- Least squares spline functions
  - Tikhonov’s regularization
  - Multiresolution spline functions
- Kriging
Classification of interpolation methods

- Different classifications are possible. In the following some of them:

  - Based on the mathematical approach:
    - Geometric
    - Statistical
    - Physics-based

  - Deterministic / Stochastic methods
    The first ones do not use probability theory; on the contrary the stochastic ones are based on the concept of randomness: the interpolated surface is one of the many possible that could have been produced from the observed data set. Stochastic methods include different kinds of kriging.
Deterministic/ Stochastic methods

- The mechanism behind the field observations depends on a large number of factors that influence the result, but cannot be easily modeled. Think, for instance, at a temperature field in a region: it depends on various factors, as the altitude, the latitude and so on. A complete model is difficult to be implemented.

- Another example can be the modeling of the outcome of throwing a dice.

- A complex deterministic model can (in principle) predict the outcome, when the forces, the trajectory in the air, the tumbling and bouncing is modeled in great detail, including the many imperfections of the dice and the table.

- A very simple stochastic model (with the six possible outcomes having equal probability) usually works better because most parameters of the deterministic model are not known, and the process of throwing cannot be controlled in sufficient detail.
Classification of interpolation methods

- **Global / Local methods**
  The global interpolators are represented by a single function for the whole region while in the local ones an algorithm applies to a (small or great) subset of the total points in turn. Therefore when using global methods the value of a single point affects the entire map, while when using the local ones it affects only the portion of the map closer to it.

- **Exact / Approximate methods**
  With the exact methods the interpolating surface passes through all points whose values are known. But if data we have are measurements, that is quantities affected by errors, it could be better to use, for instance, methods based on least squares approach, which leads to filter and therefore reduce the effects of error on the resulting surface.
Inverse Distance Weighting (IDW)

- Inverse Distance Weighting (IDW) is a deterministic exact interpolator.

\[
\hat{Z}(P_0) = \frac{\sum_{i=1}^{n} Z(x_i)}{\sum_{i=1}^{n} \frac{1}{d_{io}^r}}
\]

\[
d_{io}^r = \left| P_i - P_0 \right|^r \quad \forall i
\]

within the moving window
Inverse Distance Weighting

- Weights are proportional to the inverse distance raised to the power value. The power value is used for calculating the weights assigned to the points located in a moving window around each point.

- As the distance increases, the weights decrease depending on the power value. If the power value is high, only the immediate few surrounding points influence the prediction.

- When the power value equals two, we have the so called *inverse squared distance weighted* interpolation.

- To optimize the power value the root mean square prediction error (RMSPE) has to be minimized.

- The RMSPE is calculated from cross-validation: each measured point is removed and compared to the predicted value for that location; the RMSPE is the summary statistic obtained computing the average of the squared residuals.
Inverse Distance Weighting

- **Advantages:**
  - There are very few decisions to be made regarding model parameters. The power parameter determines how quickly weights fall off with distance from the point. As the power parameter approaches zero, the generated surface approaches a horizontal planar surface through the average of all observations from the data file. As the power parameter increases, the generated surface becomes a "nearest neighbor" interpolator and the resultant surface is polygonal.
  - There are no assumptions required of the data.

- **Drawbacks:**
  - There is no assessment of prediction errors.
  - It can produce "bulls eyes" patterns of concentric contours around the data points.
Local Polynomial (LP)

- Local Polynomial (LP) is a deterministic inexact interpolator. It does not require assumptions of the data.
- It takes into account data within a localized moving “window” around a point in which we are estimating the surface.
- The method fits local trends by weighted least squares:

\[
\sum_{i=1}^{n} w_i [Z(x_i) - P(x_i)]^2 = \min
\]

- where
  - \(n\) is the number of points within the window;
  - \(P(x_i)\) is the polynomial surface
  - \(w_i\) is a weight given by:
    \[
    w_i = e^{-\frac{3d_{io}}{a}}
    \]
  - \(a\) is the local polynomial weight parameter that controls how fast weights decay with distance
Local Polynomial

- The minimization occurs for the coefficients of the polynomials. They are estimated for each point and, in order to compute a surface, on a regular grid. The optimal value of the local polynomial weight parameter is the one that minimizes the root mean squared prediction errors from cross-validation.

- Advantages:
  - It is more flexible than the global polynomial method

- Drawbacks:
  - There are more decisions to be made regarding parameters. We have to decide the shape of the moving window and the number of observations to be taken into account and the polynomial degree. The weight parameter has to be assessed.
  - There is no assessment of prediction errors.
Radial basis function (RBF)

- Radial Basis Functions (RBF) are deterministic exact interpolators, i.e. the surface must go through each measured sample value. Radial Basis Functions make no assumptions about the data.

- As other interpolation methods, they are based on the intuitive fact that the interpolation function should be smooth. The smoothness condition can be expressed within variational principles as a minimization of a certain smoothness functional.

- There are several different basis functions:
  - thin-plate spline
  - spline with tension
  - completely regularized spline
  - multiquadric function
  - inverse multiquadric spline
If $S(x)$ is the surface, given N values of the analysed phenomenon $Z^{(j)}$, $j=1,N$ measured at discrete points $x^{(j)}$, the exact interpolator gives

$$S(x^{(j)}) = Z^{(j)}, j = 1, N$$

The minimization criterion can be expressed through a variational condition (smoothness seminorm). For example in case of minimization of curvature:

$$E_{P}(S) = \|S\| = \int_{\mathbb{R}^2} \left[ \left( \frac{\partial^2 S}{\partial x^2} \right)^2 + 2 \left( \frac{\partial^2 S}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 S}{\partial y^2} \right)^2 \right] dxdy = \text{min}$$

Different types of splines can be derived from different forms of smoothness seminorm
The surface of minimum curvature corresponds to the minimum of the Laplacian power (or satisfies the biharmonic differential equation).

Physically, it models the behaviour of an elastic plate. In the one-dimensional case, the minimum curvature method leads to the natural cubic spline interpolation. The surface is continuous and differentiable together with continuous first derivatives surfaces. It is not suitable for representing rapid change in the modelled phenomena as the plate’s stiffness creates unacceptable features, for instance false minimum or maximum.
The stiffness of the plate can be mitigated by introducing the first derivatives terms into the minimization criteria multiplied by a constant weight (tension parameter).

Anyway in this case also we cannot compute the second derivatives in the data points because of the possible presence of singularities.

In order to solve also this problem, the third derivative terms can be incorporated into the minimization criterion multiplied by another constant weight.
Tuning the surface using tension parameter

http://skagit.meas.ncsu.edu/~helena/classwork/cegis/img3.htm
### Examples of smoothness seminorms

<table>
<thead>
<tr>
<th>Method</th>
<th>Smoothness seminorm</th>
<th>Operator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Membrane spline</td>
<td>$E_M = \int_{R^2} \left[ \left( \frac{\partial S}{\partial x} \right)^2 + \left( \frac{\partial S}{\partial y} \right)^2 \right] , dxdy$</td>
<td>Laplacian</td>
</tr>
<tr>
<td>Thin plate spline</td>
<td>$E_P = \int_{R^2} \left[ \left( \frac{\partial^2 S}{\partial x^2} \right)^2 + 2 \left( \frac{\partial^2 S}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 S}{\partial y^2} \right)^2 \right] , dxdy$</td>
<td>Biharmonic differential</td>
</tr>
<tr>
<td>Thin plate spline + tension</td>
<td>$E_{PM} = E_P + \phi^2 , E_M$</td>
<td>Harmonic + Biharmonic</td>
</tr>
<tr>
<td>Regularized spline with tension</td>
<td>$E_R = E_{PM} + \tau^2 \int_{R^2} \left[ \left( \frac{\partial^3 S}{\partial x^3} \right)^2 + 3 \left( \frac{\partial^3 S}{\partial x \partial y^2} \right)^2 + 3 \left( \frac{\partial^3 S}{\partial x^2 \partial y} \right)^2 + \left( \frac{\partial^3 S}{\partial y^3} \right)^2 \right] , dxdy$</td>
<td>Biharmonic + 6° order</td>
</tr>
</tbody>
</table>
The general solution of the two conditions:

\[
\begin{cases}
  S(x^{(j)}) = Z^{(j)}, j = 1, N \\
  \|S\| = \min \text{imum}
\end{cases}
\]

is given by:

\[
S(x) = P(x) + \sum_{j=1}^{N} \lambda_j R(x, x^{(j)})
\]

where:

\[
P(x) = \sum_{l=1}^{m} a_l p_l(x)
\]

are components of low order variations which are not penalized by the smoothness functional \[\|p_l(x)\| = 0, \quad l = 1, \ldots, m\]
Radial Basis Function

\[ R(x, x^{(j)}) = R(r) \] are various choices of radial basis functions \( R: \mathbb{R}^+ \rightarrow \mathbb{R} \) depending only on the distance between the two points \( x \) and \( x^{(j)} \)

For instance in the 2-dimensional case we can have

\[ R(x, x^{(j)}) = R(r) = (\sigma r)^2 \ln \sigma r \quad \text{thin-plate spline function} \]

\[ R(x, x^{(j)}) = R(r) = \sqrt{r^2 + \sigma^2} \quad \text{multi-quadric function} \]

\[ R(x, x^{(j)}) = R(r) = \frac{1}{\sqrt{r^2 + \sigma^2}} \quad \text{inverse multi-quadric function} \]
Radial Basis Function

\[ R(x, x^{(i)}) = R(r) = \left[ \ln\left( \frac{\sigma r}{2} \right)^2 + c_E + E_1\left( \frac{\sigma r}{2} \right)^2 \right] \]

Regularized spline function, where:

- \( E_1 \) is the exponential integral function
- \( E_1(x) = \int_{1}^{x} \frac{e^{-tx}}{t} dt = \int_{1}^{x} \frac{e^{-u}}{u} du \)
- \( C_E \) is the Euler constant (\( C_E = 0.577215… \))

Spline with tension where:

- \( I_0 \) is the modified Bessel function
- \( I_0(x) = 1 + \frac{(x/2)^2}{1!^2} + \frac{(x/2)^4}{2!^2} + \frac{(x/2)^6}{3!^2} \)

For all methods except the inverse multiquadric, the higher the parameter values, the smoother the map.
Radial Basis Function

• $a_l$ and $\lambda_j$ are found by solving the system of linear equations

$$\sum_{j=1}^{N} \lambda_j p_l(x^{(j)}) = 0, \quad l = 1, ..., m$$

$$S(x^{(j)}) = Z^{(j)}, \quad j = 1, N$$

• The optimal $\sigma$ parameter, which is unknown, can be determined by finding the value that minimizes the root mean square prediction error (RMSPE): each measured point is removed and compared to the predicted value for that location.
Least squares spline functions

- Being \( t_{\text{min}} \leq t_0 < t_1 < t_2 < \ldots \ t_M < t_{M-1} \equiv t_{\text{max}} \)

a decomposition in the interval \([t_{\text{min}}, t_{\text{max}}]\), a spline function of degree \( m \geq 1 \) relative to the decomposition is a function \( s(t) \) that satisfies the following properties:

- \( s(t) \), limited to each interval \([t_k, t_{k+1}]\), \( k = 0, \ldots, M-2 \), is a polynomial at most of degree \( m \);

- \( s^{(j)}(t) = \frac{d^j}{dt^j} s(t) \) is a continuous function

on \([t_{\text{min}}, t_{\text{max}}]\) for \( j = 0, \ldots, m-1 \).

When \( m = 1 \) and \( m=3 \) respectively we are dealing with linear and cubic splines.
Least squares spline functions

Linear spline

\[ s^{(1)}(t) = \begin{cases} 
\varphi_1(t) = \frac{2w + t}{4w^2} & -2w \leq t \leq 0 \\
\varphi_2(t) = \frac{2w - t}{4w^2} & 0 \leq t \leq 2w \\
0 & \text{elsewhere}
\end{cases} \]
Least squares spline functions

Cubic spline

\[ s^{(3)}(t) = \begin{cases} 
0 & \text{when } t < -4w \\
\varphi_1(t) = \frac{(4w + t)^3}{96w^4} & \text{when } -4w \leq t \leq -2w \\
\varphi_2(t) = \frac{(4w + t)^3 - 4(2w + t)^3}{96w^4} & \text{when } -2w \leq t \leq 0 \\
\varphi_3(t) = \frac{(4w - t)^3 - 4(2w - t)^3}{96w^4} & \text{when } 0 \leq t \leq 2w \\
\varphi_4(t) = \frac{(4w - t)^3}{96w^4} & \text{when } 2w \leq t \leq 4w \\
0 & \text{when } t > 4w 
\end{cases} \]
Interpolation with spline functions

Each observation (●) can be described as a linear combination of the spline functions included in the sub-interval in which the observation itself is located:

\[ h_0(t_i) = \sum_{k=0}^{M-1} a_k s_k^{(3)}(t_i) + \nu_i \quad i = 1, N \quad k = 0, M - 1 \]

\[ s_k^{(3)}(t) = s^{(3)}\left(\frac{t - t_k}{\Delta t}\right) \]

\( a_k \) are the unknown quantities to be estimated.

Least squares spline functions

\( h_0(t_i) \) observations
\( \nu_i \) noise \( i = 0, N-1 \)
M spline functions = M grid nodes
Least squares spline functions

When $\bar{t} = t_0$

$$h_0(\bar{t}) = a_0 s^{(3)} \left( \frac{\bar{t} - t_0}{\Delta t} \right) + a_1 s^{(3)} \left( \frac{\bar{t} - t_1}{\Delta t} \right)$$
Least squares spline functions

When $\bar{t} \in [t_0, t_1]$

$$h_0(\bar{t}) = a_0 s^{(3)}\left(\frac{\bar{t} - t_0}{\Delta t}\right) + a_1 s^{(3)}\left(\frac{\bar{t} - t_1}{\Delta t}\right) + a_2 s^{(3)}\left(\frac{\bar{t} - t_2}{\Delta t}\right)$$
Least squares spline functions

When $\bar{t} \in [t_k, t_{k+1}]$

\[
h_0(\bar{t}) = a_{k-1}s^{(3)}\left(\frac{\bar{t} - t_{k-1}}{\Delta t}\right) + a_k s^{(3)}\left(\frac{\bar{t} - t_k}{\Delta t}\right) + a_{k+1}s^{(3)}\left(\frac{\bar{t} - t_{k+1}}{\Delta t}\right) + a_{k+2}s^{(3)}\left(\frac{\bar{t} - t_{k+2}}{\Delta t}\right)
\]
The least squares method is applied:

\[
\begin{bmatrix}
  h_{0_0} \\
  h_{0_1} \\
  \vdots \\
  h_{0_{N-1}} \\
\end{bmatrix}
= \begin{bmatrix}
  s_0 & s_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
  s_0 & s_1 & s_2 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
  0 & \cdots & 0 & s_{k-1} & s_k & s_{k+1} & s_{k+2} & 0 & \cdots & 0 \\
  0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & s_{M-3} & s_{M-2} & s_{M-1} \\
  0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & s_{M-2} & s_{M-1} \\
\end{bmatrix}
\cdot
\begin{bmatrix}
  a_0 \\
  a_1 \\
  \vdots \\
  a_k \\
  \vdots \\
  a_{M-2} \\
  a_{M-1} \\
\end{bmatrix}
\]

\[
\hat{a} = (A^t A)^{-1} A^t h_0
\]
How many splines should be used?

Each spline needs at least one observation in its support. The resolution step must be greater than the maximum distance between two observations.
Least squares spline functions

Few splines ➔ the surface does not follow the trend of the field in points with high variability.

Many splines ➔ the coefficient estimation is sensitive to single observations in areas with a low density of data; unwanted oscillations are therefore generated. Moreover, if the number of splines is greater than the maximum possible number, there is a singularity in the normal matrix and thus the spline coefficients cannot be estimated.
Least squares spline functions

11 spline
Tikhonov’s regularization

If data are not regularly distributed and, in particular, if there are areas with “holes” in observations, the spline coefficients of these areas are null; this causes an innatural trend of the interpolating surface:

A “regularization” must be applied
Let us apply a generalized least squares principle:

\[(A^tA + \lambda \cdot K)\hat{\alpha} = A^t h_0\]

\((\lambda \cdot K)\) is the factor involved in the “data holes”

Let us build the regularization in such a way to minimize the slope or the curvature of the interpolating function:
Two-dimensional generalization: bilinear spline

\[ s^{(1)}(t) = s^{(1)}(x, y) = \begin{cases} 
\varphi_{11}(x, y) = \frac{(2w + x)(2w + y)}{16w^4} & \text{if } x \in [-2w, 0] \\
\varphi_{12}(x, y) = \frac{(2w + x)(2w - y)}{16w^4} & \text{if } y \in [-2w, 0] \\
\varphi_{21}(x, y) = \frac{(2w - x)(2w + y)}{16w^4} & \text{if } x \in [0, 2w] \\
\varphi_{22}(x, y) = \frac{(2w - x)(4w - y)}{16w^4} & \text{if } y \in [0, 2w] 
\end{cases} \]
Least squares spline functions

Two-dimensional generalization: bicubic spline

\[ s^{(3)}(t) = s^{(3)}(x, y) = \begin{cases} 
\varphi_{33}(x, y) = \frac{(4w-x)^3}{96w^4} \cdot \frac{(4w-y)^3}{96w^4} & \begin{cases} 
x \in [0,2w] 
y \in [0,2w] 
\end{cases} \\
\varphi_{43}(x, y) = \frac{(4w-x)^3}{96w^4} \cdot \frac{4(2w-y)^3}{96w^4} & \begin{cases} 
x \in [2w,4w] 
y \in [0,2w] 
\end{cases} \\
\varphi_{34}(x, y) = \frac{(4w-x)^3}{96w^4} \cdot \frac{4(2w-x)^3}{96w^4} & \begin{cases} 
x \in [0,2w] 
y \in [2w,4w] 
\end{cases} \\
\varphi_{44}(x, y) = \frac{(4w-x)^3}{96w^4} \cdot \frac{(4w-y)^3}{96w^4} & \begin{cases} 
x \in [2w,4w] 
y \in [2w,4w] 
\end{cases} 
\end{cases} \]
Example: Let us consider 36 spline coefficients (bilinear splines) and 37 observations: coefficients $\lambda_{15}$, $\lambda_{16}$, $\lambda_{45}$ cannot be determined.
The choice of the regularization coefficient is important

- if a very low value is chosen, in correspondence to areas without data, the normal matrix is badly conditioned -> the parameters can’t be estimated
- if a very high value is chosen, the regularization contributions are given more weight and a smoother surface is obtained.
The idea is to use different “levels” of splines, corresponding to different halving steps (a new level corresponds to halving the width of the support of the previous level spline). Taking into account the field domain, i.e. the observation locations and namely $t_{\text{min}}$ and $t_{\text{max}}$, we can define the following levels:

\[ t_{\text{min}} \quad t_{\text{min}} + \Delta/2 \quad t_{\text{min}} + \Delta \quad t_{\text{min}} + 3\Delta/2 \quad t_{\text{max}} \]
Each observation is described as a linear combination of spline functions of decreasing (halving) \( \Delta \) step

\[
h(t_i) = \sum_{h=0}^{M-1} \sum_{k=0}^{N_h-1} \lambda_{h,k} \cdot \varphi\left(\frac{2^h(t_i - t_{\text{min}})}{\Delta} - k\right) + w_i
\]

where:

- \( M \) is the number of levels
- \( N_h \) is the number of splines at level \( h \) \( ( = 2^{h+1} +1) \)
- \( \Delta = (t_{\text{max}} - t_{\text{min}})/2 \)
Least squares spline functions

Contribute of each level

Level 1

Level 2

Level 3

Level 4

Level 5

Level 1 + Level 2 + Level 3 + Level 4 + Level 5
Constraints must be introduced on $\lambda_{hk}$ coefficients in order to avoid local rank deficiency. To be very cautious, the following criterion can be adopted:

A generic $k$-th spline function at $h$ level

$$\varphi_{\Delta_h}(t-k_i\Delta_h-t_{min}) \quad \Delta_h = \frac{\Delta}{2^h}$$

is active (i.e. $\lambda_{hk} \neq 0$)

- if we have at least $f$ ($f>0$) observations for each $\Delta_h$ half-support of the spline
- if it does not exist a spline at lower level having the same application point
Generalising in two dimensions, we have

$$h(t) = \sum_{h=0}^{M} \left[ \sum_{l=0}^{N_{1h}-1} \sum_{k=0}^{N_{2h}-1} \lambda_{h,l,k} \cdot \varphi_{\Delta_h} (t - \Delta_h \tau_{lk} - t_{\min}) \right]$$

where

$$t_{\min} = \begin{bmatrix} t_{1\min} & t_{2\min} \end{bmatrix}^T$$

$$\Delta_h = \begin{bmatrix} \Delta_{1h} & 0 \\ 0 & \Delta_{2h} \end{bmatrix}$$

$$\varphi_{\Delta_h}(t) = \varphi_{\Delta_{1h}}(t_1) \cdot \varphi_{\Delta_{2h}}(t_2) = \varphi\left(\frac{t_1}{\Delta_{1h}}\right) \cdot \varphi\left(\frac{t_2}{\Delta_{2h}}\right)$$

and, to avoid local rank deficiency, we generalize the same criterion seen in 1D (we need at least \(f\) observations for each quarter of spline support)
Kriging

- Kriging is a statistical tool developed by Matheron (1963) and named in honour of D.G. Krige (a South African mining engineer who pioneered the field of geostatistics).
- Kriging can be defined as a best linear unbiased estimator of a spatial variable, because it provides linear regression estimate, which is unbiased and has minimum error variance.
- It assumes data from a stationary stochastic process
  - constant mean (or trend) throughout the region
  - variance of differences between any two samples independent of position and dependent only on the distance between pairs
There are three principal Kriging types:

- **Simple kriging** (SK) assumes constant mean \( m \) known in the study area.
- **Ordinary Kriging** (OK), when mean is constant but unknown.
- **Universal kriging** (UK) assumes that unknown mean \( m(x) \) is changing smoothly in the study area \( S \) and is considered as a trend. Trend component is modeled as a linear combination of known functions \( f_k(x) \).
Kriging

Interpolation rule

- The general formula for kriging is given by a weighted sum of the data:

\[ Z(x^0) = \sum_{i=1}^{N} \hat{\lambda}_i Z(x^{(i)}) \]

- where
  - \( Z(x^{(i)}) \) is the measured value at the i-th location;
  - \( \lambda_i \) is an unknown weight for the measured value at the ith location;
  - \( x^0 \) is the prediction location;
  - \( N \) is the number of measured values.
In Kriging the weights are based on the spatial autocorrelation: semivariogram modeling is a key step.

Empirical semivariogram provides information on the spatial autocorrelation of datasets, however, it does not provide information for all possible directions and distances.

For this reason, and to ensure that kriging predictions have positive kriging variances, it is necessary to fit a model, i.e. a continuous function, to the empirical semivariogram.

There are many semivariogram models to choose from (circular, spherical, exponential, gaussian,....)
Examples of semivariogram models

Nug(0) = \begin{cases} 
0 & h = 0 \\
1 & h > 0 
\end{cases}

Nugget model

Lin(a) = \begin{cases} 
\frac{h}{a} & 0 \leq h \leq a \\
1 & h > a 
\end{cases}

Linear with threshold

Sph(a) = \begin{cases} 
\frac{3h}{2a} - \frac{1}{2} \left( \frac{h}{a} \right)^3 & 0 \leq h \leq a \\
1 & h > a 
\end{cases}

Spherical model

Exp(a) = 1 - e^{-\frac{h}{a}} \quad h \geq 0

Exponential model
Steps in kriging

1) Calculate semivariogram values
2) Produce semivariogram model
3) Compute kriging interpolation

Rule of thumb: lag size * number of lags = half of the maximum distance between pairs
Kriging

Indexes (1)

- $\varepsilon_i = \text{prediction error} = \text{Measured} - \text{Predicted}$
- $\hat{\sigma}_i = \text{standard error} \quad \text{(the square root of the variance of a prediction, computed only by using kriging)}$
- $\text{MPE} = \text{mean prediction error} = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i$
- $\text{RMSPE} = \text{root-mean-square prediction error} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (\varepsilon_i - \text{MPE})^2}$
Kriging

Indexes (2)

- $\hat{\xi}_i = \frac{\varepsilon_i}{\hat{\sigma}_i} = \text{standardized prediction error}$
- $\text{ASE} = \text{average standard error} = \frac{1}{n} \sum_{i=1}^{n} \hat{\sigma}_i$
- $\text{MSPE} = \text{mean standardized prediction error} = \frac{1}{n} \sum_{i=1}^{n} \hat{\xi}_i$
- $\text{RMSSPE} = \text{root-mean-square standardized prediction error} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (\hat{\xi}_i - \text{MSPE})^2}$
Check the goodness of the interpolation

- MPE should be near zero
- MSPE should be near zero
- the smaller the RMSPE, the better
- uncertainty of prediction can be measured
  - by comparing average estimated standard error (ASE) with the root-mean-square prediction error (RMSPE).
    - $\text{ASE} = \text{RMSPE} \rightarrow$ good
    - $\text{ASE} > \text{RMSPE} \rightarrow$ variability overestimate
    - $\text{ASE} < \text{RMSPE} \rightarrow$ variability underestimate
  - by analysing the RMSSPE (root-mean-square standardized prediction error)
    - $\text{RMSSPE} = 1 \rightarrow$ good,
    - $\text{RMSSPE} < 1 \rightarrow$ variability overestimate
    - $\text{RMSSPE} > 1 \rightarrow$ variability underestimate