CARTOMETRIC ASPECTS OF HYBRID ANALYSIS WITHIN GIS

Stefan A. Voser

Semantische Modellierung; *SMATI 99*

KEY WORDS: Cartometry, Spatial Analysis, Hybrid Data Processing, Raster, Vector, Data Structures

ABSTRACT

Cartometry gets applied to geospatial data. This incudes making measurements and queries of coordinates, distances, areas, angles as well as counting features. These functions are basic services for the access to spatial data and for spatial operations like measurements and locational analysis.
These cartometric services have to be established into the aim of integrating GIS and Remote Sensing at structural and operational level. This integrating aim requires a bridging of raster and vector spaces called hybrid analysis. When applying this to cartometry, metrically homogenized services have to be developed.
In this paper, first a generic design of spatial analysis operations is presented. Then, a zoom into various metrics in raster and vector is given to show their systematic deformations compared to the Euclidean one. This is done for the distance function applied to a straight line. The metric deformations depend on the direction of the line and on the raster resolution. A further zoom is made into the metrics of the areal location analysis operation Buffer. Buffer is a simple geometric operation with an area as its result. A comparison of various buffering algorithms in raster and vector is done as a step towards hybrid data processing.

1 INTRODUCTION

Assuming that a geospatial database is the fundament of a digital map, calculating metric measurement like positions, distances, areas and directions based on the stored spatial data is a new form of cartometry. The classical cartometry makes measurements from paper maps using tools like a ruler, a pair of compasses, a protractor, a planimeter etc. [Maling89].
The new approach to cartometry is based on the analytical geometry founded by Descartes in 1637 [Descartes1637]. Its basics are the coordinates describing positions in a coordinate system. For spatial purposes, coordinate systems need spatial semantics. The coordinate systems with spatial semantics are called *coordinate reference systems*. Latter ones are divided in different types, varying in dimension as well as in the shape of reference surfaces and coordinate definitions. See also [Voser98], [OGC98], [OGC99-2] and [OGC99-16].
Therefore the metric nature of spatial data as well as the metric calculations depend on these different coordinate reference systems like geodetic reference systems, map projections etc. For solving and implementing this fundamental field of cartometry, several method libraries for cartometric measurements are required.

The methods for cartometric measurements do not only depend on the mathematical model but also on the *data structure* in which the data is stored. When following the approach of integrated GIS as a simultaneous processing of remotely sensed data, stored in raster structures, together with interpreted vector data, a combined processing of data with vector representation as well as with raster representation is needed. This combined processing leads to the aspect of *hybridity in spatial data processing* [Voser98b].
Spatial analysis is a part of spatial data processing. It covers the interaction of interpreted and semantically modelled spatial data to derive new information. The aspect of hybridity in spatial analysis basically concerns the data structuring and the cartometric measurements based on different metrics in the raster and vector spaces.
The focus presented here zooms into the cartometry and the geometric interaction in the hybrid space. After a short view into the problem of modelling spatial operations and structuring the hybrid space, the discussion leads to the various metrics of the hybrid space. This topic finally is projected onto the buffer operation as a simple geometric function. The buffer operation varies in raster and vector space. These two spaces are compared in order to lead to a universal buffer operation covering the hybrid space.

1.1 What is Spatial Analysis?

The several processes for spatial data may be divided into the four groups:

- data capturing and preprocessing,
- data integration,
- data analysis and
- data visualisation.

Herein, *spatial analysis* is the action and interpretation of ordered, structured and consistently stored spatial data on a geometric, semantic and thematic level. The result of spatial analysis is new information for a new cognition or spatial statement, knowledge and declaration [Fischer96], [Fortheringham94], [Longley96], [Quattrochi97], [Tomlin90]. Referring to [Albrecht96], the following groups of spatial analysis processes may be found:

- Search
- Location Analysis
- Terrain Analysis
- Distribution/Neighbourhood
- Pattern Analysis
- Measurements

Important is that the spatial analysis operations interact at the aspects of geometry, semantics and theme based on clearly defined rules. The entire process includes a combination of all the three aspects which have to be embedded into spatial operators as technical implementations of spatial operations. These operators have their components on the levels *management* (high-level), *controlling* (mid-level) and *execution* (processing at low-level). See e.g. [Voser98a] and section 2.
In the following, we focus on the execution or processing level of spatial analysis. We mainly limit this on the geometric interaction between raster and vector space. The main work is done to examine the metric differences found when comparing metric measurements made in raster and vector space. This approach is required to understand the modelling of the metric problems based on the hybridity of spatial analysis as a simultaneous processing of raster and vector data without conversion.

1.2 What Does Hybridity Mean?

Hybridity results from the different approaches to store and to represent spatial data: the vector and the raster representation. The raster approach allows to represent continuous fields as well as discrete data. Its regular representation limits the geometric degree of freedom, but fills the whole area with data. The vector representation allows field views based on irregular spatial distributions as well as the object view of data [Bartelme89], [Goepfert91], [Hake94].
Concerning the problem of hybridity of spatial data processing, the different structural natures of the representation should be bridged. The traditional implementations of spatial analysis operations are dependent on data structure and data representation. These different representations still require different tools for the processing, or they only include restricted processes allowing some combined and simultaneous processing of raster and vector data [Voser98a], [Voser98b].
Hybridity means to have at least bimorph implementations of the same function, in vector as well as in raster space. But the most important aspect is the interaction between the vector and the raster space on the geometric, semantic and thematic level whithout any conversion between these two spaces. Let us approach the aspects of metrics.

1.3 What Is Cartometry?

Cartometry is the science of *measuring maps and spatial data*. It is the inversion of field measurements. Latter ones catch and describe the geometry in the real word. In the modern GIS technology, cartometry includes the retrievement of (geo-)metric information from geospatial data and the counting of phenomena based on object selections [ICA73], [Maling89]. When focussing on the hybrid extension of data structures, the different metrics in the continuous vector space and the discrete raster space have to be compared. Cartometry includes the following tasks extending the list of [Maling89] p.1:

- measuring coordinates (geographic and projected)
- measuring distances
- measuring angles
- measuring directions
- measuring areas
- measuring heights
- counting the number of objects

2 THE GENERIC SPATIAL OPERATION DESIGN

A spatial analysis task is a simple or complex set of rules of spatial process entities as basic spatial operations. The result of such an analysis is realised by selecting the required input data sets and by combining the spatial rules as operations together with the data to a spatial work flow. Such an example of a location analysis is given in Fig. 1.

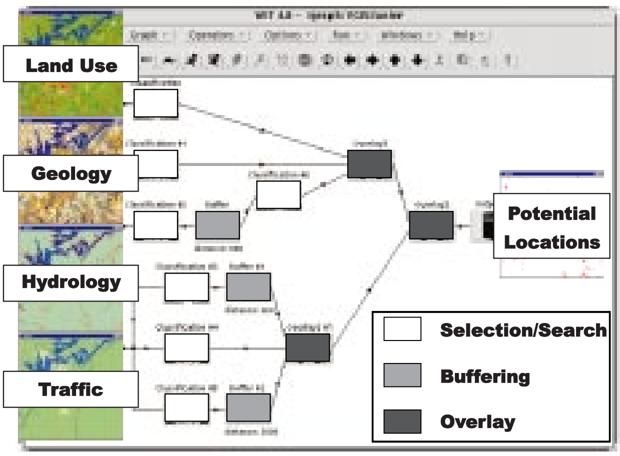


Figure 1: A spatial analysis process: location analysis

2.1 The Requirements

GIS- or environmental analysis tasks are specific to individual Information Communities (ICs) and related to their spatial questions and problems to be solved [OGC98], [OGC99-14]. The accomplishing GIS- tasks exist in a variety of complexity. These tasks can be solved through well defined GIS-Operations. The user's demands are a metadata driven process and operation design. For this, a conceptual framework and generic design of analytical GIS-Operators have to be built up. To fulfill all requirements of spatial analysis, the simultaneous and combinational processing of vector and raster data is needed. For reaching this approach, we zoom into the functionality of hybrid analysis.

All spatial data analysis systems, including image processing and GIS, offer a large amount of algorithms and analytical operations to solve spatial problems. For non-experts, this variety as well as the complexity of operations, mostly of technical nature, often impede fast digital solutions. Necessary system decisions, different data formats, data structures, and data models as well as the lack of compatibility are additional obstacles of GIS for solving different spatially related tasks. What is needed is a broad technical toolbox that is capable of supplying full functionality for a wide range of complex spatial problems. Consequently there are strong demands to design operators closer to the user's needs, but still of universal

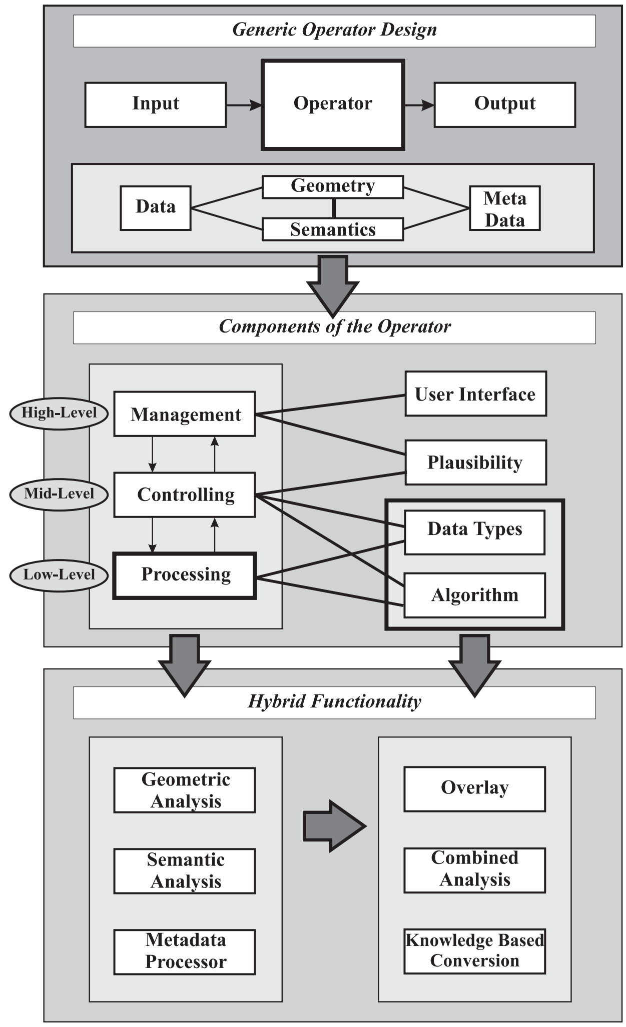


Figure 2: The generic design of an operator

nature. That requires independence from any Information Community and its specific applications.

At the design of each analysis task, these operations have to be modelled in an independent manner. This means, the operations have to be designed on a high level of abstraction as *Universal Analytical GIS Operators*. On this level, the operators are "independent" of any data catalogue, data type or data structure.

On the task level which manages the operation according to the application, the operators only work with metadata where the end-user is responsible for the conceptual task. The operation is defined by the input data and task with the underlying functionality. Both are selected from catalogues and their metadata: The input data is chosen from a data catalogue, the functionality is selected from an operation library and its metadata. The results are data in a data catalogue, attached with new metadata including the results, descriptions and characteristics of the operation (Fig. 2, Fig. 3).

Based on this conceptual, technology independent modelling of an operator, a mapping to the technological tools is needed. For that, requirements of in-

teroperability in geoprocessing have to be fulfilled [OGC98]. The generic design of the operator's functionality should be able to process any kind of geographic data types (point, line, area, nodes, edges, meshes, grid, tin, raster ...). Especially for the integration and combination of vector and raster technology (e.g. for the integration of GIS and remote sensing), there is an increasing need for hybrid analysis. This section represents an approach to design and to structure such universal and generic operators.
An overview is also given in Fig. 3: An operator needs *input*, and its results are the *output*. Input, the operator itself and the output consist of data and metadata, all concerning *geometry* and *semantics* (section 2.2). A zoom into the architecture of such an operator is focused in section 2.3, where the different levels of abstraction are classified as *management, controlling and processing*. In section 2.4 specifications of different components as *user interface, plausibility check* as well as *data types* and *algorithms* are described. The descriptive and cognitive control of the operations is made upon the *metadata* for input, output and the operator. The relationships and mapping between them have to be known (see section 2.5). The increasing need of hybrid analysis asks for *hybrid functionality*. The basic principles for that are shown in section 2.6.

2.2 The Design of Analytical GIS Operators

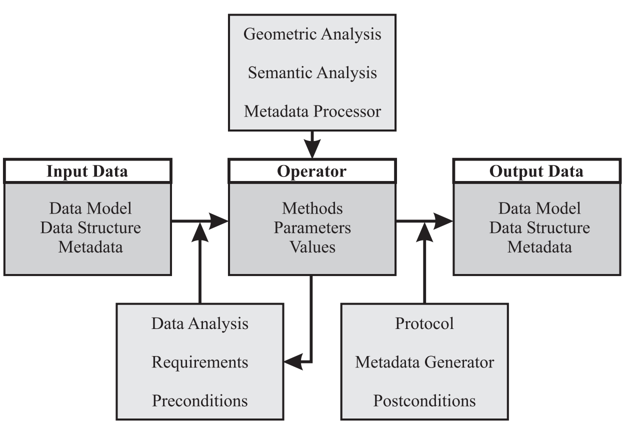


Figure 3: The conceptual design of a spatial operator

The design of Analysis-Operators in GIS predominantly concerns their structure and behaviour. A universal analytical GIS-operator is characterised by its semantics together with the implemented functionality. An operator is able to analyse the input data to choose the appropriate algorithm. As a result, new output data is generated. Such a High-Level-Operator has to be metadata driven. For selecting data and functionality, only metadata will be used. The following survey describes the concept of operators and their input and output data. Their relationships are also shown in Fig. 3.

Input Data: The input data is described through a data catalogue including its metadata and lineage information. The link to the digital representation in the related and implemented data schema has to be known.

Operator: The operator is defined by a generic method. This is specified by its semantics, its algorithms and the characteristical parameters. The operation is described by the operator's metadata. The operator generates specific metadata as well as a protocol for the lineage documentation. The underlying control has to select the correct algorithm that fits the types and structures of the input data. Each algorithm has its own profile. The process control executes the analysis of the data as an inquiry for the correct algorithm. The operator is designed in a polymorph way which includes hybrid operations.

Output Data: The output data is mapped into a data catalogue which may already exist, or which has to be newly generated or extended. It includes metadata and lineage information. Its representation is linked to the underlying data schema. The output data catalogue may be derived from the operator, or the operator has to match the conditions of the data model chosen by the user.

The functionality of a GIS-Analysis-Operator can be modelled by the following three categories: *geometric analysis, semantic analysis* and *metadata processor*. Each category concerns different kind of information.

Geometric Analysis: All geographic data has a geometric component. One of the main goals of spatial analysis is to solve geometric (metric and topological) questions. Generally a GIS analysis produces data with new geometric information. Consequently a GIS-Analysis- Operator translates the geometric input into the new geometry of the output data.

Semantic Analysis: The semantics of the data is given by the Information Community, its conceptual model with all its attributes and descriptions. It is an overlay to the underlying geometry. The geometric process of the operator is linked to the thematic component of the data, a new theme is generated during the combination process with the geometric and semantic overlay. The result is a new semantics.

Metadata Processor: The Operator is controlled and prepared for the processing, based on the metadata of the input data. The metadata is translated to process control parameters. The process produces new metadata. The generation of the output metadata is controlled by the operation's metadata.

The controlling of the operator includes the preconditions that are derived from the input data and the chosen operation with its implemented algorithms.

For the output, the postconditions include the requirements of the output data. The operation has to be protocolled including the lineage and other metadata.

2.3 Different Abstraction Levels of Analytical GIS-Operators

The technical implementation of the described conceptual design of analysis operations asks for concrete solutions. Therefore the operator is divided into three levels of abstraction. The design is a top-down concept because of its user-oriented and user-friendly approach. On *high level*, concerning the *management*, the operator works with metadata only. It is this level which the user is confronted with. On *mid level*, the *controlling* of the operation, including the handling of the data types and the metadata, is performed. On *low level*, the algorithms are *processed* and protocolled.
The universality of an operator is displayed at the highest level of abstraction because of its design which is (almost) independent of data structure and data schema. The interactions between the levels directly concern neighbour levels because of their hierarchy. The core of such a universal operator is on the control level. The performance of the operator is carried out on the processing level.

Management (High Level): On the high level, the semantics of the operation is defined. The data to be analysed are selected from the data catalogue with their associated metadata. The operator is specified upon its parametrisation which is also metadata driven. This level manages the whole operator. It is the constituent part of the analysis task. The user defines the operation upon his cognitive experience.
The data description at this level is independent of any spatial data structure (e.g. raster or vector) or data schema.
Of course, the controlling of data types at the level of user interaction is included to the management level because of the required plausibility check mechanisms.

Controlling (Mid Level): On the mid level, the controlling of the operation is carried out. The input data, selected on the high level, is linked to the corresponding data representation on the geometric and structural level. The operator analyses the metadata, given on the management level, and chooses that algorithm which matches the data types and the other requirements. Here, the requirements for hybrid analysis functionality arise because of the different data representations.
On this level, the cognitive semantics of the data and operation gets lost, the information is translated into syntactically structured information for the processing in which the cognitive context is unimportant.

Processing (Low Level): The low level is the processing level. The algorithm processes the data numerically, all characteristics of the operation are processed and sent back to the mid level, both data as well as metadata.

2.4 Specification of Analytical GIS-Operators

Based upon the operator design described above, the specification of a generic high level spatial analysis operator is necessary to create instances of an operator class "high level operator". An instance of such a class could be any high level operator, e.g. buffer, overlay, shortest path. At the current state of development, the specification introduced in this section does not claim to be complete or formal in mathematical terms. It rather serves as a more detailed description of components in order to approach a formal specification. It is the attempt to outline the general conditions necessary to create high level operators. Therefore, it has to be regarded as a step towards the realisation of operator design.
The variety of instances of high level operators, their design and specification comprise different aspects. The data structure independence on the management level has to be implemented through a data type driven polymorphism at the control level. Due to the polymorphism, a high level operator is able to execute different algorithms or sequences of algorithms on the processing level.
In addition, cognitive and semantic aspects can not be neglected in the operator design. Yet the structure of data plays a central role in the building process of high level operators concerning different parts of the specification.

2.4.1 User Interface

The user interface specifies what kind of interaction can take place between the user and the operator. User interaction includes any form of communication between the user and the operator, such as:

operator control

- Selecting data through a data catalogue
- Input or determination of parameters for the operator

user support, documentation

- Access to documentation and description (metadata) of operations and data
- Transparency of the operator by describing architecture and algorithms

process control, messages

- User comfort with messages and warnings
- Error control

2.4.2 Plausibility Control

Plausibility control of an operator should be able to determine whether an operation is correct semantically and thematically. The verification of this advanced form of semantics has to be done one level above the creation of the correct data type management. It comprises several aspects:

Semantics Control: Analysis and comparison of semantics in input and output data should enable the operator to distinguish between allowed and forbidden combinations. Predefined semantic results for given input/output combinations could also be attached to the operator (*semantic templates*).

Property Control: A relation between geometric and semantic properties of data should be established. For example, geometric accuracy control should decide which combination of different resolutions is allowed. Tolerated ranges of accuracy for different semantics of spatial data can be supplied.

Lineage Control: Lineage information attached to output data could include plausibility control documentation and results.

2.4.3 Data Types

Assuming that spatial data are organised in data types, it has to be exactly specified which data types can be processed by an operator and which data types will be produced by the operator as result. Data types distinguish between different representations of geometry within spatial data. The hybrid functionality is required for a simultaneous and combinational processing of raster and vector data types.

Catalogues: Specification of allowed data types and its definitions:

- data model (topological vector model, task specific models, ...)
- structures (raster, vector, tin ...)

Combinations: Matrix of input-output-couples (combinations of input and output data types related to an operator or its algorithms).

Restrictions: Restrictions for operator use with regard to the allowed *data couples*.

2.4.4 Polymorphism

Polymorphism stands for different algorithms related to different data types of the same high level operator. Program structures are created on the control level by decisions made according to data types, user parametrisation and semantic specifications. To include polymorphism into an operatio, the following conditions have to be fulfilled:

Uniqueness: Unambiguous decision rules.

Completeness: There must be one function for every combination of allowed input data (types).

Correctness: A method to test consistency and correctness of decision rules.

Extensibility: Operations at control level must be extensible to new data types and new algorithms.

2.4.5 Algorithms

On the processing level, standard software development guidelines have to be regarded. It is supposed to choose an object-oriented approach with certain advantages related to the described operator design.

Transparency: Understandability and clearness of algorithms by documentation.

Uniqueness: Avoiding identical processing as well as identical geometric results within different high level operators.

Redundancy Free: Non-redundant implementation of algorithms (modules, function libraries, etc.).

2.4.6 Application Specification

Optional is an extension of high level operators towards the restricted and/or expanded use in certain information communities. That would require:

Limitations or Reductions: Restriction in functionality according to a reduction of defined operations in addition to the specifications mentioned so far.

Specialisation: Particularisation of certain operators through the input of domain specific data and functionality (e.g. integration of domain specific rules and algorithms).

2.5 Metadata

Metadata is used to manage all operators on the highest level of abstraction. On the interface or management level, only metadata controls user interaction. Explicit and implicit metadata is analysed to control the operator on the mid level. The metadata has different meanings on the three levels as described in section 2.3.
In the following, the processing of metadata for high level-GIS-operators is divided into two parts: metadata of spatial data and metadata of the operator.

2.5.1 Metadata of Data

The metadata describes the data on a high level of abstraction in a data catalogue with its related information. The information includes the link to the database and to its data schema in which all other information is stored implicitly.

Management (High Level): The metadata of the data describes the content and the conceptual organisation of the data together with not in the data model included information and characteristics as data quality, lineage information, sources etc.
Metadata describes the thematic and semantic model of the data including its appropriate data type with regard to the current resolution.

Controlling (Mid Level): On the mid level, the semantics of the metadata of the high level has no direct meaning, but it serves the controlling of the operation. The information of the data types in which the data are represented mainly is important for the control of the respective polymorphism. The main aim on the mid level is the control of the structural information of input and output data. This includes the control of the polymorphism of the operator.
The operator translates the semantics of the metadata to parameters used for processing, whereas the semantics of the data is of no relevance to the parameters of a process.

Processing (Low Level): On the low level, metadata focuses on the values and categories of the parameters. The characteristic parameters of the operator and their values are instanciated and assigned to the data to be processed.

2.5.2 Metadata of Operators

The metadata of an operator controls their spatial analysis process. They map and process the metadata of input and output and protocol the operation and the lineage.

Management (High Level): The metadata of the operator describes the task, the functional behaviour and the required information. The operator's metadata includes the semantics which is related to the data.

Controlling (Mid Level): The operator's metadata controls the operation. On this level, the mapping of the data types to the algorithm of the polymorph implementation takes place.

Processing (Low Level): This level generates the input data for the lineage of the operation which is related to the derived data.

2.6 Hybrid Analysis

Hybrid analysis is part of polymorph implementation of GIS-analysis-operations. Hybrid analysis is needed for the integration of GIS and remote sensing, for terrain analysis (DTM's in raster format with overlay of vector data etc.).
As input data, we have raster and vector data which have to interact correctly. The results of a hybrid analysis are new raster data, vector data or both of them. Some examples of hybrid analysis functionality:

- Fencing off raster processes by the overlay of vector data
- Verification of geometry and attributes
- Transferring extracted geometry from raster to vector
- Transferring attributes from image interpretation to vector data
- Transferring attributes form vector databases to raster data
- Using raster information for determination of uncertainty of vector data
- Using raster to manage accuracy of both spaces
- Equalising Data

The complexity of the hybrid functionality may be shown by the following survey:

Projecting Information: (Fig. 4)

- geometric overlay
- semantic overlay

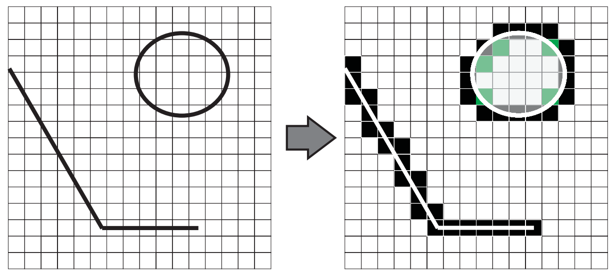
Combined Analysis:

- derivation by aggregation
- derivation by interaction
- accumulation by information transfer

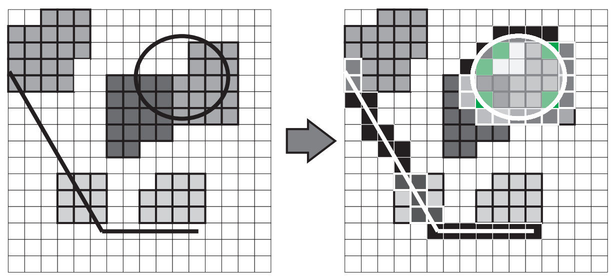
Knowledge-based Conversion:

- knowledge-based conversion
- extraction of objects

The full range of spatial analysis may be reached with operations that interact with raster and vector data without the need for conversion. The requirements of hybrid analysis are divided in geometric interaction, semantic analysis and metadata processing.



Geometric Interaction



Semantic Interaction

Figure 4: The hybrid interaction: at the top the geometric interaction, at the bottom the semantic interaction

Geometric Analysis: (see at the top of Fig. 4)

- *positional location:* generating the relation of coordinates of different data and identifying identical positions of geometric primitives.
- *linking geometric features:* identifying features covering the same location.
- *extracting and transferring geometry:* generating new geometry by extraction, transfer and interpolation.

Semantic Analysis: (see at the bottom Fig. 4)

- *geometric interpretation:* giving new semantics to the data.
- *thematic analysis:* linking, projecting and deriving new information.
- *combination of geometric and thematic interpretation:* combined analysis of geometric and thematic information.

Metadata Processing:

- analysis of metadata,
- description of the geometric, semantic and combined analysis,
- description of the semantic analysis,
- statistics about geometric overlay.

3 METRICS IN THE HYBRID SPACE

3.1 Different Metrics for Raster and Vector

Metrics is defined by rules for calculating distances. Generally, metrics fulfills the following conditions for the distance function:

- Distances between two points are positive; they are zero between a point and the point itself:

$$d(P,Q) \geq 0; d(P,P) = 0.$$

- The distance is independent of direction:

$$d(P,Q) = d(Q,P).$$

- The distance from one point to another is shorter or equal compared to the way over a third point (*triangular inequation*):

$$d(P,R) \leq d(P,Q) + d(Q,R).$$

The type of the *distance function* is the specific characteristic of each metric space. In the following, only four kinds of metrics are used for the focus on hybrid analysis of planar two-dimensional data. Metrics in vector space has no geometric resolution restrictions except when mapping data to the digital (binary) representation. In the 2D space, point, lines, areas and fields may be modelled, whereas in raster space, data is restricted in its structure, given by the choreography of the raster cells (alignment and cell size). Let us have a look at the planar metrics of raster and vector structures. The length of a straight line is given as the following:

- Euclidean metrics

$$d_e(P,Q) = \sqrt{(x_Q - x_P)^2 + (y_Q - y_P)^2}$$

- City-Block metrics

$$d_4(P,Q) = \mid x_Q - x_P \mid + \mid y_Q - y_P \mid$$

- Chessboard metrics

$$d_8(P,Q) = \max\left(\mid x_Q - x_P \mid, \mid y_Q - y_P \mid\right)$$

- Chessboard-Diagonal metrics

$$d_{8/diag}(P,Q) = \mid\mid x_Q - x_P \mid - \mid y_Q - y_P \mid\mid + \sqrt{2} \cdot \min\left(\mid x_Q - x_P \mid, \mid y_Q - y_P \mid\right)$$

The most known metric space is the Euclidean one. It is the classic one. It is defined in a coordinate system with straight and orthogonal axes with linear and continuous units. A representative of this metrics is

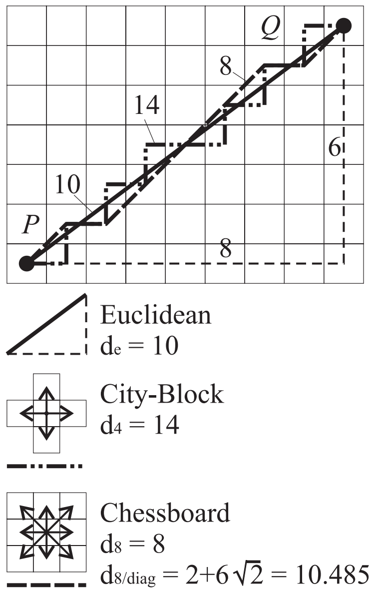


Figure 5: Metrics in the hybrid space

the two-dimensional planar cartesian coordinate system including the rules of planar geometry, trigonometry and so on. Within it, a straight line may have any beginning- node, any direction and any length. The rules of the Euclidean metrics are used in the vector space.
In raster space, point locations are restricted. Those are limited by the choreography of the regular arrangement of the raster cells (raster width and orientation). Most raster representations used in spatial data define an orthogonal grid that is arranged horizontally. In such a raster, a line is represented as a chain of neighbouring raster points or cells (Fig. 5). The length of such a line is calculated by counting the steps to the neighbouring cells. Each step has the value one. Finally the distance is counted by the multiple of the raster width:

- In the *City-Block* metrics, only horizontal and vertical jumps to the next cell is allowed. A point has four neighbours or allowed steps.
- In the *Chessboard* metrics, the counting steps are extended to the diagonal. A point has eight neighbours or allowed steps, each step gets the value one.
- The *Chessboard-Diagonal* metrics is an extension of the Chessboard metrics. The steps along the diagonal get the value $\sqrt{2}$.

In Fig. 5, a sample of these different metrics is given. In the example, the line has a horizontal difference of 8, a vertical difference of 6. The following lengths were calculated:

Euclidean	City-Block	Chessboard	Chessb.-Diagonal
10	14	8	10.485

3.2 The Length of a Straight Line

The example above shows that the various metrics of the raster space deviate from the Euclidean one. These deviations depend on the line's direction. For the following comparison, a straight line is rotated along one end point. The other endpoint describes a circle in the Euclidean space. Each line represents one radius of the circle. This circle is projected onto the raster space. For that resampling process, different resolutions are examined and the lenghts are calculated as a function of the direction. What we want to know about the different kinds of metrics is:

- How does the length depend on the resolution?
- How much do the various metrics differ within the same resolution?
- For which resolutions do the lengths fit the trendfunction of each kind of metrics sufficiently?

3.3 The Trendfunctions

The various metric distance functions in the raster space have a *systematic deformation* compared to the Euclidean one. The reason for that is the regular discretisation of the resampling process, its resolution and type of distance function. The metrics results from the resampling: There are two orthogonal main directions of true length; for the other directions, length distortion is facting.

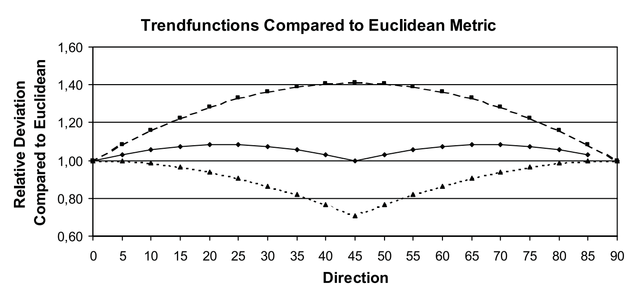


Figure 6: The metrical trendfunctions in raster space

In raster space, as shown above, the lenght is calculated by different methods regarding the resampling process. But for all kind of raster metrics, the distance function depends on the resolution of the resampling process as well as of the direction. The systematic deformations may be described by a *trendfunction* which describes the distortion regarding the

Euclidean space for an infinitesian resolution. In the example of Fig. 6, the metric trendfunctions were calculated for the resolution 10000.

- The *City-Block* metric is longer or equal to the Euclidean one. Equality is only reached for horizontal and vertical direction. The longest deformation is along the diagonal (45^o). Its scale factor is $\sqrt{2}(\approx 1,414...)$ or about 41,4% too long.
- The *Chessboard* metric is smaller or equal to the Euclidean one. Equality only is reached for horizontal and vertical direction. The maximum deformation is along the diagonal (45^o). Its scale factor is $\sqrt{2}/2(\approx\ 0,707...)$ or about 29,3% too short.
- The *Chessboard-Diagonal* metric is longer than or equal to the Euclidean one. Equality is reached for horizontal, vertical and diagonal direction. The maximum distortion is at $22,5^o$ and $67,5^o$. Its maximal value is 8,2% too long.

3.4 The Approximation of the Trendfunction

The trendfunction is the ideal deformation function of an infinitesian resolution compared to the Euclidean metrics. An infinitesian resolution is never reached for raster data. So, only approximations of the trendfunctions may be implemented. Therefore, we want to know:

- How exact can we approximate the trendfunction at a specific resolution?

For that, we analyse the metric behaviour of a straight line for the following resolutions: 1, 2, 3, 5, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 100, 150, 200, 500, 1000. The resolution is given here as the number of cells along a horizontal line of length one.

We calculate the length of the rotated line. For that, the line is rotated in steps of 5^o from 0^o to 90^o in each resolution. These lengths are compared to the corresponding trendfunction- length of each direction. For each resolution, we select from all directions the maximum relative deviation from the trendfunction as given as the following:

$$v_{max} = \max\left(\mid \max\left(\frac{d_{res}}{d_{trend}} - 1\right) \mid , \mid \min\left(1 - \frac{d_{res}}{d_{trend}}\right) \mid \right)$$

This deviation describes the maximum approximation difference from the trendfunction at a specific resolution. In Fig. 7, these deviations from the trendfunction are presented as a function of the resolution. The graphs show the maximal deviation at each resolution, first as an overview at a wide resolution spectrum, and then as a zoom to the low and high resolution spectrum.

These maximum relative approximations to the trendfunctions of the raster space look similar for all three metrics. The approximation becomes better the higher the resolution.

In table 1, the raster metrics are compared. The table shows the resolutions for which the relative deviation falls into deformation classes. E.g. for all kinds of metrics, the trendfunction is approximated better than 2% for a resolution of 30, meaning that a line with a length of thirty times the raster size along one direction approximates the trendfunction with a deformation of less than 2%.

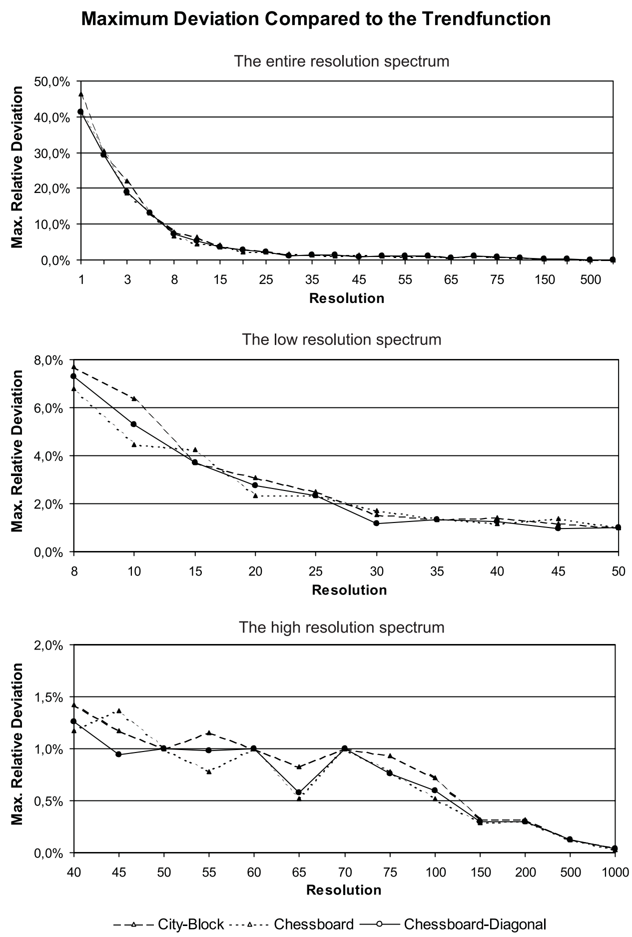


Figure 7: Maximum relative deviation from the trendfunctions - a zoom into the resolution spetrum

Relative Deviation	City-Block	Chessboard	Chessb.-Diagonal
>10%	5	5	5
<5%	12	12	12
<2%	30	30	30
<1%	60	50	50

Table 1: Maximum relative deviation from the trendfunctions at a specific resolution.

We remember that the above made examinations describe the maximum distortion of length at a specific resolution. Therefore trendfunctions and their deviations were used. In the graphs of Fig. 8, for each raster metrics, the maximum relative deviation, the mean deviation as well as the standard deviation from the trendfunction are shown. The mean and standard deviatian are calculated over all directions. These deviations from the trendfunctions seem to be approximately proportional to the reciprocal of the resolution ($\sim \frac{1}{resolution}$).

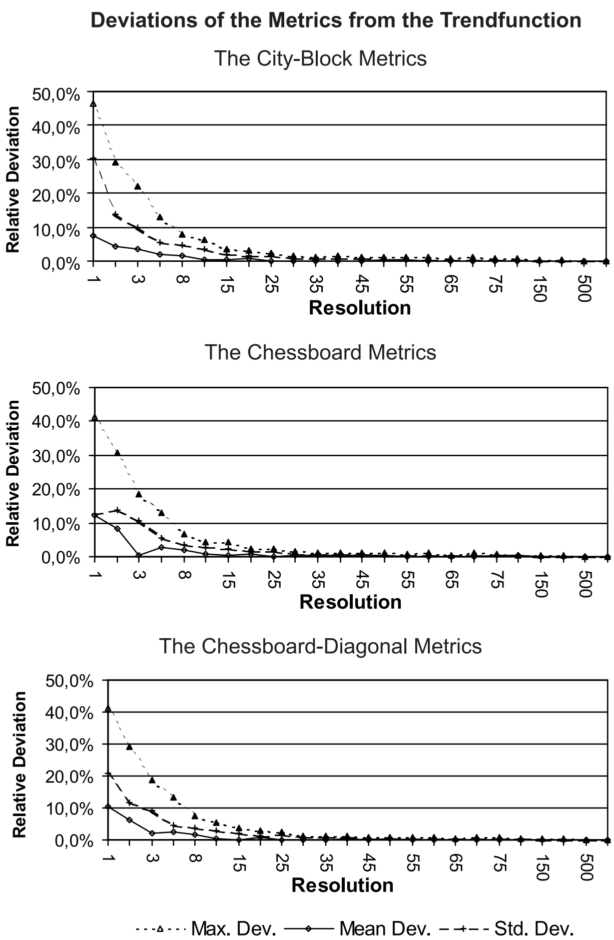


Figure 8: Intrinsics of the various metrics - the deviations from their trendfunctions

3.5 The Approximation of Euclidean

Now, we know the behaviour and the approximation of the trendfunctions as the theoretical deformation model depending on the direction. Theoretically, a reduction of the length according to the trendfunction has to be done.
We still need to know how good we approximate the Euclidean metrics without considering the trendfunction. Let us have a look at the approximation of Euclidean by the various raster metrics.

Metrics:	City-Block	Chessboard	Chessb.-Diagonal
Max. dev.:	41,4%	29,3%	8,1%
Mean dev.:	25,8%	10,0%	5,4%
Std. dev.:	$>$ 14,1%	$>$ 9,2%	$>$ 2,7%

Table 2: Maximum, mean and standard deviation of the metrics comparing to Euclidean.

In Fig. 9, the superimposition of the trendfunctions and their maximal relative deviations are given. It shows the maximal relative deviation from the Euclidean distance within each resolution. Fig. 10 and table 2 show the maximal, the mean and the standard deviation of each trendfunction to the Euclidean distance.

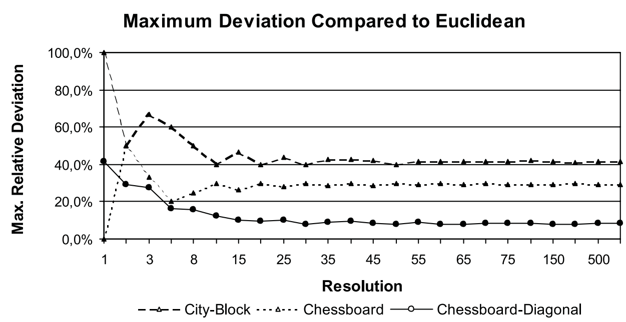


Figure 9: The maximum deviation from Euclidean

As seen in Fig. 9, the Chessboard- Diagonal metrics matches the Euclidian metrics best, but its systematic deformations still are not negligible. There is consequently is a need for including such models to consider these systematic metric deformations within spatial operations.

3.6 Hybrid Metrics and Spatial Operations

Metrics is the fundament for making measurements in the sense of cartometry. Based on it, we can assign units to measurements and, very important, we have the opportunity to use standardised size information. Nevertheless, there exist several different units that have to be converted into a homogenised unit system as e.g. the metric system.
Under the aspect of hybrid spatial processes, the metric distortion management has to be included into the spatial service "calculate distance" which is a spatial operation. Because the metric model is a function of the coordinates, all metric calculations also are operations depending on the coordinate reference system.
The problems of technical implementations of managing the metric distortions cannot be given here.

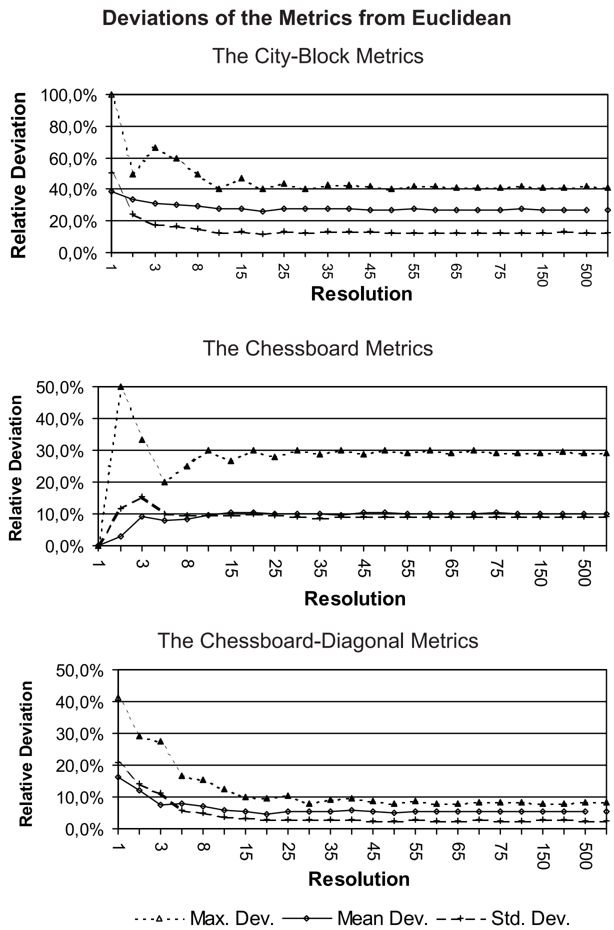


Figure 10: Extrinsics of the various metrics - the deviations from Euclidean

4 A SAMPLE OF HYBRID METRICS: THE SPATIAL OPERATION BUFFER

Until here, we met the fundamentals of metrics, especially various important metric distance functions together with their trendfunctions compared to the ”standard” metrics of the Euclidean space.
In order to understand how to embed the metric processing of hybridity into spatial analysis operations, we have to enlarge the zoom from the linear distance function to the concept of spatial operations and their implementations. In the following, the operation Buffer is chosen because it is a simple area operation. Its metric behaviour according to different kind of algorithms is discussed.

4.1 The Spatial Semantics of Buffer

Buffer is a function of locational analysis. The buffer function defines a zonal location around a feature of the type point, line, or area. Buffer describes a boundary at a constant distance around the source geometry (See Fig. 11).

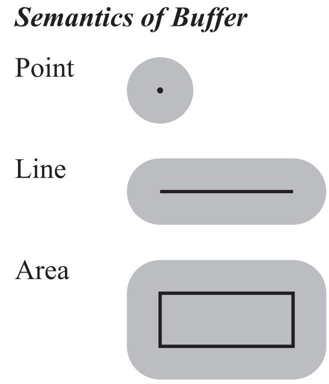


Figure 11: Low level semantics of the operation Buffer

The simplest case is buffering a *point*. Its result is a circle with the radius equal to the buffer- distance. Buffering a *line* is an extension of buffering a point. The line gets the centre line of the buffering area with the constant buffer distance. For a straight line, the buffer’s boundary is described by parallels at both sides of the line. At the endpoints, the boundary describes a circle. Buffering an *area* is similar to buffering point and line. The area is extended similar to the line. Within many implementations, a negative buffering distance diminishes the area. Negative buffer distance only is valid for areas. The semantic description given here is valid for both raster and vector space. But in vector space, there also exist more options for line and area buffering.

4.2 Implementational Aspects of Buffer

In vector space, buffer has a wider degree of freedom for modelling the buffer’s shape result. A line may be described with a direction, given by the starting point and the end point of the line. Knowing this direction of the line, it is easy to assign a one-sided buffer at the left or at the right. The only problem is the previous identification of the line’s direction. Another option for line buffering is the shape at the ending of the line. It may be a circle, a straight line or an extending square.
In raster space, there do not exist any structural differences between points, lines and areas. They differ only in the amount of identified cells and the choreography of neighbouring raster cells.
A *point* is represented as a single cell. A line is a simple chain of cells. The *line*’s shape depends on the used neighbourhood environment: four cell environment (4ce) or the eight cell environment (8ce). (See also section 3.1). For an *area*, it is similar to a line except that the neighbouring cells are not arranged as a chain. Important once again is that the cell locations are restricted to the raster definition.
As shown in section 3.1, there exist various metrics in raster space. Related to that, also various algorithmic implementations of buffer operations exist.
In Arc/Info 7.2 from ESRI which is used for the ex-

amination, two main characteristics of implementations exist. Both are capable of buffering points, lines and areas. The functions for applying Buffer within the raster tool Grid of Arc/Info 7.2 [ESRI98] are:

- Eucdistance
- Expand
- Shrink

Eucdistance calculates the shortest distance to the buffering feature according to the distance function of the Euclidean space. The buffer distance is used to classify slices in order to get the buffering zones according to the buffer distance. For this algorithm, any value (real or integer) is valid for buffer distance.
The algorithms *expand* and *shrink* work in a similar way, but different comparing to Eucdistance. Expand is buffering with a positive distance, shrink is buffering with a negative distance. They are designed to process within the horizontal and vertical main directions of the eight cell environment. These two algorithms only accept integer value as the number of cells as buffer distance. So the buffer distance is restricted according to the raster resolution.

4.3 A Metrical Zoom to the Hybrid Space

The Buffer operation is two-dimensional, its result is an area feature. We have to examine the operation in the hybrid space the same as the metric distance function. To understand the metric behaviour of the buffer operation in the hybrid space, we zoom into the line buffer similar to the distance function in section 3. We look at the direction, the raster resolution and the buffer distance. All three parameters affect the value of the buffer area in its metric behaviour.

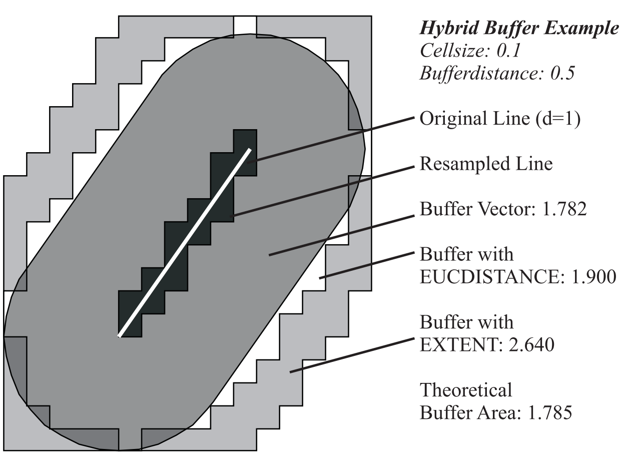


Figure 12: Extrinsics of the Chessboard-Diagonal metrics - the deviations to Euclidean

In Fig. 12, an example of the buffer operation in raster and vector is given. A line of lenght 1 and the direction 55^o is resampled with raster size 0.1. The buffer distance is 0.5 or 5 cells. The results of the theoretical area value as well as the calculated ones are given in table 3. There you also find the scale factor of the area value compared to the the theoretical area value.

Theoretical	Vector	Eucdistance	Expand
1.785	1.782	1.900	2.640
-	0.998	1.064	1.479

Table 3: Results of buffering a line of lenght 1, direction 55^o, raster size 0.1 and buffer distance 0.5.

	π	π_c	π_a
Value	3.141592	3.141553	3.141433
rel. approx.	-	0.999987	0.999949

Table 4: The approximation of π for circumference and area.

4.3.1 Buffer in Vector

As to see in Fig. 12 and table 3, the buffer area in vector always gets a too small value. This is because the circle is discretized und represented by vertices which are connected by straight lines. This fact diminishes π. The circle is approcimated by 360 line segments representing each 1^o of the centriangle. π gets two different approximations for the circumference and for the area as the following:

- π_c: approximation for the length of circumference
$$\pi_c = 360 \cdot \sin(1/2^o)$$
- π_a: approximation for the area
$$\pi_a = 360 \cdot \sin(1/2^o)\cos(1/2^o)$$

The approximation of π is given in table 4. Both approximations are smaller than π itself. π_a is even smaller than π_c. π_c is 13 ppm to small, π_a is 51 ppm to small.
The theoretical area value of the line buffer is calculated based on the length l and the buffer distance d as the following:

$$A = 2 \cdot l \cdot d + d^2 \cdot \pi$$

When looking at the example in Fig. 12, processed in Arc/Info, $\pi_{A/I}$ is even smaller.

4.3.2 Buffer in Raster with Eucdistance

As already described, the buffer area in metrics depends on the buffer distance, on the rastersize as well as on the direction of the buffered line.

The metric behaviour of the operation Eucdistance of the Grid modul of Arc/Info is shown in Fig. 13. There,

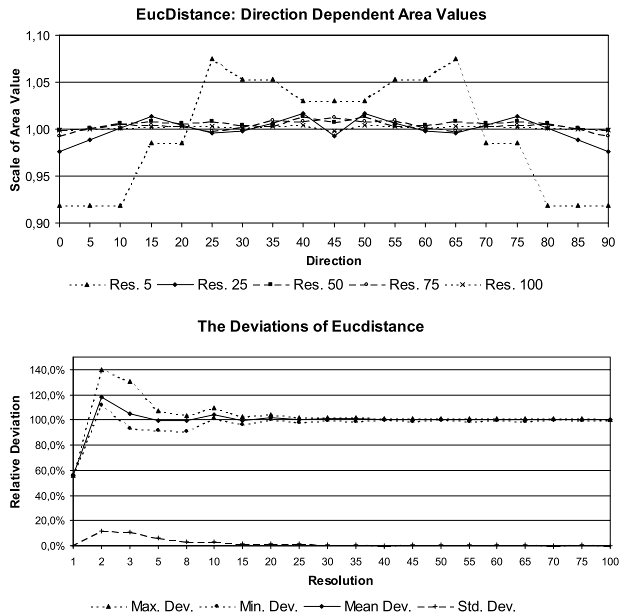


Figure 13: The metric behaviour of Eucdistance. At the top the direction dependent area values for several resolutions. At the bottom the deviations of Eucdistance.

Res.	Max.	Min.	Mean	Std.
5	107.54%	91.86%	99.52%	6.04%
25	101.71%	97.92%	99.96%	1.20%
50	101.40%	99.02%	100.52%	0.42%
75	101.23%	99.28%	100.32%	0.52%
100	100.45%	99.78%	100.18%	0.18%

Table 5: The relative deviations of Eucdistance for the sample resolutions.

the area values are compared with the theoretical one. The graph at the top shows the area scale deformation for the directions in steps of 5^o from 0^o to 90^o for the resolutions 5, 25, 50, 75 and 100. The graph at the bottom shows the various deviations for all directions within each resolution.

In table 5, the deviations for the sample resolutions are given.

It can be seen that the areas resulting of Eucdistance only have little dependence from the direction except for low resolution. The higher the resolution, the better the approximation of the theoretical area value.

For a resolution of 30 or higher, the standard deviation within a resolution gets smaller than 1%, and the maximum deviation gets smaller than 2.2% compared to the theoretical area value.

4.3.3 Buffer in Raster with Expand

For the operation Expand of the raster modul GRID of Arc/Info, a similar examination is done.

The metric behaviour of the operation Expand is

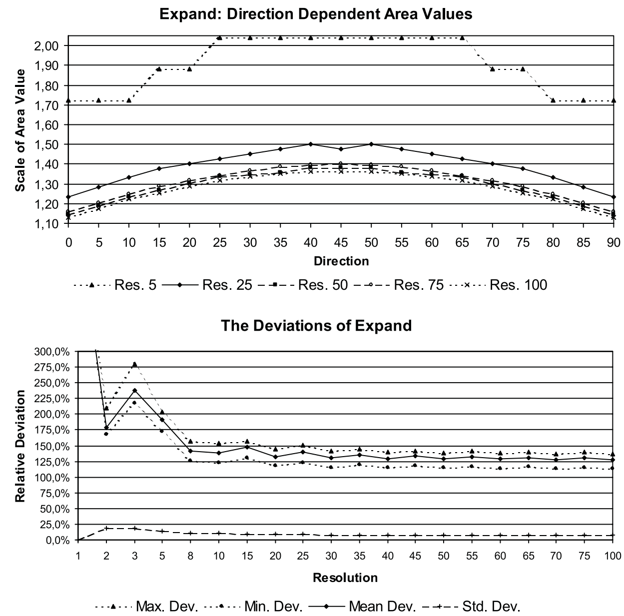


Figure 14: The metric behaviour of Expand. At the top the direction dependent area values for several resolutions. At the bottom the deviations of Expand.

Res.	Max.	Min.	Mean	Std.
5	203.88%	172.88%	190.67%	14.09%
25	150.02%	123.40%	139.19%	8.70%
50	138.25%	114.26%	128.99%	8.11%
75	140.31%	115.77%	130.62%	8.18%
100	136.33%	113.14%	127.64%	7.94%

Table 6: The deviations of Expand for the sample resolutions.

shown in Fig. 14. There, the area values are compared with the theoretical one. The graph at the top shows the area scale deformation for the directions in steps of 5^o from 0^o to 90^o for the resolutions 5, 25, 50, 75 and 100. The graph at the bottom shows the various deviations for all directions within each resolution.

In table 6, the deviations for the sample resolutions are given.

It can be seen that the areas resulting of Expand depend on the direction. The maximal deformation is along the diagonal. This is because Expand is a horizonal algorithm not concidering the diagonal. The metric behavior of Expand let us remember the City-Block metrics. Both have a strong deformation along the diagonal. Whereas the City-Block metrics has no deformation for the horizontal and the vertical, Expand has a systematic deformation for the area value also in the horizontal of more than 10%. One reason for that is because Expand does not consider the diagonal metrically. Another reason is because Expand finds the raster cells outside the initial source feature boundary chosen for buffering. When looking at buffering a line in raster, the line already represents an area feature around which the buffer was calculated. So buffering a line with Expand is handled as an area feature. This leads to a systematic deformation of buffering a line with Expand. The mean deformation within a resolution is about a factor 1.3 or about 30% too large.

4.3.4 Comparing Eucdistance and Expand

We have seen two different algorithms applied to buffer a line within the raster modul Grid of Arc/Info. The Eucdistance method considers also the diagonal direction for buffering and calculates the distances from the raster cell center. For masking the buffer zones around features with the buffer distance, real numbers are valid, and the distances are calculated from the centre of the cell. Expand however only considers the buffer distance horizontally and vertically. Expand does not consider it from the center of the cell but from the raster cell boundary. This leads to a systematic deformation of the buffer area.
Eucdistance only has few distortion, especially for low resolution. It approximates the theoretical buffer area very well.
Expand has systematic distortions, for horizontal and vertical direction as well as for the others. The minimal deformation is for the horizontal and vertical directions, the maximum for the diagonal. Comparing to Eucdistance, Expand shouldn't be used for buffering without taking its metric deformations in count.

4.4 Other Aspects of Buffer

Until here, we observed the characteristics and behaviour of the *simple feature* straight line to be buffered. As Buffer is an areal operation, one problem exist only in raster space. The buffer operation expands the geographic extent around the feature to be buffered. So before buffering a feature in raster space, the geographic extent has to be checked and appropriately expanded.
When processing *many features* together, buffer gets conflict zones for neighbouring features laying next to each other less than the double buffer distance. In vector space, those features may be handled as single entities. They may overlap other buffer entities. But in raster space when processing such features together, such an overlap is not possible because of the data structure. Because of that, the buffered zones get smaller. In both raster and vector space, the merging of features is possible.

5 FUTURE AND OTHER ASPECTS

Not discussed in this paper are the requirements for structuring the hybrid space or the problem of managing data quality like the accuracy.
A *hybrid data structure* is the infrastructural requirement for the hybrid data processing of geospatial data. Until now, hybrid data structures may only be found in science, but many systems already have bimorph structures to process raster and vector data separately. Examples are Erdas Imagine, Arc/Info and others.
In literature, two approaches to expand the traditional and bimorph structures to a hybrid one may be found. The extension of the metrical vector space are to be found e.g. in [Molenaar91], [Glemser98] or [Glemser98a]. The two-dimensional metric primitives point, line and area get a new brother: the raster element, representing both point or area cells. For this approach, the still open question is how to expand the structural primitives for managing topological relations. This hybrid model may be seen as the metric extension of the vector space.
The other approach to be found in literature as e.g. in [Winter98] expands the (metric) raster space with structural primitives for raster. This approach has point, line and area features in metric and structural sense, but with metric restrictions to the raster grid.
The hybrid space also requires an adopted *accuracy* management. In vector space, the geometric accuracy is assigned to the coordinates and the shape of the features. The thematic accuracy may be assigned to the attributes. In raster space, because of the fixed resampling geometry, the thematic as well as the geometric accuracy is assigned to raster attributes, and often, they may not be distinguished. [Fritsch98], [Glemser98a] and [Glemser98] describe an approach to translate the vectoral accuracy to the raster space.

6 CONCLUSION

For reaching real interoperability in processing raster and vector data simultaneously and combinationally without any data conversion, several aspects have to be fulfilled. For this, a hybrid data structure, a harmonised metrical concept as well as a hybrid accuracy concept are required.
The metric components of geospatial data carry the information of position, shape, size, orientation and (geo)metric morphology. These characteristics differ in the various vector and raster representations as well as in algorithms of spatial operations. Because of that a comparison of the metric behaviour of the distance functions and the simple operation buffer has been made as an approach to bridge vector and raster representation.
The discrepancies of metrics in raster comparing to vector may exceed metric tolerances if they were not considered and included into the processing and interpreting procedure of spatial analysis.

7 ACKNOWLEDGEMENT

This work is a result of the *"Virtual GIS (VGIS) for the Interaction Modelling as the Foundation for the Integrated Image Interpretation within Remote Sensing"* Project, which was member of the German Research Group *"Semantic Modelling and Extraction of Spatial Object from Images and Maps"*, funded by the German Science Foundation (DfG) from 1994 to 1999.

REFERENCES

[Albrecht96] Albrecht, J., "Universal Analytical GIS Operations: a task-oriented systematisation of data structure independent GIS functionality leading towards a geographic modelling language", *ISPA Mitteilungen 23*, University of Vechta, 1996

[Bartelme89] Bartelme, N., "GIS Technologie - Geoinformationssysteme, Landinformationssysteme und ihre Grundlagen", Springer-Verlag, Berlin, Heidelberg, New York u.a. , 1989

[Descartes1637] Descartes, R., "La géometrie", part of "Essais", 1637

[ESRI98] ESRI, "ArcDoc - Arc/Info 7.2 Online Manual", ESRI Redlands, 1998

[Fischer96] Fischer, M.; Scholten, H. J.; Unwin D. (Eds.), "Spatial Analytical Perspectives on GIS", *GIS-DATA IV*, Taylor and Francis, London, 1996

[Fortheringham94] Fortheringham, S.; Rogerson P. (Eds.), "Spatial Analysis and GIS ", Taylor & Francis, London , 1994

[Fritsch98] Fritsch, D.; Glemser, M.; Klein, U. et al., "Zur Integration von Unsicherheit bei Vektor- und Rasterdaten", *GIS - Geo-Informations-Systeme* , Vol.11., No. 4, p. 26-35, Wichmann-Verlag, 1998

[Glemser98] Glemser, M., "Ein Ansatz zur Verarbeitung hybrider Daten und deren Genauigkeit". *In: Albertz, J.; Dech, Stefan (Hrsg.): Photogrammetrie und Fernerkundung - Globale und lolake Perspektiven, Vorträge 18. wiss.-tech. Jahrestagung DGPF, 15. DFD-Nutzerseminar des DLR, 14.-16. Okt 1998, TU München*, p. 309-318, DGPF, Berlin, 1999

[Glemser98a] Glemser, M.; Fritsch, D., "Data Uncertainty in a Hybrid GIS", *Fritsch, D.;Englich, M.; Sester, M. (Eds.): GIS - Between Visions and Applications.*, ISPRS, Commission IV. 32/4, Stuttgart 1998

[Goepfert91] Göpfert, W., "Raumbezogene Informationssysteme - Grundlagen der integrierten Verarbeitung von Punkt-, Vektor- und Rasterdaten, Anwendungen in Kartographie, Fernerkundung und Umweltplanung", Herbert Wichmann Verlag, Karlsruhe, 1991

[Hake94] Hake, G.; Grünreich, D., "Kartographie", *de Gruyter Lehrbuch*, Walter de Gruyter, Berlin, 1994

[ICA73] ICA, "Multilingual Dictionary of Technical Terms in Cartography", *International Cartographic Association, Commission II*, Franz Steiner, Wiesbaden, 1973

[Longley96] Longley, P.; Batty M. (Eds.), "Spatial Analysis: Modelling in a GIS Environment", GeoInformation International, Cambridge, 1996

[Maling89] Maling, D. H., "Measurements from maps - Principles and methods of cartometry", Pergamon Press, Oxford, New York, Beijing a. o., 1989

[Molenaar91] Molenaar, M.; Fritsch, D., "Combined Data Structures for Vector and Raster Representations in Geographic Information Systems", *GIS - Geo-Informations-Systeme*, Vol. 4, No. 3, p. 26-32, Wichmann-Verlag, 1991

[OGC98] Buehler, K.; McKee, L., "The OpenGIS Guide: Introduction to Interoperable Geoprocessing and the OpenGIS Specification" *Open GIS Consortium Technical Committee*, Third Edition (Draft), Wayland 1998. (www.opengis.org)

[OGC99-2] OpenGIS Consortium, "OpenGIS Abstract Specification - Topic 2: Spatial Reference Systems" , Version 4, Project Document 99-102r1, Wayland 1999. (www.opengis.org)

[OGC99-14] OpenGIS Consortium, "OpenGIS Abstract Specification - Topic 14: Semantics and Information Communities" , Version 4, Project Document 99-114, Wayland 1999. (www.opengis.org)

[OGC99-16] OpenGIS Consortium, "OpenGIS Abstract Specification - Topic 16: Image Coordinate Transformation Services" , Version 4, Project Document 99-116r2, Wayland 1999. (www.opengis.org)

[Quattrochi97] Quattrochi, D. A.; Goodchild M. F. (Eds.), "Scale in Remote Sensing and GIS", CRC Press/Lewis Publishers, Boca Raton, New York, London u.a. , 1997

[Tomlin90] Tomlin, C. Dana, "Geographic Information Systems and Cartographic Modeling", Prentice-Hall, London, Sidney, Toronto a.o., 1990

[Voser98] Voser, S. A., "Schritte für ein automatisiertes Koordinatensystemmanagement beim Datenaustausch in GIS und Kartographie", *Nachrichten aus dem Karten- und Vermessungswesen*, BKG, NaKaVerm, Reihe I, S. 111-125, Frankfurt, 1998

[Voser98a] Voser, S. A.; Jung, S., "Towards Hybrid Analysis - Specification of High Level Operators", *1st AGILE Conference, 23.-25. April 1998, at ITC, Enschede (NL)*, 1998

[Voser98b] Voser, S. A., “Konzeptdesign für die räumliche hybride Analyse”. *In: Albertz, J.; Dech, S. (Hrsg.): Photogrammetrie und Fernerkundung - Globale und lolake Perspektiven, Vorträge 18. wiss.-tech. Jahrestagung DGPF, 15. DFD-Nutzerseminar des DLR, 14.-16. Okt 1998, TU München*, DGPF, 301-308, Berlin, 1999

[Winter98] Winter, S., “Bridging Vector and Raster Representations in GIS”. *In: Laurini, R.; Makki, K.; Pissinou, N. (Eds.): Proceedings of the 6th Symposium on Advances in Geographic Information Systems*, ACM Press ,p. 57-62, Washington, D.C, 1998