Transactions in GIS

Research Article — Transactions in GIS, 2014, 18(4): 527–543

A Generic Web Service for Ad-hoc Statistical Spatio-Temporal Aggregation

Mario Härtwig, Matthias Müller and Lars Bernard

Technische Universität Dresden

Abstract

For the analysis of geospatial data, statistical spatio-temporal aggregation (SSTA) is a frequently used functionality. Possible applications include information extraction, data fusion, generalization and schema transformation. On the basis of spatial and temporal references, SSTA transfers thematic attribute values into a coarser spatio-temporal resolution using descriptive statistical operations. Reusable geoprocessing functions are increasingly offered in Spatial Data Infrastructures (SDIs) through standardized interfaces using mainstream network technologies. Due to the variety of use cases, a Web service for SSTA promises a high potential for reuse. However, offering functionalities for geoprocessing within an SDI raises some challenges. The aggregation process itself, as well as the encapsulation of the aggregation functionalities within an SDI, require thorough design considerations. This article addresses these challenges by developing a generic framework for SSTA services within an SDI. A prototype of the framework is realized using several open-source software components. Following a modular approach, the communication between those loosely-coupled components is enabled through open, standardized interfaces. The result of the proposed work is a framework that enables users to easily perform a statistical ad-hoc aggregation on distributed spatio-temporal data.

1 Introduction

Growing amounts of geodata describing space- and time-varying phenomena require efficient methods for data densification. A required key functionality is statistical spatio-temporal aggregation (SSTA). It summarizes geodata by location, aggregating thematic values on the basis of spatial and temporal geometries. The transformation of thematic attribute values into a coarser resolution is performed using descriptive statistical operations like mode, minimum, maximum, median, arithmetic mean, standard deviation and sum. The aggregation of spatio-temporal data targets to reduce unnecessary details such that relevant relations become recognizable more easily. Hence, it is an important functionality in a broad range of use cases:

- *Data interpretation:* The calculation of statistical aggregation parameters to represent data characteristics in a condensed way is of essential importance in the field of geodata analysis (Andrienko and Andrienko 2006) and decision support.
- *Privacy and data protection:* Statistical aggregations reduce the level of detail of geodata, which is why they are utilized in the context of data protection to avoid the derivation of individual personal information from privacy-sensitive geodata (Strobl 2005).

Address for correspondence: Mario Härtwig, Technische Universität Dresden, Helmholtzstr. 10, 01062 Dresden, Germany. E-mail: mario.haertwig@geops.de
Acknowledgements: Several findings presented in this article stem from work within the research programme “Sustainable Land Management” funded by the German Federal Ministry of Education and Research (BMBF). Christin Henzen and Daniel Kadner are gratefully acknowledged for providing the TIME4MAPS implementations.

doi: 10.1111/tgis.12062

- *Data transformation:* Geodata is statistically aggregated for the purposes of generalization, schema transformation (Foerster et al. 2010) and geodata fusion (Wiemann and Bernard 2010).
- *Data reduction (storage, bandwidth):* The reduction of storage and bandwidth by SSTA plays an important role, for instance in Geosensor Networks. Observation data is aggregated either by the individual sensor nodes (Intanagonwiwat et al. 2002) or in a central data sink (Bonnet et al. 2001).

Functionalities for processing geodata are increasingly incorporated in service-based Spatial Data Infrastructures (SDIs). The provision of geoprocessing functionality via standardized interfaces using mainstream technologies offers this functionality to a broad range of users and makes it available to various applications. Easy reusable services for SSTA could be essential components in geoprocessing workflows for data analysis or in decision support systems (Bernard and Ostländer 2008; Bernard et al. 2013). However, the realization of an SSTA within a service-oriented architecture raises some challenges regarding:

1. Automatic derivation of sound or meaningful aggregation operations based on the semantics of the data;
2. Formalization of input and output data in order to deduce the geometric and thematic properties for an automatic aggregation;
3. Specification of a procedure for the automatic alignment of source and target geometries; and
4. Revealing the results and possible uncertainties by the documentation of the processes for conflation and aggregation in a structured way.

This article addresses these challenges by developing a generic framework for SSTA in the context of an SDI. Starting with the general process flow of an SSTA (Section 2), geometric and thematic aspects of the aggregation are examined individually. Section 3 deals with the transformation of geometric objects of a spatio-temporal source geometry into a coarser target geometry and Section 4 analyzes the statistical aggregation of thematic attribute values associated with the geometric objects. The system architecture and a prototypical realization is described in Section 5, followed by some conclusions.

2 A Generic Process for SSTA

The general process flow for the statistical aggregation of a spatio-temporal dataset towards a target geometry of a coarser spatio-temporal resolution is illustrated in Figure 1. The process is defined generically and supports both object- and field-based data models (Couclelis 1992). The target and output geometries always take the form of geographic objects.

As a first step, the process reads the input dataset for which the SSTA shall be performed. The subsequent aggregation procedure depends on the data model of the source and is divided into a geometric and a thematic part:

1. *Geometric aggregation*
   a. *Object based data model:* The thematic attribute values are associated with geometric objects in a spatio-temporal domain. The aggregation transforms these objects into a coarser target geometry. Therefore, spatio-temporal geometric objects from the input geometry, which can be assigned to one single geometric object of the target geometry, are summarized to a group (Figure 2). Attribute values associated with these groups are extracted for the subsequent thematic aggregation.

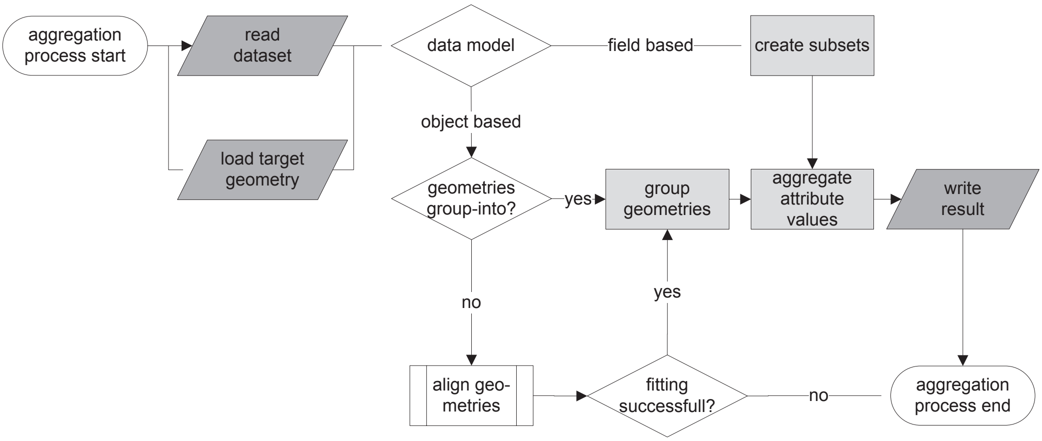


Figure 1 Process flow for the statistical spatio-temporal aggregation

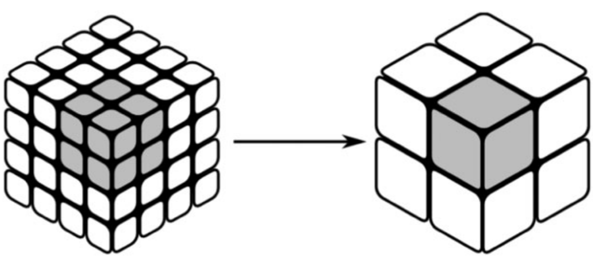


Figure 2 Geometric aggregation of discrete geometries. Geometric spatio-temporal entities in the input geometry (left) that fit into a coarser entity of the target geometry (right) form a group (grey)

   b. *Field-based data model:* It is assumed that field-based data models consist of a sampling function and an appropriate interpolation scheme (i.e. nearest neighbor, bilinear, bicubic). Thus, fitting to the spatial entities of the target geometry is not required. A geometric sub-setting procedure is sufficient to extract the values for thematic aggregation.

2. *Thematic Aggregation:* The thematic aggregation is performed in an integral fashion. Descriptive statistical measures are computed for each entity in the target geometry from the associated source values. The permissibility of an aggregation operation for summarizing attribute values depends on the measurement type. This measurement type must be defined in the source data (see Section 4).

Concepts that are relevant for the design of the service-based SSTA are presented in the following two sections. The first section deals with the geometric properties of the data sets and the related issues in geometric alignment of source and target geometries. The second discusses the statistical aggregation of thematic values in detail.

3 Geometric Aggregation

A straightforward aggregation of high resolution data into coarser geometries can only be achieved if the geometric units of the source data fit perfectly into the units of the target's

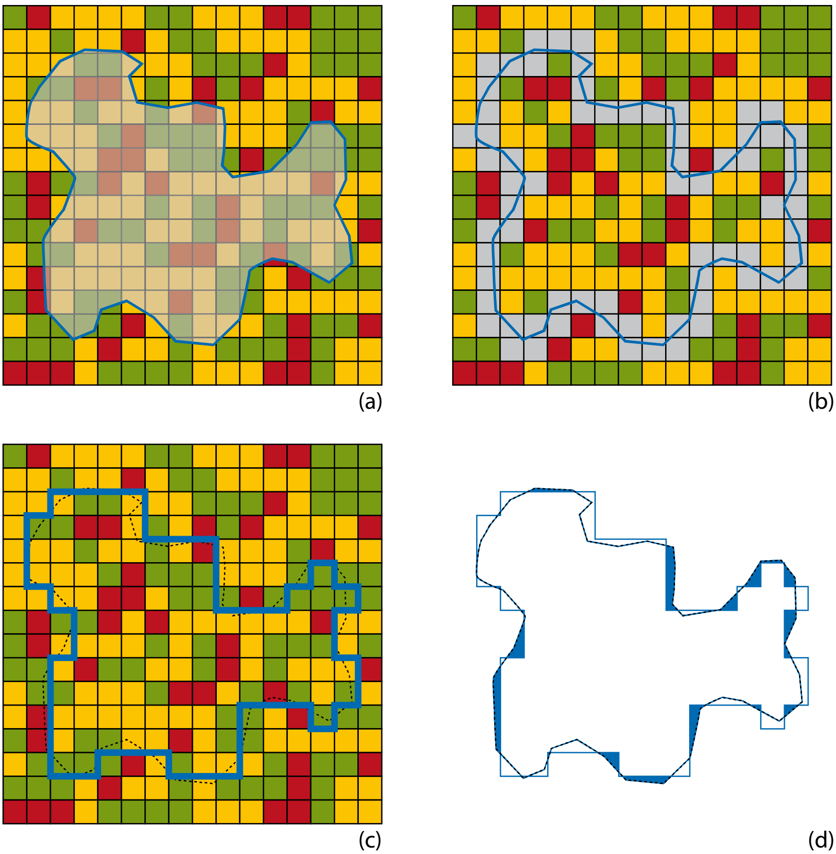


Figure 3 Geometrical alignment of a polygonal target granularity (blue) to a polygonal grid geometry: (a) Original geometries; (b) Source alignment; (c) Target alignment; and (d) Target alignment error

reference geometry. In many cases the source and target geometries may be aligned if they are using compatible reference systems. Examples for such reference systems are hierarchical administrative units like NUTS (European Commission 2012) or hierarchies of regular grids as defined by INSPIRE (2009). Here, the fine grained data can be smoothly transferred to coarser geometric units.

If source and target geometries do not adhere to such a common reference system, a perfect fit is highly unlikely (Figure 3a). In this case, some sort of geometric preprocessing is required. There are two general partitioning approaches to solve this issue: (1) the source data is repartitioned to fit the target geometries; or (2) the target partitions are adjusted to fit the partitions of the source data. Repartitioning the source data involves more than just touching the geometry: Geometric units are closely related to the thematic domain and may represent a sampling space or relate to the significance of the thematic data. Changing the original partitioning scheme may thus involve a recalculation of the thematic values which, in general, can easily lead to questionable values (Figure 3b) (Robinson 1950; Openshaw 1984). Aligning the target partitions to the source partitions may provide a slightly different output than expected but does not suffer from unpredictable effects on the thematic values (Figure 3c). The adjustment of the target

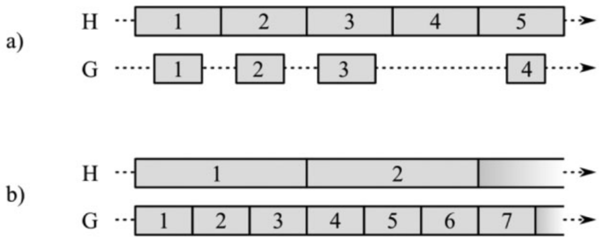


Figure 4 Granularity relationships between a source granularity *G* and a target granularity *H*

geometry is completely transparent to an analyst and can be quantified with both aerial and positional tolerance (Figure 3d). For these reasons the SSTA service proposed here follows the second partitioning approach, which is considered more suitable for automatic execution.

3.1 Geometric Aggregation in Compatible Reference Systems

In spatio-temporal datasets, thematic attributes are associated with entities in a spatial and temporal domain. These entities, such as minutes, hours, days in the temporal context or meters, miles, federal states and countries in the spatial context, can be considered as granules of a granularity. A granularity is defined as a mapping from an ordered index set to subsets of the corresponding domain (Bettini et al. 2002; Camossi et al. 2006). The granules of a granularity are non-decomposable and do not overlap. Between hierarchical structured granularities of the same temporal domain, Bettini et al. (2002) define a set of relationships. Two possible relationships are illustrated in Figure 4. Figure 4a shows the *finer than* relationship, implying that granularity *G* (e.g. the working time of a specific person) is finer than granularity *H* (e.g. a day granularity). The granularities in Figure 4b are related by the *groups into* relationship, meaning that every granule in granularity *H* (e.g. a year granularity) forms a union of a set of granules in granularity *G* (e.g. a quarter year granularity).

The compliance with the requirements of the *groups into* relationship is necessary for the grouping of granules on the basis of a coarser target granularity. It enables, for example, the grouping of day granularities to week, month or year granularities or of federal state granularities to country granularities.

Granularities of geodata are described by spatial and temporal attributes, whereas granules are represented by spatial and temporal geometric objects. All those objects, which are specified in ISO 19107 and 19108, differ in their dimensionality and complexity. For the grouping of these spatial and temporal geometric objects the following possibilities exist:

- Points can be grouped to a complex geometric object called multipoint that forms a geometric aggregation of a set of points (ISO 19107). Temporal instances could be grouped to complex temporal objects, albeit such an object is not defined for geodata (see ISO 19108).
- One-dimensional spatial curves can be grouped to a complex geometric object that forms a set of curves. Such an object is called either multicurve if the curves are loosely coupled or composite-curve if there is an additional internal structure. Furthermore, curves can be grouped to longer curves if the requirements of the *groups into* relationship are fulfilled. Similarly, a grouping of periods into longer periods in the temporal domain can be applied.
- Geometric surfaces with a two-dimensional spatial extend can be grouped to complex geometries which are, based on their semantics, either composite-surfaces or multi-surfaces. Furthermore a union into a common surface (envelope) is possible.

Table 1 Possibilities for the conversion from a source representation into a target representation according to ISO 19109 and 19123

Target Geometry / Source Geometry		Feature Type (multi / composite) Point	Feature Type (multi / composite) Curve	Feature Type (multi / composite) Surface
Feature Type	Point	✓	✗	✗
	Curve	✗	✓	✗
	Surface	✗	✗	✓
Continuous Coverage		✓	✓	✓

Table 2 Flow of the alignment process: Alignment of a target geometry (H) to a source geometry (G)

<table>
<tr><td colspan="3">For every geometric object in H do the following:</td></tr>
<tr><td colspan="3">Create an empty selection for geometric primitives</td></tr>
<tr><td colspan="3">For every geometric primitive of the geometric object in H do the following:</td></tr>
<tr><td>Point</td><td>Line</td><td>Polygon</td></tr>
<tr><td>Add the most similar point in G to the selection based on:
– (Euclidean) Distance</td><td>Add lines in G that best represent the currently analyzed line in H to the selection based on:
– Buffer intersection
– Angular information
– Relative length
– Mutual overlapping</td><td>Add polygons in G that best represent the currently analyzed polygon in H to the selection based on:
– Intersection area</td></tr>
<tr><td colspan="3">Merge all geometric objects within the selection to a new geometric object.</td></tr>
</table>

As already stated in the previous section, the geometric aggregation of field data is not an issue from a geometric perspective.

Considering the geometry models as given in ISO 19109 and 19123, the possible conversions are specified in Table 1.

3.2 Alignment Approaches for Incompatible Geometric Reference Systems

A precondition for the aggregation of discrete data is that the source geometries perfectly fit into the target geometries. To this effect, the conceptualized alignment procedure (Table 2) modifies geometric objects in the target geometry *(H)* in such a way that they form a union of a set of geometric objects in the source geometry *(G)*. For the alignment in the temporal

Table 3 Levels of measurement and permissible statistical operations

Measurement scale	Math. Structure	Additional permissible statistics
Nominal	$x' = f(x)$	*Count, Mode*
Ordinal	$x' = f(x)$	*Median, Percentiles*
Interval	$x' = ax + b$	*Mean, Standard deviation*
Ratio	$x' = cx$	*Coefficient of variation*

dimension, similar operations can be used, e.g. point distances, mutual overlaps and relative lengths. Further information about matching and alignment of geospatial data can for instance be found in Chen and Knoblock (2008), Yuan and Tao (1999) and Walter and Fritsch (1999).

Despite the several geometric alignment operations, a service-oriented SSTA process should provide simple parameters to control the alignment process. The used parameters determine the tolerance: the largest permissible offset for distances between the original and the aligned geometric object of the target geometry. For the alignment of spatial geometries, the tolerance is given by a metric length measurement whereas for temporal alignment it is defined in the form of a temporal period compliant to ISO 8601. If, at an arbitrary location, the shortest distance between the original target geometry and the aligned target geometry exceeds the tolerance, the alignment process is canceled and no aggregation measures are calculated (see process flow diagram in Figure 1).

4 Thematic Aggregation

A soon as the target geometries have been fitted to the source geometries, an instant thematic aggregation is possible. Depending on the properties of each thematic attribute, different statistical aggregation operations are permitted. Based on a brief discussion about levels of measurement, this section presents a selection mechanism for permissible statistical operations for thematic data. On the basis of the requirements for an automatic aggregation process, suggestions are made for metadata extensions in current data models.

4.1 Dealing with Measurement Types and Statistical Operations

Thematic attributes in geospatial data contain measured properties of observed phenomena. According to the rules by which numbers are assigned to observed properties during measurement, different types of measurement can be demarcated. The type of a measurement determines the semantics as well as the operations permissible for the statistical treatment of the measured values. Several approaches deal with the categorization of measurements and appropriate statistical operations. The most prominent categorization is the one proposed by the philosopher Stevens (1946), who distinguishes four measurement scales (Table 3).

Stevens (1946) proposal has been critiqued extensively. One point of criticism is that the assignment of a measurement to a scale often depends on the context (Gaito 1980). In particular the distinction of ordinal and interval scaled measures is often not obvious (Gardner 1975; Norman 2010). In many use cases ordinal scaled measurements, like school marks, are weighted and thus treated like measurements on the interval scale. Even Stevens (1959) notes that such “illegal” statistics often lead to “fruitful results”.

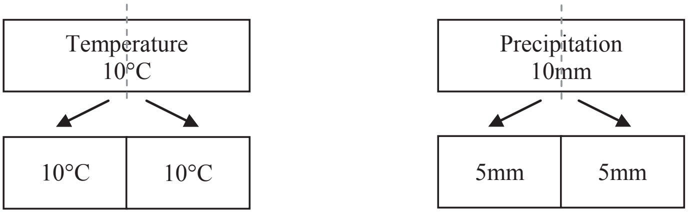


Figure 5 Differentiation between scale invariant intensive properties (left) and scale variant extensive properties (right)

A further pitfall is that not every kind of measurement can be assigned to one of the measurements scales. One example is logarithmic scaled measurements, which is why Stevens (1959) later introduced the logarithmic interval scale which is on the same level as the ordinary interval scale. Nevertheless, as for instance Chrisman (1998) points out, there are still several measurements that cannot be assigned to one of these scales (e.g. cyclical measurements).

Another disadvantage of Steven's classification scheme that particularly affects the SSTA is that it does not provide any information about the permissibility of summing up values (Chrisman 1998). On both interval- and ratio scale there are measurements that could be summed (e.g. percentage of female inhabitants; precipitation in millimeters) and measurements that could not be (e.g. temperature in degrees Celsius; migration balance of a region in percent).

Several statistical operations can be used for characterizing a group of thematic attribute values by one single value. Operations that are most frequently used for statistical aggregation are count, mode, minimum, maximum, median, arithmetic mean, standard deviation and sum. As stated above, most of these operations can clearly be assigned to one particular level of measurement. The sum, which is an often needed aggregation operation, has no single counterpart.

Whether metric measurement values can be aggregated by summing depends on whether the measured property is scale invariant or not (Figure 5). Properties that are independent from the size of the system they characterize, like temperature, pressure and density, are called intensive. On the contrary, extensive properties, like mass and volume, depend on the size of the system they characterize, since they are the sum of properties of independent subsystems. Thus, only for extensive properties can the sum be interpreted meaningfully.

Consequently, for the automated selection of common permissible statistical aggregation operations, a distinction between nominal, ordinal and metric measurement types is necessary whereas the latter must be subdivided into intensive and extensive measurements. The extended classification scheme in Table 4 enables a generic process for SSTA to automatically select permissible statistical aggregation operations for a given input dataset. Using different statistical formula, the mentioned statistical operations can be applied to thematic attribute values for both object- and field-based geodata.

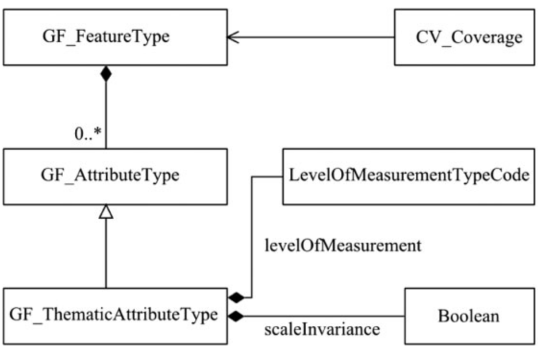
When aggregating spatially defined thematic attributes (e.g. population densities), the attribute values must be weighted according to their proportion of space or time.

4.2 Provision of Semantics

As shown in the previous section, data semantics are key prerequisites to the automatic derivation of suitable statistical aggregation operations. However, provision of semantic information

Table 4 Levels of measurements and statistical aggregation operations

Level of measurement		Additional permissible statistical aggregation operations
Nominal		*Count, Mode*
Ordinal		*Median, Minimum, Maximum*
Metric	*Intensive*	*Arithmetic Mean, Standard deviation*
	Extensive	*Sum / Integral*



	Name	Definition
1	Nominal	Qualitative measurements without ordering
2	Ordinal	Qualitative measurements with ordering
3	Intervall	Quantitative measurements without abs. zero
4	Ratio	Quantitative measurements with abs. zero

Figure 6 Definition of the measurement type in the ISO General Feature model

in SDI is still a research topic. The use of ontologies, which structure the vocabulary from different application domains based on a common conceptualization (Gruber 1993), has been suggested by several authors to enrich data with semantic information. Using ontologies as a core component, Kuhn (2003) proposes the theory of semantic reference systems which enable the semantic translation between domain-specific vocabularies. Probst (2007) demonstrates how semantic reference systems can be used for structuring measurements.

A precondition for the usage of semantic reference systems for the derivation of appropriate statistical (aggregation) operations is the specification of a common classification scheme for measurements and statistics. As shown in the previous sections, Stevens' levels of measurements in several cases do not sufficiently meet the requirements for statistical data analysis. Other classification schemes, like the one presented by Chrisman (1998) that meet the requirements for SSTA, are not well recognized. Due to the absence of a common conceptualization that meets the requirements of an automatic SSTA, no upper ontology could be used. Thus a different, pragmatic approach has been used here.

In extending an approach from Bernard and Krüger (2000) the semantic information being necessary for the SSTA application is added to the feature data model. Since the current ISO 19100 series does not offer a handle here, the ISO *GF_ThematicAttributeType* has been extended with additional mandatory attributes to define the measurement type (Figure 6): *levelOfMeasurement* and *scaleInvariance*. The element *levelOfMeausrement* is an extensible code list for the level of measurement. It currently contains the four levels of measurements defined by Stevens (1946). The Boolean element *scaleInvariance* indicates the validity of a totaling operation on the data.

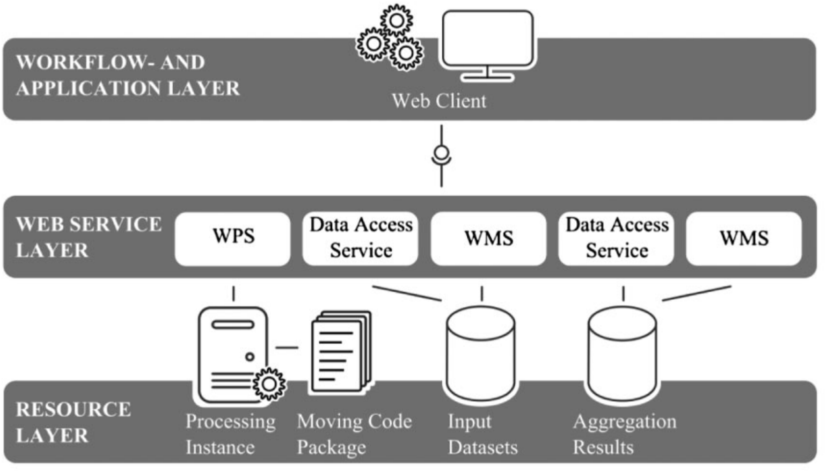


Figure 7 Three layers of the general modular system architecture

5 A Web Service for Statistical Spatio-Temporal Aggregation

In this section the modular system architecture is conceptualized on the basis of a use case. Particular attention is paid to reusability of the implemented components. This is achieved by the usage of open standardized interfaces, a clutch of open-source software products and an innovative approach for offering the computational logic in a reusable manner.

5.1 General System Architecture

With the purpose of facilitating an ad-hoc analysis of distributed spatial data sources and for making SSTA functions available to a broad user spectrum, the conceptualized aggregation algorithm is incorporated in a service-oriented SDI. Functionalities are offered through open interfaces standardized by the Open Geospatial Consortium (OGC). OGC Web services relevant to this article are the Web Processing Service (WPS) for offering geoprocessing logic, the Web Coverage Service (WCS) and the Web Feature Service (WFS) for data access and the Web Map Service (WMS) providing cartographic visualization of the data. The general architecture of a system for Web-based SSTA is illustrated in Figure 7.

The Web client at the *workflow-and application layer is* the user interface element of the application and serves to control the chain of Web services. For the implementation of the Web client, the following open source applications were used:

- *Time4Maps*: The Time4Maps application, which is the technical scaffolding of the Web client, is developed at TU Dresden and will soon be publicly available via the 52°North Initiative for Geospatial Open Source Software. Time4Maps enables the visual analysis of spatio-temporal datasets obtained via the interface of an OGC compliant WMS (version 1.1.1 or 1.3.0). The Web interface enables users to navigate spatially and temporally through datasets, to request feature information by clicking into the map and to execute a temporal animation. Since the current version of Time4Maps does not support access to geoprocessing functions offered by a WPS, it was necessary to extend the application.

- *Openlayers, Dojo Framework*: Time4Maps uses the OpenLayers JavaScript framework for accessing WMS interfaces and visualizing the returned maps. JavaScript implementations are realized using the Dojo Toolkit.

The service interfaces of the *Web service layer* offer interoperable access to the implemented functionalities and resources. They were technically realized by the usage of open source software products:

- *52°North WPS Framework*: The described architecture uses the WPS framework of 52°North for the technical realization of the OGC WPS specification. The framework is a Java-based modular and pluggable application that runs on almost every platform. One particular advantage of the framework is its ability to deploy and offer computational logic incorporated in Moving Code packages.
- *Thredds Data Server & ncWMS*: The Thredds Data Server (TDS), which is developed by the UNIDATA Community, provides access to scientific datasets, e.g. through OGC WMS and WCS interfaces. The offered datasets comply with the Unidata Common Data Model (CDM). For the visualization of NetCDF datasets, the TDS has been integrated with the open source application ncWMS, an implementation of the OGC WMS interface.
- *GeoServer*: For access to datasets which do not implement CDM, the GeoServer open source Java software is used, also implementing the OGC WCS, WFS and WMS interfaces.

Processing logic and datasets reside at the *resource layer*. Offering computational logic in a reusable manner facilitates maintenance and reduces implementation time. Since SSTA is crucial for analyzing spatio-temporal data, a generic and reusable process for SSTA would be characterized by a high level of utilization. Pursuing this target, the SSTA implementation is offered in accordance with the *Moving Code* approach (Müller et al. 2010, 2013). Following this, implemented processing logic is packed into a self-describing archive file that can be shipped to an arbitrary processing instance. Besides reusability, the *Moving Code* approach contributes further to performance enhancement and cost reduction, since computational logic can be shipped to a processing service that resides close to the data.

SSTA is performed for datasets, whose thematic attributes are associated with positions in a spatial and temporal reference system. A data format that was specially designed for the modeling of continuous spatio-temporal phenomena is the Network Common Data Form (NetCDF) which is adopted by the OGC as an open standard. The platform independent and self-describing format facilitates efficient data access and enables the access to subsets of thematic attributes, of particular advantage if dealing with large datasets as global climate scenario datasets. Apart from the data format, NetCDF also stands for a set of tools and interfaces for accessing and exchanging scientific datasets. For the modeling of discrete spatio-temporal phenomena, the XML-based Geography Markup Language (GML) defined by the OGC is used.

The interaction of the service components is illustrated in Figure 8. Before executing an SSTA, the user may want to visually inspect the input datasets of the aggregation process. A visualization of the datasets is requested by the client through the interface of a WMS. Access to datasets is offered by a data access services, i.e. WCS and WFS. For process execution for an arbitrary combination of datasets, the Web client sends an execute request to the WPS that offers functionality from the *Moving Code* package. Using the submitted dataset references the WPS requested input datasets from the data access services and performs the aggregation. Result datasets get stored and accessible via a separate data access service. A visualization of result datasets is again requested by using an appropriate WMS implementation (ncWMS, GeoServer, etc.).

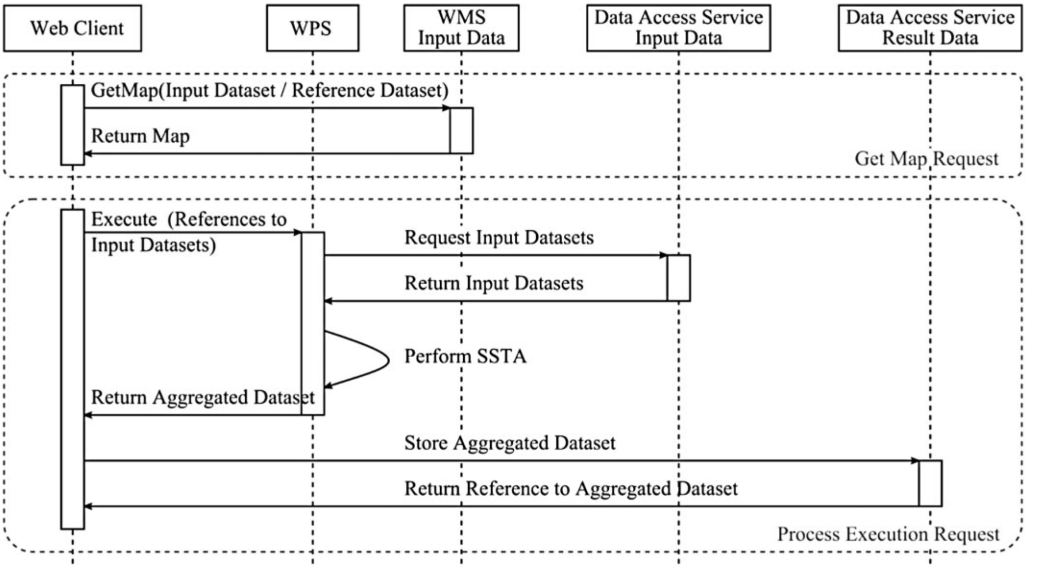


Figure 8 UML diagram sketching the SSTA service interaction

5.2 WPS Description and Implementation of the SSTA Process

The primary design goal of the SSTA process is the provision of a simple interface combined with functionality that serves a broad range of applications. Figure 9 provides an input/output oriented view on the intended WPS process which might serve as a starting point for the definition of a WPS process profile for SSTA. The process has two mandatory arguments and two optional arguments. *SourceData* represents the fine-grained or continuous spatio-temporal data which will be aggregated by the process. The intended spatio-temporal target granularity is represented by the *TargetGranularity* input. The process also provides the optional arguments *PermissibleSpatialTolerance* and *PermissibleTemporalTolerance* as discussed in Section 3.2. If these optional arguments are missing, the process will attempt an alignment process, irrespective of whether or not the result may have a very odd shape. In general, it is recommended to provide reasonable tolerance margins. For exploratory data analysis, where a visual inspection of the results is mandatory, these parameters could be omitted. If the process is unable to keep the adjustment within the desired margins, it returns an exception.

The result of the aggregation process (*AggregatedData*) provides the actual SSTA data for the possibly aligned target geometry containing the permissible statistical measures identified by the process. The aggregated data is assigned to the result of the alignment operation (*AlignedTargetGranularity*). This aligned geometry can be used in subsequent aggregation processes to ensure matching target granularities, e.g. for data comparison or further geoprocessing tasks. In this case the SSTA process may be called the *AlignedTargetGranularity* from a former execution but with different input data. The tolerance should be set to nil to assure identical output geometries.

The algorithms for alignment and for performing SSTA were implemented in Java. For accessing and analyzing spatio-temporal datasets, the following open source libraries were used.

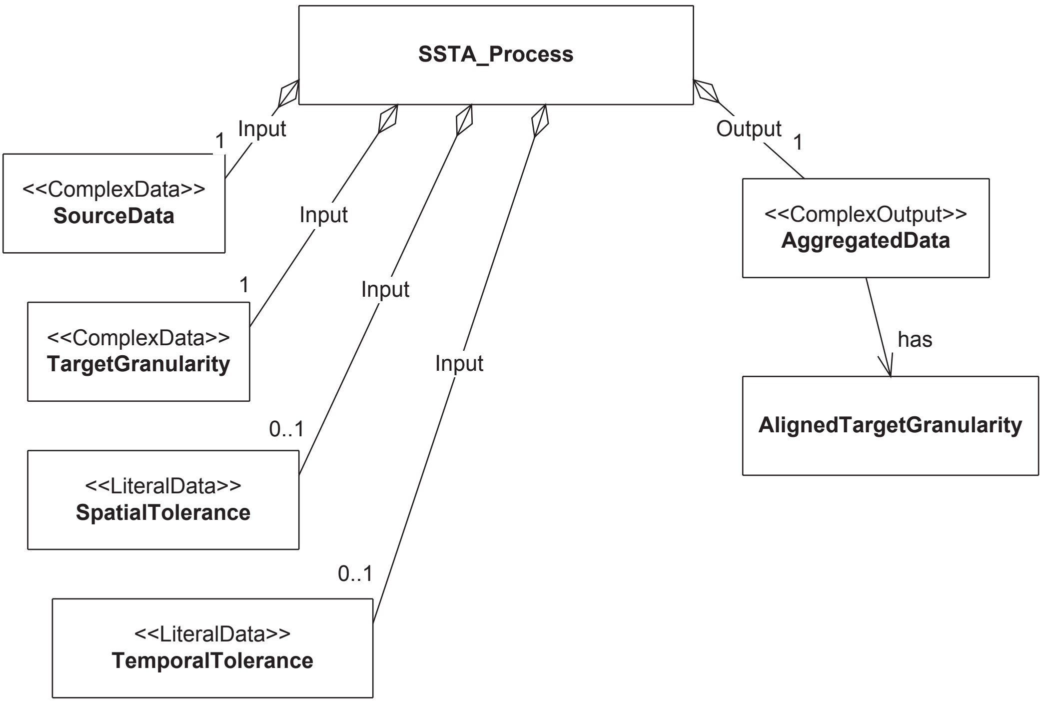


Figure 9 UML diagram displaying the interface of the SSTA WPS process

- *GeoTools*: Providing methods for accessing, analyzing and visualizing geodata, the open source library GeoTools is particularly suitable for implementing GIS functionality. The Java code library is compliant with OGC standards, which helps to structure the broad range of functions and enables a very straightforward usage. The processing algorithm described in this article primarily uses GeoTools (version 8.4) for accessing and geometrically analyzing geodata.
- *NetCDF-Java*: Since in GeoTools access to NetCDF datasets is only realized by an unsupported plugin in an experimental state, the NetCDF Java library (version 4.3.14) is used for efficient access to NetCDF datasets.
- *Joda Time*: Another open source library that was used for the realization of the service-based SSTA is Joda Time (version 2.1), a Java Date and Time API. Since the library implements ISO-compliant date and time formats, it is very appropriate for analyzing temporal aspects of geodata.

5.3 Use Case

Many decision processes require the aggregation of spatio-temporal data for getting an overview of the spatial and temporal distribution of a geographic phenomenon in a particular study area. A service for SSTA would for instance support a policymaker who wants information about the climatic conditions within an administrative area, to develop appropriate land-use policies. In a Web-based decision support system, he or she could load a dataset containing the geometry of the administrative area and mash it up with several spatio-temporal climatological data, e.g.

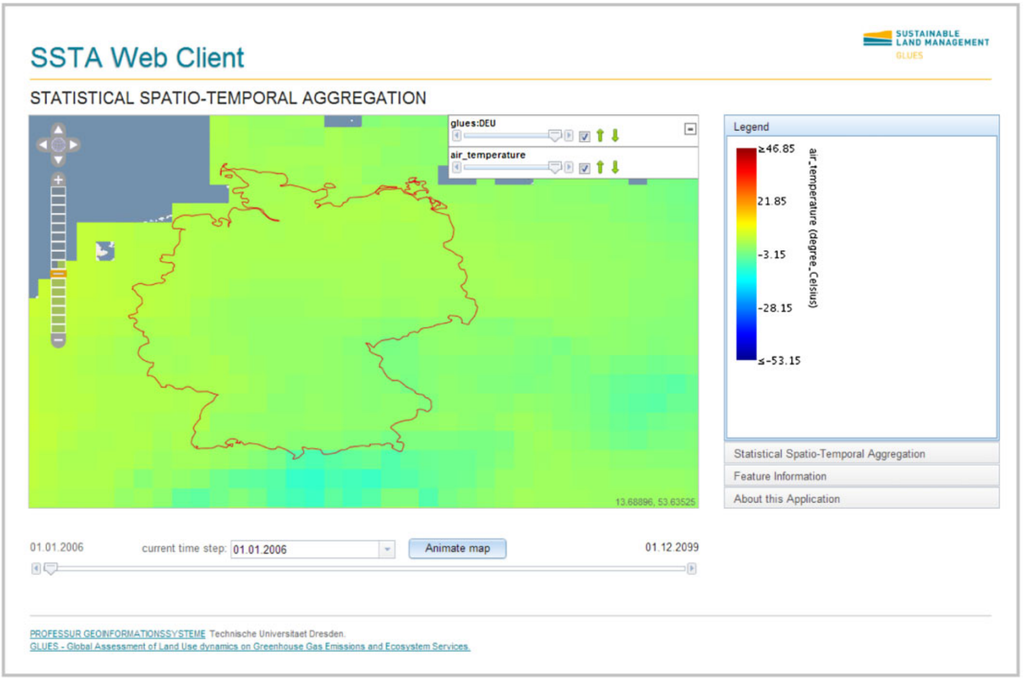


Figure 10 Initial page of the Web client

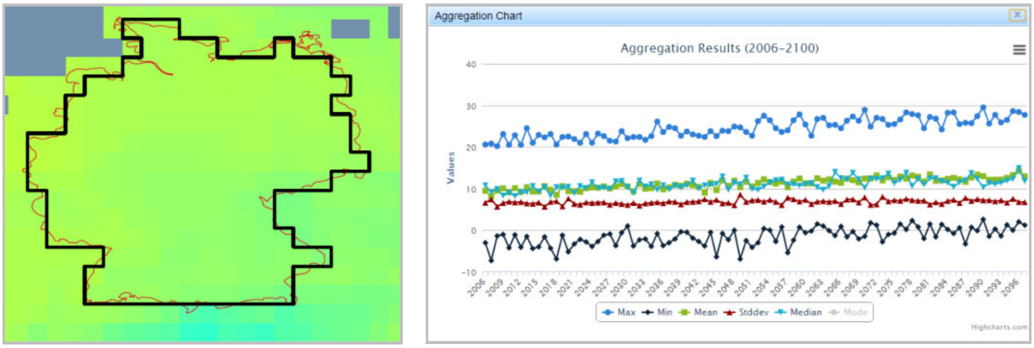
temperature and precipitation datasets. A Web client should visualize the selected data, allow spatio-temporal navigation, and give the user the opportunity to request additional information for a specific map location, such as local precipitation, temperature values, or records of time series data (Figure 10). After having visually analyzed the input datasets, the user should be enabled to easily perform an SSTA for selected time ranges with one or two mouse clicks.

To avoid fallacies or propagation of uncertainties, the Web client should visualize the geometrically aligned target geometry to indicate which geometric objects were considered by the aggregation process (Figure 11a). The Web client should further enable the user to interactively inspect the aggregation results, e.g. via diagrams (Figure 11b).

This use cases stems from the GLUES project (Global Assessment of Land Use Dynamics, Greenhouse Gas Emissions and Ecosystem Services) funded by the German Ministry of Education and Research. Within GLUES the authors implement SDI components in order to facilitate data and model exchange between scientists and to offer stakeholders appropriate tools to access the research results (http://modul-a.nachhaltiges-landmanagement.de/en/scientific-coordination-glues/).

6 Summary and Conclusions

With the objective of finding a classification scheme that enables an automatic derivation of aggregation operations from the measurement type, this article analyzes existing taxonomies of measurement types and statistics (e. g. Campbell 1928; Stevens 1946; Chrisman 1998), and



(a) (b)

Figure 11 Aggregation results: (a) Geometrically aligned target geometry; and (b) Interactive chart visualizing the aggregation results

reviews their suitability. It is proposed that in addition to Stevens's taxonomy of measurement types, a distinction of intensive and extensive properties should be introduced. The provision of necessary information for an automatic selection of permissible aggregation operations can be realized by an extension of the ISO Feature Model.

The SSTA transfers thematic attribute values on the basis of spatial and temporal references into a target geometry with a coarser spatial and temporal resolution. The aggregation process for spatial and temporal geometries is based on the concept of spatial and temporal granularities (Bettini et al. 2002; Camossi et al. 2006). For discrete samples it may be the case that the fine-grained source geometries will not perfectly fit into the predefined target geometries. Matching and alignment procedures that deal with acceptable tolerance levels have been suggested to allow a straightforward SSTA process. This capability is considered useful for quick mash-ups and exploratory ad-hoc analyses where each additional service invocation would be an overhead. Applications that require precise control of the geometric alignment should invoke separate generalization processes beforehand and apply the SSTA process with zero tolerance margins.

A traceable aggregation and alignment process requires structured documentation of the inputs, outputs and the changes that were possibly made to the data. The process interface was designed with traceability in mind and thus allows the documentation of alignment uncertainties. Future work could link this to the Uncertainty Mark Language (UncertML, http://www.uncertml.org/). The definition of a possibly standardized WPS process profile would provide additional benefits. It allows contracting the STSS functionality and interfacing so that lineage statements in metadata could reference it for replicability, which is a core requirement in scientific applications.

Reusability of the generic aggregation process is achieved by offering it as a Moving Code package which allows convenient shipping to different processing instances (Müller et al. 2013). Being an important functionality for geodata analysis, the implemented SSTA tasks can hence be used in many service-based geospatial applications. Since the SSTA process is both versatile and computationally intensive, it is unlikely that a service provider will provide it on a royalty-free public WPS. The Moving Code package can provide service-oriented processing logic in arbitrary environments. In the presented use case, the Moving Code concept was used to supply new or updated implementations to a private WPS instance. However, the package can also be plugged into other software, if the setup and invocation of a separate Web service is not an option.

As also mentioned by Rew and Davis (1990) the design of comprehensive conventions for datasets that meet the requirements of many application domains and, at the same time, allow the analysis with generic tools remains a research topic. As long as the necessary semantic information is not provided in the data models, a less rigorous process could be designed that calculates all technically permissible aggregation parameters and leaves the estimation of the aggregation results to the end-user's discretion.

References

Andrienko N and Andrienko G 2006 *Exploratory Analysis of Spatial and Temporal Data*. Berlin, Springer

Bernard L and Krüger T 2000 Integration of GIS and spatio-temporal simulation models. *Transactions in GIS* 4: 197–215

Bernard L and Ostländer N 2008 Assessing climate change vulnerability in the Arctic using geographic information services in spatial data infrastructures. *Climatic Change* 87: 263–81

Bernard L, Mäs S, Müller M, Henzen C, and Brauner J 2013 Scientific geodata infrastructures: Challenges, approaches and directions. *International Journal of Digital Earth* 6: in press

Bettini C, Jajodia S, and Wang X S 2002 *Time Granularities in Databases, Data Mining, and Temporal Reasoning*. Berlin, Springer

Bonnet P, Gehrke J, and Seshadri P 2001 Towards sensor database systems. In *Proceedings of the Second International Conference on Mobile Data Management*, Hong Kong, China: 3–14

Camossi E, Bertolotto M, and Bertino E 2006 A multi-granular object-oriented framework supporting spatiotemporal granularity conversions. *International Journal of Geographical Information Science* 20: 511–34

Campbell N R 1928 *An Account of the Principles of Measurement and Calculation*. New York, Oxford University Press

Chen C C and Knoblock C A 2008 Conflation of geospatial data. In Shekhar S and Xiong H (eds) *Encyclopedia of GIS*. Berlin, Springer: 133–40

Chrisman N 1998 Rethinking levels of measurement for cartography. *Cartography and Geographic Information Science* 25: 231–42

Couclelis H 1992 People manipulate objects (but cultivate fields): Beyond the raster-vector debate in GIS. In Frank A U, Campari I, and Formentini U (eds) *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*. Berlin, Springer-Verlag Lecture Notes in Computer Science Vol. 619: 65–77

European Commission 2012 NUTS – Nomenclature of Territorial Units for Statistics: Introduction. WWW document, http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction

Foerster T, Lehto L, Sarjakoski T, Sarjakoski L T, and Stoter J 2010 Map generalization and schema transformation of geospatial data combined in a web service context. *Computers, Environment and Urban Systems*, 34: 79–88

Gardner P L 1975 Scales and statistics. *Review of Educational Research* 45(1): 43–57

Gaito J 1980 Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin* 87: 564–67

Gruber T R 1993 A translation approach to portable ontology specifications. *Knowledge Acquisition* 5: 199–220

INSPIRE (Infrastructure for Spatial Information in Europe) 2009 D2.8.I.2 INSPIRE Specification on Geographical Grid Systems – Guidelines. WWW document, http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_Specification_GGS_v3.0.pdf

Intanagonwiwat C, Estrin D, Govindan R, and Heidemann J 2002 Impact of network density on data aggregation in wireless sensor networks. In *Proceedings of the Twenty-second International Conference on Distributed Computing Systems*, Vienna, Austria: 457–58

Kuhn W 2003 Semantic reference systems. *International Journal of Geographical Information Science* 17: 405–09

Müller M, Bernard L, and Brauner J 2010 Moving code in Spatial Data Infrastructures. *Transactions in GIS* 14: 101–18

Müller M, Bernard L, and Kadner D 2013 Moving code: Sharing geoprocessing logic on the Web. *ISPRS Journal of Photogrammetry and Remote Sensing* 76: in press

Norman G 2010 Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education* 15: 625–32

Openshaw S 1984 *The Modifiable Areal Unit Problem*. Norwich, Geo Books

Probst F 2007 Semantic Reference Systems for Observations and Measurements. Unpublished PhD Dissertation, Institute for Geoinformatics, University of Münster

Rew R and Davis G 1990 NetCDF: An interface for scientific data access. *IEEE Computer Graphics and Applications* 10(4): 76–82

Robinson W S 1950 Ecological correlation and the behavior of individuals. *American Sociological Review* 15: 351–57

Stevens S S 1946 On the theory of scales of measurement. *Science* 103(2684): 677–80

Stevens S S 1959 Measurement, psychophysics, and utility. In Churchman W C and Ratoosh P (eds) *Measurement: Definitions and Theory*. New York, John Wiley and Sons: 18–63

Strobl J 2005 Hierarchische Aggregation: Detailinformation versus Datenschutz am Beispiel adressbezogen georeferenzierter Datensätze. *Salzburger Geographische Arbeiten* 38: 163–71

Walter V and Fritsch D 1999 Matching spatial data sets: A statistical approach. *International Journal of Geographical Information Science* 13: 445–73

Wiemann S and Bernard L 2010 Conflation services within spatial data infrastructures. In *Proceedings of the Thirteenth AGILE International Conference on Geographic Information Science*, Guimarães, Portugal

Yuan S and Tao C 1999 Development of conflation components. In *Proceedings of Geoinformatics*, Ann Arbor, Michigan: 1–13

