Introduction to Spatial Modelling

D G Rossiter

Cornell University

October 5, 2022

Outline

1. Concepts of spatial analysis
2. Spatial analysis & modelling
3. The modelling paradigm
4. Example

Table of Contents

Topic: Thinking about spatial analysis

Reference: O'Sullivan, D., & D. Unwin. 2010. *Geographic information analysis*. 2nd ed. Wiley DOI: `https://dx.doi.org/10.1002/9780470549094/`

Cornell access to e-book: `https://onlinelibrary-wiley-com.proxy.library.cornell.edu/doi/book/10.1002/9780470549094`

See also: Hijmans, R. J. (2021). R Companion to O'Sullivan and Unwin.
`https://rspatial.org/terra/rosu/ROSU.pdf`

O'Sullivan & Unwin's classification

Four concepts:

1. Spatial data **manipulation**
2. Spatial data **description and exploration**[1]
3. Spatial **statistical analysis**
   - Can a statistical model represent the data?
   - This is not yet understanding, only summarizing as an empirical relation.
   - Requires **special techniques to account for spatial relations** → this course!
4. Spatial **modelling**
   - Understand **functional form** of spatial **processes**
   - Contribute to understanding the process itself (requires reasoning outside of statistics)
   - **Predict** spatial outcomes

[1] O'S & U call this "spatial data analysis"

O'Sullivan & Unwin's classification

Four concepts:

1. Spatial data **manipulation**
2. Spatial data **description and exploration**[1]
3. Spatial **statistical analysis**
   - Can a statistical model represent the data?
   - This is not yet understanding, only summarizing as an empirical relation.
   - Requires **special techniques to account for spatial relations** → this course!
4. Spatial **modelling**
   - Understand **functional form** of spatial **processes**
   - Contribute to understanding the process itself (requires reasoning outside of statistics)
   - **Predict** spatial outcomes

[1] O'S & U call this "spatial data analysis"

O'Sullivan & Unwin's classification

Four concepts:

1. Spatial data **manipulation**
2. Spatial data **description and exploration**[1]
3. Spatial **statistical analysis**
   - Can a statistical model represent the data?
   - This is not yet understanding, only summarizing as an empirical relation.
   - Requires **special techniques to account for spatial relations** → this course!
4. Spatial **modelling**
   - Understand **functional form** of spatial **processes**
   - Contribute to understanding the process itself (requires reasoning outside of statistics)
   - **Predict** spatial outcomes

[1] O'S & U call this "spatial data analysis"

O'Sullivan & Unwin's classification

Four concepts:

1. Spatial data **manipulation**
2. Spatial data **description and exploration**[1]
3. Spatial **statistical analysis**
   - Can a statistical model represent the data?
   - This is not yet understanding, only summarizing as an empirical relation.
   - Requires **special techniques to account for spatial relations** → this course!
4. Spatial **modelling**
   - Understand **functional form** of spatial **processes**
   - Contribute to understanding the process itself (requires reasoning outside of statistics)
   - **Predict** spatial outcomes

[1] O'S & U call this "spatial data analysis"

Table of Contents

Topic: Spatial analysis & modelling

- Abstracting and modelling some aspect of a **spatial reality**
  - Natural resources
  - Built environment
  - Social environment
  - Conceptual enviroment (e.g., political divisions)
- Does **not** include modelling objects in space without somehow considering their spatial position, i.e., pure feature-space analysis
  - Just displaying the results of a feature-space model on a map does not make a spatial analysis

Topic: Spatial analysis & modelling

- Abstracting and modelling some aspect of a **spatial reality**
  - Natural resources
  - Built environment
  - Social environment
  - Conceptual enviroment (e.g., political divisions)
- Does **not** include modelling objects in space without somehow considering their spatial position, i.e., pure feature-space analysis
  - Just displaying the results of a feature-space model on a map does not make a spatial analysis

Why model?

- **hypothesis formulation**: a successful model suggests that it realistically represents some real process
- **hypothesis testing**: if we can reproduce some spatial phenomenon with our model based on the hypothesis, the support for the hypothesis is increased
  - any complicated hypothesis is not tested as such, we build up evidence to support, refute or modify it
- **understanding** of "nature": a successful model increases our confidence that the model structure matches the real structure
- spatial(-temporal) **prediction**: the model results in a map which is then used for **decision-making**
- **scenario** analysis: "what if...", mainly for decision-making under **uncertainty**

Why model?

- **hypothesis formulation**: a successful model suggests that it realistically represents some real process
- **hypothesis testing**: if we can reproduce some spatial phenomenon with our model based on the hypothesis, the support for the hypothesis is increased
  - any complicated hypothesis is not tested as such, we build up evidence to support, refute or modify it
- **understanding** of "nature": a successful model increases our confidence that the model structure matches the real structure
- spatial(-temporal) **prediction**: the model results in a map which is then used for **decision-making**
- **scenario** analysis: "what if…", mainly for decision-making under **uncertainty**

Why model?

- **hypothesis formulation**: a successful model suggests that it realistically represents some real process
- **hypothesis testing**: if we can reproduce some spatial phenomenon with our model based on the hypothesis, the support for the hypothesis is increased
  - any complicated hypothesis is not tested as such, we build up evidence to support, refute or modify it
- **understanding** of "nature": a successful model increases our confidence that the model structure matches the real structure
- spatial(-temporal) **prediction**: the model results in a map which is then used for **decision-making**
- **scenario** analysis: "what if...", mainly for decision-making under **uncertainty**

Why model?

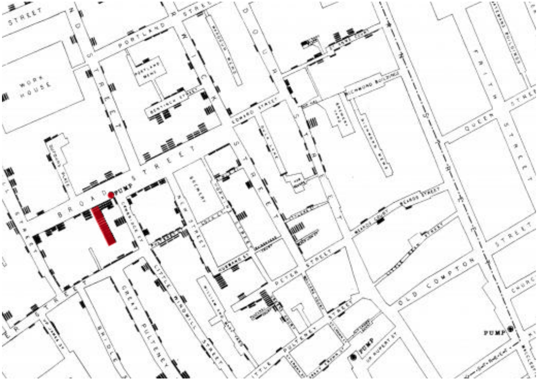
- **hypothesis formulation**: a successful model suggests that it realistically represents some real process
- **hypothesis testing**: if we can reproduce some spatial phenomenon with our model based on the hypothesis, the support for the hypothesis is increased
  - any complicated hypothesis is not tested as such, we build up evidence to support, refute or modify it
- **understanding** of "nature": a successful model increases our confidence that the model structure matches the real structure
- spatial(-temporal) **prediction**: the model results in a map which is then used for **decision-making**
- **scenario** analysis: "what if...", mainly for decision-making under **uncertainty**

Why model?

- **hypothesis formulation**: a successful model suggests that it realistically represents some real process
- **hypothesis testing**: if we can reproduce some spatial phenomenon with our model based on the hypothesis, the support for the hypothesis is increased
  - any complicated hypothesis is not tested as such, we build up evidence to support, refute or modify it
- **understanding** of "nature": a successful model increases our confidence that the model structure matches the real structure
- spatial(-temporal) **prediction**: the model results in a map which is then used for **decision-making**
- **scenario** analysis: "what if...", mainly for decision-making under **uncertainty**

Example – hypothesis formulation

Observation: clustering of cholera, relation to water sources
Hypothesis: cholera is an infectious disease, caused by an organism which lives in wastewater and cycles through humans.



source: Snow, John. *On the Mode of Communication of Cholera*, 2nd Ed, John Churchill, London, 1855

Types of models – 1

Physical: capture the essential behaviour of a physical system with equations; also called **mechanistic**

Empirical: determine relation between system components, without necessarily knowing the cause

In practice the line is blurry:

- Physical principles are often used to motivate choice of variables in empirical models
- "Physical" models have many parameters that must be empirically calibrated (very rarely deriveable from first principles)

Types of models – 1

Physical: capture the essential behaviour of a physical system with equations; also called **mechanistic**

Empirical: determine relation between system components, without necessarily knowing the cause

In practice the line is blurry:

- Physical principles are often used to motivate choice of variables in empirical models
- "Physical" models have many parameters that must be empirically calibrated (very rarely deriveable from first principles)

Types of models – 1

Physical capture the essential behaviour of a physical system with equations; also called **mechanistic**

Empirical determine relation between system components, without necessarily knowing the cause

In practice the line is blurry:

- Physical principles are often used to motivate choice of variables in empirical models
- "Physical" models have many parameters that must be empirically calibrated (very rarely deriveable from first principles)

Types of models – 2

Explanatory: the **main** purpose of building the model is to **understand the process** which gave rise to the object of study

- e.g., ecological factors controlling species distribution, based on observations of the species and co-variables

Predictive: the **main** purpose of building the model is to **predict** at unsampled locations / times (especially the future)

- e.g., identify areas to prioritize for habitat restoration of endangered species

Some models built for one purpose are (somewhat) useful for the other.

Types of models – 2

Explanatory: the **main** purpose of building the model is to **understand the process** which gave rise to the object of study

- e.g., ecological factors controlling species distribution, based on observations of the species and co-variables

Predictive: the **main** purpose of building the model is to **predict** at unsampled locations / times (especially the future)

- e.g., identify areas to prioritize for habitat restoration of endangered species

Some models built for one purpose are (somewhat) useful for the other.

Types of models – 2

Explatory the **main** purpose of building the model is to **understand the process** which gave rise to the object of study

- e.g., ecological factors controlling species distribution, based on observations of the species and co-variables

Predictive the **main** purpose of building the model is to **predict** at unsampled locations / times (especially the future)

- e.g., identify areas to prioritize for habitat restoration of endangered species

Some models built for one purpose are (somewhat) useful for the other.

Modelled geographical objects

All of these can be **model inputs** or **model outputs**:

- points (locations, possibly with attributes)
- lines (same)
- polygons / groups of polygons (same)
- continuous fields (described mathematically or discretized)

Feature-space **attributes** linked to the geographic features may be part of the model

Examples of spatial models

- Distribution of rare species in a forest (**point** pattern, no attributes)
  1. relation to spatially-distributed ecological factors (feature space, but distributed in geographic space)
  2. purely spatial relations, e.g. seed dispersal, allelopathy ...
- Spread of a disease epidemic
  - point cases, possibly with attributes e.g. age, gender, previous health
  - point or line water sources, possibly with attributes, e.g., water quality
  - point pollution sources, possibly with attributes
  - continuous fields, e.g. soil permeability or hydraulic conductivity
- Distribution of pollutants in soil or groundwater (continuous **field**)
- "Optimal" location of a new school (etc.), considering spatial factors

Examples of spatial models

- Distribution of rare species in a forest (**point** pattern, no attributes)
  1. relation to spatially-distributed ecological factors (feature space, but distributed in geographic space)
  2. purely spatial relations, e.g. seed dispersal, allelopathy ...
- Spread of a disease epidemic
  - point cases, possibly with attributes e.g. age, gender, previous health
  - point or line water sources, possibly with attributes, e.g., water quality
  - point pollution sources, possibly with attributes
  - continuous fields, e.g. soil permeability or hydraulic conductivity
- Distribution of pollutants in soil or groundwater (continuous **field**)
- "Optimal" location of a new school (etc.), considering spatial factors

Examples of spatial models

- Distribution of rare species in a forest (**point** pattern, no attributes)
    1. relation to spatially-distributed ecological factors (feature space, but distributed in geographic space)
    2. purely spatial relations, e.g. seed dispersal, allelopathy ...
- Spread of a disease epidemic
    - point cases, possibly with attributes e.g. age, gender, previous health
    - point or line water sources, possibly with attributes, e.g., water quality
    - point pollution sources, possibly with attributes
    - continuous fields, e.g. soil permeability or hydraulic conductivity
- Distribution of pollutants in soil or groundwater (continuous **field**)
- "Optimal" location of a new school (etc.), considering spatial factors

Examples of spatial models

- Distribution of rare species in a forest (**point** pattern, no attributes)
  1. relation to spatially-distributed ecological factors (feature space, but distributed in geographic space)
  2. purely spatial relations, e.g. seed dispersal, allelopathy ...
- Spread of a disease epidemic
  - point cases, possibly with attributes e.g. age, gender, previous health
  - point or line water sources, possibly with attributes, e.g., water quality
  - point pollution sources, possibly with attributes
  - continuous fields, e.g. soil permeability or hydraulic conductivity
- Distribution of pollutants in soil or groundwater (continuous **field**)
- "Optimal" location of a new school (etc.), considering spatial factors

Types of processes being modelled – 1

The idea is to **match** the **model** with a **true process** that caused the observations. Types of processes:

- **fluxes** (flows) driven by "physics": e.g., diffusion, convection / advection, radiation
  - These can be in 1D (e.g., along a river, through a road network), 2D (e.g., hillslope), 3D (e.g., soil volume above groundwater).
  - Non-physical processes, e.g., population migrations, could be modelled by physical equations, if assumptions are met.
- **physical processes**: e.g., plant growth affected by heat, light, nutrients, competition . . .
- **population dynamics**: e.g., birth / death, preditor / prey, cooperation / competition

Types of processes being modelled – 1

The idea is to **match** the **model** with a **true process** that caused the observations. Types of processes:

- **fluxes** (flows) driven by "physics": e.g., diffusion, convection / advection, radiation
  - These can be in 1D (e.g., along a river, through a road network), 2D (e.g., hillslope), 3D (e.g., soil volume above groundwater).
  - Non-physical processes, e.g., population migrations, could be modelled by physical equations, if assumptions are met.
- **physical processes**: e.g., plant growth affected by heat, light, nutrients, competition . . .
- **population dynamics**: e.g., birth / death, preditor / prey, cooperation / competition

Types of processes being modelled – 1

The idea is to **match** the **model** with a **true process** that caused the observations. Types of processes:

- **fluxes** (flows) driven by "physics": e.g., diffusion, convection / advection, radiation
  - These can be in 1D (e.g., along a river, through a road network), 2D (e.g., hillslope), 3D (e.g., soil volume above groundwater).
  - Non-physical processes, e.g., population migrations, could be modelled by physical equations, if assumptions are met.
- **physical processes**: e.g., plant growth affected by heat, light, nutrients, competition . . .
- **population dynamics**: e.g., birth / death, preditor / prey, cooperation / competition

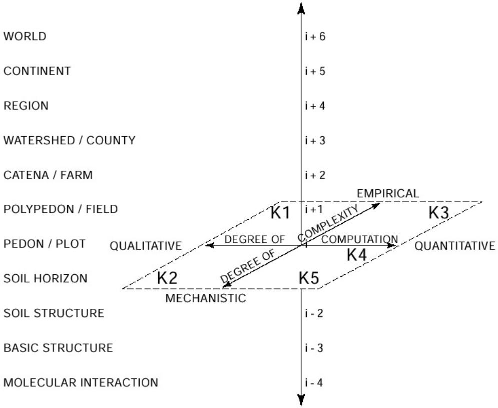
Types of processes being modelled – 2

- **“intelligent” agents** making decisions and interacting
- **decisions** (maybe under uncertainty): try to reproduce the decision-maker’s logic and criteria (e.g., site selection)

Types of processes being modelled – 2

- **“intelligent” agents** making decisions and interacting
- **decisions** (maybe under uncertainty): try to reproduce the decision-maker’s logic and criteria (e.g., site selection)

Three-axis model classification



after Bouma (1999) *Land evaluation for landscape units* In Sumner M. E. (Ed.), Handbook of soil science (pp. E393- E412). Boca Raton, FL: CRC Press

Based on Hoosbeek, M. R., & Bryant, R. B. (1992). Towards the quantitative modeling of pedogenesis – a review. Geoderma, 55, 183-210.

Most models have components spread through this diagram. They must be linked, but this brings many problems of concepts / scales.

Three-axis classification

- **degree of complexity**: Mechanistic vs. empirical – see above
- **scale** – see below
- **degree of computation**: Qualitative vs. quantitative

Degree of computation: algorithm / outputs more or less **quantified**

- e.g., "highly suitable" vs. "Net present value for intensive vegetable production \$1000 ha^{-1}"

Table of Contents

Steps in modelling

These apply to any sort of model; the terminology here is mostly from empirical-statistical ("regression") models.

1. Selecting a **functional form**, i.e. the model to be fitted;
   - may try several forms; but these should be "reasonable", based on possible mechanisms
2. Determining the **parameters** of the model; this is called **model calibration** or **parameter estimation**;
3. Determining how well the model describes reality; this is called **model evaluation**[2]
4. **Criticising** (re-examining) the assumptions and possibly re-cycling.
   - may lead to a modified or completely different **model form**

[2] often (erroneously) called **model validation**, see Oreskes (1998) https://dx.doi.org/10.1289/ehp.98106s61453

Steps in modelling

These apply to any sort of model; the terminology here is mostly from empirical-statistical ("regression") models.

1. Selecting a **functional form**, i.e. the model to be fitted;
   - may try several forms; but these should be "reasonable", based on possible mechanisms
2. Determining the **parameters** of the model; this is called **model calibration** or **parameter estimation**;
3. Determining how well the model describes reality; this is called **model evaluation**[2]
4. **Criticising** (re-examining) the assumptions and possibly re-cycling.
   - may lead to a modified or completely different **model form**

[2] often (erroneously) called **model validation**, see Oreskes (1998) https://dx.doi.org/10.1289/ehp.98106s61453

Steps in modelling

These apply to any sort of model; the terminology here is mostly from empirical-statistical ("regression") models.

1. Selecting a **functional form**, i.e. the model to be fitted;
   - may try several forms; but these should be "reasonable", based on possible mechanisms
2. Determining the **parameters** of the model; this is called **model calibration** or **parameter estimation**;
3. Determining how well the model describes reality; this is called **model evaluation**[2]
4. **Criticising** (re-examining) the assumptions and possibly re-cycling.
   - may lead to a modified or completely different **model form**

[2] often (erroneously) called **model validation**, see Oreskes (1998) https://dx.doi.org/10.1289/ehp.98106s61453

Steps in modelling

These apply to any sort of model; the terminology here is mostly from empirical-statistical ("regression") models.

1. Selecting a **functional form**, i.e. the model to be fitted;
   - may try several forms; but these should be "reasonable", based on possible mechanisms
2. Determining the **parameters** of the model; this is called **model calibration** or **parameter estimation**;
3. Determining how well the model describes reality; this is called **model evaluation**[2]
4. **Criticising** (re-examining) the assumptions and possibly re-cycling.
   - may lead to a modified or completely different **model form**

[2] often (erroneously) called **model validation**, see Oreskes (1998) https://dx.doi.org/10.1289/ehp.98106s61453

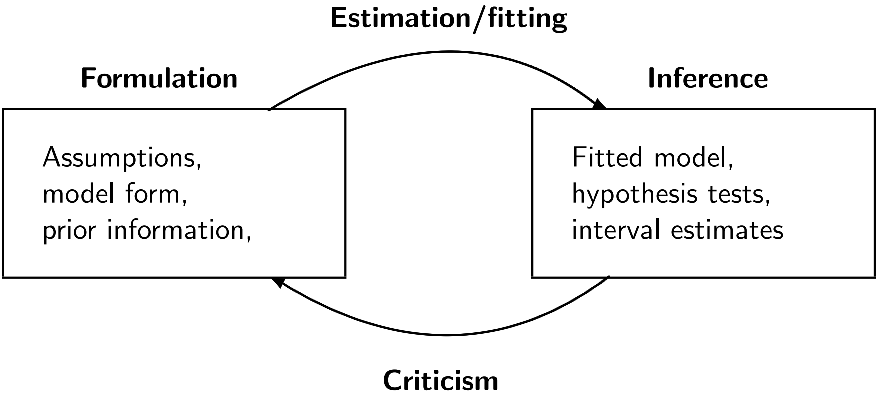
Examples of functional forms

Grain yield of a cereal crop as affected by N fertilizer:

1. **linear**: one unit of fertilizer is β units of grain yield, throughout the range;
2. **linear response with threshold**: same till λ units of N, then reaches a plateau;
3. **quadratic**: one unit of fertilizer is $\beta_1 + {\beta_2}^2$ units of grain yield, throughout the range; β_2 is negative so after a certain point yield decreases;
4. **negative exponential**: yield increases asymptotically to some limit μ at some effective range ρ.

– The Greek letters indicate **parameters** that must be fit by **calibration**
– All of these have a **plausible physical basis** within a **range of applicability**

The modelling paradigm



adapted from Cook & Weisberg (1982) *Residuals and influence in regression*
ISBN 978-0-412-24280-9

Structure vs. noise

- **Reality**
  - Reality $= f(\text{Structure, Noise})$
  - Reality $= f(\text{deterministic } \textbf{processes}, \textbf{ random } \text{variation})$
- **Observations**
  - Observations $= f(\text{Structure, Noise})$
  - Observations $= f(\text{model, unexplained variation})$
- **Observations** are a subset of **Reality**
  - The aim is to match our **model** with the true **deterministic process**
  - ... and match our estimate of the **noise** with the true **random variation**.
- It is equally an error to model the noise (**overfit** the model) as to not model the process (**underfit** the model).

Structure vs. noise

- **Reality**
  - Reality $= f$(Structure, Noise)
  - Reality $= f$(deterministic **processes**, **random** variation)
- **Observations**
  - Observations $= f$(Structure, Noise)
  - Observations $= f$(model, unexplained variation)
- **Observations** are a subset of **Reality**
  - The aim is to match our **model** with the true **deterministic process**
  - ... and match our estimate of the **noise** with the true **random variation**.
- It is equally an error to model the noise (**overfit** the model) as to not model the process (**underfit** the model).

Structure vs. noise

- **Reality**
  - Reality $= f$(Structure, Noise)
  - Reality $= f$(deterministic **processes**, **random** variation)
- **Observations**
  - Observations $= f$(Structure, Noise)
  - Observations $= f$(model, unexplained variation)
- **Observations** are a subset of **Reality**
  - The aim is to match our **model** with the true **deterministic process**
  - ...and match our estimate of the **noise** with the true **random variation**.
- It is equally an error to model the noise (**overfit** the model) as to not model the process (**underfit** the model).

Structure vs. noise

- **Reality**
  - Reality $= f$(Structure, Noise)
  - Reality $= f$(deterministic **processes**, **random** variation)
- **Observations**
  - Observations $= f$(Structure, Noise)
  - Observations $= f$(model, unexplained variation)
- **Observations** are a subset of **Reality**
  - The aim is to match our **model** with the true **deterministic process**
  - ...and match our estimate of the **noise** with the true **random variation**.
- It is equally an error to model the noise (**overfit** the model) as to not model the process (**underfit** the model).

What is “noise”?

- true randomness, chaos . . . , or
- structure we don’t suspect?
- structure that we don’t understand, or
- structure we understand but can’t model, or
- these are represented in the same way in the model formulation (“error”, “residual”)

What is "noise"?

- true randomness, chaos . . . , or
- structure we don't suspect?
- structure that we don't understand, or
- structure we understand but can't model, or
- these are represented in the same way in the model formulation ("error", "residual")

What is "noise"?

- true randomness, chaos . . . , or
- structure we don't suspect?
- structure that we don't understand, or
- structure we understand but can't model, or
- these are represented in the same way in the model formulation ("error", "residual")

What is "noise"?

- true randomness, chaos . . . , or
- structure we don't suspect?
- structure that we don't understand, or
- structure we understand but can't model, or
- these are represented in the same way in the model formulation ("error", "residual")

What is "noise"?

- true randomness, chaos . . . , or
- structure we don't suspect?
- structure that we don't understand, or
- structure we understand but can't model, or
- these are represented in the same way in the model formulation ("error", "residual")

Evidence that a model is suitable

Two levels of evidence:

1. **external** to the model:
   1. what is known about the **process** that gave rise to the data
      - this is the connection to the **reality** that the model is trying to explain or summarise;
   2. how well the model fits further data from the same population: success of **statistical evaluation** against an independent dataset
2. **internal** to the model:
   1. how well the model fits the data (success of **calibration**);
      - danger of over-fitting: fitting to "structure" in this dataset, not the population
   2. how well the model fits subsets of the data not used for model building (**cross-validation**);
   3. how well the fitted model meets the **assumptions** of that functional form (e.g. examination of regression diagnostics).

Evidence that a model is suitable

Two levels of evidence:

1. **external** to the model:
   1. what is known about the **process** that gave rise to the data
      - ★ this is the connection to the **reality** that the model is trying to explain or summarise;
   2. how well the model fits further data from the same population: success of **statistical evaluation** against an independent dataset
2. **internal** to the model:
   1. how well the model fits the data (success of **calibration**);
      - ★ danger of over-fitting: fitting to "structure" in this dataset, not the population
   2. how well the model fits subsets of the data not used for model building (**cross-validation**);
   3. how well the fitted model meets the **assumptions** of that functional form (e.g. examination of regression diagnostics).

Table of Contents

Example of a spatial modelling exercise

- **Problem**: soil contamination by heavy metals in flood plain of the Maas (Meuse) River near Stein (L), Netherlands
- **Objectives**:
  1. determine where metals came from (**explanation**). Hypotheses:
     - completely random?
     - pre-industrial river sediments, natural erosion from upstream soils?
     - recent atmospheric deposition from industry?
     - sediments deposited by recent floods from upstream industrial areas (B, F)?
     - agricultural practices (e.g., Cu-bearing fungicides)
  2. Map concentrations over an area to zone/regulate land use (**prediction**)

Example of a spatial modelling exercise

- **Problem**: soil contamination by heavy metals in flood plain of the Maas (Meuse) River near Stein (L), Netherlands
- **Objectives**:
  1. determine where metals came from (**explanation**). Hypotheses:
     - completely random?
     - pre-industrial river sediments, natural erosion from upstream soils?
     - recent atmospheric deposition from industry?
     - sediments deposited by recent floods from upstream industrial activity (B, F)?
     - agricultural practices (e.g., Cu-bearing fungicides)
  2. Map concentrations over an area to zone/regulate land use (**prediction**)

Example of a spatial modelling exercise

- **Problem**: soil contamination by heavy metals in flood plain of the Maas (Meuse) River near Stein (L), Netherlands
- **Objectives**:
  1. determine where metals came from (**explanation**). Hypotheses:
     - completely random?
     - pre-industrial river sediments, natural erosion from upstream soils?
     - recent atmospheric deposition from industry?
     - sediments deposited by recent floods from upstream industrial activity (B, F)?
     - agricultural practices (e.g., Cu-bearing fungicides)
  2. Map concentrations over an area to zone/regulate land use (**prediction**)

Example of a spatial modelling exercise

- **Problem**: soil contamination by heavy metals in flood plain of the Maas (Meuse) River near Stein (L), Netherlands
- **Objectives**:
  1. determine where metals came from (**explanation**). Hypotheses:
     - completely random?
     - pre-industrial river sediments, natural erosion from upstream soils?
     - recent atmospheric deposition from industry?
     - sediments deposited by recent floods from upstream industrial activity (B, F)?
     - agricultural practices (e.g., Cu-bearing fungicides)
  2. Map concentrations over an area to zone/regulate land use (**prediction**)

Example of a spatial modelling exercise

- **Problem**: soil contamination by heavy metals in flood plain of the Maas (Meuse) River near Stein (L), Netherlands
- **Objectives**:
  1. determine where metals came from (**explanation**). Hypotheses:
     - completely random?
     - pre-industrial river sediments, natural erosion from upstream soils?
     - recent atmospheric deposition from industry?
     - sediments deposited by recent floods from upstream industrial activity (B, F)?
     - agricultural practices (e.g., Cu-bearing fungicides)
  2. Map concentrations over an area to zone/regulate land use (**prediction**)

Example of a spatial modelling exercise

- **Problem**: soil contamination by heavy metals in flood plain of the Maas (Meuse) River near Stein (L), Netherlands
- **Objectives**:
  1. determine where metals came from (**explanation**). Hypotheses:
     - ★ completely random?
     - ★ pre-industrial river sediments, natural erosion from upstream soils?
     - ★ recent atmospheric deposition from industry?
     - ★ sediments deposited by recent floods from upstream industrial activity (B, F)?
     - ★ agricultural practices (e.g., Cu-bearing fungicides)
  2. Map concentrations over an area to zone/regulate land use (**prediction**)

Example of a spatial modelling exercise

- **Problem**: soil contamination by heavy metals in flood plain of the Maas (Meuse) River near Stein (L), Netherlands
- **Objectives**:
  1. determine where metals came from (**explanation**). Hypotheses:
     - completely random?
     - pre-industrial river sediments, natural erosion from upstream soils?
     - recent atmospheric deposition from industry?
     - sediments deposited by recent floods from upstream industrial activity (B, F)?
     - agricultural practices (e.g., Cu-bearing fungicides)
  2. Map concentrations over an area to zone/regulate land use (**prediction**)

Example of a spatial modelling exercise

- **Problem**: soil contamination by heavy metals in flood plain of the Maas (Meuse) River near Stein (L), Netherlands
- **Objectives**:
  1. determine where metals came from (**explanation**). Hypotheses:
     - completely random?
     - pre-industrial river sediments, natural erosion from upstream soils?
     - recent atmospheric deposition from industry?
     - sediments deposited by recent floods from upstream industrial activity (B, F)?
     - agricultural practices (e.g., Cu-bearing fungicides)
  2. Map concentrations over an area to zone/regulate land use (**prediction**)

Example of a spatial modelling exercise

- **Problem**: soil contamination by heavy metals in flood plain of the Maas (Meuse) River near Stein (L), Netherlands
- **Objectives**:
  1. determine where metals came from (**explanation**). Hypotheses:
     - completely random?
     - pre-industrial river sediments, natural erosion from upstream soils?
     - recent atmospheric deposition from industry?
     - sediments deposited by recent floods from upstream industrial activity (B, F)?
     - agricultural practices (e.g., Cu-bearing fungicides)
  2. Map concentrations over an area to zone/regulate land use (**prediction**)

Observations

- A set of **"point"** observations: soil samples, heavy metal concentrations measured in the laboratory
- Location (coördinates) of each observation are known
- Some other attributes could be recorded at each observation point (e.g., land use, soil type, elevation)

Observations

- A set of **"point"** observations: soil samples, heavy metal concentrations measured in the laboratory
- Location (coördinates) of each observation are known
- Some other attributes could be recorded at each observation point (e.g., land use, soil type, elevation)

Observations

- A set of **“point”** observations: soil samples, heavy metal concentrations measured in the laboratory
- Location (coördinates) of each observation are known
- Some other attributes could be recorded at each observation point (e.g., land use, soil type, elevation)

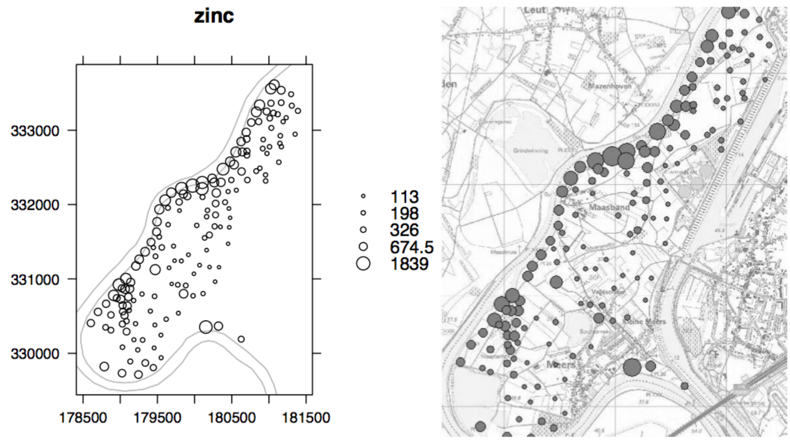


Fig. 5.2: Meuse data set and values of zinc (ppm): visualized in R (left), and in SAGA GIS (right).

Source: Hengl, T. (2009). *A Practical Guide to Geostatistical Mapping*, `http://spatial-analyst.net/book/`

Model forms - 1

Trend surface: systematic variation with geographic coördinates

- **theory**: spatial variation in metal concentration is the result of a **spatial trend**
  - e.g., atmospheric deposition from an upwind point source
- **model**: linear regression on coördinates (E, N)

Model forms - 1

Trend surface: systematic variation with geographic coördinates

- **theory**: spatial variation in metal concentration is the result of a **spatial trend**
  - e.g., atmospheric deposition from an upwind point source
- **model**: linear regression on coördinates (E, N)

Model forms - 2

Interpolation

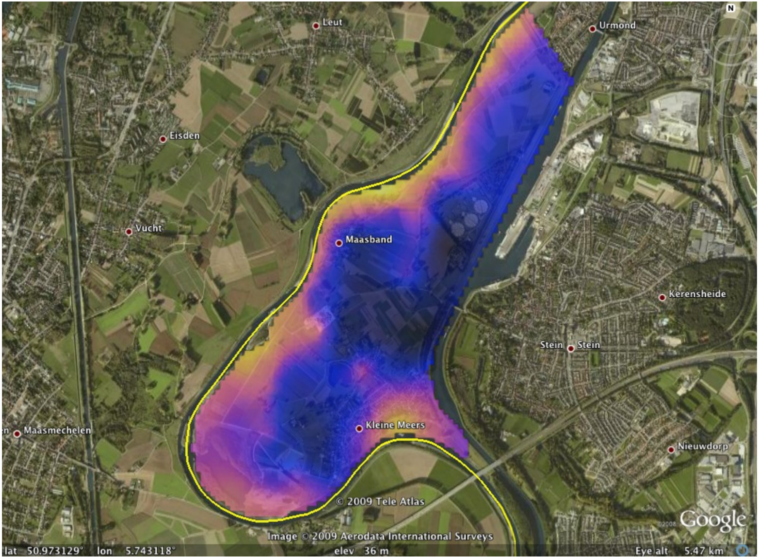
- **theory**: spatial variation in metal concentration is the result of a **spatially-autocorrelated random process**
  - purely local processes, at some local scale
- **model**: fit variogram model, predict by Ordinary kriging (OK)

Model forms - 2

Interpolation

- **theory**: spatial variation in metal concentration is the result of a **spatially-autocorrelated random process**
  - purely local processes, at some local scale
- **model**: fit variogram model, predict by Ordinary kriging (OK)

Ordinary Kriging prediction: Zn concentration



Model forms - 3

Data-driven, using machine-learning methods

- **theory**: spatial variation in metal concentration can be explained in feature space, no trend or residual local processes
- **Feature space** model: metal vs. covariables: flooding frequency, soil type, land use, elevation . . .
  - predictor variables selected because of suspected relation to the target variables
  - values of predictor variables must be known at the observation points
  - to use the model in prediction, they must be known over the entire area.
  - using a machine-learning, data-driven model, e.g., regression trees, random forest, gradient boosting . . .

Model forms - 3

Data-driven, using machine-learning methods

- **theory**: spatial variation in metal concentration can be explained in feature space, no trend or residual local processes
- **Feature space** model: metal vs. covariables: flooding frequency, soil type, land use, elevation ...
  - predictor variables selected because of suspected relation to the target variables
  - values of predictor variables must be known at the observation points
  - to use the model in prediction, they must be known over the entire area.
  - using a machine-learning, data-driven model, e.g., regression trees, random forest, gradient boosting ...

Model forms - 4

Mixed model: combine feature and geographic space, e.g., using Regression kriging (RK)

1. **theory**: some spatial variation can be explained by feature space covariables, some to local processes
2. **Feature space** model: metal vs. covariables: flooding frequency, soil type, land use, elevation . . .
3. **Geographic space** model: local spatial dependence of **residuals** from feature-space model
    - include distance from river in "feature space" model
    - OK of the residuals, i.e., variation not explained by feature-space model

Model forms - 4

Mixed model: combine feature and geographic space, e.g., using Regression kriging (RK)

1. **theory**: some spatial variation can be explained by feature space covariables, some to local processes
2. **Feature space** model: metal vs. covariables: flooding frequency, soil type, land use, elevation . . .
3. **Geographic space** model: local spatial dependence of **residuals** from feature-space model
   - include distance from river in "feature space" model
   - OK of the residuals, i.e., variation not explained by feature-space model

Model forms - 4

Mixed model: combine feature and geographic space, e.g., using Regression kriging (RK)

1. **theory**: some spatial variation can be explained by feature space covariables, some to local processes
2. **Feature space** model: metal vs. covariables: flooding frequency, soil type, land use, elevation . . .
3. **Geographic space** model: local spatial dependence of **residuals** from feature-space model
   - include distance from river in "feature space" model
   - OK of the residuals, i.e., variation not explained by feature-space model

What does the calibrated model tell us?

- Success of the model: How well did it explain/predict?
- **Feature** space: What are the most important explanatory predictors? If a regression model, what are their coefficients?
- **Geographic** space: What is the form of the local spatial dependence? What are the parameters, e.g., range of local influence?
- How well does the model fit our theories?
- Should we re-visit our theories?

What does the calibrated model tell us?

- Success of the model: How well did it explain/predict?
- **Feature** space: What are the most important explanatory predictors? If a regression model, what are their coefficients?
- **Geographic** space: What is the form of the local spatial dependence? What are the parameters, e.g., range of local influence?
- How well does the model fit our theories?
- Should we re-visit our theories?

What does the calibrated model tell us?

- Success of the model: How well did it explain/predict?
- **Feature** space: What are the most important explanatory predictors? If a regression model, what are their coefficients?
- **Geographic** space: What is the form of the local spatial dependence? What are the parameters, e.g., range of local influence?
- How well does the model fit our theories?
- Should we re-visit our theories?

What does the calibrated model tell us?

- Success of the model: How well did it explain/predict?
- **Feature** space: What are the most important explanatory predictors? If a regression model, what are their coefficients?
- **Geographic** space: What is the form of the local spatial dependence? What are the parameters, e.g., range of local influence?
- How well does the model fit our theories?
- Should we re-visit our theories?

What does the calibrated model tell us?

- Success of the model: How well did it explain/predict?
- **Feature** space: What are the most important explanatory predictors? If a regression model, what are their coefficients?
- **Geographic** space: What is the form of the local spatial dependence? What are the parameters, e.g., range of local influence?
- How well does the model fit our theories?
- Should we re-visit our theories?

Which model form(s) is/are "best"?

- How to decide?
- What criteria?
- Evaluated pre/post modelling?

End