Spatial Interpolation Comparison
Evaluation of spatial prediction methods

Tomislav Hengl
ISRIC — World Soil Information, Wageningen University
Geostatistics = a toolbox to generate maps from point data i.e. to interpolate;
Geostatistics = a toolbox to generate maps from point data i.e. to interpolate;
There are many possibilities;
Geostatistics = a toolbox to generate maps from point data i.e. to interpolate;

There are many possibilities;

An inexperienced user will often be challenged by the amount of techniques to run spatial interpolation;
Geostatistics = a toolbox to generate maps from point data i.e. to interpolate;

There are many possibilities;

An inexperienced user will often be challenged by the amount of techniques to run spatial interpolation;

...which method should we use?
Have you heard of SIC?

Automatic mapping algorithms for routine and emergency monitoring data

Spatial Interpolation comparison 2004

The Spatial Interpolation Comparison (SIC) 2004 exercise was organised during the summer 2004 within the frame of the activities of the Radioactivity Environmental Monitoring (REM) group of the Institute for Environment and Sustainability at the Joint Research Centre (JRC) of the European Commission. Its purpose was to assess the current knowledge in the field of automatic or “real-time” mapping. The underlying idea was to explore the way algorithms designed for spatial interpolation can automatically generate maps on the basis of information collected regularly by monitoring networks, in this situation of routine or emergencies. This EU report presents a collection of essays written by experts in the field of geostatistics and machine learning about the results obtained by the researchers who participated to SIC 2004. The report should provide the reader with a useful overview of the current state of the art in the field of automatic mapping of environmental variables.

Corporate author(s): Joint Research Centre, European Commission
Theme: Environmental research
The spatial prediction game

Participants were invited to estimate values located at 1000 locations (right, crosses), using 200 observations (left, circles).
Lessons learned (from SIC)

Emergency scenario

- Geostatistics
- Splines
- Neural Net.
- Support Vector Machine
- Others

Mean Absolute Errors (MAE) vs. Root Mean Squared Errors (RMSE)
How many techniques are there?

Li and Heap (2008) list over 40 unique techniques.

1. Are all these equally valid?
2. How to objectively compare various methods (which criteria to use)?
3. Which method to pick for your own case study?
There are not as many

There are roughly four main clusters of techniques:

1. splines (deterministic);
There are not as many

There are roughly four main clusters of techniques:

1. splines (deterministic);
2. kriging-based (plain geostatistics);
There are not as many

There are roughly four main clusters of techniques:

1. splines (deterministic);
2. kriging-based (plain geostatistics);
3. regression-based;
There are not as many

There are roughly four main clusters of techniques:

1. splines (deterministic);
2. kriging-based (plain geostatistics);
3. regression-based;
4. expert systems / machine learning;
The 5 criteria

1. the **overall mapping accuracy**, e.g. standardized RMSE at control points — the amount of variation explained by the predictor expressed in %;
The 5 criteria

1. the **overall mapping accuracy**, e.g. standardized RMSE at control points — the amount of variation explained by the predictor expressed in %;

2. the **bias**, e.g. mean error — the accuracy of estimating the central population parameters;
The 5 criteria

1. the **overall mapping accuracy**, e.g. standardized RMSE at control points — the amount of variation explained by the predictor expressed in %;

2. the **bias**, e.g. mean error — the accuracy of estimating the central population parameters;

3. the **model robustness**, also known as *model sensitivity* — in how many situations would the algorithm completely fail / how much artifacts does it produces?;
The 5 criteria

1. the **overall mapping accuracy**, e.g. standardized RMSE at control points — the amount of variation explained by the predictor expressed in %;

2. the **bias**, e.g. mean error — the accuracy of estimating the central population parameters;

3. the **model robustness**, also known as *model sensitivity* — in how many situations would the algorithm completely fail / how much artifacts does it produces?;

4. the **model reliability** — how good is the model in estimating the prediction error (how accurate is the prediction variance considering the true mapping accuracy)?
The 5 criteria

1. the **overall mapping accuracy**, e.g. standardized RMSE at control points — the amount of variation explained by the predictor expressed in %;

2. the **bias**, e.g. mean error — the accuracy of estimating the central population parameters;

3. the **model robustness**, also known as *model sensitivity* — in how many situations would the algorithm completely fail / how much artifacts does it produces?;

4. the **model reliability** — how good is the model in estimating the prediction error (how accurate is the prediction variance considering the true mapping accuracy)?;

5. the **computational burden** — the time needed to complete predictions;
Can we simplify this?

1. In theory, we could derive a single composite measure that would then allow you to select ‘the optimal’ predictor for any given data set (but this is not trivial!)
2. But how to assign weights to different criteria?
3. In many cases we simply finish using some naïve predictor — that is predictor that we know has a statistically more optimal alternative, but this alternative is not feasible.
Automated mapping

In the intamap package\(^1\) decides which method to pick for you:

\[ \text{meuse}\$value <- \log(\text{meuse}\$zinc) \]
\[ \text{output} <- \text{interpolate(data=meuse, newdata=meuse.grid)} \]

R 2009-11-11 17:09:14 interpolating 155 observations, 3103 prediction locations
[Time models loaded...]  
[1] "estimated time for copula 133.479866956255"
Checking object ... OK

\(^{1}\text{http://cran.r-project.org/web/packages/intamap/}\)
Hypothesis

We need a single criteria to compare various prediction methods.
Mapping accuracy and survey costs

The cost of a soil survey is also a function of mapping scale, roughly:

$$\log(X) = b_0 + b_1 \cdot \log(SN)$$  \hspace{1cm} (1)

We can fit a linear model to the empirical table data from e.g. Legros (2006; p.75), and hence we get:

$$X = \exp(19.0825 - 1.6232 \cdot \log(SN))$$  \hspace{1cm} (2)

where \(X\) is the minimum cost/ha in Euros (based on estimates in 2002). To map 1 ha of soil at 1:100,000 scale, for example, one needs (at least) 1.5 Euros.
Survey costs and mapping scale

Minimum survey costs in EUR / ha (log-scale)

Scale number (log-scale)
Total costs of a soil survey can be estimated by using the size of area and number of samples. The effective scale number (SN) is:

\[ SN = \sqrt{4 \cdot \frac{A}{N} \cdot 10^2} \quad \ldots \quad SN = \sqrt{\frac{A}{N} \cdot 10^2} \quad (3) \]

where \( A \) is the surface of the study area in \( m^2 \) and \( N \) is the total number of observations.
Converges to:

\[ X = \exp \left( 19.0825 - 1.6232 \cdot \log \left[ 0.0791 \cdot \sqrt{\frac{A}{N}} \cdot 10^2 \right] \right) \quad (4) \]
The resulting (predictions) map is a sum of two signals:

\[ Z^*(s) = Z(s) + \varepsilon(s) \]  \hspace{1cm} (5)

where \( Z(s) \) is the true variation, and \( \varepsilon(s) \) is the error component. The error component consists, in fact, of two parts: (1) the unexplained part of soil variation, and (2) the noise (measurement error). The unexplained part of soil variation is the variation we somehow failed to explain because we are not using all relevant covariates and/or due to the limited sampling intensity.
In order to see how much of the global variation budget has been explained by the model we can use:

\[
RMSE_r(\%) = \frac{RMSE}{s_z} \cdot 100
\]  

(6)

where \(s_z\) is the sampled variation of the target variable. \(RMSE_r(\%)\) is a global estimate of the map accuracy, valid only under the assumption that the validation points are spatially independent from the calibration points, representative and large enough (\(\gg 100\)).
Kriging efficiency

- Global variance
- Statistical extrapolation

part of variation that cannot be explained
Prediction accuracy and survey costs

% of variance explained vs. total survey costs

- Inherently unsolvable part of variation: short range variability, measurement errors
- Theoretical linear relationship
- Initial deterministic part of variation that can be explained with small number of samples

Geostatistics course, 25–29 October 2010, Wageningen
Mapping efficiency

We propose the measure of **mapping efficiency**, defined as the amount of money needed to map an area of standard size and explain each one percent of variation in the target variable:

\[
\theta = \frac{X}{A \cdot RMSE_r} \quad [\text{EUR} \cdot \text{km}^{-2} \cdot \%^{-1}]
\]

(7)

where \( X \) is the total costs of a survey, \( A \) is the size of area in \( \text{km}^{-2} \), and \( RMSE_r \) is the amount of variation explained by the spatial prediction model.
Comparing methods

Part of variation that cannot be explained

100 %

% of variance explained

Data driven predictor

Bulk variance

Method (A)

Suitable for small data sets

Method (B)

Sub-optimal method

Method (C)

Initial costs

Total survey costs

Geostatistics course, 25–29 October 2010, Wageningen
Exercise

To follow this exercise, obtain the GlobalSoilMap_calculus.R script. Download it to your machine and then run step-by-step.
> data(meuse)
> coordinates(meuse) <- ~x+y
> proj4string(meuse) <- CRS("+init=epsg:28992")
> sel <- !is.na(meuse$om)
> bubble(meuse[sel,], "om")
The size of the study area is 4.96 km$^2$. The cost of this soil survey is about 1150 EUR per km$^2$, or 5,700 EUR in total.
The size of the study area is 4.96 km². The cost of this soil survey is about 1150 EUR per km², or 5,700 EUR in total.

The cross-validation error for mapping organic matter is: 0.234 for OK and 0.222 for RK method (values in log-scale).
Results

- The size of the study area is 4.96 km$^2$. The cost of this soil survey is about 1150 EUR per km$^2$, or 5,700 EUR in total.
- The cross-validation error for mapping organic matter is: 0.234 for OK and 0.222 for RK method (values in log-scale).
- The mapping efficiency for RK method is hence 20.70 EUR km$^{-2}$ %$^{-1}$, compared to 22.08 EUR km$^{-2}$ %$^{-1}$ for the OK method.
The size of the study area is 4.96 km$^2$. The cost of this soil survey is about 1150 EUR per km$^2$, or 5,700 EUR in total.

The cross-validation error for mapping organic matter is: 0.234 for OK and 0.222 for RK method (values in log-scale).

The mapping efficiency for RK method is hence 20.70 EUR km$^{-2}$ %$^{-1}$, compared to 22.08 EUR km$^{-2}$ %$^{-1}$ for the OK method.

There is a gain of ca.7% in the mapping efficiency with the RK method compared to the OK method.
Whitening

(a) ordinary kriging

(b) regression-kriging
Conclusions

- Mapping efficiency (cost / area / percent of variance explained) is a possible universal criteria to compare prediction methods.
Conclusions

- Mapping efficiency (cost / area / percent of variance explained) is a possible universal criteria to compare prediction methods.
- Maps are not what they seem.
Conclusions

- Mapping efficiency (cost / area / percent of variance explained) is a possible universal criteria to compare prediction methods.
- Maps are not what they seem.
- Geostatistics really outperforms non-statistical methods (but this is area/data dependent).
Conclusions

▶ Mapping efficiency (cost / area / percent of variance explained) is a possible universal criteria to compare prediction methods.

▶ Maps are not what they seem.

▶ Geostatistics really outperforms non-statistical methods (but this is area/data dependent).

▶ It’s not about the making beautiful maps, it’s about understanding what they mean.
Mapping efficiency (cost / area / percent of variance explained) is a possible universal criteria to compare prediction methods.

Maps are not what they seem.

Geostatistics really outperforms non-statistical methods (but this is area/data dependent).

It’s not about the making beautiful maps, it’s about understanding what they mean.

If you deal with several equally valid (independent) methods, maybe you should consider combining them?

