Meaningful Spatial Statistics

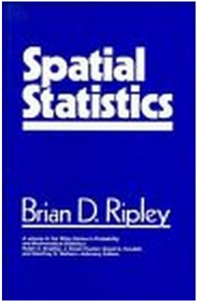
Edzer Pebesma

joint work with Simon Scheider, Ben Gräler, Christoph Stasch

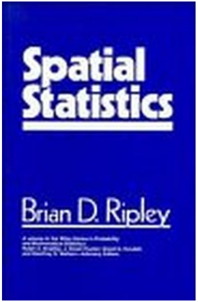
ifgi
Institute for Geoinformatics
University of Münster

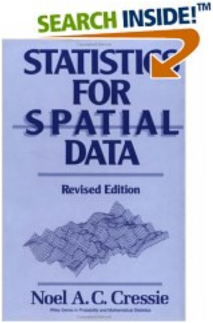
ETH Zürich, Nov 19, 2015

Spatial Statistics is...

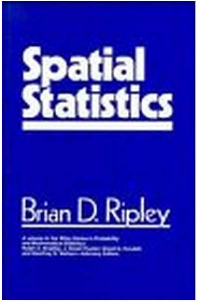


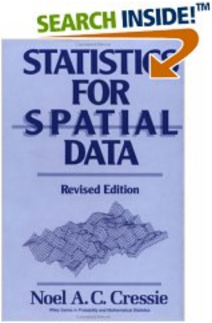
Spatial Statistics is...

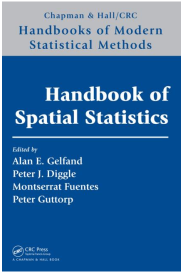




Spatial Statistics is...





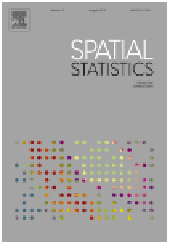


but also...

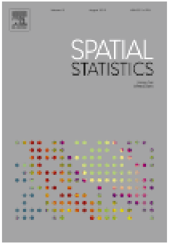

SPATIAL STATISTICS
2015
EMERGING PATTERNS
9-12 JUNE 2015
AVIGNON, FRANCE

but also...


SPATIAL STATISTICS
2015
EMERGING PATTERNS
9-12 JUNE 2015
AVIGNON, FRANCE

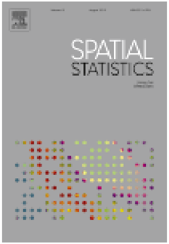

SPATIAL STATISTICS

but also...



SPATIAL STATISTICS SOCIETY

but also...



SPATIAL
STATISTICS
SOCIETY

http://www.spatialstatistics.info/

ScienceDirect Journals Books Edzer Pebesma Help

Download PDF Export Search ScienceDirect Advanced search

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

Get rights and content

Open Access

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.

Recommended articles

A comprehensive evaluation of various sensitivit...
2014, Environmental Modelling & Software more

Radiosurgery for the treatment of dominant hemi...
2013, Epilepsy & Behavior Case Reports more

Synoptic weather typing applied to air pollution m...
2013, Environmental Research more

View more articles »

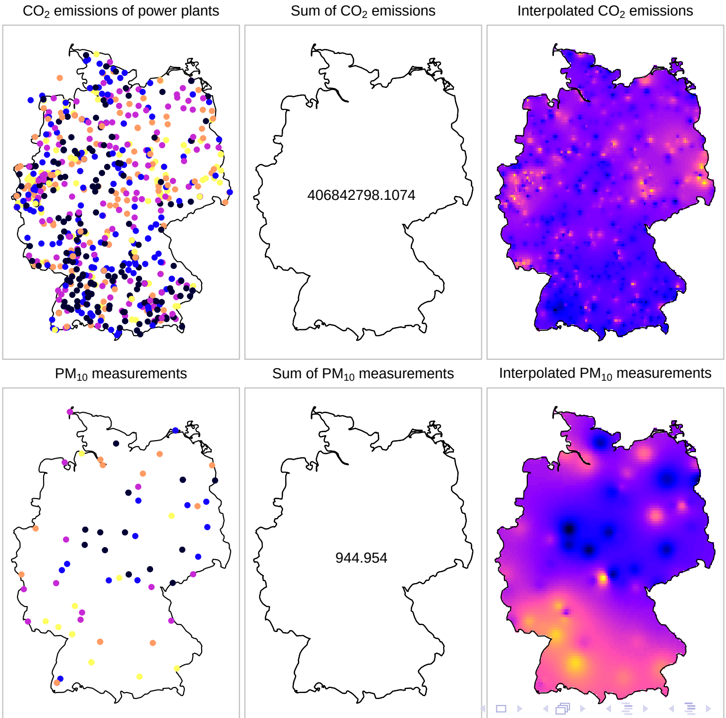
Citing articles (3)

Related book content

Data for this Article

RunMyCode
Code and data

Feedback


CO$_2$ emissions of power plants
Sum of CO$_2$ emissions
Interpolated CO$_2$ emissions
406842798.1074
PM$_{10}$ measurements
Sum of PM$_{10}$ measurements
Interpolated PM$_{10}$ measurements
944.954

Overview

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

1. Spatial Statistics

- Statistics is the science of uncertainty
- statistics (plural) vs. Statistics (singular)
- Spatial Statistics: a collection of methods/tools for analyzing spatial data (estimation, inference, prediction, simulation)
- Spatial Statistics usually adopts some kind of random field model
- Omni-presence of spatial correlation was described by Fisher (1936, The Design of Experiments), long before Tobler 1953 explained this to geographers.
- Fisher suggested to cope with it by randomizing sampling designs
- this rules out interpolation and flexible aggregation
- aren't all data spatial?

The classical types of Spatial Statistics

1. Point pattern analysis: focus on observed patterns
   - are these (crime, disease) cases clustered?
   - which process generated this pattern?
2. Geostatistics: focus on predicting missing values
   - how do I interpolate a variable at a new (point, area)?
   - how/where should I sample, to improve interpolation?
3. Areal (lattice) data analysis: focus on patterns and relations
   - Are these area values spatially correlated? Or clustered?
   - how can I analyze regression models, given this correlation?
   - (neighbours; SAR/CAR models)

The classical types of Spatial Statistics

1. Point pattern analysis: focus on observed patterns
   - are these (crime, disease) cases clustered?
   - which process generated this pattern?
2. Geostatistics: focus on predicting missing values
   - how do I interpolate a variable at a new (point, area)?
   - how/where should I sample, to improve interpolation?
3. Areal (lattice) data analysis: focus on patterns and relations
   - Are these area values spatially correlated? Or clustered?
   - how can I analyze regression models, given this correlation?
   - (neighbours; SAR/CAR models)
4. Spatio-temporal Statistics
   - extend previous themes to space-time, but also
   - state-space models, (S)PDE's
   - feature tracking, analysis of movement/trajectories

The classical types of Spatial Statistics

1. Point pattern analysis: focus on observed patterns
    - are these (crime, disease) cases clustered?
    - which process generated this pattern?
2. Geostatistics: focus on predicting missing values
    - how do I interpolate a variable at a new (point, area)?
    - how/where should I sample, to improve interpolation?
3. Areal (lattice) data analysis: focus on patterns and relations
    - Are these area values spatially correlated? Or clustered?
    - how can I analyze regression models, given this correlation?
    - (neighbours; SAR/CAR models)
4. Spatio-temporal Statistics
    - extend previous themes to space-time, but also
    - state-space models, (S)PDE's
    - feature tracking, analysis of movement/trajectories

Cognitive dissonance?

GI Science has:

- fields (coverages), and
- objects (features).

Spatial Statistics has:

- fields (geostatistics),
- objects (point patterns), and
- lattices (areal data).

Cognitive dissonance?

GI Science has:

- fields (coverages), and
- objects (features).

Spatial Statistics has:

- fields (geostatistics),
- objects (point patterns), and
- lattices (areal data).

1. is this a cognitive dissonance between Spatial Statistics and GI Science?

Cognitive dissonance?

GI Science has:

- fields (coverages), and
- objects (features).

Spatial Statistics has:

- fields (geostatistics),
- objects (point patterns), and
- lattices (areal data).

1. is this a cognitive dissonance between Spatial Statistics and GI Science?
2. is there more cognitive dissonance between them?

International Journal of Geographical Information Science
Vol. 00, No. 00, Month 2009, 1–29

Article under review

Research article

Modelling spatio-temporal information generation

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

[a]*Institut für Kartographie und Geoinformation, ETH Zürich, Switzerland;* [b]*Institute for Geoinformatics, University of Münster, Germany;* [c]*52°North GmbH*

(*Received 00 Month 200x; final version received 00 Month 200x*)

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 data in a meaningful way if information about its sources and types of conversions are discarded in the process of data gathering. A generative model of data derivation could not only help automating the description of derivation

2. Discovery

How do you discover data?

2. Discovery

How do you discover data?
Why is discovery important?

2. Discovery

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

3. 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/

3. 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. "

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

4. 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.

4. 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, pragmatics)

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

5. 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.

5. 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.

5. 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	$\text{Extent} \Rightarrow S \times T$	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	$S \times T \Rightarrow \text{Extent}$	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	$Q \Rightarrow Q \text{ set}$	map quality values to ranges of qualities
Qstat	$(Q \Rightarrow \text{bool}) \Rightarrow Q$	summarize quality values (e.g., mean, median)

6. 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

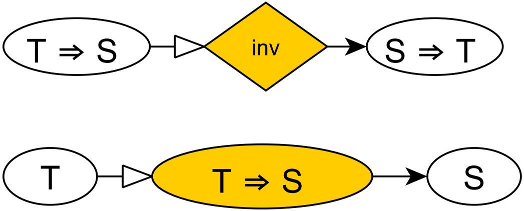
Generation procedures: Trajectories

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

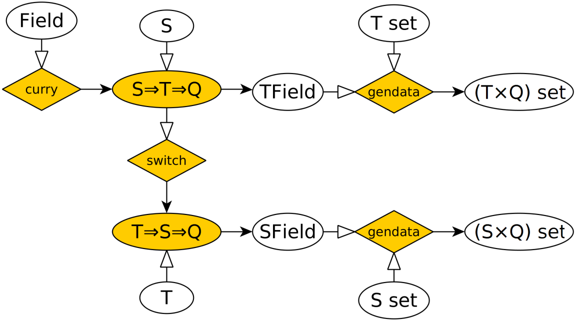
Generation procedures: Objects

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

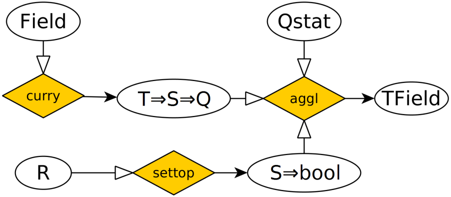
7. Data derivation

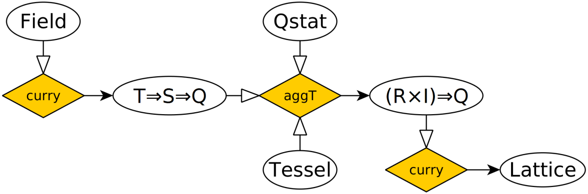


8. Data derivation graphs: generating field data



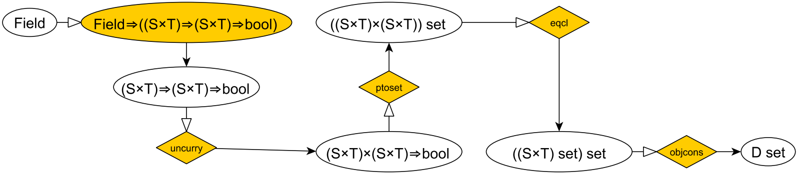
Data derivation: spatial/temporal aggregation



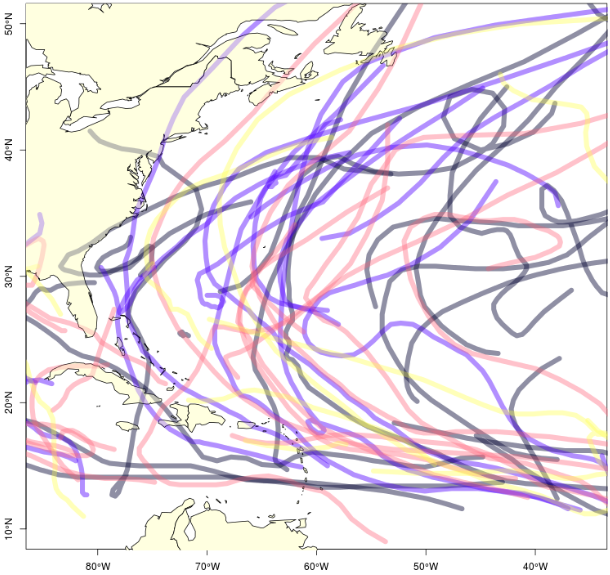


see paper for definitions of curry, aggl, aggT and settop

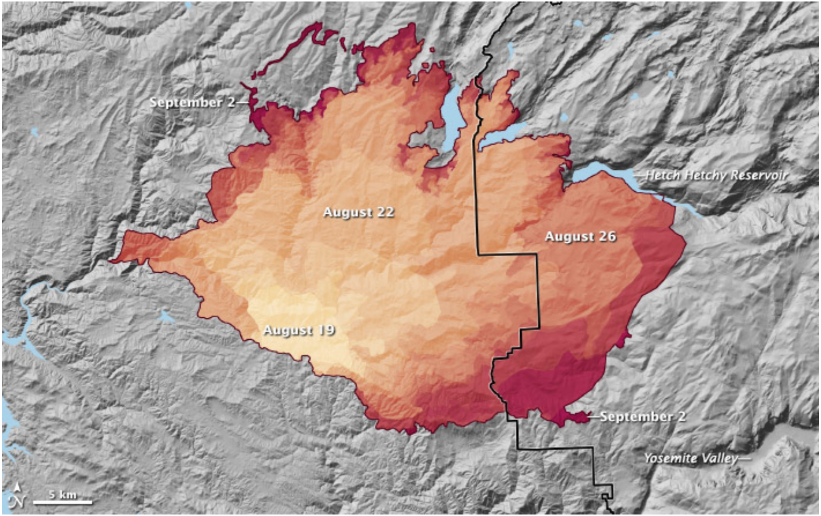
Data derivation: deriving objects from fields

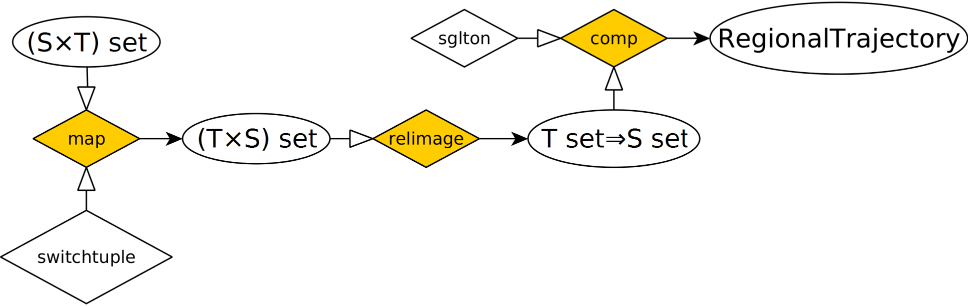


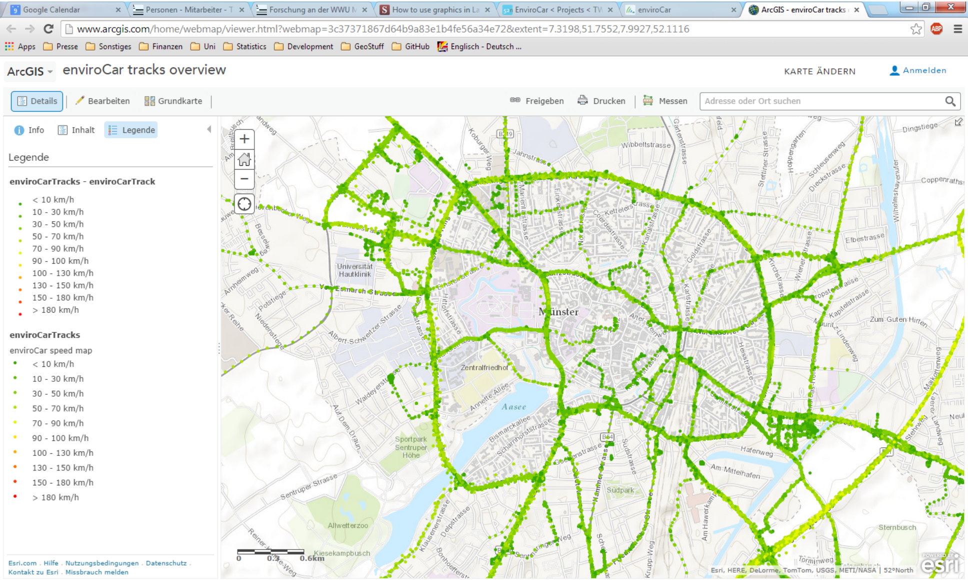
Storms: Trajectories from fields

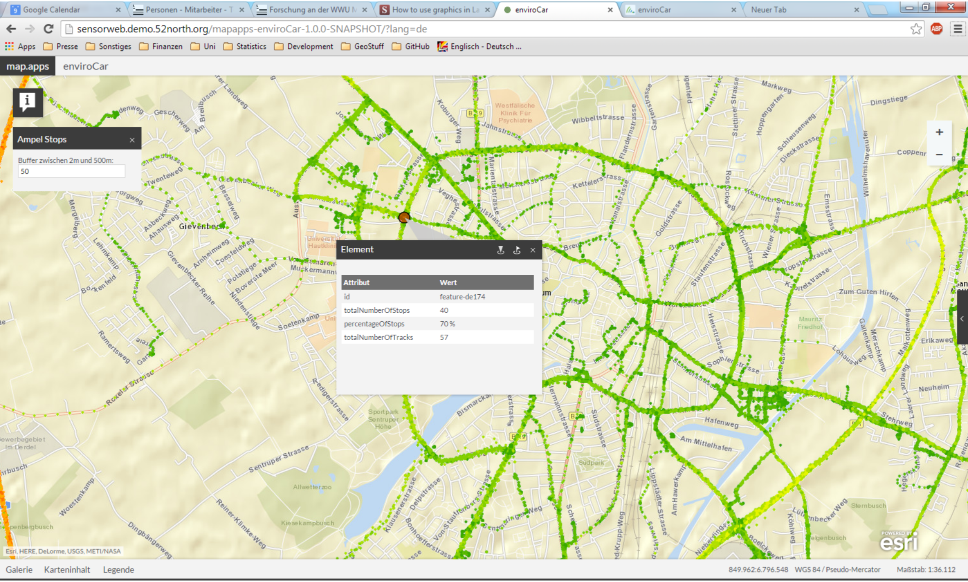


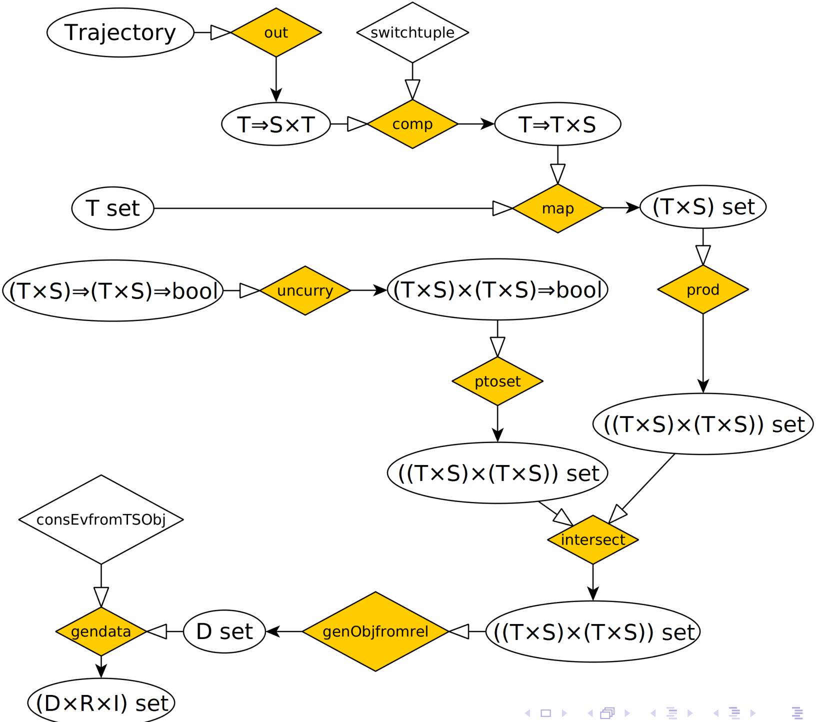
Forest fires: Regional Trajectories




(S×T) set
sglton
comp
RegionalTrajectory
map
(T×S) set
relimage
T set⇒S set
switchtuple

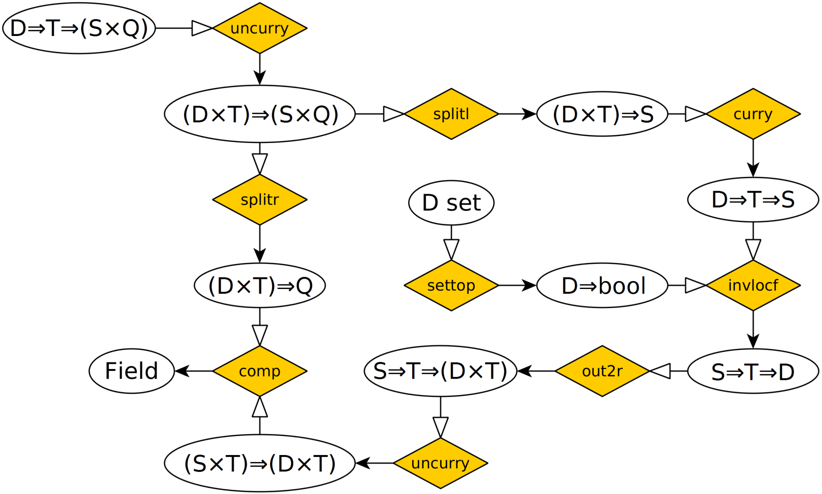

Google Calendar
Personen - Mitarbeiter
Forschung an der WWU
How to use graphics in La
EnviroCar < Projects < TV
enviroCar
ArcGIS - enviroCar tracks
www.arcgis.com/home/webmap/viewer.html?webmap=3c37371867d64b9a83e1b4fe56a34e72&extent=7.3198,51.7552,7.9927,52.1116
Apps
Presse
Sonstiges
Finanzen
Uni
Statistics
Development
GeoStuff
GitHub
Englisch - Deutsch ...
ArcGIS
enviroCar tracks overview
KARTE ÄNDERN
Anmelden
Details
Bearbeiten
Grundkarte
Freigeben
Drucken
Messen
Adresse oder Ort suchen
Info
Inhalt
Legende
Legende
enviroCarTracks - enviroCarTrack
< 10 km/h
10 - 30 km/h
30 - 50 km/h
50 - 70 km/h
70 - 90 km/h
90 - 100 km/h
100 - 130 km/h
130 - 150 km/h
150 - 180 km/h
> 180 km/h
enviroCarTracks
enviroCar speed map
< 10 km/h
10 - 30 km/h
30 - 50 km/h
50 - 70 km/h
70 - 90 km/h
90 - 100 km/h
100 - 130 km/h
130 - 150 km/h
150 - 180 km/h
> 180 km/h
Esri.com . Hilfe . Nutzungsbedingungen . Datenschutz .
Kontakt zu Esri . Missbrauch melden
Münster
Universität Hautklinik
Zentralfriedhof
Aasee
Sportpark Sentruper Höhe
Allwetterzoo
Südpark
Sternbusch
Kiesekampbusch
0 0.3 0.6km
Esri, HERE, DeLorme, TomTom, USGS, METI/NASA | 52°North
esri

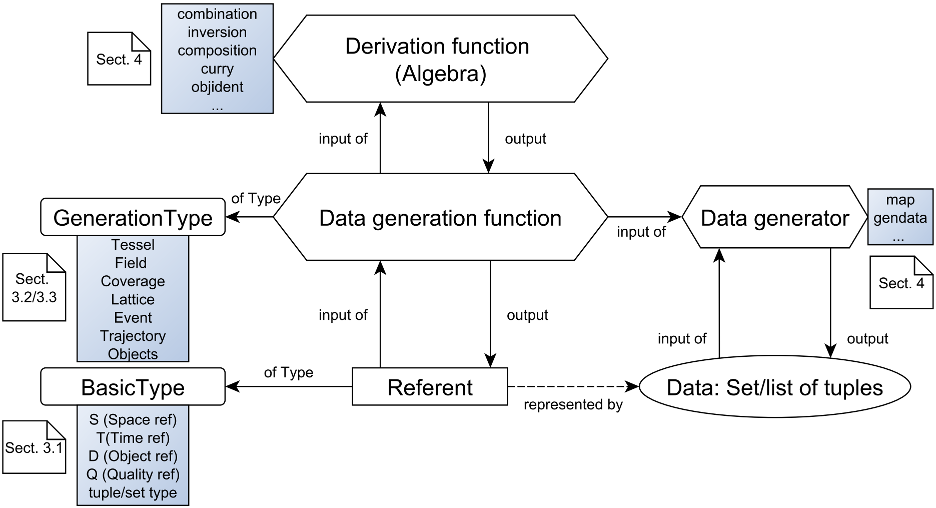

Google Calendar
Personen - Mitarbeiter - T
Forschung an der WWU M
How to use graphics in La
enviroCar
enviroCar
Neuer Tab
sensorweb.demo.52north.org/mapapps-enviroCar-1.0.0-SNAPSHOT/?lang=de
Apps
Presse
Sonstiges
Finanzen
Uni
Statistics
Development
GeoStuff
GitHub
Englisch - Deutsch ...
map.apps
enviroCar
Ampel Stops
Buffer zwischen 2m und 500m:
50
Element
Attribut | Wert
id | feature-de174
totalNumberOfStops | 40
percentageOfStops | 70 %
totalNumberOfTracks | 57
Esri, HERE, DeLorme, USGS, METI/NASA
Galerie
Karteninhalt
Legende
849.962:6.796.548
WGS 84 / Pseudo-Mercator
Maßstab: 1:36.112



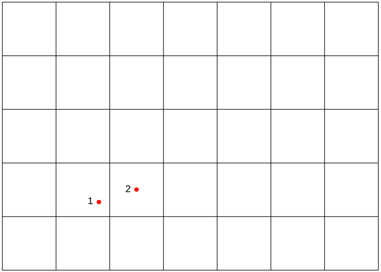
Trajectory
out
switchtuple
T⇒S×T
comp
T⇒T×S
T set
map
(T×S) set
(T×S)⇒(T×S)⇒bool
uncurry
(T×S)×(T×S)⇒bool
prod
ptoset
((T×S)×(T×S)) set
((T×S)×(T×S)) set
consEvfromTSObj
intersect
gendata
D set
genObjfromrel
((T×S)×(T×S)) set
(D×R×I) set

Generating a field from marked trajectories

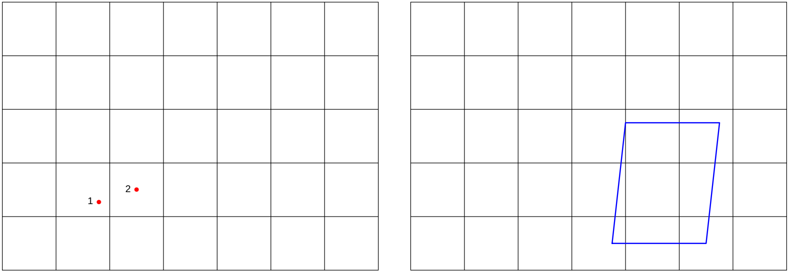



Sect. 4
combination
inversion
composition
curry
objident
...
Derivation function
(Algebra)
input of
output
GenerationType
of Type
Data generation function
input of
Data generator
map
gendata
...
Sect. 4
Tessel
Field
Coverage
Lattice
Event
Trajectory
Objects
Sect.
3.2/3.3
input of
output
input of
output
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
Sect. 3.1

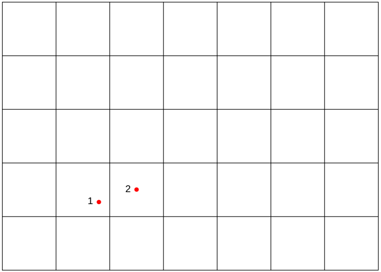
Retrieving information from raster maps

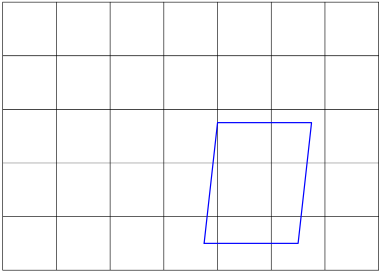


Retrieving information from raster maps



Retrieving information from raster maps

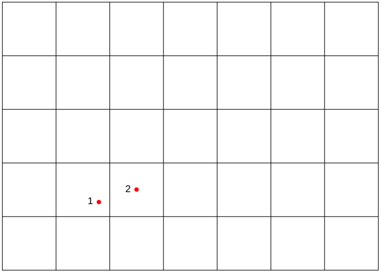


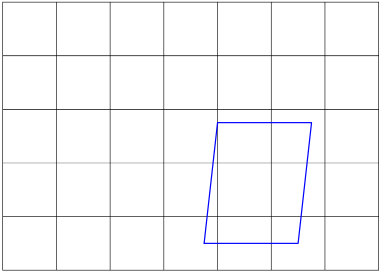


Three simplest cases:

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?

9: 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`

9: 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

10. Conclusions – I

- 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, Q, D
- 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.

Conclusions – II

The dominant GI types in use (simple features, coverages) do not inform how non-geometry attributes Q relate to geometry S, T:

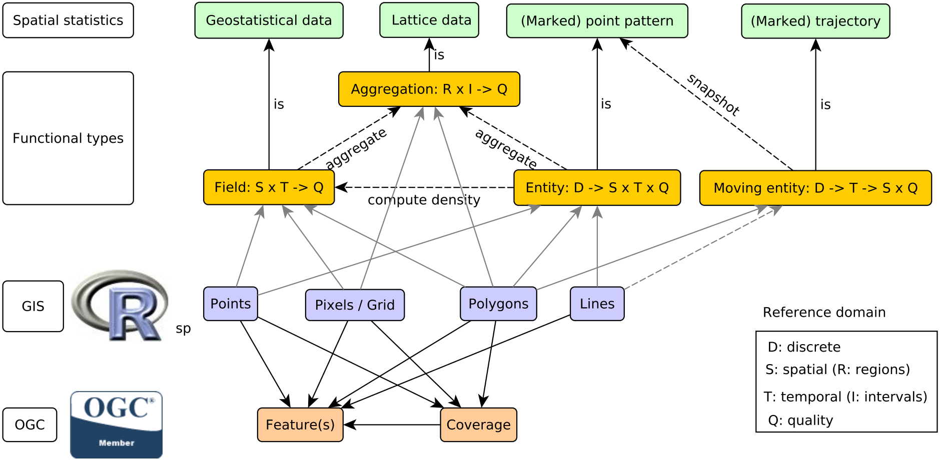
- are they *constant*,
- do they refer to an aggregate for the feature *as a whole*?
- do they refer to a single (central) location?
- are they the result of another convolution over a signal?

Conclusions – II

The dominant GI types in use (simple features, coverages) do not inform how non-geometry attributes Q relate to geometry S, T:

- are they *constant*,
- do they refer to an aggregate for the feature *as a whole*?
- do they refer to a single (central) location?
- are they the result of another convolution over a signal?

For spatial analysis, all this matters.
Thank you.



Spatial statistics
Geostatistical data
Lattice data
(Marked) point pattern
(Marked) trajectory
Functional types
Aggregation: R x I -> Q
is
is
is
is
snapshot
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
Reference domain
D: discrete
S: spatial (R: regions)
T: temporal (I: intervals)
Q: quality
OGC
OGC
Member
Feature(s)
Coverage