On generating spatio-temporal data

Edzer Pebesma, Simon Scheider, Ben Gräler, Christoph Stasch

ifgi
Institute for Geoinformatics
University of Münster

geoinformation engineering
ikg eth zurich

52north
exploring horizons

Wageningen UR, Sep 30, 2015

Environmental Modelling & Software

Volume 51, January 2014, Pages 149–165

Meaningful spatial prediction and aggregation

Christoph Stasch[a], Simon Scheider[a], Edzer Pebesma[a, b], Werner Kuhn[a]

Show more

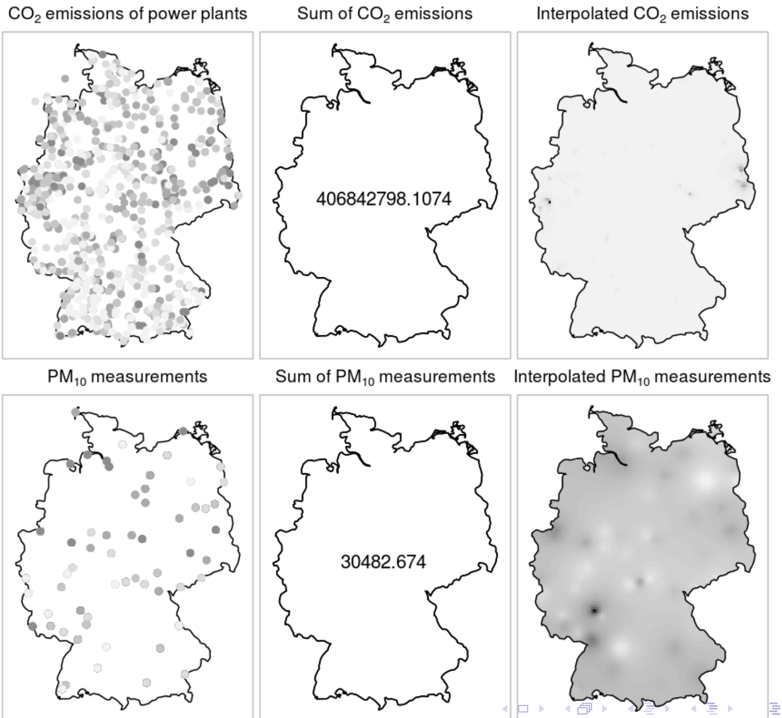
doi:10.1016/j.envsoft.2013.09.006

Highlights

- We introduce a new notion of meaningfulness of spatial prediction and aggregation.
- Observation, prediction, and aggregation procedures are formalized as functions.
- We show how datasets can be described as results of executing such procedures.
- We propose formal checks of meaningfulness based on functional correspondence.

Data for this Article

RunMyCode
Code and data


CO2 emissions of power plants
Sum of CO2 emissions
Interpolated CO2 emissions
406842798.1074
PM10 measurements
Sum of PM10 measurements
Interpolated PM10 measurements
30482.674

International Journal of Geographical Information Science
Vol. 00, No. 00, Month 2009, 1–33

Article under review

Research article

A generative algebra for spatio-temporal information

Simon Scheider[a*], Benedikt Gräler[b], Christoph Stasch[b,c], and Edzer Pebesma[b]

[a] *Institut für Kartographie und Geoinformation, ETH Zürich, Switzerland;* [b] *Institut für Geoinformatik, Universität Münster, Germany;* [c] *52°North GmbH*

(*September 2015*)

Maintaining knowledge about the provenance of data, i.e., about how it was obtained, is crucial for its further use. Contrary to what the overused metaphors of "data mining" and "big data" are implying, it is hardly possible to use

Overview

1. Discovery
2. Provenance
3. What is data?
4. Basic types
5. Data generation procedures
6. Derivation operations
7. Examples: derivation graphs
8. Potential, further work
9. Conclusions

Discovery

How do you discover data?

Discovery

How do you discover data? Why is discovery important?

Discovery

How do you discover data? Why is discovery important? Impact.

Provenance

PROV-O[1]: "Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness.

[1] http://www.w3.org/TR/2012/WD-prov-overview-20121211/

Provenance

PROV-O[1]: "Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness.
The PROV Family of Documents defines a model, corresponding serializations and other supporting defintions to enable the inter-operable interchange of provenance information in heterogeneous environments such as the Web. This document provides an overview this family of documents."

[1] http://www.w3.org/TR/2012/WD-prov-overview-20121211/

What is data?

“data are not just numbers, they are numbers with a context[2]”

[2] George W. Cobb and David S. Moore. "Mathematics, statistics, and teaching." American Mathematical Monthly (1997): 801-823.

What is data?

"data are not just numbers, they are numbers with a context[2]"
To give context, to numbers, we need

- reference systems: SI, units of measurement, datums, calenders, identifiers
- coherence: when/where/what (meaning)
- maybe also: who/why/how (intention)

[2] George W. Cobb and David S. Moore. "Mathematics, statistics, and teaching." American Mathematical Monthly (1997): 801-823.

Basic types

Basic reference system types and simple derivations thereof. Each type needs to go along with its reference system (RS). $\mathcal{P}$ denotes the power set (set of all subsets).

Symbol	Definition	Meaning	Description
S		$\mathbb{R}^3$	Set of possible spatial locations with RS.
T		$\mathbb{R}$	Set of possible moments in time with RS.
D		$\mathbb{N}$	Set of possible discrete entity identifier with RS.
Q		$\mathbb{R}$	Set of possible observed values with RS.
R	S set	$\mathcal{P}(S)$	Set of regions: bounded by polygons, or collection of isolated locations and combinations thereof.
I	T set	$\mathcal{P}(T)$	Set of collections of moments in time: continuous intervals or a set of moments in time or combinations thereof.
D set	D set	$\mathcal{P}(D)$	Sets of object identifiers
Q set	Q set	$\mathcal{P}(Q)$	Sets of quality values.
bool		$\{\mathrm{T}, \mathrm{F}\}$	Boolean, also used to express predicates for selection
Extent	$R \times I$	$R \times I$	set of spatio-temporal extent as the orthogonal product of the spatial and temporal projections
Occurs	$(S \times T)$ set	$\mathcal{P}(S \times T)$	set of spatio-temporal subsets, occurrences of events and objects, but also of certain values or conditions in a field; footprint, support

Transitions

Symbol	Type definition	Description
Select	$\mathrm{Extent} \Rightarrow S \times T$	select the centroid (or alike) of an extent
SSelect	$R \Rightarrow S$	select the centroid of a region
TSelect	$I \Rightarrow T$	select the centroid of a time interval
Tessel	$S \times T \Rightarrow \mathrm{Extent}$	map spatio-temporal locations to their ¿corresponding spatio-temporal extent
STessel	$S \Rightarrow R$	map spatial locations to regions
TTessel	$T \Rightarrow I$	map time stamps to time intervals
QPartition	$Q \Rightarrow Q\ \mathrm{set}$	map quality values to ranges of qualities
Qstat	$(Q \Rightarrow \mathrm{bool}) \Rightarrow Q$	summarize quality values (e.g., mean, median)

Generation procedures: Fields

Symbol	Type definition	Description
Field	$S \times T \Rightarrow Q$	spatio-temporal field
SField	$S \Rightarrow Q$	spatial field
TField	$T \Rightarrow Q$	temporal field (time series)

Generation procedures: Lattices

Symbol	Type definition	Description
Lattice	$R \Rightarrow I \Rightarrow Q$	spatio-temporal lattice
LatticeS	$R \Rightarrow T \Rightarrow Q$	temporal spatial lattice
LatticeT	$S \Rightarrow I \Rightarrow Q$	spatial temporal lattice
SLattice	$R \Rightarrow Q$	spatial lattice
TLattice	$I \Rightarrow Q$	temporal lattice

Generation procedures: Events

Symbol	Type definition	Description
Event	$D \Rightarrow S \times T$	spatio-temporal events
RegionalEvent	$D \Rightarrow R \times T$	events affecting a set of locations
IntervalEvent	$D \Rightarrow S \times I$	events lasting for some time interval
BlockEvent	$D \Rightarrow \mathrm{Extent}$	events affecting a set of locations and lasting for some time interval
SEvents	$D \Rightarrow S$	events' locations
TEvents	$D \Rightarrow T$	events' timestamps
MarkedEvent	$D \Rightarrow S \times T \times Q$	spatio-temporal marked events

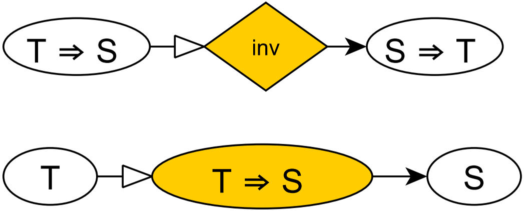
Generation procedures: Trajectories

Symbol	Type definition	Description
Trajectory	$T \Rightarrow S$	trajectory
RegionalTrajectory	$T \Rightarrow R$	trajectory of regions
IntervalTrajectory	$I \Rightarrow S$	trajectory over temporal intervals
BlockTrajectory	$I \Rightarrow R$	trajectory over temporal intervals of regions
MarkedTrajectory	$T \Rightarrow S \times Q$	marked trajectory

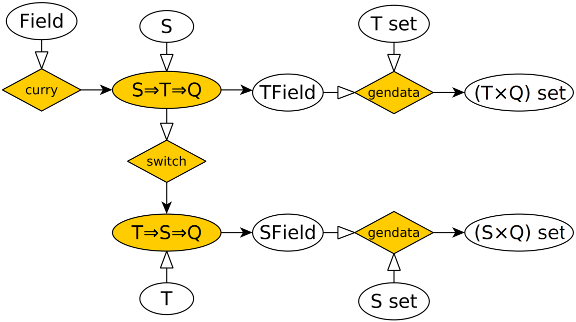
Generation procedures: Objects

Symbol	Definition	Description
Objects	$D \Rightarrow T \Rightarrow S$	objects in time and space
RegionalObjects	$D \Rightarrow T \Rightarrow R$	objects in space and time defined over regions
IntervalObjects	$D \Rightarrow I \Rightarrow S$	objects in time and space defined for collections of moments in time
BlockObjects	$D \Rightarrow I \Rightarrow R$	objects in space and time defined over regions and collections of moments in time
OjectTimeSeries	$D \Rightarrow T \Rightarrow Q$	time series associated with each object
MarkedObjects	$D \Rightarrow T \Rightarrow S \times Q$	marked object trajectories

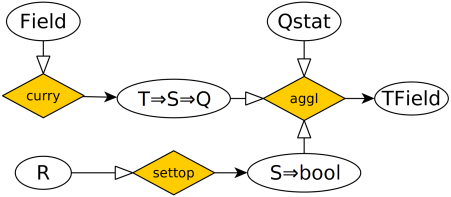
Data derivation

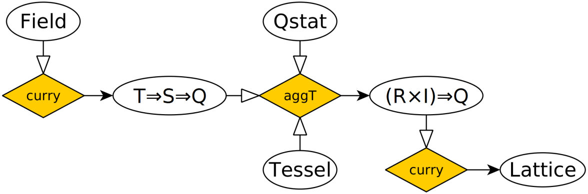


Data derivation: generating field data



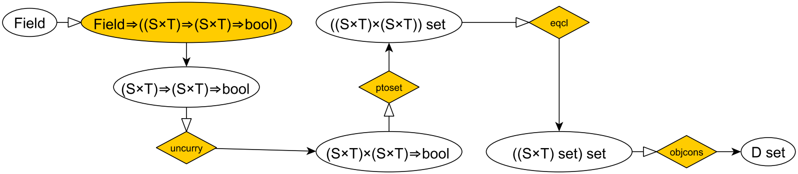
Data derivation: spatial/temporal aggregation

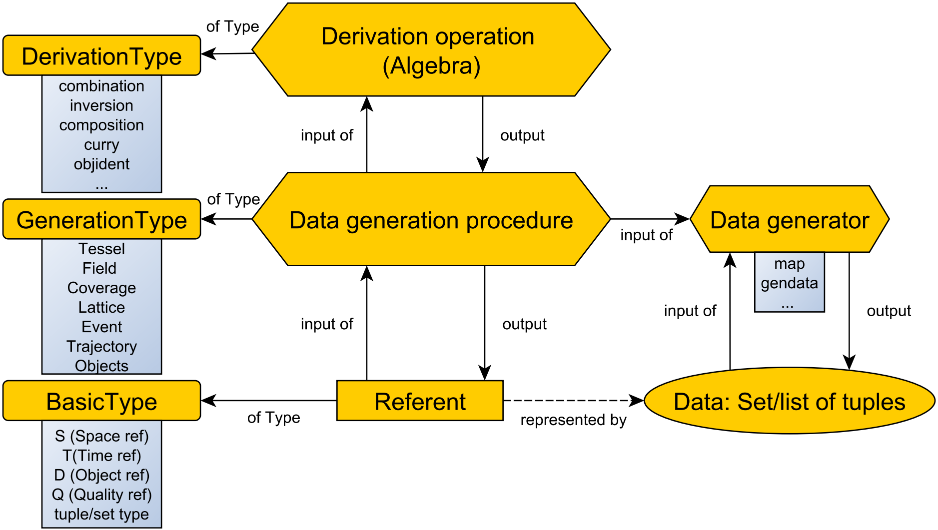




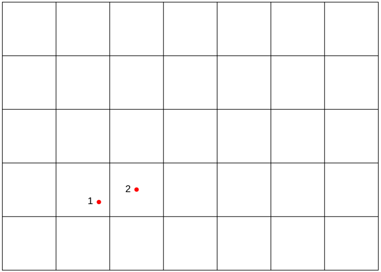
see paper for definitions of curry, aggl, aggT and settop

Data derivation: deriving objects from fields

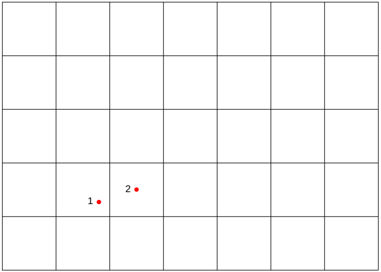


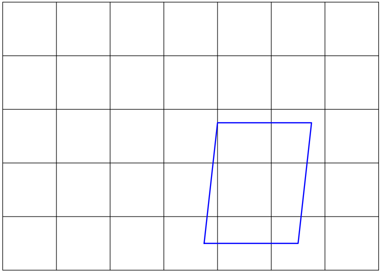

Derivation operation (Algebra)
of Type
DerivationType
combination
inversion
composition
curry
objident
...
input of
output
of Type
GenerationType
Tessel
Field
Coverage
Lattice
Event
Trajectory
Objects
Data generation procedure
input of
Data generator
map
gendata
...
input of
output
input of
output
BasicType
of Type
Referent
represented by
Data: Set/list of tuples
S (Space ref)
T(Time ref)
D (Object ref)
Q (Quality ref)
tuple/set type

Retrieving information from raster maps

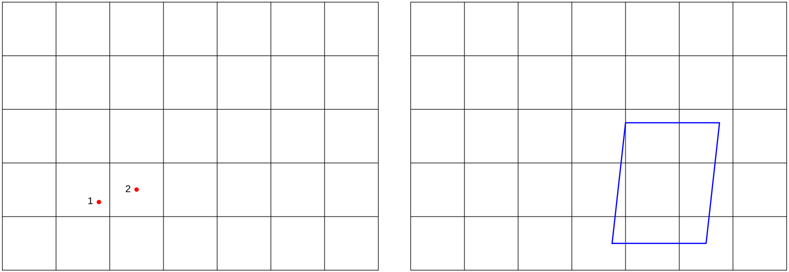


Retrieving information from raster maps





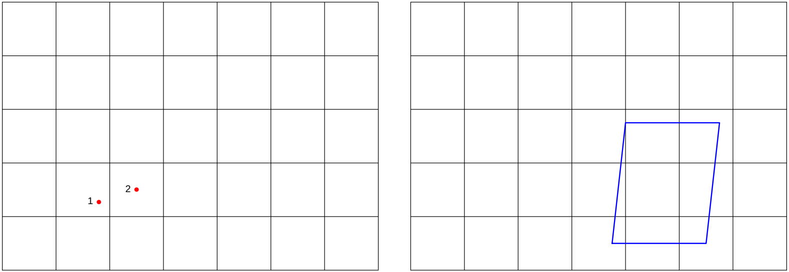
Retrieving information from raster maps



Three simplest cases:

point	cell is point	cell is constant	cell is aggregation
1	NA	cell value	NA
2	cell value	cell value	NA

Retrieving information from raster maps



Three simplest cases:

point	cell is point	cell is constant	cell is aggregation
1	NA	cell value	NA
2	cell value	cell value	NA

how *can* software decide what to do?

Potential, further work

- Discovery:
  - the theory[3] works, but does it solve problems in practice?
  - translate the abstract syntax of our algebra into tools
  - annotate data sets with derivation graphs
  - publish data with derivation graphs
  - develop discovery mechanisms (linked data, annotation tools)

[3] http://www.geographicknowledge.de/vocab/AlgebraReferenceSystems.thy

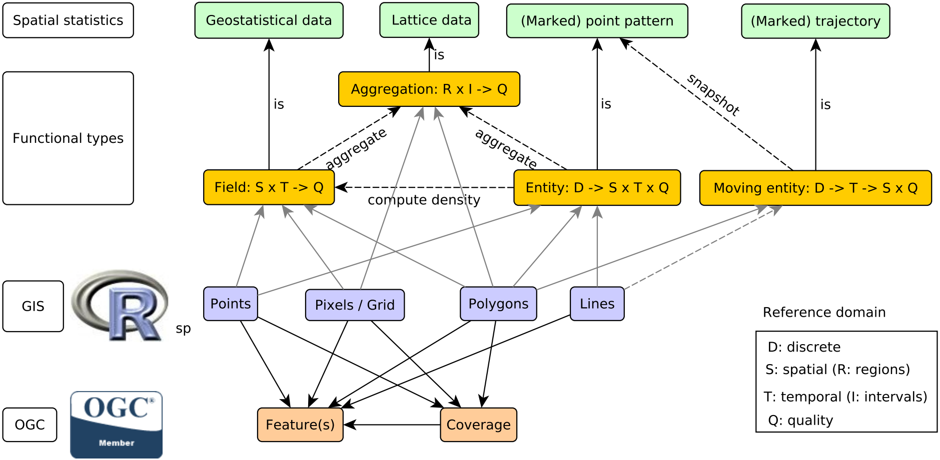
Potential, further work

- Discovery:
  - the theory[3] works, but does it solve problems in practice?
  - translate the abstract syntax of our algebra into tools
  - annotate data sets with derivation graphs
  - publish data with derivation graphs
  - develop discovery mechanisms (linked data, annotation tools)
- Generation:
  - reason about space of possible derivations
  - reason about compatibility
  - develop recommender systems

[3] http://www.geographicknowledge.de/vocab/AlgebraReferenceSystems.thy

Conclusions

- We propose a generative algebra for spatio-temporal information that describes how data is generated in a variety of derivation processes, expressed as derivation graphs.
- Data generation procedures are expressed as functions on basic types S, T, D, Q
- Possible derivations can be expressed as chains of function applications, where each function is either an operation of the algebra or a spatio-temporal data generation procedure.
- Types of data generation include tesselations, fields, coverages, lattices, events, objects, trajectories.
- We illustrate how they can be converted into each other.
- Our algebra can be used for publishing provenance of data sets in terms of a derivation graph and on a level of detail that distinguishes types of spatio-temporal information.
- Our algebra makes explicit the support of data, i.e. whether values refer to aggregated values or constant values over regions or time periods.


Spatial statistics
Geostatistical data
Lattice data
(Marked) point pattern
(Marked) trajectory
Functional types
is
Aggregation: R x I -> Q
is
snapshot
is
is
aggregate
aggregate
Field: S x T -> Q
Entity: D -> S x T x Q
Moving entity: D -> T -> S x Q
compute density
GIS
sp
Points
Pixels / Grid
Polygons
Lines
OGC
Member
Feature(s)
Coverage
Reference domain
D: discrete
S: spatial (R: regions)
T: temporal (I: intervals)
Q: quality