Optimization of Sample Configurations using Spatial Simulated Annealing

Congreso Escuela en Estadística Espacial

ALESSANDRO SAMUEL-ROSA
24 September 2019

Universidade Tecnológica Federal do Paraná
alessandrorosa@utfpr.edu.br
Introduction – Spatial Modelling
Spatial modelling is the art of constructing models – explanations – of spatial variation of geographic phenomena.
Spatial modelling is the art of constructing models – explanations – of spatial variation of geographic phenomena.

Spatial modellers aim at constructing simple yet accurate models of the spatial variation of geographic phenomena – given the available resources and the intended application.
Spatial modelling is the art of constructing models – explanations – of spatial variation of geographic phenomena.

Spatial modellers aim at constructing simple yet accurate models of the spatial variation of geographic phenomena – given the available resources and the intended application.

Spatial models (should) serve the practical purpose of producing the spatial information needed to support many of our every-day decisions.
Modern spatial modelling is based on using *statistical models* that account for:

- the *empirical correlation* between environmental conditions and the target geographic phenomenon
- the *empirical correlation* of the target geographic phenomenon itself – autocorrelation
Modern spatial modelling is based on using *statistical models* that account for:

- the *empirical correlation* between environmental conditions and the target geographic phenomenon
- the *empirical correlation* of the target geographic phenomenon itself – autocorrelation

This is the *mixed model of spatial variation*
The mixed model of spatial variation can be represented as:

\[ Y(s) = m(s) + e(s) \]
The mixed model of spatial variation can be represented as:

\[ Y(s) = m(s) + e(s) \]

- \( Y \) is the target geographic phenomenon at spatial location \( s \).
- \( m(s) \) are the fixed effects, the *deterministic* environmental conditions – that can be modelled using a (linear) trend function.
- \( e(s) \) are the random effects, the seemingly *stochastic* spatial variation – that can be modelled using a covariance function.
Spatial models – such as the mixed model – are a simplification of reality – they explain only a small part of the spatial variation of geographic phenomena.

The outcome of any spatial model – a (digital) map – will always deviate from the “truth”, i.e. be in error.
Spatial models – such as the mixed model – are a simplification of reality – they explain only a small part of the spatial variation of geographic phenomena.

The outcome of any spatial model – a (digital) map – will always deviate from the “truth”, i.e. be in error.

There are multiple sources of uncertainty in spatial modelling:

- Interpolation/extrapolation error
- Data errors (analytical error, sample design and size)
- Covariate errors (poor correlation with target phenomenon)
- Model structural error (linear or non-linear)

Today we will talk about sample design
Spatial Sampling
The usual spatial modelling *challenge*:

1. Multiple geographic phenomena have to be modelled/mapped
2. We know very little about the form of the models of spatial variation – *Terra Incognita*
3. Operational constraints limit sampling to a single phase
Sampling in *Terra Incognita*

The usual spatial modelling *challenge*:

1. Multiple geographic phenomena have to be modelled/mapped
2. We know very little about the form of the models of spatial variation – *Terra Incognita*
3. Operational constraints limit sampling to a single phase

In the mixed model context, we need an efficient spatial sample to meet three *conflicting objectives*:

1. Identify and estimate the spatial trend, \( Y(s) = m(s) + e(s) \)
2. Identify and estimate the covariance function, \( Y(s) = m(s) + e(s) \)
3. Make spatial predictions, \( Y(s) = m(s) + e(s) \)
Purposive Sampling – Free Survey

Traditional sampling method to produce area-class soil maps

- The surveyor is free to select the observation locations
- Selected based on conceptual and operational factors
- Goals: learn/verify spatial relationships and maximize the number of observations and geographic coverage
- Personal factors can play a role too, e.g. motivation

A chosen observation location
Purposive sampling is a non-probability sampling mode
Sampling locations are selected intentionally as to satisfy an *a priori* criterion
Based on the *statistical model* that will be used to infer the structure of spatial variation of $Y(s)$
Purposive sampling is a non-probability sampling mode

Sampling locations are selected intentionally as to satisfy an *a priori* criterion

Based on the *statistical model* that will be used to infer the structure of spatial variation of $Y(s)$

The modelling framework needs to be made explicit – translate objective into a function, an *objective function*

- Mathematical and heuristic rules are formalized in the form of a computer algorithm
- Find the sampling locations that minimize (or maximize) that criterion
An example
Suppose that we know before hand that the relation between a target geographic phenomenon and an auxiliary variable is linear

\[ Y(s) = \beta_0 + \beta_1 X(s) + e(s) \]

- A sample is needed to estimate the parameters $\beta_0$ and $\beta_1$ of this linear model with minimum variance – $(X^T X)^{-1} \sigma^2$
Suppose that we *know before hand* that the relation between a target geographic phenomenon and an auxiliary variable is linear

\[ Y(s) = \beta_0 + \beta_1 X(s) + e(s) \]

- A sample is needed to estimate the parameters \( \beta_0 \) and \( \beta_1 \) of this linear model with *minimum variance* – \((X^TX)^{-1}\sigma^2\)

From statistical theory: determinant of information matrix \( X^TX \)

- Search for the sample configuration that maximizes \(|X^TX|\)
- We now have an *objective function*
Spatial Simulated Annealing
There are various ways to search for a sample configuration that minimizes (or maximizes) a criterion

- Exhaustive search: check all possibilities and keep the best
There are various ways to search for a sample configuration that minimizes (or maximizes) a criterion

- Exhaustive search: check all possibilities and keep the best

Exhaustive search can be VERY time consuming

*Spatial simulated annealing* is a reasonable alternative
Spatial Simulated Annealing

A relatively simple algorithm that works by trial and error:

1. Start with a completely random sample configuration
2. Compute its objective function value
3. Select one sample and randomly shift its location
4. Compute the objective function value of the new sample configuration
5. Decide whether to accept or not the new sample configuration
6. Select another sample and randomly shift its location
7. Compute the objective function value, decide whether to accept
8. Continue till the optimum sample configuration is found
How to decide whether to accept or not the new sample configuration?

Metropolis criterion: \textit{acceptance probability} $P(X_i \rightarrow X_{i+1})$
How to decide whether to accept or not the new sample configuration?

Metropolis criterion: \textit{acceptance probability} \( P(X_i \rightarrow X_{i+1}) \)

\[
P(X_i \rightarrow X_{i+1}) = \begin{cases} 
1, & \text{if } f(X_{i+1}) \leq f(X_i), \\
\exp \left( \frac{f(X_i) - f(X_{i+1})}{T} \right), & \text{if } f(X_{i+1}) > f(X_i), 
\end{cases}
\]
How to decide whether to accept or not the new sample configuration?

Metropolis criterion: \textit{acceptance probability} $P(X_i \rightarrow X_{i+1})$

$$P(X_i \rightarrow X_{i+1}) = \begin{cases} 1, & \text{if } f(X_{i+1}) \leq f(X_i), \\ \exp \left( \frac{f(X_i) - f(X_{i+1})}{T} \right), & \text{if } f(X_{i+1}) > f(X_i), \end{cases}$$

- A better sample configuration is always accepted
- A worse sample configuration sometimes is accepted too – escape from \textit{local optima}
Optimization – Local and Global Minima

Energy space

Configuration space

Local minima

Global minima

Local minima
The Metropolis criterion has a \textit{temperature} parameter $T$

$$P(X_i \rightarrow X_{i+1}) = \begin{cases} 
1, & \text{if } f(X_{i+1}) \leq f(X_i), \\
\exp \left( \frac{f(X_i) - f(X_{i+1})}{T} \right), & \text{if } f(X_{i+1}) > f(X_i),
\end{cases}$$

The temperature decreases as the optimization goes on.
The Metropolis criterion has a \textit{temperature} parameter $T$

$$P(X_i \rightarrow X_{i+1}) = \begin{cases} 1, & \text{if } f(X_{i+1}) \leq f(X_i), \\ \exp\left(\frac{f(X_i) - f(X_{i+1})}{T}\right), & \text{if } f(X_{i+1}) > f(X_i), \end{cases}$$

The temperature decreases as the optimization goes on

- Worse sample configurations are more likely to be accepted in the beginning of the optimization
- At the end of the optimization, only better sample configurations are accepted
The Metropolis criterion has a **temperature** parameter $T$

$$P(\mathbf{X}_i \rightarrow \mathbf{X}_{i+1}) = \begin{cases} 1, & \text{if } f(\mathbf{X}_{i+1}) \leq f(\mathbf{X}_i), \\ \exp\left(\frac{f(\mathbf{X}_i)-f(\mathbf{X}_{i+1})}{T}\right), & \text{if } f(\mathbf{X}_{i+1}) > f(\mathbf{X}_i), \end{cases}$$

The temperature decreases as the optimization goes on

- Worse sample configurations are more likely to be accepted in the beginning of the optimization
- At the end of the optimization, only better sample configurations are accepted

Also, shorter random shifts in samples as the optimization approaches its end – the optimal solution is expected to be nearby
Evolution of objective function values during the optimization
Back to *Terra Incognita*
Recall the usual spatial modelling *challenge*:

1. Multiple geographic phenomena have to be modelled/mapped
2. We know very little about the form of the models of spatial variation
   - *Terra Incognita*
3. Operational constraints limit sampling to a single phase
Recall the usual spatial modelling \textit{challenge}:

1. Multiple geographic phenomena have to be modelled/mapped
2. We know very little about the form of the models of spatial variation – \textit{Terra Incognita}
3. Operational constraints limit sampling to a single phase

In the mixed model context, we need an efficient spatial sample to meet three \textit{conflicting objectives}:

1. Identify and estimate the spatial trend, $Y(s) = m(s) + e(s)$
2. Identify and estimate the covariance function, $Y(s) = m(s) + e(s)$
3. Make spatial predictions, $Y(s) = m(s) + e(s)$
Objective Functions
Various objective functions have already been proposed. Is there room for improvement?

- Spatial (nonlinear) trend estimation ($m(s)$)
- Variogram estimation ($e(s)$)
- Spatial interpolation ($Y(s)$)
Various objective functions have already been proposed

Is there room for improvement?

- Spatial (nonlinear) trend estimation \( m(s) \)
- Variogram estimation \( e(s) \)
- Spatial interpolation \( Y(s) \)

How to combine these conflicting objective functions – spatial modelling is a \textit{multi-objective combinatorial optimization problem}
Variogram Estimation \((e(s))\)
Variogram Estimation \( (e(s)) \)

**Space**  
Variogram space, i.e. the unidimensional space defined by the distances between sample points.

**Algorithm**  
Point-pairs per lag-distance class.

**Goal**  
Uniform distribution of point-pairs per equidistant lag-distance class in the empirical variogram.
Variogram Estimation \((e(s))\)

**Space** Variogram space, i.e. the unidimensional space defined by the distances between sample points.

**Algorithm** Point-pairs per lag-distance class.

**Goal** Uniform distribution of point-pairs per equidistant lag-distance class in the empirical variogram.

Example: six lag-distance classes.

Equidistant lags.
Variogram Estimation \( (e(s)) \)

**Space**
Variogram space, i.e. the unidimensional space defined by the distances between sample points.

**Algorithm**
Points Per Lag-distance class (PPL).

**Goal**
Uniform distribution of points per exponential lag-distance class in the empirical variogram.

Example: six lag-distance classes.

Exponential lags.
Variogram Estimation \((e(s))\)

<table>
<thead>
<tr>
<th>random</th>
<th>systematic</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Random Samples" /></td>
<td><img src="image2.png" alt="Systematic Samples" /></td>
</tr>
<tr>
<td>point-pairs</td>
<td>points</td>
</tr>
</tbody>
</table>

Spatial samples in a square of 500 × 500.
Spatial Interpolation \( (Y(s)) \)
Spatial Interpolation ($Y(s)$)

**Space**  Geographic space, i.e. the bi-dimensional space defined by the boundaries of the sampling region.

**Algorithm**  Spatial Coverage Sampling ($k$-means algorithm, SPCOSA).

**Goal**  Minimize the overall distance between sample and prediction points.
Spatial Interpolation ($Y(s)$)

**Space** Geographic space, i.e. the bi-dimensional space defined by the boundaries of the sampling region.

**Algorithm** Mean Squared Shortest Distance (MSSD).

**Goal** Minimize the overall distance between sample and prediction points.

Example: Regular grid with 36 observations.

Uniform coverage.
Multi-Objective Combinatorial Optimization Problem
Completely different sample configurations for

- Variogram identification and estimation
- Spatial interpolation
When solving a MOCOP, one aims at minimizing the vector of \( k \) objective functions

\[
\mathbf{f}(\mathbf{X}) = (f_1(\mathbf{X}), f_2(\mathbf{X}), \ldots, f_k(\mathbf{X})),
\]

(1)
When solving a MOCOP, one aims at minimizing the vector of \( k \) objective functions

\[
f(X) = (f_1(X), f_2(X), \ldots, f_k(X)), \tag{1}
\]

To find a single optimum solution, one can aggregate the objective functions into a single \textit{utility function}

\[
U = \sum_{i=1}^{k} w_i f_i(X), \tag{2}
\]
Multi-Objective Combinatorial Optimization Problem

When solving a MOCOP, one aims at minimizing the vector of $k$ objective functions

$$f(X) = (f_1(X), f_2(X), \ldots, f_k(X)),$$  \hspace{1cm} (1)

To find a single optimum solution, one can aggregate the objective functions into a single utility function

$$U = \sum_{i=1}^{k} w_i f_i(X),$$  \hspace{1cm} (2)

Objective functions need to be scaled to the same approximate range of values – eliminate any potential numerical dominance

How do we scale the objective functions?
The *upper-lower bound approach*:

\[ f_i'' = \frac{f_i(X) - f_i^\circ}{f_i^{\text{max}} - f_i^\circ}, \]  

(3)
The *upper-lower bound approach*:

\[ f_i'' = \frac{f_i(X) - f_i^\circ}{f_i^{\text{max}} - f_i^\circ}, \quad (3) \]

\( f_i^\circ \) is the *utopia point*, the single best solution for an objective function \( f_i^{\text{max}} \) is the single worst solution for an objective function, the *nadir point*.
The **upper-lower bound approach**:

\[ f'_i = \frac{f_i(X) - f_i^\circ}{f_i^{\text{max}} - f_i^\circ}, \]  

(3)

\( f_i^\circ \) is the **utopia point**, the single best solution for an objective function. \( f_i^{\text{max}} \) is the single worst solution for an objective function, the **nadir point**. These can be found empirically (takes time), approximated numerically (sub-optimal) or (rarely) calculated.
Spatial (Nonlinear) Trend Estimation ($m(s)$)
Spatial (Nonlinear) Trend Estimation ($m(s)$)

**Space** Attribute space, i.e. the multi-dimensional space defined by the covariates (auxiliary variables).

**Algorithm** Conditioned Latin hypercube sampling (CLHS).

**Goal** Reproduce (1) the marginal distribution of the numeric and (2) factor covariates, and (3) the linear correlation between numeric covariates.
Spatial (Nonlinear) Trend Estimation ($m(s)$)

**Space**  Attribute space, i.e. the multi-dimensional space defined by the covariates (auxiliary variables).

**Algorithm**  Conditioned Latin hypercube sampling (CLHS).

**Goal**  Reproduce (1) the marginal distribution of the numeric and (2) factor covariates, and (3) the linear correlation between numeric covariates.

Example: three samples from two covariates with three classes each.

A Latin square.

Land use map

Area-class soil map
Spatial (Nonlinear) Trend Estimation ($m(s)$)

**Space**  Attribute space, i.e. the multi-dimensional space defined by the covariates (auxiliary variables).

**Algorithm**  Association/Correlation measure and marginal Distribution of the Covariates (ACDC).

**Goal**  Reproduce (1) the marginal distribution of the covariates, and (2) the linear association/correlation between covariates.

Example: three samples from two covariates with three classes each.

A Latin square.
Spatial (Nonlinear) Trend Estimation \((m(s))\)

Numerical behaviour.
Spatial (Nonlinear) Trend Estimation \((m(s))\)

Optimized spatial sample configurations.
Sampling in *Terra Incognita*
Three sampling algorithms to meet each sampling objective:

ACDC  Spatial trend estimation, \( Y(s) = m(s) + e(s) \)

PPL  Variogram estimation, \( Y(s) = m(s) + e(s) \)

MSSD  Spatial interpolation, \( Y(s) = m(s) + e(s) \)

\(^1\text{https://CRAN.R-project.org/package=spsann}\)
Sampling in *Terra Incognita*

Three sampling algorithms to meet each sampling objective:

**ACDC**  Spatial trend estimation, $Y(s) = m(s) + e(s)$
**PPL**  Variogram estimation, $Y(s) = m(s) + e(s)$
**MSSD**  Spatial interpolation, $Y(s) = m(s) + e(s)$

General-purpose method to design sample configurations

**Space**  Attribute, variogram, and geographic spaces.
**Algorithm**  $SPAN = w_1\text{ACDC} + w_2\text{PPL} + w_3\text{MSSD}$
**Goal**  Uniformly cover the feature, variogram and geographic spaces.

---

1[https://CRAN.R-project.org/package=spsann]
Final Thoughts
Spatial sample optimization and simulated annealing are relatively well known techniques.
Optimization of Spatial Samples

Spatial sample optimization and simulated annealing are relatively well-known techniques.

1. Existing sampling algorithms can be improved, but it is not clear if this always translates into improved prediction accuracy.
Optimization of Spatial Samples

Spatial sample optimization and simulated annealing are relatively well known techniques

1. Existing sampling algorithms can be improved, but it is not clear if this always translates into improved prediction accuracy.

2. Larger sample size seems to improve prediction quality irrespective of the sampling algorithm used – is there a limit?
Spatial sample optimization and simulated annealing are relatively well known techniques

1. Existing sampling algorithms can be improved, but it is not clear if this always translates into improved prediction accuracy.

2. Larger sample size seems to improve prediction quality irrespective of the sampling algorithm used – is there a limit?

3. It is not clear what is the best sample configuration for highly nonlinear models such as random forests.
I will be happy to try answering your questions