4th International Workshop on Uncertainty in Atmospheric Emissions

7-9 October 2015, Krakow, Poland

PROCEEDINGS

4th International Workshop on Uncertainty in Atmospheric Emissions

7–9 October 2015, Kraków, Poland

PROCEEDINGS

Warszawa 2015

4th International Workshop on Uncertainty in Atmospheric Emissions
7- 9 October 2015, Cracow, Poland

Printed from the material submitted by the authors.

ISBN 83-894-7557-X
EAN 9788389475572

About the Workshop

The assessment of greenhouse gases and air pollutants (indirect GHGs) emitted to and removed from the atmosphere is high on the political and scientific agendas. Building on the UN climate process, the international community strives to address the long-term challenge of climate change collectively and comprehensively, and to take concrete and timely action that proves sustainable and robust in the future. Under the umbrella of the UN Framework Convention on Climate Change, mainly developed country parties to the Convention have, since the mid-1990s, published annual or periodic inventories of emissions and removals, and continued to do so after the Kyoto Protocol to the Convention ceased in 2012. Policymakers use these inventories to develop strategies and policies for emission reductions and to track the progress of those strategies and policies. Where formal commitments to limit emissions exist, regulatory agencies and corporations rely on emission inventories to establish compliance records.

However, as increasing international concern and cooperation aim at policy-oriented solutions to the climate change problem, a number of issues circulating around uncertainty have come to the fore, which were undervalued or left unmentioned at the time of the Kyoto Protocol but require adequate recognition under a workable and legislated successor agreement. Accounting and verification of emissions in space and time, compliance with emission reduction commitments, risk of exceeding future temperature targets, evaluating effects of mitigation versus adaptation versus intensity of induced impacts at home and elsewhere, and accounting of traded emission permits are to name but a few.

The *4th International Workshop on Uncertainty in Atmospheric Emissions* is jointly organized by the *Systems Research Institute of the Polish Academy of Sciences*, the Austrian-based *International Institute for Applied Systems Analysis*, and the *Lviv Polytechnic National University*. The 4th Uncertainty Workshop follows up and expands on the scope of the earlier Uncertainty Workshops – the *1st Workshop* in 2004 in Warsaw, Poland; the *2nd Workshop* in 2007 in Laxenburg, Austria; and the *3rd Workshop* in 2010 in Lviv, Ukraine.

Steering Committee

Rostyslav BUN (Lviv Polytechnic National University, UA)
Matthias JONAS (International Institute for Applied Systems Analysis, AT)
Zbigniew NAHORSKI (Polish Academy of Sciences, PL) – Chair

Scientific Committee

Evgueni GORDOV (Siberian Center for Environmental Research & Training, RU)
Piotr HOLNICKI-SZULC (Polish Academy of Sciences, PL)
Joanna HORABIK-PYZEL (Polish Academy of Sciences, PL)
Olgierd HRYNIEWICZ (Polish Academy of Sciences, PL)
Katarzyna JUDA-REZLER (Warsaw University of Technology, PL)
Petro LAKYDA (National University of Life and Environmental Sciences of Ukraine, UA)
Myroslava LESIV (Lviv Polytechnic National University, UA)
Gregg MARLAND (Appalachian State University, USA)
Sten NILSSON (Forest Sector Insights AB, SE)
Tom ODA (Univ. Space Research Association, NASA Goddard Space Flight Center, USA)
Stefan PICKL (Universität der Bundeswehr München, Germany)
Elena ROVENSKAYA (International Institute for Applied Systems Analysis, AT)
Kazimierz RÓŻAŃSKI (AGH University of Science and Technology in Cracow, PL)
Dmitry SCHEPASCHENKO (International Institute for Applied Systems Analysis, AT)
Anatoly SHVIDENKO (International Institute for Applied Systems Analysis, AT)
Jacek SKOŚKIEWICZ (National Centre for Emissions Management, PL)
Philippe THUNIS (EC Joint Research Centre Ispra, EU)
Marialuisa VOLTA (University of Brescia, IT)

Local Organizing Committee

Joanna HORABIK-PYZEL
Jolanta JARNICKA - Chair
Weronika RADZISZEWSKA
Jörg VERSTRAETE

Remapping gridded data using Artificial Intelligence: real world challenges

Jörg Verstraete[1]

[1] Systems Research Institute, Polish Academy of Sciences,
Warsaw, Poland
Jorg.Verstraete@ibspan.waw.pl

Abstract

Working with spatial data, regardless of the specific content (emission data, population data, land use data, etc.), requires dealing with gridded datasets. A grid is a commonly used representation method for data, where a value of interest is associated with each cell of the grid. While very adequate for representing and analysing data, combining data from different sources implies working with different grids, and this is more complicated. In this article, we present some preliminary findings of applying a novel approach to map spatial data onto a different grid. The approach simulates intelligent reasoning through the use of artificial intelligence and employs additional knowledge to help create a high quality remapping. We also present the difficulties in applying this methodology in real world applications.

Keywords: map overlay problem, grid remapping, spatial operations, artificial intelligence

1. Introduction

Data regarding atmospheric emissions are one example of data that carries a spatial dependency. For research purposes, e.g. to investigate the exposure of a population or to correlate data, it is often necessary to combine data from different sources. As data tend to come from different sources, the grids on which they are defined can be incompatible: different size of cells, different orientation or a combination. This is called the *map overlay problem* and it occurs when the data of a grid cell in one grid needs to be correlated or even compared to data that is presented an a different, incompatible grid. The simplest solution is to use areal weighting, which allows to remap one grid onto a different grid, using the amount of overlap of the cells as weights used to redistribute the data. This however implicitly assumes that the data are uniformly distributed within each single grid cell. While this assumption may hold for some data, or even for some cells, it is not always a valid assumption. The current approaches are ignorant to the fact that many other data and knowledge are available; some of this data may be known to exhibit a correlation to the data that we need to remap. In [1], we presented the first concept of an artificial intelligent system that is able to perform the remapping of one grid onto another grid, using this additional information to improve the spatial distribution of the modeled data. Following the first concept, several implementations were made and experiments were performed. Here, we present our findings regarding the challenges ahead when needing to apply this method on real world data.

In the next section, the representation of spatial data is shortly introduced, while an introduction to the artificial intelligent system used it in Section 3. The challenges with real world data are presented in Section 4, followed by the conclusion.

2. Spatial data

2.1 Spatial data representation

Spatial data can be represented in one of two commonly used models: feature based or field based [2,3]. In a feature based model, basic geometric objects are used to represent real world objects: lines are used to represent roads, polygons represent areas, etc. In a field based approach, a numeric value that carries a spatial component (e.g. emission values) are modeled over a region of interest. This can be achieved using triangular networks (commonly used for e.g. altitudes) or grids. In the application of modelling emissions, grids are more common. In a grid, the region of interest is partitioned in a number of grid cells, that completely cover the region of interest; if all cells have the same shape and size, the grid is considered regular. The cell of a grid is considered the smallest possible unit. For a grid that represents e.g. emission values, the grid provides no information regarding the distribution of the emission within each cell: the emission can be concentrated in one part of the cell, uniformly spread over the cell, or can have any other distribution. This causes problems when incompatible grids – grids that have ill aligned grid cells – as there is no easy mapping from one grid onto another grid. Exaples of incompatible grids are shown in Figure 1.

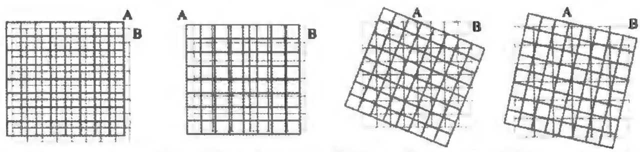

Figure 1. Examples of incompatible grids: shifted, different cell size, rotated or a combination.

2.2 Grid remapping algorithms

The map overlay problem occurs when one grid needs to be remapped onto another grid. Several approaches exist in literature, but all of them make either an implicit or an explicit assumption regarding the underlying distribution. For more details we refer to [4], and briefly describe the most common methods below.

The easiest and most commonly used algorithm for grid remapping is areal weighing. This approach implicitly assumes a uniform distribution of the data in each grid cell individually. The calculation to remap one grid onto another is very easy, as it suffices to consider the relative amount a cell of one grid overlaps with a cell of the other grid. This is illustrated on Figure 2. While effective in many cases, the method fails when the assumption does not hold.

Spatial smoothing is a second approach. Here, the modeled data is considered as a third dimension, which is subsequently smoothed and resampled. This is illustrated on Figure 2. The implicit assumption here is that the data is smooth of the entire grid and the performance of the method depends on the accuracy of this assumption.

The last method mentioned is spatial regression, where a priori statistical assumptions on the distribution of the data are used to control the grid remapping. The application of this method requires expert knowledge and quite complicated calculations. The distribution of the data is explicitly assumed here, but such knowledge may not be available.

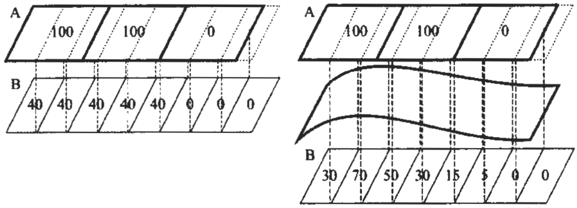

Figure 2. Illustrations of areal weighing (left) and spatial smoothing (right).

3. Artificial intelligent method and spatial grid remapping

3.1 Short introduction to fuzzy set theory

Fuzzy set theory is an extension to set theory, presented by Zadeh in [5]. In a fuzzy set, each of the elements carries a value from the interval *[0,1]*, this is the membershipgrade μ. This value can have one of three interpretations [6]: as a degree of membership – in which case it expresses "how much" the element belongs to the set, as a degree of certainty – in which case it reflects how certain it is the element belongs to the set, or finally as a degree of possibility – to indicate how possible it is the element belongs to the set. As such, a fuzzy set is defined by means of a traditional set and an associated membership function, which maps each element to its membership degree. Many applications of fuzzy sets exists [7], but for the application here we consider the possibilities of representing imprecise values and linguistic terms. An imprecise value (e.g. approximately 50) can be represented by a fuzzy set which has a membershipgrade 1 for the element 50, and decreasing membership grades as values are further from 50, as indicated on Figure 3. A linguistic term (such as "small") can be represented as illustrated on Figure 3: 0 is considered small with degree 1, larger numbers have decreasing membership grades and numbers above 50 are not considered small (they have membership grade 0). Examples for the terms "medium" and "large" are also on Figure 3. The definitions of course depend on the domain and application and the provided fuzzy sets are just an example.

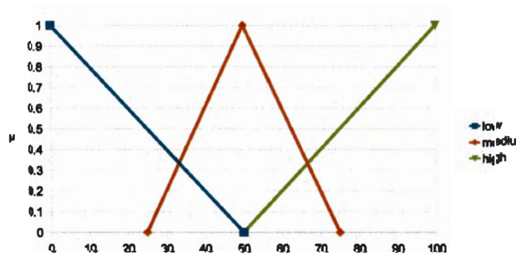

Figure 3. Examples of fuzzy sets used to represent low, medium and high numbers (on a scale from 0 to 100).

3.2 Fuzzy rulebase systems

Artificial Intelligence is a term that covers many approaches; for the presented algorithm, a rulebase system [8] is considered. Fuzzy systems have proven their effectiveness in control and applications can be found in many household appliances. A rulebase consists of a number of rules, for example:

if x is small then y is small
if x is medium then y is medium
if x is high then y is high

Here, x is an input parameter, which is a normal number, high/medium/low are linguistic terms represented by fuzzy sets, and y is an output parameter. All the rules are evaluated, so a given x can be both high and medium at the same time (e.g. using the definitions on Figure 3), but each to a different extent. Each rule results in a fuzzy set for y; the outputs of all rules are aggregated and defuzzified to yield the final result. For more details on this method, we refer to [8].

3.3 Concept of grid remapping using a fuzzy rulebase

The use of a fuzzy rulebase system to perform grid remapping requires the creation of the rulebase. The first problem is: what are the parameters used in the rulebase (x in the above example)? Several parameters can be considered as mentioned in [9], but in general parameters are what allow the additional data to be used. One example for a parameter is the amount of overlap of the auxiliary grid with the input cell under consideration. The second problem is: to define low/medium/high, it is necessary to find limits for the parameter. A lower limit could be the value of the grid cells of the auxiliary grid that are fully contained by the cell, whereas an upper value could be the total value of the grid cells of the auxiliary grid that intersect the cell.

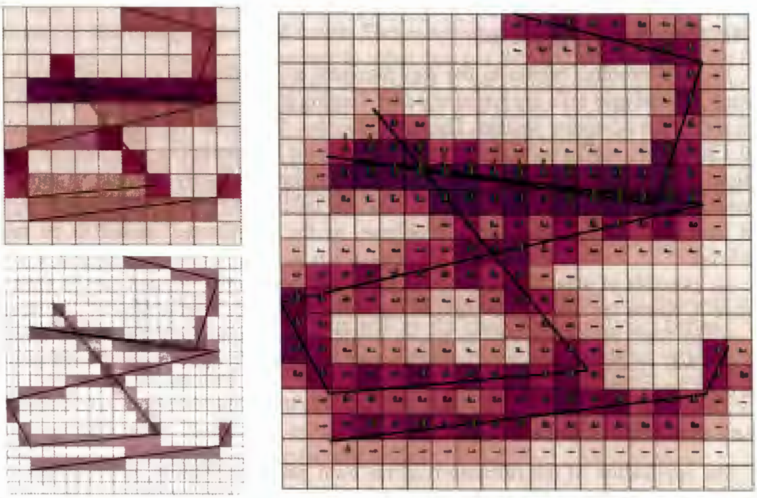

Figure 4. Example of the remapping algorithm: input (top-left), auxiliary data (bottom-left) and result (right). The line pattern shows the underlying distribution, the bar charts in the result cells show - from left to right - the result obtained through areal weighting, the ideal result and the result obtained with the presented approach.

Once it is known which parameters can be used, an appropriate rulebase can be constructed. To determine the result of the remapping, it suffices to apply the rulebase for every output, calculating the parameters and evaluating the rulebase to yield a fuzzy result for the output cell. This concept is explained in more detail in [1].

An example is shown on Figure 4. The example is still artificial, but should highlight problems that also can occur on real world data. The dark lines are the underlying lines that contain data and from which the grids were defined. The shades in the result reflect the values obtained with the presented method. Compared to areal weighting, it is clear the method is able to identify the 2 nearly horizontal lines near the bottom, whereas areal weighting sees them as one big region. The presented method tends to assign lower values to cells that are located further away from the ideal line, which is also desirable. On the two nearly horizontal lines near the bottom, an alternating pattern is visible, from left to right, which is an artifact introduced by the method. The next section aims to explain the origin of these and other problems and ties them in to real world situations.

4. Challenges related to applying the rulebase system

Various prototype implementations and proofs of concept have proven that the methodology can work. However, initial attempts at applying the prototype implementations on real world data have revealed some issues that still need resolving. First, there are problems of a more technical nature, described in the next 2 subsections; next there are problems related to the real world data itself, described in the subsequent 2 subsections. The first two problems can still be resolved using artificially generated examples, but the latter two would benefit from real world data.

4.1 Mathematical precision

The first problem relates to the way numbers are handled on a computer system. Coordinates of grid cells are represented by floating point values, which have a limited precision on a computer system. The consequences of this are very well explained in [10, Chapter 4], and in particular they pose problems for parallel or near parallel lines, which is the case for the grid cells. The presented algorithm highly depends on correctly assessing the overlap between intersections and calculating intersection areas. Initial test data suffered less from such problems, as the coordinates tended to be more artificial.

An example of what happens when values get incorrectly rounded can be seen on Figure 5, where the thick lines indicate locations where the intersection is incorrectly identified. Such errors can lead to wrong limits for the parameters. One work around for this was recently developed and is presented in [11], a more general workaround has been developed but still needs to be verified.

4.2 Parameter definitions and their limits

The application of the rulebase requires both parameter values, lower limits and upper limits. All three of them have equal importance, as a poorly defined lower and upper limit can make the parameters useless. In [9], we presented some suggestions for parameters. The example on Figure 4 was performed using a single parameter that relates to overlap of the auxiliary data. While the data does concentrate more towards the lines, the end result somewhat reflects the auxiliary grid: this explains the fluctuations in data that should be constant. Use of multiple parameters, even relating to the same data, should neutralize this effect, while still providing a result that shows a better distribution. Research in additional parameters is currently the next phase of the research.

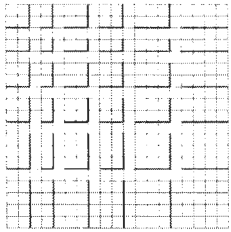

Figure 5. Example of a grid where rounding of the coordinates cause incorrect identification of intersection: the thicker lines are lines where neighbouring gridcells are incorrectly identified as intersecting.

4.3 Relative grid distributions

This topic is somewhat related to the previous topic, but there is a difference. Independent of the parameter definitions is the impact of the relative positions of the different grids. The parameter currently considered in the example is completely useless if the grid onto which the data is remapped exactly divides each grid cell. The reason for this is that in this situation, the parameter value, its lower limit and its upper limit will all be equal, and therefore no evaluation can be made by the rulebase system. The example on Figure 4 shows a fluctuating pattern along the near-horizontal lines: the values are alternating higher and lower. This is not desirable, as their values ought to be the same or at least similar: neither to input data nor the auxiliary data indicate this alternating pattern. The reason for its occurrence is the combination of the parameter that was used and the relative grid position. This is a second example that shows an effect that can happen, if the combination of the grid distributions and the parameters is not ideal. Solving this is both tied to solving the problems mentioned in 4.2, but also in making sure the auxiliary supplied data is indeed useful and of good quality for the considered problem.

4.4 Availability and quality of the data

The last aspect relates to the data itself. The assumption is made that auxiliary data are available, which for many research will be the case. The correlation between the data to be remapped and the auxiliary data should be known from prior research and not discovered on the data set at hand. The data should also be of good quality: grid obtained from down-sampling an existing grid might appear to be of higher precision, but internally is not. Regardless of the parameters implemented in the system, the use of such grids may provide unsatisfactory results.

5. Conclusions

The article shortly describes a novel approach to remap gridded spatial data and lists the challenges in bringing the approach from theory to practise. The biggest challenges are listed, along with the ideas that will be pursued to solve them.

References

[1] Verstraete J. (2014) Solving the map overlay problem with a fuzzy approach. Climatic Change, pp. 1–14.

[2] Rigaux P., M. Scholl, and A. Voisard (2002) Spatial databases with applications to GIS. Morgan Kaufman Publishers.

[3] Shekhar S. and S. Chawla (2003) Spatial databases: a tour. Pearson Educations.

[4] Gotway C. A. and L. J. Young (2002) Combining incompatible spatial data. Journal of the American Statistical Association, 97(458) pp. 632–648.

[5] Zadeh L. A. (1965) Fuzzy sets. Information and Control, 8 pp. 338–353.

[6] Dubois D. and H. Prade. (1999) The three semantics of fuzzy sets. Fuzzy Sets and Systems, 90 pp. 141–150.

[7] Zimmerman H.J. (1999) Practical applications of fuzzy technologies. Kluwer Academic Publishers.

[8] Mendel J. M. (2001) Uncertain rule-based fuzzy logic systems, Introduction and new directions. Prentice Hall.

[9] Verstraete J. Parameters to use a fuzzy rulebase approach to remap gridded spatial data. (2013) Proceedings of the 2013 Joint IFSA World Congress NAFIPS Annual Meeting (IFSA/NAFIPS), pp. 1519–1524.

[10] Hoffman C. M. (1992) Geometric and Solid Modeling. Chapter 4. https://www.cs.purdue.edu/homes/cmh/distribution/books/geo.html

[11] Verstraete J. (2015) Dealing with rounding errors in geometry processing. (accepted for publication in the Proceedings of FQAS, October 2015)