

Solving the general map overlay problem using a fuzzy inference system designed for spatial disaggregation

Jörg Verstraete^{1,2} 

Received: 24 January 2018 / Accepted: 21 June 2018 / Published online: 10 July 2018
© The Author(s) 2018

Abstract Identification of mitigation and adaptation strategies in any situation has to be a well informed decision. This decision not only has to be based on quantifiable data (e.g. amounts or concentrations of emissions), but it also needs to consider spatial aspects such as the source of emissions, the impact of policies on the population or the identification of responsible parties. As such, it is important that the decisions are based on accurate reports made by experts in the field. Many data are available to the experts, and combining this data tends to increase the inherently contained uncertainty. Novel operations that exhibit a lower increase of uncertainty can yield outcomes that contain less uncertainty, which subsequently improves the accuracy of the resulting reports. At the same time, the decision-makers are confronted with different reports, and the comparison of the contained spatial aspects requires combining the data, which exhibits similar issues relating to uncertainty. With data usually represented in gridded structures, comparing them often requires a process called *regridding*: this is the process of mapping one grid onto a second grid, a process which increases spatial uncertainty. In this contribution, a novel regridding algorithm is presented. While mainly intended as a preprocessing tool for the experts, it is also applicable for supporting the comparison of gridded datasets as used by decision-makers. In the context of this article, a grid can be irregular, allowing the presented algorithm to also be used for remapping a grid onto, e.g. administrative borders or *vice versa*. The presented algorithm for regridding is a modification that is generally applicable on spatial disaggregation algorithms. It was developed in parallel with a novel method that uses artificial intelligence (in the form of fuzzy rule-based systems) to involve proxy data to obtain better results, and it will be demonstrated using this approach for spatial disaggregation. The methodology

✉ Jörg Verstraete
jorg.verstraete@ibspan.waw.pl; jorg.verstraete@usc.es

¹ Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warszawa, Poland

² Centro Singular de Investigación en Tecnoloxías da Información, University of Santiago de Compostela, Rúa de Jenaro de la Fuente Domínguez, 15782 Santiago de Compostela, Spain

to perform regridding using an algorithm designed for spatial disaggregation is detailed in this article and the performance of the combination with the artificial intelligent system for disaggregation is illustrated by means of an example.

Keywords Fuzzy inference system · Fuzzy spatial reasoning · Map overlay problem · Gridded data · Ill-aligned grids

1 Introduction

1.1 Practical issues when combining data

In order to compare different mitigation and adaptation strategies to cope with climatic change, a proper analysis of both the available data and relevant predictions is necessary. These data (emissions, measurements, land use, etc.) are prone to uncertainty, which tends to be increased by analysis or operations. When different scenarios need to be considered and compared, too much uncertainty makes it difficult to distinguish the effects of different adaptation strategies. As such, it is necessary to try to limit the amount of uncertainty, and to ascertain its impact on the models and conclusions. Uncertainty present in the input data always propagates in the models; the aim here is to develop regridding methods that do not increase the uncertainty related to spatial aspects in the input data, resulting in less uncertainty carried over in the models and thus in the outcomes. This is particularly important when different models are used (e.g. at regional, national or global scale) and the results are not consistent with each other or the outcomes of these models exhibit too much uncertainty making comparison impossible. Examples of such different approaches are, e.g. in Jonas et al. (2014), Rafaj et al. (2014) and Boychuk and Bun (2014). Most input data is pre-processed in order to make it compatible; lowering the uncertainty of the input data is possible by performing this pre-processing differently. In this contribution, a novel method is put forward to perform a regridding of the data, which is necessary to combine different gridded datasets in what is called a map overlay. This step can be used at different stages: by experts to improve the quality of the data that are analyzed, or at decision-maker level to compare data between different reports. Various types of uncertainty exists, as well as different approaches to consider the uncertainty (Hryniewicz et al. 2014). The proposed method consists of an extension to approaches for spatial disaggregation. It was developed in parallel with a novel approach which uses related data (proxy data) and a fuzzy rule-based system to yield a grid that contains less spatial uncertainty than what would be obtained with traditional methods. By combining these methods, a more accurate regridding is possible. The proposed method is very specific for gridded data that contains a spatial aspect, but its applicability does not depend on the type of data modelled (air pollution, demographic data, climate data, ...), as long as additional knowledge can be provided. The next sections will clarify what can constitute this additional knowledge.

Climate related studies require data that exhibit a spatial aspect; researchers use data concerning air pollution, demographic data, climate data, etc., all originating from different sources. In order to perform studies, there is a need to combine data or compare data. The authors in Danylo et al. (2015) study the uncertainty of greenhouse gas emissions in the residential sector, considering fuel combustion and heat production. This type of study not only requires data on greenhouse gases, but also uses population data to define residential areas and incorporates data on heat usage. This data is acquired from different sources that are combined. A similar problem of combining data occurs when assessing the forest carbon

budget, as done by the authors in Lesiv et al. (2018). In Hogue et al. (2018), the authors in particular develop methods to deal with the inherent uncertainty of gridded datasets, referring to studies that estimate discrepancies of the combination of gridded sets to be 160%. The authors then take into account the scale at which the data are represented and consider additional knowledge regarding the underlying distribution as part of their datasets in order to lower the uncertainty of the analysis. The issue of combining datasets for analysis does not just occur when different datasets are used: the authors in La Notte et al. (2018) analyze air emission levels obtained through regional data sources for policy decisions made at a higher level.

The authors in Hogue et al. (2016) work with gridded datasets that contain emission data and combine these with population data. In total, they deal with six types of uncertainty, two of which are spatial. The underlying emission sources are point sources (e.g. factories) which surprisingly carry high levels of locational uncertainty (order of 150%). This uncertainty decreases when larger grid cells are used, but to have higher spatial accuracy in the report, smaller cells are more desirable. In Hutchins et al. (2017), the authors compare five different sources relating to the same data and study the effect of how different approaches affect the spatial distribution and magnitude of emission estimates. They also observe a strong relation between scale and uncertainty. Both these articles make a strong case for algorithms that aim to keep the uncertainty down when data are combined. In addition, the authors in Woodard et al. (2014) argue that emissions are seldom measured directly but rather estimated from other data. The presented approach can aid in the spatial distribution of these estimates through the use of proxy data.

The methodology presented in this article was developed for experts working with gridded datasets, to improve the quality of the outcome of necessary regridding operations. However, different representations for the same data can also occur in reports issued to policy makers; either depending on the time of the report or the source of the report. Advances in technology can for example allow for generating data at a higher resolution but, while this in itself is more informative, it may make the data more difficult to compare with data from previous reports. In this case, the developed algorithm can also be applied to perform a regridding of the results to make them compatible, facilitating the comparison of gridded data issued in different reports. As grids in the context of the developed algorithm (and in this article) can be irregular, it can also be used to remap the presented results on a dataset that is more suitable for policymaking, e.g. matching administrative borders rather than grids defined using arbitrarily chosen gridlines.

1.2 Regridding versus spatial disaggregation

Grids are a commonly used spatial data structure to represent values that vary spatially. When grids need to be combined, compared or analyzed, incompatible grids give rise to the map overlay problem, which can result in an increase of spatial uncertainty. The problem is commonly solved by making assumptions on the underlying spatial distribution and an operation called *regridding*; the quality of the outcome highly depends on the appropriateness of these assumptions. More recently, methods for spatial disaggregation have been developed that, rather than rely on assumptions, make use of other related datasets (proxy data) to provide information on the underlying spatial distribution. Spatial disaggregation is a special case of regridding where a grid needs to be mapped onto a finer grid that partitions each grid cell in smaller cells, a process which increases the resolution of the raster.

We developed an approach that uses a fuzzy rule-based system, a known concept in artificial intelligence, which simulates an intelligent reasoning to determine the underlying

distribution. This approach was developed for spatial disaggregation; in this contribution we will use this approach as a use-case to illustrate the method presented in this article to apply spatial disaggregation algorithms to solve general regridding. While this is in essence a general methodology usable for other spatial disaggregation algorithms, the combination with the fuzzy inference system is particularly interesting due to the specific way the latter processes the proxy data.

In order to explain the typical problems that occur when spatial datasets are combined, it is necessary to consider how spatial datasets are represented. One commonly used representation method for spatially correlated data is a grid: the region of interest is partitioned into a number of cells, and a numerical value is associated with these cells. The value is a representative value for the cell, usually obtained through an aggregation or estimation. Most common are regular grids with rectangular grid cells, but irregular shaped cells are possible; this is for example the case when administrative borders are used to define the grid cells. The gridded representation is an approximation of the real-world situation: there is a more continuous underlying spatial distribution, but as the grid cells are considered the smallest unit, the spatial distribution within each cell is not represented and thus not known to the users of the data. Different data can be defined on *incompatible* grids (Gotway and Young 2002); these are grids that are defined differently so that there is no one-one mapping between the cells of one grid and the cells of the other grid. The lack of such a one-to-one mapping makes comparing and combining grids difficult; to overcome this, usually one grid is remapped onto the other to achieve a one-to-one mapping. This process is called *regridding* and is a necessary pre-processing step to allow data to be combined: on compatible grids, map algebra (Tomlin 1994) can be used. The accuracy of this pre-processing step impacts quality of the datasets the researchers work with and thus impacts any subsequent calculations. The regridding is usually performed by either assuming an underlying distribution or by trying to estimate an underlying distribution. This contribution expands on the work in Verstraete (2016) which, using a fuzzy inference system (an approach from the field of artificial intelligence) in combination with proxy data (other available data that provides information on the spatial distribution), is capable of providing a solution to the spatial disaggregation problem.

In this article, the regridding will be performed by first translating the problem to a problem of spatial disaggregation. This is done by creating a new raster which allows for easy determination of the values of the remapped gridcells. While this intermediary raster allows for most spatial disaggregation methods to be used for regridding purposes, the combination with the developed method that uses the fuzzy rule-based system has additional benefits in this approach. The following section describes the map overlay problem for grids in more detail; Section 3 explains the methodology for the spatial disaggregation problem. The adaptation of the methodology for the general remapping problem is elaborated on in Section 4; a conclusion summarizes the results.

2 Problem description

2.1 Gridded data

In geographic information systems, two distinct models are commonly used for different purposes (Burrough and McDonnell 1998; Rigaux et al. 2002; Shekhar and Chawla 2003). Feature-based models use basic geometries with associated attributes to represent real-world entities. Such models are useful for digital maps, but not much so for representing values

that vary with location. Examples of such data are, e.g. population densities, altitudes or emission values. Such data can be represented using field-based models using on triangular networks or grids to model data spread out over a region. In this article, attention goes to gridded data: the region of interest is overlaid with a raster that partitions it into a number of cells. The raster can be regular (all cells have the same size and shape) or irregular (cells can have different shapes and sizes, e.g. when using administrative borders). Each cell of the raster is assigned a value that is considered representative for the cell: in case of, e.g. population density, the number can indicate either the average number per square kilometres for this grid cell, or the absolute number of people living inside the area marked by the cell.

The underlying distribution of the data represented by a grid is not known. To continue with the example of the population density, this means that the distribution of the people within a single grid cell is not known: they can be spread out uniformly over the cell, or be concentrated in specific parts of the grid cell. Two given grids are said to be incompatible if there is no one-one mapping of the cells of either grid. This can occur because the cell sizes are different, the rasters are shifted or rotated, or a combination of these; the latter case is shown in Fig. 1a. A special case of incompatible grids occurs when one grid is such that it partitions every cell of the other grid (Fig. 1b). In this case, the remapping problem becomes a problem of spatial disaggregation, which requires an accurate estimate of the underlying distribution locally in each cell, in this example, of grid A.

2.2 Remapping gridded data

The map overlay problem is a general name for all problems that relate to combining features in geographic data. Here, the specific problem of overlaying gridded data is considered (Gotway and Young 2002). This problem needs to be solved before applying Tomlin's map algebra (Tomlin 1994). Tomlin's map algebra provides for operations between grids, which are classified depending on which cells are considered (local, zonal, etc.), but requires the grids to have a one-one mapping between their cells. The common procedure is to achieve such a mapping by remapping one grid onto the other: the first grid will be called the *input* grid, the other grid will be referred to as the *target* grid. This is not a problem of coordinate transformation: both grids in this situation are geo-referenced and positioned correctly; they just are defined using different cells. As an example, consider the grids in Fig. 1a.

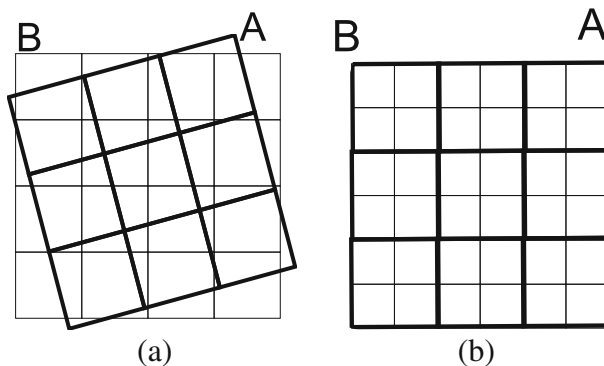


Fig. 1 Illustration of incompatible grids in the general case and the case of regriding: **a** shows the general case of incompatible grids rids A (thick lines) and B (thin lines), **b** shows the case of spatial disaggregation where each cell of A is partitioned by the cells of B

The goal of the remapping is to determine the values of the cells in the target grid, so that it best reflects the data represented in the input grid. Several approaches exist to perform this remapping, the main difference is in the assumptions made regarding the underlying distribution. In general, for every cell b_i of the output grid B , its associated value $f(b_i)$ is calculated by a weighted sum on the values $f(a_j)$ of cells a_j of the input grid A ; thus $f(b_i) = \sum_j x_j^i f(a_j)$. If it is assumed that only the cells of grid A that intersect with the target cell b_i will contribute to $f(b_i)$, the formula can be modified to only consider the intersecting cells: $f(b_i) = \sum_{j|a_j \cap b_i \neq \emptyset} x_j^i f(a_j)$. This is a reasonable assumption for most data. Remapping a grid A to a different grid B is therefor equivalent to finding weights x_j^i , so that the value $f(b_i)$ of a cell in the output grid is the weighted sum of the cells in the input grid.

The weights x_j^i are subject to the following constraints:

$$\forall i, \forall j, x_j^i \geq 0, \quad (1)$$

$$\forall i, \sum_j x_j^i = 1. \quad (2)$$

Constraint (1) states that the weights are positive: a cell of the input cannot negatively contribute to a remapped cell. For most data, such as population density or pollution data, this is logical. Note that this only applies for the cells of the input grid: proxy data as will be defined later can have a negative impact (in an emission example: presence of water implies the absence of cars). Constraint (2) guarantees that the entire value of an input cell is remapped, but also not more than the entire value.

The spatial disaggregation problem does not suffer from the issue of partial overlapping cells, as seen in Fig. 1a where cells of grid B overlap partly with cells of grid A . This simplifies the problem somewhat: the value associated with a cell from A should be distributed over its partitions defined by B . This is a more constrained problem, but even in this simplified form, it is not a trivial task. A small example on how incorrect assumptions on the underlying distribution can offset a regridding procedure is shown in Fig. 2. In the top left of this figure is a feature that is approximated using three different grids: grid G_1 (top right), grid G_2 (bottom left) and grid G_3 (bottom right). Grid G_1 has a single grid cell, whereas grids G_2 and G_3 each have two grid cells covering the same area. Grid G_2 is the ideal way of mapping the data onto these two grid cells. When only considering grid G_1 , any assumption of the underlying distribution is equally valid. When the assumption is made that the underlying distribution is uniform in each grid cell, the remapping of grid G_1 to a grid with two grid cells is done using areal weighting. This results in the grid G_3 which has both gridcells receiving the same value as they are of equal size. While the resolution is seemingly higher (more grid cells in the same area), this is actually a worse representation as it wrongly indicates that there are features in both the cell on the left and on the right. The single grid cell G_1 by comparison shows that there are features in this area but does not specify where. As the figure shows, the grid G_3 deviates from the ideal grid G_2 : the left cell has a value that is just half of what it should be, whereas the right cell has half of the value of the feature, even though it should be zero. This simple example illustrates the general problem that occurs, not only in disaggregation but also in regridding. Any algorithm that aims at disaggregating or regridding needs to make correct assumptions on the underlying distribution.

Current approaches tend to use simple assumptions based on which the weights are (implicitly) calculated. In areal weighting methods (Goodchild and Lam 1980), the underlying distribution in each grid cell is assumed to be uniform, allowing the weights to be calculated from the surface area of the overlap of the grid cells. In spatial smoothing methods (Tobler 1979), the distribution of the attribute over the region of interest is assumed to

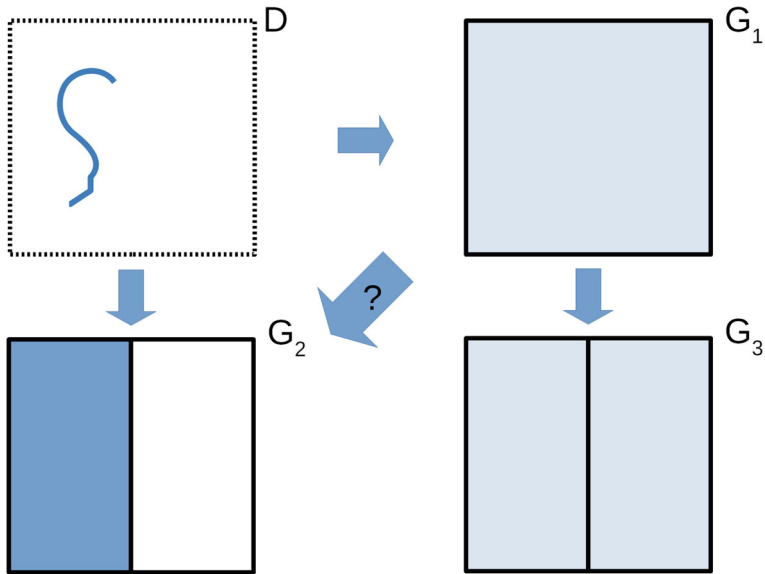


Fig. 2 Illustration of the influence of assumptions on spatial disaggregation. The real-world data is shown on the top left. This data approximated as a single grid cell; this would appear as grid G_1 (top right). Approximated as two grid cells, it could appear as grid G_2 (bottom left; the shading is darker for higher associated values). Disaggregation of the cell in G_1 should ideally result in G_2 . Without any other information, areal weighting is the most likely algorithm yielding solution is G_3 (bottom right). Proxy data can provide insights into the underlying distribution which is not known in grid G_1 , to achieve results closer to G_2 than to G_3

be smooth; the weights are calculated implicitly: the modelled attribute is considered as a third dimension, resampling of the smooth surface fitted over the grid considered results in the new values for the cells. Spatial regression methods usually involve the extraction of patterns combined with an assumed statistical distribution of the data (Flowerdew and Green 1989; Volker and Fritsch 1999). In order to make better assumptions of the underlying spatial distribution, there are spatial regression methods that try to extract information from proxy data, to derive a statistical distribution (Horabik and Nahorski 2014). The authors in Bun et al. (2018) apply this technique to improve the spatial resolution of a spatial inventory of GHG emission in Poland. In Verstraete (2014), the authors presented the concept of using a rule-based system for grid remapping. Their detailed algorithm to use proxy data to generate a fuzzy rule-based system that performs the spatial disaggregation is elaborated on in Verstraete (2016).

Most of the above methods, with the exception of the methods in Horabik and Nahorski (2014) and Verstraete (2016) are directly applicable for regridding. In areal weighting, the spatial distribution is assumed per grid cell, but its definition using the amount of overlap does not pose problems with partial overlap and can thus directly be used for regridding. In spatial smoothing or regression methods, the assumed underlying distribution is not bound by a grid layout and regridding is achieved by resampling. The method in Horabik and Nahorski (2014) was designed specifically for spatial disaggregation. The method described in Verstraete (2016) was intended for regridding but limited to spatial disaggregation. The reason is that it was difficult to maintain the constraint (2). This constraint is simpler in the case of spatial disaggregation as it translates to a constraint that is local in each input

cell. In this contribution, we extend the applicability of these last two approaches by translating the regriding problem into a spatial disaggregation problem. As this was developed together with the fuzzy rule-based approach for spatial disaggregation, these methods best complement each other; the rule-based approach is thus considered as a use-case.

3 Spatial disaggregation using the fuzzy rule-based approach

The generation and use of a fuzzy rule-based system to perform a spatial disaggregation is explained in detail in Verstraete (2016) and summarized here. In this method, the weights as mentioned in the previous section are calculated by means of a fuzzy rule-based system. A fuzzy rule-based system (Mamdani and Assilian 1975) is one of many possibilities, next to, e.g. neural networks, deep learning and genetic algorithms, that aim to create an artificial intelligent system. The method uses fuzzy set theory (Zadeh 1965; Dubois and Prade 2000) and more specifically linguistic terms (Zadeh 1975) to connect input variables with output variables. The uncertainty is contained in the linguistic terms whose associated fuzzy sets are possibility distributions (Dubois and Prade 1999), associating all the domain values with the extent to which they match with the linguistic. The artificial intelligence is captured by this set of rules that model a reasoning; these rules are of the form

```
IF  $p_1$  is  $p_{1\_FSx}$  AND  $p_2$  is  $p_{2\_FSy}$  AND ...
   THEN output IS output_FSz
```

Here, p_i are variables and p_{i_FSx} are linguistic terms (e.g. low, high) against which the variables are evaluated. The variable `output` is assigned a linguistic term `output_FSz`. A common example for the world of fuzzy control is a rule base that connects temperature values with settings of a cooling device (if temperature is high, etc.). A linguistic term is represented by a mathematical function, a fuzzy set, which associates each value in the domain of possible values with a membership grade in the range $[0, 1]$; for a fuzzy set representing *low*, the membership grade indicates how well the values of the domain match the description *low* (ranging from 0: *not at all* to 1: *completely*). In the developed approach for spatial disaggregation, the rule base will be applied once for each output cell and will return an estimate for its value. This means that the rule base will be evaluated as many times as there are output cells.

The choice for a fuzzy rule-based system over other approaches to create an artificial intelligent system was made based on the possibilities of translating the spatial problem. The rules in this context were quite intuitive and exhibited parallels with the way an expert reasons about the problem (Verstraete 2014). In addition, the system is better readable, which not only allows for a better understanding but also facilitates adding expert knowledge directly into the rule base. There are several key aspects to the rule base: what are the values and linguistic terms used, and where do the rules originate from? The rules can be constructed manually, but more practically a training set is used. A training set contains both input values and output values; the rule base is typically constructed by evaluating each data pair of input and output values against the linguistic terms and retaining the most appropriate rule(s). Aggregation of the rules for all the data pairs results in the rule base; for the standard algorithm to construct rule bases from training data, we refer to Wang and Mendel (1992). The choice of variables greatly depends on the application. In the spatial context, the variables take values calculated from properties that connect the proxy grid to the data modelled in the input grid: geometric properties such as the amount of overlap, the distance or with other more complicated geo-spatial operations (e.g. buffers).

In Fig. 3, a variable calculated by means of weighted overlap is illustrated. The value v_{b_i} of the variable for b_i is then given by the formula

$$v_{b_i} = \sum_{j|c_j \cap b_i \neq \emptyset} \frac{S(c_j \cap b_i)}{S(c_j)} f(b_j). \quad (3)$$

Here, S denotes the surface area and f is the function that returns the value associated with a gridcell. This is one example, many others are possible, with several listed in Verstraete (2013).

These values have to be evaluated against suitably defined linguistic terms, but defining these terms implies the need for a proper definition of the domain of possible values. The domain has to be such that it contains most values that appear in the rule base (the authors in Wang and Mendel 1992 present how to handle values that fall outside of this domain; it suffices to consider them equal to the closest domain-limit). For the domain of the values, different choices can be made. The standard approach is to consider the minimum and maximum of the occurring values in the training set and impose those as the domain limits. In the spatial application, this proved too limiting as regional differences on the map may cause the system to evaluate all parts of an input cell equal (e.g. all are evaluated to low, but to a slightly different extent), resulting in insufficient distinction between them. In order to overcome this, we allowed different definitions for the domain limits for each data pair and adjusted the rule-based construction (and application) to deal with this. One way to define the domain limits is not to consider all the cells of the training set, but only those that are in the direct vicinity of the considered cell. The other approach developed was to calculate suitable minimum and maximum values for a given cell directly.

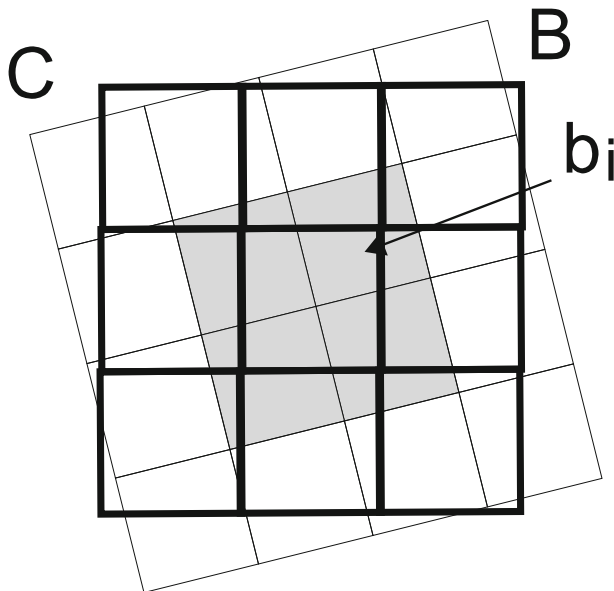


Fig. 3 Example of a variable that is obtained using the weighted overlap. The value of the cell b_i (of the output grid B) relates to the value of overlapping cells of the proxy grid C (these are shaded). The value is for example proportional to the amount of overlap

For a single value definition, different domains can be considered and the choice of the domain changes how the value is evaluated. This effectively means that the same value calculation with a differently defined domain constitutes a different variable. As a result, there are a vast number of possible choices for value definitions and domain definitions, resulting in a large number of possible values that can be considered in the rule-based system. The selection of the most suitable variable is done by checking how a grid, made from the values of the parameters and scaled using its minimum and maximum limit, resembles ideal output in the training set. For this the method presented in Verstraete (2017) is used. Apart from a minimum value for this resemblance, a maximum number of parameters is imposed to limit the size of the rule base.

The application of the rule base is quite straight forward: the values and domains for the selected variables are calculated and evaluated in the rule base. Every rule results in a linguistic term, and thus a fuzzy set. These are aggregated and a single value is extracted for each output cell. This last phase is achieved using the algorithm presented in Verstraete (2015a), to ensure satisfaction of the constraint that the partitions of an input cell should add up to that input cell.

A key benefit of this approach is that it automatically accounts for missing information in the proxy data: in the event of missing proxy data, there is no information on the underlying distribution and no parameter values can be calculated. As a result, every predicate is satisfied, which results in the system matching multiple rules. On the output, which is aggregated from the output values of all matching rules, this has the effect that it is more evenly distributed and resembles areal weighting. Barring any other information on the underlying distribution, this is one of the more conservative approaches as it assumes a uniform distribution in each grid cell. If multiple proxy datasets are used, where one or more proxy data contains no information for a location, the available data is used to redistribute the data, but the effect is weakened because of the datasets that lack the data.

4 General remapping using the fuzzy rule-based approach

4.1 Moving from spatial disaggregation

In the general remapping problem, as illustrated in Fig. 1a, the complication is that there are partial overlaps between cells of the input grid and those of the target grid. Most of the above methodology (the proxy data, the parameters, and the construction and application of the rule base) can all be considered for the general grid remapping problem without any change. The problem occurs in the final stages: the defuzzification and rescaling (or simultaneous defuzzification). Unlike in disaggregation, the aggregated total of a single target cell is unknown, as there are partial overlaps. As such, it is not possible to guarantee that the total value of the grid after remapping is the same as before the remapping. In order to overcome this, a new grid will be generated. This grid will serve as the target for a spatial disaggregation and is constructed such that

1. Remapping the input grid onto the intermediary grid is a spatial disaggregation problem, allowing for the defuzzification to be performed correctly, under the given constraint and
2. The intermediary grid can easily be mapped onto the initial target grid.

This intermediary target grid will be called the *segment grid* to avoid confusion. Its construction is elaborated on in the next section. While this method will allow any spatial

disaggregation algorithm to be applied for the general remapping problem, the rule-based approach in particular can benefit from it. The main reason for this is that the input grid can be treated similarly as proxy data, further helping to improve the estimates of the spatial distribution of the underlying data.

4.2 Segment grid

4.2.1 Construction

The general overlay problem for grids starts with a given input grid A and a given target grid B . To allow a spatial disaggregation algorithm to be applied, the partial overlaps need to be removed from the problem. For this purpose, the segment grid will be constructed. When overlaying both the input grid and the target grid, the gridlines of the target grid divide each cell of the input grid into a number of segments. The segment grid is the irregular grid comprised of exactly these segments. For a grid A defined as the set of cells $a_i, i = 1..n_A$ and B as the set of cells $b_j, j = 1..n_B$, the segment grid C is the set of cells $c_k, k = 1..n_C$ defined by

$$C = \bigcup_{\forall a_i \in A, \forall b_j \in B} \{c_k | c_k = a_i \cap b_j\}. \tag{4}$$

Determining the segment grid can be done by calculating the intersection between a cell a_i of grid A and all the cells b_j of grid B . This procedure is illustrated in Fig. 4.

In Fig. 4a, the cell a_i is highlighted: it is shaded with a dark grey colour and its vertices are marked with four bold circles. The lines of grid B are shown as dotted lines. These lines partition all cells of grid A in segments; for cell a_i , the intersection of these dotted lines with its outline are indicated by circles. A new grid is defined using both such intersection points and all vertices of both grids. For cell a_i there are four segments, but other cells may have a different amount of segments (i.e. the cell directly below a_i is partitioned in five segments). To further highlight the segments, one of the segments of a_i is shaded in light grey and labelled c_k . A similar partitioning can be performed on the other grid cells, resulting in the segment grid C as shown in Fig. 4b, with one segment highlighted.

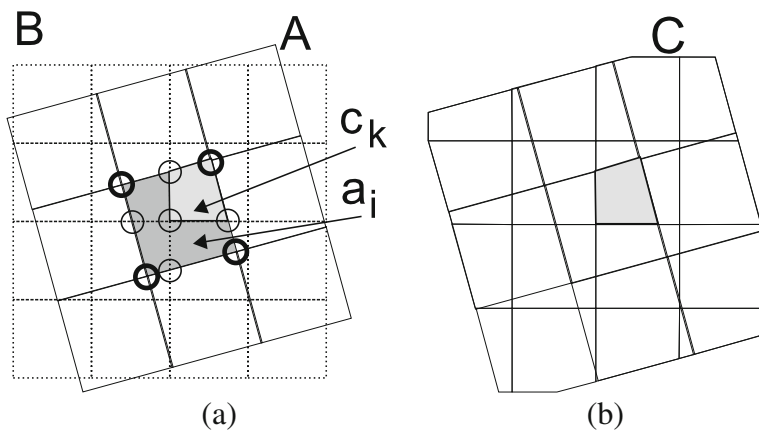


Fig. 4 Illustration of the construction of the segment grid, by using the gridlines of the target grid B to partition the cells of the input grid A . In **a**, the bold circles indicate the corners of grid cell a_i (shaded), the other circles indicate the intersection points of the sides of a_i with the gridlines of B . This results in four partitions, where one labelled c_k is highlighted. The entire segment grid C , where again c_k is highlighted, is shown in **b**. Note that this is an irregular grid

Determining the segment grid is a complex, time-consuming operation as it is necessary to find the intersecting cells between both grids. In practice however, this operation can be optimized as spatial queries and spatial indexes allow for rapid detection of the intersecting segments.

4.2.2 Properties

The segment grid has a number of interesting properties. In general, it will be an irregular grid with a finite number of grid cells as both A and B have a finite number of cells. From its definition, the segment grid partitions each cell of grid A into a number of segments. No segment overlaps with multiple cells of grid A . This means that the segment grid can be used as a target for a spatial disaggregation of grid A . For this purpose, the fuzzy rule-based approach can be used as explained earlier, resulting in fuzzy sets for the cells of the segment grid C . Once those fuzzy sets have been determined and defuzzified, it suffices to aggregate the values of the cells that together form a cell of B to determine the value for B . From the construction, the segment grid also partitions each cell of grid B into segments, so every cell b_j in grid B is the union of a number of cells of the segment grid C . This is illustrated in Fig. 5a. The segment grid C is shown in a dotted line, and one segment c_k is shaded in light grey. This segment belongs to exactly one cell of grid B , shaded in dark grey. The particular grid cell b_j is comprised of exactly four cells from the segment grid. This is the case for all cells of B and as such the values of segments in the segment grid can easily be aggregated to determine the values of the grid cells of B .

4.3 Usage and special combination with fuzzy rule-based approach

The use of the segment grid theoretically allows for all methods for spatial disaggregation to be applied for general grid remapping. The fuzzy rule-based approach in particular lends itself for using the segment grid, due to the way the result is determined. The rule-based approach calculates weights implicitly (as explained in Section 2.2) to perform the remapping. The result of these calculations are then corrected for the constraint imposed by the

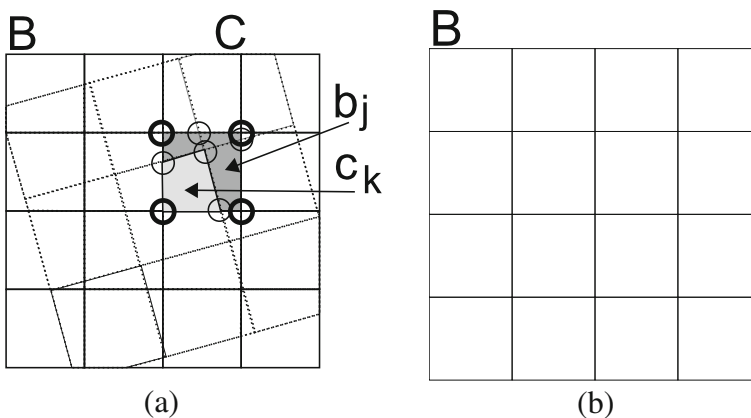


Fig. 5 Illustration of the recombination of segments of the segment grid C to yield the target grid B . In **a**, the circles indicate corners of segments of the segment grid C , where bold circles mark the corners of the target grid cell b_j . The cell c_k is one of the four segments that are contained by the cell b_j ; combining these four cells results in b_j . The same is possible for other cells of B , combining the segments in this way results in the target grid B as shown in **b**

value of the input cell, either due to simple rescaling or by using combined defuzzification (Verstraete 2015a). In case of the segment grid, the same procedure is applied for each segment: the fuzzy rule-based approach will determine values and ranges for parameters used in the rules. These rules are of the form

```
IF  $p_1$  is  $p_{1\_FSx}$  AND  $p_2$  is  $p_{2\_FSy}$  AND ...
  THEN output IS  $output\_FSz$ 
...

```

Here, both the value for a parameter (the value is shown as, e.g. p_1 in the code above) and the linguistic term against which it is evaluated (p_{1_FSx}) are calculated for the segment. The rule-based approach does not need the input grid to determine how to redistribute the data within each cell; however, it does allow for the input grid to also be treated as proxy data. Properties such as overlap with or distance to cells of the input grid can help to determine better values for the segments.

The changes to the algorithm presented in Verstraete (2016) are minimal; the pseudocode from Verstraete (2016) has been modified to work for the general map overlay problem and is listed in the Appendix. In the pseudo code, the data in the cells are considered to be absolute data: e.g. amount of some emission in this cell or the amount of people living in this area, rather than values expressed per surface unit or concepts such as temperature. As such, the aggregation is a simple summation of the values associated with the segments. In case the data would be expressed not as absolute values but, e.g. per square km, the aggregation needs to be modified accordingly by first converting the data to absolute values, performing the disaggregation and then converting the obtained results back to values per square km.

A challenging aspect is the problem of robustness errors when dealing with spatial data (Hoffmann 1989, Chapter 4); (Verstraete 2015b): errors caused by the limited accuracy of the representation of coordinates in a computer system can cause coordinates of intersection points to be rounded, and thus not accurately represent the point. This in turn causes the vertices of the segments not to be represented accurately and may cause both false positives and false negatives when calculating intersections and testing for overlaps. The determination and usage of the segment grid relies heavily on these operations. To limit the impact, the effect of this rounding was considered in the spatial operations that were implemented: intersections and overlaps that have too small an area, relative to the area of the cells of either input or output grid, are ignored (Verstraete 2015c). This guarantees that there will be no errors in identifying which cells overlap with which segments, at the cost of possibly not detecting the existence of a small piece of a cell or segment. Due to their small size, however, the omission of these pieces will not affect the results. Another measure to limit the propagation of rounding errors is that both storage of and further operations on coordinates that are the results of intersections is avoided in the algorithms.

5 Discussion

5.1 Computational performance

The presented approach consists of multiple stages that have to be considered separately in order to judge the performance. Depending on the application, not all of these stages have to be performed every time. First, the creation of the rule base is considered. This only needs to be performed once for a given combination of grid layouts for given data. This means that if the numeric data changes (e.g. pollution data represented in the same grid, but with updated values) while the grids stay the same, the rule base does not have to be recalculated.

The different stages in the rule-based creation are as follows:

1. Creation of the segment grid
2. Selection of the parameters
3. Construction of the rules

To determine the rule base, a training dataset is needed. This training set has to be sufficiently large to allow for meaningful rules to be generated, but in practice it can be much smaller than most realistic datasets. The creation of the segment grid requires geometric computations to determine the intersections, and as such is fairly inefficient, worst case quadratic complexity (number of cells in the input multiplied with the number of cells in the output grid). However, the calculations can be sped up by using spatial indexes and spatial queries (candidate intersecting cells will be in the same vicinity). Furthermore, the entire process can be parallelized.

The selection of the parameters requires all possible parameters to be evaluated; this is a process that scales linearly with the number of output cells and the number of parameters, and can also be parallelized over both cells and parameters. In most situations, the number of parameters can be limited beforehand, by having an expert select which parameters should be considered. The rule-based construction uses the evaluated values and is linear process in the number of cells. It is a lengthy step of the process, but it only needs to be repeated once for a given combination of datasets. Future combinations of the same datasets with updated numerical values can make use of the already created rule base.

Applying the rule base is the second step and constitutes of the following:

1. Creation of the segment grid
2. Calculation of the parameters
3. Defuzzification/recombination of the results

The segment grid has to be determined for the dataset. However, if the rule base is applied on the same dataset but populated with different data (as in the example of updated values, for which the rule base can be reused), the segment grid is the same and as such it was determined in the first step. The calculation of a parameter happens once per output cell, thus is linear in both the number of cells and the number of parameters, and easily parallelized. Each parameter is computable in constant time for a given cell.

The result of the rule-based application is a regridded dataset which has less uncertainty than other regridded datasets due to the use of the proxy data. This new dataset can be used whenever necessary without recalculating.

5.2 Example case

The complexity of artificial intelligent systems makes them more difficult to verify than pure mathematical approaches: even if the operation of all the components is clear and verified, the interaction between them is sometimes impossible to predict and has to be verified. Depending on the application, there are a number of indicators that can help to assess the reliability and quality of the results. The results usually should be deterministic, stable (a small deviation of the input does not yield big changes of the output), and intuitive (as the system mimics an intelligent reasoning). The novel approach is illustrated by means of an artificial example, which not only allows for a verification with a perfect solution but also offers full control over the experiments. The test presented here is one of the many cases that were tested using the novel approach and it was selected as it highlights both the strengths and weaknesses. In this case, we consider an artificial dataset created using

several datasets concerning Warsaw, Poland. These datasets are shown in Fig. 6a and hold the road network, bodies of water, and parks. The road network is further classified in three types according to size, which is reflected by the drawing style: roads are drawn using two parallel lines, with the distance between the lines indicative for the size. In addition, there is specific knowledge concerning the roads, some are restricted (shown in orange), others forbidden (shown in red). The parks and bodies of water are overlaid with this data and shown respectively in green and blue.

Each segment of road in the road network is given a value derived from its length and type. This value will be used as an indication for the amount of emission and a grid will be generated. The goal in this example case is to remap a grid that models this simulated emission (as shown in Fig. 6a) represented in the form of a 13×13 grid (Fig. 6b) onto a 20×20 grid with different orientation (this target is shown further, in Fig. 8b). An amount of traffic will also be simulated to serve as proxy data. In order to make the problem more realistic, this amount should not perfectly match the emission data. To achieve this, traffic data will be randomized from the the emission data both in value and in spatial distribution. To determine the amount of traffic t_i^s for a segment s , we consider $t_i^s = e_s + r_s \times e_s \times f_i$. Here, e_s is the emission value that was generated for a segment, r_s is a random value in the range $[0, 1]$ (a different random value is used for each segment s) and f_i is a chosen value in the range $[0, 1]$ that allows us to control how much deviation is allowed. By varying the value of f_i , it is possible to generate traffic datasets that resemble the simulated emission to different extents. To achieve the spatial randomization, the data will not be associated with the segments that make up the road network but with a buffer around each segment; by varying the size of the buffers it is also possible to consider different levels of spatial disconnection. For this example, a value of $f_i = 0.5$ and buffers shown in Fig. 7a are used to define the proxy data. As many buffers are hidden behind others, not all buffers that accompany segments are visible. The proxy data is a 31×31 grid that is defined on the dataset in Fig. 7a and is shown in Fig. 7b.

The generated data makes use of the known road network, making it possible to also generate an ideal solution for the given target grid. As such, the performance of the developed method can be assessed against an ideal solution. The output is a 20×20 grid shown in Fig. 8b. This grid, combined with the input grid results in the segment grid as shown in Fig. 8a; the points on the figure were the basis for the artificially generated training set, positioned randomly but using knowledge on the location of water and parks. The sampling of these points using the input and output rasters resulted in the training set.

The regridding algorithm is set to use up to five variables and up to five linguistic terms for each of them. In addition, we did not allow the system to use variables defined using the input grid; only variables using the proxy grid were allowed. This was done to enhance the impact of the proxy data, as this better shows how the proxy data is used. The result of the regridding algorithm is shown in Fig. 8b and can be compared with the ideal solution. In each cell, a bar chart compares the ideal value, obtained by resampling the road network's generated emission values onto the target grid with the calculated value. In each cell, the left bar (in red) shows the ideal solution, whereas the right bar (in yellow) shows the calculated solution. The closer these bars are together, the better the regridding matches the ideal solution in that cell. The grid in general resembles the ideal solution quite well; we have to ignore some of the cells near the top edge of the result grid, as those locations were not covered by the input grid and hence have no value. Looking at locations that were covered by the input grid, it shows for example that below the lower enlarged section of the map, the algorithm manages to identify the vertical pattern that is present, but distributes it over two neighbouring cells. Looking at the road network in Fig. 6, it is clear that the ideal

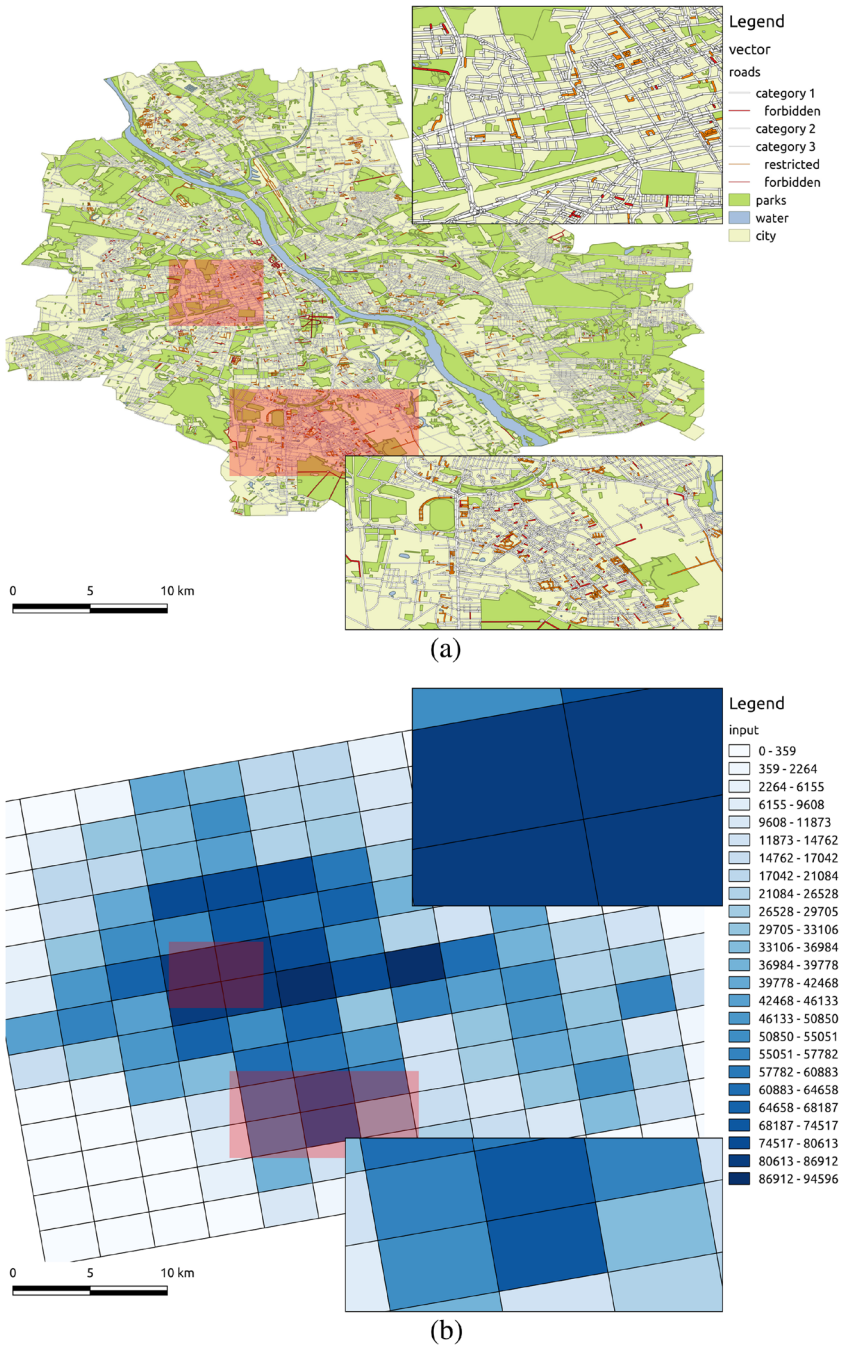
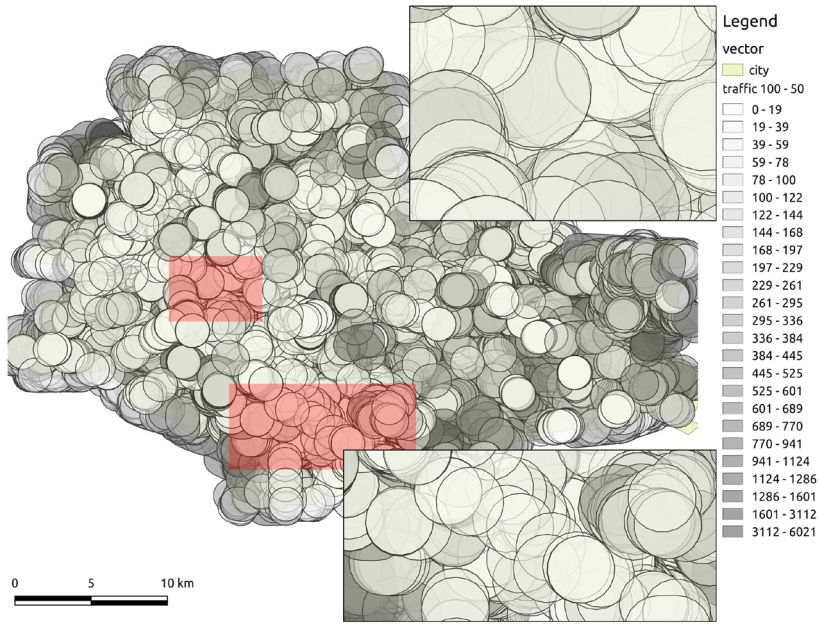
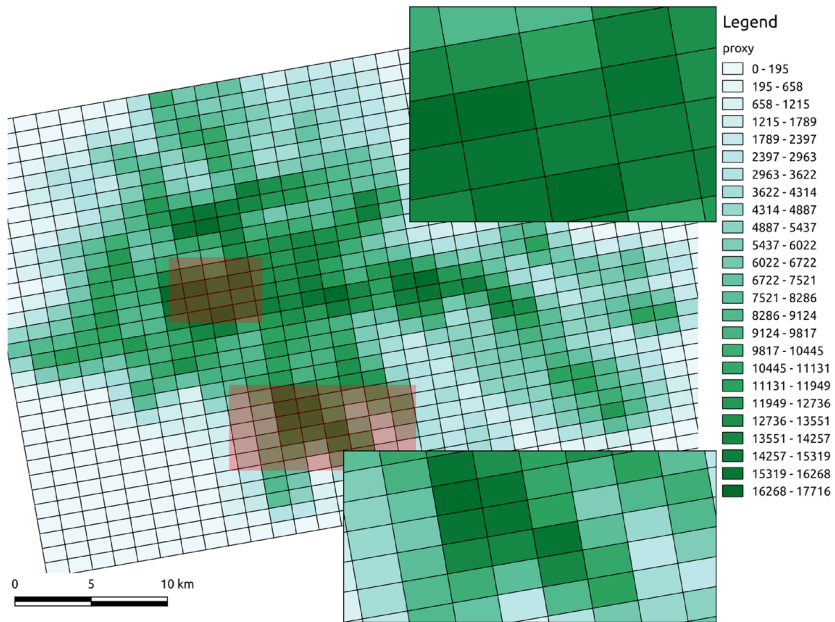


Fig. 6 Data for the generated example case. The source data is shown in **a** and contains the road network; the width of the road relates to its category, orange and red colours indicate respectively limited and forbidden (for traffic), green areas and water. From the road network, a grid as shown in **b** was generated, where each cell represents the amount of roads weighted with their category, this serves as an estimate for concentrations of a pollutant that relates to traffic

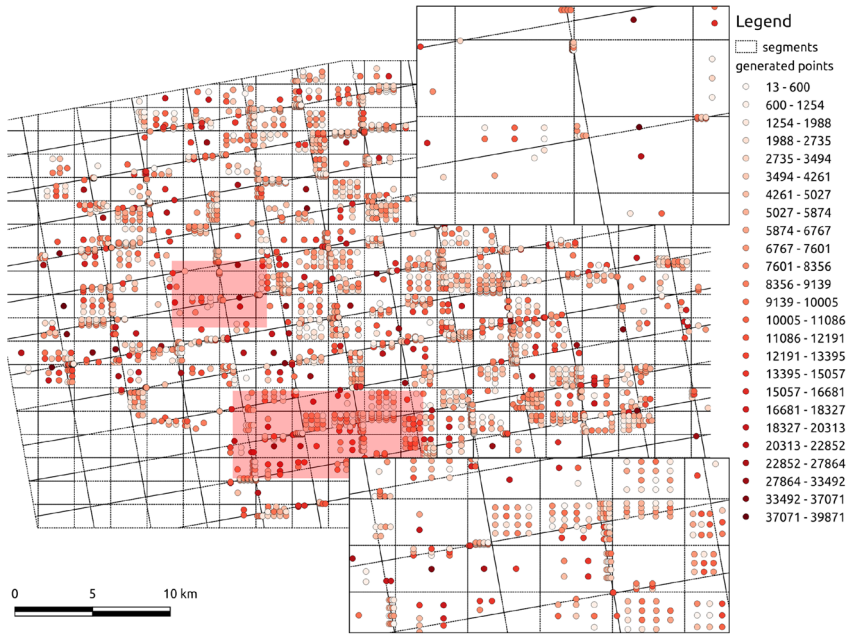


(a)

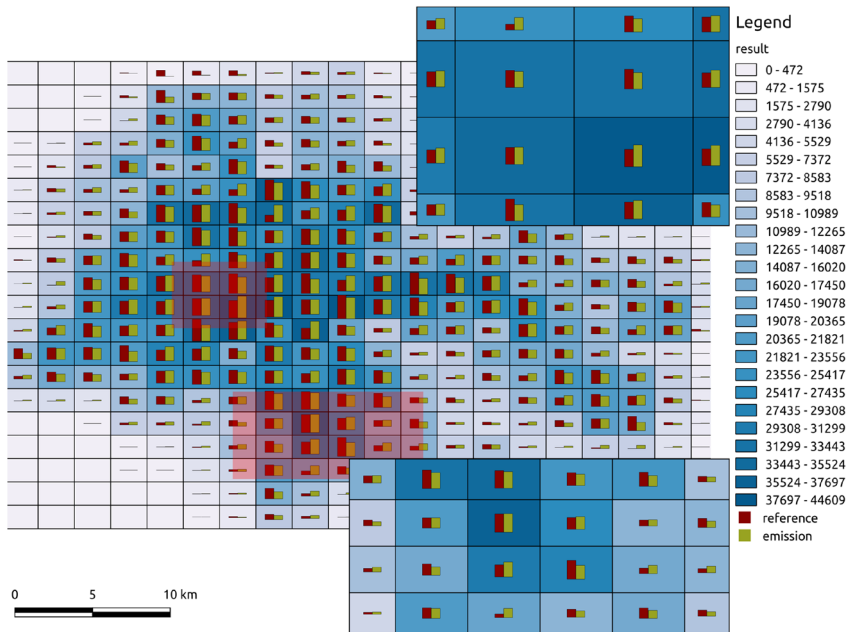


(b)

Fig. 7 Data for the generated example case. To make the artificial datasets more realistic, a random aspect was introduced to the road network of Fig. 6a: the values of the road network were multiplied with a random factor (to randomize their value somewhat) and a buffer was created around each segment of road (to randomize the location of the value). These buffers are shown in a, many are hidden below the ones that are visible. This data is used a simulated traffic data. For the example, this is sampled into a grid which will serve as proxy data b



(a)



(b)

Fig. 8 In **a**, the generated segment grid for the considered problem is shown. The system generates its own training data for a given problem; the points shown in this figure are part of this process. The result of the regriding algorithm by first disaggregating in to the segment grid are shown in **b**. In each grid cell, the bars compare the ideal solution (obtained by resampling the generated vector data on the grid), indicated by the red bar on the left, with the calculated solution indicated by the yellow bar on the right

values in that location are the result of a road, but this road is located is very close to the edge of the cell. The proxy data did not provide a distinction in this location and as such the system could not do better than distribute the way it did using this proxy data. The result is that the solution is more spread out spatially. Similarly, the region with lower values just above the topmost enlarged section is also not recognized as clearly as it shows on the ideal reference and as a result the values of the neighbouring cells are more evenly matched. This is actually the most logical choice if no other data is available; the method falls back to areal weighting if the proxy data cannot supply enough information. By contrast, the results of the region between both enlarged portions of the map much better follow the high/low-value pattern of the ideal solution, albeit with less extreme values, as is also visible near the bottom right of the topmost enlarged portion. This is of course to be expected: regrid-ding a 13×13 grid into a 20×20 carries a lot of uncertainty, particularly as the proxy data is also a grid (with the same issues of the unknown underlying distribution). Important is that the pattern is present and that the method falls back to a more conservative choice if the proxy data is not supplying much information. Additional experiments have shown that the intersection pattern between the proxy grid and the input grid plays a major role in the quality: in locations where the proxy grid is such that it partitions input cells, the algorithm is more likely to have better results as it provides the algorithm with more information on the underlying distribution in that location.

6 Conclusion

Combing or comparing gridded spatial datasets is a difficult process, prone to introducing or increasing uncertainty, yet it is necessary at different stages when working with spatial data, such as for example climate-related data. Improving the current algorithms to do so contributes to mitigation and adaptation strategies in two ways. First, when applied by experts in a preprocessing stage, it provides them with more accurate gridded data. More accurate data leads to more accurate analysis and conclusions. Second, when applied at the level of the decision-makers, it on the one hand allows for a more accurate comparison between the results of different reports (issued by different organizations as well as reports from different years), but it also allows to transform, with higher accuracy, from arbitrary grid layouts which are often chosen out of technical arguments, to layouts that are more suitable for policing, e.g. based on administrative borders.

In this contribution, an algorithm to solve the general overlay problem for gridded datasets is presented. The algorithm is a general modification of algorithms for spatial disaggregation, specifically tied together with an earlier presented algorithm that uses a fuzzy rule base to perform a spatial disaggregation. The modification consists of the creation of an intermediary grid (*segment grid*), which allows the initial problem to be solved as a spatial disaggregation problem. The segment grid as presented can theoretically be used with any method for spatial disaggregation; however, the fuzzy rule-based method makes it possible to consider the input data as proxy data for the remapping. A test case was presented, highlighting both the benefits and limitations of the approach.

The presented use of the segment grid is applicable for any spatial disaggregation method to perform a more general regrid-ding. The main disadvantage of the approach is in the calculations required to determine the geometry of the segment grid, but as in many cases, the same grids are used with varying data, these calculations need to be done only once. The segment grid in particular ties in perfectly with the author's previously presented algorithm for spatial disaggregation. This approach, where a range of possible values for each output

cell is estimated independently using artificial intelligence combined with proxy data works without modification on the segment grid. The constraints imposed on the values of the segment grid subsequently allow for extracting the most suitable values for the segments. Due to the difficulty in understanding artificial intelligent systems, many experiments on artificial data were performed; the provided example is one of these experiments aimed at verifying the approach. It shows that the rule-based approach in combination with the segment grid manages to achieve a regridding that comes close to the generated ideal solution. Future experiments are planned to perform regridding on verifiable real-world datasets and to derive an assessment of the uncertainty of the result. The methods and approaches presented in this paper can be used to improve and inform global mitigation and adaptation strategy identification and implementation in countries worldwide.

Funding information This work has received financial support from the Consellería de Cultura, Educación e Ordenación Universitaria (accreditation 2016-2019, ED431G/08) and the European Regional Development Fund (ERDF).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

The changes to the algorithm presented in Verstraete (2016) are minimal; the pseudocode from Verstraete (2016) has been modified to work for the general map overlay problem. The main benefit of using this method is that specific parameters can be defined, which in turn improves the regridding. The changes to the pseudocode are highlighted in bold below

```

determine segment grid from input/target
  # see IV.2
  rulebase = new rulebase
  FOR each parameter
    add rules to rulebase based on selected parameters
  END FOR

  FOR each input cell
    FOR each segment in input cell
      FOR each parameter
        calculate parameter value
        calculate parameter range
        define linguistic terms
        fill terms/value in rulebase
      END FOR
      evaluate(rulebase)
    END FOR
    use segment values as weights to determine final segment values
    assign final segment values to segments
  END FOR
  FOR each target cell
    # see IV.3
    target cell value = 0
    FOR each segment in target cell
      target cell value += segment value
    END FOR
  END FOR

```


References

- Boychuk K, Bun R (2014) Regional spatial inventories (cadastres) of ghg emissions in the energy sector: accounting for uncertainty. *Clim Change* 124(3):561–574. <https://doi.org/10.1007/s10584-013-1040-9>. ISSN 1573-1480
- Bun R, Charkovska N, Danylo O, Halushchak M, Horabik-Pyzel J, Kinakh V, Lesiv M, Nahorski Z, See L, Topylko P, Valakh M (2018) Development of a high resolution spatial inventory of ghg emissions for Poland from stationary and mobile sources. *Mitigation and Adaptation Strategies for Global Change* this issue
- Burrough PA, McDonnell RA (1998) Principles of geographical information systems. Technical report
- Danylo O, Bun R, Charkovska N, See L, Topylko P, Tymkow P, Xianguang X (2015) Accounting uncertainty for spatial modeling of greenhouse gas emissions in the residential sector: fuel combustion and heat production. In: Proceedings of the 4th International workshop on uncertainty in atmospheric emissions, pp 193–200. Systems Research Institute, Polish Academy of Sciences
- Dubois D, Prade H (1999) The three semantics of fuzzy sets. *Fuzzy Set Syst* 90:141–150
- Dubois D, Prade H (2000) Fundamentals of fuzzy sets. Kluwer Academic Publishers
- Flowerdew R, Green M (1989) Statistical methods for inference between incompatible zonal systems. *Accur Spatial Datab*, 239–247
- Goodchild MF, Lam S-N (1980) Areal interpolation: a variant of the traditional spatial problem. *Geo-Process* 1:297–312
- Gotway CA, Young LJ (2002) Combining incompatible spatial data. *J Am Stat Assoc* 97(458):632–648
- Hoffmann CM (1989) Geometric and solid modeling: an introduction. Morgan Kaufmann Publishers Inc., San Francisco. ISBN 1-55860-067-1
- Hogue S, Andres RJ, Marland E, Marland G, Woodard D (2016) Uncertainty in gridded co2 emissions estimates. *Earth's Fut* 4(5):225–239. <https://doi.org/10.1002/2015EF000343>
- Hogue S, Boden T, Marland E, Marland G, Roten D (2018) Gridded estimates of co2 emissions: uncertainty as a function of scale. *Mitigation and adaptation strategies for global change* this issue. <https://doi.org/10.1007/s11027-017-9770-z>
- Horabik J, Nahorski Z (2014) Improving resolution of a spatial air pollution inventory with a statistical inference approach. *Clima Change*, 575–589. ISSN 0165-0009. <https://doi.org/10.1007/s10584-013-1029-4>
- Hryniewicz O, Horabik J, Jonas M, Nahorski Z, Verstraete J (2014) Compliance for uncertain inventories via probabilistic/fuzzy comparison of alternatives. *Clim Change* 124(3):519–534. <https://doi.org/10.1007/s10584-013-1031-x>. ISSN 1573-1480
- Hutchins MG, Colby JD, Marland E, Marland G (2017) A comparison of five high-resolution spatially-explicit, fossil-fuel, carbon dioxide emission inventories for the United States. *Mitig Adapt Strat Glob Chang* 22(6):947–972. <https://doi.org/10.1007/s11027-016-9708-x>
- Jonas M, Krey V, Marland G, Nahorski Z, Wagner F (2014) Uncertainty in an emissions-constrained world. *Clim Change* 124(3):459–476. <https://doi.org/10.1007/s10584-014-1103-6>. ISSN 1573-1480
- La Notte A, Nocera S, Tonin S (2018) The effects of uncertainty for policy decisions. An initial screening procedure for air emission estimates undertaken at regional level. *Mitigation and adaptation strategies for global change*, this issue
- Lesiv ML, Fritz SF, Schepaschenko D, Shvidenko A, See L (2018) A spatial assessment of the forest carbon budget for Ukraine. *Mitigation and adaptation strategies for global change* this issue
- Mamdani E, Assilian S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. *Int J Man-Mach Stud* 7(1):1–13. [https://doi.org/10.1016/S0020-7373\(75\)80002-2](https://doi.org/10.1016/S0020-7373(75)80002-2). ISSN 0020-7373
- Rafaj P, Amann M, Siri J, Wuester H (2014) Changes in european greenhouse gas and air pollutant emissions 1960–2010: decomposition of determining factors. *Clim Change* 124(3):477–504. <https://doi.org/10.1007/s10584-013-0826-0>. ISSN 1573-1480
- Rigaux P, Scholl M, Voisard A (2002) Spatial databases with applications to GIS. Morgan Kaufman Publishers
- Shekhar S, Chawla S (2003) Spatial databases: a tour. Pearson Educations
- Tobler WR (1979) Smooth pycnophylactic interpolation for geographic regions. *J Am Stat Assoc* 74(367):519–536
- Tomlin C (1994) Special issue landscape planning: expanding the tool kit map algebra: one perspective. *Landsch Urban Plan* 30(1):3–12. ISSN 0169-2046
- Verstraete J (2013) Parameters to use a fuzzy rulebase approach to remap gridded spatial data. In: Proceedings of the 2013 Joint IFSA World congress NAFIPS annual meeting (IFSA/NAFIPS), pp 1519–1524
- Verstraete J (2014) Solving the map overlay problem with a fuzzy approach. *Clim Change*, 591–604. ISSN 1573-1480. <https://doi.org/10.1007/s10584-014-1053-z>

- Verstraete J (2015a) Algorithm for simultaneous defuzzification under constraints. In: Proceedings of the 2015 conference of the international fuzzy systems association and the European society for fuzzy logic and technology. Atlantis Press, ISBN (on-line): 978-94-62520-77-6. <https://doi.org/10.2991/ifsa-eusflat-15.2015.48>
- Verstraete J (2015b) Remapping gridded data using artificial intelligence: real world challenges. In: Proceedings of the 4th international workshop on uncertainty in atmospheric emissions. Systems Research Institute, Polish Academy of Sciences, Warszawa, pp 130–136. ISBN 83-894-7557-X
- Verstraete J (2015c) Dealing with rounding errors in geometry processing. In: Flexible Query answering systems 2015 - Proceedings of the 11th international conference FQAS, 2015, Cracow, Poland, October 26-28, 2015, pp 417–428. https://doi.org/10.1007/978-3-319-26154-6_32
- Verstraete J (2016) The spatial disaggregation problem: simulating reasoning using a fuzzy inference system. *IEEE Trans Fuzzy Syst* PP(99):1–1. <https://doi.org/10.1109/TFUZZ.2016.2567452>. ISSN 1063-6706
- Verstraete J (2017) Fuzzy quality assessment of gridded approximations. *Appl Soft Comput* 55:319–330. <https://doi.org/10.1016/j.asoc.2017.01.051>. ISSN 1568-4946
- Volker W, Fritsch D (1999) Matching spatial datasets: a statistical approach. *Int J Geogr Inf Sci* 13(5):445–473
- Wang L-X, Mendel JM (1992) Generating fuzzy rules by learning from examples. *IEEE Trans Syst Man Cybern* 22(6):1414–1427
- Woodard D, Branham M, Buckingham G, Gosky R, Hogue S, Hutchins M, Marland E, Marland G (2014) A spatial uncertainty metric for anthropogenic co2 emissions. *Greenhouse Gas Measur Manag* 4(2-4):139–160. <https://doi.org/10.1080/20430779.2014.1000793>
- Zadeh LA (1965) *Fuzzy Sets Inf Control* 8:338–353
- Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning i, ii, iii. *Inf Sci* 8(3, 4, 9):199–251, 301–357, 43–88