CEPS INSTEAD

WORKING PAPERS

An Approach for spatial and temporal data analysis: application for mobility modeling of workers in Luxembourg and its bordering areas

Hichem OMRANI
Omar CHARIF
Olivier KLEIN
Philippe GERBER

Working Paper No 2010-16
July 2010

An Approach for spatial and temporal data analysis: application for mobility modeling of workers in Luxembourg and its bordering areas

Hichem Omrani[†‡], Omar Charif[‡], Olivier Klein[†] and Philippe Gerber[†]
[‡]HEUDIASYC, UTC, CNRS, Centre de Recherche de Royallieu
BP 20529, F-60205 Compiegne, France
[†]GEODE department, CEPS/INSTEAD, Luxembourg
Email: firstname.lastname@ceps.lu

***Abstract*—In this paper, we propose a general visual analytic approach to synthesis very large spatial data and discover interesting knowledge and unknown patterns from complex data based on Origin-Destination (OD) matrices. The research studies of Tobler constitute a good basis in this topic. This paper is interested in the proposal of 2 methods entitled respectively "Weighted Linear Directional Mean: WLDM" and "DS-WLDM". The latter incorporates the Dempster-Shafer theory of evidence with WLDM. Both of the developed methods are an extension of "Linear Directional Mean: LDM" for mobility modeling. With classical techniques such as LDM among others, the results of data mapping are not intelligible and easy to interpret. However with both WLDM and DS-WLDM methods it is easy to discover knowledge without losing a lot of information which is one of the interests of this paper. This proposal is generic and it intends to be applied for data mapping such as for geographical presentation of social and demographic information (e.g. mobility of people, goods and information) according to multiple spatial scales (e.g. locality, district, municipality). It could be applied also in transportation field (e.g. traffic flow). For the application, administrative data is used in order to evaluate spatial and temporal aspects of the daily and the residential mobility of workers in Luxembourg and its bordering areas.**

***Index Terms*—Mobility modeling, data mapping, spatial mobility, geographic knowledge discovery, location uncertainty, daily and residential mobility.**

I. Introduction

The local mobility system in Luxembourg has received little attention in previous research, in spite of the attractiveness of Luxembourg's labor market for workers residing in neighboring countries and the important residential mobility characterizing the Greater Region (Luxembourg and its surrounding regions in France, Belgium and Germany). In recent years, the analyses carried out at CEPS/INSTEAD, especially in the framework of the "MOBILLUX" project (funded by the National Research Fund in Luxembourg) [1], have revealed a certain number of characteristics that are peculiar to daily and residential mobility in Luxembourg and its bordering regions. The most striking of these is the cross-border nature of Luxembourg's employment pool [3], which is sustained by the sizeable differences in wages and in land prices between Luxembourg and its neighboring countries. These differences generate extensive cross-border travels [2], with an estimated 140000 workers crossing daily borders. In this context, this proposal sets out to assess the local mobility in Luxembourg and its neighboring regions.

The goal of this research is to develop a general visual analytic approach to synthesis very large spatial data and discover interesting and unknown patterns. The proposed framework entitled "Weighted Linear Directional Mean vector: WLDM" is an extension of "Linear Directional Mean: LDM" for mobility modeling. With classical techniques such as LDM among others (i.e. for flow mapping see Tobler research work), the results of data mapping (e.g. the flow of mobility) are not intelligible and easy to interpret. However with the WLDM tool, it is much easier to discover knowledge without losing a lot of information which is the interest of this paper.

In this research, the data serving the mobility modeling and the development of the tool is provided by Luxembourg's social security authority (IGSS). It includes the whole active persons (around 300000, during the period 2002-2008) working in Luxembourg and living in Luxembourg and its surrounding regions in France, Belgium and Germany. In this research paper, our theoretical proposal will be used for modeling daily and residential mobility in Luxembourg and its neighboring regions. The developed tool could provide much needed information (indicators with spatial representation) to decision-makers in matters of urban planning and development in the Greater Region. Up to now, the development of such tool has not been common, although it could represent an asset for policy makers. For instance, in the case of our study of the Greater Region, accessibility of working places and homes in Luxembourg could be assessed and potential solutions for the improvement of accessibility could be tested as to their efficiency.

The remainder of the paper is organized as follows. Related work is reviewed in the next section. Section 3 provides an overview of the data and the proposed methodology. Section 4, presents the developed tool and the application with its modified flow map visualization.

II. Mobility modeling: an overview

In existing literature, the modeling of spatial dynamics has been developed under various frameworks, depending on the scientific field. In geography, mathematical models have been proposed, among others, in [11]. In computer science, models ranging from classical artificial intelligence to distributed artificial intelligence have been used. Mobility studies consist to discover new knowledge. However, it is a challenging problem to analysis massive spatial interaction data (e.g. residential and daily mobility), discover useful patterns, and facilitate the decision making process. Individual-level movement data sets are often very large, containing millions of events that involve millions of people and locations [10]. For instance, in Luxembourg, there exit (n=116) municipalities and more than 500 localities, therefore several thousands of potential links to be represented (i.e. $(n(n-1) = 13340$ links) for municipality level and more than 250000 links for locality level).
Few existing methods can deal with data of such a large volume and complexity. Traditionally, movement or interaction data are visualized with flow maps, as used in migration studies; files transfer analysis between net servers or economic analysis of the flow of goods and services [8] [9]. However flow maps is limited to relatively small data sets, e.g. migration among 50 US states [8]. Indeed, a flow map has limited capacity in visualizing very large data sets.

A. *Related work*

There are two types of flow maps, a directed and not directed flow maps. The disadvantage of the cartographic representation by flow map is that it becomes illegible and non clear with a large density of movement. We can meet this type of problem with a matrix of exchange of dimension more 10×10. This limit can be easily exceeded when we study the analysis of mobility on the level of a country. The figure 1 shows clearly the matter of mobility data mapping.

A method suggested to solve this problem is to transform the representations of linear flow into a surface representation. This last representation is made by choosing colors for the places of origin, and others for the destinations. Moreover, the degradation of these colors shows the volume of flow (see fig. 2). In this type of map, only movements are represented and the specifications of dynamic flow are lost.

A series of methods were proposed by [8] to solve the problem of the readability of a flowmap. These methods, consist to reduce the information presented in flow maps:

- By decreasing the surface covered by the studies (e.g. instead of making the study on the greater region, we take only Luxembourg).
- By gathering the individuals: aggregation, passage on a higher level "of address" (scale of larger representation) (e.g. passage from the level of locality to the level of municipality).
- By eliminating flows having tiny density which are lower than a given threshold. ‘
- By removing the orientation: use only of total or net flow. We recall that the net flow between two municipalities (i) and (j) is equal to $(F_{ij} - F_{ji})$ and total flow is equal to $(F_{ij} + F_{ji})$, where F_{ij} is the flow between (i) and (j).

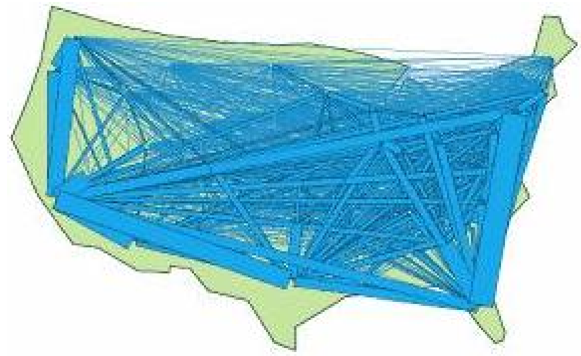

Fig. 1. Example of flow map of an OD matrix: 1128 total two-way migrations (2000 US Census 48 by 48 matrix:source [8])

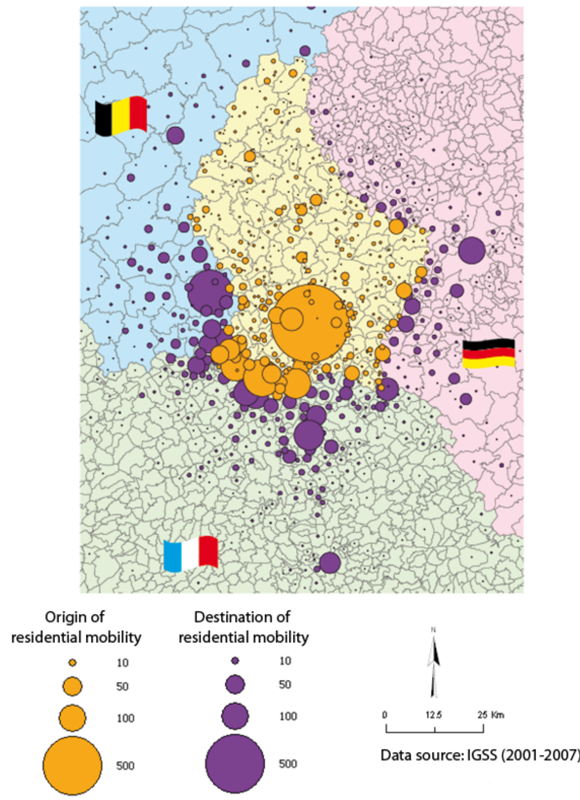

Fig. 2. Flow cartography of residential mobility, (data source: IGSS, 2001-2007)

Tobler [8] proposed also a series of methods founded on the principle of vector computation (isolated vector, regular vector, etc). The isolated vectors (noted by v_i as shown in Eq. 1) are mean vectors which are calculated by using the following equation:

$$v_i = \frac{1}{n-1} \times \sum_{i=1, i \neq j}^{m} \frac{F_{ij} - F_{ji}}{F_{ij} + F_{ji}} \times \frac{1}{d_{ij}} \times [(X_i - X_j), (Y_i - Y_j)] \quad (1)$$

with $d_{ij} = (X_i - X_j)^2 + (Y_i - Y_j)^2$

The method of the isolated vectors proposed by Tobler is inappropriate in our work, since it is based on incoming and outgoing flows. Here, we were interested to study only outgoing flows. Moreover, actually we do not have data to apply this method. In addition to the isolated vectors, Tobler proposed the method of representation by regular vector fields which it determines starting from known points, by using the interpolation, the regular vector with each node of a regular grid covering the surface of study. The method of the vector field allows the modeling of dynamic and continuous movement on all the surface of study. The continuity makes the method of the regular vector fields inapplicable in our case of study. The computation of the vector field at any point of the space of study is distorted because flows exist only on the centroid of the analyzed spatial unit. We treat here the mobility data between spatial units (e.g. municipality) which are not continuous.

B. Context and mobility data

In our case study, the flow of mobility is the number of active peoples moving between spatial units (e.g. locality, region, municipalities. Generally, in literature, the movement of peoples is represented via mapping techniques, for example the famous Tobler mapping techniques. The latter tool can be generalized to model the movement between cantons, districts and even between countries (for example for analyzing migration or import-export trade, etc.).
The data needed to construct the mobility maps (or flow maps) are the spatial coordinates of Origin and Destination (OD) places and the OD matrix of flows between the different places. A flow is represented by a link joining the origin with the destination where the width must be proportional to the number of movements. The alternative of mapping data can be unidirectional (i.e. total or net movement) or bidirectional (simultaneous two-way moves). For unidirectional cartographic flow data, some methods can be used like before mentioned. In particular the use of Tobler method on very large data sets is not very valuable. It has limited capacity in visualizing complex and large data sets and in deducing some conclusions or ideas about the mobility system. Visualizing large and bidirectional flows is promising but still problematic. In this paper, we propose a tool for data mapping both for unidirectional and multidirectional witch is able to treat very large data sets.

III. Proposed approach: Spatial and temporal analysis of the mobility data

We present hereafter a novel approach for mobility mapping based on the Linear Directional Mean (LDM) [12]. The LDM is a spatial and temporal tool able to visualize and discover new knowledge from complex mobility data. The latter is already developed under existing GIS (e.g. ArcGIS). However, it is not appropriate for mobility data and for our application. In fact, for instance, ArcGIS offer a procedure by taking a set of vector in inputs and computes the linear directional mean of these vectors. This kind of representation is useful to show, for instance, the direction mean of wind movement. But, we would like to compute and visualize for each spatial unit a mean vector for studying the direction of the movement of the set of persons inside a given spatial unit. In addition, the GIS software (e.g. ArcGIS) does not take into account the volume of vectors (e.g. number of workers moving from an origin to one destination) in the computation of LDM.
We present hereafter two methods as an extension of LDM proposal. The first method WLDM computes a mean direction taking into account the volume of vectors. The second method entitled DS-WLDM is a further development on the proposed WLDM in order to treat uncertainty and ignorance under the framework of Demspter-Shafer Theory (DS).

A. Method I: Weighted Linear Directional Mean (WLDM)

From the method of the LDM computation, we adapted the LDM method and we propose an extended version. It is entitled Weighted Linear Directional Mean (WLDM) and it is calculated as follows (see Eq. 2).

$$WLDM = atan\frac{\sum_{j=1}^{n} sin(\theta_j) \times F(ij)}{\sum_{j=1}^{n} cos(\theta_j) \times F(ij)} \tag{2}$$

Where atan, sin and cos present respectively the arc-tangent, sinus and cosinus of the angle; $F(ij)$ represent the amount of movement (known by flow) from an origin (o_i) to a given destination (d_j).

1) Algorithm implementation: In order to facilitate subsequent discussion, the main symbols used through the paper and their definitions are summarized in table I. The method of the WLDM is summarized bellow in algorithm 1.

TABLE I
SUMMARY OF SYMBOLS AND DEFINITIONS

Symbols	Definitions
coordinates	The coordinates of points inside the study space. A point is composed of 2 coordinates: X,Y
start_point	The start points of mean vectors
end_point	The end points of mean vectors, to be computed
flow: F(ij)	A table including the volume of flow from the starts points
n	The number of destination points of the OD matrix
l	Length of the mean vector

B. Method II: DS-WLDM

We incorporate here the Dempster-Shafer theory (DS) with the WLDM method. This method is a further development of the WLDM described before when we are facing uncertainty, ignorance and even missing values.

1) Theoritical framework: In this section, we briefly recall the bases of DS theory, first introduced by Dempster [5], and formalized by Shafer [4]. The interest of this theory is that it generalizes the probability theory, by introducing an explicit representation of uncertainty and ignorance. A subjectivist interpretation of Dempster-Shafer theory was proposed by

Algorithm 1: Weighted Linear Directional Mean (WLDM)

Data: coordinates, start_point, F(ij), n, l
var: sumCos, sumSin, angle, mean_angle: Real
Result: end_point: Point
begin
  for $j \leftarrow 1$ **to** n **do**
    angle $\leftarrow$ atan($\frac{|coordonnees(j).Y - start\_point.Y|}{|coordinates(j).X - start\_point.X|}$)
    if *(coordinates(j).X $<$ start_point.X)* **then**
      if *(coordinates(j).Y $>$ start_point.Y)* **then**
        angle $\leftarrow$ 180 - angle
      else
        angle $\leftarrow$ 180 + angle
      end
    else if *(coordinates(j).Y $<$ start_point.Y)* **then**
      angle $\leftarrow$ 360 - angle
    end
    sumCos $\leftarrow$ sumCos + $cos(angle) \times F(ij)$
    sumSin $\leftarrow$ sumSin + $sin(angle) \times F(ij)$
    mean_angle $\leftarrow$ atan ($\frac{sumSin}{sumCoS}$)
  end
  if *(SumCos $<$ 0)* **then**
    mean_angle $\leftarrow$ 180 - mean_angle
  else
    mean_angle $\leftarrow$ 360 - mean_angle
  end
  else
    mean_angle $\leftarrow$ 180 + mean_angle
  end
  end_point.X $\leftarrow l\times$ cos(mean_angle)
  end_point.Y $\leftarrow l\times$ sin(mean_angle)
  return end_point
end

Smets [6], under the name of Transferable Belief Model (TBM). In the TBM theory, there exists a two-level structure composed of a credal level where beliefs are entertained, and a pignistic level where decisions are made. In this section, we only define the concepts that are used in our method. Further details can be found in [6], [7].

a) Frame of Discernment: In Dempster-Shafer theory, a problem is represented by a set Θ of mutually exclusive and exhaustive hypotheses θ_i. Θ is called the frame of discernment such as: $\Theta = (\theta_1, \theta_2, \ldots, \theta_n)$.

b) Basic belief assignment (bba): A Basic Belief Assignment (denoted bba) is a function m from 2^Θ to $[0,1]$ that assigns a value to each conjunction in the frame of discernment:

$$m : 2^\Theta \longrightarrow [0,1] \tag{3}$$

such that:

$$\sum_{A \subseteq \Theta} m(A) = 1 \tag{4}$$

The basic belief mass $m(A)$ represents the measure of the belief that is committed exactly to A, given the available evidence, and that cannot be committed to any strict subset of A because of lack of information. Every $A \subseteq \Theta$ such that $m(A) > 0$ is called a focal proposition. A bba verifying $m(\emptyset) = 0$ is said to be normal. If $m(\emptyset) \neq 0$, $m(\emptyset)$ can be interpreted as the part of belief committed to the assumption that none of the hypotheses in Θ might be true (open-world assumption).

c) Decision making: In the DS theory, when a decision has to be made, the bbas are generally transformed into probabilities (called pignistic probability). Let P_{Bet} be the so-called pignistic probability distribution, defined by uniformly distributing the mass of belief given to each subset of Θ among its elements:

$$P_{Bet}(A) = \sum_{B \subseteq \Theta, A \in B} \frac{m(B)}{|B|}, \forall A \in \Theta \tag{5}$$

Where $|A|$ denotes the cardinality of A

This definition relies on the idea that, in the absence of additional information, $m(B)$ should be equally distributed among the elements of B. This solution is a classical probability measure and it is not reliable in our case study. Indeed, we propose here a weighted probability function (noted by P) as follows:

$$P(A) = \frac{m(A) \times m(\Theta)}{\sum\limits_{B \subset \Theta} m(B)} + m(A) \tag{6}$$

2) Mathematical analysis of the DS-WLDM method: The frame of discernment is: $\Theta = (d_1, d_2, \ldots, d_n, d_{n+1})$ where n is the number of destinations. Each singleton $d_j, (j = 1, \ldots, n{+}1)$ corresponds to a destination of work. The destination d_{n+1} corresponds to the absence of destination (ignorance in the data or missing values). Hereafter, we describe the DS-WLDM proposal. Let Fij be an OD matrix, with m origins $(o_i, i = 1, \ldots, m)$ and $n{+}1$ destinations $(d_j, j = 1, \ldots, n{+}1)$. For each origin (o_i), we construt the bba as follow:

$$\left\{ \begin{array}{lcl} m(j) & = & \dfrac{F(ij)}{\sum\limits_{j=1}^{n+1} F(ij)} \\ m(\Theta) & = & \dfrac{F(in+1)}{\sum\limits_{j=1}^{n+1} F(ij)} \end{array} \right. \tag{7}$$

The decision rule corresponds to the probability, that takes into account a normalization step (as shown in Eq. 6). The ignorance is assigned to one of the n destinations. We adapted the WLDM method proposed in Eq. 2 and we propose in Eq. 8 a Weighted Linear Directional Mean with DS theory(DS-WLDM).

$$\text{DS-WLDM} = atan \frac{\sum\limits_{j=1}^{n} sin(\theta_j) \times P(ij)}{\sum\limits_{j=1}^{n} cos(\theta_j) \times P(ij)} \tag{8}$$

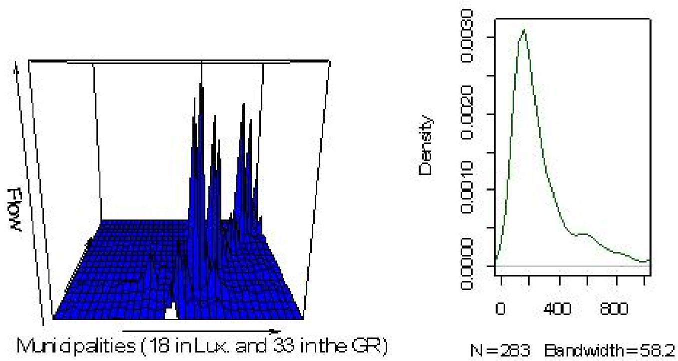

Fig. 3. A set of data related to daily mobility in Luxembourg and its bordering area (or Greater Region (noted here by GR)), density of flow larger than 100, in this case the flow of movement follows normal distribution with mean= 447.735 and standard variation= 1074.069

Where $DS - WLDM$ is the Weighted Linear Directional Mean with DS theory (DS-WLDM); atan, sin and cos present respectively the arc-tangent, sinus and cosinus of the angle and $P(ij)$ represent the probability distribution, as defined in Eq. 6, from an origin (o_i) to a given destination (d_j). The algorithm of DS-WLDM is similar to WLDM, by using Eq. 8 instead of line 3 in algorithm 1.

IV. Application: Daily and Residential Mobility of Workers in Luxembourg and Its Bordering Areas

A. Examining Data

Before analyzing the data of mobility, a step of addresses correction and geocoding is needed. We have proposed an algorithm and a tool for addresses correction, from which, we obtained at individual level corrected addresses (from the IGSS data source) related to work and residential places. After the step of addresses correction, we have generated with the developed tool, OD matrices (Origin-Destination, with two spatial units: locality by locality and municipality by municipality). For the daily mobility, the origin represents the locality or the municipality of residence and the destination represents the locality or the municipality of work. These matrices are used here to model the home to work mobility of actives persons working in Luxembourg and living in its bordering areas. Moreover, by using several data sets, within different periods, we generated OD matrices presenting the residential mobility over a pair of periods where the origin presents the address during the first period and the destination for the second period. We show in the figure below an example of an OD matrix (18×33) which is a sub matrix of the original one (482×117) representing the data of daily mobility in 2008 in Luxembourg (117 municipalities) and its bordering areas (482 in municipalities the Greater Region).

B. Results: spatial analysis and data mapping

We present hereafter below three flow maps. The first map (see fig. 4) presents the vectors of flows resulting from the raw data without treatment. From this first map, it is difficult to draw conclusions. This is why, the interest of the suggested method of the mean vector which consists to summarize the data of daily (home to work) and residential mobility (see figs. 5 and 6). According to the map in fig. 5, it is easy to conclude that Luxembourg is an attractive city, and it has a considerable potential of employment. As a result, active people coming from the Greater Region to work daily, particularly in Luxembourg City, like drawing in fig. 5. The map in fig. 6, underline that the majority of the residents come close to the border of Luxembourg without living inside. This is due to the high prices of housing in Luxembourg.

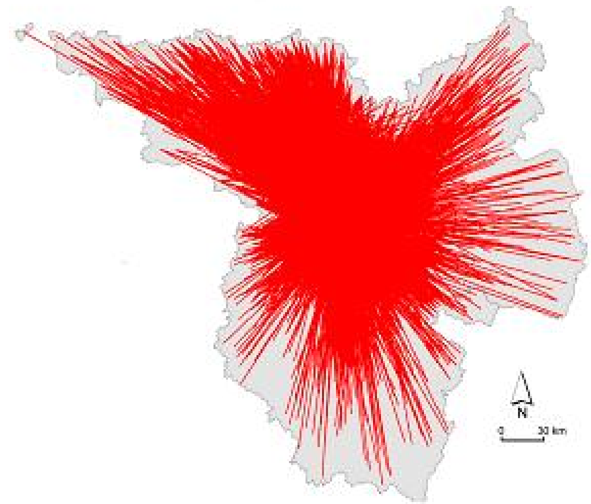

Fig. 4. Flow map movement from daily mobility, before treatment (data source: IGSS, 2008)

V. Conclusion

This paper presents theoretical and technical aspects for the modeling of mobility. At the theoretical level, we proposed 2 methods entitled WLDM and DS-WLDM able to discover new knowledge. At the technical level, we developed a tool, ergonomic and easy to use, to model spatial and temporal dynamics of the daily and residential mobility on the basis of IGSS data sets, using the developed methods. The results of the proposed WLDM method are represented using the techniques of maps. Future work will consist in applying methods of artificial intelligence like the cellular automata and multi agent systems for the dynamic analysis (at microscopic level) of the daily and residential mobility of active people in Luxembourg.

Acknowledgment

This work was supported by GEODE department of CEPS/INSTEAD institute. The authors would like to thank anonymous reviewers for their valuable comments.

References

[1] P. Gerber, *MOBILLUX. Quelles mobilits au Luxembourg et dans son espace transfrontalier? Comprendre les interactions entre les mobilits quotidienne et rsidentielle*. Rapport final pour le FNR, Luxembourg, 120 p, 2009.

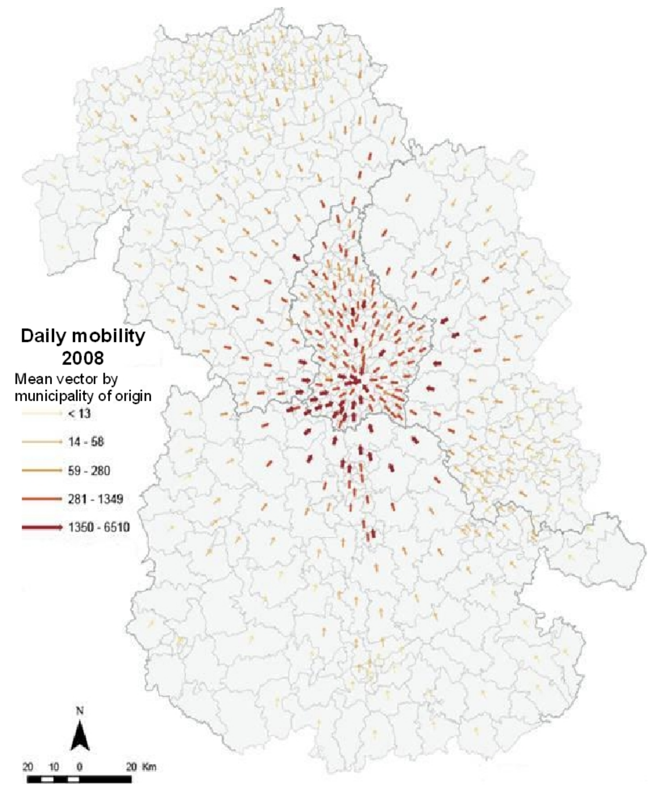

Fig. 5. Flow map of daily mobility after treatment by WLDM (data source: IGSS, 2008)

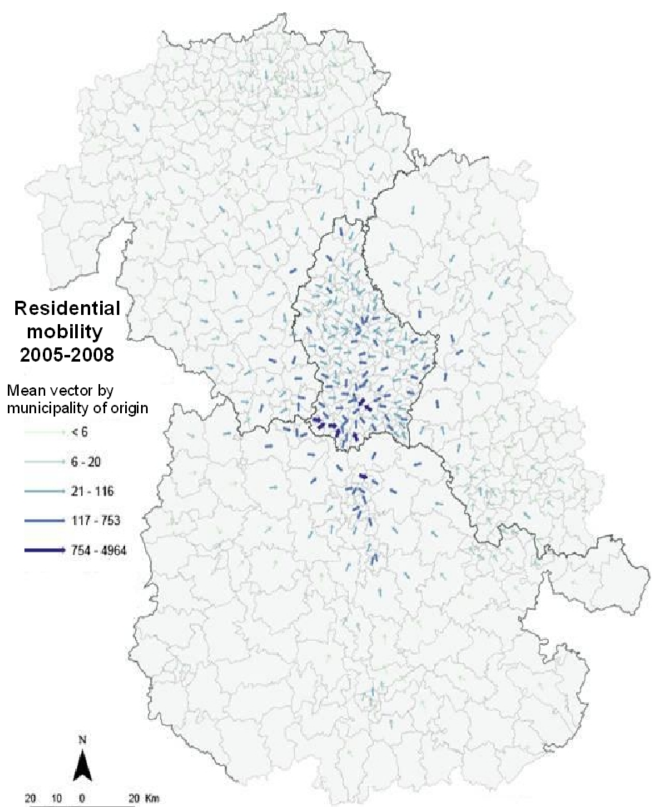

Fig. 6. Flow map of residential mobility from 2005 to 2008, after treatment by WLDM, (data source: IGSS, 2005-2008)

[2] S. Carpentier and P. Gerber, *Les dplacements domicile-travail : en voiture, en train ou pied? Vivre au Luxembourg*, n 53, mars, Luxembourg, 2 p, 2009.

[3] M. Schneider and I. Pigeron-Piroth, *Une aire rsidentielle transfrontalire en expansion*. In Bousch P. Chilla T. Gerber P. et al. : Der Luxembourg Atlas / Atlas du Luxembourg, Ed. Emons, Allemagne, pp. 180-182, 2009.

[4] G. Shafer. *A mathematical theory of evidence*, Princeton University Press, Princeton, N.J., 1976.

[5] A.P. Dempster. *A generalization of bayesian inference*, Journal of the Royal Statistical Society, 30:p.205247, 1968.

[6] Ph. Smets and R. Kennes. *The transferable belief model*, Artificial Intelligence, vol.66:p.191234, 1994.

[7] T. Denoeux, Z. Younes and F. Abdallah. *Representing uncertainty on set-valued variables using belief functions*, Artificial Intelligence, Accepted for publication, 2010.

[8] W. Tobler, *Experiments in migration mapping by computer*, American Cartography, 1987.

[9] T.C. Bailey and AC. Gatrell, *Interactive Spatial Data Analysis*, Addison Wesley Longman, Harlow, Essex, 1995.

[10] D. Guo, *Visual analytics of spatial interaction patterns for pandemic decision support*, International Journal of Geographical Information Science vol. 21, no. 8, 859877, 2007.

[11] R. J. Huggett and R. W. Thomas, *Modeling in Geography*, 1980.

[12] A. Mitchell. ESRI Guide to GIS Analysis, Geographic Patterns and Relationships and Zeroing In: Geographic Information Systems at Work in the Community, Volume 1, 2005.

CEPS
INSTEAD

B.P. 48
L-4501 Differdange
Tél.: +352 58.58.55-801
www.ceps.lu