

Generating French virtual commuting network at municipality level

Maxime Lenormand¹, Sylvie Huet¹ and Floriana Gargiulo²

¹ Cemagref, LISC, 24 avenue des Landais, 63172 AUBIERE, France
(maxime.lenormand, sylvie.huet)@cemagref.fr

² INED, 133 boulevard Davout, 75020 PARIS, France
floriana.gargiulo@gmail.com

Abstract. We aim to generate virtual commuting networks in the French rural regions in order to study the dynamics of their municipalities. Since we have to model small commuting flows between municipalities with a few hundreds or thousands inhabitants, we opt for a stochastic model presented by [Gargiulo et al. \(2012\)](#). It reproduces the various possible complete networks using an iterative process, stochastically choosing a workplace in the region for each commuter living in the municipality of a region. The choice is made considering the job offers in each municipality of the region and the distance to all the possible destinations. This paper presents how to adapt and implement this model to generate French regions commuting networks between municipalities. We address three different questions: How to generate a reliable virtual commuting network for a region highly dependant of other regions for the satisfaction of its resident's demand for employment? What about a convenient deterrence function? How to calibrate the model when detailed data is not available? We answer proposing an extended job search geographical base for commuters living in the municipalities, we compare two different deterrence functions and we show that the parameter is a constant for network linking French municipalities.

1 Introduction

Some rural areas have an increasing population while others continue to suffer from depopulation in Europe ([Johansson and Rauhut, 2007](#); [Champetier, 2000](#)). This is what the European project PRIMA¹ has been trying to understand studying the dynamics of a commuting network of virtual rural municipalities through microsimulation ([Huet et al., 2012](#)). In this framework, as in many studies through microsimulations or agent-based simulations, we need generation models capable of building reliable virtual commuting networks. Indeed, new economic theories assume local positive dynamics can be explained by implicit geographical money transfers made by commuters or retired people (see

¹ PRototypical policy Impacts on Multifunctional Activities in rural municipalities - EU 7th Framework Research Programme; 2008-2011; <https://prima.cemagref.fr/the-project>

for example [Davezies \(2009\)](#)). It is thus necessary to have virtual commuting networks of individuals for the micro simulation approaches increasingly used ([Birkin and Wu, 2012](#); [Parker et al., 2003](#); [Bousquet and Page, 2004](#)) to study the economic and land use dynamics.

We are interested in understanding the dynamics of the French rural municipalities. 95% of them have at most 3000 inhabitants. This means most of the commuting flows we want to study are weak, with a spatial distribution highly decided by chance. It's why we opt for a stochastic model proposed recently by ([Gargiulo et al., 2012](#)). Moreover, we want to consider the commuting network at different dates. Detailed data regarding flows between couples of municipalities are available in France only for 1999. At the other dates, the only reliable data is aggregated data, which for each municipality describes how many people go to work outside of the municipality and how many come from the outside of the municipality to work in without precisions about the various places of work or the various municipalities of residence. Then we also choose the model of [Gargiulo et al. \(2012\)](#) for its ability to generate a population of individuals on a commuting network, starting from this data. This model reproduces the complete network using an iterative process stochastically choosing a workplace in the region for each commuter living in the municipality of the region. The choice is made considering the job offers in each municipality of the region and the distance to all the possible destinations. It differs from the classical generation models presented in [Ortúzar and Willumsen \(2011\)](#) since it is a discrete choice model where the individual decision function is inspired from the gravity law model which is not used usually at an individual level ([Haynes and Fotheringham, 1984](#); [Ortúzar and Willumsen, 2011](#); [Barthélemy, 2011](#)). Moreover, the model ensures that for every municipality the virtual total numbers of commuters coming in and going out are the same as the ones of the data. This paper presents how to adapt and implement this model to generate French regions commuting networks between municipalities. This implementation has forced us to address three different questions: How to generate a reliable virtual commuting network for a region highly dependant of other regions to satisfy the need for jobs of people living in their municipalities? What about a convenient deterrence function? How to calibrate the model when detailed data is not available.

A first problem we have to solve is that our French regions are not islands as for example in [De Montis et al. \(2007, 2010\)](#). Indeed, some of the inhabitants, especially those living close to the border of the region, are likely to work in municipalities located outside the region of residence. This part, especially if it is significant, make the generated network largely false if we only consider that people living in the region also work in the region. A way to solve this problem is to generate the commuting network only for people living and working in the region. However, this means the modeller has to know the quantity and the place of residence of individuals who work outside the region and live in the region. The data giving this detail is very rare, and so is being able to benefit from an expertise that could give this knowledge. Then, we address this issue by extending the job search geographical base for commuters living in the

municipalities to a sufficiently large number of municipalities located outside the region of residence. We compare the model without municipalities from outside (called only outside later in this paper) and the model with outside on 23 French regions and conclude about the quality of our solution.

The second problem relates to the form of the deterrence function which rules the impact of the distance on the choice of the place of work relatively to the quantity of job offers. The initial work done by [Gargiulo et al. \(2012\)](#) proposes to use a power law. However, [Barthélemy \(2011\)](#) says the form of the deterrence function varies a lot, sometimes it can be inspired from an exponential function as in [Balcan et al. \(2009\)](#) or from a power law one as in [Viboud et al. \(2006\)](#). To choose the much more convenient deterrence function, we have compared the quality of generated networks for 34 French regions obtained on the one hand with the exponential law, and with the power law on the other hand. We showed that we obtained better results with the exponential law.

The last, but not the least problem to solve is the one related to calibration. The generation model, as most of the currently used commuting network generation models, has one parameter to calibrate. This parameter rules the impact of the distance in the individual decision regarding its place of work relatively to the quantity of job offers. The only available distance we can use is the Euclidian distance. We have detailed data on the commuting network for the year 1999 which can be used for calibration, but it is not the case for early years or more recent ones. It may be possible to assume the parameter value does not change over time but we know the transportation network can evolve greatly at the local level to reduce the time distance while we can't capture such a change with the Euclidian distance. The solution was finally easy to manage. Using 34 French regions we show that every French region can be generating using a constant value for the parameter. Then, we assume that the parameter value is constant over time and space.

2 Material and methods

2.1 The French regions and the data from the French statistical office

A complete description of the regions from which the network has been generated is provided in the Table 4. These regions have been randomly chosen for their diversity in terms of number of municipalities and commuters, and surface areas. Some correspond to a region while others are closer to the county (called "departement" in French).

The French Statistical Office (called *INSEE*) collects information about where individuals live and where they work. From this collected data, the Maurice Halbwachs Center or the *INSEE* make available for every researcher the following data:

- in 1999, data about the numbers of individuals commuting from location i to location j for every municipality of a region;

- in 1990 and 2006, the total number of commuters, the total job offers and the total number of workers living in for every municipality; these data allows to compute for each municipality the number of commuters coming to work in.

The Lambert coordinates for each municipality are easy to find on internet. They allow to compute the Euclidian distance between every municipality couple.

We start from these data sets for our implementation work of the model presented in the next section.

2.2 The model of Gargiulo et al. (2012)

Consider a region composed by n municipalities. We can model the observed commuting network starting from the matrix: $R \in M_{n \times n}(\mathbb{N})$ where R_{ij} is the number of commuters from the municipality i (in the region) to the municipality j (in the region). This matrix represents the lighth grey origin-destination table presented in Table 1.

The inputs of the algorithm are:

- $D = (d_{ij})_{1 \leq i, j \leq n}$ the Euclidean distance matrix between municipalities
- I_j the number of in-commuters from the region to the municipality j of the region, $1 \leq j \leq n$ (i.e. the number of individuals living in the region in a municipality i ($i \neq j$) and working in the municipality j).
- O_i the number of out-commuters from the municipality i of the region to the region, $1 \leq i \leq n$ (i.e. the number of individuals working in the region in a municipality j ($j \neq i$) and living in the municipality i).

I_k and O_k can be respectively assimilated to the job offers for the employed of the region and the job demand of the employed of the region for the municipality k , $1 \leq k \leq n$. The algorithm starts with:

$$I_j = \sum_{i=1}^n R_{ij} \quad (1)$$

and

$$O_i = \sum_{j=1}^n R_{ij} \quad (2)$$

The purpose of the model is to generate the lighth grey origin-destination subtable of the region described in Table 1. To do this it generates the matrix $S \in M_{n \times n}(\mathbb{N})$ where S_{ij} is the number of commuters from the municipality i (in the region) to the municipality j (in the region). It's important to note that $S_{ij} = 0$ if $i = j$. The algorithm assigns at each individual a place of work with a probability based on the distance from its place of residence to every possible places of work and their corresponding job offer. The number of in-commuters

of the municipality j and the number of out-commuters of the municipality i decreases each time an individuals living in i is assigned the municipality j as workplace. We stop the algorithm when all the out-commuters have a place of work. The algorithm is described in Algorithm 2.1 with $m = n$.

Table 1. Origin-destination table for the region; The light grey table represents the commuters living (place of residence RP) and working (place of work WP) in the region for each municipality of the region; The dark grey line represents the number of out-commuters from a municipality of the region to the region for each municipality of the region (i.e. the row totals of the light grey table); The dark grey column represents the number of in-commuters from the region to a municipality of the region for each municipality of the region (i.e. the column totals of the light grey table).

RP \ WP	M_1	...	M_j	...	M_n	Total
M_1	0	...	R_{1j}	...	R_{1n}	O_1
...
M_i	R_{i1}	...	R_{ij}	...	R_{in}	O_i
...
M_n	R_{n1}	...	R_{nj}	...	0	O_n
Total	I_1	...	I_j	...	I_n	

Algorithm 2.1 commuting generation model

Input : $D \in M_{n \times m}(\mathbb{R})$, $I \in \mathbb{N}^m$, $O \in \mathbb{N}^n$, $\beta \in \mathbb{R}_+$

Output : $S \in M_{n \times m}(\mathbb{N})$

$S_{ij} \leftarrow 0$

while $\sum_{i=1}^n O_i > 0$ **do**

 Simulate $i \sim \mathcal{U}_A$ where $A = \{k | k \in [1, n], O_k \neq 0\}$

 Simulate j from $[1, m]$ with a probability:

$$P_{i \rightarrow j} = \frac{I_j f(d_{ij}, \beta)}{\sum_{k=1}^m I_k f(d_{ik}, \beta)}$$

$S_{ij} \leftarrow S_{ij} + 1$

$I_j \leftarrow I_j - 1$

$O_i \leftarrow O_i - 1$

end while

return S

Gargiulo et al. (2012) uses a deterrence function $f(d_{ij}, \beta)$ with a power law shape:

$$f(d_{ij}, \beta) = d_{ij}^{-\beta} \quad 1 \leq i, j \leq n . \quad (3)$$

3 Statistical tools

This section presents the tools used to calibrate the model and to compare various implementation choices.

3.1 Calibration of the β value.

We used the same method as in [Gargiulo et al. \(2012\)](#) to calibrate the β value. We calibrate β in order to minimize the average Kolmogorov-Smirnov distance between the simulated commuting distance distribution and the one building from the observed data. For the basic model we compute the commuting distance distribution with the commuting distance of the individuals who are commuting from the region to the region. For the model with the outside we compute the commuting distance distribution with the commuting distance of the individuals who are commuting from the region to the region and also to the outside.

As the model of [Gargiulo et al. \(2012\)](#) is stochastic, the final calibration value we consider is the average β value over 10 replications of the generation process.

3.2 An indicator to assess the change.

We need an indicator to compare the simulated commuting network and the observed commuting network. Let $R \in M_{n_1 \times n_2}(\mathbb{N})$ a commuting network when R_{ij} is the number of commuters from the municipality i to the municipality j . Let $S \in M_{n_1 \times n_2}(\mathbb{N})$ another commuting network of the same municipalities. We can compute the number of common commuters between R and S (Eq. 4) and the number of commuters in R (Eq. 5):

$$NCC_{n_1 \times n_2}(S, R) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \min(S_{ij}, R_{ij}) \quad (4)$$

$$NC_{n_1 \times n_2}(R) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} R_{ij} \quad (5)$$

From (Eq. 4) and (Eq. 5) we compute the Sørensen similarity index ([Sørensen, 1948](#)). This index makes sense since it corresponds to the common part of commuters between R and S . Then we call it common part of commuters (*CPC*) (Eq. 6):

$$CPC_{n_1 \times n_2}(S, R) = \frac{2NCC_{n_1 \times n_2}(S, R)}{NC_{n_1 \times n_2}(R) + NC_{n_1 \times n_2}(S)} \quad (6)$$

It has been chosen for its intuitive explanatory power: it's a similarity coefficient which gives the likeness degree between two networks. It is ranging from a value of 0, for which there aren't any commuters flows in common in the two networks, to a value of 1 when all commuters flows are exactly identical in the two networks.

4 Generating French regions commuting network at municipality level

4.1 How to cope with regions which are not island or with the lack of detailed data?

A commuting network is defined by an origin-destination table (light grey table in Table 2). At the regional level, it means that we need to know, for each municipality of residence of the region and for each municipality of employment of the region, the value of the flow of commuters going from one to another. This kind of data is not always provided by the statistical offices and usually the datasets are aggregated: only the total number of out-commuters and in-commuters for each municipality is available for each municipality (dark grey row and column in Table 2). To apply the model and define the commuting network, unless we are on a really isolated region², we should need to find a way to isolate from the total number of in(out)-commuters (dark grey row and column in Table 2) the fraction strictly relating to the region (light grey table in Table 2). This is actually not a simple task.

Moreover, even if we are able to isolate these parts, it remains a problem due to the border effect. Indeed, if we consider only the region, we risk to make an error in the reconstruction of the network of the municipalities close to the border of the region. The higher the proportion of individuals working outside of the region is, the higher the error will be.

To go further, we propose to change the inputs of the algorithm. Instead of only considering the regional municipalities as possible places of work, we also consider an *outside* of the region. The outside represents the surroundings of the studied area. The following part describes how to consider this outside practically.

A new extended to outside job search base. We implement the model, taking or not into account an outside, to generate 23 various French regions (see the 23rd regions in the table 4). Their outside is composed of the set of municipalities of their neighbouring "departements".

We consider the outside of the region composed by $m - n$ municipalities, where n is the number of municipalities of the region. The inputs are the directly available aggregated data at the municipal level:

- $D = (d_{ij})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}}$ the Euclidean distance matrix between the municipalities both in the same region and in the outside
- $(I_j)_{1 \leq j \leq m}$ the total number of in-commuters of the municipality j of the region and outside of it (i.e. the number of individuals working in the municipality j of the region or the outside and living in another municipality).
- $(O_i)_{1 \leq i \leq n}$ the total number of out-commuters of the municipality i of the region only (i.e. the number of individuals living in the municipality i of the region and working in another municipality).

² an island for example, in this case grey rows and columns in Table 2 would not exist

The purpose of the algorithm with introduction of the outside is to generate the origin-destination table (light grey and grey subtable in Table 2). To do this we use the algorithm presented in Algorithm 2.1 to simulate the Table 3. Now, we can easily obtain by difference the Table 2 with the total number of in-commuters $(I_j)_{1 \leq j \leq n}$, the total number of out-commuters $(O_i)_{1 \leq i \leq n}$ and the light grey table of the Table 3.

We obtain a matricial representation of the origin-destination table presented in the light grey and grey subtable in Table 2, the simulated matrix $S \in M_{(n+1) \times (n+1)}(\mathbb{N})$ where S_{ij} is:

- the number of commuters from the municipality i (in the region) to the municipality j (in the region) if $i, j \neq n + 1$;
- the number of commuters from the outside to the municipality j (in the region) if $i = n + 1$ and $j \neq n + 1$;
- the number of commuters from the municipality i to the outside if $i \neq n + 1$ and $j = n + 1$.

Table 2. Origin-destination table; The light grey table represents the commuters living and working in the region for each municipality of the region; The grey column represent the out-commuters living in the region and working outside (Out.) for each municipality of the region; The grey line represents the in-commuters working in the region and living outside (Out.) for each municipality of the region; The dark grey line(column) represents the total number of out(in)-commuters for each municipality of the region.

RP \ WP	M_1	...	M_j	...	M_n	Out.	Total
M_1	0	...	R_{1j}	...	R_{1n}	R_{1out}	O_1
...
M_i	R_{i1}	...	R_{ij}	...	R_{iN}	R_{iout}	O_i
...
M_n	R_{n1}	...	R_{nj}	...	0	R_{nout}	O_n
Out.	R_{out1}	...	R_{outj}	...	R_{outn}		
Total	I_1	...	I_j	...	I_n		

Comparison of the two models: Assessing the impact of the outside.

We assess the impact of the outside by a comparison between the generations of the network for 23 French regions with and without the outside. The generation is made at the municipality scale with a power law deterrence function.

The inputs of the case without-outside are built from detailed data while the inputs for the with-outside case are directly the aggregated data (the total municipal number of in and out-commuters). Both implementations are compared through their *CPC* values for each region. We replicate ten times the generation

Table 3. Origin-destination table from the region to the region and the outside; The light grey table represents the commuters living (place of residence RP) and working (place of work WP) in the region for each municipality of the region; The grey table represents the commuters living (place of residence RP) in the region and working (place of work WP) outside of the region.

RP \ WP	M_1	...	M_j	...	M_n	M_{n+1}	...	M_m
M_1	0	...	R_{1j}	...	R_{1n}	R_{1n+1}	...	R_{1m}
...
M_i	R_{i1}	...	R_{ij}	...	R_{iN}	R_{in+1}	...	R_{im}
...
M_n	R_{n1}	...	R_{nj}	...	0	R_{nn+1}	...	R_{nm}

for each region and compute our indicator on each replicate. In all the presented figures, the indicator is averaging on 10 replications. The variation of the indicator over the replications is very low, 1.02% of the average at most. Consequently, it is not represented on the figures. Fig. 1 presents the common part of commuters $CPC_{n \times n}(S, R)$ between the simulated network S and the observed network R obtained with the regional job search base (square) and obtained with a job search base comprising the region and its outside (triangle). It's important to note that for the implementation without outside $S \in M_{n \times n}(\mathbb{N})$ while for the implementation with outside $S \in M_{(n+1) \times (n+1)}(\mathbb{N})$. In order to compare the two models we just consider the regional network (commuters from the region to the region). Indeed, in the without outside case $NC_{n \times n}(S) = NC_{n \times n}(R)$ but it's not necessarily true for the with outside case.

Fig. 1 shows that the two job search bases give results which are not really different. Thus, introducing the outside solves the problem linked to the lack of detailed data without changing the quality of the resulted simulated network. Indeed we have to keep in mind that the inputs of the with-outside case does not require to have detailed data on the contrary to the without outside case.

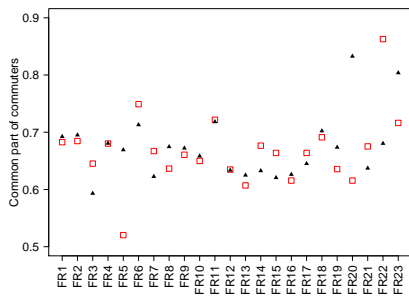


Fig. 1. Average CPC for 23 regions. The squares represent the basic model; The triangles represent the model with outside.

4.2 Choosing a shape for the deterrence function

The second problem relates to the form of the deterrence function which rules the impact of the distance on the choice of the place of work relatively to the quantity of job offers. The initial work done by [Gargiulo et al. \(2012\)](#) proposes to use a power law. However, [Barthélemy \(2011\)](#) says the form of the deterrence function varies a lot, sometimes it can be inspired from an exponential function as in [Balcan et al. \(2009\)](#) or from a power law one as in [Viboud et al. \(2006\)](#). To choose the much more convenient deterrence function, we compare the quality of generated networks for 34 French regions obtained with the model with outside on the one hand with the exponential law, and with the power law on the other hand.

A deterrence function following an exponential law is introduced:

$$f(d_{ij}, \beta) = e^{-\beta d_{ij}} \quad 1 \leq i \leq n \text{ and } 1 \leq j \leq m . \quad (7)$$

To compare the two deterrence functions, we have generated the networks of 34 various French regions (see table 4 for details) replicating ten times for each region. The networks were generated with a job search base for the algorithm considering the outside.

As an example, the Fig. 2 shows we obtained a better estimation of the Auvergne commuting distance distribution with the exponential law.

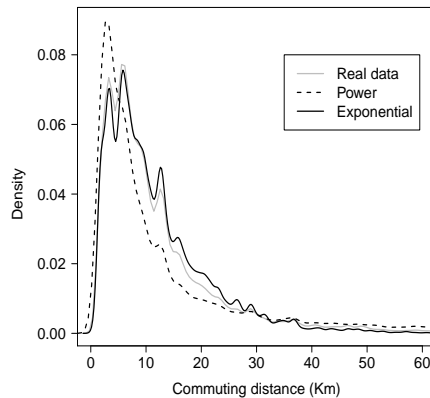


Fig. 2. Density of the Auvergne commuting distance distribution; the solid line represents the observed commuting distance distribution; the dotted line represents the commuting distance distribution obtained with the calibrated model with a job search base comprising the outside and the exponential law; the dashed line represents the commuting distance distribution obtained with a job search base comprising the outside and the power law. The two simulated commuting distance distribution are computed for one replication each.

More systematically, we plot for the two different deterrence functions, exponential law and power law, the average on the replications of the common part of commuters $CPC_{(n+1) \times (n+1)}(S, R)$ in the Fig. 3. It clearly shows that the average proportion of common commuters is always better with the exponential law represented by the squares.

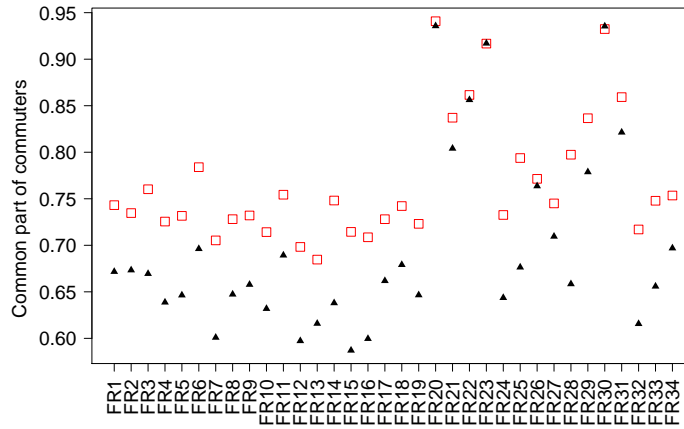


Fig. 3. Average *CPC* for the power shape (triangle) and the exponential shape (square) for 34 french regions.

4.3 Calibrating the model for French regions

The last difficulty to solve is about the calibration process which requires until now to have detailed data to be accurate.

Fig. 4 shows the calibrated β values for each of the 34 French regions. We can see these values weakly vary from about $1.7 \cdot 10^{-4}$ to $2.4 \cdot 10^{-4}$ with an average β valued ($C = 1.94 \cdot 10^{-4}$) corresponding to the dark line.

Then we hypothesize that it is possible to directly calibrate the algorithm to generate the 34 French regions, simply using a constant equal to C . To study the influence of this approximation on the common part of commuters we have computed the *CPC* with C as the parameter value for the 34 regions. We observe in the Fig. 5 that the influence of the β 's approximation on the *CPC* is very weak. We note that the average *CPC* obtained with C is, for some regions, higher than the *CPC* obtained with the not averaging β value. It's possible that the common part of commuters are better with another β value because it's not the calibration criterion.

We don't need to study the influence of the β 's approximation on the calibration criterion. Indeed from the studies made in Gargiulo et al. (2012), we know the *CPC* and the calibration criterion are highly correlated. The *CPC* and the calibration criterion have the same evolution in terms of β . The β value for the minimization of the *Kolmogorov-Smirnov* distance is very close to the one obtained for the maximization of the *CPC* (see the figure 7 in Gargiulo et al. (2012) which perfectly illustrates this relation). Then, as *CPC* values remain quasi identical with $\beta=C$ or with β valued from the calibration process

presented in 3.1, the quality of the approximation of the calibration criterion, i.e. the commuting distance distribution, remains the same.

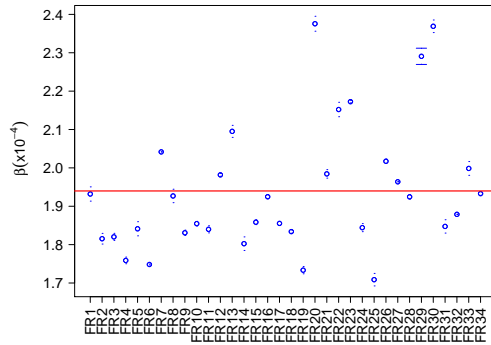


Fig. 4. The circle represents the average calibrated β values for ten replications (The confident interval is composed of the minimum and the maximum) for each regions; the line represents the average β value for the 34 regions.

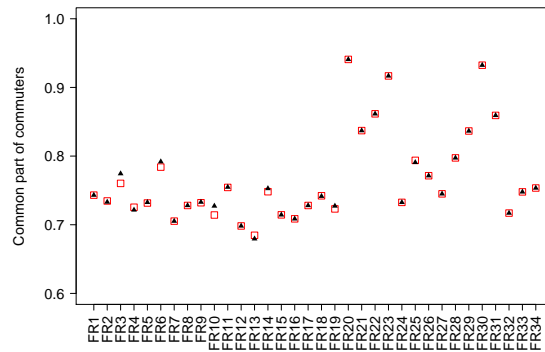


Fig. 5. Common part of commuters for the 34 regions; The squares represents the average *CPC* (10 replications) obtained with the calibrated β value; The triangles represents the average *CPC* (10 replications) obtained with the estimated β values (average β value over the 34 calibrated β values).

5 Discussion and conclusion

To study the rural areas dynamics by microsimulation, we need virtual commuting networks linking individuals living in the municipalities of various French regions. As the studied scale is very low, we have small flows and decided to opt

for a stochastic generation algorithm. The one recently proposed by (Gargiulo et al., 2012) is relevant for our problem. Starting from this model, we implement the commuting network of 34 different French regions. The implementation work leads us to solve three practical problems.

The first problem we have to solve is that our French regions are not islands. Indeed, some of the inhabitants, especially those living close to the border of the region, are likely to work in municipalities located outside the region of residence. However the classical approaches to generate commuting network consider only residents of the region working in the region. That is also the case of ours. The data giving details, or the knowledge, allowing the modeller to suppress people living in the region and working outside is hard to obtain. Then, we address this issue by extending the job search geographical base for commuters living in the municipalities to a sufficiently large number of municipalities located outside the region of residence. We compare the model without municipalities from outside and the model with outside on 23 French regions. We conclude about the relevance of our solution which keeps the value of our quality indicator identical while it does not obliged to know about people who don't work in the region and permit to generate networks only starting from aggregated data.

(Gargiulo et al., 2012) model is based on the gravity law. Then, the second problem relates to the deterrence function which is more a power law or an exponential law depending on the study. Moreover, as the empirical studies comparing the generated networks to "real" data are very rare (Barthélemy, 2011), no one knows about the better shape. To choose the much more convenient one for our French regions, we have compared the quality of generated networks for 34 regions obtained on the one hand with the exponential law, and with the power law on the other hand. We showed that we obtained better results with the exponential law whatever the region is. Indeed our 34 regions vary a lot in surface areas, number of municipalities and number of commuters.

The last problem we solved is the one related to calibration. Applying the model with an extended job search base and an exponential deterrence function, we found a constant equal to $1.94 \cdot 10^{-4}$ is a perfect parameter value to generate commuting network of French administrative regions, whatever they are. However, we didn't test this result for other countries having different types of administrative regions. The robustness of this result to commuting network described at very different scales than the municipality one remain a question we want to address in the future.

References

- Balcan, D., Colizza, V., Gonçalves, B., Hud, H., and Ramasco, J.J. and Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51):21484–21489.
- Barabási, A. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- Barrat, A., Barthélemy, M., Pastor-Satorras, R., and Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747–3752.
- Barrat, A., Barthélemy, M., and Vespignani, A. (2005). The effects of spatial constraints on the evolution of weighted complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, (5):49–68.
- Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499:1–101.
- Birkin, M. and Wu, B. (2012). A review of microsimulation and hybrid agent-based approaches. In Heppenstall, A. J., Crooks, A. T., See, L. M., and Batty, M., editors, *Agent-Based Models of Geographical Systems*, pages 51–68. Springer Netherlands.
- Bousquet, F. and Page, C. L. (2004). Multi-agent simulations and ecosystem management: a review. *Ecological Modelling*, 176(34):313 – 332.
- Champetier, Y. (2000). The (re)population of rural areas. *Leader Magazine*, Spring. Special Issue.
- Davezies, L. (2009). L'économie locale "résidentielle". *Géographie Economie Société*, 11(1):47–53.
- De Montis, A., Barthélemy, M., Chessa, A., and Vespignani, A. (2007). The structure of interurban traffic: A weighted network analysis. *Environment and Planning B: Planning and Design*, 34(5):905–924.
- De Montis, A., Chessa, A., Campagna, M., Caschili, S., and Deplano, G. (2010). Modeling commuting systems through a complex network analysis: A study of the italian islands of sardinia and sicily. *The Journal of Transport and Land Use*, 2(3):39–55.
- Gargiulo, F., Lenormand, M., Huet, S., and Baqueiro Espinosa, O. (2012). Commuting network: going to the bulk. *Journal of Artificial Societies and Social Simulation*, page 13 pages.
- Gitlesen, J. P., Kleppe, G., Thorsen, I., and Ube, J. (2010). An empirically based implementation and evaluation of a hierarchical model for commuting flows. *Geographical Analysis*, 42(3).
- Haynes, K. E. and Fotheringham, A. S. (1984). *Gravity and spatial interaction models*. Sage Publications, Beverly Hills.
- Hensen, M. and Bongaerts, D. (2009). Delimitation and coherence of functional and administrative regions. *Regional Studies*, 1:19–31.
- Huet, S. and Deffuant, G. (2011). Common framework for the microsimulation model in prima project. Technical report, Cemagref LISC.

- Huet, S., Dumoulin, N. and Deffuant, G., Gargiulo, F., Lenormand, M., Baqueiro Espinosa, O., and Terns, S. (2012). Micro-simulation model of municipality network in the auvergne case study. Technical report, PRIMA Project, IRSTEA(Cemagref) LISC.
- Johansson, M. and Rauhut, D. (2007). The spatial effects of demographic trends and migration. Technical report, European Observation Network for Territorial Development and Cohesion (ESPON) - ESPON Project 1.1.4.
- Konjar, M., Lisec, A., and Drobne, S. (2010). Method for delineation of functional regions using data on commuters. Guimares, Portugal. 13th AGILE International Conference on Geographic Information Science.
- Lemercier, C. and Rosental, P.-A. (2008). Les migrations dans le nord de la France au XIXe siècle. In *Nouvelles approches, nouvelles techniques en analyse des réseaux sociaux*, Lille France.
- Ortúzar, J. and Willumsen, L. (2011). *Modeling Transport*. John Wiley and Sons Ltd, New York.
- Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., and Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers*, 93(2):314–337.
- Pastor-Satorras, R. and Vespignani, A. (2004). *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press, New York, NY, USA.
- Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. *Biol. Skr.*, 5:1–34.
- Stillwell, J. and Duke-Williams, O. (2007). Understanding the 2001 UK census migration and commuting data: The effect of small cell adjustment and problems of comparison with 1991. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 170(2):425–445.
- Viboud, C., Bjrnsstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A., and Grenfell, B. T. (2006). Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, 312(5772):447–451.

Table 4. Description of the regions

ID	Region	Number of municip. (region)	Number of municip. (outside)	Region area (km ²)	Average municip. area (km ²)	Number of commuters
FR1	Auvergne	1310	3463	26013	19.86	295776
FR2	Bretagne	1269	1447	27208	21.44	653710
FR3	Ain	419	2809	5762	13.75	162370
FR4	Alsace	903	3081	8280	9.17	440961
FR5	Aquitaine	2296	2835	41309	17.99	700452
FR6	Mayenne	261	3124	5175	19.83	69915
FR7	Lozère	185	1859	5167	27.93	12273
FR8	Poitou-Charente	1464	2467	25810	17.63	375363
FR9	Centre	1842	4718	39151	21.25	624693
FR10	Midi-Pyrénées	3020	3845	45348	15.02	546162
FR11	Limousin	747	3169	16942	22.68	139481
FR12	Franche-Comté	1786	3317	16202	9.07	268399
FR13	Haute-Normandie	1420	3536	12317	8.67	469335
FR14	Haute-Marne	433	3914	6211	14.34	42690
FR15	Vosges	515	3808	5874	11.41	92053
FR16	Lorraine	2339	3067	23547	10.07	547457
FR17	Creuse	260	1814	5565	21.40	23949
FR18	Languedoc- Roussillon	1545	3046	27367	17.71	409116
FR19	Charente-Maritime	1948	1983	25606	13.14	375363
FR20	Haut-de-Seine	36	1245	176	4.89	973173
FR21	Yveline	262	1543	2284	8.72	618741
FR22	Val d'Oise	185	1707	1246	6.74	526600
FR23	Val de Marne	47	1234	245	5.21	642092
FR24	Haut-Rhin	377	2283	3525	9.35	183504
FR25	Tarn et Garonne	195	2338	3718	19.07	41600
FR26	Pyrénées-Atlantique	547	449	4116	7.52	65469
FR27	Alpes-Maritimes	163	353	4299	26.37	163445
FR28	Loire	327	2788	4781	14.62	178828
FR29	Territoire de Belfort	102	2031	609	5.97	45185
FR30	Seine-Saint-Denis	40	783	236	5.90	655200
FR31	Essonne	196	1597	1804	9.20	518321
FR32	Ardennes	463	2588	5229	11.29	59963
FR33	Aube	433	2728	6004	13.87	75561
FR34	Corrèze	286	2088	5857	20.48	49815