Raport Badawczy
Research Report

RB/9/2014

Statistical methodology for verification of GHG inventory maps

J. Verstraete, Z. Nahorski

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Prof. dr hab. inż. Zbigniew Nahorski

Warszawa 2014

Appendix 2: Parameters to use a fuzzy rulebase approach to remap gridded spatial data

Jörg Verstraete

Systems Research Institute - Department of Computer Modelling
Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland jorg.verstraete@ibspan.waw.pl
http://ibspan.waw.pl

Abstract. Geographically correlated data are often represented in a gridded format: a grid that covers the region of interest divides the map in cells, and a value is associated with each cell of this grid. This is for instance the case for land use information, air pollution data, etc. The value is considered representative for the entire grid cell, but the grid cell is considered to be the smallest unit. Grids often need to be combined to perform data analysis, and in general do not line up nicely (known as the *map overlay problem*). In traditional methods, data are considered to be spread uniformly or following some other mathematical distribution over the grid cell, which often is too crude an approximation of the real situation. Treating a cell in such a way immediately introduces errors that are carried on and possibly amplified during subsequent analysis. In general, the problem can be reduced to the problem of remapping data that is presented on one grid onto another grid. To perform this remapping, a novel approach using a fuzzy rulebase has been developed. In this article, the parameters this method are discussed and determined. This discussion gives better insight in the data that is needed to determine the rulebase, which is a first step to an optimization of the rulebase parameters.

1 Introduction

1.1 Problem description

Many spatially correlated data, such as air pollution, land use, etc. are presented to the researchers as gridded data. This means that the region of interest of the geographical map is overlaid with a — commonly rectangular — grid, dividing the region of interest in a number of cells [1], [2]. Data are associated with each of those cells, and represent the value of this cell.

This is the first problem: the cells give a discrete representation of the real world; a large area source over the area of a cell would be aggregated to a single value for the cell and would look the same as a cell that contains single point with a high value. Some examples of this are on figure 1. The second problem stems from the fact that data are gathered from different sources, and as such can be provided on incompatible grids: grids can use different cell sizes and can have a different orientation. To combine data, or draw conclusions on the relation between different data, it is necessary to first transform them to the same grid. This is however a non-trivial problem, as the underlying (real world) data is not obvious from the gridded data. Providing an adequate and accurate remapping of one grid to another grid will support all research that is dependent on the combination of gridded data; this includes research regarding climatic change, land use, pollution, and various other socio-economic fields.

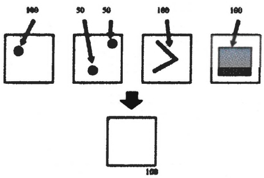


Fig. 1. Example showing the different sources that yield a similar grid cell: single point source, two point sources, line source and area source.

Current solution methods make simple assumptions about the data distributions, or assume distributions that may or may not be correct. In [3], we presented a method that is novel from two points of view: first, it uses data that is known to be related in order to *steer* how the data will be remapped on a new grid, and second it makes use of a fuzzy processing engine to achieve the goal. The method is explained more elaborately in [4], in this article attention will go to the parameters. First, the current solution methods will be briefly explained. The next section (2) briefly summarizes the new methodology and implementation introduced in [4]. The core of this article is in Section 3, where different choices of parameters are considered and their impact investigated. After this, conclusions are drawn.

1.2 Current solution methods

The general outline of the current solution methods is described below. For a more detailed overview, we refer to [5] and [6]. Current solution methods use the input grid A, and attempt to remap this to a target grid B. In this section, A and B are notations for grids with grid cells $A_i, i : 0..m$ respectively $B_j, j : 0..n$. The notation f is used for the function that maps gridcells to their associated value. Typically, the value of a grid cell B_j in the target grid B is determined by the values in the grid A of those cells A_i that overlap cell B_j. Mapping grid A to grid B using the overlapping cells means finding values x_i^j such that:

$$f(B_j) = \sum_i x_i^j f(A_i) = \sum_{i|A_i \cap B_j \neq \emptyset} x_i^j f(A_i) \quad (1)$$

The output should still adhere to the input, meaning that the data distribution over the original grid should still hold. As such, this puts a constraint on the values x_i^j.

$$\forall A_i, \sum_j x_i^j = 1 \tag{2}$$

Areal weighting In areal weighting, the input grid is overlaid with the target grid. For each grid cell, the amount a cell in the input grid overlaps with the cell in the target grid, determines the portion of the value of the cell in the input grid that is assigned to the target grid. As such, x_i^j can be defined as followed:

$$x_i^j = S(A_i \cap B_j)/S(A_i) \tag{3}$$

where S is the notation for the surface area.

Consequently, it is assumed that the data associated with a grid cell is spread uniformly over the area of the cell. While this can be the case for some data, it results in big errors when this is not the case (e.g. point sources). Due to its simplicity though, this method is most commonly used.

Spatial smoothing Spatial smoothing is slightly more advanced than areal weighting: data associated with the cells are first represented as a third dimension. A smooth surface is then fitted over this three dimensional data. Lastly, the smooth surface is sampled using the output grid. While this in general performs better than areal weighting, the method has problems differentiating different point sources that are too close (i.e. within one cell or in neighbouring cells). The added step of fitting a smooth surface makes this method less efficient from a computational point of view.

Spatial regression There are different approaches that use spatial regression, but the principle is the same in all of them. The overlap of the input grid and data in the target grid are examined and patterns are extracted. These patters are then used in combination with an underlying assumption of the distribution of the data in order to optimize the resampling. This method is quite complex to calculate, and some assumptions regarding the data are made. In general, it is not known whether these assumptions hold and if they can be made, nor if the relations have real world relevance.

2 Rule base approach

2.1 Reasoning with added knowledge

The key difference between our new methodology and existing methodologies, is that additional data is used in order to *steer* the resampling of the input grid. For many data, there can be other data that is known to be related and that can be used. One example could be the resampling of a grid with concentrations for some pollutant that is known to relate to traffic: information of the road network can then be used to further optimize the adapting the input grid to another grid format. This auxiliary data should be used as a guideline, but cannot be followed too strictly: not all values in the input grid could be attributed to known features in the auxiliary grid. As a result, there can seem to be contradictions and the remapping may not necessarily be performed uniquely. Consider the example on figure 2a. The input grid A has quite large cells; the target grid C has smaller cells. In general, the grids do not have to overlap as nicely as they do in the example. This however facilitates the methodology and if necessary, the intersection between A and C is considered as the new intermediary target grid. The auxiliary grid B has very small cells (as we assume that it stems e.g. from vectorial data that represents a road network). The data from grid A should be resampled such that most data appears in the vicinity of the overlaying road network. Intuitively, some simple relations regarding the value of a cell x in the output can be observed:

- it is very likely proportional with the input cell in A that overlaps with cell x
- it is very likely is proportional with the auxiliary cell in B that overlaps with cell x

Many other relations can be observed; in [4], a fuzzy inference system was developed to mimic an intelligent reasoning of this problem. In section 2.2, the method will be briefly described.

2.2 Fuzzy Inference System

Description A fuzzy inference system is a technology often used in fuzzy control that allows to derive values based on a set of of inputs, reflecting an artificial intelligent behaviour ([7], [8]). This is done by a set of rules of the form

$$\text{"if } \underbrace{x \text{ is } A}_{\text{premise}}, \text{ then } \underbrace{y \text{ is } B}_{\text{conclusion}} \text{"}$$

Here "x is A" is the premise and "y is B" is the conclusion; x is the input value and and y the output value. These values can be both crisp or fuzzy, but in general will be crisp. The terms A and B are linguistic terms represented by a fuzzy sets, commonly representing concepts such as *high* or *low*. The *is* in the premise is a fuzzy comparator that determines how well x satisfies A, commonly resulting the membership grade A has for the value x. The *is* in the conclusion is an assignment, and assigns y the fuzzy set B. As multiple rules can match (x can be *high* and *very high* at the same time but to a different extent), y should be assigned multiple values by different rules: all these values are aggregated using a fuzzy aggregator to result in one single fuzzy value. The fuzzy output value is finally defuzzified to obtain a crisp result. Key issues to making the rulebase work are the definition of the rules, and the definition of the fuzzy sets used in the different rules.

Parameters and rulebase In [4], the output grid was determined by aggregating a grid with smaller cells: the intersection of the input grid and the output grid. This was done to make it easier to satisfy an important constraint: the value of an input cell contributes different output cells, but the sum of all its contributions should add up to the value of the input cell. By defining the output segments, this constraint is easier to maintain in the end result. In the very last step, the different segments that overlap with an output cell are combined. Using on this segment grid, four parameters that were considered to influence the output value were determined. The first two parameters were obtained by considering only the values of overlapping cells of the input and auxiliary grids (formulas 4, 5). The third parameter was obtained by considering the values of the input cells that overlap the same output cell as the segment, but not the one containing the segment (formula 6). The fourth parameter was obtained by considering the auxiliary cells that overlap the same input cell as the segment, but not the auxiliary cell that contains it (formula 7). Following the simple reasoning that a higher overlapping input or auxiliary value should yield a higher result, the first two are considered to be proportional to the output value. Similarly, the latter two are considered to be inverse proportional. For a segment x, these parameters are:

$$par_1^x \propto f(A_i)|A_i \in A \wedge x \cap A_i \neq \emptyset \tag{4}$$

$$par_2^x \propto f(B_i)|B_i \in B \wedge x \cap B_i \neq \emptyset \tag{5}$$

$$par_3^x \propto \begin{array}{l} 1/\sum f(A_i)|A_i \in A \wedge \exists C_j \in C : A_i \cap C_j \neq \emptyset \\ \wedge x \cap C_j \neq \emptyset \wedge x \cap A_i = \emptyset \end{array} \tag{6}$$

$$par_4^x \propto \begin{array}{l} 1/\sum f(B_i)|B_i \in B \wedge \exists A_j \in A : B_i \cap A_j \neq \emptyset \\ \wedge x \cap A_j \neq \emptyset \wedge x \cap B_i = \emptyset \end{array} \tag{7}$$

Next, it was necessary to determine when these parameters should be considered high, and when they should be considered low. For this, a very simple approach was used: the minimum and maximum of respectively input and auxiliary related parameters were considered as the limits

The rule was determined manually in an intuitive way. The four parameters described above were considered, each parameter was given three possible fuzzy sets (low, medium and high, evenly distributed between minimum and maximum), and all possible combinations were generated. Assuming that each value is equally important, the impact of an increase/decrease of each value is the same. The lowest possible value for the output is when both proportionally related variables are low and the two that carry an inverse proportional relation are high; and vice versa. This all combined yielded nine possible fuzzy sets for the output. The output value was then scaled to represent the fraction of the input segment value that will be assigned to the segment.

Results The system as described shows promising results but also suffers some shortcomings. The first shortcoming can be seen from the experiments: while the desired effect is there, it still appears too weak to be considered very valuable.

Furthermore, there are some strange effects that can be seen. Both these things can either be explained by either the simple rulebase used, or the parameters at hand. The second shortcoming is the fact that there is no control over the strength of the effect. In later stages, it should be possible to specify multiple auxiliary grids, and indicate which one has a bigger impact. At the moment, there is no way of indicating the strength of the effect for a single grid.

3 Discussion of parameters

3.1 Current

When considering the parameters as defined in 2.2, it is obvious that each parameter either equals the maximum, or the minimum, so each parameter effectively has only 2 possible values. Furthermore, from their definition, two parameters will always be the minimum, and two always the maximum. This means that from a possible 4^2 parameter combinations, there only are 2^2 possible combinations, which is due to the choice of minimum and maximum values. Interestingly, the current parameters completely ignore any mathematical connection between values in the auxiliary grid and values in the input grid: it just considers high and low predicates, and these are local concepts. For some segment x, high can be a value of e.g. 100, where for another it can be a value of 1000. This currently unused information can also provide more information in steering the remapping. At the moment, no distance measures are included: data from all neighbours are aggregated. A spatial aspect to indicate where higher or lower data occurs should improve the result. Lastly, the result of the remapping is not necessarily unique. As fuzzy sets are employed, it may be possible to supply multiple outputs with possibilities rather than an arbitrary chosen output. These different parameters will be discussed using examples.

3.2 Data study

Overlapping data Consider the data on figures 2a and 2b. In these simplified examples, the input grid consists of a single cell, the auxiliary grid consists of 10 cells that together completely overlap the input cell. The target grid consists of 4 output cells, that together overlap the same area as the input and auxiliary grids. The grids are represented below each other but overlap the same area. It is assumed that the data around these cells has completely no impact. The desired result is represented as a graph, which serves purely as an illustration. Two possible outcomes are shown, but the strength of the pulling effect (i.e. how wide the base below each peak is) is an additional degree of freedom that will be discussed later. The graph gives rise to the values in the output cells C_i, which at this point are estimates purely for illustration purposes. These are the results that we would like to obtain.

As mentioned, the input grid is first remapped to a grid which is obtained from the intersection of input and output grids. As such, the situation where a single output cell overlaps with multiple input cells cannot occur.

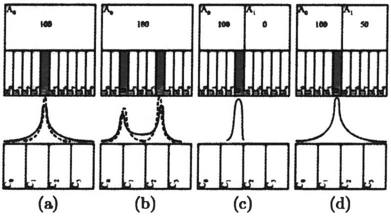


Fig. 2. Example cases for the case study. Cases (a) and (b) are used to show how auxiliary cells should influence the output, cases (c) and (d) are used to show how the input cells that neighbour the overlapping input cell influence the output.

Input The overlapping input value for all involved cells C_i is 100, as it is the same for all, it cannot contribute to the distribution of the data.

Auxiliary The overlapping auxiliary value for all involved cells C_i varies. In 2a, C_1 is the only cell that overlaps with B_4, and as a result should get the highest value. The cells neighbouring C_1 should get lower values, but as C_2 is closer to B_4 than C_0,it is expected to have a higher value. The values should decrease as cells are located further from the auxiliary cell. This can be achieved by considering the distance to the closest cell in B that has a high value, this distance is expressed as the number of cells of grid B between B_4 and the cell C_i.

cell	C_0	C_1	C_2	C_3
input	100	100	100	100
auxiliary	0	100	0	0
distance to B_4	1.5	contains	0	2.5
expected 1	10	50	35	20
expected 2	0	80	20	0
distance to B_4	medium	contains	very close	far
expected 1	lowest	highest	medium	low
expected 2	lowest	highest	low	lowest

Table 1. Example and expected values for the examples in figure 2a

From this, a proportional relation between auxiliary grid and result is observed, while greater distances decrease the influence of the proportional relation. Instead of using numerical values, linguistic terms can be used, which gives the bottom section of table 1. The issues with defining the fuzzy sets to represent the linguistic terms will be covered in section 3.2. For now, it can be observed that altering the definition of a linguistic term (*closeby*, *far*, ...) would allow some modification of the strength of the pulling effect.

In 2b, two overlapping auxiliary cells have an associated value. As a result, distances to both should be considered. The results are summarized in table 2.

cell	C_0	C_1	C_2	C_3
input	100	100	100	100
auxiliary	100	100	100	0
percentage	50%	50%	100%	-
distance to B_2	intersect + closest	intersect + closest	rather close	very far
distance to B_6	far	close	contains	very close
expected 1	high	low	highest	lowest
expected 2	high	low	highest	lowest

Table 2. Example and expected values for the examples in figure 2b

Neighbouring data

Input Consider the data on figure 2c and 2d: 2 input cells are considered, each has 10 overlapping auxiliary cells and the data should be remapped on 8 output cells. Again, it is assumed that the data around these cells has completely no impact, and the notations are the same as above. The only auxiliary cell that has a value is B_4. In figure 2c, the overlapping input value for cells C_0 and C_1 is 100, for C_2 and C_3 it is 0; in figure 2d, these numbers are respectively 100 and 50.

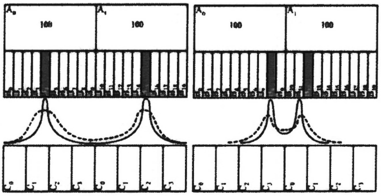
The situation in figure 2c implies that there is no influence of the neighbouring auxiliary cell B_4 to input cell A_1, which in turn implies that the *pulling effect* should be quite strong (if a symmetrical effect is present). While the situation in 2d at first seems to indicate that the auxiliary grid B is not fully related to grid A, this is not necessarily the case. The data from grid B can have a wider impact, which reaches outside of the overlapping gridcell of A, as illustrated by the graph.

cell	C_0	C_1	C_2	C_3
auxiliary	0	100	0	0
distance A_1	far	close	-	-
distance A_0	-	-	close	far
fig 1c	0	high	0	0
fig 1d	low	high	medium	low

Table 3. Example and expected values for the examples in figure 2c and 2d

This situation illustrates that the value of an input cell should influence the value of output cells that are close to it. The effect impacts the strength of the pulling effect, rather than a direct correlation to a specific value. The value of the output cell therefore has to depends on the distance to the neighbouring input cell.

Auxiliary To illustrate the influence of auxiliary cells that are close to the examined output cell, consider the examples in figure 3. As before, the assumptions are the same and the grids are as defined on figure 3.



(a) (b)

Fig. 3. Example cases for the case study, (a) and (b) are used to show the influence of auxiliary cells that overlap the neighbouring input cell.

In figure 3a, the grid cells with a higher value in the auxiliary grids are located too far apart to have an effect (this of course depends on the desired strength of the pulling effect, as shown with the solid line and dashed line). In figure 3b, the auxiliary cells are much closer together. The cells output cells that are close to both should get a higher value than before.

This example already illustrates the next issue. In figure 3, cell C_3 is the only cell that overlaps A_0 and an auxiliary cell that has a value (B_8). However, in input cell A_1, the cell B_{12} overlaps with two output cells(C_4, C_5). As such, it is intuitive that the highest value of those is lower than that of C_3. In table 4, this means that both interpretations of ***high*** should be different: high in the context of A_0 is not high in the context of A_1. This can be achieved through an appropriate definition of the fuzzy sets.

Defining fuzzy sets In Section 3.2, the transition from numbers to linguistic terms was made, without really considering the modelling of this. One thing to consider is how many fuzzy sets will be defined for each variable. This is important from a computational point of view: a larger number tends to increase the number of rules in the rulebase, but also should provide a better distribution of the result. This is however a relatively easy thing to adjust, and we chose to determine this such that the rulebase is still kept relatively small for performance

cell	C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7
aux	0	100	0	0	0	0	100	0
dist. to B_{15}	very far	very far	far	far				
dist. to B_4						far	very far	very far
fig 2a	low	high	med	low	low	med	high	low
aux	0	0	100	100	100	0	0	0
dist. to B_{12}	very far	far	far	very close				
dist. to B_8					close	far	far	very far
fig 2b	0	0	low	high	med	high	low	0

Table 4. Example and expected values for the examples in figure 3

reasons. Using more powerful hardware would allow us to increase the number of sets for the different parameters, but to illustrate the workability of the methods, 3-5 sets for each variable such suffice. A more important aspect is the range: what is the lower limit, and what is the upper limit. There are several options on how to define the limits, these will be illustrated using figure 4.

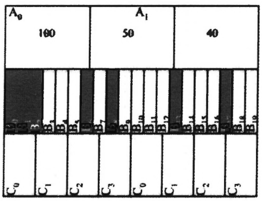


Fig. 4. Example to illustrate possible definitions for the limits of the fuzzy sets.

A first option is to determine high and low values considering the overlap with the input cells. For the input cells themselves, these values are straight forward. For the values of the auxiliary grid, the overlap is checked. If an auxiliary cell overlaps partly with an input cell, its value is not added to generate the lower limit, but is added to result in the upper limit. Table 5 shows the result for the lowest and highest input values of the example, the low and high limits for the auxiliary grid for each input cell, as well as the smallest and largest auxiliary overlap that occurs globally.

The second option is to look from the point of view of overlap with the output cell. The methodology is the same as before, with an overlapping cell

cell	A_0	A_1	A_2
global input low	30 (A_3)		
global input high	100 (A_0)		
local aux low	3	1	1
local aux high	4	3	2
global aux low	1 (from A_2)		
global aux high	4 (from A_0)		

Table 5. Different range definitions over the input cells

not contributing to the lower limit, but only to the upper limit. The results are summarized in table 6

cell	C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7
global aux low	0 (from C_4)							
global aux high	3 (from C_0)							
local input low	0		0	0	0	0	0	0
local input high	100	100	150	50	50	80	30	30
local aux low	2	0	1	1	0	1	0	0
local aux high	3	1	1	1	0	1	1	1

Table 6. Different range definitions over the output cells

The first approach has the benefit of staying close to the concept of *the input grid that will be remapped*, the downside to this is that one extreme value will immediately result in the other values being more averaged. This is clear when considering the cells that overlap A_1. The value of the auxiliary grid that overlaps output cell C_3 would be classified as low: it is value 1 in a range $[0, 3]$, even though it is the highest possible value for this input cell. This in turn will mean that the effect of the rulebase will not be very strong. On the other hand, the calculation provides for a nice scale on which other data that relates to the input cell can be mapped, this includes neighbouring auxiliary data as mentioned before.

The second approach has the benefit of being more closely related to the output cell considered, but has the downside of possibly narrowing down the intervals too much. For cells C_2 through C_5, the local auxiliary low limits and high limits are equal. As such, they don't provide that much useful information. It also requires different ranges to be applied for neighbouring data, which is not straight forward to combine.

Further study is still required on which range definitions would prove to be most useful. A different use can also be considered: the local auxiliary range from the table 6 is a range-estimate for the output cell, which can for instance serve as a first approximation or as verification.

Mathematical connection and constraint One aspect that was only briefly covered in the above discussion was a true mathematical connection between input grid and auxiliary grid. In most of the examples, the auxiliary grid was considered in quite a binary way (black or white, road or no road). In reality, this grid will also contain quantitative information (size of the road, amount of roads present), which should also be taken into account. The table entries so far were either 0 or 100. In figure 2d, an example of a different value is shown, and the impact is different. In figure 4, a different scale for the auxiliary grid is used. This illustrates that the absolute value of the auxiliary is not important, but only the relative value. However, if the auxiliary grid is closely linked to the input grid, the mathematical correlation between both grids can be exploited further. To determine the correlation, the whole grid (or several regions of interest) can be used; using the whole area in figure 4 yields that a value of 190 in the input grid relates to a value of 7 in the auxiliary grid. Or that 1 cell in the auxiliary grid reflects an input value of about 180/7=25.7. Such information can also aid the remapping of the input grid to the output grid, but only if the auxiliary grid has a strong correlation.There is also a possibility to consider the correlation at a local level rather than globally.

The mathematical correlation can be used to decrease the number of parameters: rather than considering the values of input and output, it is also possible to consider the ratio of both. This would allow us to eliminate one parameter, thus simplifying the rulebase (which in turn might allow us to define a larger number of fuzzy sets that define the parameter).

Another mathematical restriction to be considered is: how *far* is input data allowed to migrate? One criterion we enforce is the output grid should result in the input grid when resampled using areal weighting. While this restriction seems very natural as the data is still exactly the same as in the original input grid, it does impact the possible approaches. In figure 4, this means that

$$f(A_0) = f'(C_0) + f'(C_1) + \frac{2}{3}f'(C_2)$$
$$f(A_1) = \frac{1}{3}f'(C_2) + f'(C_3) + f'(C_4) + \frac{1}{3}f'(C_5)$$
$$f(A_2) = \frac{1}{3}f'(C_5) + f'(C_6) + f'(C_7)$$

where f is the function the returns the value value with cells of grid A and f' the function that returns the associated values for grid C. This restriction is important for the final interpretation of the output of the rulebase, and if compliance with it is not guaranteed, values should be rescaled. At the moment, this restriction is imposed, but research is ongoing whether the constraint can also be valid when a different resampling method would be used. Still, even without this restriction, there still is the constraint that the total value of grid C should equal to total value of grid A, which — as it requires data of the whole grid — may even be a more difficult constraint to impose than the local restriction.

4 Conclusion

In this article, the parameters used in [4] were examined, additional parameters were determined and their influence on the output result was observed. Different parameters have been examined, and can be classified in parameters that exhibit a proportional behaviour, and parameters that exhibit an inverse proportional behaviour. The value of an output cell is proportional to the value of the overlapping input cells en overlapping auxiliary cells. It can be proportional to the value of neighbouring input and auxiliary cells, but in this situation the distance matter. If the output cell is close to the neighbouring data, the influence appears to exhibit some proportional behaviour. The value is inverse proportional to the value of neighbouring input cells and neighbouring output cells, but only of they are located far enough.

Local overlapping data can provide limits for the defining fuzzy sets, and for the output values in general (section 3.2). Additional mathematical correlation can also provide more information, but should only be used when there as a strong correlation between auxiliary data and input data.

Future work will go towards developing different rulebases with the newly defined parameters, in order to compare the performance and to find optimal parameters (or cases in which some parameters are more desirable than others). Under consideration also is an automatic determination in real time of the parameters that would be best suited for the current problem being solved, but this requires a deeper insight in both the parameters and methodologies to derive them. A last part of the future research is outputting the possibility distribution of candidate values rather than a single crisp output value. The big issue is that the distributions of neighbouring cells are correlated, however, outputting the possibility distribution along with a single defuzzified value provides more information on how certain the crisp value is. The possibility distribution also allows the ambiguity regarding the value to be resolved using algorithms that are more advanced than simple defuzzification or possibly using additional or expert knowledge.

References

1. Shekhar S. and Chawla S. *Spatial databases: a tour.* Pearson Educations, 2003.
2. P. Rigaux, M. Scholl, and A. Voisard. *Spatial databases with applications to GIS.* Morgan Kaufman Publishers, 2002.
3. Jörg Verstraete. Using a fuzzy inference system for the map overlay problem. In *3rd International Workshop on Uncertainty in Greenhouse Gas Inventories*, pages 289 – 298, 2010.
4. Jörg Verstraete. Solving the map overlay problem with a fuzzy approach - using a fuzzy inference system to optimize the mapping of values in mismatched grids. *Climatic Change*, submitted January 2013.
5. R. Flowerdew and M. Green. *Spatial analysis and GIS; eds. Foterhingham S. and Rogerson P.*, chapter Areal interpolation and types of data, pages 141–152. Taylor & Francis, 1994.

6. Carol A. Gotway and Linda J. Young. Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458):632–648, 2002.
7. Jerry M. Mendel. *Uncertain rule-based fuzzy logic systems, Introduction and new directions.* Prentice Hall, 2001.
8. George J. Klir and Bo Yuan. *Fuzzy sets and fuzzy logic: theory and applications.* Prentice Hall, New Jersey, 1995.