



**HAL**  
open science

# Individual based models of social systems: data driven hybrid micro-models of rural development and collective dynamics of filtering or rejecting messages

Sylvie Huet

## ► To cite this version:

Sylvie Huet. Individual based models of social systems: data driven hybrid micro-models of rural development and collective dynamics of filtering or rejecting messages. Other [cs.OH]. Université Blaise Pascal - Clermont-Ferrand II, 2013. English. NNT : 2013CLF22332 . tel-00856248

**HAL Id: tel-00856248**

**<https://theses.hal.science/tel-00856248v1>**

Submitted on 4 Sep 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



N° d'ordre D.U : 2332  
E D S P I C : 604

Université Blaise Pascal – Clermont II  
Ecole Doctorale  
Sciences pour l'Ingénieur de Clermont-Ferrand

Thesis presented by

**Sylvie Huet**

to obtain the Doctor of Philosophy degree,  
Speciality: Computer science

**Individual based models of social systems:  
data driven hybrid micro-models of rural development and  
collective dynamics of filtering or rejecting messages**

Presented on January, the 15th 2013 before the jury:

*Reviewers:* **Nigel Gilbert**, Professor, University of Surrey, United Kingdom  
**Timotéo Carletti**, Professeur, FUNDP, Namur, Belgique

*Examiners:* **Guillaume Deffuant**, Directeur de recherche, IRSTEA, LISC  
Clermont-Ferrand, France  
**Laurent Deroussi**, Maître de conférence, LIMOS  
Université Blaise Pascal, Clermont-Ferrand, France  
**Baptiste Hautdidier**, Ingénieur Agriculture Environnement,  
IRSTEA, Bordeaux, France

*Thesis Director:* **Michel Gourgand**, Professeur, LIMOS  
Université Blaise Pascal, Clermont-Ferrand, France



## Abstract

This thesis is composed in two parts, both dedicated to individual-based modeling of social systems. While the first part is very practical, decision-support oriented, presenting a model which studies the evolution of a rural population, the second part is more theoretical, interested in various mechanisms allowing individual to accept or resist to social influence.

In the first part, we propose an individual-based model of the European rural municipalities and describe its implementation for a French region: the Cantal *département*. We use a new sample-free algorithm for generating the initial population, while classical methods require an initial sample. We design and parameterize the individual activity dynamics with data from the European Labour Force Survey database. The individual dynamics includes an original heuristic for labour statuses and employments changes, based on individual age, profession and activity sector when she is occupied. The last part of the model deals with dynamics that we have not been able to derive from data, mainly the demographic dynamics. Based on the Occam razor principle, we test very simple dynamics and choose them on their capacity to lead to model results close to reference data from the French National Statistical Office. In particular, we propose a simple residential mobility model, partly ruling the emigration, which integrates decision to move and location choice.

In the second part, with a more theoretical approach, we study the collective effects of various mechanisms leading individuals to resist or accept social influence. A first mechanism leads individuals to neglect some features of an object if they are not important enough or incongruent. These individuals exhibit the primacy bias because their attitudes are determined by the first accepted feature. We show that this bias increases when individuals directly exchange about features compared to when they only get the features from the media, in a random order. The second mechanism is a rejection reaction that we suppose occurring because of the discomfort taking place when individuals are close on one dimension of attitude and far on another dimension. The main effect of this rejection mechanism is to lead to a lower number of clusters than with the attraction mechanism alone.

Finally, I discuss the complementarity between the approaches presented in the two parts of this document and try to identify some perspectives based on this complementarity.



## Acknowledgements

First of all, many thanks to Nigel Gilbert and Timotéo Carletti who have accepted to report on this document. How brave they are to have read all these pages and written about them! Merci beaucoup à Baptiste Hautdidier qui a accepté d'être examinateur dans mon jury et m'a fait profiter de son amabilité et de son professionnalisme durant tout le projet PRIMA. Merci à Laurent Deroussi et Michel Gourgand qui m'ont délicatement accompagnée depuis mon stage ingénieur et ont su me permettre de m'exprimer tout en m'invitant à poursuivre.

Un grand merci à Gérard Delance et Clarisse Lemaire pour m'avoir aidé à préparer ma soutenance. Merci à Nicolas Dumoulin pour notre collaboration dans PRIMA, pour m'avoir supportée et tout de même filmée pour la postérité. D'aucuns auraient opté pour la dépression. Merci à Max et Flo pour tous ces bons moments passés ensemble. Merci beaucoup aussi à tous mes petits collègues du LISC, partis ou encore là et ayant eu à m'apprécier, particulièrement ces derniers mois. Heureusement qu'il y a le « respect de l'âge ».

Nous avons grandi ensemble et vous m'avez fait grandir. Que de tendres souvenirs, de rires et d'émotion en pensant à Michel Penel, Benoît Lesaffre, ..., à tous les collègues du drainage du préfa au bord du ruisseau tout au fond du parc. On a détruit ce bidonville mais il est encore bien vivant en moi. Que de chances j'ai eu de ces amitiés. Michel, je m'en souviendrai longtemps de ce déplacement en avion.

Tendres souvenirs loufoques et incroyables à Philippe Huet. Merci à Gérard Jacquin pour m'avoir offer le LISC. Merci au Cemagref, à tous les gens de cet institut qui a su accompagner mon évolution et au sein du quel j'ai fait presque toute ma carrière professionnelle.

Il est là depuis 15 ans. Il s'est inscrit dans cette merveilleuse lignée de ceux qui m'ont tout rendu possible. Notre complicité n'a d'égal que l'agacement réciproque qu'elle peut provoquer. Merci Défi de ne m'avoir jamais arrêté, de toujours t'être adapté et d'avoir mis à tout cela ton infinie patience qui participe tant de notre amitié.

Un gros sourire de bonheur pour mes quatre ovules ayant accepté de se coupler. Je n'ai pas toujours été facile bien que la meilleure mère du monde. Mais « bon dieu » que de thèses et de romans vous pourriez m'inspirer.

Merci aux hommes de ma vie, aux sentiments mêlés et persistants, à leur piètres compétences en bricolage et en cuisine. C'est sans doute un faible prix à payer pour en avoir bénéficié. Encore merci à mon incroyable maman.

Mille souvenirs à Jojo et à Serge.

Merci infiniment enfin à tous ceux qui ont cherché leur nom et ne l'ont pas trouvé alors qu'ils auraient dû ! Le doctorat ne fait pas l'homme, il ne le rend qu'oublié et imbu de lui-même. Alors imaginez la femme....



## Contents

Overview.....	9
Data driven integrated modelling of social dynamics.....	11
Theoretical individual based models of social influence .....	16
Part 1. Data driven integrated modelling of social dynamics .....	23
Chapter 1.1 The SimMunicipality model.....	25
Main entities, state variables and scales .....	27
Process overview and scheduling .....	30
Chapter 1.2 Generating the initial population .....	41
Materials and Methods .....	45
Results .....	53
Conclusion .....	55
Chapter 1.3 Parametrisation of the individual activity dynamics.....	59
Designing and parameterising the individual activity.....	61
Lessons / Experience .....	75
Acknowledgements.....	77
References.....	78
Chapter 1.4 Parametrisation of the unknown laws of demography .....	81
The Cantal and its demography .....	84
How to model couple and birth dynamics in Cantal.....	88
How to model moving.....	99
Conclusion .....	122
Acknowledgements.....	128
Part 2. Theoretical individual based models of social influence.....	131
Chapter 2.1 Disregarding information – a model exhibiting the primacy bias.....	133
The models .....	136
Diffusion of a two-feature object.....	137
Diffusion of more than two features .....	141
Discussion.....	148
Chapter 2.2 Disregarding information – a double modelling approach .....	153
The individual based model (IBM) .....	156
Bird's-eye view of the IBM for complex "interaction" cases .....	162
The aggregated model helping to predict the IBM.....	166
Conclusion, discussion.....	171
References.....	173



Chapter 2.3 Attraction-Rejection – designed from theories.....	175
1 Introduction.....	177
2 Overview of the model.....	180
3 Analysis of several examples.....	183
4 Systematic analysis of the number of clusters.....	187
Discussion and conclusion.....	191
Acknowledgement.....	193
References.....	193
Chapter 2.4 Attraction-rejection – a double modelling approach.....	197
Number of clusters in the ABM when the uncertainty varies .....	200
Aggregate model at the limit of infinite population .....	202
Discussion and conclusion.....	205
References.....	207
Chapter 2.5 Attraction-Rejection – designed from experiments.....	209
1 Introduction.....	211
2 The Dynamical Model of Interacting Individuals .....	213
3 Typical Evolutions of the population.....	214
4 Systematic experiments .....	220
5 Discussion and conclusion.....	225
References.....	227
Conclusions and perspectives .....	229
A methodological point of view on data and design .....	229
Data and design in data-driven modelling .....	229
Data and design in theoretical modelling .....	231
Perspectives.....	233
Developing the loop between data-driven and theoretical modelling .....	233
Social dynamics .....	235
References.....	239

## Overview

This document gathers a selection of my research contributions in the field of modelling and simulating social dynamics. In this introduction, in addition to an overview of these contributions, I would like to provide the reader with some information about the context of this research and to explain how the different contributions are connected. I group them into two parts. The first part is very practical, data driven and decision-support oriented, presenting a model which studies the evolution of a rural population. The second part is more theoretical, interested in various mechanisms allowing individual to accept or resist social influence. This document ends up with a conclusion organised from a methodological point of view on data and design and some perspectives.

The two parts of this document can be unified by the chosen way to model social dynamics: the individual-based approach. A part of my education is in social psychology and the other one is in computer science. Thus this naturally leads me to develop individual-based models in order to try to understand how individuals, groups, institutions and societies influence each other. The idea is far from being new. Since a long time, modellers have been using models for supporting decisions about economic and social issues. But the level of complexity and the lack of data constrained modellers to aggregated approaches, considering the evolution of low spatial scales and large object such as nations and institutions. That is why in 1957, Guy H. Orcutt outlined “the limited predictive usefulness of models which have little to say about such fundamental things as the size and location of the population of individuals, of households, or of firms” on the one hand and on the other hand “the severe difficulties of testing hypotheses and of estimating relations by use of highly aggregative time series”. He is the father of the microsimulation which aims at generating the evolution of the attributes of each individual of a population. In microsimulation, a baseline population is built from data or estimates and updated at a point in a following time. But it has been mainly a static approach in which each individual is independent from the others. However, in the seventies, the computing power started to be sufficient for simulating the dynamics of individuals.

Today micro modeling (Gilbert and Troitzsch 2005) appears particularly relevant for studying the evolution of areas involving heterogenous objects. It includes three different approaches: cellular automata, agent-based models and microsimulation. The two first modelling tools consider individuals interacting with each other and with their environment. This is why they are generally used to study social dynamics. Cellular automata is a dynamic discrete system and can be defined as a lattice of discrete variables or “cells” that can exist in different states. The evolution of each cell’s state depends on the states of its neighbouring cells. Agent-based models simulate the simultaneous operations and interactions of multiple agents. Individual agents are typically characterized as boundedly rational, presumed to be acting in what they perceive as their own interests, such as reproduction, economic benefit, or social status, using heuristics or simple decision-support rules. Most agent-based models are composed of: numerous agents specified at various scales; decision-support heuristics; learning rules or adaptive processes; an interaction topology; and a non-agent environment. A common practical aim of microsimulation is to help anticipate the impact of a policy. It remains static until the end of nineties. The dynamic microsimulation (O'Donoghue 2001) can be considered as a technique where entities change

their characteristics as a result of endogenous factors within the model. Few dynamic spatial microsimulation models have been more recently proposed (Birkin and Wu 2012) and the research domain has some difficulties to progress due to the lack of documentations and papers (Li and O'Donoghue 2012). The specificity of the microsimulation remains that it is developed with high estimation and validation ambitions, close to observables that facilitate empirical tests. However, their strong requirement for data can also be seen as a limitation.

In this document, we gather our selected papers in two groups in relation with these three modelling tools and their current uses. They are:

- data driven models of concrete social case studies, ultimately aiming at supporting policy decisions;
- theoretical individual based models investigating collective effects of a specific psycho-social hypothesis at individual level.

The first group of models aims at providing an insight on the impacts of changes in environmental, economic, and/or policy conditions on the evolution of a given population. Their purpose is to support decisions through retrospective and prospective analyses of events. The events and changes that have to be tested are organized through scenarios. To anticipate the possible impacts, the model has to include every object and process possibly impacted by the tested changes, and the interactions between these processes and changes. However, these processes and objects are generally only partially known. Indeed, the number of dimensions of the problem is generally very high. To support their decisions, policy makers generally try to gather data and expertise. They organize specific collections of data, complementing the National Censuses. Using data is a key aspect of the considered modelling approach that requires designing the model for taking the best advantage of available data.

Cellular automata and microsimulation are the classical tools for this modelling approach. Many examples of microsimulation for decision-support about socio-economical and/or demographical questions can be found in ((INSEE 1999; O'Donoghue 2001; Waddell, Borning et al. 2003; Holme, Holme et al. 2004; Morand, Toulemon et al. 2010; Turci, Bringé et al. 2010; Li and O'Donoghue 2012). Similarly, numerous are the examples using cellular automata (Verburg, Soepboer et al. 2002; Moeckel, Spiekermann et al. 2003; Rindfuss, Walsh et al. 2004; Verburg, Schot et al. 2004; Ballas, Clarke et al. 2005; Ballas, Clarke et al. 2006; Brown, Aspinall et al. 2006; Verburg, Schulp et al. 2006; Ballas, Clarke et al. 2007; Coulombel 2010). More recently, agent-based models were also used for this modelling approach (Grimm 1999; Deffuant and al. 2001; Deffuant, Huet et al. 2002; Parker, Manson et al. 2003; Bousquet and Le Page 2004; Deffuant, Huet et al. 2005; Brown, Aspinall et al. 2006; Matthews, Gilbert et al. 2007; Deffuant, Skerrat et al. 2008; Fontaine and Rounsevell 2009).

The list of topics studied is very large: from the evaluation of pensions, grants and loans for students, various declinations of the social security system to the viability of irrigated systems, various problems linked to land use and land cover change, urban development, transportation network, impact of various policies on agriculture and farmers...

The models of the second group are more theoretical and they are generally used for a better understanding of the link between basic entities' dynamics and observed collective properties. The most classical example of such an approach is a cellular automata studied by (Schelling 1971). He considered virtual individuals living in various parcels of the space represented by a grid. These individuals of different colours move when the part of people living around them having the same colours as theirs is not sufficient. The threshold of intolerance is a parameter of the model. Schelling showed that, even for small intolerance leading individuals to accept a majority of different-colour neighbours, the space becomes segregated. Hence the space segregation cannot be surely explained by the individual intolerance! As we can see in this example, the studied individual dynamics is well identified and simple. Its design is constrained by the Occam's razor: "Plurality must never be posited without necessity". Indeed, as the purpose is to know what a particular dynamics is susceptible to explain, it is necessary to isolate and control it as much as possible. It leads implicitly to consider other processes as non-significant for the studied question. For example, in the Schelling model, it is assumed the difference in average wages of people having different colours and the variations over the space of the lodging price are non-significant. In other words the modeller assumes to be in a situation where every other possible process having an impact is under control. The dynamics defined, the results consist in the collection of reached states and trajectories (not always) explained by this dynamics. What is observed is generally defined through the question the modeller is interested in and/or the literature already existing on the issue.

The classical tool for this modelling approach is the cellular automata and the agent-based models. A review of this type of modelling for social sciences can be found in (Castellano, Fortunato et al. 2009). Classical topics of this approach are opinion dynamics, cultural and language dynamics, crowd behaviour, hierarchy formation, individual dynamics synchronization, social spreading as innovation diffusion with the seminal studies of (Valente 1995) for example.

## Data driven integrated modelling of social dynamics

My research on data driven models of social dynamics has been taking place in the context of European projects. It starts in 2000 with my participation to the European project IMAGES (1997 – 2001). This project aimed at developing a decision-support model for defining agricultural-environmental measures proposed by the European Commission to favour agricultural environmental practices. This model targeted in a first step a better understanding of the dynamics of diffusion of these agri-environmental measures among farmers. As it was an individual based model, various dynamics had to be identified and modelled and an initial population had to be generated consistently with the chosen dynamics. For example, for the particular case of the diffusion of the organic farming measure in Allier, we faced the following questions: how and when had the information about the measure been diffused to farmers? How do farmers decide? How many organic farmers are there initially in the population? How to generate an initial population convenient to study the diffusion?

The project made some progresses in answering these questions, but we also identified shortcomings motivating new directions of research that were put forward in the European project PRIMA project (2008 – 2011). The aim of this project was evaluating how rural municipalities evolve under European policies for agriculture and rural areas, and I have been strongly involved in the cycles of elaboration, implementation and tests of its model. I was in charge of the day to day management of the LISC team involved in this project (comprising a PhD student, post-doctoral student and an engineer) under the supervision of G. Deffuant for the main orientations. Moreover, in the modelling process, I identified how the different data sources could be best used for our modelling purposes. The part 1 of this document is dedicated to my work within this project, and more particularly the implementation of the conceptual model for the rural municipalities of the French *département* called Cantal.

An important shortcoming in the IMAGES model was that the farm population was not evolving. Indeed, in this model the farms include precise economic parameters (Deffuant, Amblard et al. 2000; Deffuant and al. 2001; Deffuant, Huet et al. 2002) but the farms and the population of farmers remain the same during the 10 years of the simulation. However, during this period, a lot of farmers retire while much less new farmers settle. The average size of farms increases because the abandoned farms are bought by other farmers. Yet, the size of the farms is an important parameter to determine the economic impact of adoption of organic farming: the larger the farm; the better the capacity of the farmer to feed his livestock without external intakes, the higher is the benefit of the conversion. This shortcoming points out the need for a dynamic population ageing and changing labour status. That is why, when we elaborated the PRIMA project, we considered this question of the evolution of the population as central.

Modelling the evolution of countryside populations turned out to be very rich and interesting. While for decades the countryside in many regions of Europe was synonymous with inevitable decline, nowadays, some areas experience a “rebirth, even in areas where, until recently, development was not considered possible” (Champetier 2000). A recent EPSON (European Observation Network for Territorial Development and Cohesion) project report (Johansson and Rauhut 2007), concludes that “since the 1970s a global process of counter-urbanization has become increasingly manifest”. However, this general rebirth of the countryside hides deep heterogeneities. It can be observed in the Cantal “*département*” in France where some municipalities are increasing and others are decreasing.

Part 1 of this document presents the main steps of this modelling work. I organise them in a chronological order corresponding to the tasks for designing and parameterising the model. Only the three first chapters have been published: the conceptual model (chapter 1.1) has been approved by the European Commission as a deliverable; it has also been a part of chapter of a book (the other part being chapter 1.3); chapter 1.2 has been published in PLOS One. The chapters have been written at different stages of the model elaboration, and some minor parts are different from one chapter to the other. That is especially the case of chapter 1.3 in which the heuristic to search for a job or a residence is deprecated (some comments have been added to the paper to point out the deprecated methods). The other chapters are up to date.

Chapter 1.1 presents the conceptual model we have designed (Huet, Dumoulin et al. 2012) as a basis for implementations of various case studies in Europe. We follow prescriptions of recent reviews dedicated to land use and land cover change modelling which recommend hybrid approaches (Boman and Holm 2005; Birkin and Clarke 2011; Birkin and Wu 2012), and more particularly coupling microsimulation and agent-based modelling. This choice allows us to include some individual dynamics, poorly known and about which no direct data are available, such as the residential location decision (Coulombel 2010) and also to derive other dynamics from data when it is possible.

Our conceptual model considers individuals, members of households located in municipalities of a region, and their state transitions expressing demographic and activity events: birth, finding a partner, moving, changing job, quitting their partner, retiring, dying ... The municipalities include offers for jobs and dwellings which constrain the possible state transitions. Because we are interested in understanding better the dynamics leading to the development or, on the contrary, to the decline and possible disappearance of municipalities and settlements, two sets of cruxes can be identified in the model: the individual dynamics which determine the needs for residence and jobs; the dwelling and the job offers dynamics at the local (i.e. municipality) level. The municipality offers for jobs and dwelling can be parameterised with usually available data. The individual dynamics however is much more difficult to define and parameterise.

Thus, chapters 1.2 to 1.4 are dedicated to the parameterisation of individuals of this conceptual model for simulating the evolution of municipalities from the French *département* of the Cantal (150 000 inhabitants in 260 municipalities on 5726 km<sup>2</sup>). In practice, the purpose is to find out some submodels that correctly describe the evolution of the chosen objects. The design and the choice of relevant submodels is data driven, and if the link to data is straightforward in the basic microsimulation, it is not so easily manageable with individual based approach. Indeed in the dynamic microsimulation (which remains rare (Birkin and Wu 2012)), the most common way to introduce change into the demographic structure is to apply static ageing techniques consisting in reweighting the age class according to external information. Such approaches avoid considering functions of evolution of the behaviour of the individual and their parameterisation. The multi-agent modelling, (Berger and Schreinemachers 2006) holds the promise of providing an enhanced collaborative framework in which experimental designers, modellers, and stakeholders may learn and interact, but the fulfilment of this promise, depends on the model empirical parameterization. Although multi-agent models have been widely applied in experimental and hypothetical settings, only few studies have strong linkages to empirical data (Fernandez, Brown et al. 2005) and the literature on methods of empirical parameterisation is still limited.

Chapter 1.2 focuses on the problem of the initialisation of the population, which is to be solved as a first step for every individual-based model. In theoretical studies, the population can be drawn from an arbitrary distribution. It cannot be the case for a model aiming at reproducing the evolution of a particular population. We need building a population as close as possible from the data of reference using a set of indicators chosen for their relevance with the general purposes of the model. In case of a human population, the reference data generally comes from censuses. A particular problem has to be solved by a model

considering households and individuals. Indeed, some decisions, such as the residential move, concern households while other processes, as ageing or labour status, are specific to the individual only. In this case, the problem is hence building an initial virtual population fitting simultaneously reference data about households and individuals.

The classical method consists in starting from a sample of households, subset of the population for which all the attributes of each household and its individuals are known, and associate weights to each of these households in order to get the best fitting with available regional statistics. This is done using the classical Iterative Proportional Fitting process (Deming and Stephan 1940). However, this is not possible when no initial sample is available, which was our case. We propose a method starting from aggregated data and creating on the one hand the right number of individuals with their own properties and on the other hand the right number of households with their adequate size. Then, a heuristic allows filling the households with the created individuals while respecting the constraints given by available data about the relationship between individuals and households. Chapter 1.2 describes in more details the method and evaluate its efficiency. A more recent work, comparing an improvement of the classical sample-based IPF method with our sample-free method, shows this latter tends to be slightly better (Lenormand and Deffuant 2012).

Chapter 1.2 presents only a part of the initialisation of the population; it left out the initialisation of the individual labour status and place of work. The initialisation of the place of work is based on a new algorithm modelling the commuting. Several papers are dedicated to this work: one presenting the algorithm and several use cases (Gargiulo, Lenormand et al. 2011), an improvement of the algorithm solving the problem of closure of the system and making the algorithm universal (Lenormand, Huet et al. 2012b), and a comparison with other universal algorithms (Lenormand, Huet et al. 2012a). We did not select these papers in the present document because they are devoted to the initialisation of the model whereas we preferred to focus more this document on the dynamic modelling.

Chapter 1.3 focuses on the design and the parameterisation of the individual dynamics in the labour market. After a first step in which we collected various possible sources, we chose the European Labour Force Survey (LFS) and the National Censuses as our main data sources. They avoid making a lot of assumptions because a large part of their variables have the same definitions in both surveys. They contain data on age and situation (student, retired, actives, occupied or not, inactive)... allowing us to make a connection between the two sources of data. Moreover, they are “official” data sources which are regularly used by policy-makers and stakeholders. Hence, their variables correspond to the common knowledge they have about the social system. This makes the communication around the model easier and clearer.

We consider the basic classical statuses: student, unemployed, employed, inactive and retired. Moreover, we give the individual attributes describing her job: the socio-professional category (SPC) and the activity sector. In France, the socio-professional category (SPC) is available in the LFS and the French Censuses. The job offers of every municipality use also the SPC as a description of the job. The activity sector completes the description of the jobs. For example in France, we consider 24 different possible jobs (6 SPC in 4 activity sectors). The Labour Force Survey, particularly the Employment Survey which is the French

declination of this survey, allows extracting the probabilities of transition between this various statuses. The European Labour Force Survey (LFS) is a continuous survey following the state of individuals over several years (3 years and more recently 18 months) during which they are interviewed several times. It is based on a very large representative sample and gives the weights for projecting it at different scales. We use these weights for extracting data for municipalities less than 50000 inhabitants which is more relevant in our study of rural areas.

We also extract from this survey the probability of the first profession of a young individual depending on the profession of the father. Depending on this first profession, we then extract the age distribution for entering the labour market.

Chapter 1.4 presents the parameterisation of the demographical dynamics. They are related to the formation and the disruption of couples, to the birthrate and the residential mobility sometimes leading to out-migration. We don't have enough data at the Cantal level for using them to directly extract dynamics. Therefore we have to design them, and link them with the dynamics defined from data. For example, regarding a "giving birth" process, we have to decide if an increase of births is due to the increase of the number of births per individual or to a structural change of the population (more people in age and condition to have children). We made hypotheses allowing us to distinguish between these two possibilities. Then, for each case, we check if the number of births given by the data of reference is a possible result of the model. From this first phase, we conclude we need increasing the number of births per individual. We also perform an analysis of the variance of the number of births in order to identify the parameters having the biggest impact on the possibility for the model to be compatible with the reference data. More generally, the method is as follows. We assume different hypotheses in the dynamics and study their capacity to produce results close to data of reference chosen as they are directly impacted by the tested dynamics (for example the number of births for testing the "giving birth" process). In practice, we check that data of reference is a possible result of the model considering a large set of different possible values of the parameters of all the dynamics. In a second step, we study the sensitivity of other indicators to chosen dynamics considering only a subset of them. Indeed, this last study is restricting to selected parameters (and implicitly their related dynamics) on the basis of an analysis of variance. Applying this approach to the elaboration of the demographic processes, we conclude that:

- The implementation for Cantal requires an increasing number of births by individual to reproduce the number of births.
- A two-step dynamics should be considered for couple formation to reproduce the migratory and natural balances in Cantal: first, the annual decision of an individual searching for a partner; second, the searching strategy in terms of effort to meet a convenient partner. In practice, to fit the indicators of reference for the Cantal, a single should be limited in her motivation to search for a partner (i.e. the probability to search) at the same time the level of effort produced when she has decided to search one year has to be restricted (i.e. the maximum annual number of trials to meet someone convenient for her). If these conditions are not respected, couples, and then children, are too numerous.



- A constant probability for couples to split appears sufficient, to match natural balances and migratory balances in Cantal.
- A dynamics based on a limited spatial search for partner and dwelling (i.e. research at a maximum distance) and a probabilistic avoidance of the largest municipalities as a possible place of residence is necessary to reproduce the spatial characteristics of the evolution of the population.

In addition, in chapter 1.4, we collect some information about the relevant segments of value for each parameter. At the same time we identify the indicators which can probably be correctly reproduced by the model and those which cannot.

The work on this model is still in progress. A lot of investigations are still needed for a better understanding of the impact of basic dynamics and their interactions. The problem of modelling the evolution of the population, coming from the IMAGES project, led us to many others. Actually, the IMAGES project can draw a link between the two parts of this document, because the research presented in the second part can also be seen as initiated with questions and problems raised by this project.

## **Theoretical individual based models of social influence**

During the IMAGES project, I developed a submodel computing the economic impact of organic farming adoption depending on the type of farms (Huet and Deffuant 2001; Deffuant, Huet et al. 2002; Lenormand, Huet et al. 2012a). This model, validated by the organic farming technician of the Allier Agricultural department showed that a large part of Allier farmers got an economic benefit if adopting the organic farming measure. However, very few farmers have adopted the measure, apparently because of an important cultural resistance and/or lack of information. This led me to study the dynamics of filtering or rejecting messages that could be the cause of the low adoption level. A theoretical approach appeared more relevant to better understand these mechanisms. Indeed while the decision-support models are generally applied, driven by data describing the problem, a theoretical approach can focus on the impact of a particular dynamics without being constrained by the data. It can be done with a model coupling several processes as we did in (Deffuant, Huet et al. 2005), for the adoption of agro-environmental measures by farmers on various prototypical case studies, with stylised farmers and stylised measures. It is more often done using very simple models involving only one or two dynamics as I did with Margaret Edwards on a binary decision model (Edwards, Huet et al. 2003; Edwards, Ferrand et al. 2005). In this approach, the dynamics is simple enough to explore extensively model trajectories.

The chapters of part 2 are devoted to the questions of filtering information and/or cultural resistance studied through a theoretical modelling approach. Four of them have been published in scientific journals, one has been presented in a conference and the last one is a discussion we wrote recently and which has not been published.

Chapter 2.1 proposes a simple model of information filtering. The model considers an object defined by a set of features, each feature being characterised by utility (or attitude,

supposed shared by all individuals). The model supposes that individuals tend to ignore the features of an object which are not important enough or which contradict their current view. A feature which has not been ignored is saved in memory, and changes the individual's global attitude towards the object. The global attitude is the sum of an a priori attitude toward the object and the attitudes towards the saved features.

The model assumes that an individual has filters which select only important features. The importance of a feature is assessed by comparing the absolute value of attitude towards the feature with a threshold. When the attitude towards the feature and the global attitude towards the object are of the same sign (congruent feature), the threshold is smaller than when the signs are different (incongruent feature).

Our individuals are all in contact with a media, communicating randomly chosen features of an object. An individual can hear about a feature from the media or from a peer with which she regularly discusses. An individual only communicates about her known congruent features because she is reluctant to talk about her incongruent features. As we know little about the incongruence threshold, we decided to consider two variants (Deffuant and Huet 2007):

- A constant threshold which is supposed to be an attribute of the individual expressing when someone considers something as important; it is called the Constant Incongruence Threshold model (CIT);
- A dynamic threshold which depends on the current global attitude value; it is called the Dynamic Incongruence Threshold model (DIT).

We firstly compared our model with the case of individuals informed by the media but not discussing and exchanging features between them. The media diffuses in a random order the features of the object. In our study, we consider a neutral object, meaning that the sum of all the feature attitudes is zero. It comprises:

- A set of negative features with an absolute attitude higher than the incongruence threshold of the individual –called major features;
- A set of positive features with an absolute attitude comprised between the congruence and the incongruence threshold –called minor features.

All the individuals have an initial attitude valued 0 and considered as positive, making the positive features congruent and the negative ones, incongruent. With the constant incongruent threshold model, negative features are always saved by the individual; on the contrary, the positive features are saved only if the sign of the individual's global attitude is positive - if the individual's global attitude becomes negative, the positive features are ignored.

The rational model assumes that all features are saved and the object is considered as neutral, whatever the order of the features. However, in our model, when the negative features are at the beginning, the object is finally perceived as negative, because once the major negative feature is saved, the positive features are ignored. That is why we say the model exhibits the primacy bias. We compared the two versions (constant or dynamic

incongruence threshold). They slightly differ in their impact but both exhibit the primacy bias.

Chapter 2.1 also shows that the interactions between individuals, exchanging features, can significantly increase the probability of such non rational attitudes.

Chapter 2.2 enlarges the study of the interaction effect. It also considers an object involving more features and focuses on the CIT model. This new study adopts the double modelling approach (Deffuant 2004) which aims at better understanding the aggregated dynamics of an individual based model by developing a specific model of these aggregated dynamics, derived from the individual based dynamics. In ecology, (Grimm 1999) encourages researchers to compare IBMs with aggregated models and numerous examples of the double-modelling come from ecology where researchers have various motivations to apply it:

- to identify the conditions in which an individual based model can be replaced by its aggregated approximation (which often require less computer capacity and time) (Mabrouk 2010; Mathias 2011).
- to understand the impact of a particular modelling choice : (De Angelis and Gross 1992) study the influence of transforming continuous variables into discrete distributions in models of ecological dynamics.
- to understand what to get out of IBM compared to other modelling approaches: (Picard and Franc 2001) show that space-dependent individual based models and aggregated models (regarding either spatial influence or description of the population) of forest dynamics lead to different results.
- (Fahse, Wissel et al. 1998; Duboz, Ramat et al. 2003) use individual-based models to extract parameters for population-level dynamics...

In social dynamics models, focusing for example on opinion dynamics studies, this double-modelling approach has been applied to the bounded confidence model (Deffuant, Neau et al. 2001; Hegselmann and Krause 2002) by (Ben-Naim, Krapivsky et al. 2003; Deffuant and Weisbuch 2007; Lorenz 2007). Their purpose was to develop an exhaustive knowledge of this model's asymptotic behaviour. (Deffuant and Weisbuch 2007) used the same approach, to improve the understanding of the extremist effect for this bounded confidence model. (Martin, Deffuant et al. 2004) applied this method to the study of a binary vector version of the bounded confidence model. They show a limitation of the double modelling approach: the state space can be too large to be tractable. (Edwards, Huet et al. 2003; Edwards 2004; Huet, Edwards et al. 2007) applied this method to study a stochastic IBM of binary behaviour diffusion individual model. We particularly aim at understanding the interaction effect in a random Erdős network (Erdős and Rényi 1960).

In the case of our information filter model, we built a differential equation model from the individual-based model. This aggregated model rules the evolution of probabilities that individuals belong to different groups defined by a set of retained features. We solve it numerically and show that this aggregated model approximates very well the IBM results. Moreover, the analysis of the graph structuring the groups shows how interactions can favour the diffusion of the negative major feature. Indeed, the qualitative analysis of the

equations shows that the negative major feature is more frequently communicated by the population than by the media if the frequency of the diffusion by the media is large enough. This increases the quantity of primacy bias in the final population compared with the case of isolated individuals. When the frequency is large enough, the sensitivity to the order of reception of the first individuals receiving the features becomes crucial.

In (Deffuant and Huet 2009), we also check if the effect of the interaction on the primacy bias depends on the structure of the network. We test different types of networks: random networks with a given average of links per node (Erdős and Rényi 1960), random networks with a constant number of links per node, or scale free networks (Barabasi and Albert 1999). For each type of network, we vary the number of average links by node (2, 4, 6, 8 and 10 links). The effect decreases when the number of links decreases, but it is still significant for networks with 2 links per node on average (75% of negative people against 85% in the basic case, but only 66% of negative people when people don't exchange their features). We note that the networks give very similar results when they have the same average number of links. Moreover, when this average number reaches 6 or 8, the increase of primacy bias is very close to the one observed in a totally connected population. In practice, this means that the aggregated model can be relevant to approximate the effect of interaction for some networks.

After this section on our results on the impact at the population level of an individual information filter, we focus in chapters 2.3 to 2.5 on another mechanism: the rejection of others' opinion. These chapters study the impact of the introduction of a rejection mechanism into a social influence model based on attraction between discussing agents having sufficiently close opinions. This is the bounded confidence (BC) model (Deffuant, Neau et al. 2001; Hegselmann and Krause 2002) which considers that two individuals, each having an opinion defined by a segment, influence each other and become closer in opinion, when their opinions differ less than a threshold. The above principle is based on homophily implying that close enough individuals tend to become closer. I selected some papers that I co-authored on this topic which are representative of the models and the method we use to study them.

Chapter 2.3 proposes a first model of the rejection coupled to the attraction which is inspired from theories in social science (Huet, Deffuant et al. 2008 d). It corresponds, except for the rejection, to a bounded confidence model in two dimensions. It is mainly based on a need for consistency of beliefs and leads to a preference for interactions with similar partners. Overall, the model dynamics can be summed up as follows: Individuals are characterised by two-dimensional continuous attitudes, each associated with an uncertainty  $u$ . Individuals interact by random pairs. If their attitudes are closer than  $u$  on both dimensions, or further than  $u$  on both dimensions, or closer than  $u$  on one dimension and closer than  $u+\delta u$  on the other dimension, then the rules of the BC model are applied. But if their attitudes are closer than  $u$  on one dimension and further than  $u+\delta u$  on the other dimension, then the individuals are in a dissonant state that they tend to solve by shifting away their close attitudes. The dynamics has three parameters: the uncertainty  $u$ ,  $\delta$  the level of intolerance, and a parameter ruling the speed of the opinion move when individuals become closer or shift away.

As for the information filter model, the study is guided by a comparison with a reference model. The reference to which we compare our models is the bounded confidence model in two dimensions (2D BC). As for the 2D BC, since the dynamic is launched, individuals tend to group each other into clusters of individuals having almost the same opinion. The model shows metastable clusters, which maintain themselves through opposite influences of competitor clusters. Our analysis and first experiments support the hypothesis that, for a large range of uncertainty values, the number of clusters grows linearly with the inverse of the uncertainty  $u$ , whereas this growth is quadratic in the 2D BC model.

Chapter 2.4 is dedicated to the study of this first model (Huet and Deffuant 2008 a) using a double modelling approach. This new study completes the study by simulations that we present in the previous chapter. This implementation of the double-modelling method was particularly enlightening. The results were presented in two conference papers (Huet and Deffuant 2008 b; Huet and Deffuant 2008 c). However, only one was selected in this document as a chapter. Indeed, the aggregated model of the IBM is the same in the two papers except the way the population is initialised. We sum-up in the following paragraphs the results we obtain in the two papers.

The aggregated model of the IBM is based on the socio-dynamic principles (Weidlich 2002). They come from physics and consist in deriving a master equation that rules the dynamics of groups with a similar state. In our case a group gathers individuals having the similar opinions. It requires discretising the opinion space with a sufficiently small resolution. Then we observe the evolution of the probability distribution until it becomes stable. For the present model exhibiting metastable states in the IBM version, the detection of the stationary state is easier than with the IBM. Indeed, in IBM, at the stationary state, individuals continue changing opinion, even if they remain member of the same group. This continuous move does not facilitate the detection of this stationary state because it is often based on the size of the individual opinion change which tends to 0 in the standard BC model. In the aggregated model, individuals do not exist; only groups are considered and, in this case the detection of the stationary state is easy and not ambiguous. It corresponds to a flow of 0 between the groups.

The first aggregated model considers a totally uniform initial population of groups. This means that every group contains exactly the same part of the population. This work is presented with more details in (Huet and Deffuant 2008 b) and it:

- Shows that rejection favours consensus because in most cases the final state is only one centered cluster of individuals having the same opinion; It occurs for a sufficiently large uncertainty  $u$  like in the 2D BC. However, while everyone agree on the opinion at the end of the simulation for an uncertainty valued at least 0.54 in the 2D BC, only an uncertainty of 0.2 is required to obtain the same final state in this first aggregated model. The conditions for a total consensus are then more numerous or less restrictive when we add rejection into the model.
- Exhibits the conditions for which the clustering is impossible because of the incapacity of the model to generate local maxima required for launching the dynamics of clustering.

However, this first aggregated model is quite different from the IBM because it has only these two “final” states: one centred cluster or the distribution remains uniform. That is due to the uniformity of the initial distribution of probabilities over the groups having a same opinion. It makes every interaction symmetric and every change is exactly compensated by the symmetric one.

In a second step, we study the same aggregated model giving to an initial group a slightly higher probability than the others (Huet and Deffuant 2008 c). This small initial asymmetry allows the distribution to change and the results of the aggregated model are equivalent to those of the IBM for a very large population. This work is presented in chapter 2.4. It:

- Shows a small local initial asymmetry is a necessary and sufficient condition to obtain the clustering obtained in the IBM;
- Confirms the existence for the aggregated model of conditions of low uncertainty values for which the distribution remains unchanged. A new experimental design on the IBM exhibits this behaviour also exists in the IBM but appears only for very large population (about 10000). While for this “no-clustering” state in the aggregated model, we simply observe no change into the distribution of probabilities, in the IBM, we observe continuous random moves of individuals who are always more rejected than attracted;
- Confirms the particular behaviour of the IBM with a sufficiently small finite population size (< 10000); it is able to cluster, even for a low uncertainty value.

Chapter 2.5 proposes a second model of the rejection coupled to the attraction (Huet and Deffuant 2010). It is inspired from a particular set of experiments (Wood, Pool et al. 1996). It can be considered as a particular case of the previous one. As outlined by (Wood, Pool et al. 1996), it corresponds to the particular behaviour of people who are highly self-engaged in one dimension and consider it as defining themselves. From the modelling point of view, it assumes that one dimension is more stable and more important than the other. This more stable dimension, called main dimension, represents the highly self-relevant issue for the individual and rules the attitude change on a secondary dimension. If two individuals are close on the main dimension, then they attract each other on the main and on the secondary dimensions, whatever their disagreement on the secondary dimension. If they are far from each other on the main dimension, then too much proximity on the secondary dimension is uncomfortable, and generates rejection on this dimension. The proximity is defined by comparing the opinion distance with  $u_m$  the attraction threshold on the main dimension and with  $u_s$  the rejection threshold on the secondary dimension.

The stationary states exhibited by this new model are somewhat close to the stationary states obtained from the first model. The differences are:

- the number of clusters is smaller than in the first attraction-rejection model; it corresponds to the number of clusters of the 1D BC. The dynamics is ruled by the main dimension and the final number of clusters is approximately  $1/u_m$ .
- the dynamics always lead to clusters: it seems that no uncertainty conditions or population size lead to general random walks in the opinion space.

- fluctuations of opinions are observed only on the secondary dimension – these fluctuations can push groups to polarize, adopting a more and more extreme opinion in order to find out a position where its members are not too much rejected. These fluctuations remain in-group fluctuations if  $u_s$  is not too large. If not, groups become less and less cohesive on this secondary dimension. The fluctuations can remain continuous because an individual cannot have an opinion which does not imply to be rejected by the large number of members of other groups compared to the small number of members of its own group. However, the groups remain stable because they are defined on the main dimension and remain very cohesive on this dimension.

The models are overall very close since both lead globally to fewer clusters than the BC model. We can interpret it as a higher level of consensus. Both can exhibit polarisation while it is not the case for the BC model. They also both have individual fluctuations in their equilibrium state. Moreover, these fluctuations don't affect the population level state, since they don't change the membership. However, the fluctuations take place only on the secondary dimension for the second model, while they occur on both dimensions in the first one.

We note that, in the first model (the two dimensions being equivalent), for the same small values of  $u$  on the two dimensions, and large population, every individual adopts a continuous random move without being able to find out a stable opinion position. This does not occur in the second model (with main and secondary dimensions). To better understand the reason of this difference, we did some experiments aiming at reducing the difference between the two models in order to identify the mechanism responsible for the non-clustering state. We conclude that the uniform random state is reached only when the rejection is possible on all the attitudinal dimensions. If it is possible only on the secondary dimension, a clustering process always occurs on the main dimension. The other difference between the two models, which was suspected to explain the clustering process in the second model, was the unbounded attraction on the secondary dimension, in case of agreement on the primary dimension (while the attraction is bounded on the two dimensions for the first model). This complementary work also allows us to conclude that this unbounded attraction is not responsible for the clustering process for low uncertainties in the second model.

The present document ends with a synthesis of our work, organised around methodological modelling questions and some perspectives.

# Part 1. Data driven integrated modelling of social dynamics

## Abstract

In chapter 1.1, we describe the structure of the individual-based model of the European rural municipality evolution developed in the PRIMA project, and we focus on its implementation on a French region: the Cantal *département*. The other chapters describe different parts of this model in more details, in particular their link with the data. A first part of the model generates the initial population. It is based on a sample-free innovating algorithm while classical method required an initial sample that has to be projected at the level of the studied region at the convenient date. The model of the individual activity dynamics is derived from the European Labour Force Survey database. This model rules how the individual changes labour statuses and employments, essentially based on her age, her profession and her activity sector when she is occupied. The modelling part deals with dynamics that were more difficult to derive from data, mainly the demographic dynamics. Based on the Occam razor principle, we test very simple dynamics and select them on their capacity to produce results close to reference data from the French National Statistical Office. In particular, we propose a simple residential mobility model integrating decision to move and location choice and partly ruling the emigration.

## Contents

Chapter 1.1 The SimMunicipality model	25
Chapter 1.2 Generating the initial population	41
Chapter 1.3 Parametrisation of the individual activity dynamics	59
Chapter 1.4 Parametrisation of the unknown laws of demography	81

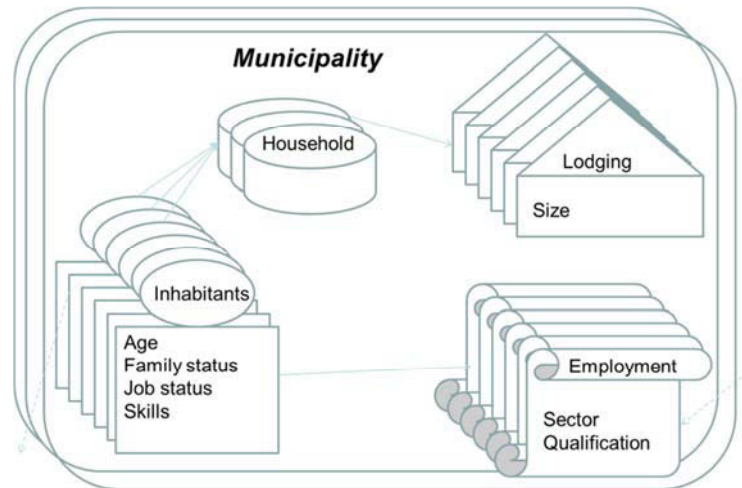




## Chapter 1.1 The SimMunicipality model

Authors: Sylvie Huet, Nicolas Dumoulin, Guillaume Deffuant, Floriana Gargiulo, Maxime Lenormand, Omar Baqueiro Espinosa, Sonia Ternes

Extracted from the European project PRIMA final report (Huet, Dumoulin et al. 2012) approved by the European Commission



### Abstract

During the PRIMA European project, we conceived a conceptual model dedicated to the study of the evolution of the European Rural Areas in terms of demography and occupation of their populations. This is a hybrid agent-microsimulation model considering inhabitants as the basic elements composing in households. On the other hand, municipalities are the basic spatial element and the providers for employment and housings. This chapter presents the main objects and their relationships are identified as a first generic way to parameterize the dynamics of such a conceptual model for an implementation of a European region.

### Contents

Main entities, state variables and scales .....	27
MunicipalitySet.....	27
Municipality.....	28
The job and the residence.....	29
Household .....	29
Individual .....	29
Process overview and scheduling .....	30
The main loop.....	30
Dynamics of offer for jobs, services and housing .....	31
Dynamics of labour status and job changes.....	32
Demographic dynamics.....	34



We have adopted a micro-modelling approach. The presentation of the model globally follows the requirements of the ODD (Overview, Design concepts, and Details) framework (Grimm, Berger et al. 2006). Indeed, this recently updated protocol (Grimm, Berger et al. 2010) has proved its utility to describe properly complex individual-based models, for example in (Polhill, Parker et al. 2008).

The purpose of the model is to study how the population of rural municipalities evolves. We assume that this evolution depends, on the one hand, on the spatial interactions between municipalities through commuting flows and service, and on the other hand, on the number of jobs in various activity sectors (supposed exogenously defined by scenarios) and on the jobs in proximity services (supposed dependent on the size of the local population). Indeed, in the literature, the most cited experimental designation for the evolution of the rural municipalities is what is called the residential economy (Blanc and Schmitt 2007; Davezies 2009). It argues that rural areas dynamics is linked to the money transfers between production areas and residence locations. These money transfers are for instance performed by commuters, or by retirees who move from the urban to the rural areas. Indeed migrations from urban to rural areas are also considered as a very important strand for rural areas evolution (Perrier-Cornet 2001). The residential economics studies particularly how an increasing local population (and money transfers) increases the employment in local services. The geographic situation plays also a role in the municipality evolution (Dubuc 2004). To summarise, existing literature stresses the importance of the different types of mobility between municipalities, commuting, residential mobility (short range distance), migration (long range distance) (Coulombel 2010) and the local employment offer generated by the presence of the local population.

These two aspects have to be properly taken into account in our model, since our objective is to study through simulations the dynamics of rural areas. Obviously, it appears also essential to model the demographic evolution of the municipality considering the strands explaining the local natural balance.

## **Main entities, state variables and scales**

The model represents a network of municipalities and their population. The distances between municipalities are used to determine the flows of commuting individuals (for job or services). Each municipality comprises a list of households, each one defined as a list of individuals. The municipalities also include the offers of jobs, of residences and their spatial coordinates. Here is the exhaustive list of the main model entities with their main attributes and dynamics.

### **MunicipalitySet**

The set of municipalities can be of various sizes. It can represent a region of type NUTS 2 or NUTS 3<sup>1</sup>, or more LAU or intermediate sets of municipalities such as "communauté de communes" in France.

---

<sup>1</sup> Eurostat defines the NUTS (Nomenclature of Territorial Units for Statistics) classification as a hierarchical system for dividing up the EU territory: NUTS 1 for the major socio-economic regions; NUTS 2 for the basic regions for the application

**Parameter:** a threshold distance called "proximity" between two municipalities; beyond this distance the municipalities are considered too far from each other, to allow commuting between them without considering to move for instance

## Municipality

It corresponds to LAU2<sup>2</sup>. The municipality is the main focus of the model. It includes:

- A set of households living in the municipality. The household corresponds to the nuclear family<sup>3</sup>. It includes a list of individuals who have an occupation located inside or outside the municipality).
- The set of jobs existing on the municipality and available for the population of the model (i.e. subtracting the jobs occupied by people living outside the modelling municipality set).
- The distribution of residences, or housings, on the municipality.

There is a particular municipality, called "Outside": it represents available jobs accessible from municipalities of the considered set, but which are not in the considered set. The job offer of Outside is infinite and the occupation is defined by the process ruling the probability of individuals to commute outside the set.

### Parameters:

- An initial population of households composed of individuals with their attribute value and their situation on the labour market (see chapter 12 for an example of building of such a population)
- A residence offer: available number of residences for each type. A type corresponds to the number of rooms
- A job offer: number of jobs offered by the municipality for each type of job; the exogenously defined part of job offers is distinguished from the endogenously defined part in order to update this last part easily
- The neighbourhood: each municipality has rings of 'nearby' municipalities (practically every 3 Euclidian kilometres) with a maximum distance of 51 Euclidian km where individuals can find out jobs and partners while households can find lodgings.
- Distance of the municipality to the border of the region
- Spatial coordinates

As said earlier, in the case of special municipality called "Outside", all variables, except job offer and job occupation, are empty.

---

of regional policies; NUTS 3 as small regions for specific diagnoses; LAU (Local Administrative Units 1 and 2) has been added more recently to allow local level statistics

<sup>2</sup> consists of municipalities or equivalent units

<sup>3</sup> A nuclear family corresponds to the parents and the children; that is a reductive definition of the family corresponding on the most common way to define the family in Europe nowadays.

## The job and the residence

A job has two attributes, a profession and an activity sector in which this profession can be practiced. It is available in a municipality and can be occupied by an individual. The profession is an attribute of the individual at the same time it defines a job. For example in France, it takes six various values. There are four activity sectors: Agriculture, Forestry and Fishing; Industry; Building; Services and Commerce. Overall, considering the six professions for four activity sectors, we obtain 24 jobs to describe the whole diversity of jobs in the region we study (i.e. the Cantal "département", called only Cantal later in this chapter).

The residence has a type which is classically its size expressed in number of rooms. A residence is available in a municipality and can be occupied by 0, one or more households. Indeed several households can live in one residence for instance when a couple splits up and one of the partner remains in the common residence for a while. It is also the case in some European countries where it is customary for several generations to live under the same roof.

## Household

Table 1. Attributes defining the household state

Name	Type	Values
<i>Members</i>	<i>List of Individuals</i>	
<i>Couple</i>	<i>Boolean</i>	<i>True, false</i>
<i>Leader</i>	<i>Individual</i>	
<i>Residence</i>	<i>Residence</i>	
<i>Residence need</i>	<i>Boolean</i>	<i>True, false</i>
<i>Municipality of residence</i>	<i>Municipality</i>	

For the initialisation, residences are associated randomly with households. Then, new households are created when new couples are formed or when people from outside the set of municipalities migrate into the municipality. Households are eliminated when their members die, or when the couple splits up, or when they simply migrate outside the municipality set. When a behavior of an individual has an impact on the household, a leader is assigned randomly, or designed depending on the process. This leader will be the one deciding for the household. That is for example the case when an individual finds a job very far: she becomes the leader to make the household moving and finding a residence close to her new job.

## Individual

The individual is instantiated via one of the adults of a household having the "couple" status in the birth method, or directly from the initialisation of the population, or by immigration. The age to die, the age the person will enter the labour market, and the age of retirement are attributed to the individual when it is created. These ages are assigned by a probability method. The activity status defines the situation of the individual regarding employment, especially whether or not she is looking for a job. The individual can quit a job, search for and change jobs ...

The profession is an attribute of the individual indicating at the same time her skills, level of education and the occupation she can aspire to. In France for example, professions take the value of the French socio-professional categories categorised in six modalities that define at the same time a kind of occupation, an average level of education and an approximate salary.

Table 2. Attributes defining the state of an individual

	Type	Values
<i>Activity status</i>	<i>Enum</i>	<i>student, inactive, retired, employed, unemployed (only the two last can search a job)</i>
<i>Profession</i>	<i>Enum</i>	<i>farmers; craftsmen, storekeepers, business owners; top executive managers, upper intellectual profession (senior executives); intermediary professions; employees; workers.</i>
<i>Job</i>	<i>Couple of values</i>	<i>24 couples (profession, activity sector) (see 0 for details)</i>
<i>Place of work</i>	<i>Municipality</i>	<i>Nil or a Municipality</i>
<i>Household status</i>	<i>Enum</i>	<i>Adult, Child</i>
<i>Age to die</i>	<i>Integer</i>	<i>Drawn from a distribution</i>
<i>Age in labour market</i>	<i>Integer</i>	<i>Drawn from a distribution</i>
<i>Age of retirement</i>	<i>Integer</i>	<i>Drawn from a distribution</i>

## Process overview and scheduling

### The main loop

The main loop calls processes ruling demographic evolution, the migrations, the job changes, and their impact on some endogenously created services and/or jobs. First, the scenarios are applied to the municipalities. Then, endogenously available jobs and services are updated in municipalities. Finally, demographic changes are applied to the list of households. The following pseudo code sums-up the global dynamics:

```

At each time step:
  For each municipality
    municipality.update external forcings: offer of jobs, residence
    municipality.update endogenous job offer for services
    to residents
    municipality.compute in-migration
  For each household:
    household.members.job searching decision (this process can make
    free some jobs from people becoming retired or inactive)
  For each household:
    household.members.searching for a job
    household.members events (coupling, divorce, birth, death)
    household.residential migration
    household.members.individual ages

```

Time is discrete with time steps corresponding to years. The households are updated in a random order during a time step.

## Dynamics of offer for jobs, services and housing

In the municipality objects, jobs, services and housing offers are ruled. Changes in housing offers are specified in scenarios. Various sizes are considered in order to match the needs of households.

The job offer process is optionally twofold:

- it is entirely defined exogenously through scenarios;
- one part defined through scenarios which specify the increase or decrease of jobs in different sectors, and a second part concerning the proximity of service jobs, which are derived by a specific statistical model.

We consider the second possibility since numerous are the researches pointing out the importance of services for the rural areas dynamism (Soumagne 2003; Dubuc 2004; Fernandez, Brown et al. 2005; Aubert, Dissart et al. 2009). Also the residential economics shows the importance of the presence of the population in rural municipalities (Davezies 2009). Practically, in France, we distinguish the proximity services which rely directly on the presence of population from the services which are decided according to other factors (assets of the location, political will at different levels, etc.). We integrated the dynamics of creation and destruction of proximity services jobs in the micro-simulation model, using a statistical model derived from the data of the region. Starting from the classical minimum requirement approach proposed by (Ullman and Dacey 1960), (Huet, Dumoulin et al. 2012) propose a model which takes into account the distance between a municipality and its closest centre of services (i.e. most frequented municipality, called MFM). This new model has been grounded on detailed data related to jobs and centres of services (Huet, Dumoulin et al. 2012). Therefore, we use the extracted statistical relation to adjust the number of jobs in proximity services in the municipalities of the model. It is  $E = \beta_0 + \beta_1 \ln P + \varepsilon$  with  $E$  = minimum employment offer in the municipality to satisfy the need for services of one resident;  $P$  = the population of the municipality;  $\beta_0$  and  $\beta_1$  = parameters.

For each municipality, this function is computed every year in order to update the service sector job offer depending on the distance of the municipality to the closest centre of service (called MFM). The form of the function for different municipality sizes with various distances to the MFM indicates that:

- in any case, the job offer is higher in the centre of services and decreases in the surrounding;
- however further from the centre of services, the number of jobs increases again until reaching a plateau at a distance higher than 10 minutes;
- the larger is the municipality, the higher is the number of jobs in proximity services.

The other creations and destructions of jobs are ruled by scenarios.

### Parameters:

- exogenously defined scenarios of job offer by municipality
- optional, for endogenously defined job offer: for example for the French case study, distances to the Most Frequented Municipality of every municipality of the Cantal (given



by the French Municipal Inventory of 1999); class of distance to the most frequented municipality (MFM) for every municipality and regression coefficients  $\beta_0$  and  $\beta_1$  extracted of the analysis of the French Census of 1990, 1999 and 2006 (see (Huet, Dumoulin et al. 2012)).

The proportion of proximity service jobs offer over professions is assumed to be the same than the one for the whole service sector job offers (which is probably a strong approximation). This allows us to distribute the proximity service jobs in the different jobs in the service sector.

## **Dynamics of labour status and job changes**

A new individual can be generated in a household having the “couple” status with the birth method, or directly from the initialisation of the population, or from the immigration method. A newly born individual is initialised with a student status that she keeps until she enters the labour market with a first profession. Then, she becomes unemployed or employed with the possibility to look for a job. She may also become inactive for a while. When she gets older, she becomes a retiree. We describe rapidly these dynamics to situate them in the global picture of the model. It has been parameterised from data. The parameterisation is described with more details in (Huet, Lenormand et al. 2012) which is also our chapter 1.3.

### ***Entering on the labour market***

The individual stops being a student at the age to enter on the labour market and becomes unemployed. She searches immediately for a job and can get one during the same year. A first profession she looks for has to be defined at the same time the first age of research is determined.

**Parameters:** probabilistic laws to decide the age a student enters on the labour market and the first profession she is going to look for.

### ***Job searching decision***

The decision for searching a job is a two-step process. First, an individual has an activity status indicating if she is susceptible to search for a job or not. She can change her status and then her probability to seek a job. When she decides searching, she has also to decide what type of job to search for. Five different activity statuses define the individual situation regarding the labour market in the model:

- The **student**: an individual is a student in the first part of its life, until the age she enters on the labour market. We consider the probability of a student to look for a job is 0 since we are only interested in rural municipalities. Students in age working mainly look for a job in the large cities where they study.
- The **unemployed**: an individual is unemployed when she is considered active (on the labour market) and has no job. For sake of simplicity, we assume an unemployed has a probability 1 to look for a job.
- The **employed**: she is an individual who has a job. She can decide searching for another job, in the same profession or not. Her probability willing to change job classically depends at least on her age.

- The **inactive**: she can be inactive for a long time or just stopping to work for one year, having a baby for example. During this period, her probability to search for a job is 0.
- The **retired**: at the age of retirement, an individual retires. Her probability to look for a job is then assumed to be 0.

We have seen the probability to search for a job (or the law ruling this probability) depends on the activity status. Figure 1 describes the way an individual changes activity status and thereby the probability to search.

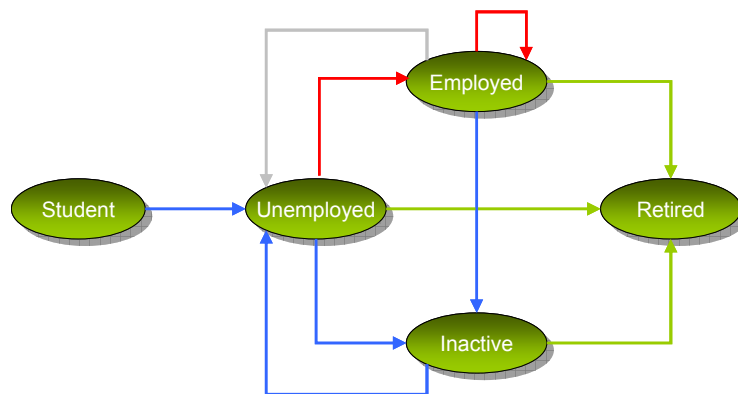


Figure 1 - Transitions of status and their link to the data. Red arrows: change by finding a job; grey arrows: when she is fired; green arrows: at the age of retirement (picked out from a law extracted from data); yellow arrows: due to a probabilistic decision of becoming inactive extracted from the Labour Force Survey data; purple arrows: due to probabilistic decisions extracted from the Labour Force Survey data.

Entering the labour market, the student becomes unemployed and searches for a job with a probability 1. An unemployed, as an employed, can find a job through processes presented in the following sections and become employed. If an unemployed always searches for a job by assumption that is not the case for an already employed individual (her probability to search has to be extracted from data). Employed and unemployed individuals can also become inactive. Then we assume that they stop searching for a job the time they remain inactive. Every activity states, except student, can be followed by the retirement state in which we assume the individual stops searching for a job. An inactive, if she doesn't retire, either can come back on the labour market adopting an unemployed status to search for a job or can remain inactive.

Most of the laws ruling the activity status changes have to be parameterised. The grey-arrows transitions are much more endogenously defined. That is the employed to unemployed transition which is due to the decreasing availability of job offer implying a sacking. It can also be, for instance a resignation of an individual leaving her municipality to follow her partner to another place of residence.

Knowing an individual searches for a job, we have to compute which profession she looks for. One can notice that an individual only looks for a profession; we neglected to take into account the activity sector in her choice. The activity sector will be defined by the found job among the set of possible job offers for the individual. We expect the job offer to be a sufficient constraint on the activity sector to allow the model exhibiting a statistically correct distribution of occupied jobs by activity sector.

**Parameters ruling the job research decision:** probability becoming inactive; probability to stop being inactive; probability laws defining what profession to search for; parameters for entering the labour market and to retire

### ***Searching for a working place***

The question for the individual is now to decide where to search for a job. If the leader of the household has already found a job far (further than the proximity attribute) from the place of residence and the household is trying to move close the leader's place of work, then the other household members, waiting for a change of residence, do not try to change job since they do not know where they will be living. Until the household finds out a new residence place, nobody is going to change jobs.

In the other cases, if the individual is searching for a job, she looks from the closest offer to the furthest considering successive rings of distance relevant to describe the average space between municipalities, for example 3 in France, starting from her place of residence. Indeed, she considers at first the job offer located in her place of residence and at most at distance 3 in our example from this place. If she can't find a job, she continues looking from a distance 3 to a distance 6 (in our example) from her place of residence. She continues the procedure until finding a not empty list of possible jobs or considering the ring at the maximum distance far from her place of residence. She can also search for a job outside the set. This depends on the parameterisation. For example, in France, as presented in (Huet, Lenormand et al. 2012), she decides to commute outside using the probability to commute outside knowing her place of living since the searching procedure finished and only if she had not found out a job in her place of residence and found one elsewhere (to be coherent with the data available for parameterisation). Finally, if she does not commute outside the set of municipalities, she chooses at random a municipality as a place of work in the list of the possible job offer she has collected.

**Parameters:** length of a ring; maximum distance to search for a job inside the set; probability to commute outside for an inhabitant of every municipality (example for France).

### ***Become a retiree***

At a given age, the individual becomes a retiree. We assume, for sake of simplicity, that a retiree does not search for a job.

**Parameter:** probability law to decide the individual's retirement age.

### ***Demographic dynamics***

A new household can be created when an individual becomes an adult or when a new household comes to live in the set of municipality (i.e. in-migration). The main reasons for household elimination are out-migration and death. Three main dynamics change the household type (single, couple, with or without children and complex<sup>4</sup>): makeCouple; splitCouple and givingBirth. These processes have to be parameterised depending on the case study and its available data. We describe them using an example of implementation for a French region allowing the reader to form a clearer idea about them. Moreover this implementation is described and discussed in chapter 1.4.

---

<sup>4</sup> A complex household is a household which is not a single, a couple with or without children.

### ***BecomingAnAdult***

Becoming an adult means an individual creates her own household. This can lead her to move from parental residence because of a low housing satisfaction level, but it's not always the case. An individual loses her child status and becomes an adult when: she finds her first job; or she is chosen by a single adult as a partner; or she remains the only children in a household after her parents leave or die while her age is higher than parameter `firstAgeToBeAnAdult`.

**Parameter:** first age to become an adult – 15 is the age considered by the French or other European National Statistical Offices

### ***Household migration and mobility***

In changing residence process, we include both residential migration and mobility without making a difference, between short and long distance move, as it is often the case (Coulombel 2010) in the literature. The submodel we propose directly manages both types of moving. However, it turned out easier for us to distinguish two categories of migration: the migration of people coming from outside to live inside the set; the migration of people who already live inside the set.

The immigration into the set is an external forcing. Each year, a number of potential immigrants from outside the set are added to the municipalities of the set. These potential immigrants can really become inhabitants of the set if they find a residence by themselves or by being chosen as a partner by someone already living in the set in case they are single (with or without children). Thus, looking for a place or a partner of residence are the only action they execute until they become an inhabitant of the set. Until the potential immigrant becomes a real inhabitant, she cannot search for a job. Indeed, the job occupied by people living outside the municipality set are already taken into account through the scenario and allowing potential immigrants to find a job directly would be redundant. The definition of who are potential immigrants, how numerous they are, and when they are introduced is specified exogenously. Since they are created, the potential immigrants are temporarily located into a municipality from which they can find a residence or being chosen as a partner. They are placed in a municipality following a probability to be chosen, which is computed for each municipality depending on the population size of the municipality and its distance to the frontier of the set. A particular attraction of young people for larger municipalities is also taken into account.

The mobility of people already living inside the set of municipalities is mainly endogenous. Such a mobility can lead the household simply to change residence, municipality or to quit the set of studied municipalities. Overall, a household decides to look for a new residence when:

- a new couple is formed: the couple chooses to live initially in the largest residence among the ones of the partners;
- a couple splits: one of the partners, randomly chosen, has to find out another residence even if she remains for a while in the same residence (creating her own household);
- an adult of the household finds a job away from the current place of residence (beyond the proximity parameter of the `MunicipalitySet`);

- a student or a retiree decide to move;
- the residence is too small or too large. This can be due to a birth, a new couple or to someone who left the residence for example. The too small or too large characteristic is assessed through a satisfaction function which has to be parameterized. It is at least based on the difference of size between the occupied size and an ideal size for this household but can also take the age into account as in France.

The principle for the search of a new residence is the same as the one for searching a place to work. The household through her leader (chosen at random among the adults each time a decision has to be taken for the household) looks from the closest offer to the furthest considering successive rings of distance<sup>5</sup> x (i.e. the same basic distance used to search for a job), valued 3 as example for France. She starts from her place of work (or residence if she does not work), meaning at a distance at most 3. If she can't find a satisfying place to live in, she continues looking from a distance 3 to a distance 6 (in our example). She iterates the procedure until finding a not empty list of possible jobs or considering a ring of a distance valued proximity far from her place of work (or residence if she does not work). Above this proximity parameter, the possible lodgings are considered too far from the starting place and the research stops. Before accepting to consider an offer, she checks the residence offer is not too far from the place of work of her partner (if she has one). She can also move outside. The decision moving outside depends on the parametrisation. As an example, it has been designed as follows for France. The searching procedure finished, if she had not found out a residence in her place of residence and if she has found one elsewhere, she decides to move outside using the probability to move outside knowing her place of living. Finally, if she does not move outside the set of municipalities and has found a residence, she chooses at random a municipality of residence in the list of collected possible housings. The probability to move out of the set of municipalities varies with the age of the individual. What is an acceptable housing to collect during the search procedure depends on a satisfaction regarding the size of the housing. The level of acceptation of a possible size is 0 if the size does not respect the fact the household want to increase or decrease her actual size. If the proposed size respects this tendency her probability to be collected for the list of choices decreases with the increasing of the difference of size between the proposed and the ideal size.

The way a household decides if a housing proposal is satisfying has to be parameterised also. Always as an example, it has been designed as follows for the implementation of this model for a French region. The level of satisfaction of a possible place of residence depends on two dimensions. Firstly, a municipality is examined as a place to reside if it satisfies the need for a house (and not for an apartment) for the household of size higher than one. Practically, a municipality has a ratio of house offer over its residence offer and she is considered in the research procedure with a probability equal to this ratio for every household larger than one member. The second dimension influencing the place of residence is the size of the residence offer. Indeed she has at least to respect the need for increasing size or decreasing size of the household. Practically, if a household needs a larger housing, she will consider an offer only if it is a least as large as the current one. That is the same for the need for a

---

<sup>5</sup> the distance definition depends on the parameterization of the model

smaller housing. Coupled to a dissuading effect of age in the moving decision, it makes rarer people decrease their size housing than increase it (in conformity with literature).

**Parameters for immigration:** yearly migration rate; number of out of the set migrants in year  $t^0 - 1$ ; probabilities for characteristics of the immigrants (size of the households, age of individuals...); distance to the frontier of the region of each municipality.

**Parameters within the set of municipalities and out-migration:**

- function defining the level of satisfaction of the size of the current housing;
- functions defining if a municipality, if a housing proposal is satisfying
- laws for migration extracted from data (as for example in France, for students and retirees)

### **Death**

The death age of the individual is determined when she enters the simulation (through birth, initialisation or immigration). When an individual dies, its household status is updated depending on the number of remaining members and their statuses, parent or children. Households are eliminated when all their members die, when the couple splits up, or when they simply out-migrate.

**Parameter:** probability to die by a certain age - made available by INED<sup>6</sup> from the various French Census at the national level.

### **MakeCouple**

This method has to be parameterised depending on available data. As an example, we describe what has been chosen for the implementation of a French region:

- During each time step, each single individual (with or without children) has a probability to search for a partner;
- If the individual tries to find a partner, she tries a given number of times in her place of residence before trying in every municipality close to her own and her place of work to find someone who is also single and whose age is not too different (given from the average difference of ages in couples and its standard deviation). She begins searching very close and goes to search further if she can't find out a partner (in the same way she looks for a residence or a job). She can search among the inhabitants or the potential immigrants. The furthest municipalities where she can found out a partner are defined through the threshold parameter "proximity";
- When a couple is formed, the new household chooses the larger residence (the immigrating households always go into residences of their new partners; this move can force one member to commute very far (at a distance higher than the MunicipalitySet parameter proximity). This situation can change only when she is becoming the leader triggered by the job search method and implying that the household will aim to move closer to her job location.

---

<sup>6</sup> French Institute of Demographic Studies

**Parameters:** probability to search for a partner; maximum number of trials; average difference of age of couples and its standard deviation. For the French implementation, the last one is given by the INSEE<sup>7</sup> at the national level based on the data from Census. For the two first, they have to be calibrated since they do not correspond to existing data.

### *SplitCouple*

All couples, except the potential immigrants have a probability to split up. When the split takes place, the partner who works further from the residence leaves the household and creates a new household, which implies that she searches for a new residence. When there are children, they are dispatched among the two new households at random.

**Parameter:** probability to split (in the French example: no possible data source then the parameter has to be calibrated)

### *Giving birth*

To simplify, we made the assumption that only households with a couple can have children, and one of the adults should be in age to procreate. We assumed that an individual in couple have an average number of children over her life. This number has to be parameterised. The other parameters are the minimum and maximum ages to have a child. From all these parameters, we compute for each couple the probability to have a child during that particular year if one randomly chosen individual's age allows reproduction.

**Parameters:** minimum and maximum age to give birth, number of children an individual can have during her life on average. Usually ages for reproduction ranges from 18 to 45.

## References

- Aubert, F., J. C. Dissart, et al. (2009). "Facteurs de localisation de l'emploi résidentiel en France." XLVIème Colloque de l'Association de Science Régionale de Langue Française (ASRDLF) 6-8 juillet, Clermont-Ferrand, France: 27.
- Blanc, M. and B. Schmitt (2007). "Orientation économique et croissance locale de l'emploi dans les bassins de vie des bourgs et petites villes." *Economie et Statistique* 402: 57-74.
- Coulombel, N. (2010). Residential Choice and Household Behavior: State of the Art, Ecole Normale Supérieure de Cachan. Working Paper 2.2a: 69.
- Davezies, L. (2009). "L'économie locale "résidentielle"." *Géographie, économie, société* 11(1): 47-53.
- Dubuc, S. (2004). "Dynamisme rural: l'effet des petites villes." *L'Espace Géographique* 1: 69-85.
- Fernandez, L. E., D. G. Brown, et al. (2005). "Characterizing location preferences in an exurban population: implications for agent-based modeling." *Environment and Planning B: Planning and Design* 32(6): 21.
- Grimm, V., U. Berger, et al. (2006). "A standard protocol for describing individual-based and agent-based models." *Ecological Modelling*(198): 115-126.
- Grimm, V., U. Berger, et al. (2010). "The ODD protocol: A review and first update." *Ecological Modelling* 221: 2760-2768.
- Huet, S., M. Lenormand, et al. (2012). Parameterisation of individual working dynamics. *Empirical Agent-Based Modeling: Parameterization Techniques in Social Simulations*. A. Smajgl and O. Barreteau, Springer: 22.
- Lenormand, M., S. Huet, et al. (2012). "Deriving the number of jobs in proximity services from the number of inhabitants in French rural municipalities." *PLoS One* 7(7): 13.
- Perrier-Cornet, P. (2001). "La dynamique des espaces ruraux dans la société française : un cadre d'analyse." *Territoires* 2020 3(Etudes et perspectives - Data): 61-74.

---

<sup>7</sup> French Institute of Statistics and Economical Studies

- Polhill, J. G., D. C. Parker, et al. (2008). "Using the ODD Protocol For Describing Three Agent-Based Social Simulation Models of Land-Use Change." *Journal of Artificial Societies and Social Simulation* 11(2 3): 30.
- Soumagne, J. (2003). "Les services en milieu rural, enjeu d'aménagement territorial." *Revista da Faculdade de Letras - Geografia I série XIX*.
- Ullman, E. and M. Dacey (1960). "The minimum requirement approach to the urban economic base." *Papers and Proceedings of the Regional Science Association* 6(192).



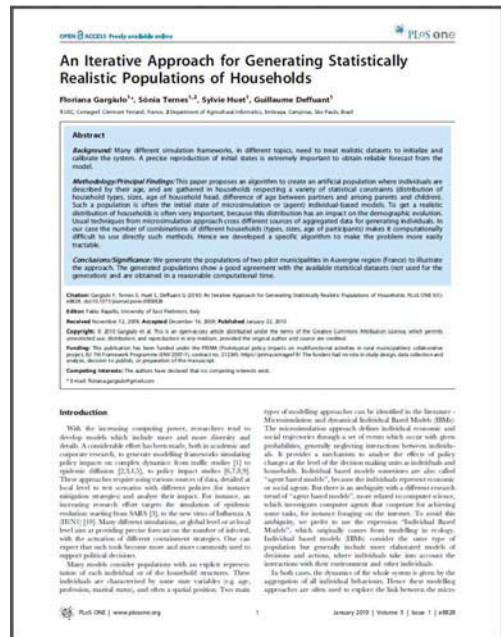


# Chapter 1.2 Generating the initial population

Title : An Iterative Approach for Generating Statistically Realistic Populations of Households

Authors: Gargiulo Floriana, Ternes Sonia, Huet Sylvie, Deffuant Guillaume

Published in PLOS, 2010.



## Abstract

This second chapter describes a method for generating the initial population, which does not require an initial sample population (whereas most methods for generating virtual populations do require such a sample). To summarize, we build for each municipality a list of agents with the exact number of individuals being each age and a list of households with the exact number of household members. Then, we try to fill one by one each household with individuals taking into account the probability of households having some particular properties, such as being a couple or having a given number of children. Each time a household is completed, another one is selected to be filled. At the end, we get a virtual population of households following the exact distribution of sizes, showing good statistical household properties and composed from individuals following the exact distribution of ages. We generate the populations of two pilot municipalities in Auvergne region (France), to illustrate the approach. The generated populations show a good agreement with the available statistical datasets and are obtained in a reasonable computational time.

## Contents

Materials and Methods .....	45
General Formulation of the problem .....	45
Calculating the probability of a household .....	47
An iterative algorithm avoiding to generate all possible households .....	51
Results .....	53
Conclusion .....	55



With the increasing computing power, researchers tend to develop models which include more and more diversity and details. A considerable effort has been made, both in academic and corporate research, to generate modelling frameworks simulating policy impacts on complex dynamics: from traffic studies (Nagel, Beckman et al. 1999) to epidemic diffusion (Eubank, Guclu et al. 2004; Colizza, Barrat et al. 2007a; Colizza, Barrat et al. 2007b; Ciofi degli Att, Merler et al. 2008), to policy impact studies (Gotts, Polhill et al. 2003; Holme, Holme et al. 2004; Ballas, Clarke et al. 2005; Ballas, Clarke et al. 2006). These approaches require using various sources of data, detailed at local level to test scenarios with different policies (for instance mitigation strategies) and analyse their impact. For instance, an increasing research effort targets the simulation of epidemic evolution: starting from SARS (Colizza, Barrat et al. 2007a), to the new virus of Influenza A (H1N1) (Balcan, Hu et al. 2009). Many different simulations, at global level or at local level aim at providing precise forecast on the number of infected, with the actuation of different containment strategies. One can expect that such tools become more and more commonly used to support political decisions.

Many models consider populations with an explicit representation of each individual or of the household structures. These individuals are characterised by some state variables (e.g. age, profession, marital status), and often a spatial position. Two main types of modelling approaches can be identified in the literature - Microsimulation and dynamical Individual Based Models (IBMs)<sup>8</sup>: The microsimulation approach defines individual economic and social trajectories through a set of events which occur with given probabilities, generally neglecting interactions between individuals. It provides a mechanism to analyse the effects of policy changes at the level of the decision making units as individuals and households. Individual based models (IBMs) consider the same type of population but generally include more elaborated models of decisions and actions, where individuals take into account the interactions with their environment and other individuals.

In both cases, the dynamics of the whole system is given by the aggregation of all individual behaviours. Hence these modelling approaches are often used to explore the link between the micro and macro dynamics. For instance models of evolving human populations yield demographic patterns in geographical space, which can be compared with census-based data (Mahdavi, O'Sullivan et al. 2007).

In both approaches, the first step for the simulation is to initialize the system with a realistic population: the state variables defining the agents or the individuals, must replicate, as closely as possible, the statistical properties of the targeted population. In particular, the demographic evolution must take into account the structure of the distribution of households. Indeed, for the same age structure of the population, different household structures evolve differently.

---

<sup>8</sup> Sometimes they are also called "agent based models", because the individuals represent economic or social agents. But there is an ambiguity with a different research trend of "agent based models", more related to computer science, which investigates computer agents that cooperate for achieving some tasks, for instance foraging on the internet. To avoid this ambiguity, we prefer to use the expression "Individual Based Models", which originally comes from modelling in ecology.

If individual data were available about the household structure, the problem would be solved quickly by creating a one to one correspondence between the agents and the real persons. However, such a situation rarely occurs, because the institutes managing statistics usually provide aggregated datasets, describing the global properties of the households and individuals. Therefore we must use these aggregate data to generate the artificial sets of individuals and households.

This paper focuses on the specific case of generating a population distributed in households to initialise a dynamical microsimulation model for the PRIMA project. PRIMA – Prototypical Policy Impacts on Multifunctional Activities in Rural Municipalities – is a European project (FP7-ENV-2007) which aims to model the impact of European policies on land use at municipality level in a set of case study regions. Hence in this project, the microsimulation process represents a population of individuals at municipality level, living in households of different types. Once generated, the initial synthetic population evolves through different processes such as birth, death, marriage, divorce, leaving parental house, getting a job and retirement. The quality of the final results depends heavily on the accuracy with which the initial synthetic population represents the available real data.

According to the literature, there are two approaches commonly used to create a synthetic population. In the first approach, some data at individual level are used to create the synthetic population. For instance in the SVERIGE model (Holme, Holme et al. 2004), the whole population of Sweden in 1990 is the starting population, and large longitudinal data sets are used for estimation of many equations for the demographic process. In a similar way, DYNAMOD (King, Baekgaard et al. 1999) is a dynamic model designed to project population characteristics over a 50-year period, using a 1% sample. A second approach uses the Iterative Proportional Fitting (Norman 1999) to estimate the joint probability of characteristics belonging to different sets of aggregated data. This approach is used in the SMILE model (Ballas, Clarke et al. 2005) where the synthetic population is generated from Census of Small Area Population Statistics (SAPS) in 1996 in Ireland, considering characteristics as gender, age, employment status and industry, for a given group of the population in a specific location. IPF can be applied when the Census data, describing the aggregate properties of individuals and households, are integrated with individual data, extracted by surveys on samples that can be bigger or smaller than the size of the desired artificial population. Thus, the initialization process consists in finding the good weight to attribute to each sub-element of the analyzed sample to make it representative of the objective population. Some methods to solve the up-scaling or downscaling initialization problem, with stochastic and deterministic approaches, are described in (Williamson, Birkin et al. 1998; Birkin, Turner et al. 2006; Ballas, Clarke et al. 2007; Ballas, Kingston et al. 2007b; Smith, Clarke et al. 2009).

In our problem, individual data to cross with the aggregate properties are not available. This situation does not allow us to apply the IPF method. Moreover computing the joint probability of characteristics of households, including size, type and age of members, implies heavy computations. In this paper, we propose an iterative semi-stochastic algorithm, involving a sequence of stochastic extractions, which considerably decreases the computational cost of the population generation. This algorithm uses only aggregated

datasets from Census, and the missing crossings between the data are obtained through testing procedures.

The algorithm is adjusted for data from the Auvergne region (France), but the general concept can be easily adapted to different uses. The next section describes the details of the problem to solve. Section 3 describes the available data in Auvergne region, as well the attributes of the synthetic population to be generated. The iterative algorithm is described in detail in section 4. Sections 5 and 6 present the results and conclusions.

## Materials and Methods

### General Formulation of the problem

The classical generation approach only considers one micro level (individuals or households). The specificity of this work is that we need to respect statistical constraints on the distribution of the individual ages, the distribution of household size and the distribution of individual ages within households.

More precisely our problem is to generate a set of households comprising individuals taken in a distribution of age of the population, and which respect all the constraints we found in the data about the distributions of:

- sizes and types of households,
- ages of the head of the household,
- differences of age between partners,
- ages of children according to mother's age.

Let us call:

- $t$  the type of household, the values of  $t$  can be: 'single', 'couple', 'single-parent', 'complex';
- $s$  the size of the household, the values of  $s$  can be: 1, 2, 3, 4, 5, >5;
- $a_r$  the age of the head of the household;
- $a_1, \dots, a_{s-1}$  the age of the children of single-parent households;
- $a_{r'}$ , the age of the head's partner, and  $a_1, \dots, a_{s-2}$ , the age of the children for couple households
- $(a_i)$  generally represents the list of the ages of the household members.

In a first approach we would suppose that we are able to compute a good approximation of the probability of a given household  $P(t, s, (a_i))$  (a possible method to compute these probabilities is described in section 2.3). Then, a straightforward way to proceed is described in algorithm 1.

#### Algorithm 1:

1. Generate all possible households, considering all possible combinations of types, sizes and ages of members;
2. Associate with each of these households, defined by the values of  $(t, s, (a_i))$ , the probability  $P(t, s, (a_i))$ ;
3. Generate a void list  $H$ . Repeat, until the size of  $H$  reaches the expected number of households:

- a. Pick a household generated in step 1 according to its probability associated in step 2;
  - b. Add the household to list  $H$ .
4. Return  $H$ .

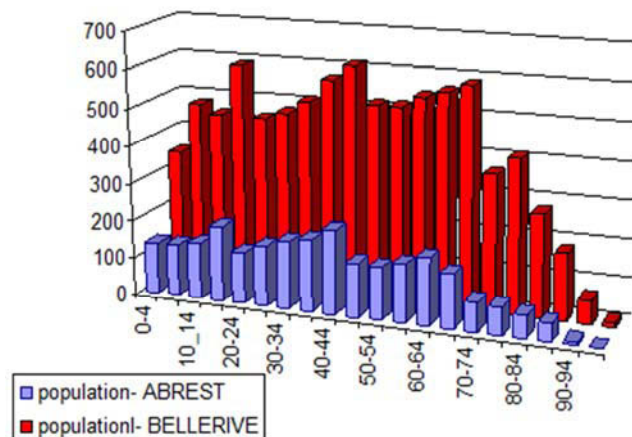
This algorithm shows a significant drawback. Although the average on a large runs of this algorithms of the distribution of age will be close to the data, one can expect significant differences between the age distribution of a specific run and the data. Since the data about the distribution of ages are reliable in our problem, we would like to keep it as precise as possible in our approach.

This leads to algorithm 2, where we use the list of ages of individuals directly taken from the data, and a probability of household  $P'(t, s, (a_i))$ , independently from the distribution of ages in the population:

**Algorithm 2:**

1. Generate a population of individuals following the age structure of the population. Let us call it the list  $I = \{a_i\}$  (to each element of the list an age is associated);
2. Generate all possible households, considering all possible combinations of types, sizes and ages of members;
3. Associate with each of these households, defined by the values of  $(t, s, (a_i))$ , the probability  $P'(t, s, (a_i))$  of the household, independently from the age distribution of the population;
4. Generate a void list  $H$ . Repeat, until list  $I$  is void or a number  $N$  of iterations is reached:
  - a. Pick a household  $h$  generated in step 2 according to its probability  $P'(t, s, (a_i))$ ;
  - b. If ages  $(a_i)$  are included in  $I$  then remove them from  $I$  and copy household  $h$  in  $H$ .
5. Return  $H$ .

With algorithm 2, we guarantee to keep the final distribution of individual ages close to the data. Generating the list of individuals following the age structure of the population is straightforward. The Census data of 1990 (INSEE 1990-2002 ; 2003-2007), the starting point at which we initialize the model for the Auvergne region, chosen in the PRIMA project as a pilot region to be studied, provides the age distribution for the population at municipality level. Two municipalities are chosen to test the algorithm: Abrest, which was composed by 964 households with a total population of 2545 individuals, and Bellerive-sur-Allier, composed by 8530 individuals organized in 3520 households. The choice of these municipalities was made arbitrary, considering the difference of sizes, for testing the algorithm.



**Figure 1:** Histogram of the number of individuals according to various age ranges of 5 years each for Abrest and Bellerive-sur-Allier. Source: INSEE, French Census data, 1990.

These data, displayed in Figure 1, allow us to generate directly the list  $I$  of individuals following the age structure of the population. Simply, going through all the age brackets, and for each one, we add to the list the corresponding number of individuals.

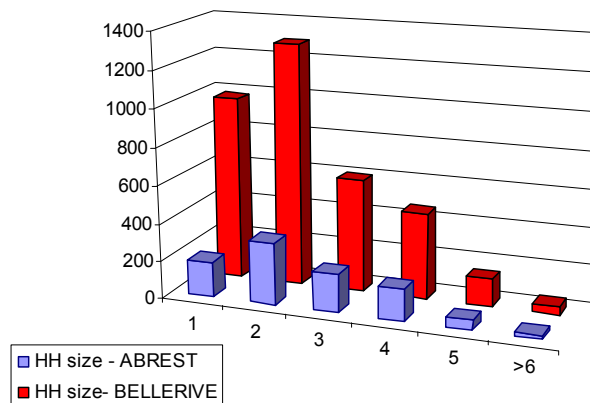
However, the other steps of the algorithms involve several difficulties:

- To evaluate the probability of a given household. This will be addressed in section 2.2.
- To manage the complexity of the set of all possible households. This will be addressed in section 2.3.
- In general, the algorithm leaves some individual ages unused at the end, and generates a smaller number of households than expected (because of the impossibility to find the necessary individuals to fit the drawn households). This is also addressed in section 2.3

### Calculating the probability of a household

Census data, (INSEE 1990-2002 ; 2003-2007), provide also some information about households: the size distribution, the age distribution for people living alone (single households) and the age distribution of the head of the household. Figures 2 to 4 show those available data for the two municipalities. From those data, we can calculate the probability of each household.

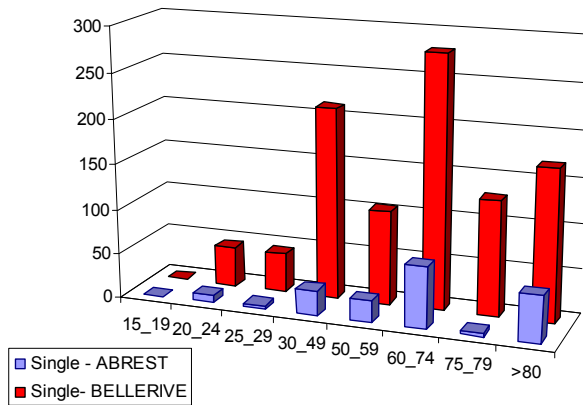
Data of figure 2 provide us with  $P(s)$ , the probability of having a household of size  $s$ .



**Figure 2:** Histogram of the number of households according to their size (number of individuals in the household) in Abrest and Bellerive-sur-Allier. Source: INSEE, French Census data, 1990.

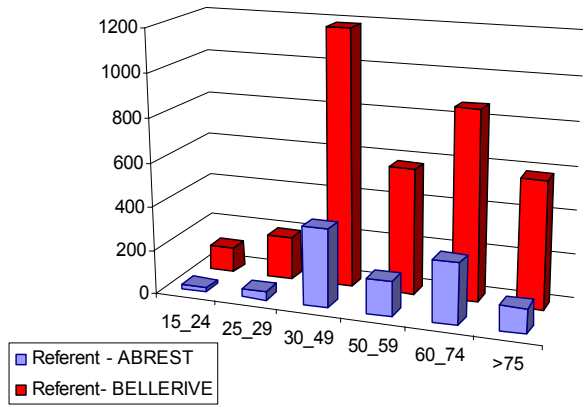
Data of figure 3 provide us with  $P(a_r | s=1)$ , the probability of age range of the head for households of size 1 (single).



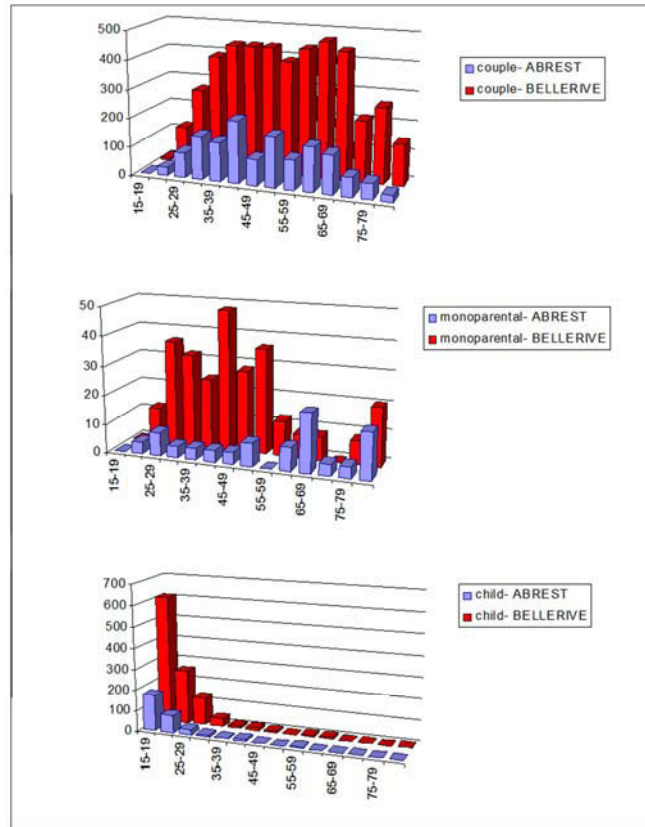


**Figure 3:** Histogram of the number of households according to the age ranges of person living alone in Abrest and Bellerive-sur-Allier. Source: INSEE, French Census data, 1990..

Data of figure 4 together with data of figure 3, provide with  $P(a_r / s > 1)$ , the probability of age range of the head for households of size superior to 1.



**Figure 4:** Histogram of the number of households according to the age ranges of the head in Abrest and Bellerive-sur-Allier. Source: INSEE, French Census data, 1990.



**Figure 5:** Histogram of the number of individuals of age > 15 according to different age ranges, from the top to the bottom, living as partners in couple, as head in single-parent households or living with parent(s) in Abrest and Bellerive-sur-Allier. Source: INSEE, French Census data, 1990

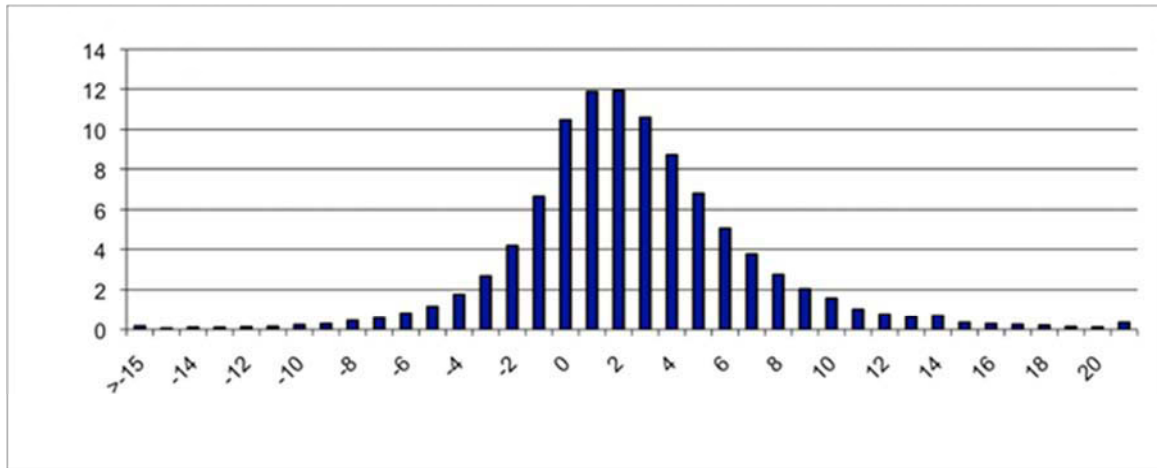
Data of figure 5 provide us with  $P(t | a_r = \alpha)$ , the probability of a household type given the age of the head equals  $\alpha$ , and the probability  $P(\text{child} | a = \alpha)$  for a individual of age  $\alpha$  to live in a household without being the head or the partner (this means, as a "child"<sup>9</sup>). Involving this constraint is very important to avoid to get households with very old parents (e.g. 90 years) and old children (around 70).

Clearly these data at local level are not sufficient to characterize a household. We lack constraints on the distribution of ages inside a given type of household. Hence we used some data at national level about the age structure inside the households regarding the ages of parents on one hand, and the ages of children on the other hand. Figures 6 and 7 show the national level data that we use to calculate the probability of the structure of ages, (INSEE 1990; Communities) 1999).

Data of figure 6 provide us with  $P(a_r | a_r = \alpha)$ , the probability of the age of the head's partner, given the age of the head.

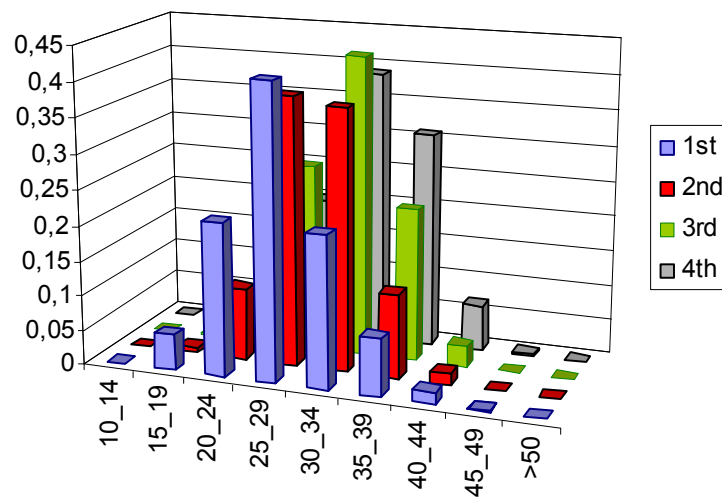
---

<sup>9</sup> That is the definition of "child" for the French Census managed by INSEE



**Figure 6:** Histogram of the number of couples according to their difference of ages in France in 1999. Source: INSEE, "Enquête sur l'étude de l'histoire familiale de 1999".

From data of figure 7, we can derive  $P((a_i) | a_m = \alpha, s = \sigma)$  the probabilities of children ages knowing the number of children and that the age of the mother is  $\alpha$ . We consider that in couple households, the mother is the partner, and in single-parent households, the head is the mother.



**Figure 7:** distribution of live births by birth order and mother's age range in France. Source: Eurostat Data 1999.

We can now use these partial probabilities to evaluate the probability of a given household  $P(t, s, (a_i))$ . We must distinguish cases 'single', 'single-parent', 'couple':

$$\begin{aligned}
 P'(t='single', s, a_r) &= P(s=1) * P(a_r | s=1) \\
 P'(t='singlep', s, a_r, (a_i)) &= P(s=\sigma) * P(a_r | s>1) * P('singlep' | a_r) * IIP(a_i | a_r) * P(child | a_i) \\
 P'(t='couple', s, a_r, a_r', (a_i)) &= P(s=\sigma) * P(a_r | s>1) * P('couple' | a_r) * P(a_r' | a_r) * IIP(a_i | a_r) * P(child | a_i)
 \end{aligned}$$

This evaluation theoretically allows us to apply the approach of algorithm 2. However, to generate all the combinations of households and picking one according to these probabilities is computationally expensive. In the next section, we propose an iterative algorithm which is more efficient computationally.

## **An iterative algorithm avoiding to generate all possible households**

The principle of the algorithm is to build progressively the household, by picking its member(s) according to the previously described probabilities, and, for each new member, to test if there is an individual of this age in the list of individuals  $I$ . If not, we stop the process for this household and begin to build another one.

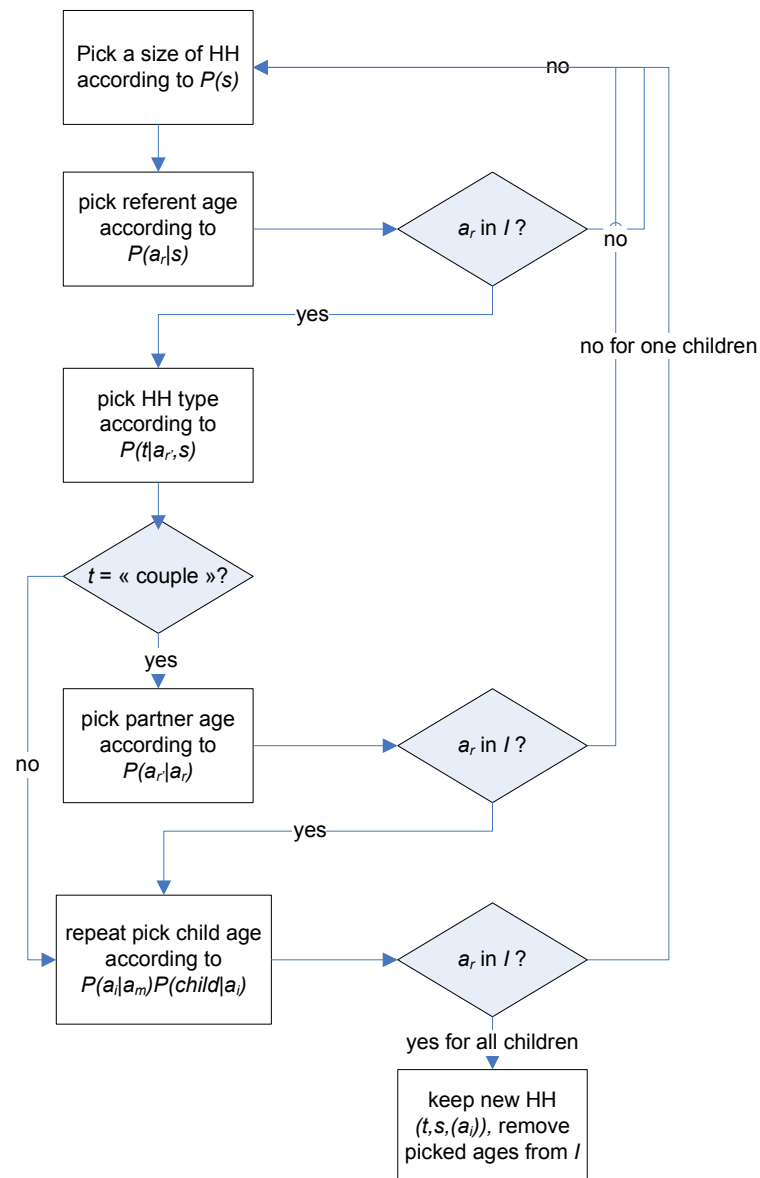
The flux diagram describing the process is represented in Figure 8.

This process is equivalent to pick one household according to its evaluated probability, and keeping it if all the ages of its members are present in list  $I$ . Indeed, the process of picking the different members of the household leads to the same overall probability to pick a household, and since the attempt is cancelled as soon as one age is lacking in list  $I$ , it changes nothing to make these tests iteratively.

Moreover, we can constrain even more the process by considering the list of household sizes which is directly derived from the data. The rest of the process remains the same. Then we are sure to have the right number of households, even though when algorithm 3 stops, some void households remain in the list.

Indeed, the described algorithm should *a priori* be repeated until all the households of the list are filled with all the individuals of the availability vector. However, this situation is never reached and after the creation of almost all the households, the program reaches a point where no more households can be achieved given the remaining individuals. For this reason, when this situation is reached, the algorithm is stopped. The remaining households can be considered as “complex structures”, namely all the housing solutions that cannot be placed into the usual categorization of household type (single, couple, single-parent). A complex household can be, for example, a group of students occupying the same housing or two familiar groups sharing the same location. Therefore, since we do not have any information about these structures from the data sets, to conclude the generation of the artificial population, the complex households are filled randomly with the remaining individuals in the availability list.

The algorithm consists of five main steps (see algorithm 3).



**Figure 8:** Flux diagram describing the algorithm for the generation of an artificial population for PRIMA project

### Algorithm 3

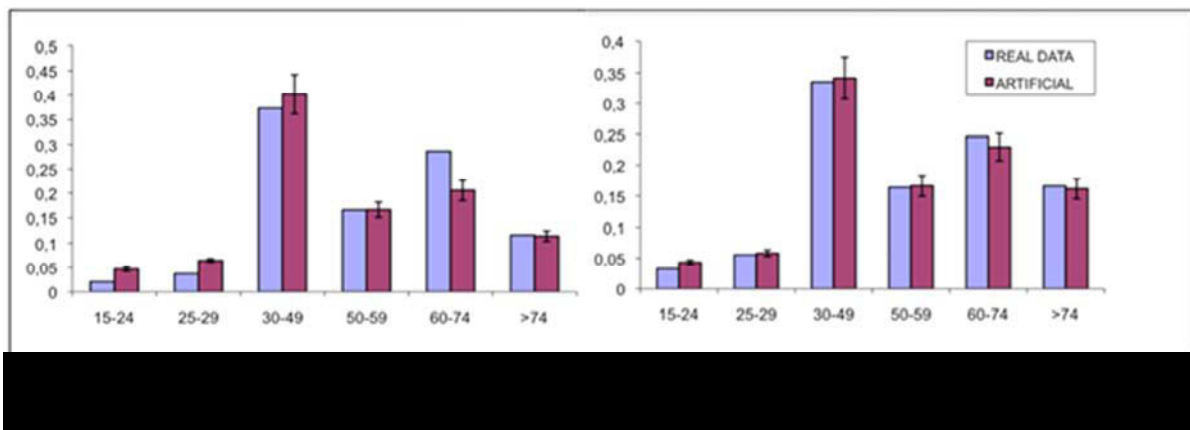
1. Pick the size of the household according to  $P(s)$ ;
2. Pick the age range of the head according to  $P(a_r|s)$ . If there is no individual in  $I$  of the age range, the process is stopped and a new attempt for building a household is launched. Otherwise an individual of the chosen age range is added to the household, and removed from list  $I$ ;
3. If  $s > 1$ , pick a household type ('couple' or 'single-parent') according to  $P(t|a_r)$ . "Complex" households are not considered at this stage.
4. If  $t = \text{'couple'}$ , pick the age of the partner according to  $P(a_r'|a_r)$ . Again, if there is no individual in  $I$  of the chosen age range, then the household is abandoned, the head is put back to list  $I$ , and a new attempt to build a household is launched. Otherwise an individual of the chosen age range is added to the household and removed from list  $I$ ;
5. We pick the age of children with probability  $P(a_i | a_r) * P(\text{child} | a_i)$  for single-parent and  $P(a_i | a_r) * P(\text{child} | a_i)$  for couples. Again, for each child, if there is no individual in  $I$  of the chosen age range, then the household is abandoned, its members put back to list  $I$ , and a new attempt to build a household is launched. Otherwise an individual of the chosen age range is added to the household and removed from list  $I$ .

## Results

We tested the algorithm for two different municipalities in Auvergne: Abrest and Bellerive-sur-Allier. The first one had a population of 2545 inhabitants in 1990, while the second one had 8530. In the following we compare the statistical properties of the artificial population with the real Census data. We use for the comparison both the data implicitly used in the building algorithm and other national and municipality level data, which were not used in the generation process. We calculate the distributions both for one single realization of the system and for a sequence of 100 realizations (the random nature of the algorithm leads to some variations from one run to the other).

By construction of the algorithm, the age distribution and the size distribution of the household are directly derived from the data for the two villages. In Figure 9 we show the distributions of the age of head for real data and the artificial population. The distribution of the age of head was used inside the generation process, but the stochastic extractions from this distribution were spaced out from various tests; for this reason we can expect some discrepancy between the real data and the generated population.

As we can observe in Figure 9, the artificial population respects quite well the real distribution.

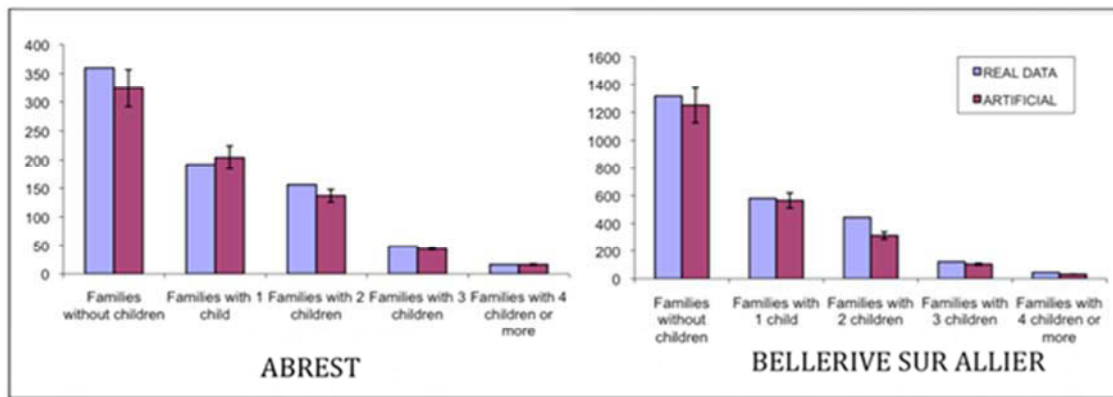


**Figure 9:** Histograms for age of head distribution for the municipality of Abrest (left plot) and of Bellerive-sur-Allier (right plot). The light purple bars represents the real data, the dark purple bars the average for 100 realizations for the artificial population. The error is the standard deviation on the 100 replicas.

In Figure 10, we compare the obtained artificial population with the real distribution of number of children in households. This particular data set was not used in the generation, so the comparison can give an idea of the accuracy of the algorithm; this data set is reported in Table 1. Also in this case we can observe a good agreement between the real data and the simulations.

**Table 1:** Distribution of households according to the number of children for the municipalities Abrest and Bellerive sur Allier. Source: INSEE, French Census data, 1990.

Type	ABREST	BELLERIVE
Household without child	360	1316
Household with one child	192	580
Household with two children	156	444
Household with three children	48	120
Household with four or more children	16	44

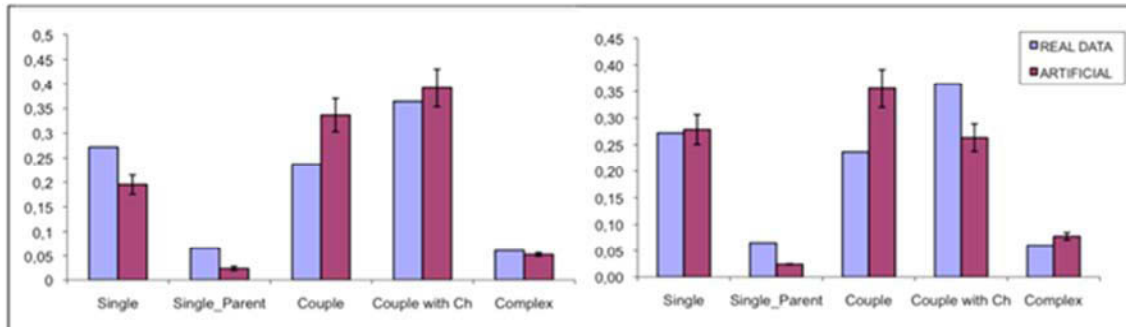


**Figure 10:** Histograms for age number of children distribution for the village of Abrest (left plot) and of Bellerive-sur-Allier (right plot). The light purple bars represents the real data, the dark purple bars the average for 100 realizations for the artificial population. The error is the standard deviation on the 100 replica.

The final comparison (Figure 11) regards the household typology. For this comparison we will not use directly the data that we have used in the generation (the probability for a person to be in a certain type of household) but another dataset containing the direct proportions of household types at national level. This dataset is reported in Table 2.

**Table 2:** Distribution of households according to the type in France. Source: INSEE, 1990.

Type	Proportion
Single	0,2720
Single-parent	0,0660
Couple	0,2370
Couple with Children	0,3640
Complex	0,0610



**Figure 11:** Histograms for the household type distribution for the village of Abrest (left plot) and of Bellerive-sur-Allier (right plot). The light purple bars represents the real data, the dark purple bars the average for 100 realizations for the artificial population. The error is the standard deviation on the 100 replica.

In this case the differences from the real data, for both municipalities, are quite significant. It could be expected: the data we are using in this case for the comparison are at national data, and therefore keep into account of the population of metropolitan areas. The discrepancy between our results and the national data, therefore, do not highlight an error in the generating process, but show the behavioural difference between metropolitan area and rural villages, with small population.

Moreover, it is noticeable that the data reported in the previous graph provide relevant information about the complex households. We lack completely this information at village level and therefore we cannot use any constraint on complex households in the building procedure. In the proposed algorithm, complex households are created randomly, grouping together the individuals that the generating procedure cannot assign to a household according to the selection/test mechanism. Nevertheless, we observe that the proportion of complex households that we obtain is close to the data at national level.

Finally, we need to stress that this kind of algorithm is strictly correlated to the data structure we have: for Auvergne region such as for France and most of occidental countries the main household structures are based on the concept of “nuclear family”: a couple of parents and a certain number of children, or a subset of this structure. In some other cultures the basic household can have completely different structure (for example many generations sharing the same housing), and therefore this kind of approach can give rise to potential bias without any additional information about the structure of complex households.

## Conclusion

In this paper we proposed an algorithm for the generation of a synthetic population organized in households that can be applied in various modeling contexts. This method gives good results without using a set of prototypical households that, in many cases, are not available. This is an advantage compared with existing methods such as IFT. Moreover it allows one to reproduce exactly some features of the real population that are particularly important for the subsequent analysis.



This algorithm is a practical implementation of a general approach where the households are picked according to their probability, among all the possible household structures. The method builds the households iteratively. It tests the availability of the age of its members at each step, and backtracks as soon as an age is lacking. This saves a lot of computations.

We presented the example of the PRIMA project, where the artificial population is needed as initialization of a dynamical microsimulation model at municipality level. We showed that the algorithm yields a good agreement between the statistics of the artificial population and the real one. Clearly, the approach can be adapted to other cases where it is necessary to generate a population organized in households. During the project, we shall have to adapt it to other sets of data that can be found in different case study regions.

The algorithm can deal with other properties of the individuals and of the households. For instance, we could add a gender variable to describe the individuals of our example. We would need to split the list *I* of individuals of different ages into two lists, one for males and one for females. Moreover, we would need to include the percentage of household where the head is a male and about the percentage of heterosexual couples. Then the principle of the method remains the same. The only difference is that to build the households, we pick either in the list of males or in the list of females.

More generally, after the set-up of the demographical structure, other characteristics can be assigned to each individual, through stochastic extractions or deterministic associations: the level of instruction, the professional activity, the favorite recreational activities, the commuting pattern, etc. According to the available datasets, these properties can be assigned to each individual independently from the household in which it is embedded, or some correlations can be considered inside the same household.

### **Acknowledgment**

This publication has been funded under the PRIMA (Prototypical policy impacts on multifunctional activities in rural municipalities) collaborative project, EU 7th Framework Programme (ENV 2007-1), contract no. 212345.

### **References**

- Duygu Balcan, Hao Hu , Bruno Goncalves Paolo Bajardi , Chiara Poletto, Jose J Ramasco , Daniela Paolotti , Nicola Perra , Michele Tizzoni , Wouter Van den Broeck , Vittoria Colizza and Alessandro Vespignani 2009, Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC Medicine* 2009, 7:45
- Ballas D., Clarke G.P., Wiemers, E. 2005. Building a dynamic spatial microsimulation model for Ireland. *Population, Space and Place* 11, 157-172.
- Ballas D., Clarke G.P., Wiemers, E. 2006. Spatial microsimulation for rural policy analysis in Ireland: the implication of CAP reforms for the national spatial strategy. *Journal of Rural Studies* 22, 367-378.
- Ciofi degli Atti M.L., Merler S., Rizzo C. , Ajelli M., Massari M., Manfredi P., Furlanello C., Scalia Tomba G., and Iannelli M., 2008. Mitigation Measures for Pandemic Influenza in Italy: An Individual Based Model Considering Different Scenarios. *PLoS ONE*. 2008; 3(3): e1790.
- Colizza V., Barrat A., Barthélemy M. and Vespignani 2007 A. Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study. *BMC Medicine* 2007, 5:34
- Colizza V, Barrat A, Barthelemy M, Valleron A-J, Vespignani A. 2007B Modeling the World-wide Spread of Pandemic Influenza: Baseline Case and Containment Interventions. *PLoS Med* 4(1): e13. doi:10.1371/journal.pmed.0040013.

- Eubank S, Guclu H, Kumar VSA, Marathe MV, Srinivasan A, Toroczkai Z, Wang N. 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* 429:180–184.
- Gotts N.M., Polhill J.G., Law A.N.R. Aspiration levels in a land use simulation. 2003. *Cybernetics and Systems*, n. 34, p. 663-683.
- Holm E., Holme K., Makila K., Mattson-Kauppi M., Mortvik G. 2004. The SVERIGE spatial microsimulation model: content, validation and example applications. Spatial Modelling Centre, Sweden. Technical Report. 55p.
- INSEE (National Institute of Statistics and Economic Studies), 1990. French Population Census of 1990. Data distributed by the Centre Maurice Halbwachs, 48 bd Jourdan, 75014 Paris.
- INSEE (National Institute of Statistics and Economic Studies), 1999. "Enquête sur l'étude de l'histoire familiale de 1999". Data distributed by the Centre Maurice Halbwachs, 48 bd Jourdan, 75014 Paris
- Eurostat (Statistical Office of the European Communities), 1999. Demography, fecundity data 1999 [epp.eurostat.ec.europa.eu/portal/page/portal/population/data/database](http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/database)
- King A., Baekgaard H., Robinson M. 1999. Dynamod2: an overview. Technical Paper 19, NATSEM, University of Canberra.
- Mahdavi B., O'Sullivan D., Davis P. 2007. An agent-based microsimulation framework for investigating residential segregation using census data. In: Oxley, L. and Kulasiri, D. (eds.) MODSIM 2007 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2007, pp. 365-371.
- Nagel K., Beckman R.L., Barrett C.L., 1999. TRANSIMS for transportation planning. In 6th Int. Conf. on Computers in Urban Planning and Urban Management
- Norman P. 1999. Putting Iterative Proportional Fitting on the researcher's desk. Working Paper 99/03, School of Geography, University of Leeds, United Kingdom. 36p.
- Ballas D, Clarke G, Dorling D, Rossiter D (2007) Using SimBritain to model the geographical impact of national government policies. *Geographical Analysis* 39: 44.
- Ballas D, Kingston R, Stillwell J, Jin J (2007) Building a spatial microsimulation-based planning support system for local policy making. *Environment and Planning A* 39: 2482–2499.
- Birkin M, Turner A, Wu B. A synthetic demographic model of the UK population: methods, progress and problems; 2006.
- Smith DM, Clarke GP, Harland K (2009) Improving the synthetic data generation process in spatial microsimulation models. *Environment and Planning A* 41: 1251–1268.
- Williamson P, Birkin M, Rees PH (1998) The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environment and Planning A* 30: 785–816.



# Chapter 1.3 Parametrisation of the individual activity dynamics

Authors: Huet S., Lenormand M., Deffuant G., Gargiulo F.

Published in: accepted as a chapter for “Empirical Agent-Based Modeling: Parameterization Techniques in Social Simulations”, A. Smajgl and O. Barreteau eds. To appear probably in 2013

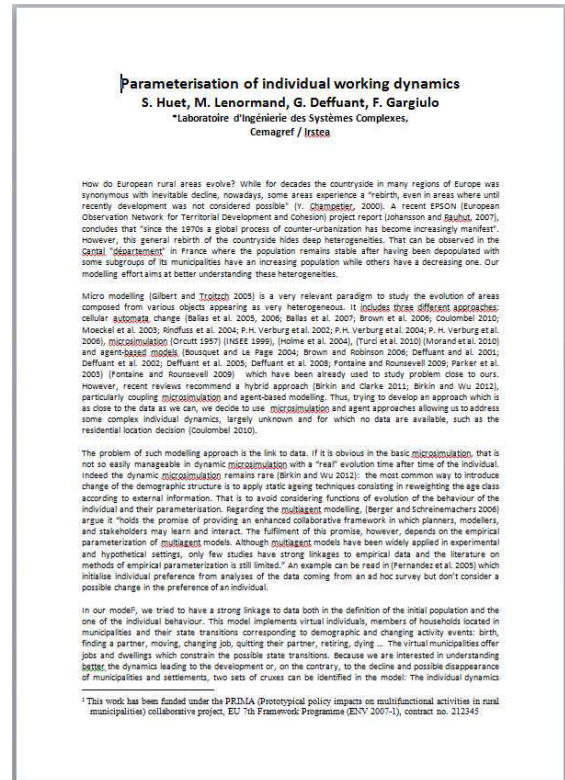
*The introductive part and the presentation of the model have been suppressed from the present chapter since they are presented in other sections of this these.*

## Abstract

This chapter explains how dynamics of the conceptual model presented in the chapter 1.1 can be parameterised from data for an implementation for a French region. It focusses on the dynamics of work statuses and professions which can be derived from the European Labour Force Survey (LFS). The individual is considered from her entering on the labour market until the moment she decides to retire. Meanwhile, she can be fired, searching for a job when she is unemployed or already employed but also deciding becoming inactive for a while. Jobs are defined by a profession and an activity sector. The job offer is located at the municipality level and can correspond to the searched profession of an individual having a municipality of residence. The French implementation is described starting from the employment survey in France, which is the French declination of the LFS.

## Contents

Designing and parameterising the individual activity.....	61
Data sources and main modelling choices.....	61
Defining the initial individual labour attributes.....	63
Defining the individual behavioural rules regarding activity.....	66
Lessons / Experience .....	75
Acknowledgements.....	77
References.....	78





## Designing and parameterising the individual activity

This part focuses on the design and the parameterisation of the individual activity. The purpose is to illustrate how to model in a micro simulation approach individuals' behaviour on a labour market utilising existing data. The European project that funded this work did not fund specific interviews or surveys for this purpose. But, even if such funding had been available, it would have been difficult to have a sufficiently large sample to ensure the statistical significance of the obtained attributes and behaviours. Therefore, it seemed better to use existing large database dedicating especially to the labour force, such as the labour force survey, which gives information on the labour force based on a very large sample and the weights for projection at various levels. Moreover these databases, developed by the National Statistical Office, have been built on a data collection model designed by experts. They represent common knowledge, largely shared by every stakeholder since they are used as references in decisions and predictions.

We start from existing databases and the objectives of the modelling to characterise our agents and their attributes and behaviours. That is what we discuss in the following first subsection. The two following subsections give details on the initialisation of the attributes and on the parameterisation of the behaviours. The link between attributes and behaviours is guaranteed as this data is implemented to ensure its compatibility with the agent attribute modalities. Similarly, the projection of attributes and behaviour for the whole virtual population is easy: an innovative generation population algorithm builds directly a robust and significant population of individuals while the link between modalities of attributes and their evolving rules allows an automatic projection at the population level.

### Data sources and main modelling choices

This is to identify the agent classes and the structure of agent behaviour in each class. The first steps have been:

- to collect all relevant data source regarding the region we want to simulate considering the exact problem (aim of the project) we need to address;
- to make a state-of-the art;

From the literature and the expertise coming mainly from economists, we identify two complementary groups of dynamics to take into account to model the evolution of a local labour market:

- Job offers and corresponding dynamics;
- Job demand and occupation, and corresponding dynamics.

We identify two possible databases to help us conceptualising and parameterizing the model:

- The Census: it gives indications about the situation of individual when being student, retired, or active and also who is occupied and who is not occupied, what occupations individual have aggregated in socio-professional categories and activity sectors; Census data are available at the municipality level for three different dates

1990, 1999 and 2006. We can also benefit from the mobility tables of the Census giving, at least in 1999, an exhaustive description of the commuting flows between municipalities; French Census data are also available for 1982 but not electronically;

- Labour force survey (from 1990) and census data;

From literature and data, we have to define agents:

- corresponding to the local level of offer: the **municipality**
- corresponding to the job demand and occupation: the **individual** is the one who is going to search for a job, deciding if and where she searches taking into account the **household** of which she is a member and her *municipality* of residence.

Then we have a municipality offering jobs, composed from households, themselves composed of individuals who decide, considering their household, if and where they are going to search for a job. A job can be found in a municipality and individuals accept jobs based on the distance.

Other available data sources include SIRENE and UNEDIC. The SIRENE database includes information on the number of societies by activity sector. The UNEDIC database includes the number of paid employees by activity sector. But both these data sources describe only a part of our problem and start only in 2000 while the simulation requires longer periods to allow for a proper calibration of the model.

The incompatible coverage also constrains the choice of agents and their attributes. However, given the available datasets we decide to start simulations in 1990. On the one hand, it means the parameterisation of some attributes is less robust than with shorter calibration periods. A later start would allow us to use the supplementary information given in more recent surveys and not available in older surveys. For example, we use only four modalities of size to describe the size of housings because only four are available in 1990 while five and more are recorded in later surveys. On the other hand, the 1990 census data gives us the cross distribution socio-professional categories x sector of activities we use to define the jobs while this cross distribution is not available later. Then, we can and have to use IPF to define the job offer after 1990 starting from the 1990 cross distribution.

The definition of a job is directly driven by the available data. Both Censuses and Labour Force Survey (or Employment survey) describe jobs with profession (socio-professional category) and activity sector. Both also contain data on age and situation (student, retired, actives, occupied or not, inactive) allowing us to make a connection between both sources of data. Moreover, when the data sources are “official”, it often corresponds to the common knowledge of stakeholders and other decision makers.

Moreover, as a general modelling good practice, it is particularly important to minimise the number of unknown parameters. Indeed, every parameter which is not derived from the data has to be calibrated. The calibration computational cost increases with the number of parameters. Moreover, the more numerous are the parameters to calibrate, the less relevant also is likely to be the model which, given its large number of freedom degrees, can produce almost any trajectory.

## Defining the initial individual labour attributes

The main source of information to define attributes and their values is Census data. The French Census is available for 1990, 1999 and 2006. The 2006 Census has to be used with caution since it is different from 1990 and 1999. It is now a continuous survey which interviews a part of the population every year. Municipalities having less than 10000 inhabitants are exhaustively surveyed by 1/5 every year. Larger municipalities have sample surveyed every year. In both cases, INSEE, responsible for the Census, give the information allowing the projection at the population level every year. A very good point is that the access to data is easy and free<sup>10</sup>.

To compute a population with sufficiently realistic local statistical properties for individuals and households, we propose an algorithm described in (Gargiulo, Ternès et al. 2010) presenting the generation of households in the Auvergne Region. An improved version has been developed for generating the Cantal population. To summarize our algorithm, we build for each municipality a list of agents with the exact number of individuals being each age and a list of households with the exact number of household members. Then, we try to fill one by one each household with individuals taking into account the probability of households having some particular properties, such as being a couple or having a given number of children. Each time a household is completed, another one is selected to be filled. At the end, we have a virtual population of households following the exact distribution of sizes, having good statistical household properties and composed from individuals following the exact distribution of ages. To build the initial population of Cantal, our algorithm uses for each municipality:

- The distribution of the size of households – available at the municipality level in 1990
- The distribution of ages of individuals – available at the municipality level in 1990
- The distribution of ages of the reference person of households – available at the municipality level in 1990
- The distribution of household types (single, couple, couple with children, single-parent, other) - available at the municipality level in 1990
- The distribution of age differences for couples – only available at the national level in 1990
- The distribution of the probability to be a child (i.e. living at parental home) by age and for each household type – available at the municipality level in 1990

This generation method is different from the nowadays used IPF (Iterative Proportional Fitting) which reweight a measured population under some constraints to obtain a virtual population representing the one the modeller is interested in. However this method cannot control the attributes at the two levels, the person and the household. Some recent work proposed a hierarchical IPF (Müller and K.W. 2011) to control the two levels but they still

---

<sup>10</sup> made available by the Maurice Halbwachs Center of the Quételet Network (<http://www.reseau-quetelet.cnrs.fr/spip>) for 1990. For 1999 and 2006, they are directly accessible through internet via the website of INSEE <http://www.recensement-1999.insee.fr/> and [http://www.insee.fr/fr/publics/default.asp?page=communication/recensement/particuliers/diffusion\\_resultats.htm](http://www.insee.fr/fr/publics/default.asp?page=communication/recensement/particuliers/diffusion_resultats.htm)



required an initial sample, which can be reweighted to fit the scale the model is interesting in.

After the virtual population has been built, individuals require a labour market status. That means the following four individual attributes have to be parameterised during the initialisation: Activity status; Profession, approximated by the socio-professional category; Sector of activity to define, with the profession, the occupied job; Place of work.

To characterize the status we distinguish between active and inactive individuals. Active people can be employed or unemployed. For non-active people we distinguish three categories: students, retired and other. No further characterization is required for non-active person. On the contrary, active people, both employed and unemployed require a socio-professional category (SPC) defining their profession. Moreover, employed individuals require a sector of activity defining the occupation (i.e. a SPC (proxi for profession in a sector of activity)). Once the municipality of employment is determined, the employed individual is successfully parameterized.

Figure 2 shows the generation algorithm. The initialization of the activities starts from the population of households previously generated for each village: each person is assigned an activity, according to the characterization presented above. All the individuals younger than 15 are automatically considered students. For all the others the first step is the decision about being active or not. This decision depends on the age of the person. If the person is not active then her age determines whether she is retired or a student. If she is neither student nor retired, she will be identified with the status "inactive". If the person is active, the first step is the selection of the socio-professional category (SPC). This choice depends on the age. Secondly it is decided whether the person is employed or unemployed, according to the age. If she is unemployed, no further choices are needed. If she is employed, the municipality of employment is determined. The municipality of employment depends on two questions: first, does she work inside her municipality of residence? If no, find at random a place of work among the possible places of work starting with her own municipality of residence if employment is available according to the SPC. The possible places of work are defined through a generated virtual network built from the mobility data of the French Census of 1999 (see the generation model proposed in (Gargiulo, Lenormand et al. 2011) and improved in (Lenormand, Huet et al. 2011)). Finding a possible place means the individual can find a free job partly defined by the same SPC as hers. A vector for available jobs is maintained (corresponding to the total number of commuters-in at the beginning of the initialisation) for each municipality and decreases with individuals filling vacancies. If no vacancies remain among the possible places of work while an individual is still looking for employment, the attribution of a place of work among the possible ones is forced. Indeed, this can occur due to the fact the generated virtual network is built under the only constraints related to the job demands and the job offers of each municipality. The virtual network doesn't consider the SPC then it can't ensure a demand with a particular SPC can be satisfied by an offer with this SPC in the set of municipalities it has fixed as possible places of work. Finally, an activity sector is attributed to the employed individual based on the cross

distribution SPC<sup>11</sup>. We have to acknowledge that the French Statistical Office, as many Statistical Offices, use two ways to count the jobs: counted on the place of residence – that means corresponding to the job occupation by people living in a municipality wherever they work; and counted on the place of work – that means counted on the municipality where people work wherever they live. The algorithm uses the following data for each municipality of the set:

- Age x activity status counted on the place of residence
- Age x SPC for actives counted on the place of residence
- Distribution of probabilities working inside her place of residence by SPC
- A generated commuting network through (Gargiulo, Lenormand et al. 2011) (Lenormand, Huet et al. 2011) given for each municipality the distribution of commuters out to each of the other municipality<sup>13</sup>
- SPC for actives x activity sector counted on the place of work

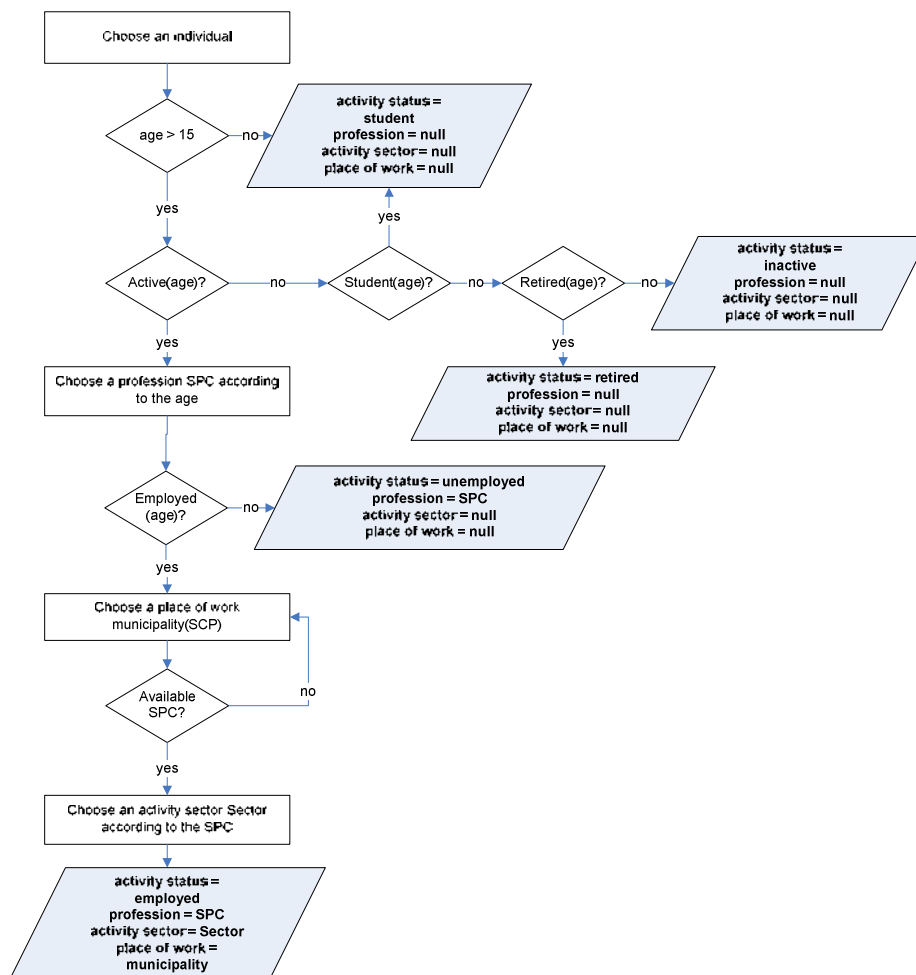


Figure 2. Algorithm for the initialization of the activities for Auvergne case study

<sup>11</sup> The initialisation was done in this way when the paper has been written. But this method is now deprecated since the initial the place of work of commuters is now determined through the job search method of the model (see chapter 11 for more details about this method)

## Defining the individual behavioural rules regarding activity

This part is dedicated to the parameterisation of events on the labour market. Characterization and parameterization is required for those rules that change the value of the individual's attributes related to its labour activity: Activity status; Profession, approximated by the socio-professional category; Sector of activity to define, with the profession, the occupied job; Place of work.

The main data source to do so is the European Labour Force Survey, and particularly its French declination called in French "Enquête Emploi", meaning "Employment survey". The data are kindly made available for free by the Maurice Halbwachs Center of the Quêtelet Network<sup>12</sup>. This Employment survey was launched in 1950. It was redesigned in 1968, 1975, 1982, 1990 and 2003. From 1982, the survey became an annual survey. Since the last redesign the survey is implemented continuously to provide quarterly results. The resident population comprises persons living on French metropolitan territory. The household concept used is that of the 'housing household': a household means all persons living in the same housing. It may consist of a single person, or of two families living in the same housing.

As our approach starts the simulation in 1990 the first period is based on annual data while from 2003 on values can be considered in quarterly time steps (Givord 2003; Goux 2003). The data to select from these two periods vary a bit due to the structural and practical changes in the survey).

Coming back to the description of the whole data, the sample sizes of the data varies from 168883 to 187326 from 1990 to 2002 each year and from 92300 to 95647 each quarter a year for the new Employment survey. The individuals are asked a very comprehensive series of questions from 1990 to 2006, related to their work. In particular, we can follow their situation year by year, and also their wishes to change job and the type of job they are looking for. Table 2 shows the variables we extract from the databases to compute the probabilities we need. However, for the sake of simplicity, we use only data from 1990 to 2002 to explain how to extract the information we need from the data.

From the databases, we considered only the population being more than 14 that is not military people of students (FI = 3 and 4).

---

<sup>12</sup> <http://www.reseau-quetelet.cnrs.fr/spip/>

Table 1. Data to extract from the various databases of the French labour force Survey to compute the probabilities related to working status of the individual

1990 to 2002	2003	2004	2005	2006	2007	Meaning of the variable
<i>ag</i>	<i>Ag</i>	<i>Ag</i>	<i>ag</i>	<i>Ag</i>	<i>Ag</i>	<i>Age</i>
<i>annee</i>	<i>annee</i>	<i>Annee</i>	<i>annee</i>	<i>annee</i>	<i>annee</i>	<i>Year of interview</i>
<i>dcse</i>	<i>csepr</i>	<i>Csepr</i>	<i>csepr</i>	<i>csepr</i>	<i>csepr</i>	<i>Socio-professional category</i>
<i>cspp</i>	<i>cspp</i>	<i>Cspp</i>	<i>cspp</i>	<i>cspp</i>	<i>cspp</i>	<i>Socio-professional category of the father</i>
<i>dcsep</i>	<i>cser</i>	<i>Cser</i>	<i>cser</i>	<i>cser</i>	<i>Cser</i>	<i>Socio-professional category one year before</i>
<i>dcsea</i>	<i>cslong</i>	<i>Cslong</i>	<i>cslongr</i>	<i>cslongr</i>	<i>cslong</i>	<i>Socio-professional category which has been occupied for most of the time [for inactive and unemployed people]</i>
<i>tu99</i>	<i>tu99</i>	<i>tu99</i>	<i>tu99</i>	<i>tu99</i>	<i>tu99</i>	<i>Urban area type</i>
<i>fip</i>	<i>eoccua</i>	<i>Eoccua</i>	<i>eoccua</i>	<i>eoccua</i>	<i>eoccua</i>	<i>Occupation one year before</i>
<i>extri</i>	<i>extriA, extriA04</i>	<i>extri99, extri04, extri05,</i>	<i>extri05, extri06,</i>	<i>extri06</i>	<i>extri06</i>	<i>Weights making the interviewed individuals representative (depending on the census done 1999 or of the first result from the last French census (in 2004, 2005, 2006))</i>
<i>rg</i>	<i>reg</i>	<i>Reg</i>	<i>reg</i>	<i>reg</i>	<i>Reg</i>	<i>Region of residence</i>
<i>fi</i>	<i>sp00</i>	<i>sp00</i>	<i>sp00</i>	<i>sp00</i>	<i>sp00</i>	<i>Occupation during the month of interview</i>
-	<i>trim</i>	<i>Trim</i>	<i>trim</i>	<i>trim</i>	<i>trim</i>	<i>For the second period of the survey, the only keep the first quarter of the year.</i>
<i>csrech</i>	<i>csrech</i>					<i>Searched socio-professional category</i>
<i>dre1</i>						<i>Situation in regards to employment (mainly to use dre1=5 meaning people looks for a job (or another job))</i>
	<i>soua ; mrec</i>					<i>Wish another job; Is the individual has searched for a job during the last four weeks?</i>

### **Entering the labour market**

A first step consists of extracting the age from which on the individual is going to look for a job. This will determine the age at which a student status changes to a "on labour market" status. We consider in the period 1990 to 2002 the value FIP=3, which means that the individual was student the year before and the value FI=all the possible values except 3 means that the individual is not a student anymore. Then, for each five-year step we compute the probability to be a given age and having entered on the labour market for every year.

We used the weights to obtain a projection of the data at the Auvergne level. Auvergne is the region containing the Cantal "département" and three others. That is the closer significant and representative level of the Cantal. Then, we assume the probabilities are the same at the regional and the "département" level.

The second step is to allocate a first SPC (proxy used for defining the profession) to the individual allowing us to approximate what she is going to look for. We know that both these variables, the age of entry and the first SPC, are not independent. Moreover, a social determinism rules the choice of the profession by children compared to the profession of their parents. Figure 3 presents such a relation for the Auvergne population. It shows, for example, that almost only farmers' children become farmers or that executives' children mainly become executives and/or adopt an intermediary profession.

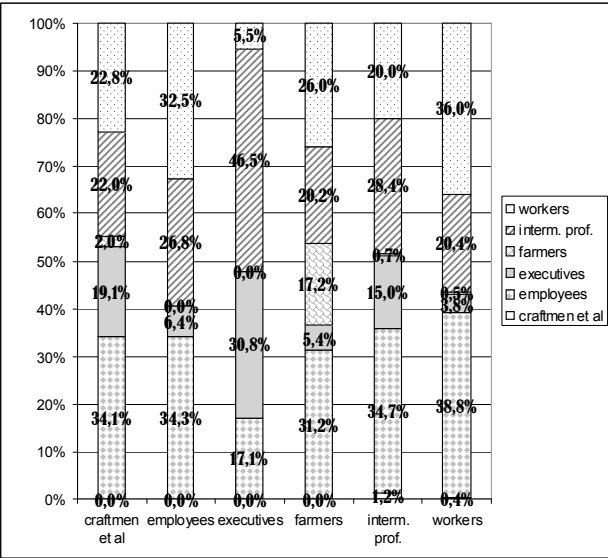


Figure 3. Distribution of SPCs choices by children regarding the father's SPC (in abscissa) for the Auvergne population. Source: French Labour Force Survey, 1990 to 2002 data.

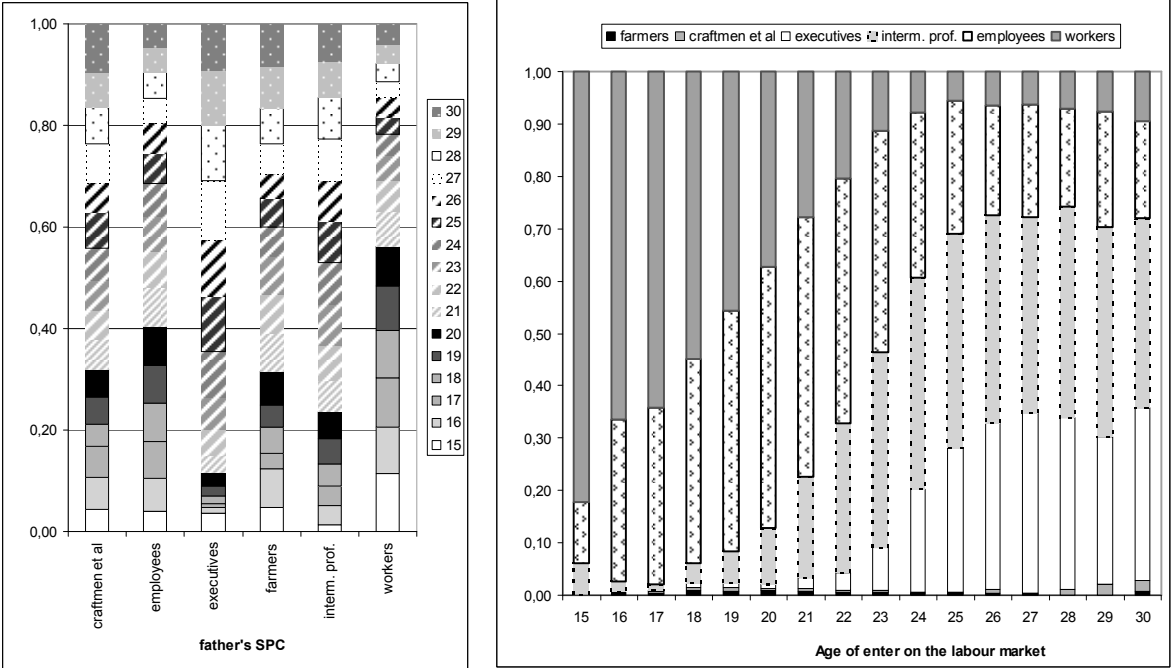


Figure 4. (a on the left) Probability of a "first" SPC depending on the age of entry in the labour market; (b on the right) Distribution of probability to enter the labour market at a given child age for each of the six father's SPC considered – French population. Source: French Labour Force Survey, 1990 to 2002 data.

Thus, starting from this social determinism, we have some indications to set the SPC of children. However, we also have to decide the age of entry in the labour market, and we know that this age is not independent from the level of education, which can be related to the SPC. Consequently, we apply a two-time process which, at first, decides the age at which to enter the labour market using the father's SPC and then determines the child's SPC depending on the age of entry.

The age of entry on the labour market is determined by the SPC of the father. Since the individual has no gender in our model, the father is randomly chosen between the two parents when there are two.

A criticism can be formulated to this approach since the SPCs of the couple members is not controlled, while we know from the literature that the partner is not chosen at random regarding her SPC. The homogamy can be explained by the constraint associated to the meeting places (Bozon and Héran 1987; Bozon and Héran 1988). It has been identified as a possible next step for modelling.

Figure 4a shows the distributions of probabilities to enter the labour market depending on the various ages of a child for each of the six SPC attributed to the father. We can for example read that if the father is an executive, the probability to enter on the labour market before 20 is only 0.1 while it is more than 0.5 if the father is a worker. Once our individual has an age to enter the labour market, we can determine her first SPC. Figure 4b shows for each age of entry on the labour market (abscissa) the distribution of probabilities over the possible SPC to provide the individual with a first SPC. For example, one can notice how high the likelihood of looking for a worker position for the individual looking at first for a job at 15 is, while at 30, she will mostly look for intermediary or executive positions. The individual who enters the labour market can decide looking for a job.

### ***Individual job searching decision***

We assume that the probabilities are stable in time for the Auvergne region. Thus, we mix the data from the years 1990 to 2007 in a single sample. Starting from the variables presented in the table 2, we count the frequencies of transitions between inactive, unemployed, employed, from one year to the following. For each counted transition, we take into account the weight of the related individual in order to have a probability quantified for the Auvergne level.

Finally, we calculate the probability to reach a given situation by dividing the total obtained for a transition starting from the situation  $x$  by the sum of all the totals related to the transitions starting from this same situation  $x$ .

We focus on the municipalities of the Auvergne region having less than 50000 inhabitants using the area type "tu99".

### ***From and to the inactive status***

The following variables are used to extract the transitions from a starting situation to an arriving situation. They are used for the transitions from and to the inactive status.

- $fi = 7$  plus  $8$  or  $EOCCUA = 6$  plus  $7$  to define the inactive status as starting situation;  $fi = 7$  or  $SP = 8$  to define the inactive status as arriving situation;
- $fi = 2$  or  $EOCCUA = 2$  to define the unemployed status as starting situation;  $fi = 2$  or  $sp00 = 4$  to define unemployed status as an arriving situation ;
- $fi = 1$  or  $EOCCUA = 1$  to define employed status as starting situation;
- $DCSP$  or  $DCSA$  are used to define to starting SCP for unemployed and employed while  $DCSE$  is used to define the arrival SCP (for unemployed).

The table 2 shows the extracted probabilities for the Auvergne region.

Table 2. Probabilities of the transitions "inactive  $\rightarrow$  unemployed", "unemployed  $\rightarrow$  inactive depending on SPC", "employed  $\rightarrow$  inactive depending on SPC"

Starting situation	Arriving situation		Inactives	Unemployed					
	Starting SCP	Arriving SCP		farmers	craftmen et al	executives	intern. profes.	employees	workers
Inactives				0,00005557	0,00055947	0,00031037	0,00172877	0,00644310	0,00604629
Unemployed	farmers		0,05462738						
	craftmen et al		0,06335331						
	executives		0,11808481						
	intern. profes.		0,06202433						
	employees		0,07066007						
	workers		0,06165634						
Employed	farmers		0,00650018						
	craftmen et al		0,01423226						
	executives		0,01729000						
	intern. profes.		0,01192824						
	employees		0,00930251						
	workers		0,01129013						

### *Probability to look for a job with a given profession*

The probabilities are computed using the same method we used to compute the probabilities of transitions of activity status. The difference is that we use the answers to the questions about the fact that the interviewee looks for another job. For the first period, we select the employed individuals ( $fi = 1$ ) looking for a job ( $dre1=5$ ). For the second period of the survey, from 2003 to 2007, we assume people look for a job if they have answered  $SOUA=1$  (want to have another job) and  $MREC = 1$  (have searched for recently) or  $SOUA=1$  and  $MREC = 2$  and  $NTCH = 1$  or  $2$  (have not recently search for because they wait for answer to recent applications or they have been ill for a while).

#### *Deciding looking for a job when unemployed*

Unemployed people are assumed to be those who search for a job. Even if, in the labour force survey, only 80% of unemployed people declare searching a job, we assume the probability to search for a job of unemployed people is one. Indeed, if we consider the whole model, it globally underestimates the job offer and the probability to find a job. This is difficult to correct as, for instance, we cannot consider that in most cases a job offer is proposed before it has been quit while the model time step is not less than one year. Also we assume the job offer equal to the job occupation. Then, the probability to search for a job of unemployed people is one in order to compensate a bit this underestimation and be able to occupy every job offer (which is the state the model has to reach). The data indicates the probability to look for a job for unemployed individuals is quite stable until 54 years of

age and dramatically decreases for older individuals. A second step of the modelling work would be to see if this dramatic decrease needs to be considered. We also analyse how different parameters describing the household (the number of unemployed in the household, the number of children, or the type of household) influence the probability to look for a job, and we did not find any clear dependency.

The probability to begin searching (i.e. becoming unemployed) if an individual did not search previously (not because she is employed) corresponds in the model to the transition from inactive to unemployed. As already mentioned, it is the complementary value for each age range of the value to make the transition from inactive to inactive.

Since an individual is unemployed, it is necessary to define which SPC she is going to search for. It varies a lot with the current SPC of the individual. As shown in Table 3 even if there is a tendency to look preferentially for her own SPC, an unemployed individual can prefer changing SPC. That is particularly the case of farmers and craftsmen. Then, we parameterise the process from the computation of the probability distribution to choose a SPC knowing the current SPC.

Table 3. Probability for unemployed people to search for a job with various SPCs knowing the current SPC of the individual

SPC/ Looks for	Farmers	craftsmen et al	executives	interm. prof.	Employees	Workers
Farmers	0.000	0.000	0.000	0.177	0.376	0.447
craftsmen et al	0.000	0.079	0.012	0.088	0.443	0.377
Executives	0.000	0.037	0.499	0.256	0.171	0.037
interm. prof.	0.000	0.009	0.053	0.591	0.273	0.074
Employees	0.003	0.007	0.006	0.063	0.808	0.113
Workers	0.006	0.010	0.003	0.056	0.251	0.674

### *Deciding looking for a job when already employed*

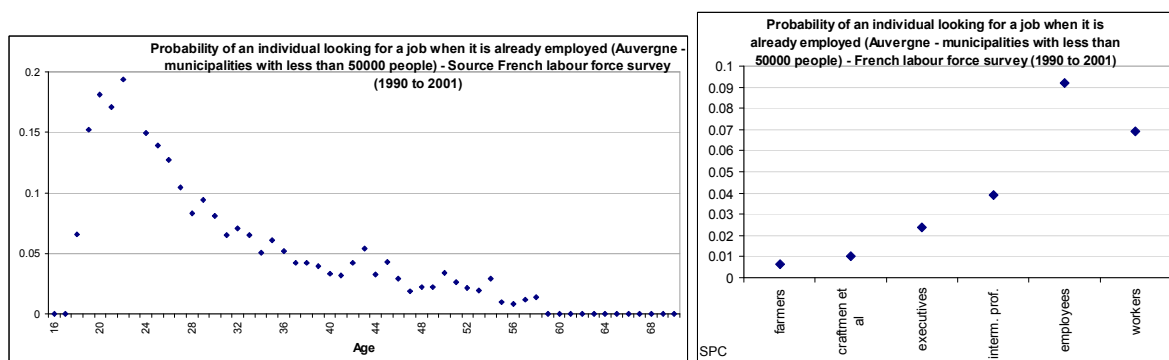


Figure 6 – (a) Probability for an already employed individual to look for another job according to the age (on the left); (b) Probability that an already employed individual looks for another job according to socio-professional category (on the right).

We consider those respondents being employed who answered that they are looking for another job. We have the age of these people, as well as the type of their current job. The analysis shows that the age is a very significant variable for determining if an employed



individual looks for another job (see Figure 6a). Young people are more susceptible to look for another job and this tendency decreases with age.

The SPC is also a significant variable to predict the probability to look for a job (see Figure 6b). Some SPC, such as employed farmers or craftsmen are not very susceptible to look for another job. On the contrary, others, such as workers and especially employees have quite a high probability to look for another activity.

Table 4 shows the parameter values for the decision searching for a given profession when the individual is already employed for some age ranges. For employed people, we built a probability containing the both information “have decide to search for a job” and “what she searches for”. It is important to point out that the probabilities presented in Table 5 do not add up to one but to the overall probability to search, which is quite low for already employed people.

Table 4. Extract of probabilities for employed people with a given SPC and a given five-year old age to look for a job within a given SPC.

<i>Age Range</i>	<i>Looks for/ Is a</i>	<i>farmers</i>	<i>craftmen et al</i>	<i>executives</i>	<i>interm. prof.</i>	<i>employees</i>	<i>workers</i>
15	<i>Farmers</i>	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002
	<i>craftmen et al</i>	0.0000	0.0000	0.0000	0.0000	0.0011	0.0014
	<i>executives</i>	0.0000	0.0000	0.0000	0.0000	0.0010	0.0000
	<i>interm. prof.</i>	0.0000	0.0000	0.0000	0.0000	0.0143	0.0040
	<i>employees</i>	0.0000	0.0000	0.0000	0.0000	0.1319	0.0168
	<i>Workers</i>	0.0000	0.0000	0.0000	0.0000	0.0162	0.0498
...	<i>Farmers</i>	...	...	...	...	...	...
	<i>craftmen et al</i>	...	...	...	...	...	...
	<i>executives</i>	...	...	...	...	...	...
	<i>interm. prof.</i>	...	...	...	...	...	...
	<i>employees</i>	...	...	...	...	...	...
	<i>Workers</i>	...	...	...	...	...	...
55	<i>Farmers</i>	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	<i>craftmen et al</i>	0.0000	0.0000	0.0000	0.0000	0.0002	0.0002
	<i>executives</i>	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000
	<i>interm. prof.</i>	0.0000	0.0000	0.0000	0.0000	0.0030	0.0005
	<i>employees</i>	0.0000	0.0000	0.0000	0.0000	0.0274	0.0021
	<i>workers</i>	0.0000	0.0000	0.0000	0.0000	0.0034	0.0062

### **Individual searches for a job**

Since the individual knows which profession she wants to search for, she has to find a place where to look for a job. Firstly, the individual selects an accepted distance she would want to commute. The next section presents how to the related probabilities<sup>15</sup>. If the chosen distance is higher than zero, the individual has to decide if she is going to work outside her set of municipalities. The law allowing this decision and the way to extract it from data is the subject of what follows in the next section. In case the individual has not found a job, she revises the maximum distance. She revises the distance up to 10 times.

### *The probability to accept a distance to cross over to work<sup>13</sup>*

The distance of search for a job is selected from a probability law giving the probability to accept a certain distance between the residence and the work place. The principle is very simple: the probability to commute at a given distance  $i$  [ $pc(i)$ ] is assumed to be the product of a probability to accept a certain distance  $i$  [ $pa(i)$ ] by the pay offered at  $i$  [ $O_i$ ] with a renormalisation coefficient  $k$ :  $pc(i) = k pa(i) * O_i$ .

Then, it is possible to extract the probability to accept a given distance ( $pa$ ) to work place, which will be used in the model. This procedure, coupled to an appropriate job offer, will allow maintaining the statistical properties of the  $pc$  distribution over the time of the simulation.

We extract from the mobility data of the 1999 Census for every municipality of the Auvergne region data on commuting ( $pc$ ) and data on job occupations, which we assume to be equivalent to job offers ( $O$ ). Evidently, the number of occupied jobs is used as a relevant proxy for the job offer of a municipality. An exhaustive description of the work allowing to build this probability law is given in (Felemou 2011).

Figure 5 shows an example of commuting data probability distribution ( $DDC = pc$ ) and of job offer probability distribution ( $DOE = O$ ) for one randomly chosen municipality.

A classification of acceptable distance distributions shows municipalities can be classified in three different groups, apparently depending on the size of the municipality of residence (see Figure 6 on the right). Thus, we assume for this parameter three probability distributions shown on the left of Figure 6 for three different size-dependent classes of municipalities (to the right of Figure 6). The data suggests that the larger the municipality, the lower the probability to work in the place of residence and the longer the commuting distance.

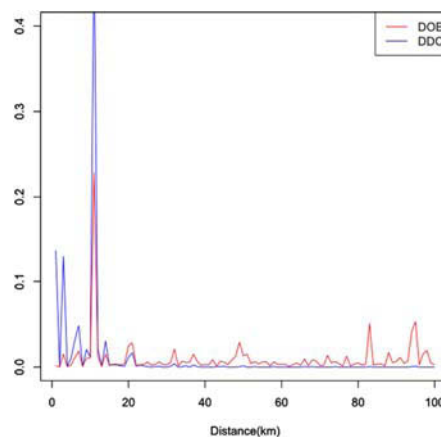


Figure 5 - Example for one municipality of the density distribution of job offers ( $DOE=O$ ) and the one of commuters ( $DDC=pc$ )

---

<sup>13</sup> This method, used at the time the paper was written, is now deprecated. See chapters 11 or 14 for the last version.

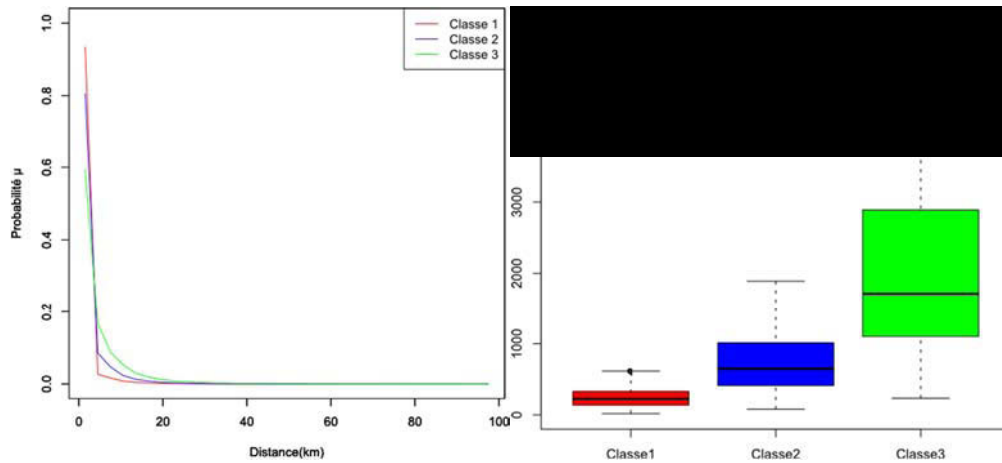


Figure 6 - Probability laws that an individual accept to a certain commuting distance knowing that a job is available for it.(on the left) - Different population sizes for the municipalities of each sub-group (on the right)

It is important to emphasise that only if the selected distance is higher than zero, the individual has to decide if she is going to outside or inside the set.

### *Going to work outside the set*

When the individual is commuting – meaning she has picked out a distance of research higher than 0 – she has to check if she has a chance to commute outside considering her place of residence. Indeed, an individual living close to the border of the set has a higher probability to commute outside the set. Then, the individual chooses at random to work outside depending on the probability associated with her municipality of residence. Each municipality has such a probability which is a function of its distance to the border of the set. This function is extracted from the mobility data from 1999 (Source: INSEE). Figure 7 shows this function for the Cantal department and the whole Auvergne region of which Cantal is a part. Both laws are quite close and it appears relevant to use as a parameter the law extracted for the whole region since it is probably less noisy.

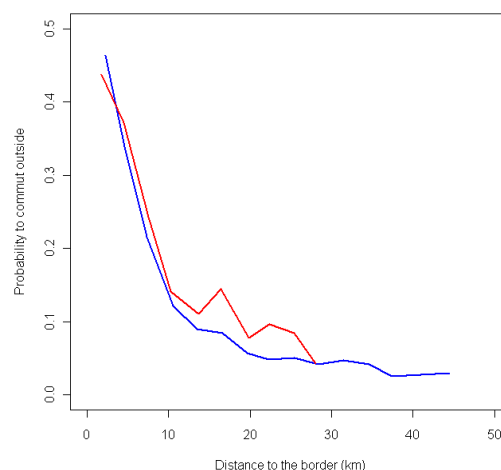


Figure 7 - Probability to commute outside the set (ordinate) depending on the distance of the municipality of residence to the frontier of the set (abscissa in Euclidian kilometers) - Red: Cantal; Blue: Auvergne

We are now describing how to extract the probability law for the final event which is going on retirement.

### *Going on retirement, and stop searching for a job*

To extract the transition to the retirement, we consider, in the period 1990 to 2002, the value FIP=all except 5 or 6, which means that the individual has not yet retired and the value FI=5 or 6, which means that the individual is now retired. We assume that the retiree does not search for a job anymore since this is generally the case true in France. Figure 8 shows that the speed of transitioning into retirement varies a lot from one SPC to another: we can read for example that at 60, 63 % of workers are retired while only 17 % of farmers are retired. Then, instead of considering a generic retirement law for all the individuals we consider a law for each SPC. Indeed, as these laws influence the job availability at a given moment it is very important to be sufficiently precise.

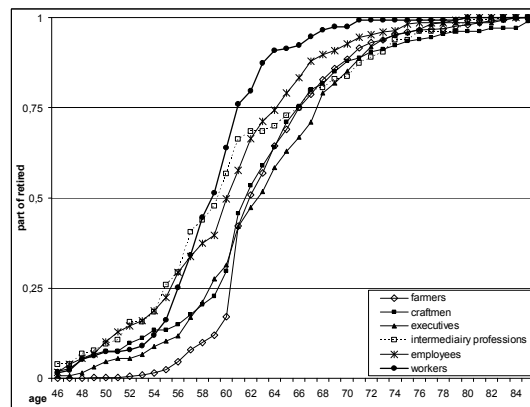


Figure 8. Speed of going into retirement by SPC (source LFS) – France level

## Lessons / Experience

First, we want to stress the necessity to not only consider the objectives of the model during the design, but from the very beginning exploring existing data sources and studying the implicit model beside the existing databases. The availability of data and the more or less implicit model guiding the collection of data constrain the definition of agents, their attributes and behaviours.

Using large existing databases can appear more relevant, especially the “official” ones from the National Statistical office, than collecting a small sample and reweighting it to obtain a statistically significant artificial population.

For these large databases, the models guiding the collection of data represent the expertise knowledge and generally assume some dynamics, particularly if time series are collecting during the survey. Moreover, if the data sources are collected by the National Statistical Office, they probably represent the commonly used information and knowledge by the stakeholders and policy makers. A model which aims to inform decision making is more useful if it can be easily understood and discussed by the relevant decision makers. This is easier if the model starts with common knowledge.

More generally, the modeller has to identify the rationale behind the considered data sources and use it to build the dynamic model. Indeed, this rationale often makes some implicit assumptions on the dynamics. Let's take the definition of a household as an example. *"In surveys prior to 2005, people were required to share the same main residence to be considered as households. It was not necessary for them to share a common budget. De facto, a household corresponded to a housing (main residence)".* Thus, until 2005, the French National Statistical Office (INSEE) assumes the household/family is defined by the place where it lives, which is unique. Indeed, following the INSEE definition, each person in a household may belong to only one family. In this framework, residential mobility is a household/family decision and the number of occupied housings in a place corresponds to the number of resident households. That is also what we assume in the model. *"Since 2005, a housing can include several households, referred to as "living units". Every household is composed of the people who share the same budget, that is who contribute resources towards the expenses made for the life of the household; and/or who merely benefit from those expenses."* The new definition is based on the fact that related or unrelated individuals can share the same budget and have a habitual residence (the housing in which they usually live). This new definition takes into account some cultural evolutions and allows a European homogenization of the way households are defined. However, it modifies the way the dynamic of move can be considered since each individual of the household can have more than one housing. This is to point out that the choice between one data source and another corresponds to a representation of the world to which some particular dynamics can be linked. If the first definition of household is more related to the idea that relationships between people can be identified by the concept of family and/or the identical of place of living, the second definition puts the economic constraints (i.e. the sharing budget) much more at the heart of the dynamics of closeness. A modeller, having the choice between a data source containing data built on the first definition and another one based on the second definition, should be aware of the choice to make and communicate about it.

A practical example of the implication of the choice of variables the modellers made is the one done to define what a job is. Choosing to only use data on the SCP and the activity sector to describe a job while it is possible to use the salary, which is available in some databases, makes having an occupation much more important than the level of salary. It also implies, for example, that an individual can change jobs just to change their working environment. Differently, the classical economic models considering job change start from the salary and assume an individual changes to increase their salary. We simply assume our individual wants to change jobs, without necessarily changing SCP at the same time. However, one can notice our assumption is relevant due to the existence of a minimum salary in France which ensures a minimum amount of money to live with.

The choice of existing databases for facilitating model design and parameterisation needs to consider:

- a longer as possible period of calibration: indeed it is not sufficient to strongly link the model to data if the model is not calibrated or calibrated with poor data compromising the robustness of the trajectory of underlying model dynamics;

- a sufficient number of modalities for each attribute in order to be able to reproduce the diversity of relevant agent types and behaviours. For example, we chose to aggregate in our work jobs in 24 types; at the end this depends on data availability;
- a minimum number of variables to calibrate: too many unknown parameters implies we don't know much about the dynamics and every experimental designation for observed trajectories can be valuable;
- the possibility to use them simultaneously for initialising agent attributes and defining agent behaviours: that means in particular that they have to have common variables allowing for a link between them. The challenge is to make an easy fit between attributes and behaviours.

Finally, starting from large national databases makes it likely that the model can be easily implemented and parameterized in another country. For instance, the example on the individual dynamics of activities indicated the possibility to apply the model in another European country even if some small adaptations are required. Indeed, Europe tends to harmonise the data bases in order to have common indicators at the European level. Then, large national databases have been designed or redesigned for answering the European demand. For example, the French "Employment survey" is the data source for the French contribution to the European Labour Force Survey. That is why (Baqueiro Espinosa, Unay-Gailhard et al. 2011) proposes a way to parameterise our model directly starting from the data of this European survey. For the same reason, national census data in Europe tend to consider more and more comparable or identical variables. That makes it possible to use them to parameterise our model even if a particular attention to the definition of used concepts remains: while to be a retiree in France (at least until a very recent period) means not looking for a job, it is not the case in UK for example.

Taking into account data at an early stage is not an easy task. It is at the same time laborious and confusing since the modeller is confronted with a very large set of information and more or less implicit knowledge. Finding a way to use the data and to choose the object, their attribute and the dynamics in order to remain simple as possible is much more demanding than developing a theoretical model. However, for such complex systems and models as ours that focus on the dynamics of interacting municipalities, the approach allows defining and controlling properly some sub-dynamics, even if they are not independent from other dynamics in order to test hypothesised system properties. For our concerns, we expect the expertise we developed for the labour market in conjunction with the robust parameterisation of the individual activity dynamics and job offer dynamics, will allow us to better understand how the demography impacts on the population/depopulation phenomena and how these phenomena impact on demography in return.

## **Acknowledgements**

This work has been funded under the PRIMA (Prototypical policy impacts on multifunctional activities in rural municipalities) collaborative project, EU 7th Framework Programme (ENV 2007-1), contract no. 212345

## References

- Aubert, F., Dissart, J.C., and Lépicier, D. (2009), 'Facteurs de localisation de l'emploi résidentiel en France', XLVIème Colloque de l'Association de Science Régionale de Langue Française (ASRDLF), 6-8 juillet, Clermont-Ferrand, France, 27.
- Ballas, D., Clarke, G.P., and Wiemers, E. (2005), 'Building a Dynamic Spatial Microsimulation Model for Ireland', *Population, Space and Place*, 11, 157-72.
- (2006), 'Spatial Microsimulation for Rural Policy Analysis in Ireland: The Implications of CAP Reforms for the National Spatial Strategy', *Journal of Rural Studies*, 22, 367-78.
- Ballas, D., et al. (2007), 'Using SimBritain to Model the Geographical Impact of National Government Policies', *Geographical Analysis*, (39), 44-77.
- Baqueiro Espinosa, O., et al. (2011), 'Two adaptations of a Micro-Simulation Model to Study the Impacts of Policies at the Municipality Level', in PRIMA European project (ed.), Working paper (IAMO, Newcastle University, Cemagref), 62.
- Berger, T. and Schreinemachers, P. (2006), 'Creating Agents and Landscapes for Multiagent Systems from Random Sample', *Ecology and Society*, 11 (2), 19.
- Birkin, Mark and Clarke, Martin (2011), 'Spatial Microsimulation Models: A Review and a Glimpse into the Future', in John Stillwell and Martin Clarke (eds.), *Population Dynamics and Projection Methods (Understanding Population Trends and Processes, 4: Springer Netherlands)*, 193-208.
- Birkin, Mark and Wu, B. (2012), 'A review of Microsimulation and Hybrid Agent-Based Approaches', in A.J. Heppenstall, et al. (eds.), *Agent-Based Models of Geographical Systems (Springer)*, 51-68.
- Blanc, M. and Schmitt, B. (2007), 'Orientation économique et croissance locale de l'emploi dans les bassins de vie des bourgs et petites villes', *Economie et Statistique*, 402, 57-74.
- Bousquet, F. and Le Page, C. (2004), 'Multi-Agent Simulations and Ecosystem Management: a review', *Ecological Modelling*, 176, 313-32.
- Bozon, M. and Héran, F. (1987), 'La découverte du conjoint : I. Evolution et morphologie des scènes de rencontre', *Population*, 42 (6), 943-85.
- (1988), 'La découverte du conjoint : II. Les scènes de rencontre dans l'espace social', *Population*, 43 (1), 121-50.
- Brown, D.G. and Robinson, D.T. (2006), 'Effects of Heterogeneity in Residential Preferences on an Agent-Based Model of Urban Sprawl', *Ecology and Society*, 11 (1), 46.
- Brown, D.G., Aspinall, R., and D.A., Bennett (2006), 'Landscape Models and Explanation in Landscape Ecology - A Space for Generative Landscape Science', *The Professional Geographer*, 58 (4), 369-82.
- Coulombel, N. (2010), 'Residential Choice and Household Behavior: State of the Art', (Working Paper 2.2a: Ecole Normale Supérieure de Cachan), 69.
- Davezies, L. (2009), 'L'économie locale "résidentielle"', *Géographie, économie, société*, 11 (1), 47-53.
- Deffuant, G. and al., et (2001), 'Rapport final du projet FAIR 3 2092 IMAGES : Modélisation de la diffusion de l'adoption de mesures agri-environnementales par les agriculteurs (1997-2001)'.<sup>1</sup>
- Deffuant, G., Huet, S., and Amblard, F. (2005), 'An individual-based model of innovation diffusion mixing social value and individual payoff dynamics', *American Journal of Sociology*, 110 (January) (4), 1041-69.
- Deffuant, G., Skerrat, S., and Huet, S. (2008), 'An agent based model of agri-environmental measure diffusion: what for?', in A. Lopez Paredes and C. Hernandez Iglesias (eds.), *Agent Based Modelling in Natural Resource Management (Universidad de Valladolid: INSISOC)*, 55 - 73.
- Deffuant, G., et al. (2002), 'How can extremism prevail ? A study based on the relative agreement interaction model', *Journal of Artificial Societies and Social Simulation*, 5 (4).
- Dubuc, S. (2004), 'Dynamisme rural: l'effet des petites villes', *L'Espace Géographique*, 1, 69-85.
- Felemou, Mamourou (2011), 'Analyse de données de flux de navetteurs. Extractions de modèles', (Rapport de stage de 1ère année de Master Statistiques et Traitement de Données; Aubière: Cemagref), 31.
- Fernandez, L.E., et al. (2005), 'Characterizing location preferences in an exurban population: implications for agent-based modeling', *Environment and Planning B: Planning and Design*, 32 (6), 21.
- Fontaine, Corentin and Rounsevell, Mark (2009), 'An agent-based approach to model future residential pressure on a regional landscape', *Landscape Ecology*, 24 (9), 1237-54.
- Gargiulo, F., et al. (2010), 'An Iterative Approach for Generating Statistically Realistic Populations of Households', *PLoS One*, 5 (1), 9.
- (2011), 'Commuting network model: going to the bulk', *JASSS*, 15.
- Goux, D. (2003), 'Une histoire de l'enquête Emploi', *Economie et Statistiques*, 362, 41-57.

- Grimm, V., et al. (2010), 'The ODD protocol: A review and first update', *Ecological Modelling*, 221, 2760-68.
- (2006), 'A standard protocol for describing individual-based and agent-based models', *Ecological Modelling*, (198), 115-26.
- Holme, E., et al. (2004), 'The SVERIGE spatial microsimulation model: content, validation and example applications. ', (Technical report: Spatial Modelling Centre, Sweden), 55 p.
- Huet, S., et al. (2011), 'Micro-simulation model of municipality network in the Auvergne case study', in *PRIMA Working paper* (ed.), (Aubière, France: IRSTEA), 63.
- INSEE, division "Redistribution et Politiques Sociales" (1999), 'Le modèle de simulation dynamique DESTINIE', *Série des documents de travail de la Direction des Etudes et Synthèses Economiques*, 124.
- Lenormand, M., Huet, S., and Deffuant, G. (2011a), 'Deriving the number of jobs in proximity services from the number of inhabitants in French rural municipalities', *Regional Science*, submitted, 13.
- Lenormand, M., Huet, S., and Gargiulo, S. (2011b), 'A commuting generation model requiring only aggregated data.', in *European Social Simulation Association* (ed.), *ESSA 2011* (Montpellier), 16.
- Lenormand, M., et al. (2011c), 'Evaluating the number of proximity service jobs per inhabitant in small French municipalities', (PRIMA report: Cemagref, LISC), 12.
- Moeckel, R., et al. (2003), 'Microsimulation of land use', *International Journal of Urban Sciences*, 71 (1), 14-31.
- Morand, E., et al. (2010), 'Demographic modelling: the state of the art', (FP7-244557 Projet SustainCity; Paris: INED), 39.
- Müller, K. and K.W., Axhausen (2011), 'Population synthesis for microsimulation: State of the art', 90th Annual Meeting of the Transportation Research Board (Washington D.C.).
- Orcutt, G.H. (1957), 'A new type of socio economic system', *Review of Economics and Statistics*, 58, 773-97.
- Parker, Dawn C., et al. (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), 314-37.
- Perrier-Cornet, Ph. (2001), 'La dynamique des espaces ruraux dans la société française : un cadre d'analyse', *Territoires 2020*, 3 (Etudes et prospectives - Data), 61-74.
- Polhill, J.G., et al. (2008), 'Using the ODD Protocol For Describing Three Agent-Based Social Simulation Models of Land-Use Change', *Journal of Artificial Societies and Social Simulation*, 11 (2 3), 30.
- Rindfuss, R., et al. (2004), 'Developing a science of land change: Challenges and methodological issues', *PNAS*, 101 (39), 13976-81.
- Soumagne, J. (2003), 'Les services en milieu rural, enjeu d'aménagement territorial', *Revista da Faculdade de Letras - Geografia I série*, XIX.
- Turci, L., et al. (2010), 'Provisional demographic outline', (FP7-244557 Projet SustainCity; Paris: INED), 24.
- Ullman, E. and Dacey, M. (1960), 'The minimum requirement approach to the urban economic base', *Papers and Proceedings of the Regional Science Association*, 6 (192).
- Verburg, P. H., et al. (2006), 'Downscaling of Land Use Change Scenarios to Assess the Dynamics of European Landscapes', *Agriculture, Ecosystems and Environment*, 114, 39-56.
- Verburg, P.H., et al. (2004), 'Land Use Modelling: Current Practice and Research Priorities', *GeoJournal*, 61, 309-24.
- (2002), 'Modelling the Spatial Dynamics of Regional Land Use: The CLUE-S Model', *Environmental Management*, 30 (3), 391-405.





## Chapter 1.4 Parametrisation of the unknown laws of demography

Authors: Huet S., Lenormand M., Deffuant G.

### Abstract

In this chapter we design demographic dynamics which are impossible to derive directly from data: forming couple, splitting, giving birth, deciding moving, deciding where to search for a new housing. We assume that these dynamics can be inspired from expert knowledge about the study region and try to minimise their number of parameters. For each parameter, we identify a segment of relevant values to obtain results that are compatible with the reference values given by the National Statistical Office. We also identify which indicators are sensitive to which parameter. This preliminary work is a prerequisite for defining parameter bounds and indicators for the calibration.

### Contents

The Cantal and its demography .....	84
A quick overview .....	84
The natural balance.....	86
The migratory balance .....	86
The increasing and decreasing municipalities .....	87
Commuting and residential mobility.....	87
How to model couple and birth dynamics in Cantal.....	88
Formalizing the questions and the dynamics to test.....	88
The experimental designs .....	93
Choosing the dynamics .....	94
Parameter bounds and other indicators.....	96
How to model moving.....	99
Which dynamics for deciding to move and modelling couples .....	100
Which dynamics for deciding where searching a residence or a partner?.....	111
Conclusion .....	122
Couple creation .....	123
Couple splitting.....	123
Birth .....	124
Moving.....	124
Indicators and parameters .....	126
Segments of relevant values .....	128
Acknowledgements.....	128



To implement the conceptual model presented in the chapter 1.1 (Huet and Deffuant 2010) for the Cantal French “*département*”, we have to build an initial population, as presented in the chapter 1.2, and to derive from data the dynamics of the chosen objects when it is possible, as it is presented in the chapter 1.3 for the activity dynamics (Huet, Lenormand et al. 2012). However, deriving dynamics from data is not always possible. Required data may not be available, especially longitudinal data about the individual behaviour. It is also common that available data can be used for checking and validating a chosen model but are not sufficient for guiding the hypotheses about the model. Thus, in this case, we must make such hypotheses and assume basic mechanisms. This chapter describes examples of such a work. Its purpose is to determine the unknown demographic dynamics that can be used to adapt the conceptual model to the Cantal case study.

In absence of data from which one could derive the dynamics, we test several hypotheses and check when the model is able to produce results close to available data. We then select the best compromise between the model simplicity and its ability to reproduce reference data. The first step consists in designing hypothetical dynamics using the knowledge and data available for the Cantal keeping in mind we try to minimise the number of parameters required for each of them. Then we study, through a simulation approach based on a sufficiently large experimental design, the capacity of the dynamics to produce results close to reference value of chosen indicators. Thus, we can decide which dynamics are relevant. In a second step, we enlarge the checking to other indicators which should be impacted by the chosen dynamics. We start from the results of analysis of variance allowing selecting, for each indicator, a subset of parameters (and implicitly of dynamics) and we check again the relevance of the chosen dynamics. Practically, as in the first step, we check if for some values of the parameters, the results of the model are close to the reference for the chosen indicators. At the same time we try to identify bounds for the parameters values leading to relevant results. These bounds, as well as the identified link between indicators and parameters, are useful in the calibration process which will identify the best values for the parameters for all the indicators simultaneously.

The unknown demographic dynamics can be grouped into two sets:

- the “family” dynamics with mechanisms linked to couple, creation and split, and giving birth;
- the “moving” dynamics with mechanisms ruling how a household decides moving, where it searches for another housing and how a proposal can appear as satisfying for her.

Overall, they are responsible for the value of the main indicators describing the evolution of a population:

- The natural balance;
- The migratory balance.

The moving parameters are also responsible for the internal moves into the studied region and the way a subregion of Cantal evolves due to people moving from other subregions. Some indicators give details about the impact of the moves:

- The commuting distribution;

- The residential moves: moving distribution;
- Increasing and decreasing municipalities in terms of number of inhabitants.

The first stage consists in analysing these indicators in Cantal in order to formalize hypotheses about dynamics that can yield the same indicator values.

## The Cantal and its demography

Before giving details about demographic indicators in the Cantal, we present a quick overview of a spatial representation of the Cantal. This overview is built from the data measured by INSEE, the French Statistical National Office. Such data is called reference in the following.

### A quick overview

The Cantal is a French *département* comprising 260 municipalities, 27 *cantons*. Figure 1 and 2 show the Cantal in France and a spatial representation of the administrative organisation of the Cantal in 3 *arrondissements* and 27 *cantons*.



Figure 1. French “départements”. The Cantal is represented in red.



Figure 2. The Cantal is organised in 3 arrondissements (boundaries represented at the top on the right) having main cities Aurillac, Saint-Flour and Mauriac and 27 cantons (delimited on the blue map).

The Cantal has about 150000 inhabitants. Following the document "Assises des territoires ruraux<sup>14</sup> », the Cantal has all the features of a very deep rural *département*:

- Average density: 26 inhabitants/km<sup>2</sup>;
- Decreasing population from 1990 to 2006;
- 50% of the road network at an altitude higher than 800 meters;
- An economy based on agriculture (constrained by the severe climate);

<sup>14</sup> Report on Cantal *département*. Ministère de l'espace rural et de l'aménagement du territoire. 26 novembre 2009, 28 pages

- A main town distant from more than one hour of the closest highway, from more than 4 hours of a TGV railway station.

The figure 3 shows the Cantal subdivided into sections from the ZAUER typology. This typology distinguishes urban from rural areas as well as municipalities depending on centres from these two areas for the employment. We notice that there is only one urban centre around Aurillac and three rural employment centres around St Flour (on the right), Mauriac (on the top left) and Champs-sur-Tarentaine which is situated on the top border.

The histogram presented on the figure 4 shows how numerous are the small municipalities among the 260 municipalities of this *département*. The very small size of the spatial object, i.e. the municipality, we choose as the base to model commuting and residential move, is a real challenge for a model partly based on a probabilistic approach.

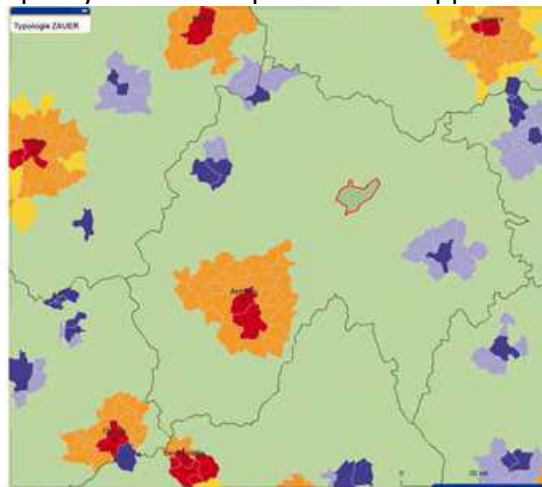


Figure 3. The Cantal and the surroundings declined through the ZAUER typology: in red, municipalities from an urban center; in orange, municipalities depending on an urban center; in yellow, municipalities depending on several urban centers; in dark blue, municipality from an employment center from the rural area; in light blue, municipalities which are the surrounding of the municipalities in dark blue; in green, isolated municipalities from the rural area. Source: INSEE, DATAR (1990)

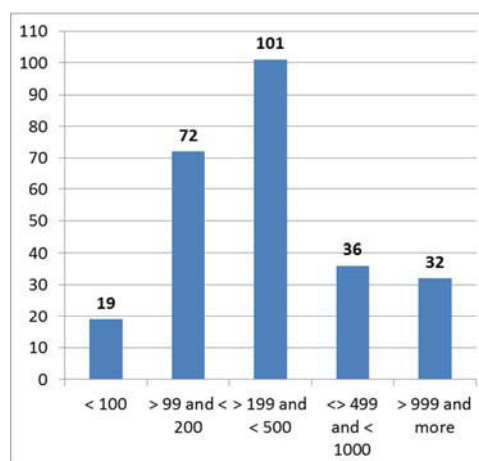


Figure 4. Number of municipalities of Cantal in 1990 in various size ranges in terms of number of inhabitants in 1990. In abscissa : number of inhabitants in 1990

## The natural balance

The natural balance remains negative for the two dates at which we have a reference value from the National Statistical Office. Their values are large compared to those of the migratory balance (see next section). They are:

- -4979 for 1990-1999 with 17225 deaths and 12246 birth; annual natural balance: -498;
- -3974 for 2000-2007 with 15032 deaths and 11058 births; annual natural balance: -497 (the natural balance for 2000-2006 was -3554).

One can notice from these figures that the computed average annual natural balance remains constant all over the time. Indeed, even if the average number of annual deaths increases, the average number of annual births also increases.

## The migratory balance

The migratory balances of the Cantal were slightly negative between 1990 and 1999 and became positive between 2000 and 2006. We notice that it tends to compensate the negative natural balance in 2000-2006 even if not totally. Their values are:

- -50 for 1990-1999 with 17075 outmigrants and 17025 inmigrants;
- 2091 for 2000-2006 with 9814 outmigrants and 11905 inmigrants;

The figure 5 gives details about the age structure of the migratory balance. The balance is better in 2000-2006 for individuals from 15 to 60. It especially improves for young people below 30.

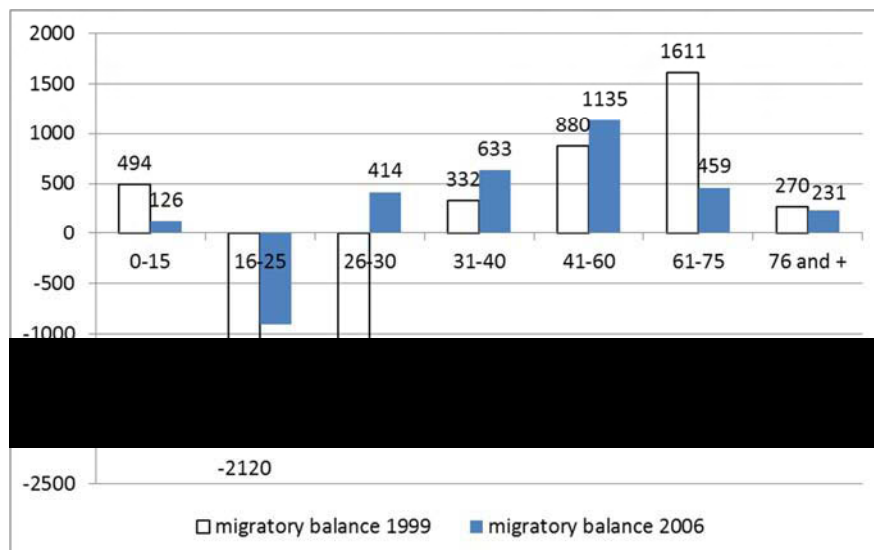
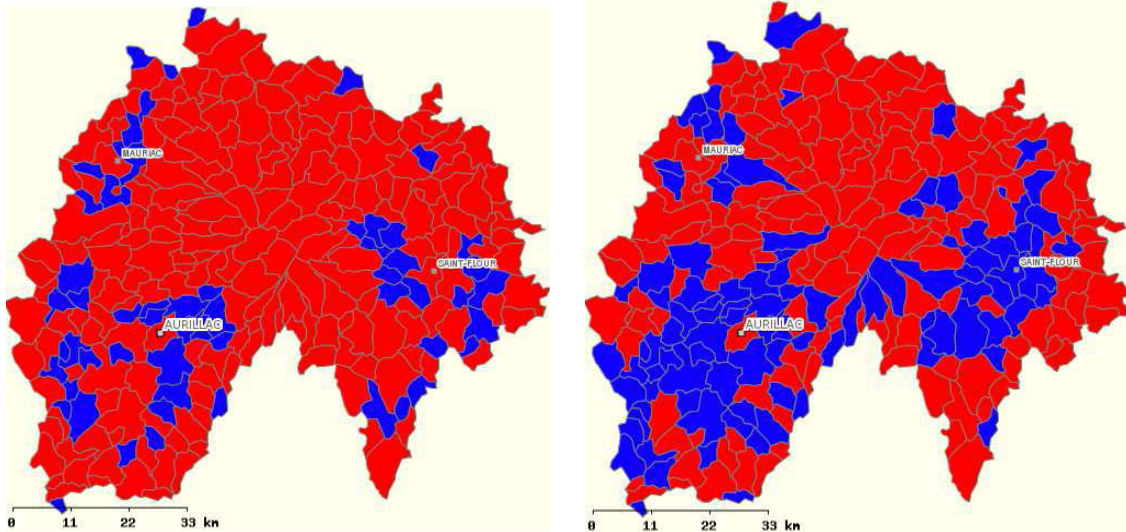


Figure 5. Migratory balances for 1990-1999 (white rectangles) and 2000-2006 (blue rectangles) in number of individuals and by age ranges (in abscissa)

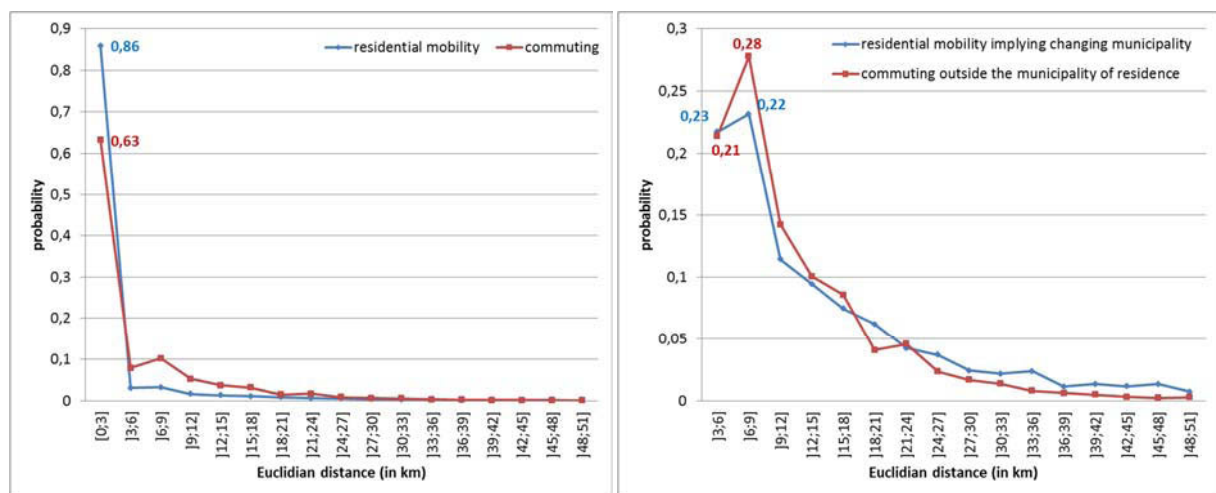
## The increasing and decreasing municipalities



Figures 6. Evolution of the size of the municipalities in Cantal with decreasing municipalities in red and increasing municipalities in blue: on the left, between 1990 and 1999; on the right, between 2000 and 2006.

The maps of the figures 6 show the positive or negative evolution of the number of inhabitants of the 260 municipalities of the Cantal. The increases are quite rare from 1990 and 1999 and occur in the surroundings of Aurillac, St Flour and Mauriac even if the urban (Aurillac) or rural (St Flour and Mauriac) don't increase themselves. From 2000 to 2006, the increases are quite numerous, much more diffused in and around the "urban centres" areas. However, the top centre remains all the time a decreasing area.

## Commuting and residential mobility



Figures 7. Distance of commuting and residential mobility distributions of probabilities (the distance is the Euclidian distance in kilometers): on the left, the total distribution comprising the move inside the same municipality; on the right, the distribution only for the moves outside the starting municipality

Figures 7 shows how short are the moves inside Cantal, be it commuting or residential mobility. A large majority of people work in their municipality of residence (63 %) and move inside the municipality where they already live (86 %) or close to it (at most 3 km, knowing



the average distance between two French municipalities is about 4 km). Moreover, the figure on the right shows that if we only consider the moves outside the starting municipality, almost 50 % of moves are below 9 kilometres (45% for residential mobility and 49% for commuting). The average Euclidian distance in kilometres of moving (without considering moving at most 3) is 14.83 km, of commuting (without considering commuting at most 3) is 11.08 km.

Moreover, we notice that the Cantal population simultaneously moves very close and increases spatially in and around the urban centres and the rural employment centres.

Another interesting figure is the number of individual moves from a Cantal municipality:

- From 1990 to 1999 : 133459
- From 2000 to 2006 : 116371

Considering the population is about 150000 individuals in 1999 and 2006, it means that on average every individual moves almost once in ten years.

## **How to model couple and birth dynamics in Cantal**

### **Formalizing the questions and the dynamics to test**

Couple creation and split are very important mechanisms in the model because, following the analysis of data coming from various French surveys (Debrand and Taffin 2005; Debrand and Taffin 2006), creating and splitting up couples are the first factors explaining the residential mobility. Moreover the model assumes that only couples can have children. This means that only individual in age of having children and having a partner can impact the natural balance in a positive way.

#### ***Couple creation***

We have a closed model, meaning that the couples are made of individuals both existing in the population. Indeed, in an open model, when the individual enters in union, a partner is created with suitable characteristics. This partner is not a full member of the population; she only exists by her relationship with a full member. In a closed model, as ours, two individuals have to be put into relation for creating a couple: they are both full members of the population and their union implies the creation of a new “couple” household.

From the various state-of-the-arts (Abelson and Miller 1967; O'Donoghue 2001; Bacon and Pennec 2007; Morand, Toulemon et al. 2010; Li and O'Donoghue 2012), we notice that the classical way to model this couple formation is a two-step model. First, individuals are selected to enter into partnership in a given year depending of some variables. They enter in a marriage market created to match those entitled to partner. This procedure matches the partners according to some characteristics (for example age, education level, socio-economic status...). Maximisation techniques are used and some behavioural rules are applied for those who can't find a perfect partner. They may relax their wishes in terms of partner's characteristics or wait another time period in the marriage market.

In France, a reference model is the DESTINIE model (Blanchet and Chanut 1998; INSEE 1999; INSEE 1999; Robert-Bobée 2001; Duée 2005; Blanchet, Buffeteau et al. 2011), developed by the French Statistical Institute and which reproduces the demographic evolution of French people in order assessing the retirement policies. The second large dynamic microsimulation model for France, PRISME, relates to social health security system and use a different data source from DESTINIE ((Poubelle 2006). However, it is hard to find out details about its implementation and those found are not different from DESTINIE. Then, in this chapter we give details on DESTINIE. In this model, the candidates to couple are selected a given year and put onto the marriage market. Marriage only means union. Then, a man and 20 women are randomly chosen without any geographical constraints since the model is not spatial. Among the picked out women, the woman closest to the man considering their ages and their ages of ending study is chosen as the partner of the man, if the difference of ages between the woman and him is less than 20 years.

This procedure appears us very heavy from a computational point of view, especially because we have to consider some spatial constraints. Moreover, it generally forbids that the individual lives other events during the year she tries forming a couple, as moving for example. Thus we opt for a very close procedure which is more computationally economical. We decompose the couple creation dynamics in two subdynamics: the decision searching for a partner; the search for the partner. We assume they can be modelled as described in the following algorithm.

Algorithm 1. Pseudo-code algorithm describing the searching for a partner process

```

Loop over HouseholdList hh
  ...
  if random < probaToSearchForPartner && hh.Type == single
    createNewCouple(hh)
  ...
=====
createNewCouple(Household hh)
  Individual part = findPartner(hh)
  if part != null
    move partners in the same larger housing
=====
findPartner(Household hh)
  // preference for searching in her place of residence
  for nbJoinTrials
    return partner = trial(hh, place = hh.placeOfResidence)
  // searching around the place of residence and the place of work comprised these latter places
  munWhereSearch = collectMunicipalitesAtMax(proximity)
  for nbJoinTrials
    return partner = trial(hh, place = picked up at random in munWhereSearch)
=====
trial(Household hh, Municipality place)
  otherHH = pick up at random in place.HouseholdList
  return partner = checkOneMemberMatchConditions(hh, otherHH)

```

During each time step, each single individual (with or without children) has a probability to search for a partner. If the individual tries to find a partner, the model draws at random a given number of trials (which is a parameter) other households living in her place of residence. Each time he has drawn a household; he tests his compatibility with each

member of the household in terms of age and single status. If no partner passes this test, the individual do the same but in a municipality which can vary from one trial to the other. This municipality is drawn at random from a collection of municipalities containing every municipality located in two circles: one having her place of residence as centre and radius *proximity* (parameter); and another one having her place of work as centre and radius *proximity* (parameter). Her place of residence is also included in the collection, as well as her place of work. It has to be noticed a municipality close from both the place of residence and the place of work is two times added in the collection.

Like in the “marriage” market, the procedure does not ensure a couple is formed. However, it runs properly in practice, probably because the single status (single and single-parents) are the most frequent ones of household in the population and because an individual can form a couple with a household having the status of potential immigrant. Then the potential immigrant becomes a full resident of the region.

We distinguish three elements in the couple creation process: the decision of a single to search for a partner; the number of trials to find out a partner; the age compatibility.

The third element can be designed from data considering the average difference of age between couple members and its standard deviation. That is given by the INSEE at the national level. Then, we assume that a single considers an individual as a compatible partner if she is also a single and if the difference of age between them is not higher than a difference picked out at random in the distribution of reference given by the INSEE.

The two first elements are much more difficult to design because we have no data about them. However they are important. Suppose for instance that couples are not numerous enough in the population. This can be because singles do not decide to search for a partner frequently enough; or because singles do not find compatible partners when they search for one. We assume that the process can be modelled as described in the algorithm n°1.

We wonder if the two elements, probability to search for a partner and number of times the process draws at random a possible partner, are necessary to fit the data corresponding to natural and migratory balance. Thus, our first question is: **do we need one or two parameters to model the couple creation?** In other words, assuming that the singles have always a probability 1 to search for a partner is the number of trials to find a compatible partner sufficient to fit the reference data given by INSEE.

To answer this question, **we test that a single searches for a partner with an annual probability value 1 versus other possible values less than 1.** If the model is able to fit the Cantal reference using the value 1, it means the Cantal case study does not require considering two processes to model the couple creation: only the number of trials finding out a partner is sufficient. Practically, it means that the number of singles is limiting, rare enough to explain the limited number of couple leading to the number of births.

### ***Couple split***

The more recent study we found on union dissolutions in France has been carried out by (Vanderschelden 2006). She shows that for the unions formed a given year, the risk to split in the following year is almost the same whatever the union duration. The probability to split tends to increase for more recently formed couple. The annual probability to split varies from about 0.004 for older couples (formed from 1950 to 1954) to 0.04 for those formed from 1990 to 1994. Moreover, the more recent unions considered by the author and having duration of at most 4 years are those formed between 1990 and 1994. She said that nothing ensures that the risk to split remains a constant over the duration of union for the most recently formed unions. These unions are those formed between 1990 and 1994; in other words that are all the unions formed during our simulation time since it starts in 1990.

Regarding dynamic modelling, the various state-of-the-arts done by (O'Donoghue 2001; Bacon and Pennec 2007; Morand, Toulemon et al. 2010) list the co-variables used for modelling the dissolution of unions in the existing microsimulation models. Duration of union and or ages of the partner, particular of the wife, are the most frequently cited co-variables. Children, employment status, level of education and others can also be used. In France, the model DESTINIE (INSEE 1999; Robert-Bobée 2001; Duée 2005) considers the union dissolution is woman driven and relates to the duration of union, the age at union (assuming the union formed in an early age is less stable), the number of children and the education.

Thus, even if the union duration is often used to model the union dissolution, it appears the observed splitting distribution of probabilities comes from an attribute of cohorts, evolving over the cohort, and not from the duration of the union. Some models, in their more recent design, have taken this into account. That is for example the case of DYNASIM<sup>15</sup> II or DYNACAN<sup>16</sup>.

We do not have in our data, and thus in our initial population the year when the existing couple in 1990 was formed. Then, applying a rule based on couple duration to model the split in Cantal is impossible. Thus, to limit the number of parameters and the complexity of the model, we use a constant probability to split that we will calibrate, trying to find out a kind of average probability to split over the different unknown year-meeting couples. Consequently, the second question is: **is a constant probability enough to model couple splitting and fit the reference data?** Once again, the number of births will be used to appreciate the accuracy of the chosen model because the impact of couple splitting on the natural balance is what we are interested in. We remind that the death process is parameterised from data and we expect that the number of deaths does not vary much with the parameters ruling couple formation and splitting.

### ***Giving birth***

Following the reviews of the literature (O'Donoghue 2001; Bacon and Pennec 2007; Morand, Toulemon et al. 2010), the most common co-variables used for the fertility are the age, the

---

<sup>15</sup> Microsimulation model projecting demographic and labour market events over time, interested in Social Security in USA

<sup>16</sup> Microsimulation model of Canada Pension Plan assessing its impact on the population

marital status and the parity (i.e. the number of children born by one woman). However, in practices the total number of co-variables used to define the probability of birth is often very high. The same type of approach is developed in DESTINIE to simulate the evolution of the French population (Duée 2005). We have not enough data related to the Cantal to envisage such a data based approach.

As already mentioned, the conceptual model assumes that only households with a couple can have children, and one of the adults should be in age to procreate. Moreover, a couple has an annual probability to have a child computed from an “average number of children by individual”. That is a starting point and the conceptual model itself explains that this “average number of children by individual” has to be calibrated since it is unknown. It cannot be directly derived from existing data such as the fertility rate available from the National Statistical Office. The total fertility rate (TFR), sometimes also called the fertility rate of a population is the average number of children that would be born to a woman over her lifetime if:

- she were to experience the exact current age-specific fertility rates through her lifetime, and
- she were to survive from birth through the end of her reproductive life.<sup>1</sup>

It is obtained by summing the single-year age-specific rates at a given time. This computation is equivalent to give an identical weight to each age range, whatever their real weight in the population. It suppresses the structural effect linked to the distribution by age ranges of women in age to procreate.

The French fertility rate increases over the period. The figure 8 shows that it is also the case in Cantal even if the tendency is not as high as in the whole country.

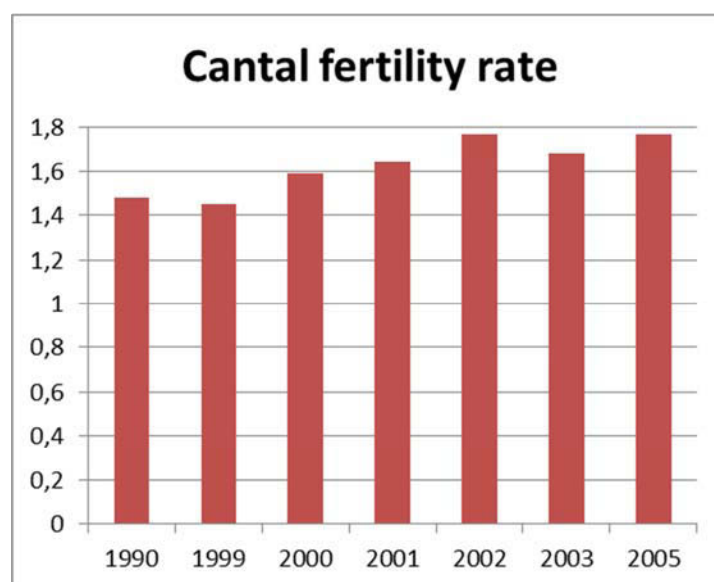


Figure 8. Fertility rate of the Cantal for various years for which the data is available

While the fertility rate increases in Cantal from 2000, the migratory balance of the 2000 to 2006 improves for people in age of procreation (see figure 5). It becomes less negative for 16

to 25 years old, positive for 26 to 30 years old and even more positive for 31 to 40 years old. The improvement of the migratory balance, especially the one of people around 30 could explain the increase of the fertility rate.

The third question is then: **is a constant average number of children by individual sufficient to generate a number of births corresponding to the reference or is it necessary considering an increasing average number of children by individual?**

To answer this question, **we consider a linearly increasing average number of children by individual parameterised with an intercept and a slope.** A slope valued 0 means the average number of children is a constant while a slope higher than zero means it is necessary to increase the parameter value over the time to obtain the right numbers of births. Figure 9 shows how the annual value of average number of children evolves with a slope 0.09.

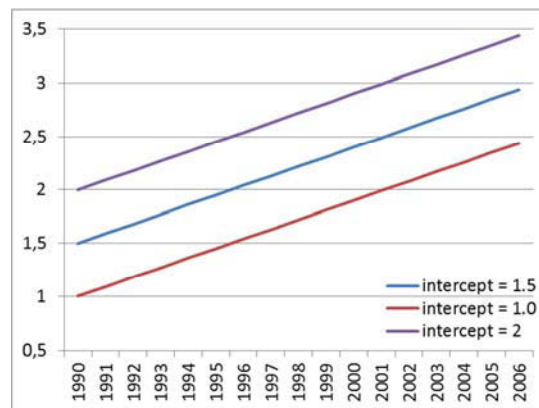


Figure 9. Temporal values of the giving birth linear parameter with a slope 0.09 and three different values of intercept.

## The experimental designs

Code param	Name	Values experimental design A	Values experimental design B
<i>nb child slope</i>	<i>slope of the giving birth function</i>	0	0.09
<i>nb child intercept</i>	<i>intercept of the giving birth function</i>	1.5; 2.0; 2.4	1; 1.2; 1.4; 1.6; 1.8; 2.0
<i>P search partner</i>	<i>probability to search for a partner</i>	0.1; 0.35; 0.6; 1	0.1; 0.35; 0.6; 1
<i>split p</i>	<i>probability to split</i>	0.002; 0.025; 0.034; 0.05	0.002; 0.018; 0.025; 0.034; 0.05
<i>join trials</i>	<i>number of join trials</i>	1; 5	1; 5
<i>Proximity</i>	<i>proximity threshold</i>	21; 33; 45	21; 24; 27; 30; 33; 36; 42; 45
<i>dispo threshold</i>	<i>house availability threshold</i>	0; 0.1; 0.15; 0.2; 0.25; 0.3; 0.35; 0.4	0; 0.1; 0.15; 0.2; 0.25; 0.3; 0.35; 0.4
<i>res satis p</i>	<i>residence satisfaction</i>	0.03; 0.07; 0.09; 0.11; 0.14; 0.16	0.03; 0.07; 0.09; 0.11; 0.14; 0.16

We used several experimental designs to study the model. Two are presented in this section because they allow us to answer the questions introduced in the previous section.

Experimental design A considers a constant average number of children by individual. Experimental design B assumes the slope to be 0.09.

The total number of experiments of the design A is 13 824, of B is 92160. Each experiment performed only once (no replication) because we want to “quickly” have an idea of the responses of the model.

## Choosing the dynamics

The figure 10 shows the results of the experimental design A.

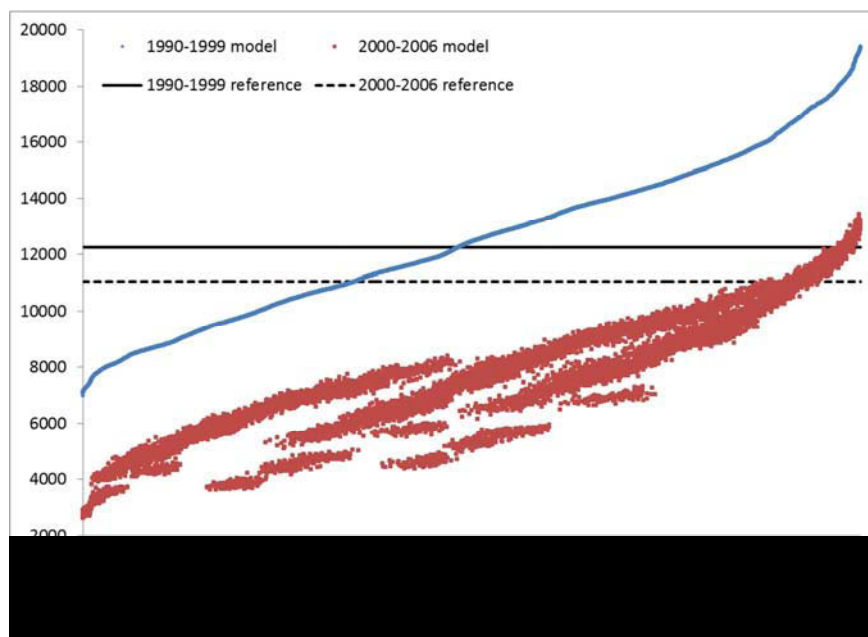


Figure 10. Every number of births for 1990-1999 (in blue) and 2000-2006 (in red) from the experimental design A with the slope = 0 for the giving birth process ordered following the births obtained for 1990-1999. In black are the references: plain for 1990-1999 and dotted for 2000-2006. No experiment fits these references simultaneously

Figure 10 shows that the annual number of average children by individual never increases enough between 1999 and 2006 to fit the reference value. Indeed, the difference of births between the two periods for the reference is -1188. Even if the number of births given by the model is measured from 2000 to 2006 (instead 2007), the minimum difference between the two periods for the model (experimental design A) is about -3200. This means that **a constant average number of children is not compatible with the reference value.**

It now remains checking that this becomes possible when the slope is higher than 0. That is what we test with the experimental design B with an arbitrary chosen slope at 0.09. Figure 11 shows the results.

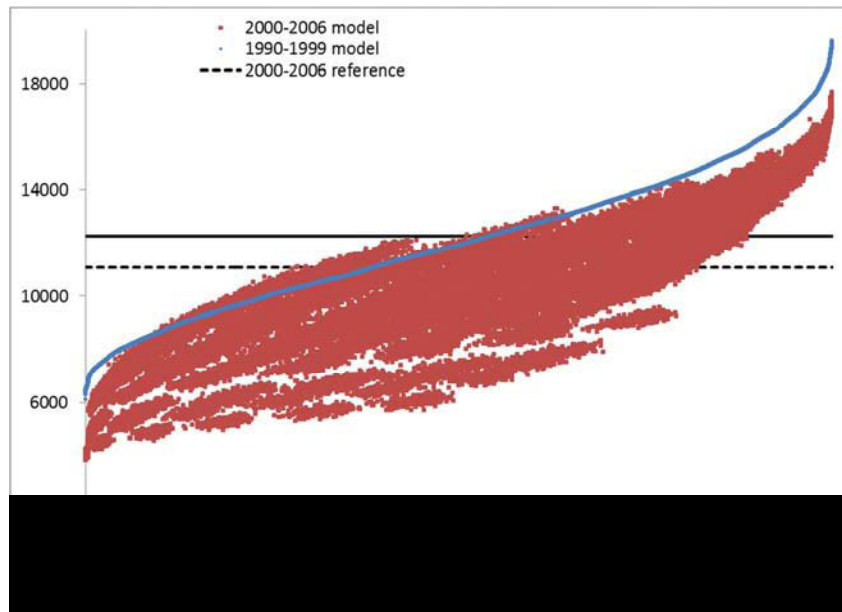


Figure 11. Every number of births for 1990-1999 (in blue) and 2000-2006 (in red) from the experimental design B with the slope = 0.09 for the giving birth process ordered following the births obtained for 1990-1999. In black are the references: plain for 1990-1999 and dotted for 2000-2006.

That is clear that the reference values are reachable for a slope equal to 0.09.

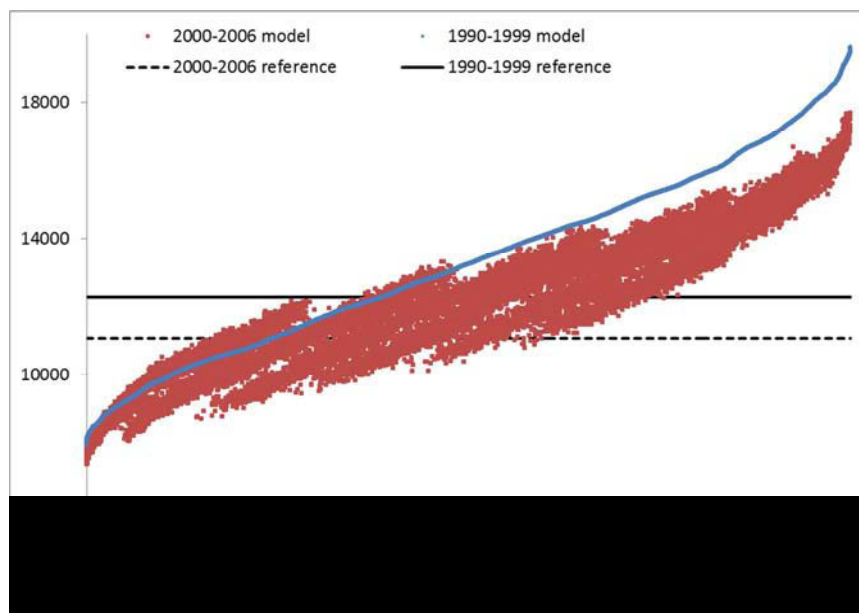


Figure 12. Every number of births for 1990-1999 (in blue) and 2000-2006 (in red) from the experimental design B limiting value for  $p$  search for a partner to 1 for the giving birth process ordered following the births obtained for 1990-1999. In black are the references: plain for 1990-1999 and dotted for 2000-2006.

However, the experimental design B does not directly allow us to answer the question about the necessary number of parameters to model the couple creation. Indeed, it considers several values for the probability to search for a partner from 0.1 to 1. Then, we have to check if the births of reference are a possible result of the model when  $p$  search for a partner is 1. That is why we select a subset of results from the experimental design B respecting this constraint. Figure 12 shows the reference is a possible result with a probability to search for a partner equal to 1. But it is too soon to conclude because we have



to check other indicators depending on couple dynamics. Similarly nothing allows us to conclude that the couple splitting process is not well modelled as a constant.

### Parameter bounds and other indicators

In order to get further, trying to identify with which values the model reaches the references; we performed an ANOVA over the results of the experimental design B. The purpose is to identify the parameters ruling the natural balances, and especially the number of births.

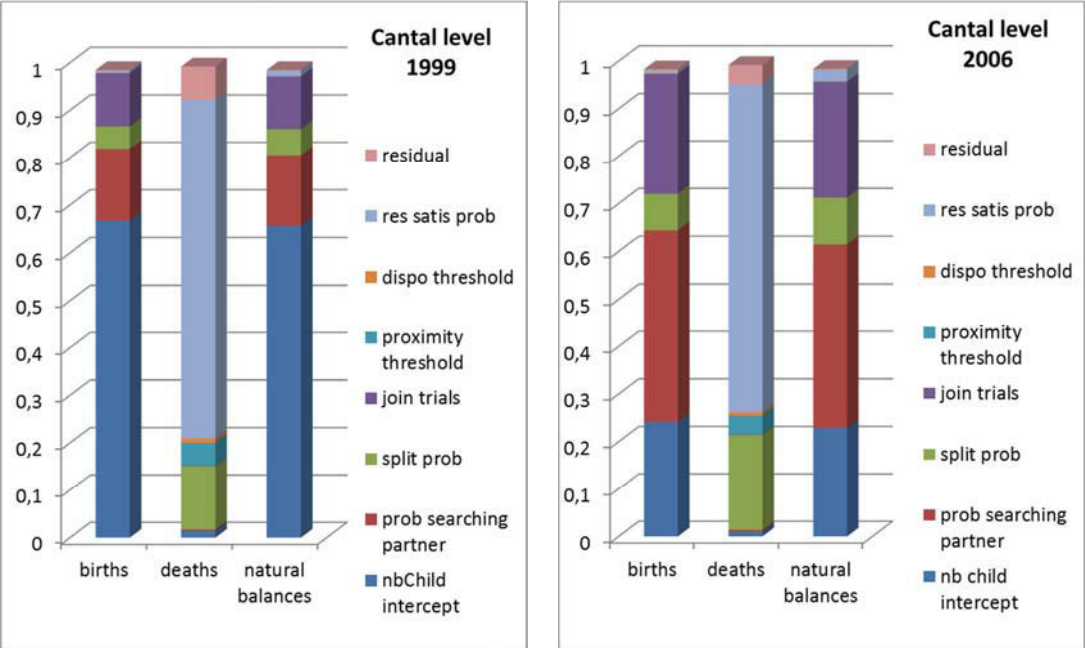


Figure 13. ANOVA for the births, deaths and natural balances measured in 1999 and 2006 (for 1990-1999 and 2000-2006), obtained from the results of the experimental design B. It represents effect of one parameter and residual (cross effect of two parameters are not represented).

Figure 13 shows the results of the ANOVA for the indicators births, deaths and natural balances. The natural balances are sensitive to the same parameters as the births. As expected, the main parameters responsible for the values of these indicators are: the intercept of giving birth process (nbChild intercept), the probability to search for a partner, the *join trials*, the residence satisfaction parameter and the *split* probability.

From these selections of main parameters for each indicator, we can have a look to a subpart of the results of the experimental design B considering the graphs shown in figures 14.

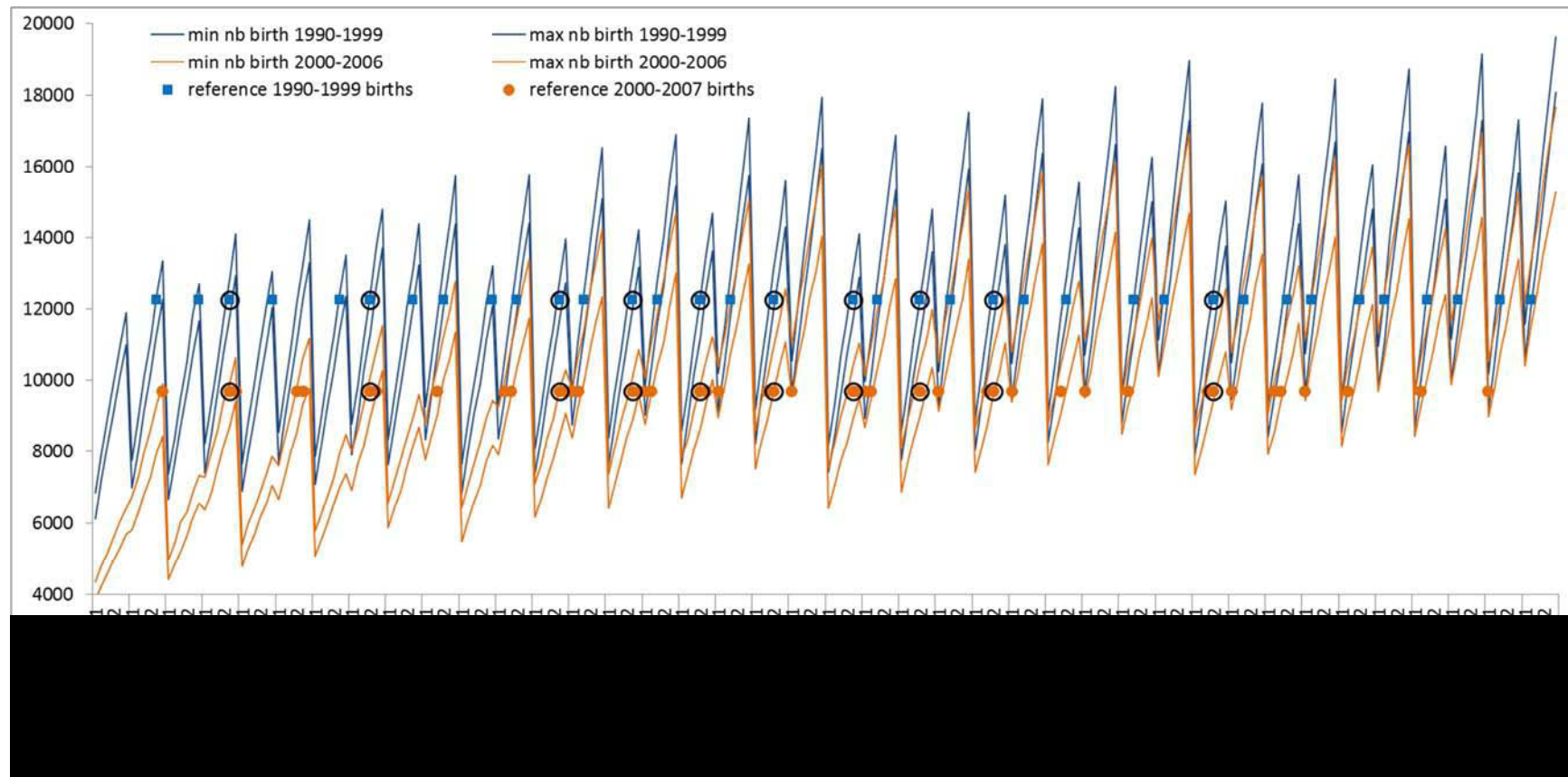


Figure 14a. Selection of births obtained with the experimental design B, for reference for 1990-1999 and 2000-2006. The black circles indicate the parameter sets for which the two references can be yielded by the model

The figure 14a is dedicated to birth. The black circles indicate the parameter sets for which the two references can be yielded by the model. Convenient nbChild intercepts are 1.6 or 1.8 (for a slope of 0.09). The reference can be yielded also for a probability to search for a partner equal to one but only for a large splitting probability (0.05) which at the same increases the number of deaths (while they are already too numerous). We can also notice that the right number of births is never produced by a splitting probability equal to 0.002. That is a useful control since it almost corresponds to no split (it means a couple duration is about 500 years).

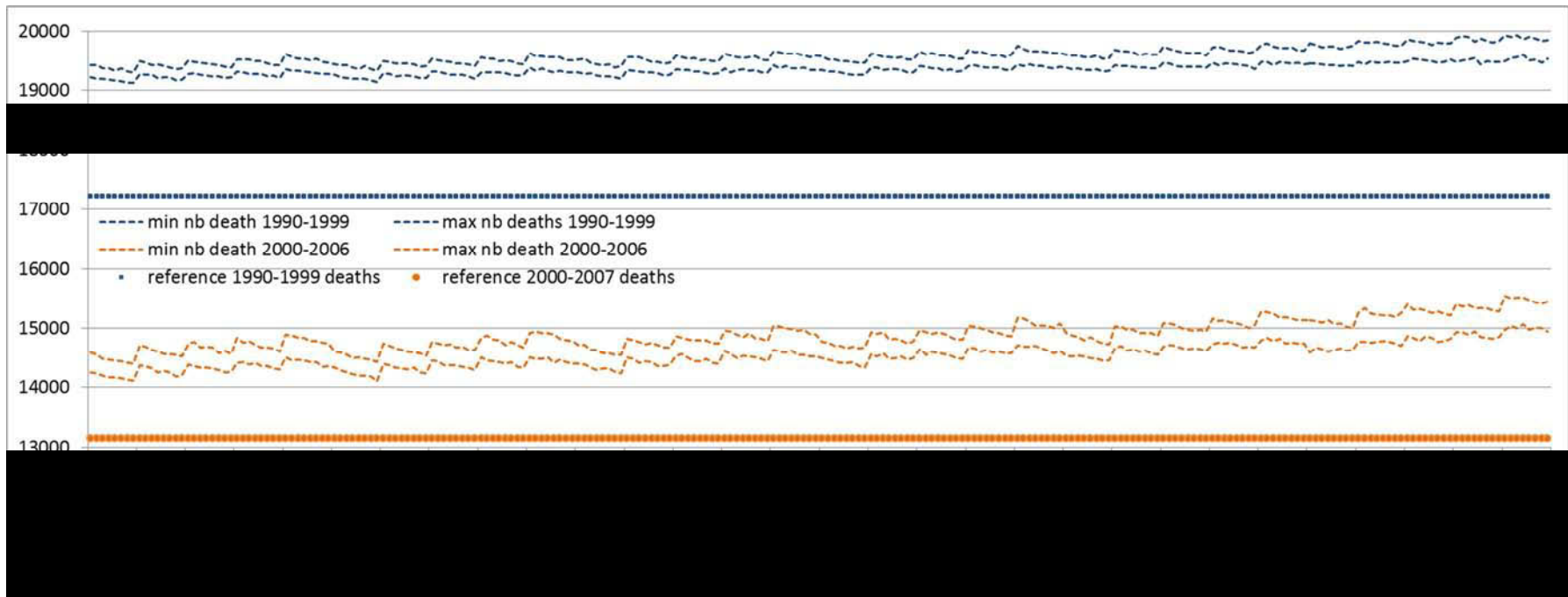


Figure 14b. Selection of deaths for the model obtained with the experimental design B, for reference for 1990-1999 and 2000-2006.

The figure 14b shows why the natural balance has the same sensitivity as the births: the deaths don't change much. It is expected because parameterisation of death is exogenously defined. The mortality in the model is always too high for the 90-99 period (about 2000 too much because it corresponds to the national data. Such a law is not available at the Cantal level. Thus we cannot correct this bias.

Even if the right number of births can be a result of the model, that is not the case of the number of deaths. Thus it means that it is not possible to get at the same time the right number of births and the right natural balance.

A last checking we did, which is not only related to the natural balance but also to the migratory balance, is the capacity of the model to reach reference values of number of inhabitants by age range. Figure 15 shows the interval defined by the maximum and minimum number of inhabitants for various age ranges given by the model contains the reference except for people being more than 60. However, that can be easily explained by the number of deaths a bit too high.

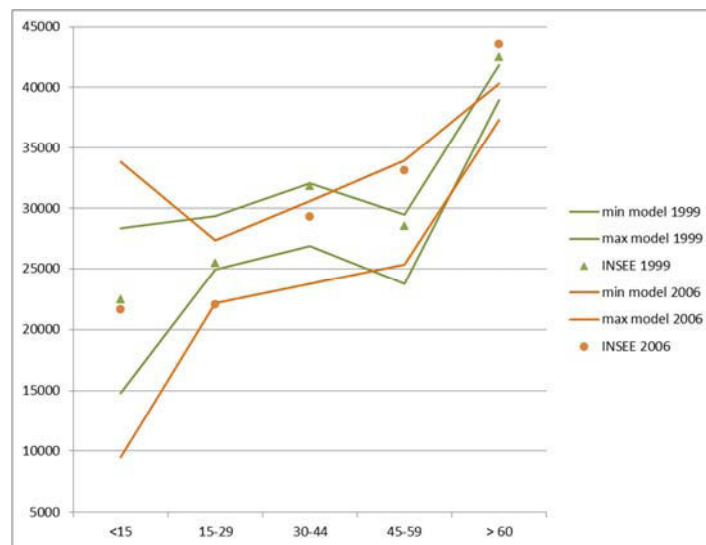


Figure 15. Minimum and maximum number of inhabitants by age range (in abscissa) and the reference values from INSEE in 1999 and 2006

To conclude about the choice of the dynamics for modelling birth in the Cantal, we retain that it is necessary to increase the annual number of children by individual over time. Thus we keep the function we choose to test our hypothesis of a linear increase over the period (model with a slope higher than 0). This means in practice we will exclude 0 as a possible value for the slope parameter in the calibration. Moreover, we will explore children.intercept around 1.6, 1.8 (for a slope of 0.09).

Regarding the couple creation, a probability looking for a partner valued 1 allows the model to fit reference data. The couple splitting process does not appear limiting. Then, we can say that two parameters (and two processes) to model the couple creation are perhaps not necessary and a constant is sufficient to model the split process. But it is too soon to conclude only on the base of the natural balance. Indeed, in the model, changes in couple have a large impact on residential moves, in conformity with the literature. Thus, we have to check the impact of these parameters on the migratory balance. That is what we do in the next section which is also dedicated to modelling the moving decision.

## How to model moving

Following the literature (Morand, Toulemon et al. 2010), dynamic microsimulation models are generally not spatial. Then, they do not consider the residential mobility. That is the case

of DESTINIE in France which simply considers the migration by adding individuals corresponding to the migratory balance. On the contrary, agent-based model and cellular automata can be easily spatial. However, agent-based models which take into account all main demographic processes have not yet been developed (Morand, Toulemon et al. 2010) even if they have been involved in particular dynamics where they appear as more convenient, especially partnership formation and spatial mobility. More recently hybrid models have been developed to combine advantages of the two modelling approaches. (Birkin and Wu 2012) review the existing hybrid approaches as for example SVERIDGE (Holme, Holme et al. 2004) which integrates inter and intra migration but requires a lot of data. (Coulombel 2010) in a very interesting review on residential choice and household behaviour outlines several shortcomings of the current approaches:

- The independence between the decision to move and the residential choice per se, and more generally the lack of retroaction between the demographic model and the mobility model;
- The decision to move is undoubtedly the most neglected aspect in the residential process, most models putting much more emphasis on the location choice;
- The location is generally decided using a discrete choice model which includes housing prices and housing and neighbourhood characteristics, then it is unclear whether this is a direct or indirect utility function;
- Regarding the location, the subset of alternatives is randomly and uniformly drawn from the whole set of vacant housings, disregarding any strategic consideration in the search process of the household;
- Migration and residential mobility are considered apart from each other, as independent decision while migration and residential mobility are often the same for people living close to the border.

Thus, we propose an integrated model of the mobility caused by distance of commuting, family events and satisfaction of the size of housing and susceptible to become a migration if a household decides to move outside the studied region. We studied possible functions for the decision to move and the searching procedure.

## **Which dynamics for deciding to move and modelling couples**

### ***Formalizing the questions and the dynamics to test***

We aim at parameterising the model in order to obtain a migratory balance, at least at the Cantal level, close to the reference. Thus we have to parameterise the moving decision of an individual living in Cantal. One can wonder why the individual decision to move is responsible for the whole migratory balance. That is what we begin to explain to allow the reader to understand how various processes depend on each other.

The elements to compute the migratory balance are:

- the number of in-migrants. In the model, the in-migrants correspond to the part of potential in-migrants which enter into Cantal by finding housing or a partner (to form a couple). The potential in-migrants are computed from adding the annual migratory balance (which is a parameter extracted from data) to the out-migrants of the previous

year; their distribution of ages is controlled following the distribution of ages given by data;

- The outmigrants are partly exogenously and endogenously defined :
  - students and old people move out from Cantal following probabilistic laws extracted from available data;
  - active people move following endogenous processes and are susceptible to move outside following probabilistic laws extracted from available data since they have decided moving. Then a large part of out-migration is explained though the decision to move.

From (Gobillon 2001; Minodier 2006), we know that in France the main reasons to move are firstly related to family events. Following (Debrand and Taffin 2005; Debrand and Taffin 2006) based on the analysis of data coming from various French surveys, creating a couple and splitting up a couple are the factors explaining most of the residential mobility. The second set of reasons is professional. (Debrand and Taffin 2006) notice that moving decreases with age. They point out that the short distance mobility is rather linked to the modification of the family structure while the long distance mobility is more often associated to professional changes. The third type of reasons concerns the change in the tenure (mainly between renters and owners) (Djefal and S. 2004) however this is not considered in the model for now because the decision to buy a house is a source of complexity that we chose to neglect. From this literature review, we retained some mechanisms which imply a decision to move occurs due to:

- the formation of a new couple;
- the split of a couple;
- a too long commuting time (higher than the *proximity* parameter) after a change of job, a new partnership or a move;
- a change in the housing satisfaction level due to a family event implying that the household is unsatisfied: we decide to capture this change through the compatibility of the housing size with the family size.

This stresses out that the creation and splitting couple processes have an impact on the migratory balance, not only on the natural balance. Thus, **we ask again the first and the second question presented in the previous section about the relevance of a two parameters approach for modelling couple creation and a constant probability to model the couple splitting process.**

All the listed events of the enumeration set at true the need for residence change attribute of the household. A part of them are age dependant in the model, changing job or having children for example. Then we can suppose that the age dependency of the move can emerge from the dynamics of the model without considering a complementary dependence to age in the decision. At the same time, literature also tells us the move to a large housing is much more common than the move to a smaller one. It means that probably the departure from a household does not lead to the same probability to move as for the arrival of a new member. It can be for example the case of the split of a couple, a young adult quitting the

parental housing or a death. If the first example cannot be linked to age easily – and is not in the model, the two last are age dependant, even in the model. Thus, it is possible that we need to take into account explicitly the age in the moving decision in order to avoid too much moves at an old age.

Thus the question is: **should the age be explicitly taken into account in at least part of the decision to move in order to generate a number of out-migrants and in-migrants for Cantal corresponding to the reference?**

To answer this question, we parameterise the moving decision related to the satisfaction about the current size of residence using an exponential law with a parameter and taking into account the average age of the adults of the household which has to decide. This parameter is called residence satisfaction parameter (*res. satisf. p* or  $\beta$  in the following mathematical formalisation). When it is valued 0, it means that the age is not directly taken into account in the moving decision. This average age enters into account in the decision when the parameter is higher than 0.

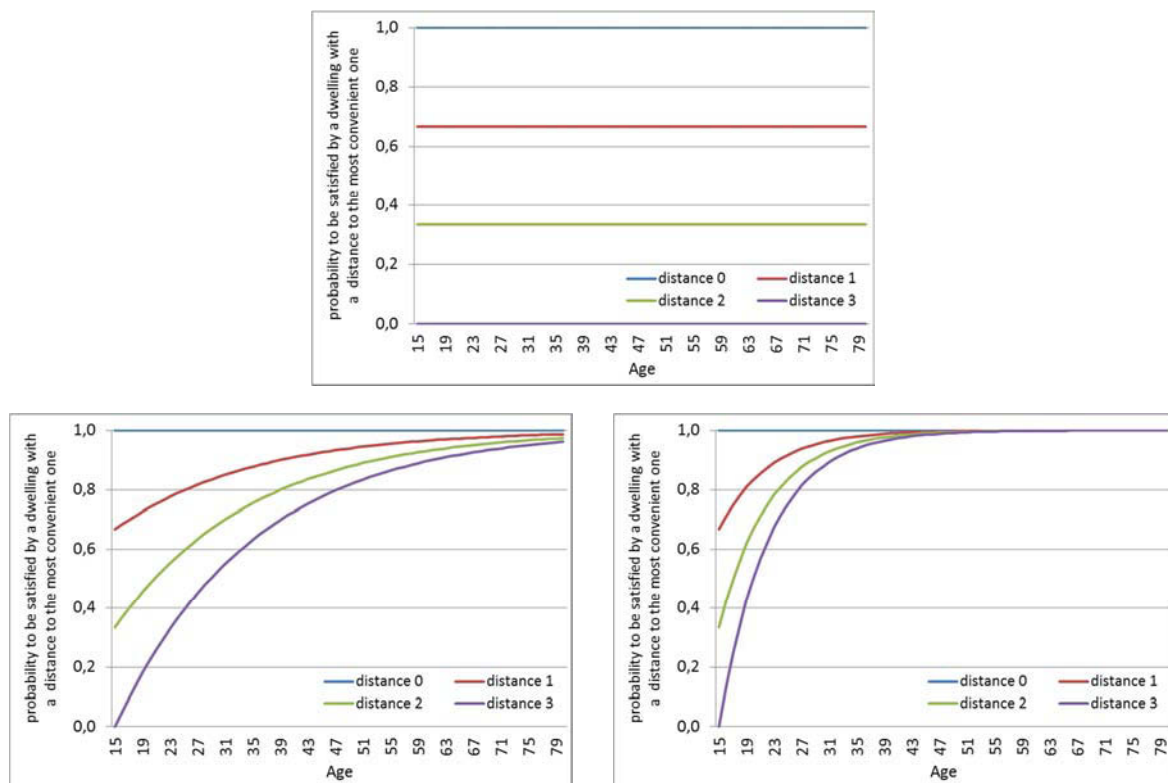


Figure 16. Forms of the probabilistic function deciding about the satisfaction on the current housings for various parameter values: on the top, *res satisf p* = 0 meaning the age does not play a role in the decision; on the left bottom, *res satisf p* = 0.05 and on the right bottom, *res satisf p* = 0.14 showing an increasing consideration for age leading individuals to move less when older.

We define:

- $l_d$ : ideal size of housing for the household size (i.e. number of rooms equals to number of individuals composing the household)
- $l_c$ : size of the currently occupied housing



- $i_p$ : size of a proposed new housing
- $nbSizes$ : total number of sizes considered in the model
- $\beta$ : parameter ruling the impact of the age;
- $a$ : average age of the adults in the household

from which, the probability to be satisfied  $ps$  by a current housing is:

$$ps = 1 - \left( \frac{|i_d - i_c|}{nbSizes} \exp(-\beta(a - 15)) \right)$$

The household is satisfied if  $ps$  is higher than, or equal to, a random number. The household need of residence attribute is set to true when the household is not satisfied of its residence. Figure 16 shows how the chosen function and the parameter value push the older household to decide not to move since they are more easily satisfied with the current size of their housing.

### **The experimental design**

The used experimental design is also the B corresponding to *res satisf p* values higher than 0. We add an experimental design C for *res satisf p* value equal to 0.

#### **Experimental design C:**

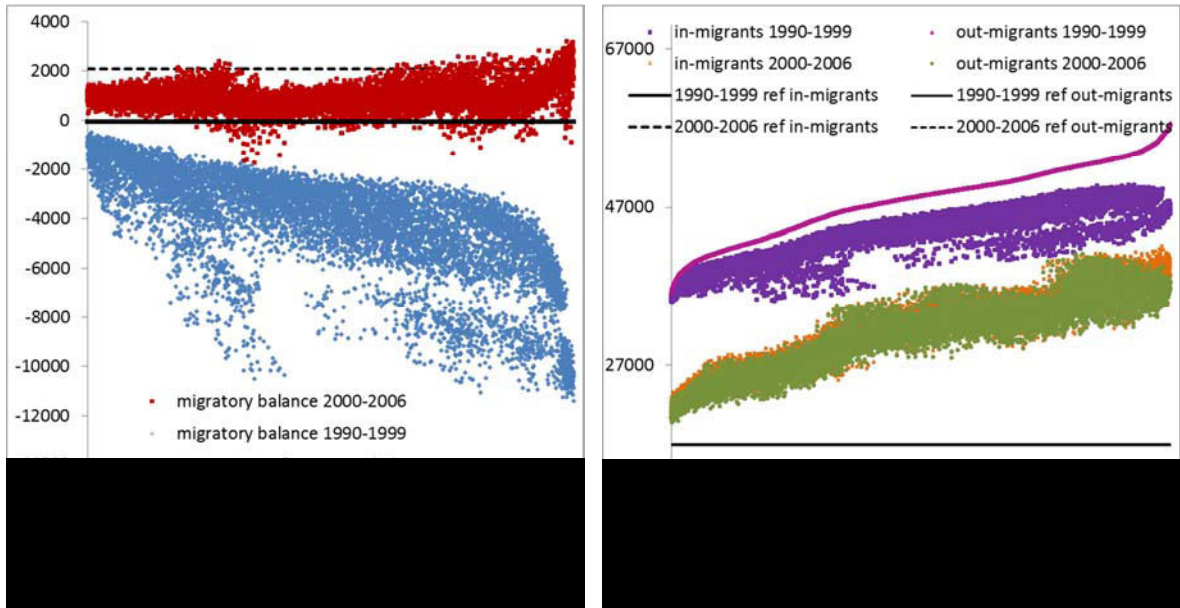
<i>Code param</i>	<i>Name</i>	<i>Values</i>
<i>nb child slope</i>	<i>slope of the giving birth function</i>	0.09
<i>nb child intercept</i>	<i>intercept of the giving birth function</i>	1.0; 1.2; 1.4; 1.6; 1.8
<i>P search partner</i>	<i>probability to search for a partner</i>	0,1; 0,35; 0,6; 1
<i>split p</i>	<i>probability to split</i>	0.002; 0.025; 0.034; 0.05
<i>join trials</i>	<i>number of join trials</i>	1; 5
<i>Proximity</i>	<i>proximity threshold</i>	21; 33; 45
<i>dispo threshold</i>	<i>house availability threshold</i>	0; 0.1; 0.15; 0.2; 0.25; 0.3; 0.35; 0.4
<i>res satisf p</i>	<i>residence satisfaction</i>	0

The total number of experiments of the design C is 8960. Each experiment is replicated once.

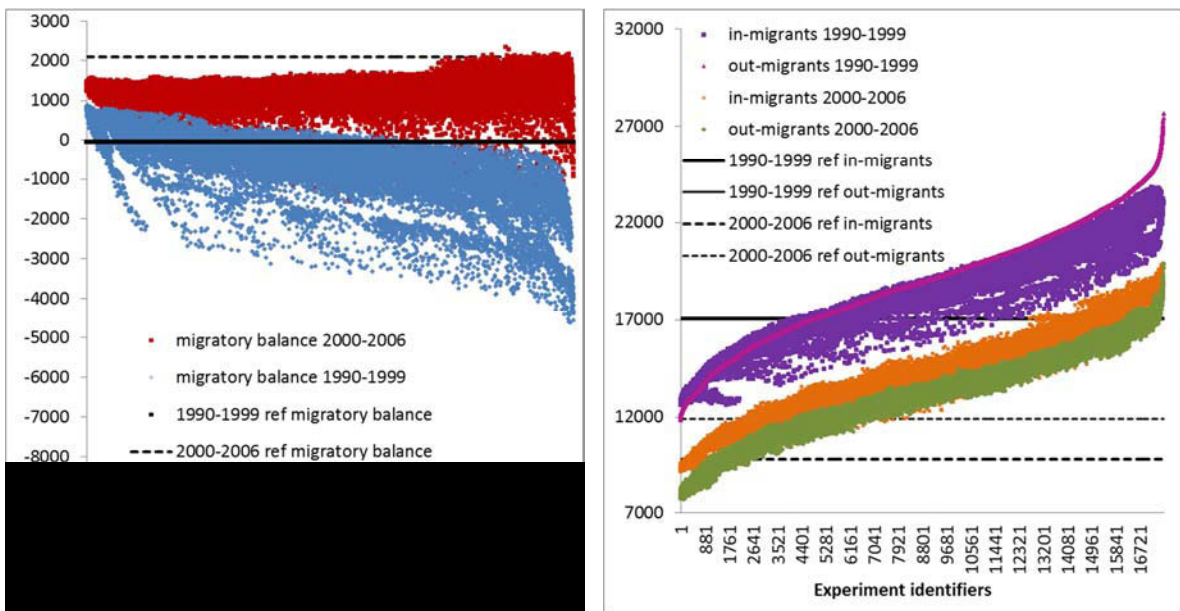
### **Choosing the dynamics**

Figure 17a shows that the reference migratory balances are not possible results of the model. The balance is too much negative for 1990-1999 (the least negative value is -487, to compare to the reference -50). The reference for 2000-2006 is a possible result of the model (reference value 2091) but only for very negative value of the balance for 1990-1999. To better understand why the migratory balance is not a possible result of the model, let's have a look on the in and out migrants. Figure 17b shows that there are too much in and out migrants for the two periods. References are not possible results from the model since the minimum results for the model is at least 3000 individuals too large.





Figures 17. For the experimental design C (*res satisf*  $p = 0$ ): (a) on the left, every values of migratory balance for 1990-1999 (blue) and 2000-2006 (red) - the reference values are -50 for 1990-1999 and 2091 for 2000-2006; (b) on the right, every values of out (light purple) and in (dark purple) migrants for 1990-1999 and out (green) and in (orange) migrants for 2000 - the reference values are 17075 outmigrants and 17025 inmigrants for 1990-1999 and 9814 outmigrants and 11905 inmigrants for 2000-2006. All the results are ordered following the increasing number of out-migrants for 1990-1999. In black are the references: plain for 1990-1999 and dotted for 2000-2006.



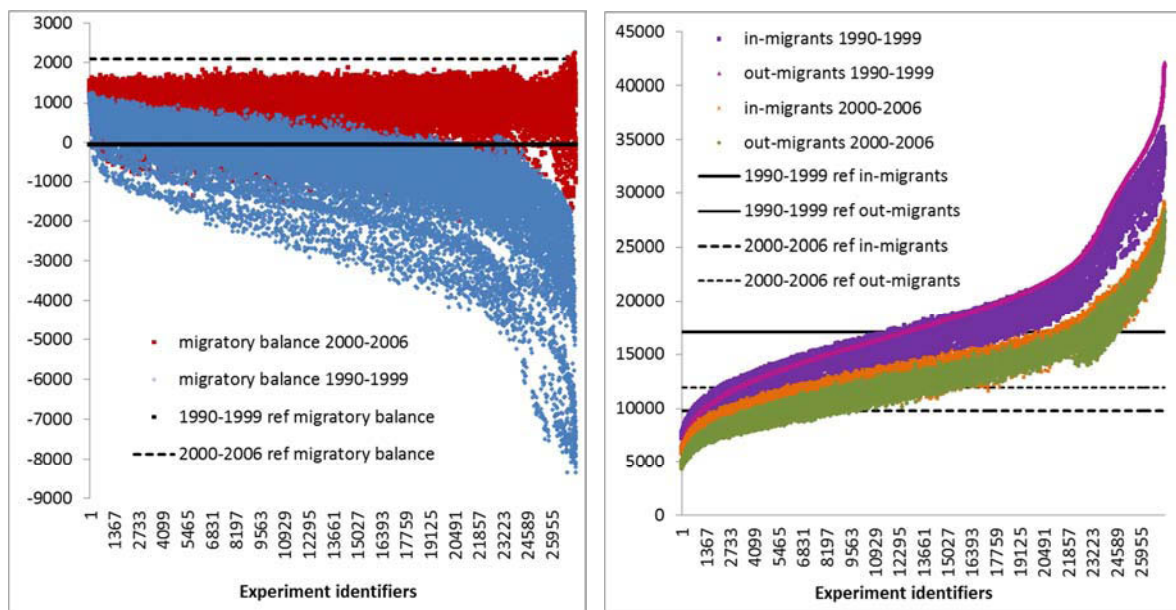
Figures 18. For the experimental design B (*res satisf*  $p > 0$ ): on the left, every values of migratory balance for 1990-1999 (blue) and 2000-2006 (red) - the reference values are -50 for 1990-1999 and 2091 for 2000-2006; on the right, every values of out (light purple) and in (dark purple) migrants for 1990-1999 and out (green) and in (orange) migrants for 2000 - the reference values are 17075 outmigrants and 17025 inmigrants for 1990-1999 and 9814 outmigrants and 11905 inmigrants for 2000-2006. All the results are ordered following the increasing number of out-migrants for 1990-1999. In black are the references: plain for 1990-1999 and dotted for 2000-2006.

Figures 18 show the results for  $res\ satisf\ p > 0$ , all ordered following the increasing value of the out-migrants for 1990-1999. We clearly see that there are several sets of parameter values allowing the model to obtain a result close to the references.

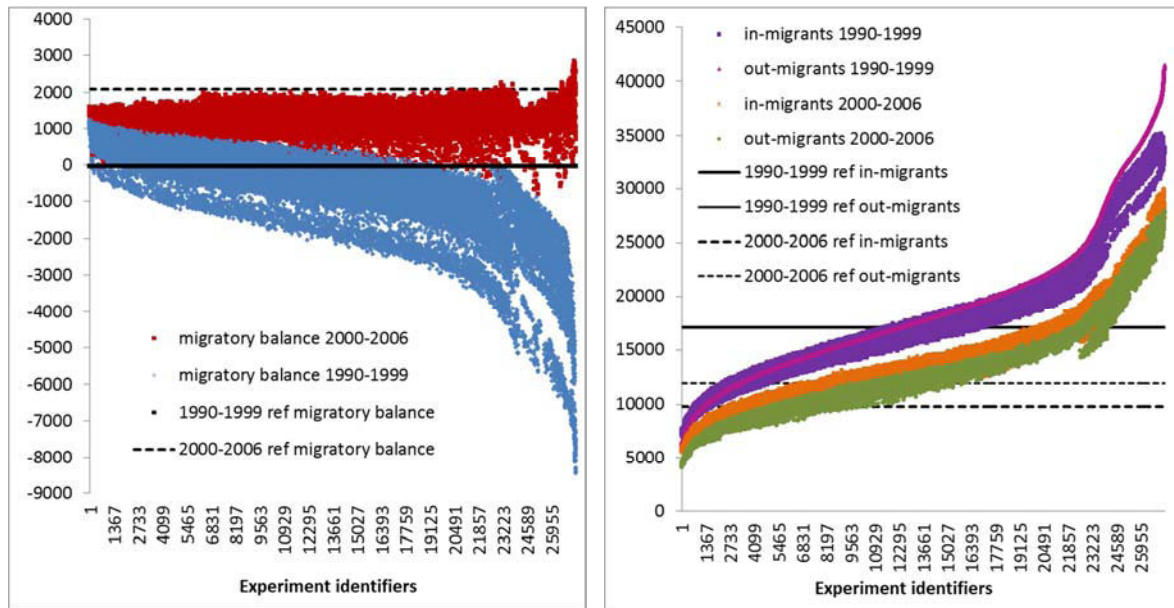
The references are not results of the model for  $res\ satisf\ p = 0$ , probably because too many people move outside Cantal. It means that the size of the current housing is not sufficient to model the moving decision. The purpose is now to check if taking into account the age allows the model to reach the reference value in the results of the experimental design B. That is what we expect since it should decrease the number of out-migrants at the same time it decreases the number of individuals deciding to move.

We now consider the couple creation processes to see if suppressing the probability to search for a partner (by putting the probability to 1) allows the model to yield results close to the references.

From the figure 19, we conclude that the references for 2000-2006 are not possible results from the model with a probability to search for a partner equal to 1. The migratory balance in 2000-2006 (2091) can be obtained only for very large number of in and out migrants, between 25000 and 30000, while the references are respectively 11905 and 9814.



Figures 19. For the experimental design B and a probability to search for a partner = 1: on the left, every values of migratory balance for 1990-1999 (blue) and 2000-2006 (red) - the reference values are -50 for 1990-1999 and 2091 for 2000-2006; on the right, every values of out (light purple) and in (dark purple) migrants for 1990-1999 and out (green) and in (orange) migrants for 2000 - the reference values are 17075 outmigrants and 17025 inmigrants for 1990-1999 and 9814 outmigrants and 11905 inmigrants for 2000-2006. All the results are ordered following the increasing number of out-migrants for 1990-1999. In black are the references: plain for 1990-1999 and dotted for 2000-2006.



Figures 20. For the experimental design B and a probability to search for a partner  $< 1$ : on the left, every values of migratory balance for 1990-1999 (blue) and 2000-2006 (red) - the reference values are -50 for 1990-1999 and 2091 for 2000-2006; on the right, every values of out (light purple) and in (dark purple) migrants for 1990-1999 and out (green) and in (orange) migrants for 2000-2006 - the reference values are 17075 outmigrants and 17025 inmigrants for 1990-1999 and 9814 outmigrants and 11905 inmigrants for 2000-2006. All the results are ordered following the increasing number of out-migrants for 1990-1999. In black are the references: plain for 1990-1999 and dotted for 2000-2006.

On the contrary, figures 20 show that fitting the references is possible for a probability to search for a partner  $< 1$  but we have to notice that probably few parameter sets allow the model to reach them.

### ***Parameter bounds and other indicators***

In order to get further, trying to define which values allow the model to reach the reference value, we perform an ANOVA over the results of the experimental design B. The purpose is to identify the parameters ruling the migratory balances, the in-migrants and the out-migrants.

Figures 21 show the results of the ANOVA for the indicators in-migrants, out-migrants and migratory balances resulting from the execution of the experimental design B. In and out migrants are mainly sensitive to *res satisf p*, *split* probability and *proximity* threshold. The migratory balance is sensitive to many different parameters changing from the first period to the second one: *res satisf p*, *dispo threshold*, *split* probability, *p search for partner* and *join trials*.

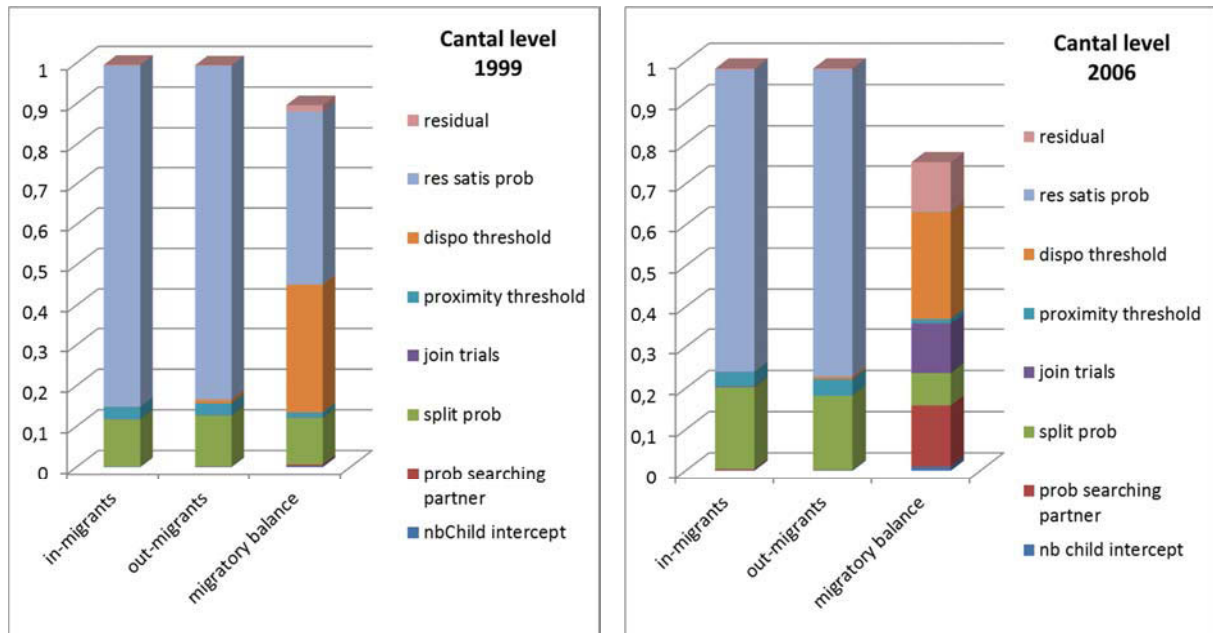


Figure 21. ANOVA for the in-migrants, out-migrants and migratory balance in 1999 and 2006 (for 1990-1999 and 2000-2006), obtained from the results of the experimental design B. It represents effect of one parameter and residual (cross effect of two parameters are not represented).

From these selections, we can consider a subpart of the results of the experimental design B, shown on the figures 22 and 23. 1990-1999 results are represented in blue while 2000-2006 are represented in orange. Straight lines represent the reference value in order to visualise the distance between the references and the results of the model. Plain lines represent the in-migrants and dotted lines represent the out-migrants or the migratory balance.

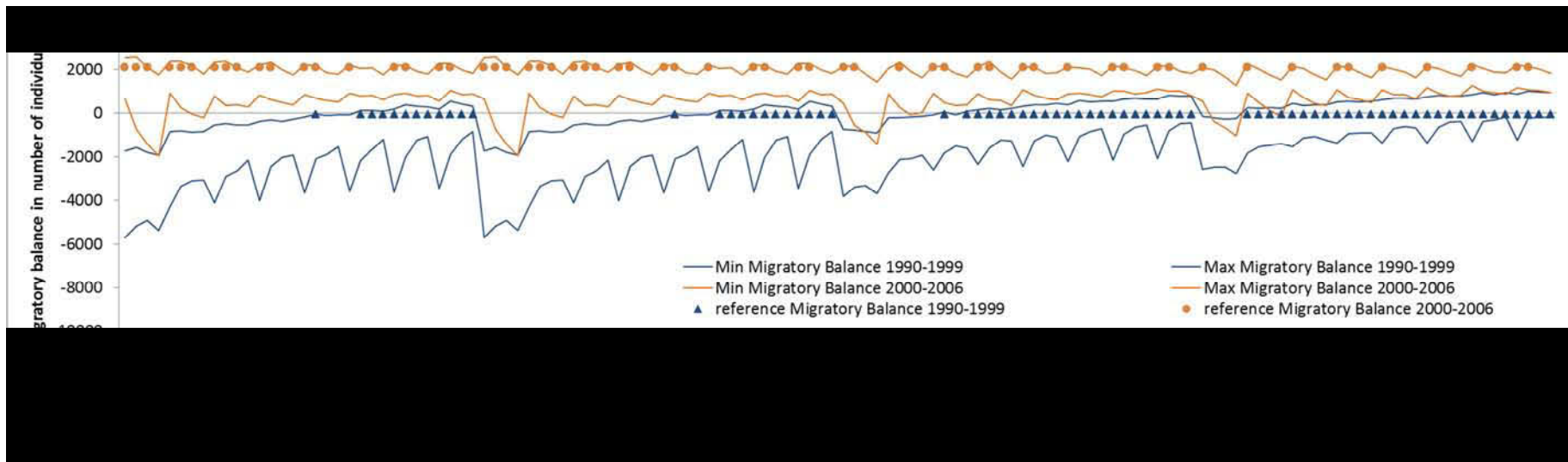


Figure 22. Selection of migratory balances obtained from the execution of the experimental design B and references for 1990-1999 and 2000-2006. Abscissa: from the bottom to the top the *res satisf p*, the *dispo. threshold*, the probability to search for a partner

Figure 22 shows the references for the migratory balances are not possible results of the model. Good values of the parameter *res. satisf p* are 0.09 and 0.11 and *dispo. threshold* should at least be valued 0.2. The migratory balance for 2000-2006 cannot be fitted for a probability to search for a partner equal to 1. A deeper analysis shows that the larger the probability to search for a partner, the more numerous are couples and less numerous are singles. Indeed, potential in-migrants can enter into Cantal as effective in-migrants only by finding a partner or a housing. The smaller number of singles resulting from a larger probability to search for a partner makes potential in-migrants less susceptible to enter into the Cantal. Then the migratory balance cannot be high enough to be equal to the reference.



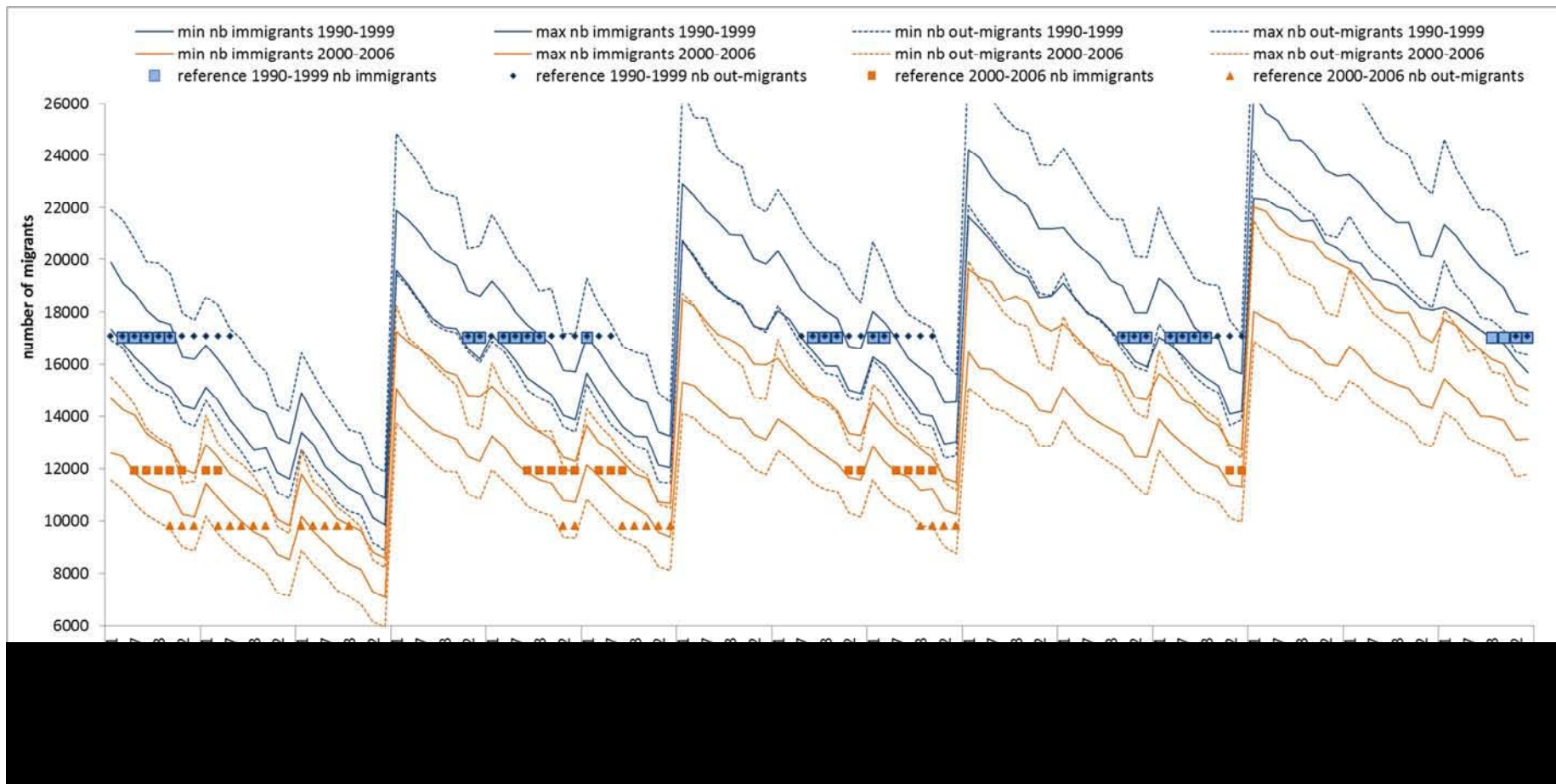


Figure 23. Selection of in and out migrants obtained from the execution of the experimental design B and references for 1990-1999 and 2000-2006. Absissa: from the bottom to the top the *res satis p*, the *split p* and *proximity*.

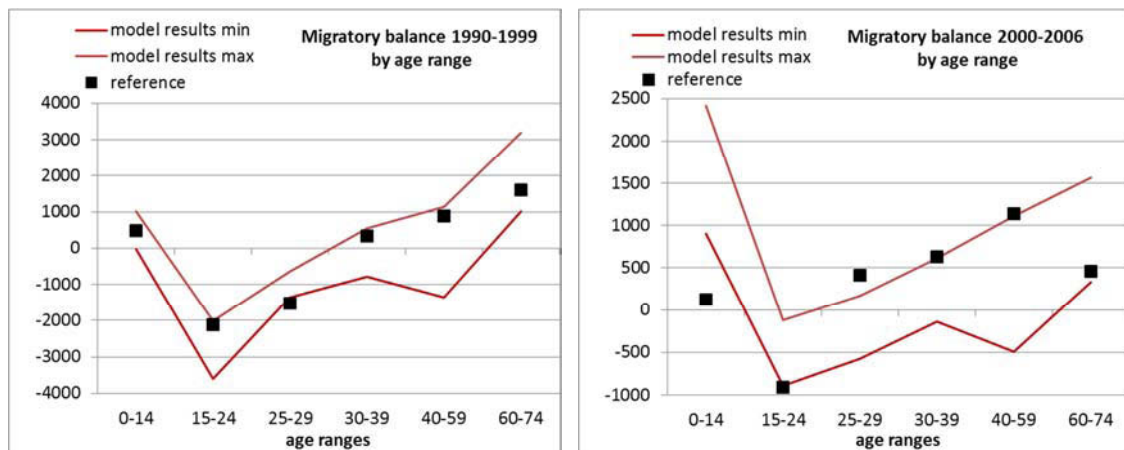
Also, the figure 23 showing the results of the model and the references indicates that some results are close to the reference even if they do not fit exactly. It is especially the case when *split* is valued 0.018 or 0.025 and *res satis p* is equal to 0.09 or 0.11. The *split* mechanism based on a constant doesn't seem to limit the capacity of the model to produce results close to the references.

Some complementary checking can be done to better assess our choice by looking how varies the number of movers (number of individuals moving in the Cantal, wherever they move) and the migratory balance by age range. We did these checking considering a subset of results coming from the execution of the experimental design B restricting to the values we conclude as good from the previously presented work (*res satisf p* = 0.09 or 0.11; *dispo threshold* equal or higher than 0.2; *nbChild.intercept* = 1.2 or 1.4 or 1.6 or 1.8; probability to search for a partner higher than 0.1).

Regarding the movers, the references are far from being comprised in the interval of results defined by the minimum and the maximum results for this indicator. Indeed:

- while the reference from 1990 to 1999 is 133459, the results vary from 43202 to 120230;
- while the reference from 2000 to 2006 is 116371, the results vary from 29332 to 86684.

However, figures 24 show it is possible to obtain a relatively correct migratory balance by age range.



Figures 24. Minimum and maximum migratory balance by age ranges obtained from a relevant subset of parameter set from the experimental design B (red lines) compared to reference (black squares) for 1990-1999 on the left and 2000-2006 on the right

To conclude regarding the choice of the dynamics for modelling the decision to move in the Cantal, we retain that it is necessary to directly take into account the age in the decision function. Thus we keep the function we choose to test our hypothesis with a parameter value higher than 0.

This means in practice the 0 won't be kept as a possible value for the parameter *res satisf p* in the next phase of the definition of the parameter value which is the calibration. Moreover, we know good values for *res satisf p* are around 0.09 and 0.11. Similarly, it seems *dispo threshold* has to be chosen equal or higher than 0.2.

We come back to the couple creation mechanism. Even if we don't need to limit the individual probability to search for a partner to obtain a close-to-reference number of births, we shall do it to obtain convenient results for the migratory balance and the in and out migrants for 2000-2006.

Regarding the splitting couple mechanism, a constant seems sufficient at least to allow the model to produce results close to the chosen references.

## **Which dynamics for deciding where searching a residence or a partner?**

### ***Formalizing the questions and the dynamics to test***

The residential location decision and research remains a challenge in modelling. That is what (Coulombel 2010) concludes in his exhaustive state-of-art regarding residential choice and household behaviour.

In our model, we have decided from data analysis and literature that three dynamics lead to moving in the space of Cantal:

- finding a partner to form a couple;
- finding a job leading to commuting distance creating a spatial relation between two municipalities: the one of residence and the place of work;
- finding a new housing.

From the section presenting the Cantal and its demography, we retain several indications allowing us to hypothesize on the way people move:

1. the residential and the commuting moving distances are very small, mainly at a distance 0 (see the figure 7), or very close (mainly less than a distance 9).
2. we observed an heterogeneous spatial evolution of the number of inhabitants (see the figure 6): municipalities around the three largest municipalities tend to increase while the other decrease;

The second indication is in coherence with the first one: then a first assumption would be that people search in the space starting from the place they live or the place they work and going further and further only if they are not able to find out what they are searching for. In other words, they take a job, a residence or a partner as close as possible from their starting place of research.

Thus, a first question is: **is a dynamic only considering a research as close as possible sufficient to obtain an increase of the population only situated around the larger municipalities? or is it necessary to identify a specific dynamic for larger cities?**

In order to answer this question, we consider a mechanism excluding the three largest cities from the possible places of search with a given probability each time an individual searches for housing. In practice, this probability, which is a parameter of the model, is compared to a random number to decide if the city is excluded or not from the space of research. This parameter is called *dispo. threshold*. From a semantic point of view, we assume cities are mainly composed from flats; and houses are rare and expensive. Then, a household which wants to have a house with a garden should search elsewhere due to the low availability leading to a high price of such housing. It is particularly the case for young couples who want



to form a family and to have a house. That is why the exclusion of the searching space is only applied for household having a least the size two.

We are able to answer our question because if the parameter *dispo. threshold* is 1, the largest municipalities are never excluded from the searching space. If it values less than 1, they are sometimes excluded. Then, we will be able to know if the heterogeneous spatial evolution can be reached only from *dispo. threshold* equal 1.

Another question is much more related to the computational cost. Indeed, searching further and further, collecting every possible offer is very costly from a computational point of view. We assume that an individual searches for a job at most 51 km, considering that the part of people commuting further is negligible. We can thus consider the maximum distance to search for housing can be smaller and try to gain in computational cost considering a maximum distance for searching for housing or partner.

Thus the second question is: **is it possible to limit the search for a partner and a residence at a maximum distance smaller than the commuting distance? Does it allow the model to reach the reference values collected by the French National Statistical Office in 1999?**

In order to answer this question, we use a parameter called *proximity* threshold which bounds the spatial research for housing and/or partner. If a search reaches this threshold without any success, the individual stops searching until the next year.

Working and residence places are also determined by commuting. We have seen in the previous chapter the decision to search for a job is parameterised by laws extracted from data. Regarding the place where individuals search for a job, we assume the same simple heuristic as for searching housing: the individual aims at finding a job as close as possible from her place of residence. We arbitrary limit the area of research to a maximum Euclidian distance of 51 kilometres. Indeed, from the commuting distance distribution extracted from data of the 1999 Census for Cantal, we know that 99.5% of people living in Cantal work at most at 51 km from their residence location. As for housing and partner searching, a searcher is constrained by the spatial distribution of the availability of what she searches for, a job. The offer of jobs is exogenously defined while the occupation is endogenously updated.

The whole searching residence procedure is described in the algorithm 2 in pseudo-code.

### ***The experimental design***

The experimental design D is larger and considers more values for *dispo threshold* while the experimental design E consider more values for *proximity*. The total number of experiments of the design D is 103680 and the total number of experiments of the design E is 6912. Each experiment is replicated once.

Algorithm 2. Pseudo-code describing how a household search for a new residence (the finally chosen new place of residence is picked out at random in muns – the household moves there if it does not decide moving outside the region)

```

// hh is a household which has decided to move; center is the starting place of research defined by
the leader (one of the adult of the household, randomly chosen or design through the process telling
she works at a distance higher than proximity
getPlacesToReside(Household hh, Municipality center)
  for i=3 ; i <= proximity ; i=i+3
    subnet = collectEveryMunLocatedAt(minDist i-3, maxDist i, from center)
    List muns = getAvailableResidenceMun(hh, subnet)
    if muns.size() > 0
      for each Municipality munic of muns
        if leader.hasAWorkingPartner
          if dist(munic, partner.workingPlace) > proximity
            muns.remove(munic)
      return muns
=====
getAvailableResidenceMun(Household hh, List subnet)
  List possiblePlaces
  dispolev=true
  for each Municipality mun of subnet
    if hh.getSize() > 1
      if mun.name == 15014 or 15087 or 15012 or 15120 // (the four largest cities of Cantal)
        if random > dispoTreshold
          dispolev = false
  if (dispolev)
    count = mun.getFreeConvenientResidence(hh).size()
    for 0 to count
      possiblePlaces.add(mun)
  return possiblePlaces

```

#### Experimental design D

Short name of the parameter	Description of the parameter	Values of the experimental design D	Values of the experimental design E
<i>nb child slope</i>	<i>slope of the giving birth function</i>	0.09	0.09
<i>nb child intercept</i>	<i>intercept of the giving birth function</i>	1; 1.2; 1.4; 1.6; 1.8; 2.0	1.2; 1.6; 2.0
<i>P search partner</i>	<i>probability to search for a partner</i>	0.1; 0.35; 0.6; 1	0.35; 1
<i>split p</i>	<i>probability to split</i>	0.002; 0.018; 0.025; 0.034; 0.05	0.002; 0.05
<i>join trials</i>	<i>number of join trials</i>	1; 5	1; 5
<i>proximity</i>	<i>proximity threshold</i>	21; 24; 27; 30; 33; 45	9; 15; 18; 21; 24; 27; 30; 33; 36; 42; 48; 51
<i>dispo threshold</i>	<i>house availability threshold</i>	0; 0.1; 0.15; 0.2; 0.25; 0.3; 0.35; 0.4; 0.55; 0.7; 0.85; 1.0	0; 0.2; 0.4; 0.55; 0.85; 1.0
<i>res satis prob</i>	<i>residence satisfaction</i>	0.03; 0.07; 0.09; 0.11; 0.14; 0.16	0.03; 0.07; 0.09; 0.14

## Choosing the dynamics

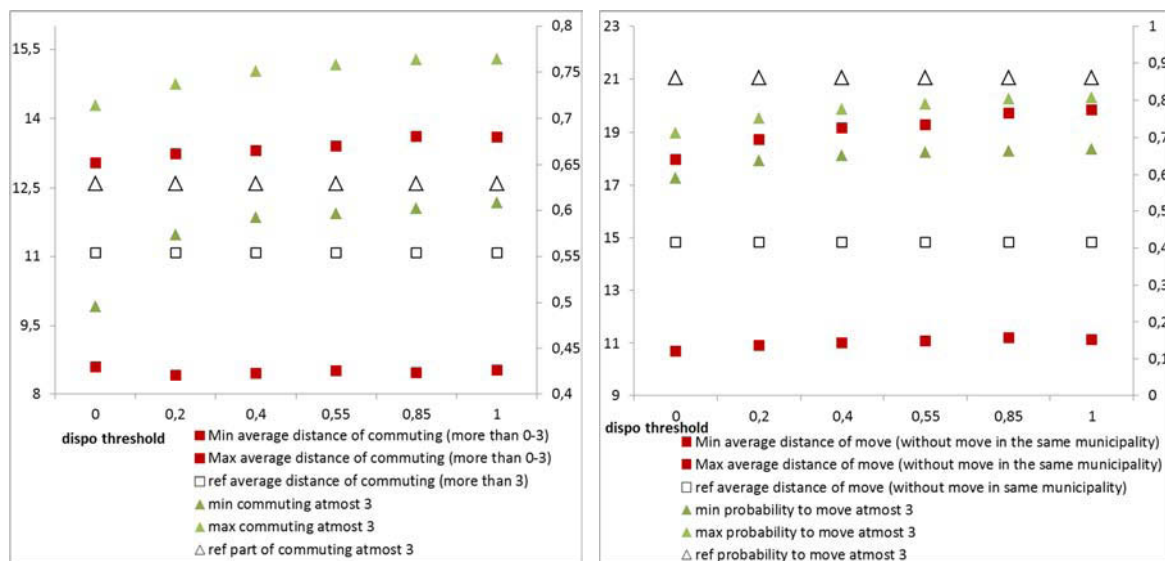
In order trying to disentangle the impact of the modelling choice regarding residential mobility and search for a partner dynamics, we are going to study how parameters impact the following indicators:

- The probability to move at most 3 Euclidian km knowing the individual has already decided to move;
- The average residential move distance (excluding move at most 3 km corresponding roughly to remain in the same municipality);
- The probability to commute at most at 3 km knowing the individual has a job;
- The average commuting distance (excluding distances lower than 3 km which correspond roughly to commuting in the municipality of residence).

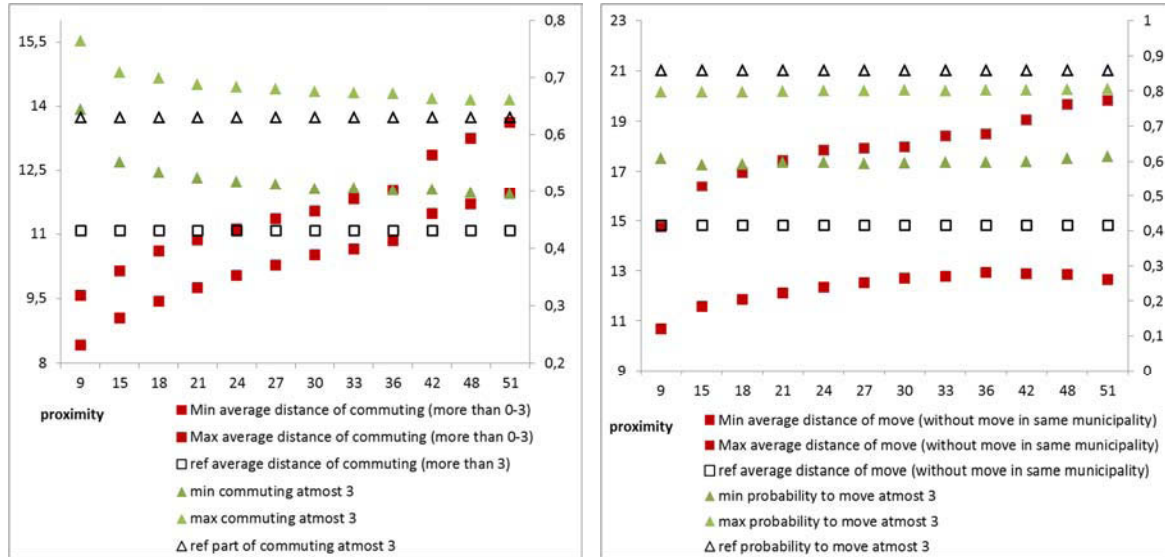
We know from INSEE the values in 1999 of these indicators (called later references), but not for 2006. As previously, we begin by an overview of the model behaviour for these indicators in comparison with the references using the experimental design E.

Figures 25 investigating the impact of the parameter *dispo threshold* on our indicators do not argue in favour of a specific dynamic for larger cities. Indeed, the model shows the same capacity to produce (or not to produce) results close to references either for *dispo threshold* equal 1 and for *dispo threshold* lower than 1.

On the other hand, we notice that the probability to move in the same municipality (roughly equal to move at most a distance 3) cannot be reached by the model (figures 25b and 26b). The temporality of offers and demands for housing is very simplified in the model. Indeed, in the model, a housing is available only when it is empty, whereas we can say that in France renters begin to search for a housing about three months before quitting their own housing, at the same time announcing their housing as soon available. That can perhaps explain why it is difficult to fit the reference value.



Figures 25. Minima and maxima values of commuting indicators (a on the right) and moving indicators (b on the left) for various value of the parameter *dispo threshold* from 0 to 1: red square, average distance of commuting (at a distance more than 0-3) (on right), average distance of moving (without move in the same municipality) (on left), black square, corresponding references; green triangle, probability to commute (on right), to move (on left) at most a distance 3, black triangle, corresponding references. Model results come from the experimental design E.



Figures 26. Minima and maxima values of commuting indicators (a on the right) and moving indicators (b on the left) for various value of the parameter *proximity* from 9 to 51: red square, average distance of commuting (at a distance more than 0-3) (on the right), average distance of moving (without move in the same municipality) (on the left), black square, corresponding references; green triangle, probability to commute (on right), to move (on the left) at most a distance 3, black triangle, corresponding references. Model results come from the experimental design E.

Figures 26 clearly show that it is not only possible, but necessary to limit the search for a partner and a residence at a maximum distance smaller than the maximum distance to search for a job (51). On the left, the reference for the average distance of commuting is embedded by the model results only when *proximity* is valued from 24 to 36. As previously, we see the reference probability to move at a distance lower than 3 km is not a possible result of the model. Moreover, *proximity* equal to 9 is not a relevant value.

### Parameter bounds and other indicators

To go further, an ANOVA selects the parameters impacting the indicators about spatial moves.

### The commuting indicators

We begin by considering the results of an ANOVA performed on the results obtained from the experimental design E. Figures 24 are dedicated to the indicators related to commuting.

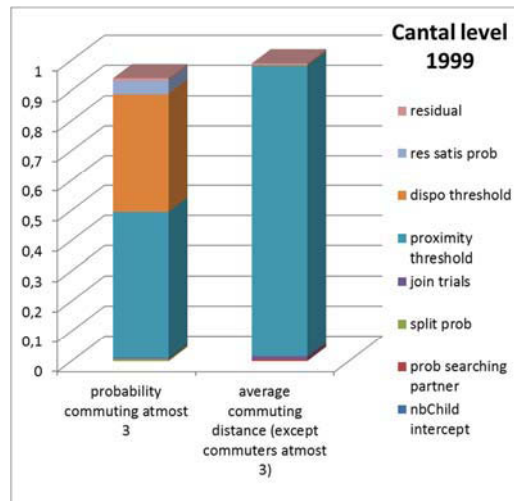


Figure 27. ANOVA for the probability commuting at most 3 km and the average commuting distance (except commuters at most 3 km) for results in 1999 and 2006, obtained from the execution of the experimental design E. It represents effect of one parameter and residual (cross effect of two parameters are not represented).

Figure 27 shows the probability of commuting at a distance lower than 3 km is sensitive to *proximity* and *dispo threshold*. This sounds right for the *dispo threshold*. Indeed, the commuting inside her place of residence is for numerous active individuals working in the three largest municipalities ruled by the possibility to choose these same largest municipalities as a place of residence. This can explain why the probability of commuting shorter than 3, meaning quasi in her municipality of residence, is sensitive to *dispo threshold*. It is less intuitive for the *proximity*: how can the largest distance at which someone can search for a housing impact the probability to work and live in the same municipality? Looking further in the results, we observe the probability to commute at a distance lower than 3 decreases when the threshold *proximity* (limiting the largest distance of search) increases. Enlarging the area of research makes the chance of finding a convenient housing at the first trial (or for a small number of trials) higher. Therefore the household searches for a housing less frequently when the threshold *proximity* is large and thus she tries a smaller number of times to find out one in her working municipality (from where occupied active individuals begin to search for a housing) leading to a smaller probability to commute very close.

Also, figure 27 shows that the average commuting distance is only sensitive to *proximity* as we can expect. That means that the *dispo. threshold* has no impact on the commuting distance distribution except at a distance 0 since it rules the capacity of workers of main municipalities to live where they work.

We are now able to check if the results of the experimental design E are close to the references when the parameter impacting the indicators varies. We only investigate the probability of commuting shorter than 3 since the figure 26a already presents the evolution of the average commuting distance when the *proximity* varies.

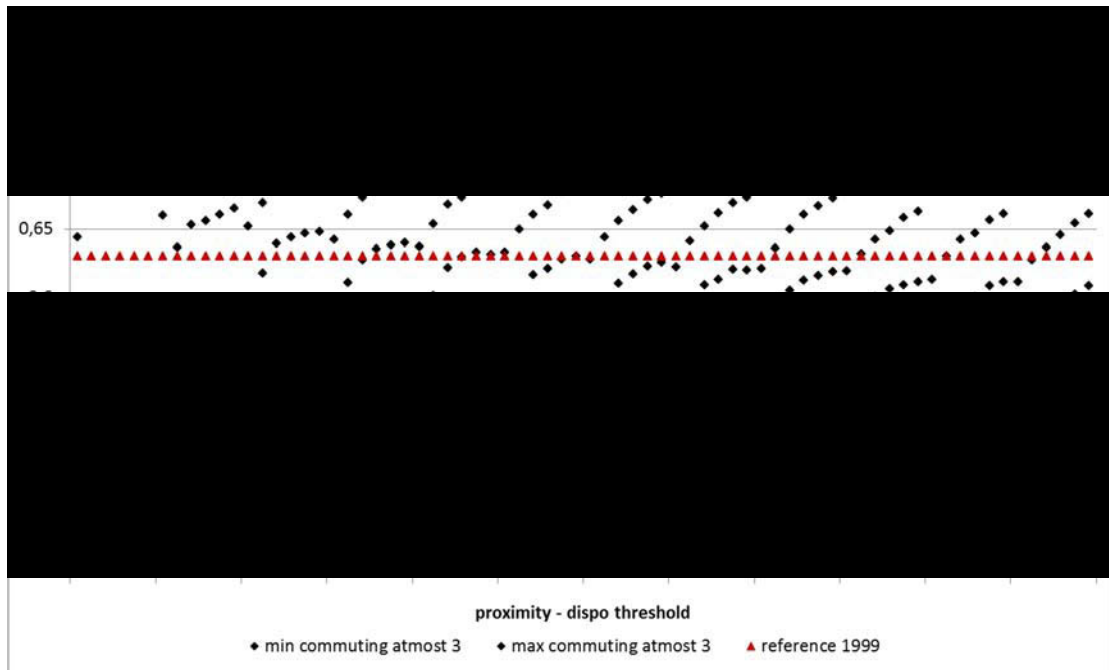


Figure 28. Probability of commuting shorter than 3 km in 1999 (reference in red, minimum and maximum simulated values from experimental design E in black) – Abscissa, from bottom to the top: proximity, dispo threshold

Figure 28 shows in red the reference value and in black the minimum and maximum probabilities to commute at most 3 km obtained from the variations of the *proximity* and *dispo threshold* presented in abscissa and the variation of the other parameters of the experimental design E. We observe that the reference seems easily reachable except for small values of *proximity* (for which only very small values of *dispo threshold* contains the reference). The values of *proximity* which are higher than 30 seem more reliable because they match the reference for more *dispo threshold* values (except when *dispo threshold* equals 0).

Coupled to what we already know from the figure 26a, it means that the *proximity* should be valued from 24 to 36 and *dispo threshold* should be higher than 0 to fit the references for commuting.

### *The moving indicators*

Figure 29 presents the results of the ANOVA for the residential move indicators. It shows the probability of moving shorter than 3 km is ruled by *dispo threshold* and *res satisf p*. The average moving distance (without considering distances smaller than 3) is sensitive to *proximity*, *res satisf p* and split probability. Split has an impact since it is the member of the couple working the furthest from the place of residence who moves elsewhere in the model. The impact of *res satisf p* is related to the fact it partly defines the number of movers. Then it impacts the availability of residences.

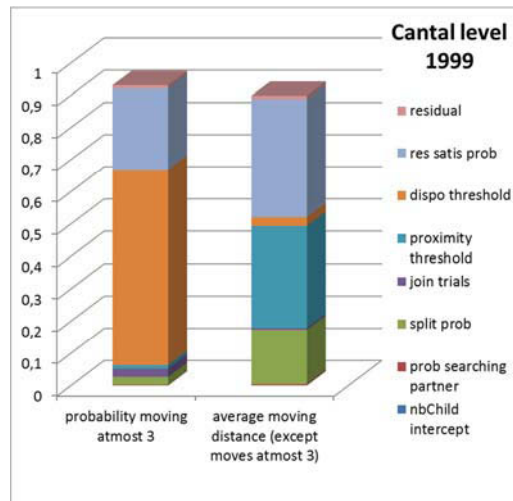


Figure 29. ANOVA for the probability of moving shorter than 3 km and the average moving distance (except commuters at most 3 km) for results in 1999 and 2006, obtained from the execution of the experimental design E. It represents effect of one parameter and residual (cross effect of two parameters are not represented).

From this selection of parameters, we can study the variation of indicators due to these parameters. We already know from the analysis of figures 25b and 26b that the model is not able to produce a probability to move at a distance smaller than 3 km equal to the reference (even if it is not so far). Figure 30 shows it is closer when *res satisf p* is low and when *dispo threshold* is high. That makes sense since the population is low when *res satisf p* is low and the largest municipalities offering numerous housing are never excluded from the space search for housing when *dispo threshold* is 1.

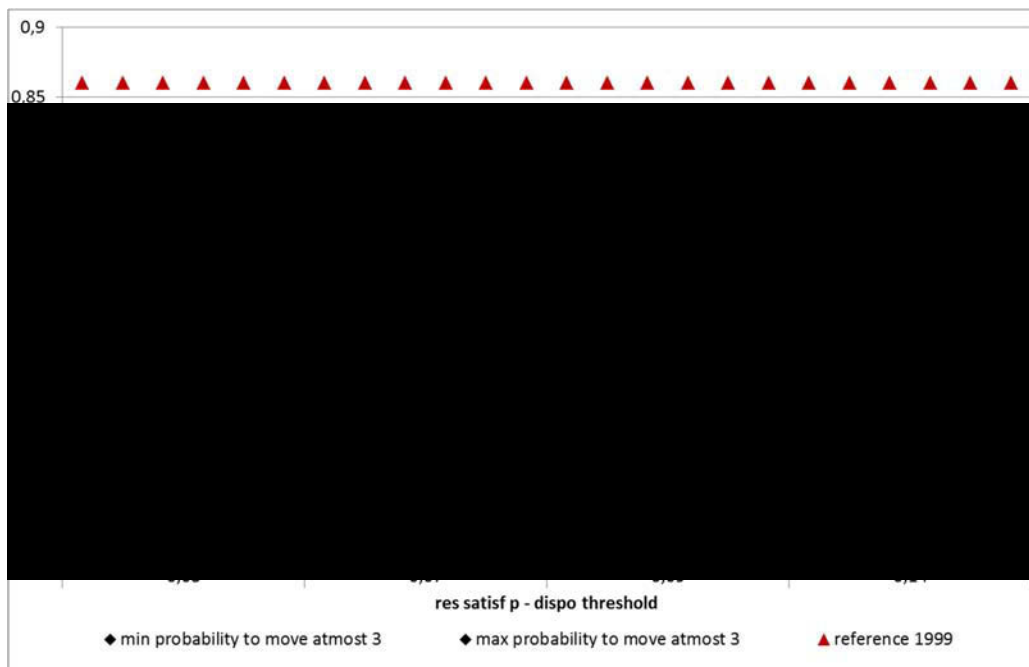


Figure 30. Probability of move shorter than 3 km in 1999 (reference in red, minimum and maximum simulated values from experimental design E in black) - Abscissa, from bottom to the top: *res satisfy p*, *dispo threshold*

Figure 31 shows there are large interaction between the three parameters *split*, *res satisf p* and *proximity*. Numerous various sets of values of these three parameters allow producing the reference. Then we can't conclude about relevant values.

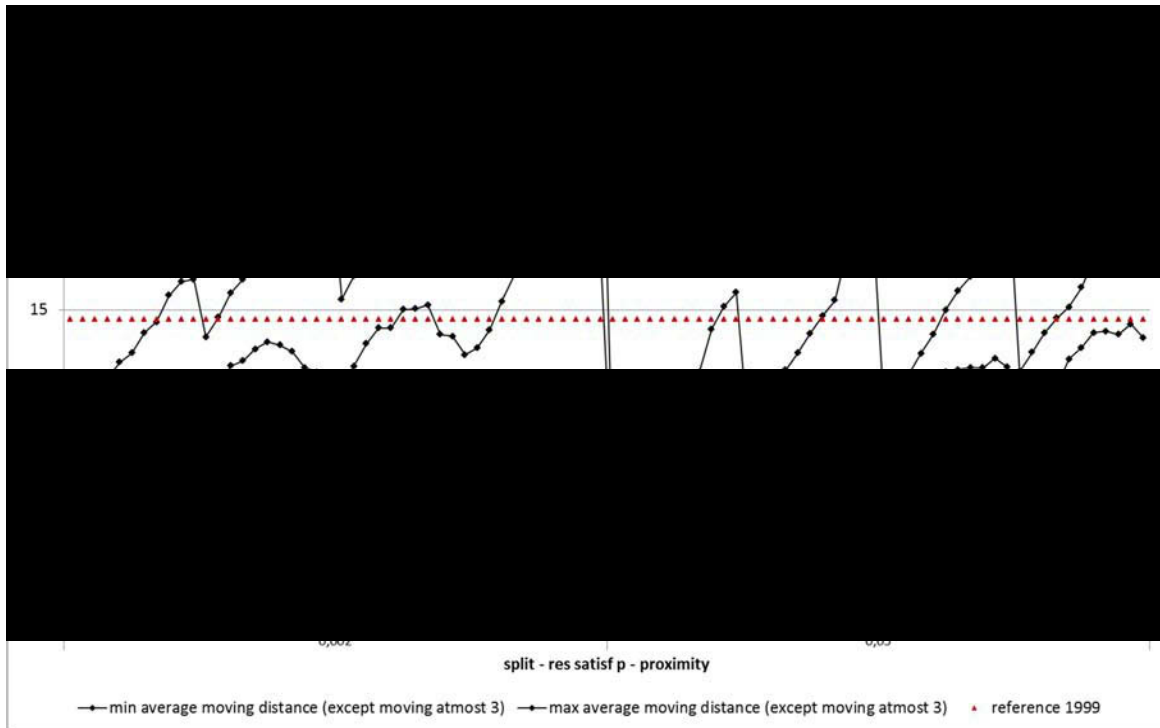


Figure 31. Average moving distance (except moves shorter than 3 km) in 1999 (reference in red, minimum and maximum simulated values from experimental design E in black) - Abscissa, from bottom to the top: *split*, *res satisf p*, *proximity*

From what we saw until now, it appears that *proximity* has to be comprised between 24 and 36 km. We don't know a lot about *dispo threshold* except it must not value 0. In particular, we don't know if the largest municipality exclusion mechanism is necessary to fit the reference. In order to answer this, we check if the results of simulations are close to the references at a larger spatial scale: the canton level (see figure 2 for a representation on a map). We compare the results of the experimental design D for *dispo. threshold* = 1 to those with *dispo. threshold* < 1.

Figure 32 presents the minimum and maximum results for 1999 (on the left) and 2006 (on the right) for every canton of the Cantal. References are represented by symbols while lines figure the bounds of the model results. It shows that *dispo. threshold* = 1 does not allow the model to reach the reference values for 2006 for a lot of cantons. Globally, it seems that the whole population is not sufficient when *dispo. threshold* = 1 meaning the three largest municipalities (Aurillac, Arpajon-sur-Cère and St Flour) are never excluded from the search space for housing of households.



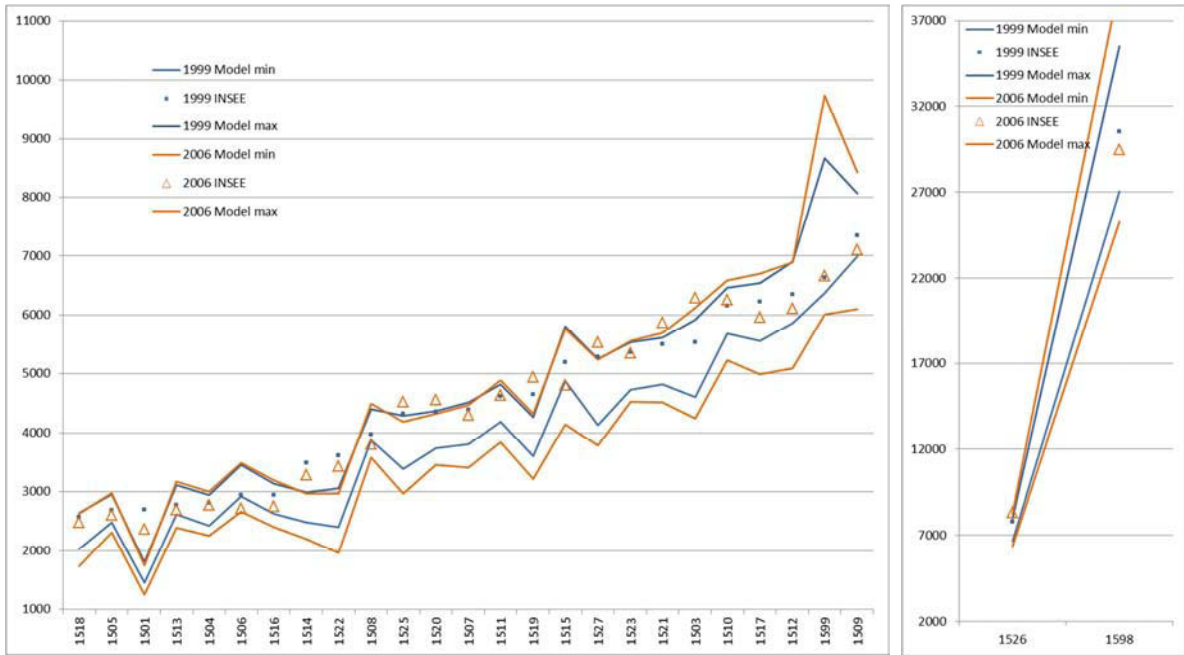


Figure 32. Minimum and maximum number of inhabitants by canton (on abscissa) given by the execution of the experimental design D limited to *dispo. threshold* = 1. The results of the model are represented by lines while dots and triangles represent the reference values (in blue: 1999 values; in orange: 2006 value). The two largest “cantons” are represented on the right allowing the smallest ones to be better represented on the left (due to the scale of the results)

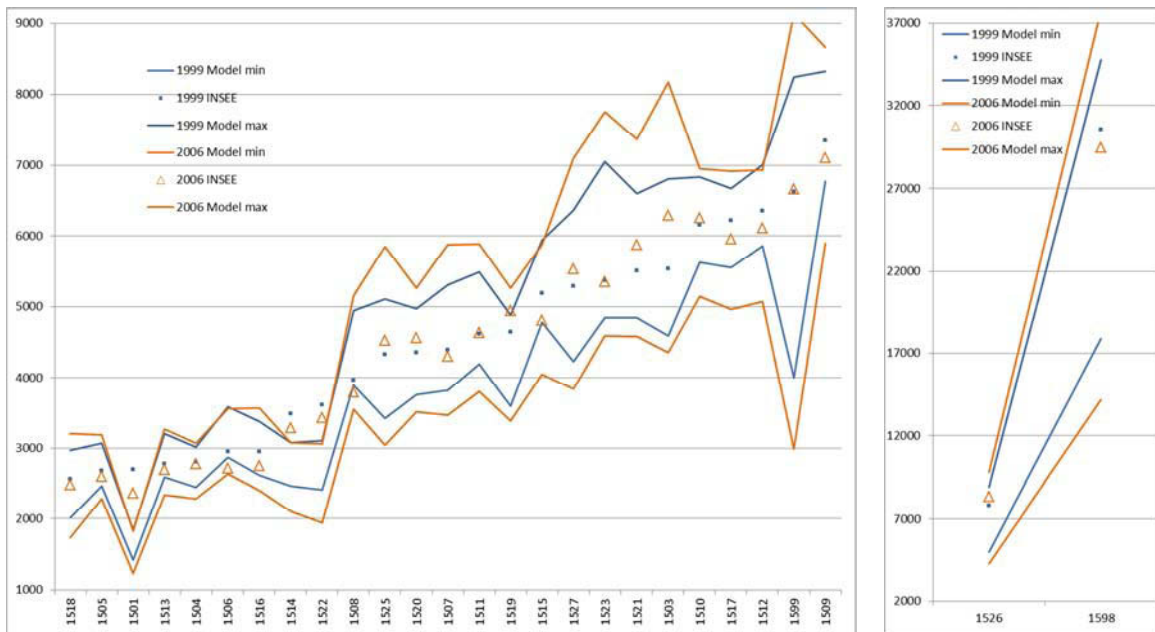


Figure 33. Minimum and maximum number of inhabitants by canton (on abscissa) given by the execution of the experimental design D limited to *dispo. threshold* < 1. The results of the model are represented by lines while dots and triangles represent the reference values (in blue: 1999 values; in orange: 2006 value). The two largest “cantons” are represented on the right allowing the smallest ones to be better represented on the left (due to the scale of the results)

Differently from *dispo. threshold* = 1, most of the references are comprised into the range defined by the minimum and the maximum results of the model when *dispo. threshold* < 1 (execution of the experimental design D) as shown in figure 33. The remaining non reachable reference values concern small cantons (1501 Allanches, 1514 Salers, 1522 Pleaux). Their

size and the fact that two of them are partly on the border of Cantal make them difficult to simulate properly.

To assess some relevant values of *dispo. threshold*, we extract a subset from the experimental design results with already defined “good” parameters: *res satisf*  $p = 0.09$  and *proximity* = 33. Figures 34 show the minimum and maximum results for various *dispo threshold* value (0.2 in blue, 0.4 in green, 0.55 in red) for 1999 (on the left) and 2006 (on the right) for every municipality of the Cantal (except Aurillac which is too large for the graph making the other municipalities readable). References are represented by a black dot. *Dispo threshold* = 0.2 appears more relevant for 1999 and 2006 than 0.4 and larger values.

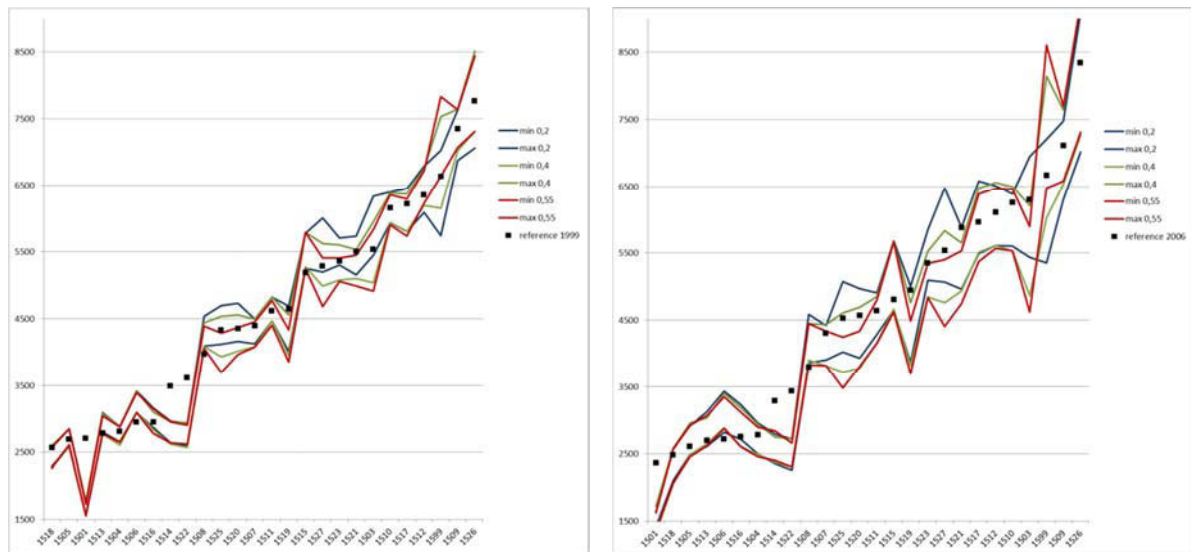


Figure 34. Simulated number of inhabitants for every canton from Cantal (in abscissa) and references for 1999 (on the left), 2006 (on the right) for *dispo threshold* = 0.2 (in blue), *dispo threshold* = 0.4 (in green) and *dispo threshold* = 0.55 (in red); the references are represented by the black dots. Aurillac, the largest canton ?, is not represented on the graph.

Regarding Aurillac, the references are contained in the minimum and maximum interval for *dispo threshold* higher than 0.2 in 1999, and higher than 0 and lower than 1 in 2006. We can conclude that for *proximity* equals 33 km, *dispo threshold* should be comprised between 0.2 and 0.4.

To go further, we check if the model is susceptible to reproduce the reference evolution of every canton. Figures 35 show that is probably almost possible. Only four cantons have references outside the bounds defined by the results of the model for the evolution between 1999 and 2006. However, it seems that the sign of the evolution can be respected. We notice the size of the bounds for the canton 1598, St Flour, corresponding to the second largest municipality of the Cantal.

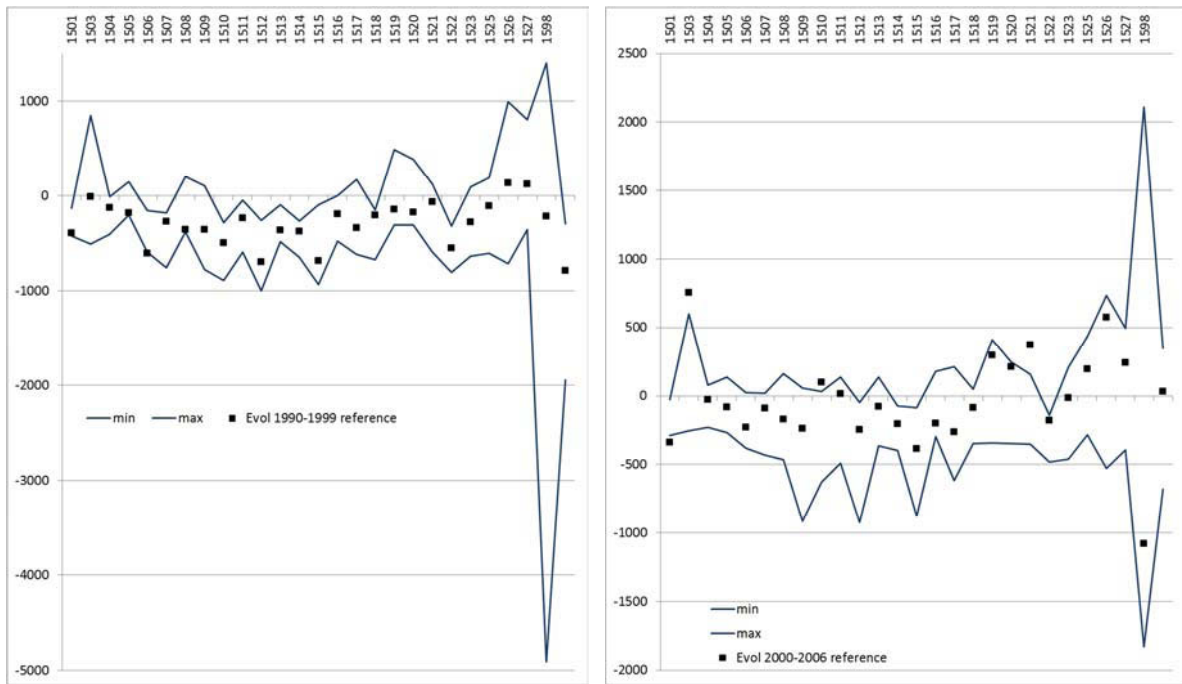


Figure 35 : Reference evolution (black dot) for every canton (in abscissa) on the left, evolution 1990-1999 and on the right, evolution 1999-2006, and minimum and maximum evolution (blue lines) given by the model for the selected range of relevant values of the experimental design B (see the table 2 in the following to read a synthesis of these values)

Coming to our first question: **is a dynamics only considering a research as close as possible is sufficient to obtain an increase of the population only situated around the larger municipalities or is it necessary to identify a specific dynamics for larger cities?** We conclude that a simple research further and further until the individual finds out a partner and/or a residence is not enough to fit the reference. It is necessary to add a mechanism excluding the three largest municipalities from the search space for housing of the households. It should be based on a probability which excludes the three largest (in 1990) municipalities at a frequency around 0.2, 0.4 and anyway smaller than 1.

The second question was: **is it possible to bound the search for a partner and a residence at a maximum distance smaller than the commuting distance?** We saw that the references are possible results from the model only when *proximity* values are from 24 to 36 km. It means that not only it is possible but it is also necessary to limit the search for a partner and residence at a smaller distance than for commuting.

## Conclusion

We start this implementing phase of our conceptual model knowing we have five unknown dynamics:

- Couple creation;
- Couple splitting;
- Birth;
- Move decision;
- Searching a new housing or a partner.

We assume a dynamics for each of them inspired from what we know from Cantal and try to minimise the number of parameters required for each of them. For each of these parameters, we tried identifying bounds of relevant values to obtain a result from simulation close to the reference values given by the National Statistical Office. These parameters have to be calibrated to find out their exact value giving the closest results from the reference.

We now state the finally chosen dynamics and discuss them.

## **Couple creation**

We wondered if the couple creation required a two-step dynamics to model the Cantal: a first step for the individual's decision to search for a partner; a second step for the searching itself when the individual has decided searching. We consider a parameter ruling the annual individual annual probability deciding searching, and a parameter ruling the maximum annual individual number of trials to meet someone convenient. The first parameter corresponds to a frequency searching while the second corresponds to the level of research in searching.

Another way to tell the question we are interested in is "is it necessary considering a frequency of research for partner lower than annual to obtain a number of couples leading to the number of births of the Cantal?". It is possible that potential partners are rare enough in the space to limit the number of couples, limiting itself the number of births. Then, in this case, it does not matter single decides searching for a partner every year since we can limit to 1 or very few the number of trials meeting someone and obtain a proper number of couples. It is exactly the situation where we can suppress the parameter probability to search for a partner which is implicitly valued 1. But that was not the case for the Cantal. To be close to the reference, it is necessary to choose a probability less than one to search for a partner. This is not necessary to obtain the right number of births in the population but it becomes necessary to obtain out-migrants and migratory balances close to the reference. Indeed mechanisms around formation and disruption of couple have an important impact on residential mobility as outlined by the literature of demographers (Debrand and Taffin 2005; Debrand and Taffin 2006).

One shortcoming of our approach relates to the consistency of the chosen dynamics. Indeed, we choose to define the socioprofessional category of a new worker using the socioprofessional category of one randomly chosen of their parents. We assume a kind of homophily in level of education transmission, but we did not include any homophily of partners in couple. Such a homophily does exist, however, we have notice from data analysis that it tends to be less strong than in the past.

## **Couple splitting**

We assume the simplest way to split: an annual constant probability. It is equivalent assuming, even if we know it is false from the literature (Vanderschelden 2006) that the average duration is constant. The purpose was to define if such a simple hypothesis is sufficient to produce natural and migratory balances close to the references. It appears it is. We don't find any indicator at this stage of the study which leads us to reject the hypothesis.

In practices, it probably means the union dissolutions of older people are overestimated. The main impact of the split probability is on outmigration which is essentially the fact of young people. Moreover, we observe on the figure 24 that the model estimates the migratory balance of individuals being at least 30 with difficulty. It tends to underestimate the positivity of the migratory balance for age ranges higher or equal to 30. It is possible that 30 years' old and older split too much and thus, move outside more often than in the reference.

This is a shortcoming of our work. A future version of the model should compare this splitting model to a dissolution model based on the age of the partner (considered here as a proxy for the partner cohort). It remains however to find a way to parameterise it using basic data.

## **Birth**

The fertility increases during the considered periods in Cantal even if it increases less than at the French level. At the same time, the migratory balance was improved for people in age of giving birth. Thus, the question was to know if it is necessary to consider an increasing probability to give birth for each couple to obtain the right number of births (remaining almost constant over the time), or if a constant probability to give birth for each couple can be convenient to model match the references. It finally appears from our study that the model has to consider an increasing individual probability to give birth (chosen as a linear function) in order obtaining enough births to be close to the references given by INSEE.

The model is able to give a correct number of births and natural balance. However, they don't correspond to the same parameter values because the number of deaths is systematically too high compared to the reference (about 2000 "extra" deaths). The death is ruled by a law extracted from French data at the national level. It seems people from Cantal lives longer than other French people on average. However, we don't know exactly how to correct the law or mortality by age given at the French level to obtain the one for Cantal. This is a shortcoming of our work and it remains to solve.

## **Moving**

We propose an integrated model of the mobility caused by distance of commuting, family events and satisfaction of the size of housing and susceptible to become a migration if a household decides to move outside the studied region. We studied possible functions for the decision to move and the searching procedure.

### ***The decision***

We assume that the decision to move depends on:

- the distance of commuting;
- some family events as union dissolution and couple formation;
- the adequacy of the size of the housing expressed in number of rooms to the size of the household.

For the last point, we consider that a household is totally satisfied when the number of rooms of the housing is equal to the number of members of the household. The dissatisfaction increases with the difference between these two quantities (number of members and number of rooms). However, we know from data and literature there is a significant link between the age and the tendency to move: a lot of old people live alone in big houses for example. It means that the satisfaction should also depend on the age of the household members. That is what we demonstrate when considering an exponential satisfaction function taking into account the average age of the adults of the household. Indeed, the model cannot show a result close to the reference if the age is not taken into account. Especially, the model with the age function is able to produce correct migratory balances and numbers of in and out migrants. These indicators directly depend on the decision to move of households.

A question that we have not answered remains. Perhaps a decision function only based on the age would be sufficient? We have not tested this simpler model and it is certainly a shortcoming of our work.

### ***Where searching housing and/or partner***

We assume a model searching for a housing and for a partner (as finding a partner implies changing residence for at least one of the partner) starting from the best location and enlarging the search progressively if the object of search (house or partner) is not found. For an occupied active individual, the best location of housing is the place of work. For the other individuals, the best location is the current place of residence. We had to deal with two issues. Firstly, we had to define if the individual searches until she finds or if there is a maximum distance where she stops searching. Secondly, we had to decide if a specific mechanism for avoiding the three largest municipalities of Cantal as residence locations is necessary.

Our study shows that the model needs such a mechanism because otherwise, whatever the maximum distance for searching is, the model is not able to yield a result close to the reference for the population of every canton (close to a county) of the Cantal. In fact, this exclusion of searching area concerns the individual wherever they work. However it impacts many people working in the three largest municipalities since they are employment centres.

The necessity for a bounded search at a maximum distance is evaluated through the capacity of the model to give values describing the commuting distance and the residential mobility distance distributions close to the references. The study shows that the maximum Euclidian distance for searching housing or partner has to be from 24 to 36 kilometres. For other values of the parameter *proximity* corresponding to this maximum distance, the model is not able to yield results close to references.

To conclude, it appears both mechanisms, one pushing people to live outside the main municipalities, the other bounding the area where people search for new housing or partner, are necessary to reproduce the spatial evolution of the population. It is probable, even if the parameter can change from a case study to another that these two dynamics should be always present to model the spatial evolution.



However, the identification and the number of municipalities to exclude probabilistically from the searching area is not so easy to determine. For example, if we exclude also the fourth largest municipality, Mauriac, the main indicators we use don't change significantly. Hence, we decide to restrict the exclusion to the three largest municipalities. But choosing a same probability of exclusion over the municipalities and time is a strong assumption which deserves to be more studied and criticised. This is certainly a shortcoming of our work.

The research heuristic coupled to the maximum distance called *proximity* appears quite convenient and very simple. However, we can wonder what exactly means this *proximity*. Is it a geographical attribute expressing the easiness to move inside a given area (limited by type of roads, climatic conditions ...), an individual attribute which depends on the household type or of preference for a given mean of transport (bike, train, car...)? It would be relevant to better understand the impact of this parameter and how do the indicators sensitive to it evolve over time.

A last limitation we want to point out can be read on the graph showing the results of the model at the canton level. It appears that the canton is a good spatial level to evaluate if the model is able to reproduce the spatial heterogeneity of the evolution of the population. Indeed, the municipality is a too small element in terms of number of inhabitants to be correctly modelled (on average 440 inhabitants without the two largest). The canton composed on average of 4500 inhabitants (without the two largest) is more appropriated. However, we have observed that three cantons cannot be correctly modelled. Moreover, the model has some difficulties to reproduce the positive evolution of some cantons from 1999 to 2006. It seems that the impact of the parameters *proximity* and *dispo threshold* should be studied more deeply for explaining these difficulties.

Before closing this chapter we synthetize our main findings about parameters and indicators and about the relevant values of the parameters.

## Indicators and parameters

The ANOVA establishes clear links between some indicators and the parameters. This is synthesised in table 1. This synthesis constitutes a knowledge base for choosing indicators for the calibration. The references of these indicators will be the values the model should fit as close as possible. The parameter values giving the closest values for these indicators will be considered as the most relevant.

From what we studied, we expect the model is susceptible giving results close to the reference of the following indicators:

- Births
- Natural balance
- In-migrants
- Out-migrants
- Migratory balance
- Migratory balance by age range

- Probability to commute at a distance smaller than 3
- Average commuting distance (considering only distances larger than 3)
- Average moving distance (considering only distances larger than 3)
- Population of every “cantons”
- Inhabitants by age range (except more than 60)

The list of indicators the model can't be very close is the following:

- Deaths
- Inhabitants being more than 60
- Total number of movers
- Probability to move shorter than 3

Table 1. Parameters and main indicators sensitive to these parameters

<i>Code param</i>	<i>Name</i>	<i>Indicators</i>
<i>nb child slope</i>	<i>slope of the giving birth function</i>	<i>Births</i>
<i>nb child intercept</i>	<i>intercept of the giving birth function</i>	<i>Births</i>
<i>P search partner</i>	<i>probability to search for a partner</i>	<i>Births, migratory balance</i>
<i>split p</i>	<i>probability to split</i>	<i>Deaths, births, in-migrants, out-migrants, migratory balance, average moving distance (except moving atmost 3)</i>
<i>join trials</i>	<i>number of join trials</i>	<i>Births, migratory balance</i>
<i>Proximity</i>	<i>proximity threshold</i>	<i>Probability commuting atmost a distance 3, average commuting distance (except commuters atmost 3), average moving distance (except moving atmost 3)</i>
<i>dispo threshold</i>	<i>house availability threshold</i>	<i>Migratory balance, probability commuting atmost a distance 3, probability moving atmost a distance 3, average moving distance (except moving atmost 3)</i>
<i>res satis p</i>	<i>residence satisfaction</i>	<i>Deaths, in-migrants, out-migrants, migratory balance, probability commuting atmost a distance 3, probability moving atmost a distance 3, average moving distance (except moving atmost 3)</i>

The lack of people being more than 60 is explained by the excess of deaths. It is difficult to correct this error because we do not have the death distribution of probabilities by age for Cantal.

The mobility of people moving very close is underestimated by the model. Obtaining the right number of movers appear too difficult for our approach (i.e.: a stochastic individual-based model at a high scale) considering the number of processes impacting them and the few knowledge we have about them and their interdependency. For example, we know that the correspondence between the size of the housing and the size of the household is not a reliable approximation. It appears from the data analysis in France that there are very few housings of size 1 compared to the number of households of size 1. The average number of individuals by room varies from 0.5 to 0.7 in 1990 and 1999. The number of housings of size 1 is 2304 in the Cantal *département* in 1999 while the number of households comprising only



one member is 19142. Identically, the housings of size 2 correspond to 9.4 % of all the housings while the households comprising two members represent 33.02 % of the households in 1999. Other aspects impact the move as for example how parents “share” children between them when they quit each other, what are their needs in terms of size of housing? Overall, if we parameterise the model in order to get more movers, the population decreases too much because the emigration becomes too high. Then, the out-migration law should perhaps be studied further and changed, particularly the one attributing a probability to out-migrate depending linearly on the distance to the border of Cantal. We know that this dependence it is not the same all along the border: some directions are privileged to get out from the region.

Some contradictions appear, requiring making choices for the calibration. For example, is it better to fit the births or the natural balance knowing the number of deaths can't be modelled correctly? There are still about 2000 deaths in excess.

### Segments of relevant values

Some relevant segments of values have been identified for the parameters. They are presented in the following table. They constitute a base for the calibration process which has to define more precisely the parameter values giving results as close as possible to chosen references. Our study does not allow us to conclude on the couple splitting probability but the value segment to test can be limited from the semantic of the parameter, the duration of a couple. For example, 0.1 means that the couple lasts 10 years on average while 0.02 means that it lasts 50 years on average. Moreover, the literature exhibits that the probability to split depending on the cohort varies from about 0.004 to 0.04 (Vanderschelden 2006).

Table 2. Synthesis about the segments of relevant values to parameterise the model in order trying to obtain a result close to reference for the indicators which can be fitted.

<i>Code param</i>	<i>Name</i>	<i>Segment of relevant values</i>
<i>nb child slope</i>	<i>slope of the giving birth function</i>	<i>&gt; 0</i>
<i>nb child intercept</i>	<i>intercept of the giving birth function</i>	<i>]1.2;2.0[ with expected good values around 1.6, 1.8</i>
<i>p search partner</i>	<i>probability to search for a partner</i>	<i>]0.1;1[</i>
<i>split p</i>	<i>probability to split</i>	<i>]0.01;0.04]</i>
<i>join trials</i>	<i>number of join trials</i>	<i>1; 5</i>
<i>Proximity</i>	<i>proximity threshold</i>	<i>[24;36]</i>
<i>dispo threshold</i>	<i>house availability threshold</i>	<i>[0.2;0.4]</i>
<i>res satis p</i>	<i>residence satisfaction</i>	<i>]0.03; 0.14[ with probably good values around 0.09 and 0.11</i>

### Acknowledgements

This work has been funded by the PRIMA (Prototypical policy impacts on multifunctional activities in rural municipalities) collaborative project, EU 7th Framework Programme (ENV 2007-1), contract no. 212345.

We thank a lot the following institutions who made available data and tools for maps:

- Maurice Halbwachs (<http://www.cmh.ens.fr/>)
- IRSTEA, unité de recherche Développement des Territoires Montagnards, Grenoble (<http://siddt.grenoble.cemagref.fr/>)



# Part 2. Theoretical individual based models of social influence

*When men wish to construct or support a theory, how they torture facts into their service!*<sup>17</sup>

### Abstract

With a more theoretical approach, we also study, especially through a double-modeling approach comparing aggregated and individual-based models, the impact of some innovative individual attitude dynamics. We add various mechanisms to resist the social influence to the classical attraction for congruent attitude. A first studied mechanism leads individuals to neglect some pieces of information if they are not enough important regarding their incongruence. These individuals exhibit the primacy bias since their attitudes is defined by the first accepted messages. We show that this bias reducing the rationality of the individual can be increase or decrease when individuals directly exchange messages compared to when they are only informed by a media. The second mechanism is a repulsion one we add to the classical positive influence occurring where two interlocutors are close enough in attitude. Actually, a negative influence occurs due to the discomfort of a meeting where individuals are close on one attitude and far on another attitude. The main impact of this repulsion mechanism is certainly the less number of clusters of individuals having the same attitudes emerging from the interactions, compared to the number obtained only with the attraction mechanism.

### Contents

- Chapter 2.1 Disregarding information – a model exhibiting the primacy bias..... 133
- Chapter 2.2 Disregarding information – a double modelling approach ..... 153
- Chapter 2.3 Attraction-Rejection – designed from theories..... 175
- Chapter 2.4 Attraction-rejection – a double modelling approach..... 197
- Chapter 2.5 Attraction-Rejection – designed from experiments..... 209

---

<sup>17</sup> John Mackay, *Extraordinary Popular Delusions and the Madness of Crowds*, 1852



# Chapter 2.1 Disregarding information – a model exhibiting the primacy bias

Title : Propagation effects of filtering incongruent information

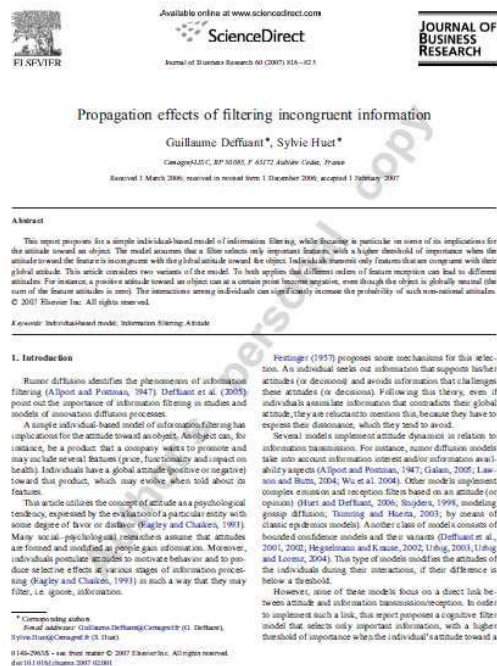
Authors: Guillaume Deffuant and Sylvie Huet  
Published in Journal of Business Research, 2007

## Abstract:

This chapter presents a proposal for a simple individual-based model of information filtering, while focusing in particular on some of its implications for the attitude toward an object. The model assumes that a filter selects only important features, with a higher threshold of importance when the attitude toward the feature is incongruent with the global attitude toward the object. Individuals transmit only features that are congruent with their global attitude. This paper considers two variants of the model. To both applies that different orders of feature reception can lead to different attitudes. For instance, a positive attitude toward an object can at a certain point become negative; even though the object is globally neutral (the sum of the feature attitudes is zero). The interactions among individuals can significantly increase the probability of such “non-rational” attitudes.

## Contents

The models.....	136
Constant incongruence threshold (CIT) .....	136
Dynamic incongruence threshold (DIT).....	137
Diffusion of a two-feature object.....	137
A single individual.....	137
Interconnected individuals.....	140
Diffusion of more than two features .....	141
One major negative and p minor positive features .....	141
Simulating interacting individuals.....	143
Discussion.....	148





Rumor diffusion identifies the phenomenon of information filtering (Allport & Postman, 1947). Amblard, Deffuant & Huet, (2005) point out the importance of information filtering in studies and models of innovation diffusion processes.

A simple individual-based model of information filtering has implications for the attitude toward an object. An object can, for instance, be a product that a company wants to promote and may include several features (price, functionality and impact on health). Individuals have a global attitude (positive or negative) toward this product, which may evolve when told about its features.

This paper utilizes the concept of attitude as a psychological tendency, expressed by the evaluation of a particular entity with some degree of favor or disfavor (Chaiken & Eagley, 1998). Many social-psychological researchers assume that attitudes are formed and modified as people gain information. Moreover, individuals postulate attitudes to motivate behavior and to produce selective effects at various stages of information processing (Chaiken & Eagley, 1998) in such a way that information may be filtered, i.e. ignored. Festinger (1957) proposes some mechanisms for this selection. An individual seeks out information that supports his/her attitudes (or decisions) and avoids information that challenges these attitudes (or decisions). Following this theory, even if individuals assimilate information that contradicts their global attitude, they are reluctant to mention this, because they have to express their dissonance, which they tend to avoid.

Several models implement attitude dynamics in relation to information transmission. For instance, rumor diffusion models take into account information interest and/or information availability aspects (Allport & Postman, 1947; Butts & Lawson, 2004; Galam, 2005; Adamic, Huberman, Tyler & Wu, 2004). Other models implement complex emission and reception filters based on an attitude (or opinion) (Deffuant & Huet, 2006; Snijders, 1998 modeling gossip diffusion; Huerta & Tsimring, 2003; by means of classic epidemics models). Another class of models consists of bounded confidence models and their variants (Deffuant, Neau, Amblard, & Weisbuch, 2001; Deffuant, Amblard, Weisbuch & Faure, 2002; Hegselmann & Krause, 2002; Urbig, 2003). This type of models modifies the attitudes of the individuals during their interactions, if their difference is below a threshold.

However, none of these models focuses on a direct link between attitude and information transmission / reception. In order to implement such a link, this report proposes a cognitive filter model that selects only important information, with a higher threshold of importance when the individual's attitude toward a feature contradicts his global attitude toward the object. In other words, the model filters out a feature that is incongruent with the general attitude, except if it is very important. Individuals only transmit information that is congruent with their global attitude.

This study uses an individual-based approach, defining the dynamics of individuals and their interactions and simulating the evolution of the system under different initial conditions and parameter values. Beginning with very simple models so as to gain an insight into the global dynamics, the models become progressively more realistic, following a general approach advocated by Deffuant, Weisbuch, Faure & Amblard (2003). The two models differ with



regard to the threshold of incongruent features. In one model, this threshold is constant, and in the other the threshold is the sum of feature attitudes currently retained.

The aim of this paper is to identify the conditions under which these simple hypotheses on information filtering lead to “non rational” attitude changes, in particular, cases where different orders of reception lead to different global attitudes. In some cases, an initial positive population attitude can become negative when the sum of all the feature attitudes is zero. This report initially describes these models and then studies individual behavior in the case of two-feature objects, comparing the results to simulations of interacting individuals. The study then extends to objects with more features, supposing the existence of one major feature, and several minor ones.

## The models

This study deals with two models for implementing the information filter, one with a constant threshold and one with a dynamic threshold.

### Constant incongruence threshold (CIT)

The assumptions of this model are that in a population of  $N$  individuals who discuss an object, consisting of a set of features  $F = (1, 2, \dots, d)$ ,

all the individuals share:

- The *initial* global attitude  $g$ ;
- The attitudes toward the object features  $(u_1, \dots, u_j, \dots, u_d)$  with  $u_j \in \mathfrak{R}$ .

an individual  $i$  is characterized by:

- $L_i$ : a subset of  $F$  containing the features currently retained by the individual; this list is supposed to be empty at the beginning.
- $G_i = g + \sum_{j \in L_i} u_j$ : the global attitude toward the object. This choice can be related to the information integration theory of Anderson (1971),
- A neighborhood corresponding to the subset of individuals with whom  $i$  can communicate.

When  $G_i \cdot u_j \geq 0$ , the feature  $j$  is said to be congruent for individual  $i$ , and incongruent otherwise.

Two positive numbers determine the dynamics of filtering,  $\theta$ , the congruence threshold and  $\Theta$ , the incongruence threshold. We assume  $\Theta > \theta$ . Being told about feature  $j$ , an individual will react as follows:

- If  $j$  is congruent:
  - If  $|u_j| > \theta \rightarrow$  “retain the feature”.
  - Otherwise  $\rightarrow$  “ignore the feature”.
- If  $j$  is incongruent:
  - If  $|u_j| > \Theta \rightarrow$  “retain the feature”
  - Otherwise “ignore the feature”

Here, “retain the feature” means that  $j$  is added to  $L_i$  (if  $L_i$  does not yet include  $j$ ), “ignore the feature” means that the feature is filtered (not added to  $L_i$ ).

A time step involves the following computations:

- $N$  times repeat: Choose individual  $i$  at random, with probability  $f$ ,  $i$  is being told about a feature chosen at random in  $F$ ,
- $N$  times repeat: Choose a pair  $(i, j)$  of individuals at random, individual  $i$  tells individual  $j$  about a feature chosen at random in the *congruent* features of  $L_i$ , and  $j$  does the same.

### Dynamic incongruence threshold (DIT)

The variant of the previous model, assuming that the incongruence threshold varies with the global attitude, proposes that:  $\Theta = \theta + |G_i|$

Following this variant, retaining incongruent information then implies  $|u_j| > |G_i|$ . Therefore, in this model the global attitude sign changes each time the individual receives incongruent information. This is not the case with the constant incongruence threshold, which allows individuals to retain incongruent information without changing the global attitude sign. This is a major difference between the two models.

### Diffusion of a two-feature object

With respect to the diffusion of an object involving two features with attitudes  $(u_1, u_2)$ , this report investigates the possible successive states of a single individual when the feature attitudes vary, while testing the models for particular values.

#### A single individual

In the case of a two-feature object, it is possible to represent all the states of the model when  $u_1$  and  $u_2$  vary, supposing that an individual is first told about feature 1 and then about feature 2, and that the initial global attitude  $g$  is positive. The graphs take  $u_1$  as the horizontal axis and  $u_2$  as the vertical axis.

In Figures 1, 2 and 3, the areas of  $(u_1, u_2)$ , corresponding to a given sequence of states of the individual are delimited by dotted lines. In white areas the individual’s global attitude sign

does not change. In the light grey areas the global attitude sign changes once (in the end, the sign is negative). In the dark grey areas, the global attitude changes twice (in the end, the global attitude is positive).

Moreover, in each area, a set of characters of the following format appear:

{C1, C2, S1}

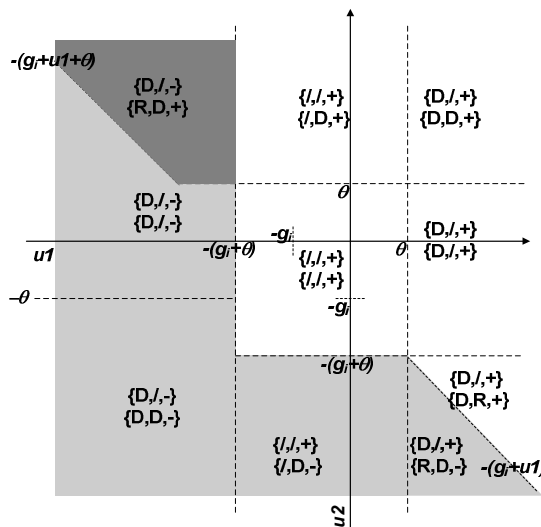
{C3, C4, S2}

In the first line, the characters indicate the following:

- Character C1 indicates whether the individual retains feature1 when he/she is told about it. The meaning of the characters is the following:
- “R” means the feature is retained but incongruent.
- “D” means the feature is retained and congruent, and therefore the individual diffuses the feature.
- “/” means the feature is not in the individual’s list  $L_i$ .
- Character S1 is the sign of the global attitude just after the individual is told about feature 1 (not yet about 2).

The second line of characters ({C3, C4, S2}) follows the same principle, but after the individual has been told about features 1 and 2.

Figure 1: Dynamic incongruence threshold,  $g > 0$ .



Figures 2 and 3 illustrate the differences between the cases  $g < \Theta$  and  $g > \Theta$  by means of the constant incongruence threshold model.

Figure 2: Constant incongruence threshold,  $g < \Theta$

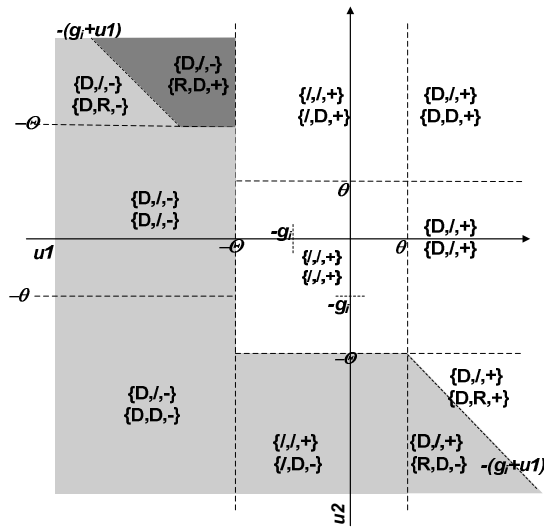
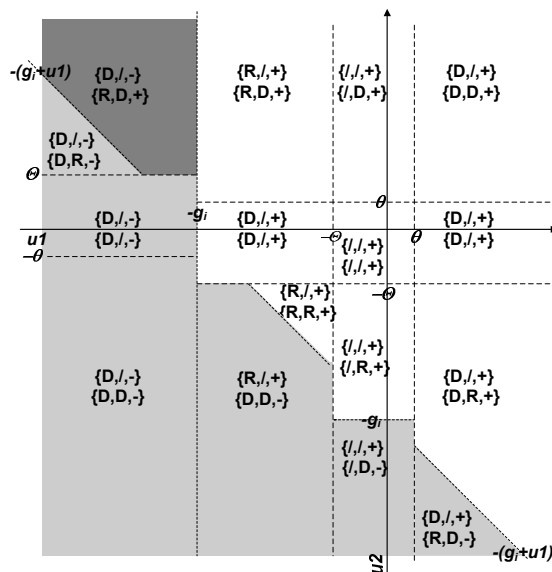


Figure 3: Constant incongruence threshold,  $g > \Theta$

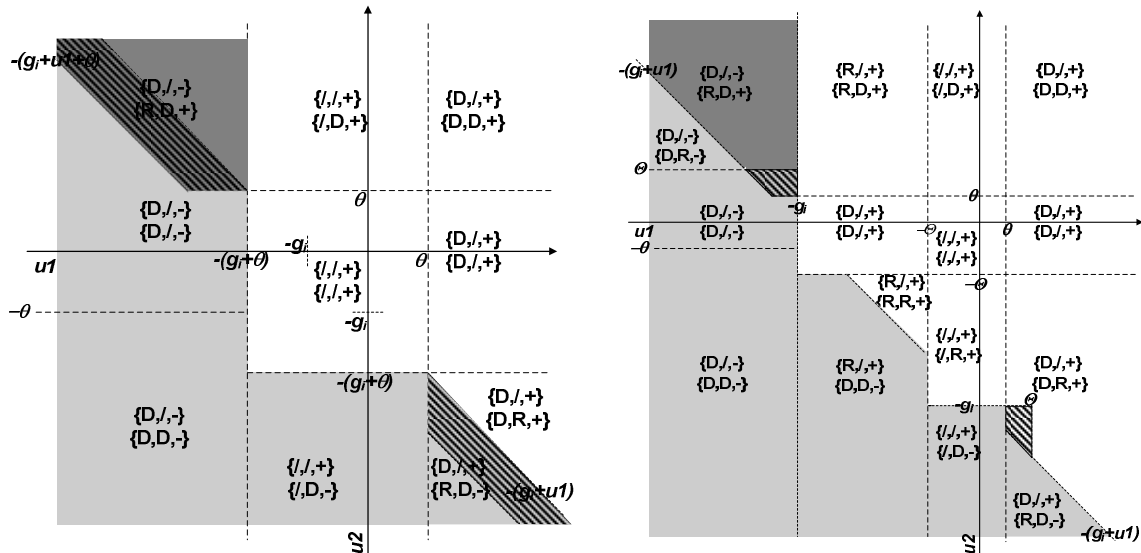


The dynamic incongruence threshold model shows 10 different areas, while the constant incongruence threshold shows one more when  $g < \Theta$  ( $\{D,R,-\}$  in the top left corner of figure 2) and three more when  $g > \Theta$  (as in Figure 2,  $\{R,R,+ \}$  and  $\{I,R,+ \}$  in the middle low and left corner). Therefore, the constant incongruence threshold model is richer as regards the number of different state sequences than the other model.

When  $g > \Theta$ , figure 1 and 3 show significant differences with respect to the boundary between positive and negative final attitude. This is in line with the different behaviours that the models considered display.

In addition, the grey area is not symmetric with the diagonal (hatched zones in Figure 4). This means that the change of the global attitude sign is sensitive to the order in which the features are presented to the individual.

Figure 4. In the hatched zones, the model change of the attitude sign depends on the order of the features.  
Left: Dynamic IT. Right: Constant IT with  $g > \Theta$ .



For example, if the  $(u_1, u_2)$  couple is in the right-hatched area:

- For the DIT model, the final state is then  $\{R,D,-\}$  if the order of reception is (1,2), and  $\{R,D,+ \}$  if the order is the opposite;
- For the CIT model, the final state is  $\{D,R,+ \}$  if the order of reception is (1,2), and  $\{D,I,- \}$  if the order is the opposite.

In the CIT model, the individual retains both features in one order of reception, and in the opposite order, he/she retains only one. Following the DIT model, the individual retains both features in both orders.

This behavior is typically “non rational”. Depending on the order of presentation of its features, the global attitude toward the object can be either positive or negative, and the features retained will be different. In these dissymmetric areas of the parameters, a single individual receiving the features at random has on average a 50% chance to have a positive global attitude, and a 50% chance to have a negative one.

Focusing on these parameter areas, the next step is to test this model with interconnected individuals.

### Interconnected individuals

The authors have tested the model with a population of 1000 individuals who have an initial positive global attitude (which is the same for everyone) and who are fully connected, i.e. each individual can discuss with any other (it will of course be worth to consider different types of networks in the future). The model includes two frequencies on which the media

communicates about a randomly chosen feature:  $f = 0.1$  (on average 100 individuals are told about a feature in each time step) and  $f = 0.0001$  (on average 1 individual is told about a feature every 10 time steps).

The model suggests that when the frequency is high, the results will be close to those of a single individual connected to the media, because the effect of the interactions will be small compared to the effect the media have. In this instance the prediction is that about half of the individuals will develop a negative global attitude.

Conversely, when the frequency is low, the order in which the first individuals receive the features is crucial, because they have more time to influence the rest of the population that may follow their position. Therefore, the distribution of the final states will be less concentrated, with a significant probability of getting extreme results (a final large majority of negative global attitudes or a large majority of positive global attitudes).

Figure 5. CIT model. Distribution of the final proportion of negative global attitudes. 5000 steps.

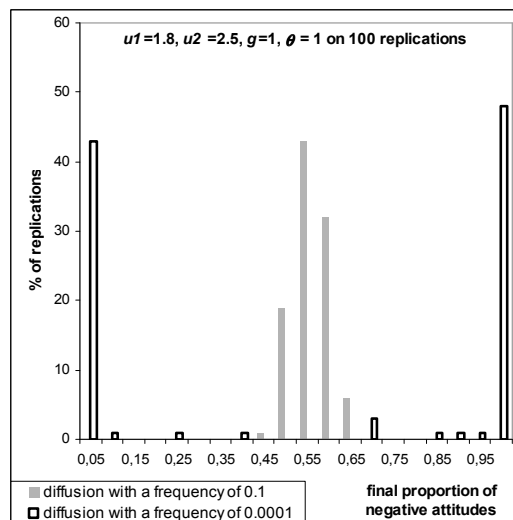


Figure 5 confirms these expectations. If frequency  $f = 0.1$ , the population generally has about 50% of negative final attitudes. If frequency  $f = 0.0001$ , we have about a 50% chance to get 100% of negative final attitudes, and a 50% chance to get 100% of positive attitudes. The diffusion process can therefore have a dramatic effect.

## Diffusion of more than two features

### One major negative and $p$ minor positive features

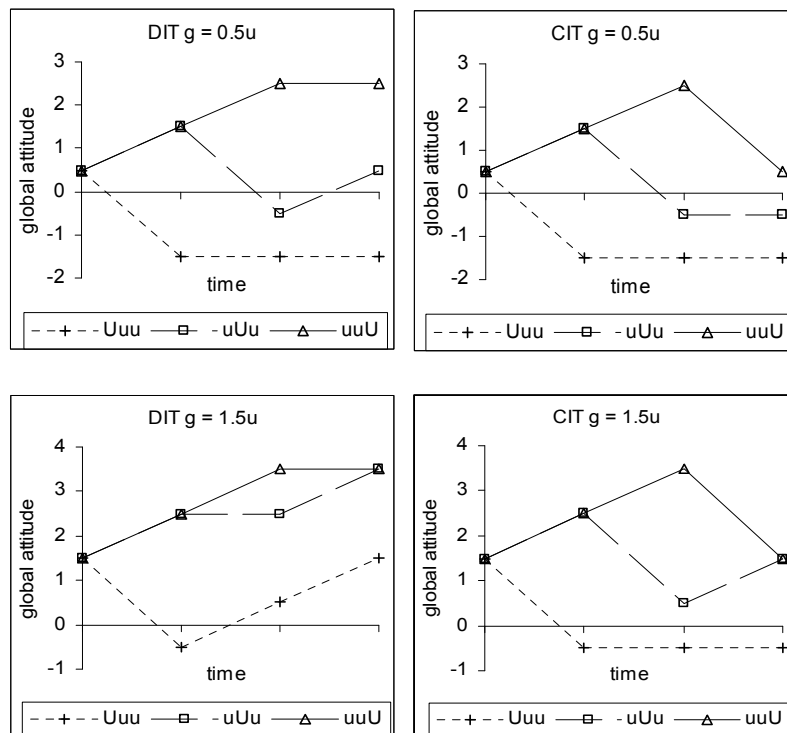
When an object consists of one major negative feature  $U$  and  $p$  minor positive features  $u$ , the major feature is such that  $U > \Theta$ , and the minor features verify  $\theta < u < \Theta$

Assuming that the global message, when taking all features into account, is neutral:  $U = pu$ . In this case, the global attitude of the individual is sensitive to the order in which the features are being communicated (as Figure 6 shows), that is, for two minor features and one major feature. This figure shows the possible trajectories of the global attitude in the

different orders of feature reception. The horizontal axis represents the time of reception of each feature.

One can derive from these individual trajectories the likely proportion of global negative attitudes in a population where individuals only communicate with the media, because each trajectory is equally likely to take place. Therefore, in the case of the above graphs of Figure 6 ( $g = 0.5u$ ), for the CIT model, two thirds of the population have a final negative attitude, and for the DIT model one third has a negative final attitude (with the DIT model two global attitude sign changes occur in trajectory  $uUu$ ). In the case of the below graphs of Figure 6 ( $g = 1.5u$ ), for the DIT model all trajectories lead to a final positive attitude, and for the CIT model we expect that one third of the population will have a negative final attitude. Obtaining such proportions of final negative attitudes is surprising, considering that all individuals have an initial positive global attitude, and that the message is globally neutral. Note that the same type of results can occur even with a globally positive message.

Figure 6. DIT (left) and CIT (right) models - Individual trajectories of global attitude for  $p=2$ ,  $\theta=0.3u$ ,  $\Theta=1.3u$ , for different features of delivery orders and different values of initial  $g$  (above:  $g = 0.5u$ , below  $g = 1.5u$ )



Computing the individual trajectories for 10 minor features and with different values of  $g$  shows similar outcomes. If  $g = 2.5u$ , the results are 8 trajectories over 11, leading to a final negative attitude for the CIT model, and 7 over 11 for the DIT. If  $g = 6.5u$ , there are 4 and 3 trajectories respectively over 11 leading to a final negative attitude for both the CIT and DIT model. One can directly determine the probabilities in the case of non-connected individuals.

The following sections investigate the effect of interactions with different frequencies of media communication.

## Simulating interacting individuals

Using the same protocol as in section 3.2:  $N = 1000$ , total connection,  $f = 0.1$  or  $f = 0.0001$ , the authors compared the results with those that would be obtained if the individuals were only connected with the media (receiving the features in a random order). In a first exploration of the model, the authors considered two cases,  $p = 2$  and  $p = 10$ .

### Two minor features

Using the parameter values of section 4.1:  $\theta = 0.3u$  and  $\Theta = 1.3u$ , the researchers tested two values for  $g$ :  $g = 0.5u$  and  $g = 1.5u$ . For isolated individuals, the expected results were 66% ( $g = 0.5u$ ) and 33% ( $g = 1.5u$ ) of negative final attitudes in the case of the CIT model, and 33% ( $g = 0.5u$ ) and 0% ( $g = 1.5u$ ) for the DIT model.

Figure 7: CIT  $g = 0.5u$ .

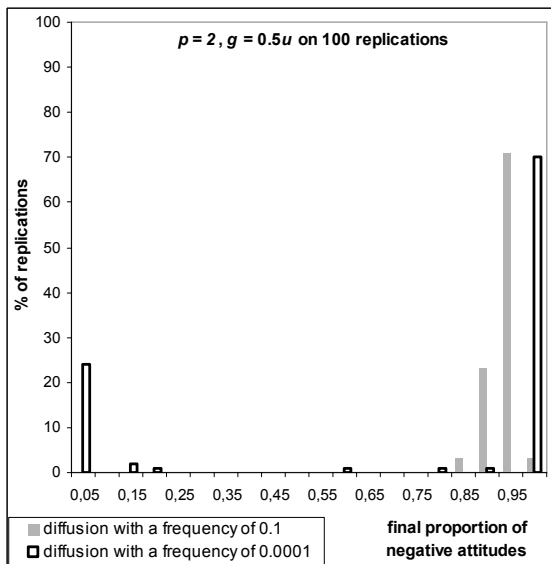


Figure 8: CIT  $g = 1.5u$ .

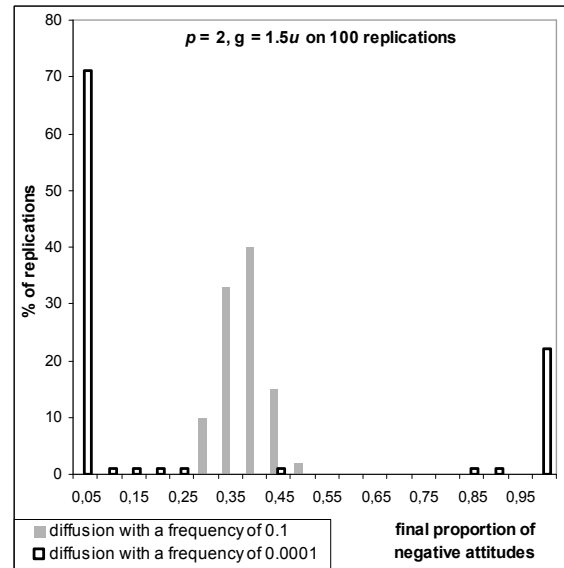


Figure 9: DIT  $g = 0.5u$ .

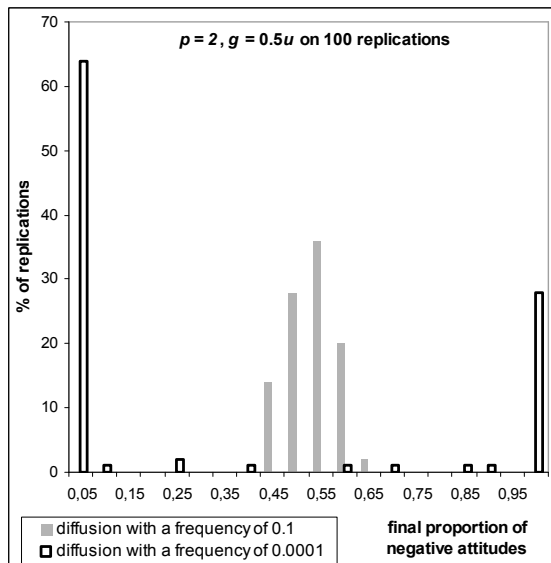
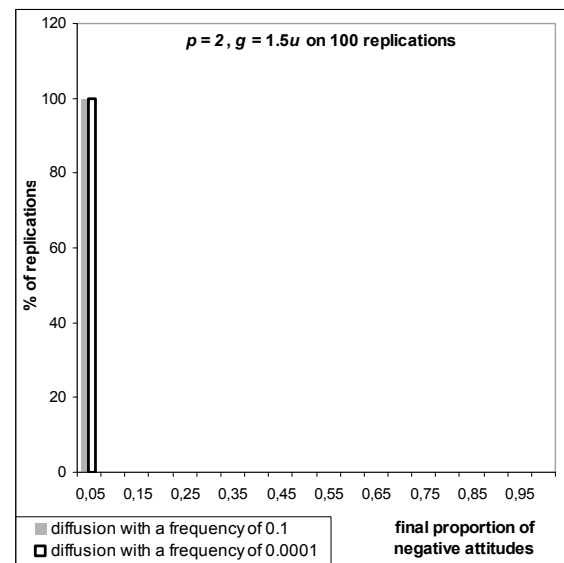


Figure 10: DIT  $g = 1.5u$ .





Figures 7, 8, 9 and 10 show the distribution of the final proportion of negative attitudes in the population after 5000 steps, for 100 replications.

When  $f = 0.1$ , for the CIT the final proportion of a negative attitude is significantly higher than expected without interactions with  $g = 0.5u$  (figure 7, 91% negative on average compared to 66% expected without interactions) and in the case of the DIT with  $g = 0.5u$  (figure 9, 50% negative on average compared to 33% expected without interactions).

One can explain this discrepancy by differences in the transmission of the various features: in the cases considered, the major feature tends to be transmitted more often than the others. Considering the CIT model, if  $g = 0.5u$ , all individuals who received the major feature first or second only transmitted this feature. Therefore, other individuals were more likely to receive the major feature first or the second afterwards. This led to a rapid increase in the transmission of the major feature  $U$  shown on the left-hand graph of Figure 12. The small peak in the transmission of the minor features at step 2 corresponds with the individuals who received a minor feature first, followed by the major feature. They first transmitted a minor feature, and only then a major one.

This explanation does not hold when the major feature is retained when received first, because in that case the probability of transmission is no longer biased, as the right graph of Figure 12 shows. The features are transmitted with an equal probability, leading to the expected probability of negative attitudes in the isolated case (33%) as shown in Figure 8.

The DIT model shows a significant discrepancy, even in the case where the major feature is retained when received first (Figure 9,  $g = 0.5u$ ). The difference with the CIT model lies in the double change of the attitude sign of the  $uUu$  trajectory on the above left-hand graph of Figure 6. The individuals who received the major feature in the second position, only talked about this feature after they had received the last minor feature. This explains the peak in the major feature transmission on the left-hand graph of Figure 13. This significantly biases the number of individuals receiving the major feature first.

When  $f = 0.0001$ , the picture is very different: the general outcome is then that all individuals have an attitude of the same sign. The probability that the sign will reach the population is very similar to the probability of the individual trajectories without interactions.

The explanation of this result is the same as the one given in the two-feature case: on average one individual receives a feature every ten-iteration, which is enough to propagate information from one individual to the whole population. In each simulation, the whole population has therefore the same trajectory, corresponding to that of an individual only connected to the media. Figure 11 illustrates this by showing the evolution of the feature transmission in particular simulations. The left-hand graph shows how the first individual receives the major feature and communicates the feature to the rest of the population in a few steps. In the central graph, the first individual receives a minor feature, and communicates this to the rest of the population. Then, an individual receives the major

feature and communicates it to the rest of the population. In the right-hand graph, the individual receives the minor features first.

The results of Figure 10 are as expected, because in this case, all trajectories lead to a positive final attitude.

Figure 11. CIT model,  $f = 0.0001$ . Examples of evolution of the proportion of each feature as subject of discussion,  $p = 2$ ,  $g = 0.5u$

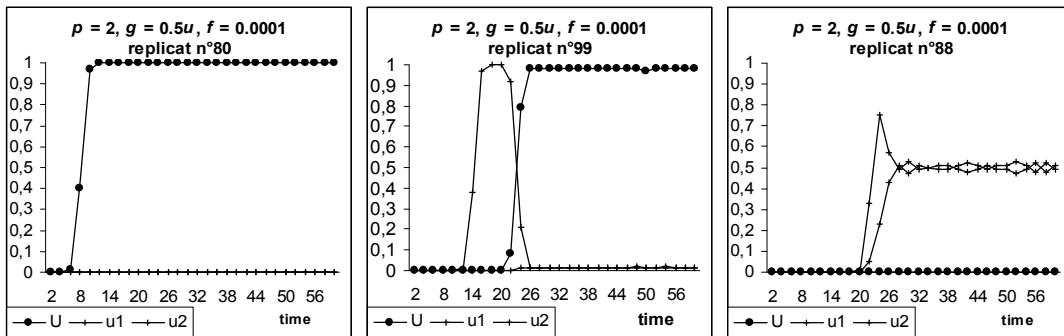


Figure 12. CIT model. Mean proportion of each feature as subject of discussion in case of  $p = 2$  and  $f = 0.1$  for (left  $g = 0.5u$ , on right  $g = 1.5u$ )

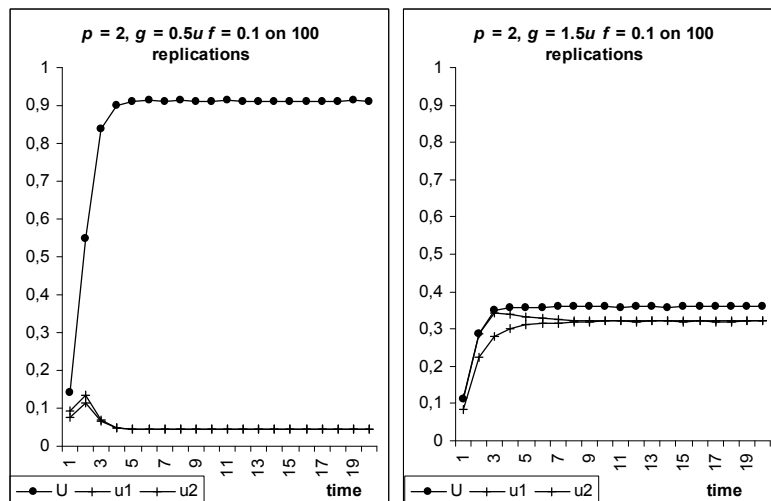
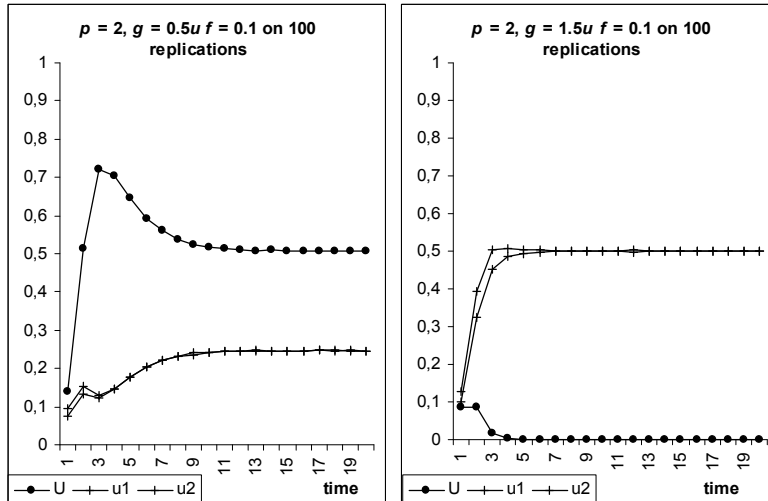


Figure 13. DIT model. Mean proportion of each feature as subject of discussion in case of  $p = 2$  and  $f = 0.1$ . (on left  $g = 0.5u$ , on right  $g = 1.5u$ )



### Ten minor features

Retaining the values  $\theta = 0.3u$ ,  $\Theta = 1.3u$ , the authors tested two values of  $g$ ,  $g = 2.5u$  and  $g = 6.5u$ . Without connection, the expected proportions of negative final attitudes are respectively 8/11 and 4/11 for the CIT model, 7/11 and 3/11 respectively for the DIT model.

Figures 14, 15, 16 and 17 show an even more dramatic effect when a connection with more negative final attitudes occurs.

Figure 14. CIT.  $g = 2.5u$

Figure 15. CIT.  $g = 6.5u$

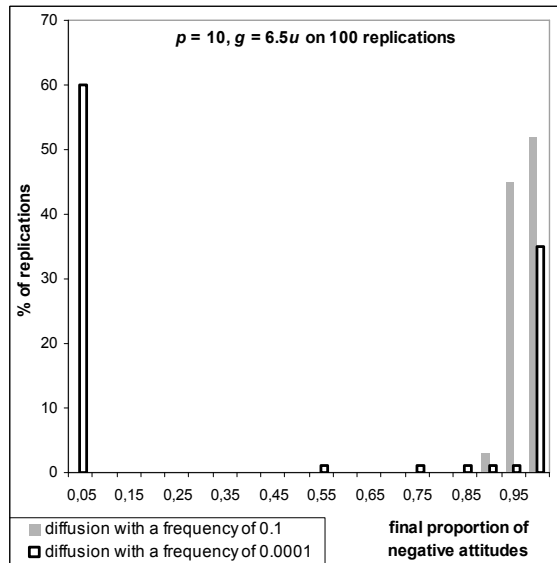
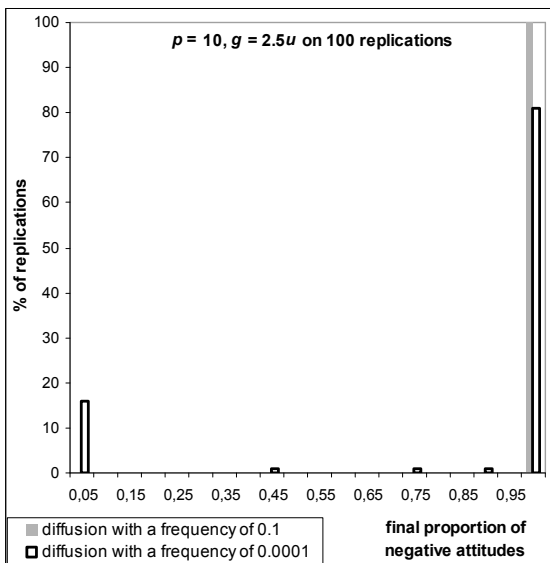
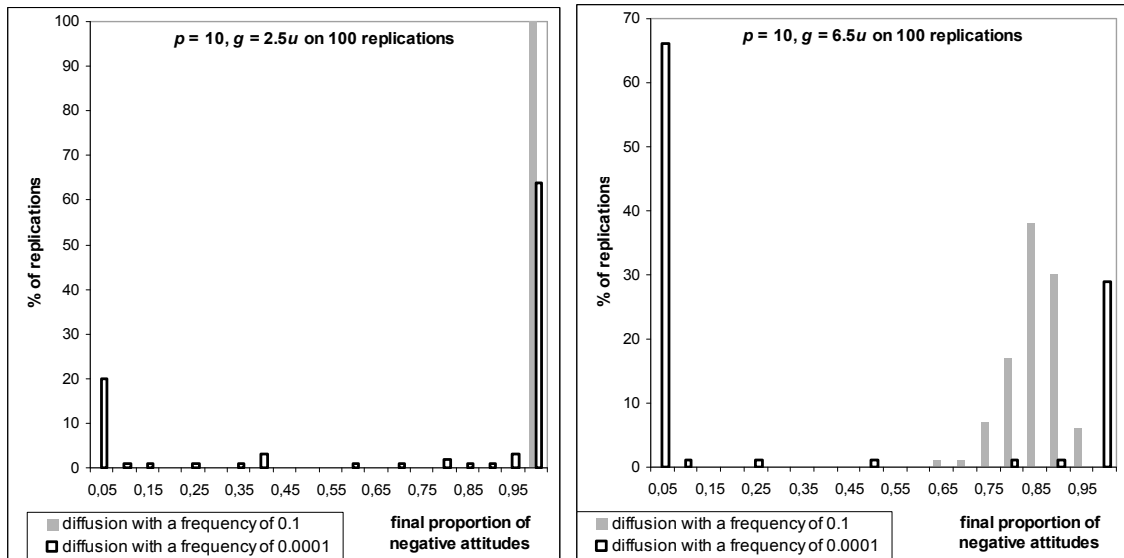


Figure 16. DIT.  $g = 2.5u$

Figure 17. DIT.  $g = 6.5u$



The proportion of negative final attitudes is approximately 100% for  $g = 2.5u$  and  $f = 0.1$  for both models, compared with an expected 73% for the CIT and 64% for the DIT without connection. The phenomenon observed in the previous case is enhanced: globally, the individuals are more inclined to transmit the major feature, because in the beginning of the process, they only transmit the feature as soon as they receive this characteristic. The shift toward negative attitudes is also very high in the case of  $g = 6.5u$ : 83% of negative attitudes for the CIT model compared with 36% (4/11) expected in the case of no connection, and 73% of negative attitudes for the DIT compared with 27% (3/11) expected in the case of no connection. Figures 18 and 19 clearly illustrate this; the very rapid increase in the major feature transmission occurs at the expense of the minor feature transmission.

Again, the frequency  $f = 0.0001$  yields an average (for all the replicas) of final negative attitudes, which is similar to the expected average without connection, and the distribution of the results shows that the population has always almost completely the same attitude sign. The explanation proposed in the previous paragraph holds.

Figure 18. CIT model. Mean proportion of each feature as subject of discussion in case of  $p = 10$  and  $f = 0.1$  (left,  $g = 2.5u$ , right,  $g = 6.5u$ )

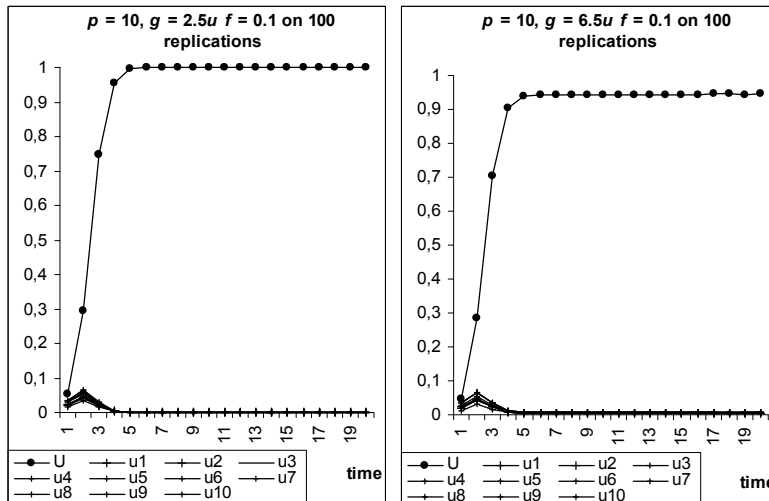
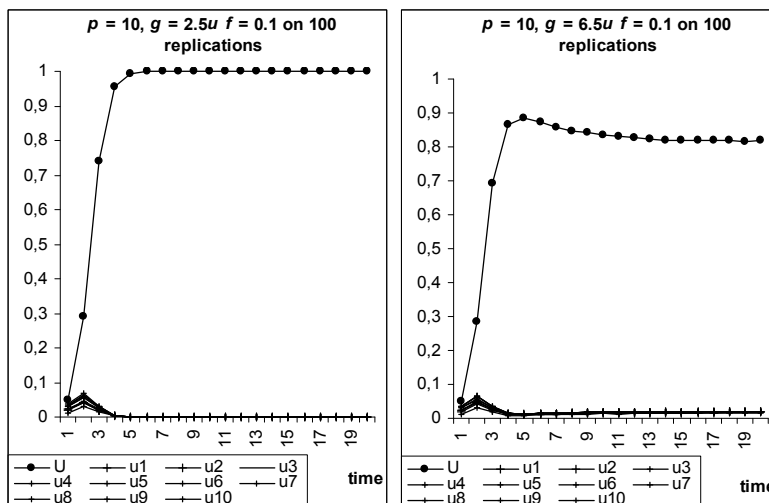


Figure 19. DIT model. Mean proportion of each feature as subject of discussion in case of  $p = 10$  and  $f = 0.1$  (left,  $g = 2.5u$ , right,  $g = 6.5u$ )



## Discussion

This paper has explored two versions of a simple model for filtering out incongruent features of an object. The models have their basis in a general assumption, which is in agreement with a number of observations in social psychology: people tend to pay less attention to the features of an object that contradict their global feeling, and are inclined to speak less about them. Both versions of the model can show “non-rational” behaviors: they are sensitive to the order in which the individual is told about the object features. Predicting the behavior of a population of isolated individuals that receive the features in a random order is quite easy: each possible trajectory (corresponding to an order of feature reception) take place with the same probability. However, this may lead to counter-intuitive results. For instance, a majority of a population develops a negative final attitude toward an object, whereas each agent’s initial global attitude was positive, the sum of the object feature attitudes being zero (the object was globally neutral).

The dynamics of interacting individuals is of course quite difficult to access analytically. This study has investigated these dynamics by simulation, restricting the scenario to the case of a

totally connected network and the consideration of objects with one major negative feature and several positive ones in such a way that the sum of the feature attitudes is zero.

The finding that the major negative feature tends to be transmitted more often than anticipated is important. The consequence is that the final attitude of the population tends to be more negative than it would be if the individuals received random messages about the features while being isolated. This effect becomes more important when the number of minor features increases, which can be related directly to the findings of Mowen (1998) who suggests that word-of-mouth communications have a negative bias and that consumers give more weight to negative than to positive information.

Several differences between the versions of the models are noticeable. The main one is that the CIT model can retain an incongruent feature. In the DIT model, retained features can become incongruent, but as soon as they are retained, they are necessarily congruent. So the DIT model may seem more restrictive than the CIT model. Moreover, the DIT model can lead to an incongruence threshold that is very close to the congruence threshold, in which case the individual can very easily change of attitude sign.

The models remain simplistic, embracing as they do an incremental approach, allowing users to study and understand precisely the properties of the models. Despite this simplicity, the models did not easily predict the cumulative effect of the major feature propagation among the interacting individuals observed in the simulations.

One important test would be to confront the model with real observation. For instance, during the recent campaign of the referendum about the EU constitution in France, the population was initially globally positive (polls indicated about 60% of favourable opinions), but in the end a majority voted against. The model would then suggest that the effect of a major negative feature leads to the filtering of a set of minor positive ones (of course as perceived as such).

Future improvements should focus on an attempt to include psycho-sociology theories and observations. Particularly relevant are theories of minority influence (Moscovici, 1979; Moscovici, 2000; Mugny & Pérez, 1998), dealing with the assimilation of incongruent information that for some time would seem to have no effect, but can suddenly trigger a complete change of attitude. In view of the future evaluation and improvement of the models, possible references to consider include the work of Chaiken and Eagly (1998, p. 201) on the stability of attitudes, or the elaboration-likelihood model (Chaiken & Eagly, 1998; Foss & Littlejohn, 2005; Cacioppo & Petty, 1981).

## REFERENCES

- Allport, G.W. & Postman, L. (1947). *The psychology of rumor*. New York: Holt, Rinehart & Winston.
- Anderson, N.H. (1971). Integration theory and attitude change. *Psychological Review*, 78, 171-206.

- Deffuant, G., F. Amblard, G. Weisbuch & Faure, T. (2002). "How can extremism prevail? A study based on the relative agreement interaction model." *Journal of Artificial Societies and Social Simulation* vol. 5, n°4, <<http://jasss.soc.surrey.ac.uk/5/4/1.html>>
- Deffuant, G., Huet, S. and Amblard, F. (2005). An individual-based model of innovation diffusion mixing social value and individual payoff dynamics. *American Journal of Sociology* 110: 1041-1069.
- Deffuant, G., Neau, D., Amblard, F. & Weisbuch, G. (2001). Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3, 87-98.
- Deffuant, G., Weisbuch, G., Amblard, F., Faure T. (2003). Simple is beautiful... and necessary. *Journal of Artificial Societies and Social Simulation* vol. 6, n°1, <<http://jasss.soc.surrey.ac.uk/6/1/6.html>>.
- Eagly, A.H.& Chaiken, S. (1993). *The Psychology of Attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
- Galam, S. (2005). Modeling rumors: The No Plane Pentagon French Hoax Case. *Physica A*, 32, 571-580.
- Hegselmann, R. & Krause, U. (2002). Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation. *Journal of Artificial Societies and Social Simulation*, vol.5, n° 3.
- Huet, S.& Deffuant, G. (2006) Effets d'un filtre cognitif sur la diffusion d'information. *MOSIM Conference, May 2006, Rabat*.
- Lawson, G.C. & Butts, C.T. (2004) Information Transmission Through Human Informants: Simulation. In K. Carley (Ed) *Proceedings of North American Association for Computational Social and Organizational Science*. Pittsburg PA: Carnegie Mellon University
- LittleJohn, S.W., Foss, K.A. (2005) *Theories of Human Communication*.(8<sup>th</sup> Edition), London, Thomson Wadsworth.
- Moscovici, S. (1996) *Psychologie des minorités actives*. Presses Universitaires de France Quadrige, 288 p.
- Moscovici, S. (2000) *Psychologie sociale des relations à autrui*. Nathan Université, France.
- Mowen, J.C. (1998) Group processes I: Group-Dyadic and Diffusion Processes. In *Consumer Behavior*, Upper Saddle River, New Jersey, Prentice-Hall, Inc. Pp483-518.
- Mugny, G. & Pérez, J.A.(1998) Recherches sur l'influence sociale. In J-L.Beauvois, R.V. Joule, J-M. Monteil. *20 ans de psychologie sociale expérimentale francophone. La psychologie sociale*, 6, Presse Universitaire de Grenoble.13-66.
- Petty, R. & Cacioppo, J. (1981). Epilog: A General Framework for Understanding Attitude Change Processes. In: *Attitudes and Persuasion: Classic and Contemporary Approaches*. Dubuque, Iowa: Wm. C. Brown,
- Snijders, T.A.B. (1998) Methodological Issues in Studying Effects of Networks in Organizations. *Computation & Mathematical Organization Theory*, 4, 205-215.
- Tsimring L.S. & Huerta, R. (2003) Modelling of contact tracing in social network. *Physica A*, vol. 325, n°1-2, 33-39.
- Urbig, D. (2003) Attitude Dynamics with Limited Verbalisation Capabilities. *Journal of Artificial Societies and Social Simulation*, 6, 1, <<http://jasss.soc.surrey.ac.uk/6/1/2.html>>.
- Urbig, D. & Lorenz, J.(2004). Communication regimes in opinion dynamics: Changing the number of communicating agents; *Proceedings of the Second Conference of the*

*European Social Simulation Association (ESSA), Valladolid, Spain, 16-19 September 2004.*

Wu, F., Huberman, B.A., Adamic, L.A. & Tyler, J.R. (2004) Information flow in social groups. *Physica A*, 337, 327-335.





## Chapter 2.2 Disregarding information – a double modelling approach

Title: Differential equation models derived from an individual-based model can help to understand emergent effects

Authors: Sylvie Huet and Guillaume Deffuant

Published in JASSS, 2008



### Abstract

We study a model of primacy effect on individual's attitude. Typically, when receiving a strong negative feature first, the individual keeps a negative attitude whatever the number of moderate positive features it receives afterwards. We consider a population of individuals, which receive the features from a media, and communicate with each other. We observe that interactions favour the primacy effect, compared with a population of isolated individuals. We derive a differential equation system ruling the evolution of probabilities that individuals retain different sets of features. The study of this aggregated model of the IBM shows that interaction can increase or decrease the number of individuals exhibiting a primacy effect. We verify on the IBM that the interactions can decrease the primacy effect in the conditions suggested by the study of the aggregated model. We finally discuss the interest of such a double-modelling approach (using a model of the individual based model) for this application.

### Contents

The individual based model (IBM) .....	156
The dynamics of attitudes .....	156
Impact of interactions on the primacy bias .....	158
Bird's-eye view of the IBM for complex "interaction" cases .....	162
Building the model .....	162
Comparison of the aggregated model with the individual-based model .....	165
Using the aggregated model to better understand the individual-based model..	165
The aggregated model helping to predict the IBM.....	166

The differential equation model for $2u \leq g < 3u$ .....	166
Comparison of the aggregated model with the individual-based model .....	167
Impact of the diffusion by the media.....	171
Conclusion, discussion.....	171
References.....	173

This paper focuses on a recently proposed simple individual based model (IBM) of the primacy effect (Deffuant and Huet 2007). The primacy effect (Asch 1946; Miller and Campbell 1959) occurs when somebody, who encounters a positive and then a negative message forms judgments which tend to be positive (of course positive and negative can be inverted). The model makes the assumption, that the primacy effect is related to the tendency to maintain internal consistency: the features contradicting the current global attitude tend to be filtered. Deffuant and Huet (2007) used individual based simulations to show that interactions can increase the primacy effect in a population.

This paper mainly focuses on the methodology used to study the IBM: the derivation of a differential equation model ruling over time the probability of individuals to belong to different groups. This approach can be called "double-modelling", because it needs to develop a model (differential equations) of an IBM (Deffuant 2004). The expected interest of the differential equation model is to provide explanations of the collective effects observed in IBM simulations, through an aggregated view of the IBM behaviour. We illustrate and discuss this on the particular case of the primacy effect IBM.

This approach is expressed and investigated in different researches. In ecology, Grimm (1999) encourages researchers to compare IBMs with aggregated models. Deffuant (2004) formalises a "double-modelling" which precisely advocates for the development of aggregate models to "theorise" IBMs. Focusing on attitude dynamics study, this double-modelling approach has been applied to the bounded confidence model (Deffuant et al. 2001, Hegselman and Krause 2002) by Ben-Naim P., Krapivsky L. et al (2003), Lorenz (2007), and Deffuant and Weisbuch (2007). Their purpose was to develop an exhaustive knowledge of this model asymptotic behaviour. Deffuant and Weisbuch (2007), using the same approach, improve the understanding of the extremist effect for this bounded confidence model. Martin, Deffuant et al (2004) applied this method to the study of a bit vector version of the bounded confidence model. They show the bigger limitation of the double modelling approach: the state space can be too large to be tractable. Edwards, Huet et al (2003), Edwards (2004), Huet, Edwards et al (2007) applied this method to study a stochastic IBM of binary behaviour diffusion individual model. They particularly aim at understanding the interaction effect in a random Erdős network. We present here a new application of this approach, which gives a particularly clear insight on the effect of the interactions.

The model, on which we apply this approach, uses the concept of attitude, understood as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favour or disfavour" (Eagly and Chaiken 1998). It is widely observed that attitudes exert selective effects at various stages of information processing (Eagly and Chaiken 1998): information may be filtered (ignored) by the individuals. In his theory of cognitive dissonance (1957), Festinger proposes some mechanisms for this selection: people seek out information that supports their attitudes and avoid information that challenges them, in order to minimise their cognitive dissonance. Following this theory, even if they assimilate information which contradicts their global attitude, people are reluctant to talk about it, because they avoid expressing their dissonance. Such selection mechanisms can imply sensitivity to the order of information delivery. In the present work, we are interested in people who are motivated to form an attitude about a particular object. Thus, they seek out relevant information and are sensitive to the media diffusion. Regarding the more recent

literature, the attitude strength perspective assumes, among others, that the more you have knowledge on a particular object, the better you resist to a counter attitudinal attack, particularly when the messages require a cognitive effort and when you are motivated to think (Visser, Bizer and Krosnick 2006). This means that one tends not to consider an argument against one's own current attitude. Haugtvedt and Wegener (1994) conclude in their seminal paper: "*when participants (to the experiment) were motivated to elaborate on the message content, primacy effects occurred.*"

Huet and Deffuant (2006) and Deffuant and Huet (2006, 2007) proposed a simple individual based model (IBM) of the individual primacy effect, which abstracts from the cited researches in social psychology. We particularly focussed on the following question: do the interactions between agents modify the likelihood of individual primacy effect in the population? With the simple model we consider, the answer is clearly positive. In some cases, the number of agents showing primacy effect is significantly higher, and in other cases significantly lower when agents interact than when they are isolated.

In Deffuant and Huet (2006), we derive and solve numerically a differential equation model of the individual based model, in order to better understand this particular impact of the interactions exposed to a short message (comprising a major negative feature and two positive features). It stressed that interactions favour the broadcasting of the major feature, which increases its probability to be received, thus giving an advantage to globally negative attitudes. Moreover, it appears that the differential equation model cannot reproduce the individual based model results when the frequency of the diffusion of the message by the media is too weak.

This paper extends this study by considering a more complex message, including two major negative and three positive features. We show that in this case, interactions can also decrease the number of individuals exhibiting the primacy effect. We study again this particular configuration through the corresponding differential equation model.

First, we describe the individual-based model and the impact of the interactions. In section 3, we present the methodology to build the differential equation model of the IBM and apply it to the case of a population exposed to a neutral complex message, composed of two major negative and three minor positive features. The analysis of the differential equation model allows us to understand better the impact of interactions and to assume that in some cases, interactions tend to decrease the primacy effect. In Section 4, we derive a new differential equation model, to test this assumption, and check its compatibility with the IBM. Finally, we conclude and discuss the benefit from the double-modelling approach.

## **The individual based model (IBM)**

### **The dynamics of attitudes**

Our model is strongly inspired by the dissonance theory (Festinger 1957) on the one hand, and on Allport's work on rumor diffusion (1947) on the other hand. To summarise, we assume that individuals avoid incongruent information, and, keep only important information.

We consider a population of  $N$  individuals forming a global attitude about an object. We define this object by a set of features  $F=(1,2,\dots,d)$ , which are associated with positive or negative real values  $(u_1,\dots,u_j,\dots,u_d)$  with  $u_j \in \mathfrak{R}$ . An individual can have a partial view of the object, in which case it has a real value for some features and nil for others. To simplify we use feature instead of feature value in the following.

The model is based on the congruency principle. A feature is congruent when it has the same sign as the individual's global attitude to the object, incongruent otherwise.

An individual  $i$  is characterised by:

- $g$ : An initial attitude (suppose the same for all individuals in the following).
- $L_i$ : A subset of  $F$  containing the features currently retained by the individual (empty at the beginning).
- $G_i = g + \sum_{j \in L_i} u_j$  : The global attitude about the object (related to information integration theory of Anderson (1971)).
- $\Theta_i$ : A threshold defining the absolute value since when an incongruent feature is judged enough high not to be filtered.
- A neighbourhood corresponding to the subset of individuals with whom  $i$  can communicate.

The dynamics of the model have four main aspects:

1. **Exposure to feature values.** We suppose that, at each time step:

- A media sends a randomly chosen feature to the individual following a delivery frequency  $f$ , which is, on average, the number of individuals who are exposed to a feature per iteration.
- An individual is exposed to feature values proposed by its neighbours during regular meeting (see 2.2.2.1. for details).

2. **Selective retaining:** The dynamics of filtering are determined by the individual incongruence threshold  $\Theta_i$ . Being told about feature  $j$ , the individual  $i$  will react as follows:

- If  $j$  is congruent  $\rightarrow i$  "retains" the feature  $j$ . This means that  $j$  is added to  $L_i$  (if  $L_i$  does not include  $j$  yet),
- If  $j$  is incongruent:
  - if  $|u_j| > \Theta_i \rightarrow i$  "retains" the feature  $j$ ;
  - otherwise  $i$  "ignores" the feature  $j$ . This means that  $j$  is filtered (not added to  $L_i$ ).

3. **Selective emission:** individuals only talk about congruent features

4. **Computation of attitude:** an individual computes its global attitude each time it retains a new feature. As presented in the characteristics of the individual, the global attitude to

the object is the sum of the feature values the individual retained and its initial attitude,  $g$ .

In the following, we consider that the individual incongruence threshold  $\Theta_i$  is a constant  $\Theta$  which is the same for all individuals of the population. Deffuant and Huet (2007) have considered various choices for  $\Theta_i$ , which can be dynamic. They showed the particular interaction effect we are interested in occurs for all of the various studied  $\Theta_i$ .

## **Impact of interactions on the primacy bias**

As explained in Deffuant and Huet (2007), by an analysis of the definition of the dynamics presented above, we have to distinguish two main cases: a first very simple one, which is not sensitive to order of feature exposure, and a second one, which is sensitive to this order. Thus, when the model is sensitive to delivery order, the interactions, modifying this order, can have a particular effect on the final state of the population.

### ***Individual trajectories can be sensitive to the order of feature exposure***

#### ***The trajectories with a message including major negative and minor positive features***

We now consider an individual with an initial attitude  $g > 0$ , and an object with at least one negative feature of absolute value higher than  $\Theta$  (called major incongruent feature), and positive features lower than  $\Theta$  (called minor congruent feature). Notice that the same reasoning can be done with inverted signs. In this case, the final attitude depends on the reception order:

- If the individual receives the negative feature first, if  $g$  is low enough, it can change its global attitude, and the positive features become incongruent. As they are lower than  $\Theta$ , they are not retained.
- If the individual receives the positive features first, they are necessarily retained.

When the individual attitude is sensitive to the feature reception order, we can observe primacy effect: first few received features define the individual's attitude sign.

This leads us to define a more concrete example, used in all the following. We suppose that the initial attitude  $g$  is positive. Then we consider an object described by 5 features: two major negative ones, valued at  $-U$ , such that  $U > \Theta$ , and three positive ones, valued  $u$ , such that  $u < \Theta$ . We suppose that the object is globally neutral, that is:  $3u - 2U = 0$ . For instance, we choose  $U = 6$  and  $u = 4$ , with  $\Theta = 5$ . We are interested in this paper in a more complex "message" example than in Deffuant and Huet (2007), which treats of a 3-feature message composed from one major negative feature and two minor positive features. One more time, remain that this more complex message of 5 features is the one used in the following to study the individual based model dynamics.

Figure 1 shows the evolution of a global attitude, for a particular reception order of the features. Initially, the individual has an attitude  $g = 6.5$ . First, it receives a positive feature,

which is retained because it is congruent, and its attitude increases to 10.5. Second, it then receives a negative feature, which is incongruent, but it is retained because its absolute value is higher than the threshold, and its attitude decreases to 4.5. Next, it receives the second negative feature, which is incongruent and also retained and its attitude decreases to -1.5. It is then exposed to the fourth and the fifth positive features, which are incongruent with an absolute value below the threshold, and therefore they are not retained. Its attitude thus does not change anymore. It has finally a negative attitude although the object is globally neutral. On figure 2, the individual receives firstly the three positive features consecutively, and then the two major negative features. All features are retained in this case, and the attitude follows another trajectory, leading to a final positive attitude.

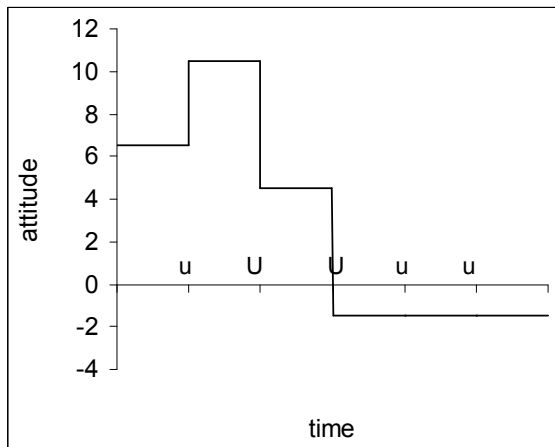


Figure 1: Examples of temporal evolution of the attitude of an individual with  $g = 6.5$ . On the left, the individual receives the features in order  $u, U, U, u, u$ . Its final attitude is negative.

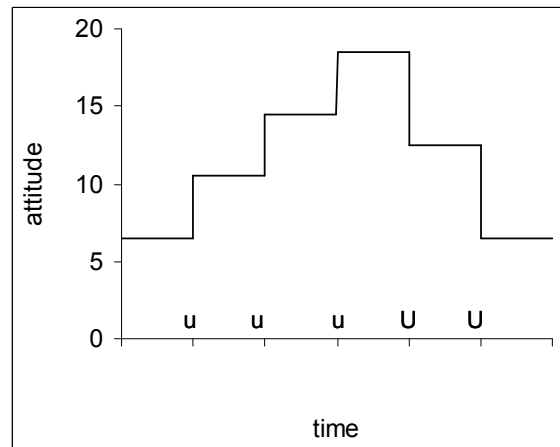
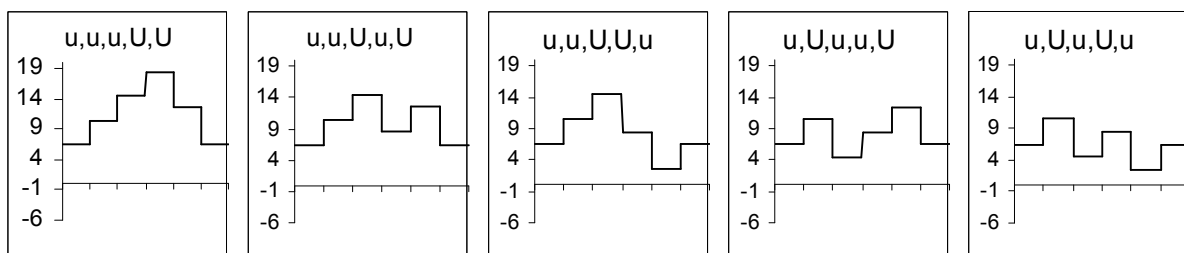


Figure 2: Trajectory for order of exposure  $u, u, u, U, U$ , and the final attitude is positive

Figures 3 shows the ten possible trajectories of attitude. The first, the two first, or at the most, the three first features determine the final sign of attitude: this is the primacy effect.

If we consider a population of isolated individuals, each individual trajectory has the same probability of occurring. It is thus very easy to predict the final part of negative versus positive people. It simply corresponds to the relative part of trajectories leading to a final negative attitude. From the figures 3, we can predict the final state of the population: 70 % of individual with a final positive attitude, because we observe 7 final positive trajectories out of 10 total trajectories (the presented seventeenth trajectories) , and, 30 % of individual with a final negative attitude, because we observe 3 final negative trajectories out of 10 total trajectories (the last three trajectories of figures 3).





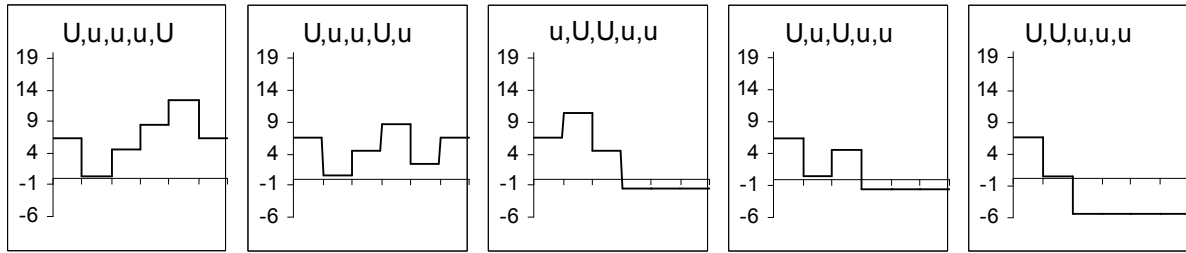


Figure 3. The 10 possible individual trajectories, for case  $g = 6.5$ , 5 features composed of 2  $U$  and 3  $u$  with  $U = -6$  and  $u = +4$ . Three trajectories over ten lead to a final negative attitude.

We know that the equiprobability of exposure is true for the media we chose. Nevertheless, when individuals interact, an individual can be exposed to a feature proposed by the media or by another individual. The interactions can modify the probability of presentation, and consequently the part of the population with a final negative attitude. We now investigate this impact of interactions.

### *The impact of interactions on final attitudes*

Deffuant and Huet (2006, 2007), showed that interactions can increase the final part of the population with a negative attitude. This is the case for a message, composed of three features with one major negative attitude  $U$  and two minor positive attitudes  $u$  when  $g$  has a value in a range from 0 to  $u$ . We are now looking for the same "increase" effect of the interaction between people for our more complex neutral 5-feature message.

Before that, let us describe in more details the model of interactions.

### *Interaction model*

The interaction mechanism is very simple. The aim is to ensure that, on average, one individual meets another individual in each iteration. As it is a stochastic process, it remains possible that one individual does not meet any other, or meets several others during an iteration.

We consider two cases:

- an individual talks only about the congruent features it retained (only congruent feature transmission);
- an individual talks about all the features it retained (any feature transmission).

The complete algorithm, containing exposure to the diffusion by media and exposure from interaction is:

For a population of  $N$  individuals, at each time step:

$N$  times repeat:

- *Media diffusion.* choose individual  $i$  at random with probability  $f$ , choose feature  $j$  at random in the object, send feature  $j$  to individual  $i$ .

- *Interactions*: choose couple of individuals  $(i,j)$  at random:
  - ◇ *Only congruent feature transmission case*:  $i$  tells  $j$  about one of its randomly chosen congruent features
  - ◇ *All feature transmission case*:  $i$  tells  $j$  about one of its randomly chosen features (congruent or not).

### *Interactions can change the number of primacy effects*

We run simulations of our IBM with our 5-feature message and an initial positive attitude  $g = 2.5$ . We have  $0 < g < u$ . Such an attitude value should allow us, following what we have found in Deffuant and Huet (2007) to observe that interactions increase the population part exhibiting the primacy effect. For these parameters, figure 4 shows a comparison between the number of final negative attitudes for isolated individuals and this number for interacting individuals (for both transmission of congruent features only, and transmission of all features). We observe that interactions induce more negative final attitudes than the isolated case. Indeed, we obtain 83% of negative individuals with transmission of only congruent features, but only 70% for isolated individuals. This impact of interactions is even higher when individuals can transmit any retained feature, even though it is not statistically significant

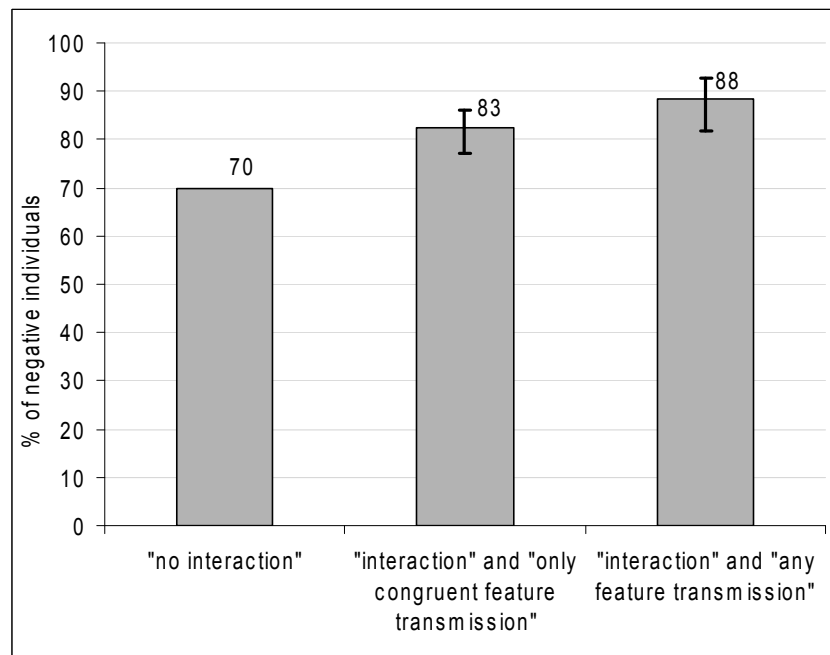


Figure 4. Final percentage of negative individuals, averaging on 100 replicas, for  $g = 2.5$  for three various dynamics: isolated individuals; interacting individuals transmitting only congruent features and interacting individuals transmitting any feature.

In Deffuant and Huet (2006) for the 3-feature message, we used a differential equation model of the IBM to analyse this impact of interactions. We are now considering again this approach in the case of the 5-feature message.

## Bird's-eye view of the IBM for complex "interaction" cases

We now consider the probability of individuals to belong to different groups (defined by the features they retained) over time. We assume that the dynamics of these probabilities (corresponding to a infinite population) reflect the evolution of the IBM. We already know that this assumption has a limited validity: Deffuant and Huet (2006) showed that it supposes a significant level of media diffusion. For weak frequency of diffusion (0.001 and less), there are many time steps without any diffusion from the media, which contradicts the assumptions behind the differential equation model. Therefore, we now suppose that the diffusion frequency is high enough.

First, we write down the differential equations ruling the probabilities to belong to the different groups, solve it numerically and compare it with the IBM. Secondly, we analyse the differential equation model to learn more about the interaction impact. In the following, we also use "aggregated model" to denote the differential equation model.

### Building the model

The general idea to build the aggregated model is to consider groups of agents and to define the transfer equations ruling the flows of probability densities between the groups. First, we need to determine the groups, and the possible flows between them.

A group is defined by a possible list of retained features, which may appear at any moment of the simulation. The list of feature retained depends on the order of feature exposure. Thus, we have to begin with the study of what the different exposure orders imply. Table 1 lists all the various orders possible for our message (defined in 2.2.1.1.). It also shows that the final global attitude sign of an individual exposed to a particular order of features depends on the value of  $g$ . Thus, considering all the ten possible orders, the initial attitude  $g$  splits into different value segments to define a unique distribution of final global attitude sign on all orders.

The table shows that the primacy effect can be observed (i.e. that individuals are sensitive to event order) for  $0 \leq g < 3u$  because some trajectories are finally positive while others are negative. For this given particular message, the exhaustive study of the interaction effect in complex cases needs to build six different aggregated models of group dynamics. Indeed a value segment of  $g$  defines a particular final distribution of positive and negative global attitude and, thus, as we will see later, a particular given simplification of groups to model. Each segment defining a particular list of groups, each segment leads to a particular aggregated model.

Exposure order	$g < 0$	$0 \leq g < 0.5u$	$0.5u \leq g < u$	$u \leq g < 1.5u$	$1.5u \leq g < 2u$	$2u \leq g < 3u$	$g \geq 3u$
UUUUU	-	-	-	-	-	-	+
UuUuu	-	-	-	-	-	+	+
UuuUu	-	-	-	-	+	+	+
UuuuU	-	-	-	-	+	+	+
uUUuu	-	-	-	-	-	+	+
uUuUu	-	-	-	+	+	+	+
uUuuU	-	-	+	+	+	+	+
uuUUu	-	-	-	+	+	+	+
uuUuU	-	+	+	+	+	+	+
uuuUU	-	+	+	+	+	+	+

Table 1. Final sign of the global attitude for the ten different trajectories and all different values of  $g$

We select one set of values of  $g$  to construct the corresponding aggregated model:  $u \leq g < 1.5u$ . Figure 5 shows the transition graph of the different groups to model. In this case, 7 groups have to be considered. We start from a group having no features, able to receive a  $U$  or a  $u$ . In one hand, receiving a  $U$  implies the individual will always has a final negative global attitude, whatever you receive after. Thus, the second group is the one of people having received  $U$  at first. In the other hand, receiving a  $u$  is a third group to consider. As this third group does not allow the decision about the final global attitude sign, we continue to develop the branch. Having received  $u$ , it is possible to receive  $u$  or  $U$ ...We continue until each branch can be stopped because it defines without ambiguity the final sign of the global attitude.

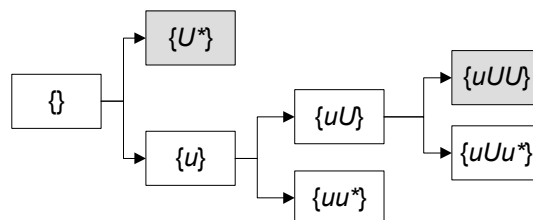


Figure 5. The graph of transitions between the groups for  $u \leq g < 1.5u$ , defined by the set of retained features. The groups with a negative attitude are in grey.

The second stage of the modelling approach is to determine the flow through each transition (i.e. each considered group). This requires evaluating the probability that the agents in each group retain a feature, which makes them change their group. This probability is directly related to the features, which are sent by each group. This is broken down in table 2 in the case of transmission of congruent features only. For example, you can read in the third column of the table 2 that all individuals who have received a  $U$  at first only talk to others about feature  $U$ .

Group	Media	{U*}	{u}	{Uu}	{uu*}	{uUU}	{uUu*}
Communicated features	U, u	U	u	U	u	U	U

Table 2: communicated features for each group in the case  $u \leq g < 1.5u$  or  $U$

This work done, it is possible to write down the differential equations for each group, summing up the flows in and subtracting the flows out the group. For  $u \leq g < 1.5u$ , we get:

$$\begin{aligned}
\frac{dS_0}{dt} &= -S_0(f + S_u + S_{U^*} + S_{uU} + S_{uu^*} + S_{uUU} + S_{uUu^*}) \\
\frac{dS_{U^*}}{dt} &= S_0 \left( \frac{2f}{5} + S_{U^*} + S_{uUU} \right) \\
\frac{dS_u}{dt} &= S_0 \left( \frac{3f}{5} + S_u + S_{uu^*} + S_{uUu^*} + S_{uU} \right) \\
&\quad - S_u \left( \frac{4f}{5} + S_{U^*} + S_{uUU} + \frac{2}{3}(S_u + S_{uU} + S_{uUu^*} + S_{uu^*}) \right) \\
\frac{dS_{uU}}{dt} &= S_u \left( \frac{2f}{5} + S_{U^*} + S_{uUU} \right) - S_{uU} \left( \frac{3f}{5} + \frac{1}{2}(S_{U^*} + S_{uUU}) + \frac{2}{3}(S_u + S_{uUu^*} + S_{uu^*} + S_{uU}) \right) \\
\frac{dS_{uu^*}}{dt} &= S_u \left( \frac{2f}{5} + \frac{2}{3}(S_u + S_{uUu^*} + S_{uu^*} + S_{uU}) \right) \\
\frac{dS_{uUU}}{dt} &= S_{uU} \left( \frac{f}{5} + \frac{1}{2}(S_{U^*} + S_{uUU}) \right) \\
\frac{dS_{uUu^*}}{dt} &= S_{uU} \left( \frac{2f}{5} + \frac{2}{3}(S_u + S_{uU} + S_{uu^*} + S_{uUu^*}) \right)
\end{aligned} \tag{1}$$

with:

$S_0$ : proportion of individuals with a void list of retained features,

$S_u$ : proportion of individuals with a list of retained features containing only  $u$ ,

$S_{U^*}$ : proportion of individuals following all trajectories beginning with  $U$ ,

$S_{uU}$ : proportion of individuals with a list of retained features containing only  $u$  at first and  $U$  at second

$S_{uu^*}$ : proportion of individuals following all trajectories beginning with  $uu$ .

$S_{uUU}$ : proportion of individuals following all trajectories beginning with  $uUU$ .

$S_{uUu^*}$ : proportion of individuals following all trajectories beginning with  $uUu$ .

$f$ : frequency of media feature communication.

We compute finally the evolution of groups at the end by calculating, for each group  $S_G$  with  $G \in \{0, u, U^*, uU, uu^*, UU, uUu^*\}$ :

$$S_G = S_G + \frac{dS_G}{dt} dt \tag{3}$$

The systems can be simulated considering different values for  $dt$ . After having tested various possible values for  $dt$ , it appears  $dt = 0.1$  is weak enough to approximate correctly the

differential equation system. Indeed, the results for  $dt = 0.1$  are exactly the same as the one obtained with smaller values of  $dt$ .

### Comparison of the aggregated model with the individual-based model

For the IBM, we consider a population of 5041 individuals. From runs of the IBM and aggregated model, it results that the part of final negatives in the population for  $u < g < 1.5u$  is 53.3% on average for the IBM (with a minimum of 45% and a maximum of 64% on 100 replicas) and 53.2% for the aggregated model with  $dt = 0.1$ . It appears that the aggregated model gives an accurate approximation of the average number of negative individuals in the population.

Figure 6 shows the evolution of the part of each group during the simulation for the IBM on the one hand, and for the aggregated model on the other hand. Group results for the IBM are built by a concatenation of the individual level results. One more time, we observe that both models, IBM and aggregated, give very close results.

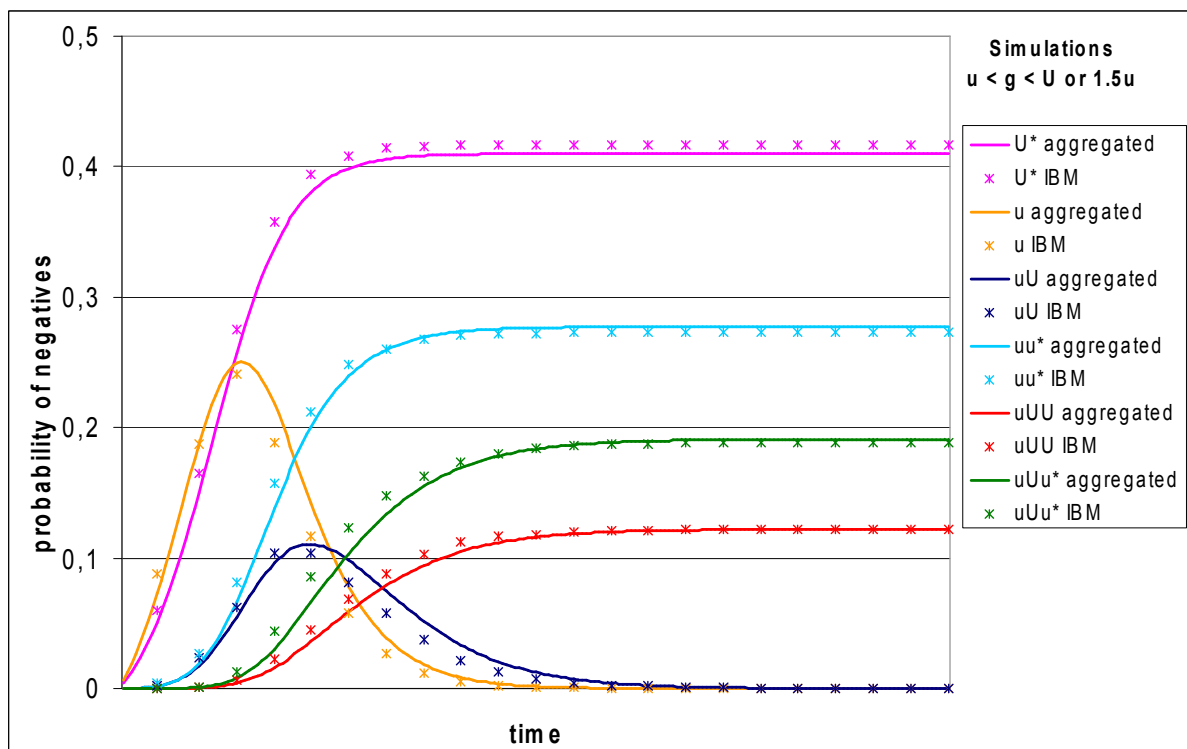


Figure 6. Comparison of trajectories of each groups of aggregated and IBM model. One measure of the IBM's replicas is put all the ten measures of the aggregated model

### Using the aggregated model to better understand the individual-based model

From the results of the aggregated model, we obtain the proportion at each time step of the negative feature  $U$  communicated during interactions. This proportion of  $U$  emission by

interaction can be compared with the proportion of  $U$  emission from the media. Figure 7 shows this comparison for the particular message and initial value of attitude we study.

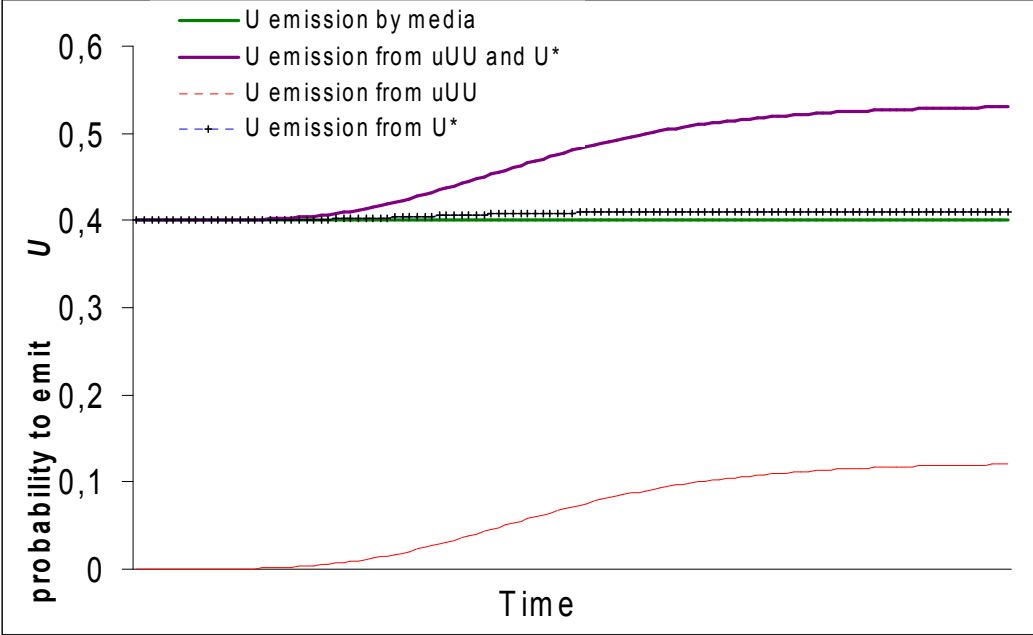


Figure 7. Comparison of probability of  $U$  emission due to interaction with the probability of  $U$  emission due to medium for  $u \leq g < 1.5u$

We see on figure 7 that the global probability of  $U$  emission by interaction begins at a value equal to probability of  $U$  emission by medium. It increases with the  $U$  emission from the negative group  $uUU$ . Thus, referring to figure 5, we can remark that the presence of the negative group  $uUU$ , following positive groups in the tree whereas no positive groups follows a negative one, induces a diffusion advantage in favour of  $U$ . This explains the increase of negative final states.

Can such an advantage be in favour of  $u$  for a different value of  $g$ ? The aggregated approach will now help us to answer to this question.

### The aggregated model helping to predict the IBM

From the previous analysis, and from the observation of the table 1, we select the segment of initial attitude  $2u \leq g < 3u$  for which we notice that the transition branch beginning with  $U$  leads to positive and negative groups whereas the one beginning with  $u$  only leads to positive groups.

### The differential equation model for $2u \leq g < 3u$

Figure 8, showing the transition graph of the different groups to model, illustrates our attitude segment choice. Indeed, one can anticipate that group  $Uu^*$  will increase the frequency of positive feature communication.

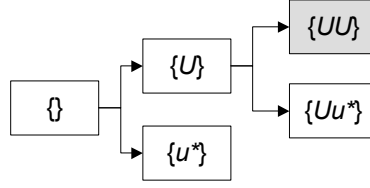


Figure 8. The graph of transitions between the groups for  $2u \leq g < 3u$ , defined by the set of retained features. The groups in grey have a negative attitude.

Table 3 determines the flow through each transition.

Group	Media	{U}	{u*}	{UU}	{Uu*}
Communicated features	$U, u$	<i>none</i>	$u$	$U$	$u$

Table 3: communicated features for each group in the case  $2u \leq g < 3u$  or  $2U$

Then we can write down the differential equations for this new case,  $2u \leq g < 3u$ :

$$\begin{aligned}
\frac{dS_0}{dt} &= -S_0(f + S_{u^*} + S_{UU^*} + S_{Uu^*}) \\
\frac{dS_{u^*}}{dt} &= S_0\left(\frac{3f}{5} + S_{u^*} + S_{Uu^*}\right) \\
\frac{dS_U}{dt} &= S_0\left(\frac{2f}{5} + S_{UU^*}\right) - S_U\left(\frac{4f}{5} + \frac{S_{UU^*}}{2} + S_{u^*} + S_{Uu^*}\right) \\
\frac{dS_{UU^*}}{dt} &= S_U\left(\frac{f}{5} + \frac{S_{UU^*}}{2}\right) \\
\frac{dS_{Uu^*}}{dt} &= S_U\left(\frac{3f}{5} + S_{u^*} + S_{Uu^*}\right)
\end{aligned} \tag{2}$$

with :

- $S_0$ : proportion of individuals with a void list of retained features,
  - $S_U$ : proportion of individuals with a list of retained features containing only  $U$ ,
  - $S_{u^*}$ : proportion of individuals following all trajectories beginning with  $u$ ,
  - $S_{UU^*}$ : proportion of individuals following all trajectories beginning with  $UU$ ,
  - $S_{Uu^*}$ : proportion of individuals following all trajectories beginning with  $Uu$ .
- $f$ : frequency of feature diffusion by the media.

### Comparison of the aggregated model with the individual-based model

As previously, we consider a population of 5041 individuals for the IBM. From runs of the IBM and aggregated models, it results that the part of final negatives in the population for  $2u \leq g < 3u$  is 0.4% on average for the IBM (with a minimum of 0.2% and a maximum of 0.8%



on 100 replications) and 0.4% for the aggregated model with  $dt = 0.1$ . The final part of negatives is very weak because, in most cases, individuals do not consider the negative features. Moreover, the interactions increase the diffusion of the positive features. Thus, it is very difficult for an individual to receive the negative information at an early stage, which is the only way for it to remain negative.

Figure 9 shows the evolution of the part of individuals in each group during the simulation for the IBM on the one hand, and for the aggregated model on the other hand. One more time, we observe that both models, IBM and aggregated, give very close results.

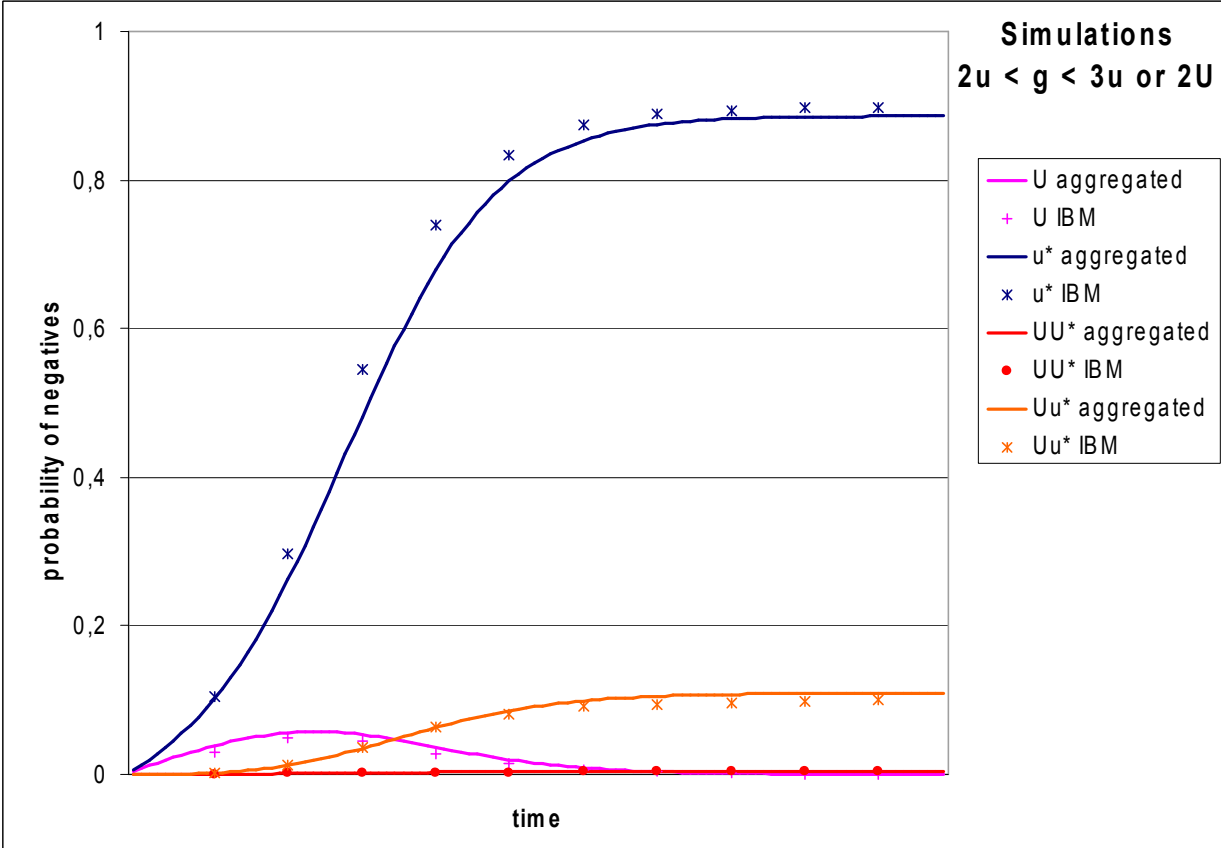


Figure 9. Comparison of trajectories of each groups of aggregated and IBM model for  $2u \leq g < 3u$ . One measure of the IBM's replicas is put all the ten measures of the aggregated model

As previously, from the results of the aggregated model, we obtain the proportion at each time step of the negative feature  $U$  communicated during interactions. This proportion of  $U$  emission by interaction can be compared with the proportion of  $U$  emission from the media.

Figure 10 shows this comparison for  $2u \leq g < 3u$ .

We see on figure 10 that the probability of  $U$  emission by interaction (from 0 to 0.0042) is always lower than the probability of  $U$  emission by the media (0.4).

Now, we can generalize our conclusions:

- From the observation of table 1, we deduce that, for  $0 \leq g < 1.5u$ , the interactions increase the primacy effect diffusion because the transition branch beginning with  $u$  contains negative groups whereas the transition branch beginning with  $U$  does not contain positive groups. Thus, for these three segments, the diffusion of the negative feature  $U$  is increased by interactions.
- Following the same reasoning, for  $1.5u \leq g < 3u$ , the interactions decrease the diffusion of primacy effect. We studied above the case  $2u \leq g < 3u$ . For  $1.5u \leq g < 2u$ , things are less clear. However, we can observe, from table 1, that half of exposure orders beginning with the negative feature  $U$  quickly lead to a positive global attitude (for the second received feature), whereas only one on six exposure orders beginning with the positive feature  $u$  leads less quickly to a negative global attitude (from the third received feature). Thus, with the transmission of only congruent features, the diffusion of  $u$  is favoured.

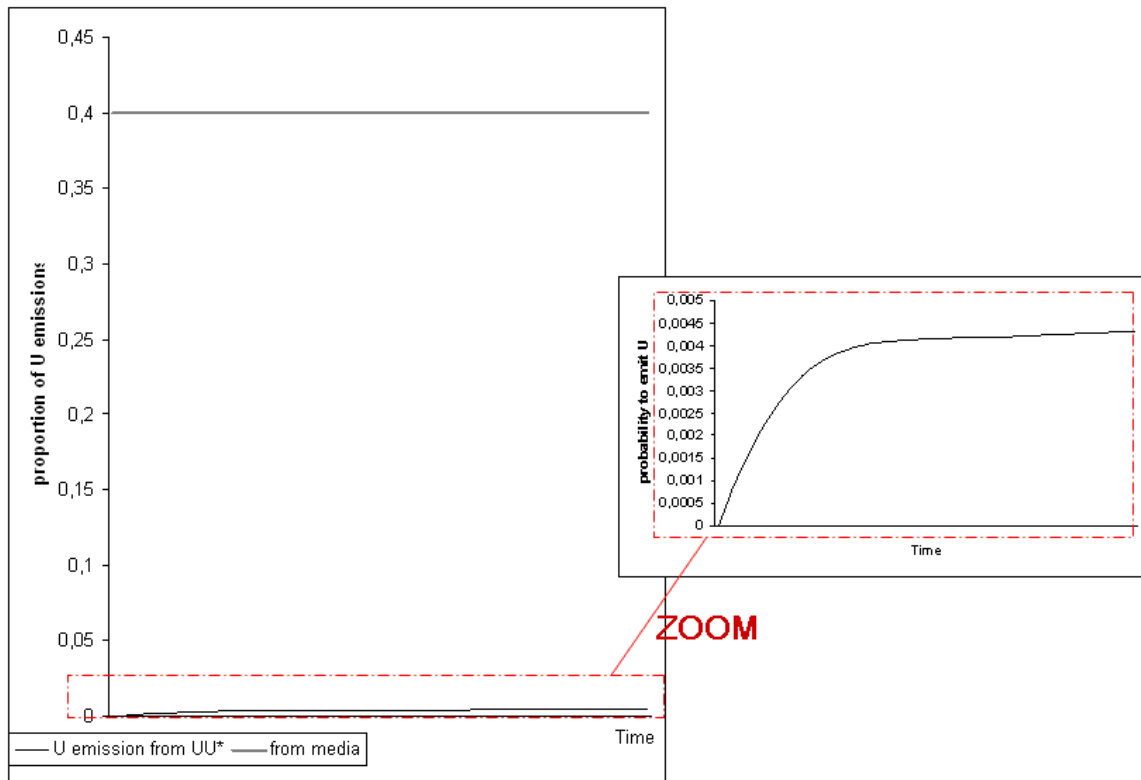


Figure 10. Comparison of probability of  $U$  emission due to interaction with the probability of  $U$  emission due to medium for the "decrease" interaction" effect case

A comparison between the IBM and the isolated case for all the considered initial values of  $g$  confirms our generalization. Figure 11 shows the number of final negative attitudes for the IBM with interactions (with transmission of only congruent features and transmission of all features) and in the isolated case. The isolated case is represented by dark bars, and the interaction cases (average results on 100 replicas) are represented in grey bars (transmission of all features) and in white bars (transmission of congruent features only).

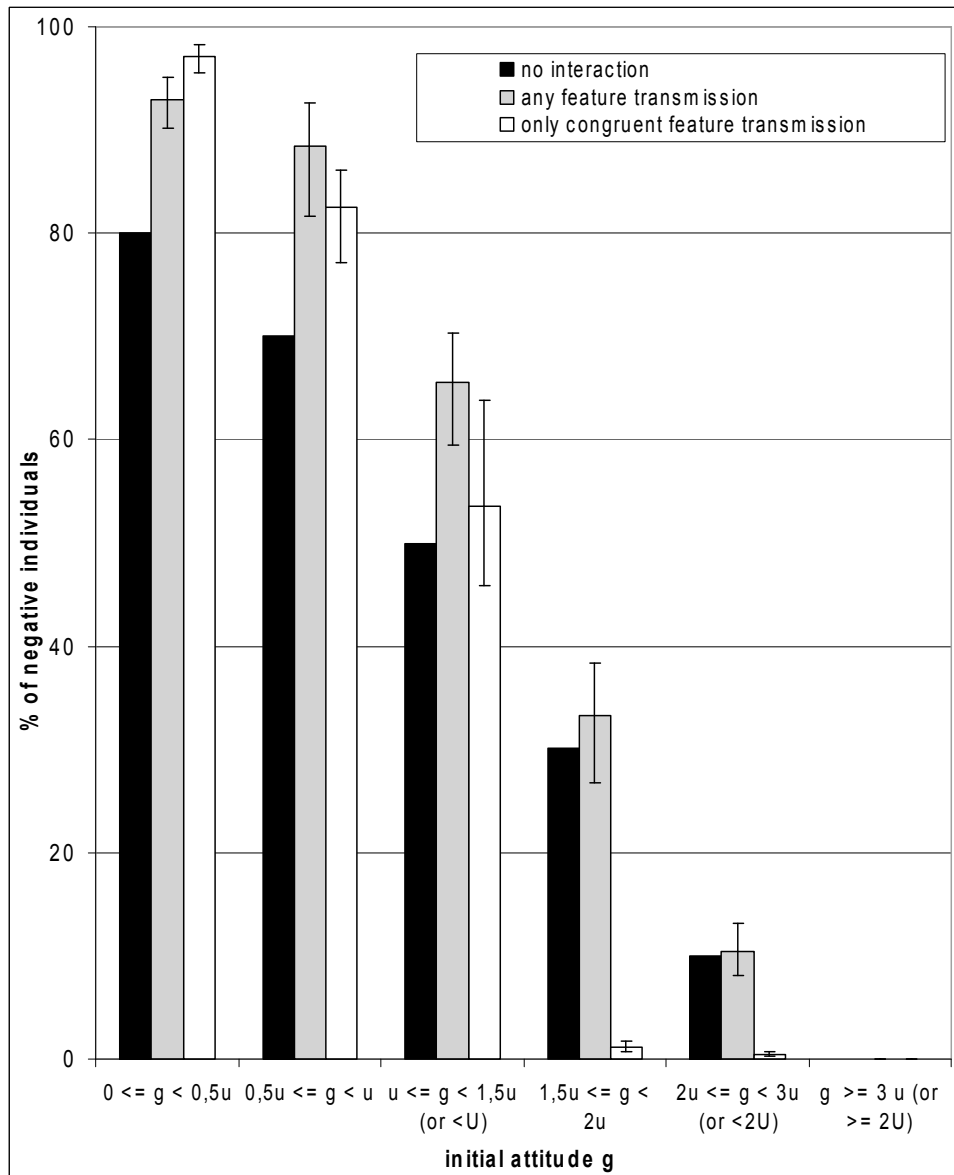


Figure 11. Final percentage of negative individuals for various value of  $g$  and for three dynamics: isolated individuals; interactions with both filters (reception and emission) and interactions reception filter only (average on 100 replicas). The errors bars represent the minimum and maximum values on 100 replicas.

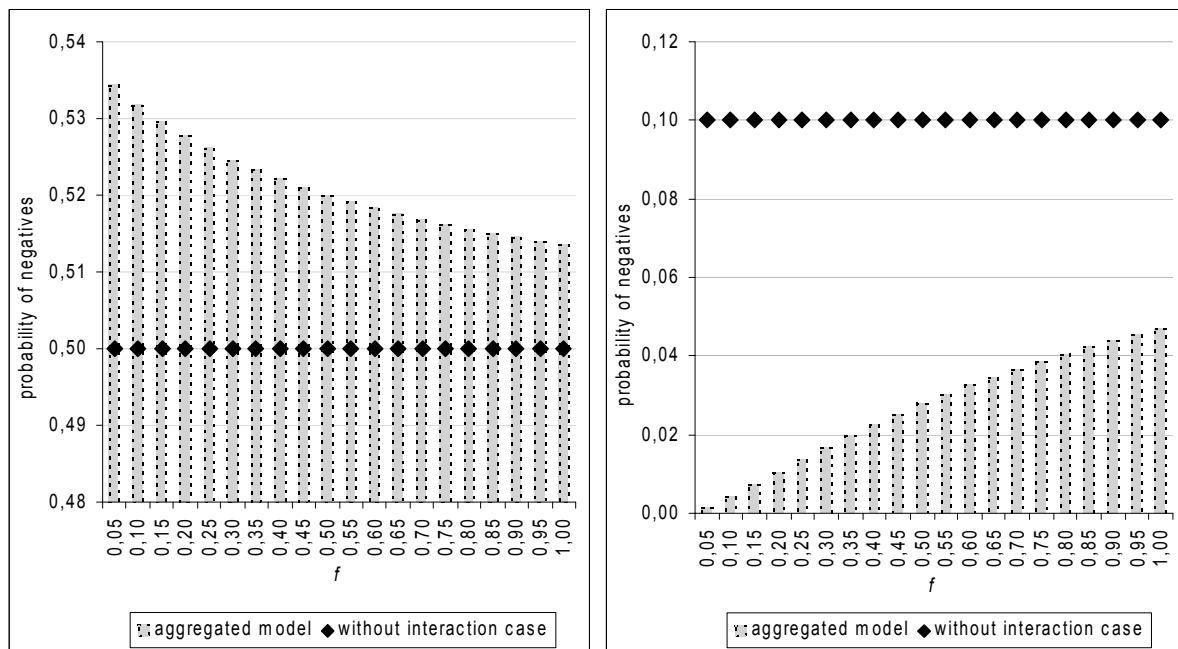
We observe on the left, for  $0 \leq g < 1.5u$ , that both interaction cases lead to a higher part of the population exhibiting the primacy effect.

On the contrary, on the right of the figure, for  $1.5u \leq g < 3u$ , we note that the interaction case with transmission of congruent features only leads to significantly less people exhibiting primacy effect than in the other cases. This effect takes place for initial attitudes between  $U$  and  $2U$ , which correspond to the values between the absolute value of the most negative feature  $U$  and the sum of the absolute values of positive features. Moreover, we note that interactions with transmission of all features do not lead to less primacy effect. In fact, as the aggregated approach shows, an individual having an initial attitude comprised between  $U$  and  $2U$  and receiving at first a negative feature has still a positive attitude. Due to the emission filter, it does not communicate about its negative retained feature while the others do communicate the positive feature. If it receives a positive feature right after the negative

one, it will never be negative. The only possibility to be negative is to receive the two negative features first. The probability of this case decreases because the interactions transmit almost only the positive features.

### Impact of the diffusion by the media

We can think that the frequency of diffusion, which defines how many individuals on average during one iteration are exposed to a feature delivered by the media (parameter  $f$ ), can change the result and suppress the interaction effect. From previous work on the IBM (Deffuant and Huet 2006), we know that for weak frequency of diffusion (0.001 and less), the model tends to yield replicas in which, either all final attitudes are positive, or all are negative. Thus, for a weak frequency of diffusion, the aggregated model cannot be equivalent to the individual based model. However, for higher frequency of diffusion, we can study the persistence of the impact of interactions.



Figures 12. Comparison of final negatives part for different value of the frequency parameter  $f$  in case "without interaction" and case "with interaction" with the aggregated models: on the left, "increase" interaction effect case (for  $u \leq g < 1.5u$ ); on the right, "decrease" interaction effect case at bottom (for  $2u \leq g < 3u$ ).

Figures 12 show the sensitivity of the results to variations of  $f$ . We notice that, the impact of interactions is higher for low frequencies, but even when the frequency is at its maximum value 1, the "increase" or "decrease" primacy effect due to interactions remain. Here the aggregated model gives the possibility to investigate very rapidly the average behaviour of the IBM.

### Conclusion, discussion

We studied an individual based model of "information filtering", which refers to the theory of cognitive dissonance of Festinger and work on rumour from Allport. In this model, for some parameters, the final attitude toward an object depends on the order of reception of the features. This can be interpreted as a variant of primacy effect, because the first

received features determine the final attitude. Supposing that a media broadcasts this features in a random order, one can easily predict the final state of a population of isolated individuals. The model is very simplistic, all individuals share the same initial attitude, the same threshold, the same feature values, and therefore its results should be seen as metaphorical. In addition to all these simplifications, we would like to stress a strong hypothesis that could remain unnoticed: all individuals are supposed motivated to process information about the object. Such a situation is very unlikely in real diffusion processes. On the contrary, the majority is often composed of poorly motivated individuals, who tend less to exhibit a primacy effect (Haugtvedt and Wegener, 1994).

When, in addition to the media, individuals can transmit some retained features to their neighbours through an interaction, the outcome is less straightforward to predict. Indeed, for some particular values of the initial attitude  $g$ , we observe that interactions modify the final part of the population exhibiting primacy effect compared to the case where individuals are isolated.

To understand better this difference, we build a differential equation model ruling the evolution of probabilities that individuals belong to different groups defined by a set of retained features. We solve it numerically and show that this aggregated model approximates very well the IBM results. Moreover, the analysis of the graph structuring the groups shows how interactions can favour the diffusion of the negative feature.

This explanation of the increase of primacy effect due to interactions led us to hypothesise that interactions can also decrease the primacy effect. The analysis of the different graphs of groups corresponding to different values of the initial attitude  $g$  allowed us to identify such a configuration. We checked on both the aggregated model and the IBM that the primacy effect is lower than in the isolated case.

The double modelling approach provided a significant enrichment for the analysis of the IBM. In particular, it guided us to formulate and verify more easily hypotheses on the IBM behaviour as in Edwards M. (2004), that we could have missed otherwise. This enrichment comes from the point of view on the dynamics brought by the aggregated model. We consider probability flows between groups instead of individuals. It provides a more compact view of the processes, which eases their understanding.

Moreover, the aggregated model provides asymptotic results, corresponding to an infinite population. In some cases, such results are useful as a reference.

However, some specific limitations of the aggregated model should be underlined:

- It is not possible as in Edwards and Huet S., Edwards M., Deffuant G. (2007) to globally substitute a single aggregated model to the IBM because the aggregated model graph of groups generally change with the message and the initial attitude.
- For low values of media diffusion frequency, this type of aggregated model is not appropriate.

In addition, it is possible to derive such aggregated models because the model is simple and has favourable properties. The task can rapidly become impossible when the model becomes more complex.

## References

- Allport, G.W., Postman, L. (1947) "The psychology of rumor." Russel & Russell, Inc., 1947, 1965.
- Asch, E.S. (1946) "Forming Impressions of Personality". *Journal of Abnormal and Social Psychology*, 41, pp. 258-290.
- Anderson, N. H. (1971) "Integration theory and attitude change". *Psychological Review*, 78, 171–206.
- Ben-Naim, E., Krapivsky P. L. and S. Redner (2003) "Bifurcations and Patterns in Compromise Processes", *Physica D: Non linear phenomena*, vol. 183, 190-204.1
- Deffuant, G., Neau, D., Amblard, F., Weisbuch, G. (2001). "Mixing beliefs among interacting agents", *Advances in Complex Systems*, 3, pp87-98, 2001.
- Deffuant, G. (2004) "Modéliser les Systèmes Complexes. Quelques pistes pour relever le défi." Habilitation dissertation, Ecole doctorale "Sciences Pour l'Ingénieur", Université Blaise Pascal de Clermont-Ferrand (France), 133 p.
- Deffuant, G. and Huet, S. (2006) "Collective Reinforcement of First Impression Bias". First World Congress on Social Simulation, Kyoto (Japan), August. Working paper, 10 pages.
- Deffuant, G. and Huet, S. (2007) "Propagation effect of filtering incongruent information". *Journal of Business Research*, 60(8): 816-825.
- Deffuant G., Weisbuch G. (2007) "Probability distribution dynamics explaining agent model convergence to extremism", in B. Edmonds, C. Hernandez and K. G. Troitzsch, eds., *Social Simulation. Technologies, Advances, and New Discoveries (Information Science Reference, Hershey, USA)*, 43-60.
- Eagly, A.H., Chaiken, S. (1998) "The psychology of attitudes". Thomson/Wadsworth; 1993, 1998, 800 pages
- Edwards M., Huet, S., Goreaud F., Deffuant G. (2003). "Comparing an individual-based model of behaviour diffusion with its mean field aggregate approximation". *Journal of Artificial Societies and social simulation*, vol. 6, n° 4,
- Edwards M. (2004) "Intérêt d'un modèle agrégé pour étudier le comportement et simplifier la simulation d'un modèle individu-centré de consommation couplé à un modèle de ressource en eau". Thèse en informatique, 17.09.2004, Université Paris 6 - Pierre et Marie Curie – Paris (France), 281 p.
- Festinger, L. (1957) "A Theory of Cognitive Dissonance". Stanford, CA: Stanford University Press.
- Grimm V. (1999) "Ten years of individual-based modelling in ecology: what we have learned and what could we learn in the future?". *Ecological Modelling*, 115:129-148.
- Hagtvedt C.P. and Wegener D.T. (1994) "Message Order Effects in Persuasion: An Attitude Strength Perspective". *Journal of Consumer Research*, vol. 21, 205-18.
- Hegselmann, R. and Krause, U. (2002) "Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation". *Journal of Artificial Societies and Social Simulation*, vol.5, 3. <http://jasss.soc.surrey.ac.uk/5/3/2.html>
- Huet, S. and Deffuant, G. (2006) "Effets d'un filtre cognitif sur la diffusion d'information". MOSIM 2006, Rabat.
- Huet S., Edwards M., Deffuant G. 2007. "Taking into account the variations of social network in the mean-field approximation of the threshold behaviour diffusion model". *Journal of Artificial Societies and Social Simulation*, volume 10, Issue 1, <<http://jasss.soc.surrey.ac.uk/10/1/10.html>>.
- Huet S., Edwards M., Deffuant G. (2007) "Taking into account the variations of social network in the mean-field approximation of the threshold behaviour diffusion model". *Journal of Artificial Societies and Social Simulation*, vol. 10, 1, <http://jasss.soc.surrey.ac.uk/10/1/10.html>.
- Lorenz J., (2007) "Continuous Opinion Dynamics Under Bounded Confidence: A Survey". *International Journal of Modern Physics C*, vol. 18 (2007), 1-20.
- Martin S., Deffuant G., Nadal J.P., Weisbuch G. (2004). "Comparing the Binary Opinion Vector Version of the BC Model with Different Approximating Aggregated Dynamics". *Proceedings of ESSA conference. Valladolid*.
- Mabrouk N., Deffuant G., Lobry C. (2007) "Confronting macro, meso and micro scale modelling of bacteria dynamics". 3rd International conference Model-to-Model – Marseille 2007
- Miller N. and Campbell D. T. (1959) "Recency and primacy in persuasion as a function of the timing of speeches and measurements". *Journal of Abnormal and Social Psychology*, 59, 1-9

Visser, P.S., Bizer, G.Y. and Krosnick, J.A. (2006) "Exploring the Latent Structure of Strength-Related Attitude Attributes". *Advances in Experimental Social Psychology*, vol. 38, 1-76.

# Chapter 2.3 Attraction-Rejection – designed from theories

November 25, 2009 17:35 WSPC/INSTRUCTION FILE

Title: Rejection mechanism in 2D bounded confidence provides more conformity

Authors: Sylvie Huet, Guillaume Deffuant, Wander Jager

Published in Advances for Complex Systems, 2008.

Advances in Complex Systems  
© World Scientific Publishing Company

## REJECTION MECHANISM IN 2D BOUNDED CONFIDENCE PROVIDES MORE CONFORMITY

HUET S.  
Cemagref, Laboratoire d'Ingénierie des Systèmes Complexes  
63172 Aubière, France  
sylvie.huet@cemagref.fr

DEFFUANT G.  
Cemagref, Laboratoire d'Ingénierie des Systèmes Complexes  
63172 Aubière, France  
guillaume.deffuant@cemagref.fr

JAGER W.  
Faculty of Economics and Business, University  
Groningen, The Netherlands  
w.jager@rug.nl

This paper explores the dynamics of attitude change on 2 dimensions (2D) as a result of social interaction. We add a rejection mechanism into the 2D bounded confidence (BC) model proposed by Deffuant et al [19]. The principle is that one shifts away from a close attitude of one's interlocutor, when there is a strong disagreement on the other attitude. The model shows metastable clusters, which maintain themselves through opposite influences of competitor clusters. Our analysis and first experiments support the hypothesis that, for a large range of uncertainty values, the number of clusters grows linearly with the inverse of the uncertainty, whereas this growth is quadratic in the BC model.

### 1. Introduction

Much behaviour, especially in conditions of a higher involvement, can be understood as originating from underlying attitudes. One may vote on an extreme national party because of a negative attitude towards immigrants, and buy organic products because of a positive attitude towards environmentally friendly production. Hence attitudes motivate behaviour and exert selective effects at various stages of information processing [1]. Consequently, in studying behavioural change it is essential to have an understanding of the underlying attitudinal dynamics that give rise to such a change. Attitude is here understood in its psychological meaning as a tendency to evaluate a particular entity with some degree of favour or disfavour. The dynamics of attitudes are closely related to social influence, which includes individ-

\*Cemagref, Laboratoire d'Ingénierie des Systèmes Complexes, BP 50085, 63172 F. Aubière

## Abstract.

This chapter explores the dynamics of attitude change on 2 dimensions (2D) as a result of social interaction.

We add a rejection mechanism into the 2D bounded confidence (BC) model proposed by Deffuant et al [19].

The principle is that one shifts away from a close attitude of one's interlocutor, when there is a strong disagreement on the other attitude. The model shows metastable clusters, which maintain themselves through opposite influences of competitor clusters. Our analysis and first experiments support the hypothesis that, for a large range of uncertainty values, the number of clusters grows linearly with the inverse of the uncertainty, whereas this growth is quadratic in the BC model.

## Contents

1 Introduction.....	177
2 Overview of the model.....	180
21 Purpose of the model.....	180
22 Start variables and scales .....	180
23 Process overview and scheduling.....	181
24 Initialisation.....	183
3 Analysis of several examples.....	183
31 Evolution with uncertainty $U = 0.2$ and intolerance threshold with $\delta = 0.7$ .....	183
32 Spatial organisation of the clusters and hypothesis of linearity of their number with $1/U$ .....	184
33 Influence of intolerance threshold $\delta > 0$ .....	185
34 Different values of uncertainty $U$ with intolerance threshold $\delta = 0$ .....	186
4 Systematic analysis of the number of clusters.....	187
41 Computing the number of clusters .....	187
42 Final number of clusters on the multiple clusters zone of the BC model .....	188
43 Final number of clusters on the one major cluster zone of the BC model.....	190
Discussion and conclusion.....	191
Acknowledgement .....	193
References.....	193





Advances in Complex Systems  
© World Scientific Publishing Company

## REJECTION MECHANISM IN 2D BOUNDED CONFIDENCE PROVIDES MORE CONFORMITY

HUET S.

*Cemagref, Laboratoire d'Ingénierie des Systèmes Complexes  
63172 Aubière, France\*  
sylvie.huet@cemagref.fr*

DEFFUANT G.

*Cemagref, Laboratoire d'Ingénierie des Systèmes Complexes  
63172 Aubière, France  
guillaume.deffuant@cemagref.cf*

JAGER W.

*Faculty of Economics and Business, University  
Groningen, The Netherlands  
w.jager@rug.nl*

This paper explores the dynamics of attitude change on 2 dimensions (2D) as a result of social interaction. We add a rejection mechanism into the 2D bounded confidence (BC) model proposed by Deffuant et al [19]. The principle is that one shifts away from a close attitude of one's interlocutor, when there is a strong disagreement on the other attitude. The model shows metastable clusters, which maintain themselves through opposite influences of competitor clusters. Our analysis and first experiments support the hypothesis that, for a large range of uncertainty values, the number of clusters grows linearly with the inverse of the uncertainty, whereas this growth is quadratic in the BC model.

### 1. Introduction

Much behaviour, especially in conditions of a higher involvement, can be understood as originating from underlying attitudes. One may vote on an extreme national party because of a negative attitude towards immigrants, and buy organic products because of a positive attitude towards environmentally friendly production. Hence attitudes motivate behaviour and exert selective effects at various stages of information processing [1]. Consequently, in studying behavioural change it is essential to have an understanding of the underlying attitudinal dynamics that give rise to such a change. Attitude is here understood in its psychological meaning as a tendency to evaluate a particular entity with some degree of favour or disfavour. The dynamics of attitudes are closely related to social influence, which includes individ-

\*Cemagref, Laboratoire d'Ingénierie des Systèmes Complexes, BP 50085, 63172 F-Aubière

2 *Huet S., Deffuant G., Jager W.*

ual influence on feelings, beliefs and behaviours of others [2]. These dynamics are studied, by experiments in laboratory on individuals and small groups, and are the subject of a variety of theories and assumptions. The most common assumption is a tendency of attitudes to get closer to already similar ones (attraction). A less usual assumption is a tendency to reject the other's attitude if it is psychologically uncomfortable (rejection).

Whereas an abundance of studies have been published in social psychology on the processes leading towards attitude change, relative little attention has been devoted to the interactions between multiple attitudes in social interactions. Yet, the issue of interactions between attitudes in a social interaction context seems to be highly relevant, as people often discuss different (unrelated) issues, and shifts on one attitude dimension may also have an impact on other dimensions as well. For example, if a friend, who is having similar attitudes on different issues, is speaking favourably about organic food, on which you have a negative attitude, the resulting dissonance may be resolved by either developing a more positive attitude on organic food as well, or by shifting away from the attitude position of the friend on the other attitude. In contrast, if a person you disagree with on many issues also advocates in favour of organic food, your attitudes are not likely to change as no dissonance is experienced. This paper aims to study, through computer simulations, how individuals with both these opposite tendencies (attraction and rejection in some conditions) produce different global patterns in a two dimensional attitude space. Our main result is that we observe fewer clusters than in the case of dynamics only based on attraction for a large range of uncertainty values. Before going through this result in more detail, we briefly present related research in social simulation and social psychology.

To begin, we consider the assumption of homophily. It assumes that people, especially if they are uncertain about their capacity and knowledge to evaluate a particular object, are more likely to adopt opinions and attitudes of similar others. For example, [3] shows that people like to have opinions similar to the ones of people they interact with. Similarity between receiver and source has a strong impact on the influence level of Word of Mouth [4]. Additionally, [5] suggest that homophily, facilitates the flow of information between people because of perceived ease of communication. Secondly, besides a perspective on what drives people's attitudes towards each other, some experiments and theories focus on the forces that may drive people's attitudes apart. At the individual level, the reactance theory [6], the balance theory [7], the motivation to protect oneself [8], and the social judgement theory [9] indicate that a persuasive intend can induce a rejection reaction: the behaviour, and/or the attitude changes in the direction opposite to the persuasion effort. In groups, the social identity theory [10], the self-categorization theory [11] and the optimal distinction theory [12] consider a capacity to differentiate from the individuals who are members of the same group by rejecting their opinions. This rejection is usually called the 'boomerang effect'. The conditions of its oc-

currence vary from one theory to another. Furthermore, social psychologists admit that the boomerang effect remains poorly understood [13]. The social judgement theory states that uncertainty plays an important role in both attitude attraction and rejection. The social identity theory stresses that attitude rejection is linked to the salience, at a given time, of the individual social identity. At the individual level, the theories link attitude rejection to loss of control or freedom, or a negative relation with others. From these theories, we retain that attitude rejection occurs when several attitudes are implicitly or explicitly activated. Moreover, it is favoured by a 'dissonant' situation, such as agreement on some attitudes and disagreement on others. As an example, [14] reports about students who, informed that their attitudes regarding a particular issue are close to the one of the Ku Klux Klan, decide to reinterpret this issue and to finally adopt an attitude further away from the one of the Ku Klux Klan.

Another group of interesting results for our purpose comes from the social influence paradigm which has exhibited two important group behaviours: the average consensus [15, 16] and the polarized consensus [17]. The average consensus occurs when the value of an object given by a group after discussion, is close to the average of the values given by individuals before discussion. The polarized consensus takes place when the value given by the group after discussion is significantly more extreme than the average of individual opinions before discussion. Following these studies, Nowak [18], in the social simulation domain, has recommended to investigate the tendency of individual attitudes to become more extreme (polarisation) as well as the tendency of individual to aggregate themselves in groups (clustering).

A large number of computer models are based on homophily. They postulate the existence of an attractive force between agents having close attitudes, which can be formulated using thresholds that determine when agents move towards each other's position [19-23], see [46] for an interesting review on opinion dynamics. This attraction threshold, also called uncertainty, can be fixed or dynamic [24, 25].

Other models, less numerous and more recent, also include a rejection mechanism in addition to assimilation. In formalising the Social Judgement Theory [26, 27, 35], an individual has two thresholds on an attitude dimension: a first for assimilation and a second one for rejection (the second is assumed higher than the first). In [28], based on the theory of self-categorisation and the meta-contrast principle, an individual tends to minimise the distance to a prototypical opinion which defines his own group and, at the same time, he maximises the distance to an external group. Moreover, a rejection effect appears in [29, 30] as an emerging effect of homophilic individual interactions. This effect is due to the fact that getting closer in the 2-dimensional attitude space may in some cases result in a shift away on the global attitude (which is a weighted aggregation of the attitudes).

Another form of rejection mechanism can be found in the 'contrarians' of Galam [38, 39] who decide, after having participated to the formation of a local consensus, to adopt the opposite behaviour to the majority. Quite close from the 'contrarians', the stochastic Sznajd model proposed in [40, 41] exhibit a similar dynamic without

considered a priori contrarians. They just take into account a social temperature which implies with a probability  $p$  the application of the appropriate Sznajd rule for the opinion choice of an agent, or the application of the opposite rule with the probability  $1 - p$ . Both of these works consider one-dimensional binary models and exhibit a particular final state, due to the "contrarians" effect, for which 50% of the population adopts one opinion, and the other 50%, the other opinion.

The attitude dynamic model we propose postulates multidimensional attitudes, like in [27, 29-34, 43-44]. Considering two dimensional attitudes with an equal importance, our main assumption is that, if you strongly disagree with someone on one of your two attitude-dimension, and are close on the other one attitude-dimension, you tend to solve the dissonance by shifting away on the closeness attitude-dimension. More precisely, when attitudes on both dimensions are far or close from each other, we follow the hypotheses of bounded confidence (BC) models [19, 21, 23-25] and [45] for a review: when both are close, the attitudes tend to get closer, when both are far apart, there is no influence. Two models are usually identified as bounded confidence models: the Deffuant-Weisbuch model [19] and the Hegselmann-Krause model [21]. These two models differ regarding their communication regime. Agents of the Hegselmann-Krause model adopt the average opinion of all agents which lie in her area of confidence. Agents of Deffuant-Weisbuch model meet in random pairwise encounters after which they comprise or not. For our model and for this first work, we retain the Deffuant-Weisbuch model. Therefore, our model is similar to a multi-dimensional bounded confidence model, except that we added the rejection mechanism when people are close on one attitude and far apart on the other.

The next part of this paper describes the model in a simplified version of the ODD framework [37] which is a standard protocol for describing individual and agent based models in three blocks (Overview, Design concepts, and Details). Following that, we present examples of simulation runs for different parameters, which lead to the hypothesis that the number of clusters grows linearly with the inverse of the uncertainty. Then, we show some results of more systematic exploration of the parameter space which support this hypothesis. Finally we will discuss the results and conclude.

## 2. Overview of the model

### 2.1. *Purpose of the model*

The purpose of the model is to test the collective effects of a particular rejection dynamics in 2-dimensional bounded confidence models which are based on individual attraction dynamics. The rejection takes place when individuals are close on one attitude and far on the other.

### 2.2. *State variables and scales*

We consider a population of  $N$  individuals, each having a 2-dimensional attitude or two different attitudes  $x_1$  and  $x_2$ , represented by real numbers between  $-1$  and  $+1$ ,

and the related uncertainties  $u_1$  and  $u_2$ . Uncertainty is a term used for convenience, because this variable may represent confidence in one's own attitude position as well as the motivation to comply with other's attitude positions (social susceptibility). It corresponds also to the latitude of acceptance of the Social Judgement Theory and represents the level of ego-involvement in the value of the attitude. In the following experiments all individuals have the same uncertainties  $U$  on both attitudes ( $u_1 = u_2 = U$ ).

### 2.3. Process overview and scheduling

At each time step, we choose a pair of individuals A and B at random, and they may influence each other. More precisely, at each time step, the algorithm is as follows:

N times repeat:

- choose couple of individuals (A,B) at random;
- A may influence B.

The influence depends on the conditions describing the values of attitudes and uncertainties. Suppose A has attitudes  $a_1$  and  $a_2$  with uncertainties  $u_1$  and  $u_2$ , and B has attitudes  $b_1$  and  $b_2$  with uncertainties  $u'_1$  and  $u'_2$ . Then, A compares its attitudes with the ones of B. Three cases arise.

2.3.1. *Case 1: B is close to A on both attitudes:*

$$|a_1 - b_1| \leq u_1 \text{ and } |a_2 - b_2| \leq u_2 \quad (1)$$

Then both attitudes of A get closer to the ones of B:

$$a_1^{t+1} = a_1^t + \mu (b_1^t - a_1^t) \quad (2)$$

Here  $\mu$  is a kinetic parameter of the model, representing the velocity of the attraction or the rejection. In our following study,  $\mu$  has the same value for all individuals.

2.3.2. *Case 2: B is far from A on both attitudes:*

$$|a_1 - b_1| > u_1 \text{ and } |a_2 - b_2| > u_2 \quad (3)$$

Then, there is no influence of B on A.

2.3.3. *Case 3: B is far from A on one attitude and close to A on the other.*

*Without loss of generality, we suppose:*

$$|a_1 - b_1| \leq u_1 \text{ and } |a_2 - b_2| > u_2 \quad (4)$$

Then two cases arise, depending on whether A and B differ strongly on attitude 2. We introduce the positive parameter  $\delta$ , ruling the intolerance threshold which globally depends on the uncertainty, i.e. on the ego-involvement level:

6 Huet S., Deffuant G., Jager W.

*Case 3.1: A and B do not differ strongly on attitude 2*

$$|a_2 - b_2| \leq (1 + \delta)u_2 \quad (5)$$

Then, the disagreement is not strong enough to trigger the rejection. A approaches B on attitude 1 and ignores B on attitude 2:

$$a_1^{t+1} = a_1^t + \mu (b_1^t - a_1^t) \quad (6)$$

*Case 3.2: A and B differ strongly on attitude 2*

$$|a_2 - b_2| > (1 + \delta)u_2 \quad (7)$$

Then, A shifts away from B on attitude 1. The movement is proportional to the distance needed to get  $b_1$  out of A's range of uncertainty around  $a_1$ .

$$a_1^{t+1} = a_1^t - \mu \text{psign}(b_1^t - a_1^t) (u_1 - |b_1^t - a_1^t|) \quad (8)$$

Where  $\text{psign}()$  is a particular sign function, which returns -1 if its argument is strictly negative, +1 otherwise. The particularity is that it returns +1 if the argument is 0. Moreover, we confine the attitude within the bounds (-1, +1) of the attitude space:

$$\text{If } |a_1| > 1 \text{ then } a_1 := \text{sign}(a_1) \quad (9)$$

The following figures illustrate the different types of interactions (attraction, rejection or indifference). Note that we suppose that uncertainties are the same on both dimensions and for all individuals. This means that we only get symmetrical interactions: if A attracts B, B attracts A; if A rejects B, B rejects A.

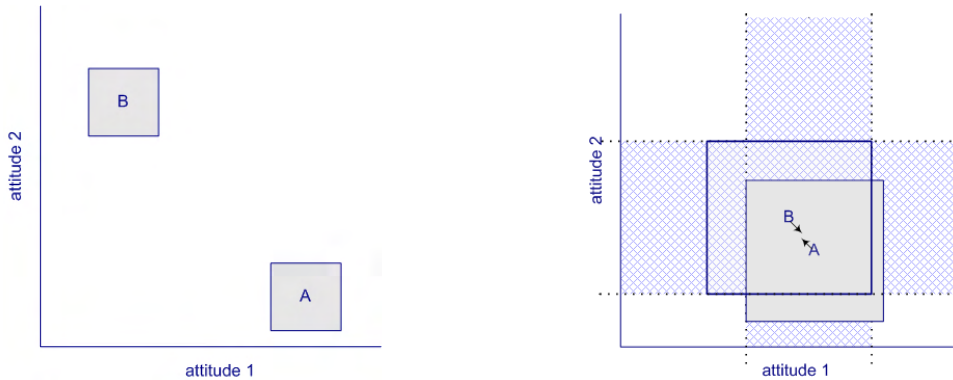


Fig. 1. A and B in a situation of no influence on both dimensions (left) and in situation of attraction (right)

Figure 1 shows on the left the case where A is not influenced by B: they are far from each other on both dimensions. On the right, figure 1 shows the case where A

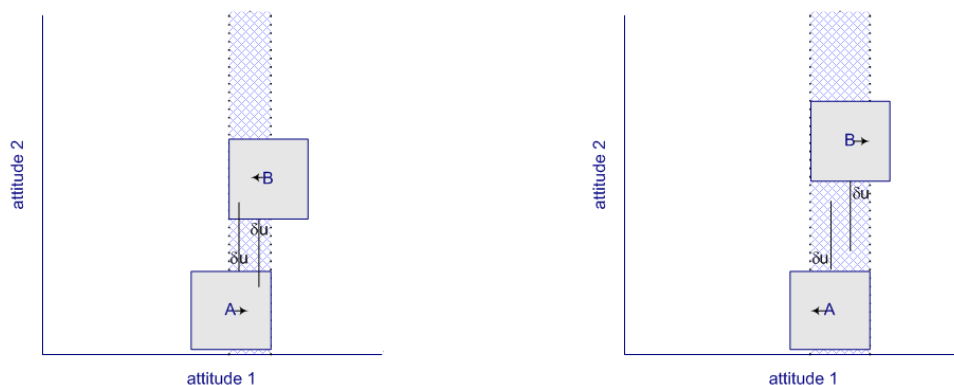


Fig. 2. Left:  $A$  and  $B$  in a situation of attraction on one dimension (on attitude 1 dimension here) and indifference on the other dimension. Right:  $A$  and  $B$  in a situation of rejection on one dimension (on attitude 1 dimension here) and indifference on the other dimension

is attracted by  $B$  and vice-versa because they are close to each other. This means each one has attitude in the other's acceptance zone.

Figure 2 left shows another case where people are close to each other on only one dimension. People are far from each other on one dimension but not far enough to consider the proximity on the other dimension as unacceptable. Thus, they assimilate each other on the dimension where they are close. On the contrary, figure 2 right shows the cases where people are far enough from each other on one dimension. The proximity on the other dimension is perceived as unacceptable. Thus, they move away from each other on this dimension.

#### 2.4. Initialisation

We consider a population of 1000 individuals with two attitudes. On each dimension, the attitude is randomly initialised following a uniform distribution comprised between  $-1$  and  $+1$ . Uncertainty  $U$  is constant and identical on each dimension. Identically, has the same value for all individuals.

### 3. Analysis of several examples

In this section, we observe several simulation examples, and this analysis leads to formulate the hypothesis that the number of clusters is a linear function of  $1/U$ .

#### 3.1. Evolution with uncertainty $U=0.2$ and intolerance threshold with $\delta=0$

Figure 3 shows an example of evolution for uncertainties  $U=0.2$ , and intolerance parameter  $\delta=0$ , and the kinetic parameter  $\mu=0.3$ . The number of time steps  $t$  appears on the top of each picture.



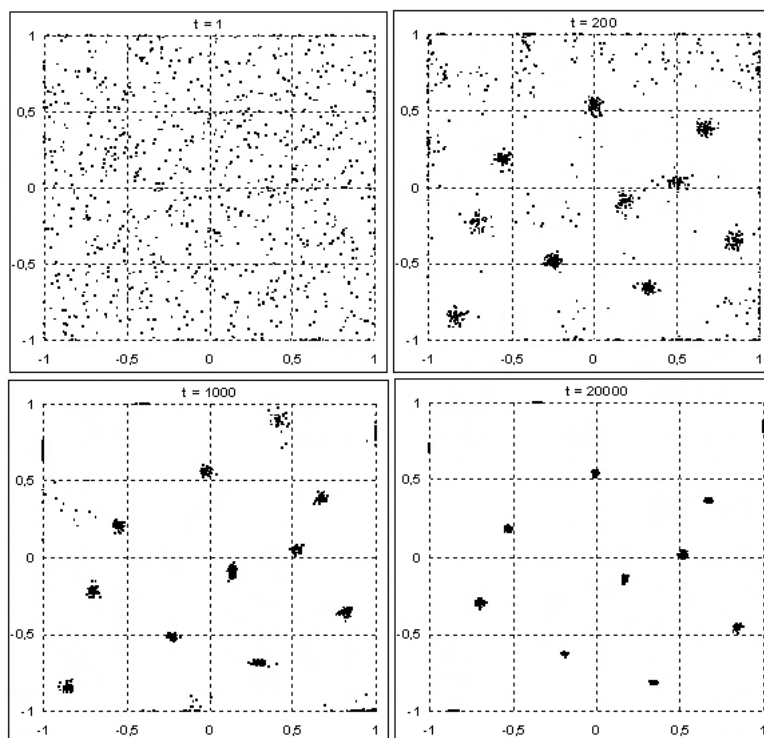


Fig. 3. Initial population uniformly distributed in 2D attitude space.  $U=0.2$ ,  $\mu=0.3$ ,  $\delta=0$ . We observe the emergence of metastable clusters, with remaining oscillations of individuals within the clusters. Moreover, some flat clusters are located on the borders of the attitude domain, containing radicalised individuals

### 3.2. Spatial organisation of the clusters and hypothesis of linearity of their number with $1/U$

The spatial organisation of the clusters can be further analysed. In this particular case where  $\delta = 0$ , we note that there is only one cluster on a horizontal or vertical line. Indeed, two clusters on the same horizontal or vertical line is an unstable situation. If the clusters are far, they tend to push each other from the line. If the clusters are close, they tend to merge. This can be checked by considering the histogram of presence of the individuals on each axis on figure 4. We note that 13 clusters appear on the projection of both axes. Moreover, the distance between the clusters is too small to prevent the rejection to play (11 clusters is the maximum, to provide a distance of at least  $U$  between two consecutive clusters), which explain why the individuals oscillate in the clusters.

In this case, the number of clusters can be analysed on a single axis: there should be a minimum interval between the clusters on each axis which is about the value of  $U$ . As we have seen, because of the metastability, it is possible to get slightly

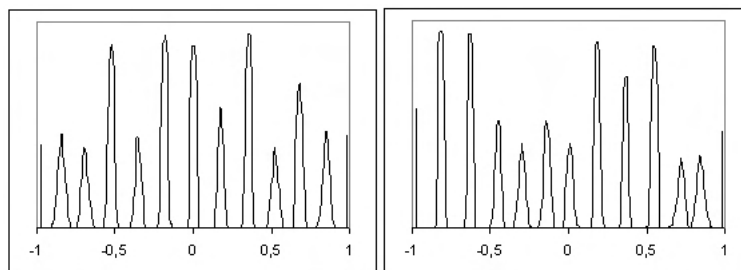


Fig. 4. Kernel density estimator on horizontal axis (left) and vertical axis (right), for the final situation of figure 3 ( $U=0.2$ ,  $t=20000$ ). One notes that the 13 final clusters are regularly distributed on each axis.

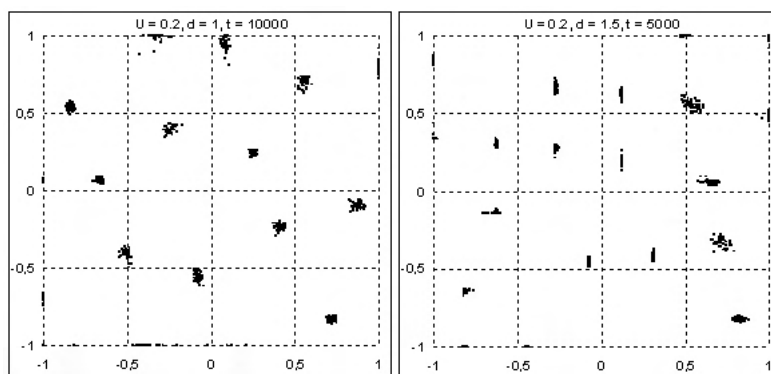


Fig. 5. Example of final configuration for  $U = 0.2$ ,  $\mu=0.3$ ,  $N = 1000$ ,  $\delta=1$  (left),  $\delta=1.5$  (right). It is possible to get 2 clusters on the same horizontal and vertical line, which is unstable when  $\delta=0$ . Moreover, for  $\delta=1.5$ , some clusters are flat inside the attitude domain.

smaller intervals. Nevertheless, one can expect a number of clusters varying linearly with  $1/U$ .

### 3.3. Influence of intolerance threshold $\delta > 0$

When the intolerance threshold gets higher, the conditions for rejection are more restricted: the disagreement on one attitude must be higher. Figure 5 shows two examples of final attractors, for  $U=0.2$ ,  $\delta=1$  (left) and  $\delta=1.5$  (right). The number of clusters appears to increase with  $\delta$ .

We observe that for these values of  $\delta$ , it becomes possible to get two clusters on the same horizontal or vertical line, when they are not too far apart (they remain in the tolerance zone). This explains why there are more clusters. Nevertheless, we can hypothesise that this number should still vary linearly with  $1/U$ , but with a higher coefficient.

Moreover, for  $\delta = 1.5$ , we observe flat clusters inside the attitude domain,

10 *Huet S., Deffuant G., Jager W.*

whereas this did not take place for  $\delta = 1$ . Such a flat cluster appears when all the neighbour clusters are on the same line in the tolerance zone, or far on both attitudes. The rejection interactions are therefore only in one direction.

### 3.4. Different values of uncertainty $U$ with intolerance threshold $\delta=0$

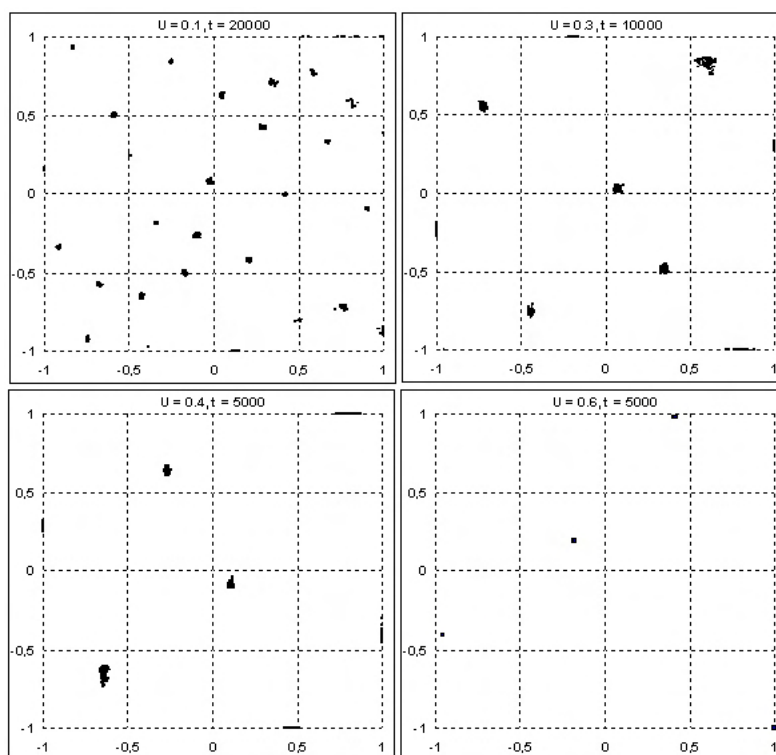


Fig. 6. Examples of attractor configurations for different values of uncertainty  $U$  and intolerance parameter  $\delta = 0$ ,  $\mu = 0.3$ . Population size  $N = 1000$ .

Figure 6 shows several attractor configurations for different values of uncertainty  $U$ . This first exploration suggests that the number of clusters decreases with  $U$ , like with the BC model. The observations made on our first simulation extend to these cases: Oscillations of individuals remain, with higher oscillations when  $U$  increases, and spatially organised to avoid two clusters on the same horizontal or vertical line. In each case, we get flat clusters with the maximum value for one attitude (polarisation).

For  $U = 0.6$ , we observe that the clusters become very concentrated, like in the simple BC model. The reason is that with 4 clusters, the intervals between

the clusters on a same horizontal or vertical line can easily be higher than  $U$ , and therefore avoid to generate a competition between the attraction in the cluster and the rejection from the neighbouring clusters.

#### 4. Systematic analysis of the number of clusters

We are interested in comparing the final number of attitude clusters with the one generated by the standard BC model proposed by [19]. First we describe how we compute the final number of clusters. Then, we analyse this final number of cluster regarding two different behaviour zones of the BC model (see [19, 47] for more details). The first zone is a zone for which the population is organized in several clusters; it is the object of our second point. The second zone is a zone for which the wide majority of people go in one cluster; it is the object of our third point.

##### 4.1. Computing the number of clusters

From the individual-based simulations, we collect the average, minimum and maximum final number of clusters. To compute the number of clusters, we define a minimum distance  $\epsilon$  between attitudes, below which we consider that they belong to the same cluster. We compute the clusters as groups of agents such that between any couple of agents of opinions  $x$  and  $x'$  in the group, there is a list of agents in the group of opinions  $(x_1, x_2, \dots, x_k)$  making a chain of couples distanced from each other of at most an Euclidian distance lower than  $\epsilon$ . The following pseudo-code can be used to compute the clusters; `necessaryToLookAt` is a table containing the identification number of each individual for all the population:

```
for all i of the population
  if necessaryToLookAt[i] > 0
    currentCluster.add(i)
    compt++;
  necessaryToLookAt[i] = 0
  while currentCluster.isNotEmpty()
    for all j of the population
      if necessaryToLookAt[j] > 0
        if distance(pop[currentCluster.get(0)],pop[j]<epsilon)
          necessaryToLookAt[j] = 0
          currentCluster.add(j)
          compt++
    currentCluster.remove(0)
  nbClusters++
if compt = populationSize then i = populationSize
```

In practice, we chose  $\epsilon=0.2U$  and we neglected the clusters of size lower or equal to 3 individuals. The simulations are stopped after 1,000,000 iterations. They can

12 *Huet S., Deffuant G., Jager W.*

be stopped before if the number of clusters has not changed after 100,000 iterations. Even if [47] have demonstrated the importance, for the 2D BC model, the presence of minor clusters in wide population and for high value of  $\mu$  [48], we do not look for them in our model in this first study.

#### 4.2. Final number of clusters on the 'multiple clusters zone' of the BC model

The BC model, in one dimension, yields a final number of clusters  $n_c$  in a population initialised with a uniform law on an attitude space of width  $2M$ , with all the same uncertainty  $U$ , which can be approximated by:

$$n_c \approx \frac{M}{U} \quad (10)$$

In the 2-dimensional case, when both attitude axes are adjusted independently and all have the same uncertainty  $U$  on both attitude dimensions, this rule is repeated on all lines of the space, therefore we get:

This result is confirmed by figure 7 which presents on abscissa  $1/U^2$  and on y-axis, the average number of clusters obtained on 30 replicas.

$$n_c \approx \left(\frac{M}{U}\right)^2 \quad (11)$$

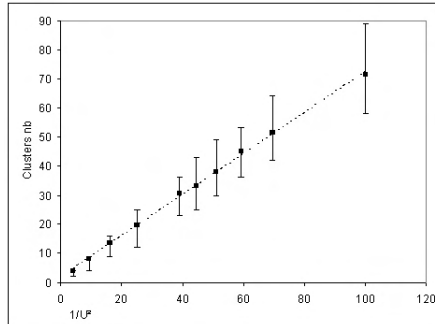


Fig. 7. Average final number of clusters of the 2D bounded confidence model as a function of  $1/U^2$ . Error bars indicates minimum and maximum obtained on 30 replicas.

We are interested in this point on the zone where the BC model exhibit a final state composed of several clusters. For our attitudinal domain, it goes from  $U = 0$  to  $U = 0.54$ . Figure 8 shows the number of clusters obtained with rejection dynamics, for different values of  $U$  and  $\delta$ . Results confirm the hypothesis of linearity of the number of clusters with  $1/U$  for  $\delta = 0$  and  $\delta = 0.5$  (left).

For  $\delta = 1, 1.5, 2$  and  $3$ , there is a non-linearity for  $U$  larger than 1 (only 1 and 2 are presented on the figure). When  $U$  is larger than 0.3, and  $\delta$  is large, the conditions

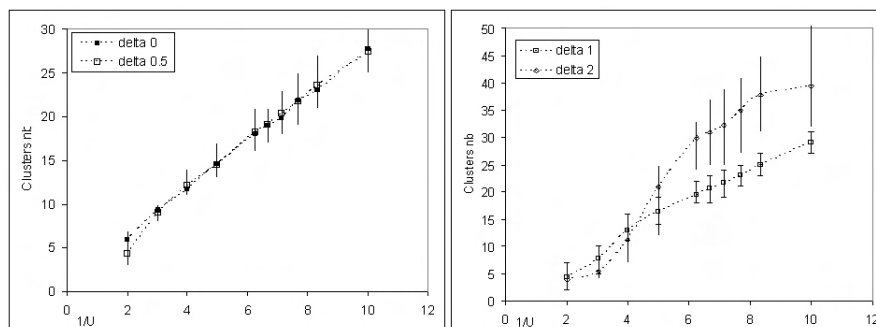


Fig. 8. Mean final number of clusters for the model with rejection as a function of  $1/U$ , for various values of  $\delta$ .  $N=1000$  and  $\delta=0.3$ . The error bars are the minimum and maximum numbers met in 30 replicas. On the left, for  $\delta=0$  and  $\delta=0.5$ , the number of clusters seems linear with  $1/U$ . On the right, the behaviour is not linear for large  $U$ .

for rejection are much constrained by the size of the domain: two individuals must be at both sides of the domain. Most of the interactions correspond therefore to the standard BC, and the curve is therefore quadratic. When  $U$  decreases ( $1/U$  grows), the rejection becomes more common and the curve becomes linear.

Let's now verify if this behaviour is efficiently robust to the population size. Figure 9 shows the number of clusters obtained with rejection dynamics for  $\delta = 0$  and different values of  $U$  and population size  $\delta$ . To be able to compare the different population size, we suppress the threshold of 3 individuals used to count the number of clusters. This means all clusters are counted, even if it is composed of only one individual. From figure 9, we see that the population size does not change the previous conclusion: the final number of clusters tends to be linear with  $1/U$ .

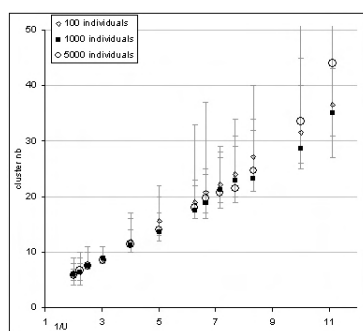


Fig. 9. Mean final number of clusters for the model with rejection as a function of  $1/U$ , for  $\delta = 0$  and various value of  $N$  (100, 1000, 5000) and  $U$  (from  $U=0.09$  to  $U=0.5$ ). The error bars are the minimum and maximum numbers met in 30 replicas. The number of clusters seems also linear with  $1/U$ .

14 Huet S., Deffuant G., Jager W.

#### 4.3. Final number of cluster on the "one major cluster" zone of the BC model

We are now interested on the zone where the BC model exhibit one major cluster. For our attitudinal domain, it begin for  $U > 0.54$ . Since for  $U > 1$ , the rejection mechanism can't work due to the bounded attitudinal domain forbidding that two people can reject each other because their attitudes can't be separated by an unbearable distance, we only study the  $U$  value range from 0.54 to 1. Indeed, for  $U > 1$ , all people go in one unique central cluster, exactly as in the BC model. Figure 10 shows the results for this zone at the same time we can appreciate the effect of the population size. We immediately see that the final number of clusters is not, for this zone, linear with  $1/U$ .

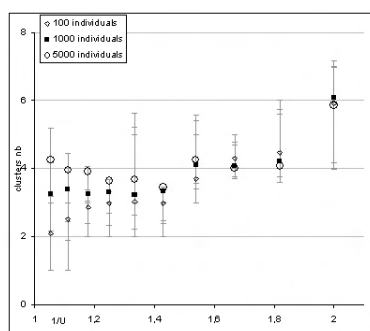


Fig. 10. Mean final number of clusters for the model with rejection as a function of  $1/U$ , for  $\delta = 0$  and various value of  $N$  (100, 1000, 5000) and  $U$  (from  $U=0.55$  to  $U=0.95$ ). The error bars are the minimum and maximum numbers met in 30 replicas. The number of clusters is not linear with  $1/U$ .

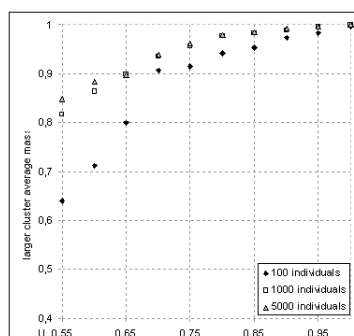


Fig. 11. Average mass of the final larger cluster for  $\delta = 0$  and various value of  $N$  (100, 1000, 5000) and  $U$  (from  $U=0.55$  to  $U=0.95$ ).

In the standard BC model proposed by [19], the final state for this zone is one major cluster containing a large majority of people with eventually a lot of very minor clusters when the population is very large [46]. Figure 11 shows that our model has also, on the zone of  $U$  values, one major cluster containing a majority of people, from a part of 0.5 to 1 of the population depending on the parameter value. In our model with a rejection mechanism, we finally obtained, on average on replications, between two and six final clusters as shown on figure 10. Are the non-major clusters the same than those of the BC model. From [48], we know the very small clusters of the standard BC model are very numerous and do not exist for low values of  $\mu$ . In our attraction-rejection model, minor clusters are not numerous, from one to five on average, and larger than those of the classical BC model. Moreover, they remain when we run simulation with of value of  $\mu$  equal to 0.01. Finally, our population size is not enough wide to really observe if 'minor clusters' in the sense used by [47] exist in our model.

## 5. Discussion and conclusion

In the model of 2-dimensional attitude dynamics we propose an agent shifts away from a close attitude on one axis when the interlocutor is far on the other axis. We assume that this is a way to solve the dissonance between the attitude axes. The distance threshold to trigger rejection depends on the intolerance parameter  $\delta$  and on the uncertainty  $U$ , which may define a non-commitment zone, in which the dissonance is tolerated. When the conditions of rejection are not met, that is when we exclude the case 3.2. for which two individuals differ strongly on one attitude and are similar on the other attitude, the model behaves exactly like the 2D bounded confidence (BC) model.

The first explorations of this model, in the simple case where all uncertainties are the same, show several striking results, in comparison with the 2D BC model:

- Since the uncertainty is not higher than 1 allowing the rejection to occur, the dynamics leads to several clusters, which are generally in competition and tend to reject each other. Finally, the system reaches a metastable state: the stability is due to contradictory rejections from neighbouring clusters, which compensate each other. If one of its neighbouring clusters is removed, the position of a cluster changes significantly, and it may even disappear. Moreover, individuals belonging to a cluster are in constant movement, with amplitudes depending on the cluster size and on the proximity of competing clusters. In this respect, the configuration is very different from the one obtained with simple BC model where, after a while, clusters keep concentrating with time, each independently from the other.
- Several clusters are moving towards the limits of the attitude domain. This may be interpreted as a radicalisation of a part of the population, which reaches the maximum absolute value of one of the attitudes. This never happens with the 2D BC model.



- In the case where the intolerance threshold  $\delta = 0$ , two clusters cannot be maintained on the same horizontal or vertical line. Therefore, the clusters tend to occupy points of the space where they are as far as possible from other clusters on each axis. This analysis suggests a number of clusters growing linearly with  $1/U$  for values of  $\text{textitU}$  for which the 2D BC model exhibits several clusters called 'major' and 'central' clusters by [47]. However, for the 2D BC model, for this same range of  $\text{textitU}$  values, the cluster number grows quadratically with  $1/U$  in the 2D BC model. When  $\delta$  grows, configurations with more than one cluster on a line may be stabilised, but this number is limited by the size of the tolerance zone. Therefore, the growth of the cluster number should still be linear, but with a factor growing with  $\delta$ . First systematic experiments support this statement. This behavior is the same for various population sizes.
- For values of  $\text{textitU}$  for which the 2D BC model exhibits only one cluster called 'central' clusters by [47], our model do not follow the same law and tends to have less consensus than the 2D BC model depending on the population size. Indeed, depending on the parameter value, it exhibits from two to six clusters on average with one larger containing a majority of people. The other clusters are generally on the limits of the attitude domain. [47] and [48] show that the 2D BC model has, for a subpart of this zone of  $\text{textitU}$  values, numerous very minor clusters when  $\mu$  is high and when the population size is wide. However these very minor clusters, even if they are located close to the bound of the attitudinal space, are different from the minor clusters of extremists of our model. These results suggest several points to discuss.

The metastability of the clusters is due to the bounds we impose on the attitude values. Indeed, without these bounds, the attitudes grow until the distance between the clusters is higher than the uncertainty in all directions. Then, the clusters do not influence each other, and they keep concentrating as in the BC model. First simulations realised on the same model applied on an unbounded attitude domain indicates us that the final number of clusters is close to the one obtained with the bounded one. However, the unbounded case has to be the object of a particular study. In any case, the metastability of the clusters is an interesting feature of this model, which better fits real group dynamics than the perfect similarity obtained without a bound (or by a standard BC model).

Even without bounds, we obtain a global result which shows strong similarities with social identity and self-categorization theories. Our individuals tend to minimise their in-group distance and maximise their out-group distance (to competing groups). We also get some polarized groups (which have more extreme opinions than all the individuals initially). This reminds of the results Moscovici and Zavalloni [17] obtained. Therefore, with a model considering only paired interactions, we get group dynamics which seem to make sense in a social psychology perspective.

However, the model remains very simplified, and a challenge that remains is checking if these interesting properties last when adding more sophisticated hypotheses. In particular, in our model, all attitudes are considered to have the same weight on the behaviour, whereas one expects that only disagreements on attitudes deeply related to social identity can lead to rejection. To take this aspect into account, we thus should consider attitudes different of different types.

For our model, we have chosen the particular communication regime of the Deffuant-Weisbuch model. Considering the communication regime as a parameter of the bounded confidence model as in the formulation proposed by [42]. Moreover, it would be a logical extension to relate the chance of interacting to the attitude similarity between the agents, thus reflecting principles of preferential attachment.

In future research, we plan to continue to explore the properties of this model. In particular, we suspect interesting effects of the population size on the number of clusters. Furthermore, introducing extremists like in [24] could produce unexpected effects.

## 6. Acknowledgement

We thanks the anonymous reviewers for their invaluable recommendations and their relevant remarks.

## References

- [1] Eagly, A.H. and S. Chaiken, *The Psychology of Attitudes* (Fort Worth, TX: Harcourt Brace Jovanovich, Thomson/Wadsworth ed. 1993).
- [2] Mason, W.A., F.R. Conrey, and E.R. Smith, Situating Social Influence Processes: Dynamic, Multidimensional Flows of Influence Within Social Networks. *Personality and Social Psychology Review*, 2007(11): pp. 279-300.
- [3] Festinger, L., *A Theory of Cognitive Dissonance* (Stanford, CA: Stanford University Press ed., 1957).
- [4] Moschis, G.P., Social Comparison and Informal Group Influence. *Journal of Marketing Research*, 1976. 13(3): pp. 237-244.
- [5] Price, L.L. and L.F. Feick, The role of interpersonal sources in external search: an informational perspective. *Advances in Consumer Research*, 1984(11): pp. 250-253.
- [6] Brehm, J.W., *A theory of psychological reactance* (New York: Academic Press, 1966).
- [7] Heider, F., Attitudes and Cognitive Organization. *Journal of Psychology*, 1946(21): pp. 107-112.
- [8] Rogers, R.W. and S. Prentice-Dunn, *Protection motivation theory*, in Handbook of health behavior research, D. Gochman, Editor (New York: Plenum, 1997) pp. 113-132.
- [9] Sherif, M. and C.I. Hovland, *Social Judgement* (New Haven, CT: Yale University Press, 1961).
- [10] Tajfel, H., *Differentiation between social groups: Studies in the social psychology of intergroup relations* (London: Academic Press, 1978).
- [11] Turner, J.C., et al., *Rediscovering the social groups: A self-categorization theory* (Cambridge, MA: Basil Blackwell, 1987).
- [12] Brewer, M.B., The Social Self: On Being the Same and Different at the Same Time. *Pers. Soc. Psychol. Bulletin*, 1991. 17(5): pp. 475-482.

18 Huet S., Deffuant G., Jager W.

- [13] Eagly, A.H. and S. Chaiken, *Process Theories of Attitude Formation and Change: Attribution Approaches and Social Judgment Theory*, in The Psychology of Attitudes, A.H.C. Eagly, S., Editor (Fort Worth, TX: Harcourt Brace Jovanovich, Thomson/Wadsworth, 1993) pp. 351-388.
- [14] Wood, W., et al., Self-definition, defensive processing, and influence: the normative impact of majority and minority groups. *Journal of Personality and Social Psychology*, 1996(71): pp. 1181-1193.
- [15] Allport, G.W., *Social Psychology* (Boston: Houghton Mifflin, 1924).
- [16] Sherif, M., A study of some social factors in perception. *Archives of Psychology*, 1935. 27(187).
- [17] Moscovici, S. and M. Zavalloni, The Group as a Polarizer Of Attitudes. *Journal of Personality and Social Psychology*, 1969. 12(2): pp. 125-135.
- [18] Nowak, A. and R.R. Vallacher, *Dynamical Social Psychology* (The Guilford Press, 72 Spring Street, New York, NY 10012. 318, 1998).
- [19] Deffuant, G., et al., Mixing beliefs among interacting agents. *Advances in Complex Systems*, 2001(3): pp. 87-98.
- [20] Galam, S. and S. Moscovici, Towards a theory of collective phenomena: consensus and attitude changes in groups. *J. Soc. Psychology*, 1991(21): pp. 49-74.
- [21] Hegselmann, R. and U. Krause, Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation. *Journal of Artificial Societies and Social Simulation*, 2002. 5(3).
- [22] Latané, B., Dynamic Social Impact: The Creation of Culture by Communication. *Journal of Communication*, 1996. 46(4): pp. 13-25.
- [23] Weisbuch, G., et al., Meet, discuss and segregate ! *Complexity*, 2002. 7(3): pp. 55-63.
- [24] Deffuant, G., et al., How can extremism prevail ? A study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation*, 2002. 5(4).
- [25] Weisbuch, G., G. Deffuant, and F. Amblard, Persuasion dynamics. *Physica A*, 2005(353): pp. 555-575.
- [26] Jager, W. and F. Amblard, A dynamical perspective on attitude change. *NAACSOS Conference 2004*. June 27-29, 2004, Pittsburgh PA <http://www.casos.cs.cmu.edu/naacsos/>.
- [27] Jager, W. and F. Amblard, Multiple Attitude Dynamics in Large Populations. *Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms*. 2005: co-sponsored by Argonne National Laboratory and The University of Chicago.
- [28] Salzarulo, L., A continuous Opinion Dynamics Model Based on the Principle of Meta-Contrast. *Journal of Artificial Societies and Social Simulation*, 2006. 9(1).
- [29] Urbig, D. and R. Malitz, Dynamics of structured attitudes and opinions. *Third Conference of the European Social Simulation Association*. 2005. September 5-8, Koblenz, Germany.
- [30] Urbig, D. and R. Malitz, Drifting to more extreme but balanced attitudes: Multidimensional attitudes and selective exposure. in *4th Conference of the European Social Simulation Association*. 2007. Toulouse, 10-14 September.
- [31] Axelrod, R., The Dissemination of Culture - A model with local convergence and global polarization. *Journal of Conflict Resolution*, 1997. 41(2): pp. 203-226.
- [32] Deffuant, G. and S. Huet, Propagation effects of filtering incongruent information. *Journal of Business Research*, 2007. 60(8): pp. 816-825.
- [33] Fortunato, S., et al., Vector Opinion Dynamics in a Bounded Confidence Consensus Model. *International Journal of Modern Physics*, 2005. 16(10): pp. 1535-1551
- [34] Huet, S. and G. Deffuant, When do Interactions Increase or Decrease Primacy Effect?

- in *Model-2-Model*. 2007. Marseille.
- [35] Jager, W. and F. Amblard, Uniformity, Bipolarization and Pluriformity Captured as Generic Stylized Behavior with an Agent-Based Simulation Model of Attitude Change. *Computational and Mathematical Organization Theory*, 2004(10): pp. 295-303.
  - [36] Lorenz, J. and D. Urbig, About the power to Enforce and Prevent Consensus by Manipulating Communication Rules. *Advances in Complex Systems*, 2007. 10(2): pp. 251-269.
  - [37] Grimm, V., et al., A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 2006(198): pp. 115-126.
  - [38] S. Galam, Contrarian deterministic effects on opinion dynamics: 'the hung elections scenario'. *Physica A*, 2004, 333: pp. 453-460.
  - [39] C. Borghesi and S. Galam, Chaotic, staggered and polarized dynamics in opinion forming: the contrarian effect. *Physical Review E*, 2006, 73: 9 pages.
  - [40] M. S. de La Lama, J. M. Lopez and H.S. Wio, Spontaneous emergence of contrarian-like behaviour in an opinion spreading model. *Europhys. Lett.*, 2005, 72: pp. 851-7.
  - [41] H. S. Wio, M. S. de la Lama and J. M. Lopez, Contrarian-like behavior and system size stochastic resonance in an opinion spreading model. *Physica A*, 2006, 371: pp. 108-11.
  - [42] D. Urbig and J. Lorenz, Communication regimes in opinion dynamics: Changing the number of communicating agents, Second Conference of the European Social Simulation Association (ESSA), (Valladolid, Spain (September 16-19): 2004).
  - [43] J. Lorenz, Continuous opinion dynamics of multidimensional allocation problems under bounded confidence: More dimensions lead to better chances for consensus, *European Journal of Economic and Social Systems*, EJESS, 2006(19): pp. 213-27.
  - [44] J. Lorenz, Fostering Consensus in Multidimensional Continuous Opinion Dynamics under bounded Confidence, in R.P. Würtz, ed., *Organic computing*, (Budapest Springer Berlin / Heidelberg), 2008: 321-34.
  - [45] J. Lorenz, Continuous Opinion Dynamics Under Bounded Confidence: A Survey, *International Journal of Modern Physics C*, 2007(18): pp. 1-20.
  - [46] C. Castellano, S. Fortunato and V. Loreto, Statistical physics of social dynamics, (2007), <http://arxiv.org/abs/0710.3256>
  - [47] E Ben-Naim, P. L. Krapivsky and S. Redner, Bifurcations and Patterns in Compromise Processes, *Physica D: Non linear phenomena*, 2003(184): pp. 190-204.
  - [48] M.F. Laguna, G. Abramson and D. H. Zanette, Minorities in a Model for Opinion Formation, *Complexity*, 2004(9): pp. 31-6.



## Chapter 2.4 Attraction-rejection – a double modelling approach

Title: Bounded Confidence with Rejection: Clusters or Scattered Opinions?

Authors: Sylvie Huet, Guillaume Deffuant

Accepted and presented at the European Association of Social Simulation Conference, 2008. Brescia, Italy, September 1-5

### Abstract

Huet and Deffuant (2007) propose a new opinion dynamics model based on the bounded confidence principles, with a rejection mechanism. We study new simulations of the agent-based model for a population of 10000 individuals. We generally observe fewer clusters than the classical bounded confidence model, and below a threshold of the uncertainty, no cluster appear: all opinions remain scattered. We build an aggregated model of this agent-based model (ABM), in the limit case of an infinite population, in order to better understand this dynamical behaviour. When adding a small perturbation in the initial distribution, the aggregated model and ABM yield similar numbers of clusters for the same parameters. Below a critical value of uncertainty, the aggregated model forms no cluster and the distribution remains uniform. The critical value is higher for the aggregated model than for the ABM, indicating a finite size effect that we discuss.

### Contents

Number of clusters in the ABM when the uncertainty varies .....	200
The ABM .....	200
Experiments: methods and results .....	201
Aggregate model at the limit of infinite population .....	202
The aggregate model.....	202
Comparison between the aggregate and the AB models .....	203
Discussion and conclusion.....	205
References.....	207



Wood et al in (Wood, Pool et al. 1996) report the following experiment: First, students give their opinion about a particular issue. Then they are told that this opinion is similar to the one of people they disagree strongly with on other issues (Ku Klux Klan members). Then, generally, the students change their opinion on the initial issue, to differentiate from Ku Klux Klan members. In this paper, we propose and study a simple model of this dynamics of attitudes.

Several experiments and theories focus on the forces that may drive people's attitudes apart. At the individual level, the reactance theory [11], the balance theory [12], the motivation to protect oneself [13], and the social judgement theory [14] indicate that a persuasive effort can induce a rejection reaction: the behaviour, and/or the attitude changes in the direction opposite to the persuasion effort. In groups, the social identity theory [15], the self-categorization theory [16] and the optimal distinction theory [17] consider a capacity to differentiate from the individuals who are members of the same group by rejecting their opinions. This rejection is usually called the 'boomerang effect'. The conditions of its occurrence vary from one theory to another. Furthermore, some social psychologists admit that the boomerang effect remains poorly understood [18]. The social judgement theory states that uncertainty plays an important role in both attitude attraction and rejection. The social identity theory stresses that attitude rejection is linked to the salience, at a given time, of the individual social identity. At the individual level, the theories link attitude rejection to loss of control or freedom, or a negative relation with others. From these theories, we retain that attitude rejection occurs when several attitudes are implicitly or explicitly activated. Moreover, it is favoured by a 'dissonant' situation, such as agreement on some attitudes and disagreement on others.

The study of the link between individual interaction processes and opinions or attitude of a whole population can only be done by a virtual simulation approach. Classically, opinion dynamic models do not consider a "boomerang" effect. Very few recent models include a rejection mechanism in addition to an assimilation or attraction process between people. Firstly, a form of rejection mechanism can be found in the "contrarians" of Galam [19, 20] who tend to adopt an attitude which is opposite to the one of the majority (attitudes are supposed binary). The stochastic Sznajd model [21, 22] also includes individuals who oppose to the majority following a given probability. Both of these models consider one-dimensional binary attitudes and tend to a particular final state, due to the "contrarians" effect, for which 50% of the population adopts one opinion, and the other 50%, the other opinion. Secondly, other recent works consider multi-dimensional continuous attitudes. In formalising the Social Judgement Theory [23, 24, 25], an individual has two thresholds on an attitude dimension: a first for assimilation and a second one for rejection (the second is assumed higher than the first). In [26], based on the theory of self-categorisation and the meta-contrast principle, an individual tends to minimise the distance to a prototypical opinion which defines his own group and, at the same time, he maximises the distance to an external group. Moreover, a rejection appears in [27, 28] as an emerging effect of homophilic individual interactions. This effect is due to the fact that getting closer in the 2-dimensional attitude space may, in some cases, result in a shift away on the global attitude (which is a weighted sum of the attitudes). None of these works base the rejection mechanism on an unacceptable dissonance feeling.



It is the particular focus of this work. It aims at better understanding the impact of a rejection mechanism, added to attraction, on the organisation in groups of a whole population. Inspired from the dissonance theory (Festinger 1957), our model of opinion dynamics presented in (Huet, Deffuant et al. 2007) is based on the bounded confidence (BC) model (Deffuant, Neau et al. 2001). Considering two dimensional attitudes<sup>18</sup> (concerning object 1 and object 2) with an equal importance, our main assumption is that, when two individuals have very different attitudes about object 1, and have close attitudes about object 2, they tend to solve the dissonance by shifting away attitudes about object 2. Thus, we add the rejection mechanism when people have close attitudes about one object and far attitudes about the other object. In [8], a rejection rule is also added to the BC model, but it is different: It is assumed that the rejection takes place when the opinions are too far apart. The rejection is still poorly understood by social psychologists, and we believe that there is room for testing different hypotheses.

First simulations of the agent-based model showed that, for a large interval of the uncertainty, the rejection mechanism leads to fewer clusters than with the standard bounded confidence: the number of clusters is linear with  $1/u$  for the model with rejection, whereas it is linear with  $1/u^2$  for standard BC [2]. In this paper, we investigate with more care the behaviour of the model for small values of  $u$ , and a larger population. We observe that, below a critical value of the uncertainty, the opinions remain scattered instead of organizing clusters.

In order to better understand this observation, we developed an aggregated model of the individual-based model. We follow the "double-modelling" approach: developing an analytical model of an ABM, in order to provide explanations of the collective effects observed in individual-based model simulations, through an aggregated view of the individual-based model behaviour (Deffuant 2004). We build the aggregated model in the limit case of infinite populations (similarly to [9]), and we write the differential equations ruling the evolution of the probability density of opinions. Practically, we have to discretise the opinion space to solve numerically these equations. The results help to understand why the behaviour of the models is different for low values of uncertainty.

The first part of the paper describes the ABM, and its dynamical behaviour for different values of uncertainty and a population of 10000 individuals. A second part presents the aggregated approximation of the ABM and compares it with the ABM. A third part discusses this comparison and concludes.

## **Number of clusters in the ABM when the uncertainty varies**

### **The ABM**

The ABM considers a population of agents with bidimensionnal attitudes (or opinions), supposed initially uniformly distributed. To each attitude is associated an uncertainty, which

---

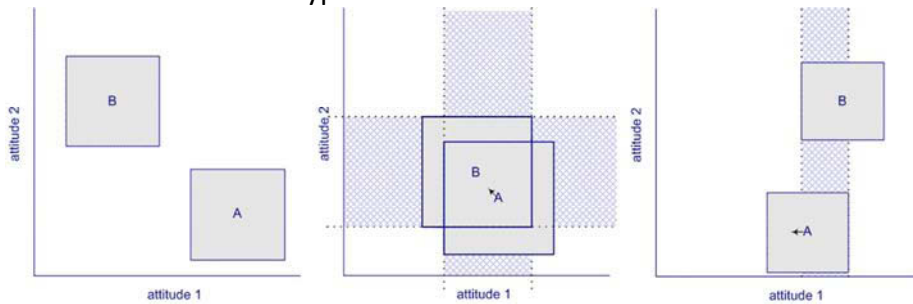
<sup>18</sup> An attitude toward an object is defined here as a psychological assessment of this object Fishbein, M. and I. Ajzen (1975). Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research, Reading MA: Addison-Wesley.

is supposed constant  $u$  in this study, for sake of simplicity. We suppose that agents meet by randomly chosen pairs. Suppose that individual  $A$  of attitudes  $a_1$  and  $a_2$ , meets with individual  $B$  of attitudes  $b_1$  and  $b_2$ . Let  $(\delta a_1, \delta a_2)$  be the bidimensionnal vector of the changes of attitudes of  $A$ , because of  $B$ 's influence (i.e, after the meeting  $a_1$  becomes  $a_1 + \delta a_1$ ,  $a_2$  becomes  $a_2 + \delta a_2$ ). This influence can be tuned with parameter  $\mu$ .

Three cases occur:

1. If  $a_1$  is close to  $b_1$  and  $a_2$  is close to  $b_2$ :  $|a_1 - b_1| \leq u$  and  $|a_2 - b_2| \leq u$ , then, the rules of the BC model apply on both dimensions, and  $A$ 's attitudes are moved towards  $B$ 's on both dimensions (attraction effect):  $\delta a_1 = \mu(b_1 - a_1)$ ,  $\delta a_2 = \mu(b_2 - a_2)$ .
2. If  $a_1$  is far from  $b_1$  and  $a_2$  is far from  $b_2$ :  $|a_1 - b_1| > u$  and  $|a_2 - b_2| > u$ , then, the rules of the BC model also apply on both dimensions, and there is no influence:  $\delta a_1 = 0$ ,  $\delta a_2 = 0$
3. If the attitudes are close on one dimension (suppose:  $|a_1 - b_1| \leq u$ ) and far on the other ( $|a_2 - b_2| > u$ ), then  $a_1$  moves away from  $b_1$ . The movement is the highest when these opinions are equal, and tends to zero linearly, when their difference approaches  $u$ .  
 If  $a_1 - b_1 < 0$  then:  $\delta a_1 = -\mu(u - (b_1 - a_1))$ ,  $\delta a_2 = 0$ .  
 If  $a_1 - b_1 \geq 0$  then:  $\delta a_1 = \mu(u + (b_1 - a_1))$ ,  $\delta a_2 = 0$ .

Fig. 1 illustrates the different types of interactions.



**Fig. 1.** Situations of interaction. Left: no influence. Middle: B attracts A. Right: B rejects A on attitude 1 and does not influence A on its attitude 2. The centres of squares  $A$  and  $B$  are opinion vectors  $(a_1, a_2)$  and  $(b_1, b_2)$ . The size of the squares is the uncertainty  $u$ . The arrow shows the direction of change of  $A$ .

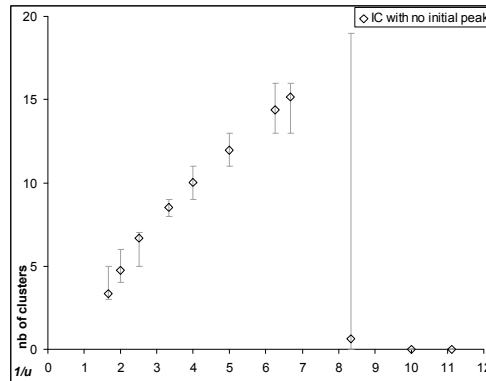
In [2], we suppose in addition that the attitudes are confined within the initial bounds of the distribution of attitudes (when the rejection pushes an attitude outside the bounds, we bring it back on the boundary). In these conditions, we observed that this model converges towards a set of metastable clusters. The number of these clusters is linear with  $1/u$ , whereas it linear with  $(1/u)^2$  in the standard 2-dimensional bounded confidence model.

In [10], we studied a symmetrical aggregate version of this ABM, in the limit of infinite population. We observed that, for low uncertainty, this model does not create clusters, and keeps its initial uniform state. This result was surprising, because in our first experiments on the ABM (with a population size of 1000), we always got clusters, even for low uncertainty. In this paper we perform new experiments on the ABM, with a larger population.

## Experiments: methods and results

We consider a population of 10000 individuals with two attitudes. On each dimension, the attitude is randomly initialised following a uniform distribution comprised between -1 and +1. Uncertainty  $u$  is constant for all individuals and identical on each dimension.  $\mu$  is equal to 0.5. Simulations run during 500000 iterations. Each iteration represents 10000 random pair interactions, hence in one iteration, each individual is picked once on average.

We use a particular procedure to count the clusters obtained in the final state of the model in order to compare our ABM more easily with the aggregate model in the following sections. We discretise the bi-dimensional attitude space into a grid of 100 X 100 boxes. The proportion of agents in each box measures the local density. We identify a cluster as a set of such boxes of a density higher than 0.005, and connected on the grid. We chose this value of 0.005 after 100 tests on randomly uniformly distributed opinions, showing that the maximum of the boxes was always lower than 0.005. Therefore, this value is adequate to discriminate noisy uniform distributions from clusters.



**Fig. 2.** Number of clusters in the final state of the ABM for various values of  $1/u$  (the values of  $u$  are 0.09, 0.1, 0.12, 0.15, 0.16, 0.2, 0.3, 0.4, 0.5, 0.6), and a population of 10000 individuals. The abscissa is  $1/u$ . The points represent the average number of clusters over 30 replicas, and the error bars indicate the minimum and the maximum.

Fig. 2 shows that the behaviour of the model changes radically from  $1/u$  equals 8: beyond this value, we obtain no clusters. For  $u$  equals 0.12, only one replica out of 30 yields 18 clusters, for all the others, all opinions remain scattered (no cluster).

Now, we consider an infinite aggregated model of this ABM in order to get a better understanding of this dynamical behaviour transition.

## Aggregate model at the limit of infinite population

### The aggregate model

As in [10], we consider the limit case of an infinite population, with an initially perfectly uniform distribution of attitudes on  $[-1,1] \times [-1,1]$ . To solve numerically the aggregated model, we discretise the compact  $[-1,1] \times [-1,1]$ , in a regular grid of size  $m \times m$  (typically, we take 100 x 100). On each point  $g(i,j)=(-1+i/m,-1+j/m)$  of the grid,  $\rho(i,j)$  represents the probability that an agent of the population has its attitudes inside a square of center  $(i,j)$ , and of size  $1/m$ . Moreover, we discretise the boundary of the domain, with values of  $i$  and  $j$  equal to 0 or  $m+1$ .

Distribution  $\rho$  is initially perfectly uniform and is null on the boundary, thus:

$$\begin{aligned} \rho(i, j) &= \frac{1}{m^2}, (i, j) \in \{1, \dots, m\} \times \{1, \dots, m\} \\ \rho(i, j) &= 0 \text{ if } i = 0 \text{ or } i = m + 1 \text{ or } j = 0 \text{ or } j = m + 1 \end{aligned} \quad (1)$$

The principle of the model dynamics is to compute the flows of distribution from one site  $(i, j)$  to any other site  $(k, l)$ , and to sum them up to compute the distribution change (all sites are updated at the same time).

More precisely, for point of the grid  $g(i, j)$ , we consider all the points of the grid  $g(k, l)$ , and we compute the influence of  $g(k, l)$  on  $g(i, j)$ . Let  $(\delta\alpha_1, \delta\alpha_2)$  be the change of attitude on dimension 1 and 2, computed with the rules presented above. Let  $\lfloor a \rfloor$  be the integer part of number  $a$ , we define:

$$\delta i = \left\lfloor \frac{\delta\alpha_1}{2} m \right\rfloor, \text{ and } \delta j = \left\lfloor \frac{\delta\alpha_2}{2} m \right\rfloor \quad (2)$$

The probability that agents from site  $(i, j)$  encounter agents from site  $(k, l)$  is proportional to the product  $\rho(i, j)\rho(k, l)$ . Therefore, the global change of the distribution  $d\rho$ , due to the systematic encounters between all pairs of sites is computed as follows:

Computation of  $d\rho$

```

For  $(i, j) \in \{0, \dots, m+1\} \times \{0, \dots, m+1\}$  do:
  For  $(k, l) \in \{0, \dots, m+1\} \times \{0, \dots, m+1\}$  do:
    If  $\delta i = \left\lfloor \frac{\delta\alpha_1}{2M} m \right\rfloor \neq 0$  or  $\delta j = \left\lfloor \frac{\delta\alpha_2}{2M} m \right\rfloor \neq 0$ 
       $d\rho(i + \delta i, j + \delta j) := d\rho(i + \delta i, j + \delta j) + \rho(i, j)\rho(k, l)$ 
       $d\rho(i, j) := d\rho(i, j) - \rho(i, j)\rho(k, l)$ 
    end if
  end for
end for

```

Then, the global evolution of the probability density  $\rho$  is simulated numerically, as follows.

Repeat:

  Compute  $d\rho$

  For  $(i, j) \in \{1, \dots, m\} \times \{1, \dots, m\}$  do:  $\rho(i, j) := \rho(i, j) + d\rho(i, j)$

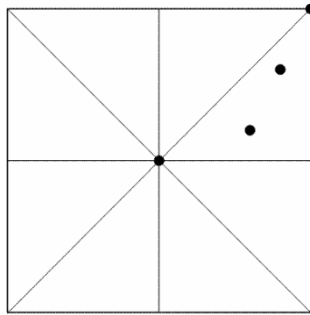
  Reset  $d\rho$  to 0.

## Comparison between the aggregate and the AB models

### *Increasing the initial density at a point*

The infinite population approximation with a perfectly uniform distribution cannot reproduce the behaviour of the individual based model, because the initial symmetry always remains [10]. Therefore, one must introduce an asymmetry into the initial distribution. We chose to increase the density at one point  $\mathbf{p}$ . We tried 4 different positions for this point in a quadrant as shown on Fig. 3 (note that by symmetry we only need to test positions in one quadrant). We found no significant change in the final number of clusters when changing

the position of the point. Therefore, we limited the tests to the case of the initial peak at (0.6, 0.2).



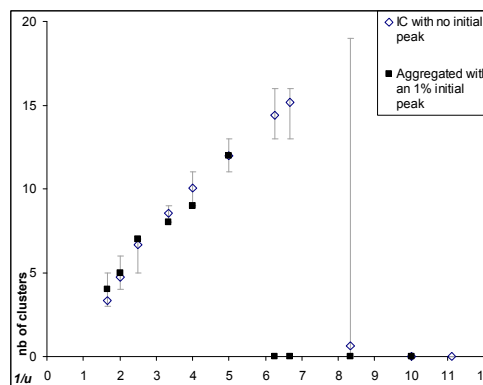
**Fig. 3.** Tested positions of the initial peak  $\mathbf{p} = (0,0), (0.6,0.2), (0.8,0.6), (1.0,1.0)$ .

### ***Varying the uncertainty $u$ with the initial density at point $\mathbf{p}$ of 1%***

We know from the study of the individual-based model that the main parameter ruling the final number of clusters is the uncertainty  $u$ . Therefore, in our first experiment we choose a high value for the initial density at point  $\mathbf{p}$  of 1%. This value is high because, on this grid, the initial density is 0.0001 on each point for a uniform initialisation. As for the ABM, we vary the uncertainty  $u$ : 0.12, 0.15, 0.16, 0.2, 0.3, 0.4, 0.5, 0.6. The size of the grid does not allow us to test lower uncertainty (such uncertainties require a more precise grid and simulations lasting a dissuasive time). The aggregated model runs between 20000 and 30000 iterations depending on the value of  $u$ . We count the clusters with the procedure presented in 2.2.

Fig. 4 shows the results. Black squares represent the result obtained with the aggregated model with an initial 1% peak. Empty rhombuses represent the average ABM results on 30 replicas. We note that the models behave very similarly. We observe that:

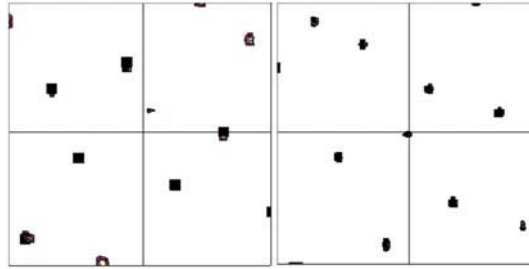
- when  $u$  is below a critical value, there is no clustering. However the ABM stops to cluster for a lower uncertainty value (0.12) than the aggregated model (0.2).
- when  $u$  is higher than this critical value, we get clusters, in a number which is linear with  $1/u$ .



**Fig. 4.** Final number of clusters of the AB and aggregate model for various values of  $u$  (0.12, 0.15, 0.16, 0.2, 0.3, 0.4, 0.5, 0.6). The initial density of the aggregate model has an initial peak of size 0.01.

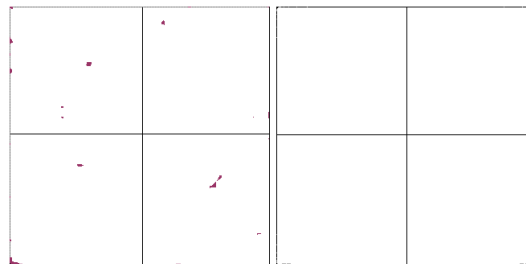
Fig. 5 shows the final state of both models for  $u = 0.2$ . The replica of the ABM (on the left) exhibits 12 clusters, and the aggregate model (on the right) exhibits 13 clusters. Both have

four "extreme" clusters on the boundary of the attitude space and, as we explain in [2], clusters are separated by a distance of about  $u$  on the two dimensions.



**Fig. 5.** Final state of both models for  $u = 0.2$ : on the left, a replica of the ABM; on the right, the aggregate model.

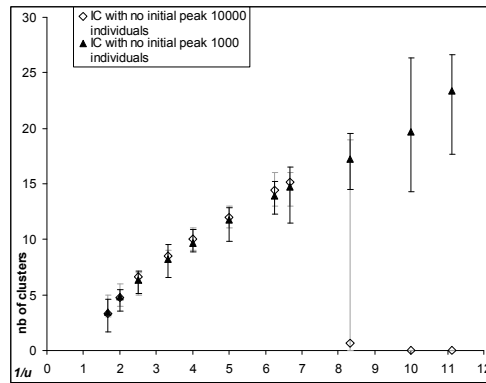
Fig. 6 shows the final state of both models for  $u = 0.12$ . None of the models yields clusters. The irregularities of density that we observe are lower than 0.005. We studied the influence of the size of the initial peak for  $u = 0.2$  and  $u = 0.5$ . It appears that the model behaviour is the same with initial peak of 0.0002 as with an initial peak of 0.01. This shows that for these values of  $u$ , any small irregularity of the initial density leads to clustering, whereas for  $u$  lower than 0.2, the dynamics completely suppresses initial peaks of 1% of the population. These observations are in accordance with the results obtained with the symmetrical aggregate model: for all value of  $u$  equal or higher than 0.2 and for  $\mu = 0.5$ , the dynamic leads to only one central cluster (see [10] for more details), whereas for lower values no cluster is formed.



**Fig. 6.** Final state of both models for  $u = 0.12$ : on the left, a replica of the ABM; on the right, the aggregate model. The irregularities of the density are lower than 0.005.

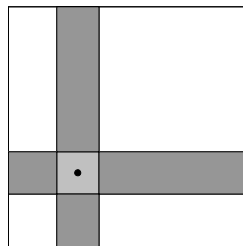
## Discussion and conclusion

We note that the critical value of  $u$  is 0.2 for the aggregated model, whereas it is 0.12 for the ABM with a population of size 10000. Here, there is a population size effect. On Fig. 7, we note that with 1000 individuals, the clustering takes place with all the tested values of  $u$  (this is the reason why we missed the non clustering behaviour in our first tests on a population of 1000). Note that, as the initial noise on the density is higher with 1000 individuals, we neglect densities of 0.01 instead of 0.005 in the procedure of peak counting. Obviously, small population favours clustering.

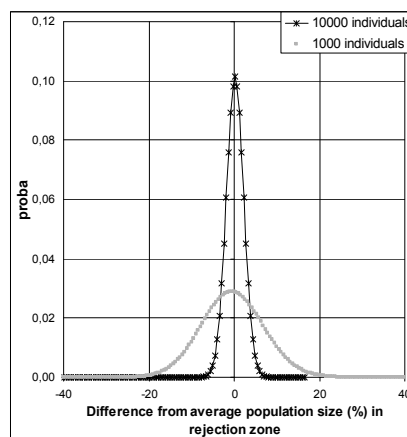


**Fig. 7.** Number of clusters in the final state of the ABM for various values of  $u$  (0.09, 0.1, 0.12, 0.15, 0.16, 0.2, 0.3, 0.4, 0.5, 0.6), and a population of 1000 and 10000 individuals. The abscissa is  $1/u$ . The error bars indicates the minimum and the maximum obtained on the 30 replicas.

Qualitatively, we can explain these results by analysing more carefully the rules of interactions. Fig. 8 represents the zones of attraction and rejection of a given opinion vector. The opinion is attracted by opinions located in a square of size  $2u$  around (thus a surface of  $4u^2$ ), and rejected by opinions located in the vertical and horizontal strips of width  $2u$ , representing a surface  $8(Mu - u^2)$ . Hence the proportion of surface of attracting opinions over the rejecting opinions is:  $u/(2(M-u))$ . When  $u$  decreases, the influence tends to be only rejection.



**Fig. 8.** Influence zones in the opinion space. The considered opinion is the black dot. The light grey square around it is the attraction zone, and the dark grey horizontal and vertical strips are the rejection zones.



**Fig. 9.** Probability distributions of the number of individuals in a rejection zone (defined on figure 8), considering a uniform distribution of individuals in the whole domain, and for  $u = 0.1$ . The variations around the average are much higher when the population size is 1000 than when it is 10000 (central-limit theorem).

Below a critical value, and when the population is large, an individual is submitted almost only to random rejections. Each individual has then a random trajectory in the opinion

space, without finding an attraction centre strong enough to keep it, as shown on Fig. 10. This figure presents the attitude trajectory of a randomly chosen individual in a population of size 10000 during 500000 iterations (its attitudes are measured every 5000 iterations). The global density remains therefore almost uniform, and all opinions stay scattered.



**Fig. 10.** Trajectory of a randomly chosen individual during 500000 iterations measured all the 5000 iterations in the ABM for  $u = 0.12$  and a population of 10000 individuals.

The aggregated model predicts a critical value of  $u = 0.2$  below which there is no clustering, and the individual trajectories are random. However, we observe that this critical value is smaller for finite population. The reason is that, with the random fluctuations of the population, configurations where a small cluster is formed, with a lower density of opinions in the rejection strips (see Fig. 8) and a higher density in the attraction square than the average density, get more likely for smaller population as predicted by the central limit theorem (see Fig. 9). In this case, the rejection forces are not strong enough to erase the cluster, which has a chance to strengthen and maintain itself. Hence, for  $u < 0.2$ , in finite populations, clustering is due to particular random events that may take a very long time to take place.

To conclude, we rapidly discuss more general issues about potential interpretations of the model in human societies. First, the results suggest that, when individuals are open to each other, adding a dynamics of rejection favours the clustering, and thus conformism. This was the main observation of our first study of this model, which was somewhat counter-intuitive. Now, we observe a radical change of this tendency when the population is large and individuals in wide interaction with each other: there is a critical level of openness to the others below which, the population becomes totally unstable, with only individualist behaviours. Each individual has its own random trajectory and there is no possibility of collective organisation. This reminds some features of the post-modern globalisation. However, we certainly don't claim to capture the essence of this complex phenomenon with our simple model.

## References

9. Ben-Naim, E., Krapivsky, P.L., and Redner, S.: Bifurcations and Patterns in Compromise Processes, *Physica D: Non linear phenomena*, (2003) vol. 183, 190-204.
10. Huet, S., Deffuant, G.: Bounded Confidence with Rejection: The Infinite Population Limit with Perfect Uniform Initial Density. MASHS 2008, *Modèles et Apprentissage en Sciences Humaines et Sociales*, 5 et 6 juin 2008, Créteil, France, 9 p.
11. Brehm, J.W.: *A theory of psychological reactance* (New York Academic Press, 1966).
12. Heider, F.: Attitudes and Cognitive Organization. *J. of Psychology*, 1946(21): pp. 107-112.
13. Rogers, R.W. and S. Prentice-Dunn: *Protection motivation theory*, in Handbook of health behavior research, D. Gochman, Editor (New York:Plenum, 1997) pp. 113-132.



14. Sherif, M. and C.I. Hovland: *Social Judgement* (New Haven, CT: Yale Univ. Press, 1961).
15. Tajfel, H.: *Differentiation between social groups: Studies in the social psychology of intergroup relations* (London: Academic Press, 1978).
16. Turner, J.C., et al.: *Rediscovering the social groups: A self-categorization theory* (Cambridge, MA: Basil Blackwell, 1987).
17. Brewer, M.B.: The Social Self: On Being the Same and Different at the Same Time. *Pers. Soc. Psychol. Bulletin*, 1991. 17(5): pp. 475-482.
18. Eagly, A.H. and S. Chaiken: *Process Theories of Attitude Formation and Change: Attribution Approaches and Social Judgment Theory*, in *The Psychology of Attitudes*, A.H.C. Eagly, S., Editor (Thomson/Wadsworth, 1993) pp. 351-388.
19. Galam S.: Contrarian deterministic effects on opinion dynamics: 'the hung elections scenario'. *Physica A*, 2004, 333: pp. 453-460.
20. Borghesi C. and S. Galam: Chaotic, staggered and polarized dynamics in opinion forming: the contrarian effect. *Physical Review E*, 2006, 73: 9 pages.
21. de La Lama M.S., J. M. Lopez and H.S. Wio: Spontaneous emergence of contrarian-like behaviour in an opinion spreading model. *Europhys. Lett.*, 2005, 72: pp. 851-7.
22. Wio, M. S. de la Lama and J. M. Lopez: Contrarian-like behavior and system size stochastic resonance in an opinion spreading model. *Physica A*, 2006, 371: pp. 108-11.
23. Jager, W. and F. Amblard, A dynamical perspective on attitude change. *NAACSOS Conference 2004*. June 27-29, 2004, Pittsburgh PA <http://www.casos.cs.cmu.edu/naacsos/>.
24. Jager, W. and F. Amblard: Multiple Attitude Dynamics in Large Populations. *Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms*. 2005, Chicago.
25. Jager, W. and F. Amblard: Uniformity, Bipolarization and Pluriformity Captured as Generic Stylized Behavior with an Agent-Based Simulation Model of Attitude Change. *Computational and Mathematical Organization Theory*, 2004(10): pp.295-303.
26. Salzarulo, L.: A continuous Opinion Dynamics Model Based on the Principle of Meta-Contrast. *Journal of Artificial Societies and Social Simulation*, 2006. 9(1).
27. Urbig, D. and R. Malitz: Dynamics of structured attitudes and opinions. *Third Conference of the European Social Simulation Association*. 2005. September 5-8, Koblenz, Germany.
28. Urbig, D. and R. Malitz: Drifting to more extreme but balanced attitudes: Multidimensionnal attitudes and selective exposure. *4th Conf. of the European Social Simulation Association*. 2007. Toulouse, 10-14 September.

# Chapter 2.5 Attraction-Rejection – designed from experiments

Advances in Complex Systems, Vol. 18, No. 8 (2010) 405–428  
© World Scientific Publishing Company  
DOI: 10.1142/S021952991000258



Title: Openness leads to opinion stability and narrowness to volatility

Authors: Sylvie Huet and Guillaume Deffuant

Published in Advances for Complex Systems, 2010.

## OPENNESS LEADS TO OPINION STABILITY AND NARROWNESS TO VOLATILITY

SYLVIE HUET<sup>1</sup> and GUILLAUME DEFFUANT<sup>2</sup>

<sup>1</sup>Centagref, Laboratoire d'Ingénierie des Systèmes Complexes,

GH12 F-Aubière, France\*

<sup>2</sup>sylvie.huet@centagref.fr

<sup>3</sup>guillaume.deffuant@centagref.fr

Received 8 December 2009

Revised 7 May 2010

We propose a new opinion dynamic model based on the experiments and results of Wood et al. (1996). We consider pairs of individuals discussing on two attitudinal dimensions, and we suppose that one dimension is important, the other secondary. The dynamics are mainly ruled by the level of agreement on the main dimension. If two individuals are close on the main dimension, then they attract each other on the main and on the secondary dimensions, whatever their disagreement on the secondary dimension. If they are far from each other on the main dimension, then too much proximity on the secondary dimension is uncomfortable, and generates rejection on this dimension. The proximity is defined by comparing the opinion distance with a threshold called attraction threshold on the main dimension and rejection threshold on the secondary dimension. With such dynamics, a population with opinions initially uniformly drawn evolves to a set of clusters, inside which secondary opinions fluctuate more or less depending on threshold values. We observe that a low attraction threshold favours fluctuations on the secondary dimension, especially when the rejection threshold is high. The opinion evolutions of the model can be related to some stylised facts.

**Keywords:** Attitudinal opinion dynamics; rejection; attraction; openness; narrowness; individual based model.

### 1. Introduction

Teenagers often tend to adopt any opinion expressed by their group or by their rock star idol, and reject any opinion expressed by their parents (or other representatives of the previous generation). Such unconditional influence and rejection mechanisms have been exhibited in several experiments [10, 24, 25]. Socio-psychological theories propose a variety of conceptual frameworks to interpret these observations: at individual level [2, 14, 26, 28] and at group level [29, 30, 3]. They generally claim influence is based on two complementary processes: attraction and rejection. Nevertheless, rejection mechanisms are still poorly understood.

\*Centagref, Laboratoire d'Ingénierie des Systèmes Complexes, BP 20085, GH12 F-Aubière.

## Abstract.

We propose a new opinion dynamic model based on the experiments and results of Wood et al (1996). We consider pairs of individuals discussing on two attitudinal dimensions, and we suppose that one dimension is important, the other secondary. The dynamics are mainly ruled by the level of agreement on the main dimension. If two individuals are close on the main dimension, then they attract each other on the main and on the secondary dimensions, whatever their disagreement on the secondary dimension. If they are far from each other on the main dimension, then too much proximity on the secondary dimension is uncomfortable, and generates rejection on this dimension. The proximity is defined by comparing the opinion distance with a threshold called attraction threshold on the main dimension and rejection threshold on the secondary dimension. With such dynamics, a population with opinions initially uniformly drawn evolves to a set of clusters, inside which secondary opinions fluctuate more or less depending on threshold values. We observe that a low attraction threshold favours fluctuations on the secondary dimension, especially when the rejection threshold is high. The opinion evolutions of the model can be related to some stylised facts.

## Contents

1 Introduction.....	211
2 The Dynamical Model of Interacting Individuals .....	213
3 Typical Evolutions of the population.....	214
31 Model initialization and experiments .....	214
32 Attraction and rejection thresholds are equal .....	215
33 Attraction larger than rejection threshold .....	215
34 The attraction threshold is smaller than the rejection threshold .....	216
35 Hypothesis about the global behaviour of the model.....	219
4 Systematic experiments .....	220
41 Initialization of the model and experiments .....	220

42 Final number of clusters.....	220
43 Presence of fluctuations and polarization.....	221
44 Size of the major clusters .....	224
5 Discussion and conclusion.....	225
References.....	227

## OPENNESS LEADS TO OPINION STABILITY AND NARROWNESS TO VOLATILITY

SYLVIE HUET<sup>†</sup> and GUILLAUME DEFFUANT<sup>‡</sup>

*Cemagref, Laboratoire d'Ingénierie des Systèmes Complexes,  
63172 Aubière, France\**

<sup>†</sup>*sylvie.huet@cemagref.fr*

<sup>‡</sup>*guillaume.deffuant@cemagref.cf*

Received 8 December 2009

Revised 7 May 2010

We propose a new opinion dynamic model based on the experiments and results of Wood *et al.* (1996). We consider pairs of individuals discussing on two attitudinal dimensions, and we suppose that one dimension is important, the other secondary. The dynamics are mainly ruled by the level of agreement on the main dimension. If two individuals are close on the main dimension, then they attract each other on the main and on the secondary dimensions, whatever their disagreement on the secondary dimension. If they are far from each other on the main dimension, then too much proximity on the secondary dimension is uncomfortable, and generates rejection on this dimension. The proximity is defined by comparing the opinion distance with a threshold called attraction threshold on the main dimension and rejection threshold on the secondary dimension. With such dynamics, a population with opinions initially uniformly drawn evolves to a set of clusters, inside which secondary opinions fluctuate more or less depending on threshold values. We observe that a low attraction threshold favors fluctuations on the secondary dimension, especially when the rejection threshold is high. The opinion evolutions of the model can be related to some stylized facts.

*Keywords:* Attitude; opinion dynamics; rejection; attraction; openness; narrowness; individual based model.

### 1. Introduction

Teenagers often tend to adopt any opinion expressed by their group or by their rock star idol, and reject any opinion expressed by their parents (or other representatives of the previous generation). Such unconditional influence and rejection mechanisms have been exhibited in several experiments [10, 24, 25]. Socio-psychological theories propose a variety of conceptual frameworks to interpret these observations: at individual level [2, 14, 26, 28] and at group level [29, 30, 3]. They generally claim influence is based on two complementary processes: attraction and rejection. Nevertheless, rejection mechanisms are still poorly understood.

\*Cemagref, Laboratoire d'Ingénierie des Systèmes Complexes, BP 50085, 63172 F-Aubière.

Social simulation can provide complementary evidences on how mechanisms at individual level generate collective regularities in large populations, whereas experiments of social psychology are limited to small groups. In this perspective, the effect of the opinion or attitude [9] attraction has been widely studied [5, 7, 8, 12, 13, 21]. Fewer works are dedicated to both attraction and rejection mechanisms [18–20, 27, 31, 32]. We have recently proposed a model inspired from the socio-psychological theories [17] (dissonance [10], the social judgement [28] and the self-categorization [30]), which couples attraction and rejection in a multidimensional approach as in [1, 22, 6, 11, 16, 19, 20, 31, 32]. We observe that this model generally leads to more consensus than when attraction is the only mechanism. In this paper, we propose a refinement of our previous model which is directly inspired from Wood *et al.* experiment [33].

Wood *et al.* describe their main results as follows: in a first study, *participants who considered a majority group relevant to their own self-definitions (but not those who judged it irrelevant), on learning that the group held a counter-attitudinal position, shifted their attitudes to agree with the source. In a second study, recipients who judged a minority group (negatively) self-relevant, on learning that the group held a similar attitude to their own, shifted their attitudes to diverge from the source. These shifts in attitudes were based on participants' interpretations of the attitude issues.* The authors suggest that these attitude shifts reflect normative pressures to align with valued groups and to differentiate from derogated groups. Globally, these experiments show that:

- one can be attracted by a far opinion (counter-attitudinal), if this opinion comes from a group (which can be majority but not necessary) sharing other fundamental values (relevant to their own self-definition),
- one tends to shift from a close opinion expressed by a group (minority or not) with significant differences about fundamental values (negatively self-relevant).

We propose a new model derived from [34], which aims at reproducing these observations. However, modeling implies interpretations and simplifications. First, we study group formation from initially scattered opinions, whereas the experiments took place in the context of already existing groups. Thus, in our model, we consider pair interactions only (we do not consider the group directly), and we reinterpret the experiment by considering two individuals who can share or not fundamental values, and influence each other on less important (secondary) opinions. Second, we consider only two attitudinal dimensions, one is for fundamental values (called main dimension), the other for more secondary opinions (called secondary dimension). On each dimension, we suppose that the opinions can take continuous values, between  $-1$  and  $+1$ . Then, we introduce influence dynamics. When the individuals are close to each other on the main dimension, they tend to attract each other on the secondary dimension, even if their disagreement is strong. When they disagree on the fundamental dimension, and are close on the secondary, then they tend to reject each other's secondary opinions. To model a disagreement, we compare the

distance between opinions with a threshold (called attraction threshold on the main dimension, and rejection threshold on the secondary dimension).

We consider populations of individuals with opinions initially drawn at random, and simulate their evolutions with these mechanisms, for different values of the parameters (mainly the attraction and rejection thresholds). The stationary state shows a set of clusters, with fluctuations on the secondary dimension. These fluctuations are lower or absent when the attraction threshold is high. They cover the whole length of the opinion axis when both the attraction and rejection thresholds are low. Moreover, a high rejection threshold tends to generate some clusters with extreme opinions on the secondary dimension (polarization). We propose a theoretical relation between the thresholds values which allows us to predict large fluctuations and polarization.

This article firstly presents the proposed model. A second section describes typical opinion evolutions of the population. These first observations lead us to formulate our main hypothesis on the parameter values leading to fluctuations or polarization on the secondary dimension. Then we propose a more complete experiment design to check this hypothesis. A last section discusses and concludes this work.

## 2. The Dynamic Model of Interacting Individuals

We consider a population of  $N$  individuals. The model includes three parameters:  $u_m$  and  $u_s$ , the attraction and rejection thresholds, respectively, and  $\mu$  ruling the intensity of influence at each meeting (comprised between 0 and 0.5).

An individual has two opinions  $x_m$  (on the main dimension) and  $x_s$  (on the secondary dimension) taking real values between  $-1$  and  $+1$ .

During an iteration, a couple of individuals  $X$  and  $Y$  is randomly chosen and can influence each other. The algorithm is the following:

- Choose randomly a couple  $(X, Y)$  of individuals in the population;
- $X$  and  $Y$  change their opinions at the same time, according to the influence function.

We present the calculation of the influence of  $Y$  on  $X$  (of course the influence of  $X$  on  $Y$  is found by inverting  $X$  and  $Y$ ). Let  $(x_m, x_s)$  and  $(y_m, y_s)$  be the opinions of  $X$  and  $Y$  respectively. We first consider the main opinion dimension:

- If  $|x_m - y_m| \leq u_m$ , individual  $X$  agrees with  $Y$  on the main dimension. Both attitudes of  $X$  are going to get closer to those of  $Y$ , proportionally to the attitudinal distance on each dimension:

$$x_m(t+1) = x_m(t) + \mu(y_m(t) - x_m(t)) \quad (1)$$

$$x_s(t+1) = x_s(t) + \mu(y_s(t) - x_s(t)). \quad (2)$$

Indeed, whatever the agreement level on the secondary dimension is,  $X$  is going to be globally closer to  $Y$ . If it was already close, it gets closer. Even if it was

far on this dimension, following the spirit of the experimental observations, the closeness on the main dimension leads to get closer on the secondary dimension.

- If  $|x_m - y_m| > u_m$ , individual  $X$  disagrees with  $Y$  on the main dimension and if  $|x_s - y_s| \leq u_s$ : Individual  $X$  feels it is too close to  $Y$  on the secondary dimension, because of their disagreement on the main dimension. To solve the conflicting situation,  $X$  moves away from  $Y$  on this dimension. The attitude change is proportional to the distance to reach the rejection threshold:

if  $(x_s - y_s) < 0$  then

$$x_s(t+1) = x_s(t) - \mu\{u_s - (y_s(t) - x_s(t))\} \quad (3)$$

else

$$x_s(t+1) = x_s(t) + \mu\{u_s + (y_s(t) - x_s(t))\}. \quad (4)$$

In the other cases,  $X$  is not modified by  $Y$ .

Moreover, we confine the attitude in the interval  $[-1, +1]$ : if  $|x_i| > 1$  then  $x_i := \text{sign}(x_i)$  where  $\text{sign}(\ )$  is a function which returns  $-1$  if its argument is strictly negative,  $+1$  if otherwise.

The attitude of  $Y$  is calculated in the same way considering the situation of the meeting with  $X$ .

### 3. Typical Evolutions of the Population

As known for the classical bounded confidence model and its extensions, the most significant parameter is the threshold limiting the confidence (sometimes called uncertainty). Here the attraction and rejection thresholds  $u_m$  and  $u_s$  are the main parameters. We consider three situations, for which we study the population opinion evolutions:

- $u_m = u_s$ ;
- $u_m > u_s$ ;
- $u_m < u_s$ .

We firstly describe the initialization parameters and the experimental design. Then, each possible typical evolution of the population is shown by several two-dimensional graphs representing each the two-attitude space. Each graph draws at different times the attitudes of individuals of the population. Globally, a figure, composed of a set of graphs, shows the evolution of the attitudes over time for a run of the model. Finally, we formulate some hypothesis about the global behavior of the model.

#### 3.1. Model initialization and experiments

In all simulations, the following values are fixed:

- All individuals have the same speed of attitude change  $\mu = 0.5$  on the two attitudinal dimensions;

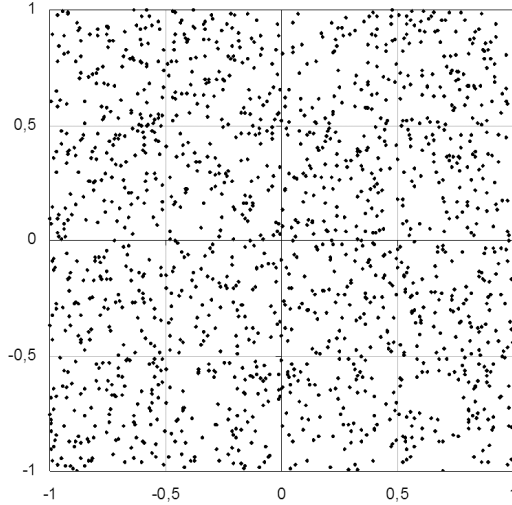


Fig. 1. Initial state for of a 1500-individual population.

- The main attitude dimension is the horizontal one (the secondary dimension is the vertical one);
- The size of the population is equal to 1500 individuals.

The attitudes  $x_i$  on each dimension are initialized following a Uniform law and comprised between  $-1$  and  $1$ . Such an initialization is presented in Fig. 1 for which each axe represents one attitudinal dimension varying from  $-1$  to  $1$ . Thus, each point corresponds to the coordinates of an individual's attitudes.

We run the model until it reaches a stationary state.

### 3.2. Attraction and rejection thresholds are equal

Figure 2 shows an example of the evolution of the opinions when the attraction and rejection thresholds are equal. We firstly observe an attraction between individuals on the secondary dimension ( $y$ -axis) due to their proximity on the main dimension ( $x$ -axis). Then, clusters appear on the main dimension. While groups are forming close to the middle of the secondary dimension, they begin to reject each other. This is due to the disagreement between groups on the main dimension, which lead to rejection between close opinions on the secondary dimension. Thus, most of the groups polarize on the secondary dimension. This means that the average  $y$ -attitude value of each group increases in absolute value. We finally observe that the number of clusters is defined by the attraction threshold  $u_m$ . This value defines the minimum distance between two groups preventing them to merge.

### 3.3. Attraction threshold larger than rejection threshold

Figure 3, shows an example of simulation with  $u_m = 0.3, u_s = 0.1$ . Again, we observe the attraction on the secondary dimension first, and then, clusters appear



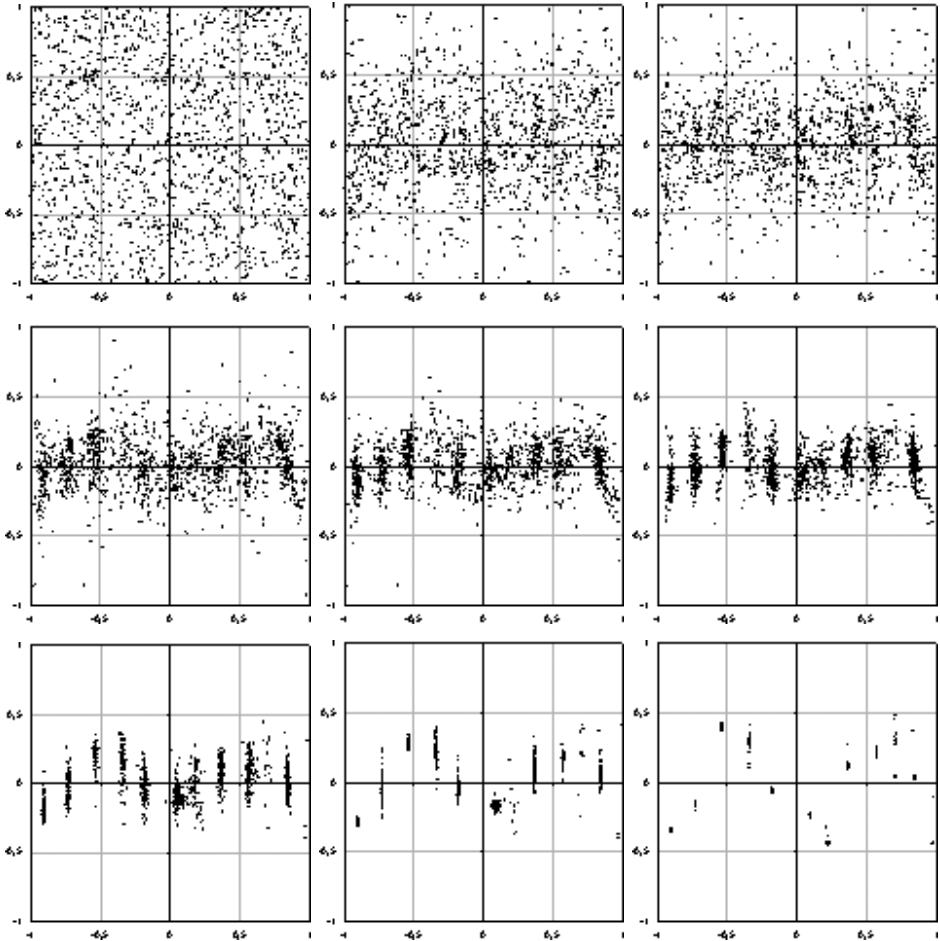


Fig. 2. Population evolution (from upper left to lower right) at step number 0, 30,000, 37,500, 52,500, 67,500, 90,000, 105,000, 172,500, 375,000, of a 1500-individual population for  $u_m = 0.1$ ,  $u_s = 0.1$ . Main dimension is horizontal.

on the main dimension, with a gravity center close to 0 on the secondary dimension. As previously, the clusters tend to reject each other on the secondary dimension, but this effect is lower because the rejection threshold is lower.

As in the previous section, the number of large clusters is determined by the attraction threshold. We also observe some minor clusters on the border of the main attitude space.

### 3.4. *The attraction threshold is smaller than the rejection threshold*

Figure 4 shows the evolution of the population for  $u_m = 0.1$ ,  $u_s = 0.3$ . The pattern of opinion evolution is slightly different. The initial attraction on the secondary

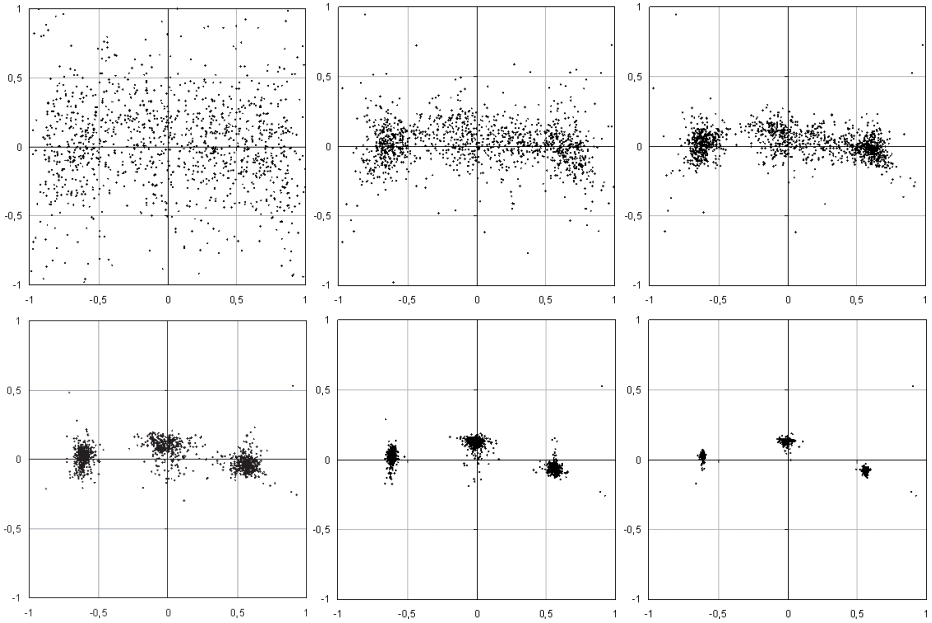


Fig. 3. Population evolution (from upper left to lower right) at step number 11,250, 22,500, 37,500, 45,000, 52,500, 67,500, of a 1500-individual population for  $u_m = 0.3, u_s = 0.1$ . Main dimension is horizontal.

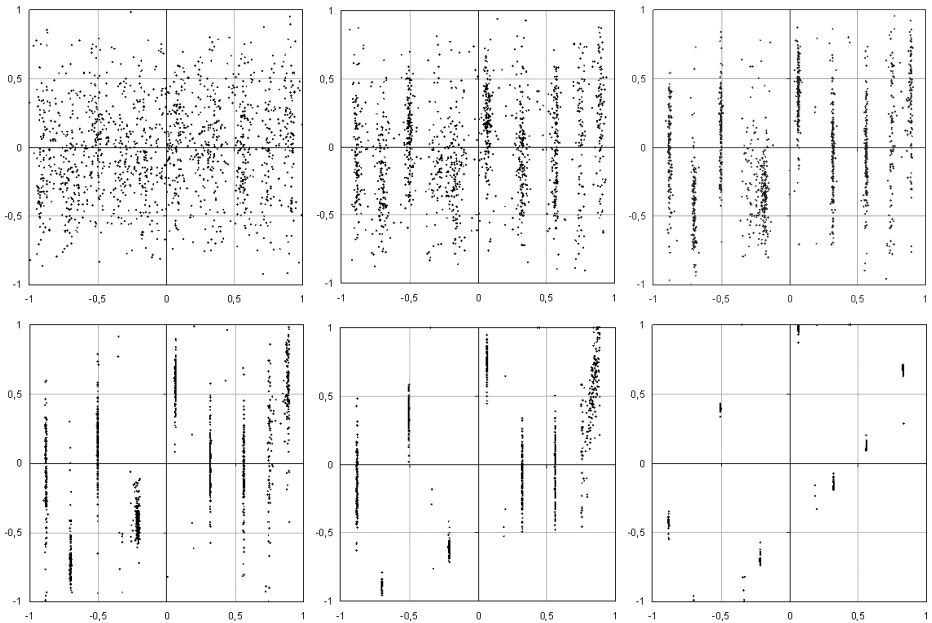


Fig. 4. Population evolution (from upper left to lower right) at time step number 30,000, 90,000, 127,500, 168,750, 225,000, 375,000, for  $u_m = 0.1, u_s = 0.3$ .

dimension is very weak. The polarization on this dimension begins before the clusters have been really formed. Indeed, as the rejection threshold is high, the conditions for rejection are met more often. Moreover, the clusters must have a higher distance on the secondary dimension to reach some stability. The stability is not complete, because some fluctuations remain in the clusters, on the secondary dimension, it is not possible to keep a distance of 0.3 between 10 clusters on a distance of 2 overall. Hence some rejection continues to take place between the groups at the stationary state.

This effect increases when the rejection threshold increases (see Figs. 5 and 6). The fluctuations on the secondary dimension reach almost the whole dimension space on Fig. 6, for a rejection threshold  $u_s = 0.5$ .

For  $u_m = 0.1, u_s = 1.1$  (Fig. 6), a larger density of individuals appear on the borders of the attitude space as shown on Fig. 7 on the left. On the right with  $u_m = 0.3, u_s = 1.5$  this tendency is enhanced with two stable extreme clusters, and a central cluster where secondary attitudes fluctuate. The three groups include similar number of individuals.

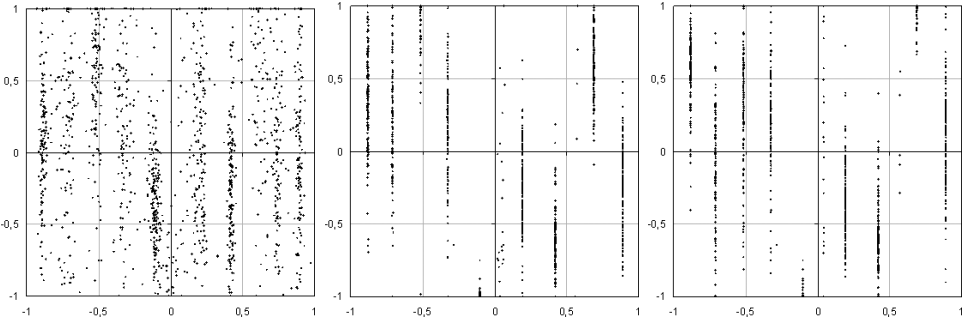


Fig. 5. Population evolution at time step number 90,000, 187,500, 375,000,  $u_m = 0.1, u_s = 0.5$

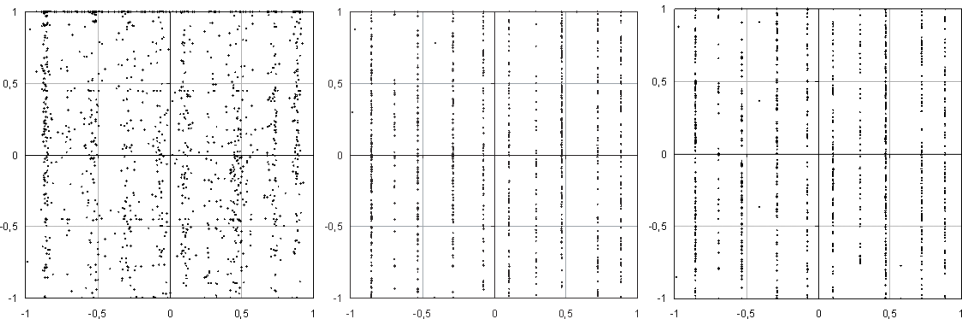


Fig. 6. Population evolution at time step number 90,000, 187,500, 375,000, for  $u_m = 0.1, u_s = 1.1$ .

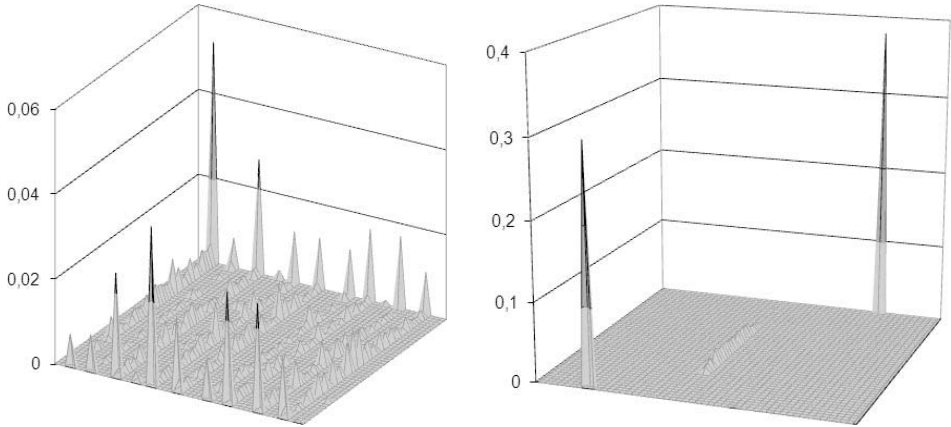


Fig. 7. Density of individuals on the attitude space at the equilibrium for a 1000-individual population: for  $u_m = 0.1, u_s = 1.1$  on the left; for  $u_m = 0.3, u_s = 1.5$  on the right.

### 3.5. Hypothesis about the global behavior of the model

In the area of our experimental plan, and from these first observations of the model evolutions, we make the following hypothesis on the global dynamics.

- (1) The final number of large clusters is approximately  $(1/u_m)$ , because this number is ruled by the dynamics on the main dimension, the classical bounded confidence model [8]. Indeed, on the secondary dimension, there cannot be two clusters on the same vertical line, because these clusters tend to merge whatever their distance. Let us notice that 1 corresponds to the mid-width of an attitudinal dimension (2 is the total width of an attitudinal dimension).
- (2) On the secondary dimension, if the space is sufficient to get clusters distant from each other of more than the rejection threshold  $u_s$ , the final state is static. Otherwise, there are constant fluctuations due to the rejection, in the stationary state. Regarding the approximation of the total number of clusters, the maximum possible distance between two clusters on the secondary dimension is:

$$\delta = \frac{2}{((1/u_m) - 1)}. \tag{5}$$

Thus, if  $u_s \leq \delta$ , there is no fluctuations on the secondary dimension. But, when  $u_s$  gets close to  $\delta$ , the clusters tend to occupy the whole length of the secondary dimension. For  $u_s > \delta$ , the clusters are necessarily rejecting each other on the secondary dimension, because their distance is less than  $u_s$ . The opinions fluctuate more and more on the secondary dimension, when  $u_s$  increases.

Note that when  $u_m$  is small,  $\delta$  is close to  $2u_m$ . In this case, one can summarize that fluctuations on the secondary dimension appear when the rejection threshold is about higher than twice the attraction threshold.

From the study of opinion evolutions, we can also claim that, in the no fluctuation zone, the larger the rejection threshold, the higher is the polarization. Indeed, after the initial attraction to the neutral position on the secondary dimension, groups tend to increase in absolute value their average opinion until they reach one avoiding the conflict with the other groups.

We are now going to check our hypothesis with more systematic experiments.

## 4. Systematic Experiments

### 4.1. Initialization of the model and experiments

The initialization is the same as the one presented in Sec. 3.1. The model runs during 40,000,000 iterations. This is always sufficient to attain the stationary state. In several experimental designs, we vary systematically the values of the attraction and rejection threshold. We study the number of clusters (for two population sizes: 1000 and 7500 individuals), the presence of fluctuations and the polarization for a population of 1000 individuals (see Fig. 8).

The presented results are the average, and sometimes the maximum and the minimum, of the measured values on 20 replicas run for each set of parameter values. The final number of clusters is computed via a classical algorithm searching for the chains of individuals separated by a maximum distance (the chosen distance for this model is  $\min(u_m, u_s)$ ).

### 4.2. Final number of clusters

Figure 9 shows the counted average, maximum and minimum, number of clusters representing each more than 1% of the population.

Firstly, we observe that the size of population does not change the final number of clusters. Indeed, the diagram for 1000 individuals (see Fig. 9 on the left) is close to the diagram for a population of 7500 individuals (see Fig. 9 on the right).

The most important observation is that, for the both population sizes, the average number of major clusters approximatively corresponds to  $(1/u_m)$ . Regarding the number of clusters containing more than 1% of the population, the model has

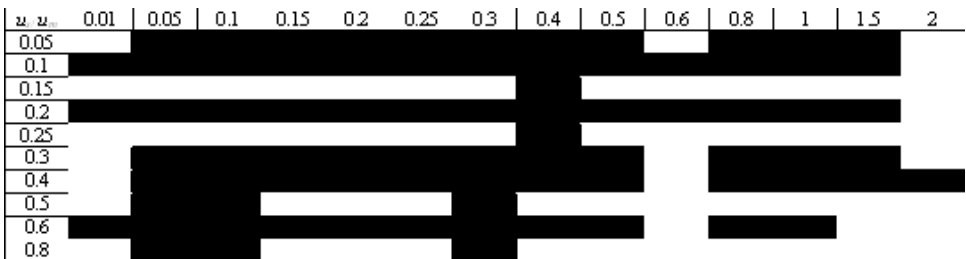


Fig. 8. Tested values (in black) for  $u_m$  (columns) and  $u_s$  (lines) with a population of 1000 individuals.

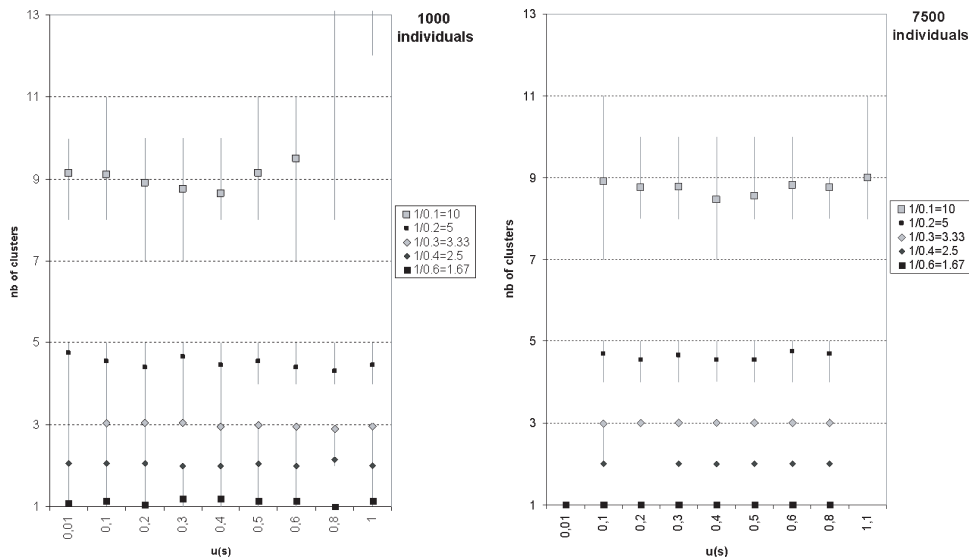


Fig. 9. (On the left) Average final cluster numbers, each representing at least 1% of the population, for various values of  $u_m$  and  $u_s$  for  $u_m$  equal to (from the bottom to the top: 0.6, 0.4, 0.3, 0.2, 0.1) for 1000 individuals (on the left) and 7500 individuals (on the right). The legend indicates the number of clusters which can be calculated from the simulated values of  $u_m$ . One can see the results of these calculations are quite similar to the simulation results.

exactly the same behavior as the classical bounded confidence model [8]. The exceptions are due to a default of the algorithm counting the cluster for small population. Indeed, when  $u_m$  is lower and lower, the algorithm becomes more and more inefficient especially when the fluctuations on the secondary dimension are large.

### 4.3. Presence of fluctuations and polarization

Figure 10 shows the results of sets of simulations defined by the values of  $u_m$  ( $x$ -axis) and  $u_s$  ( $y$ -axis).

For each replica, we calculate at the equilibrium state the opinion standard deviation of each cluster. Then, we compute for the replica the average, minimum and maximum standard deviations on the clusters. We finally compute the average, minimum and maximum standard deviations on all the replicas for the considered couple of values  $(u_m, u_s)$ . Hence we get, for each couple  $(u_m, u_s)$ , the average on the replicas of the average standard deviations on clusters (represented by a grey disc on Fig. 10, left), the minimum on the replicas of the minimum standard deviations on clusters (represented by a dark circle on Fig. 10, left), the maximum on the replicas of the maximum standard deviations on clusters (represented by a dotted circle on Fig. 10, left). This indicator gives a rough idea of the width of the clusters on the secondary dimension. Indeed, we have seen in Figs. 4 to 6 that fluctuations enlarge the clusters on the secondary dimension. The dark line represents  $\delta$ . Figure 10 on

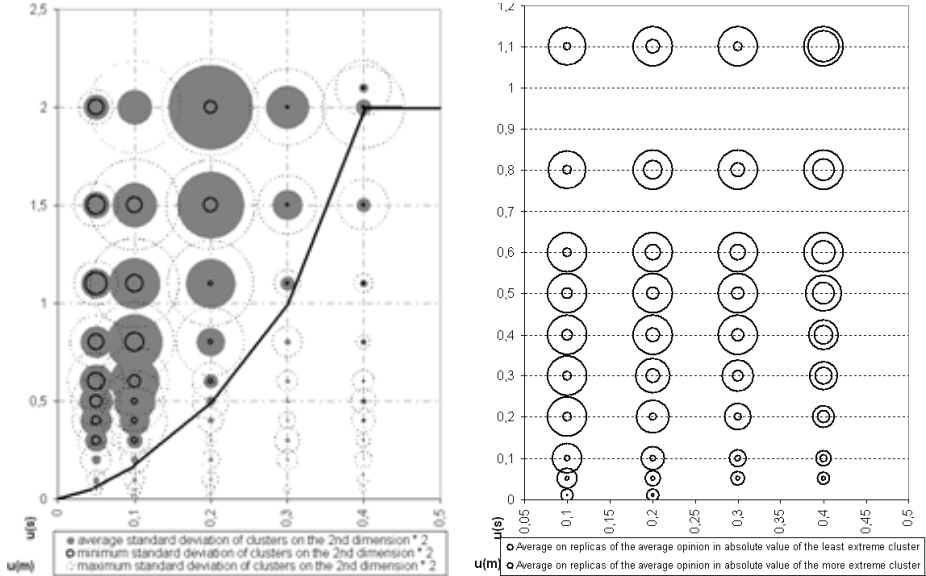


Fig. 10. Average, minimum and maximum on the 20 replicas of the standard deviation of cluster for the secondary dimension and  $\delta$  which is represented by the dark line (on the left). Average on replicas of the average attitude on the secondary dimension of the least and the more extreme cluster (on the right) for various values of  $u_m$  and  $u_s$  and a population of 1000 individuals.

the left confirms that large fluctuations occur when  $u_s > \delta$  since the width of the clusters increases in this zone. It also confirms that the fluctuations increase when  $u_s$  increases. However, while our hypothesis claims that the final state is static for  $u_s \leq \delta$ , we observe on the graph some fluctuations, especially when  $u_m$  is large.

On the right of the Fig. 10, we can observe the average attitude on replicas of the secondary dimension of the least and the most extreme clusters. One can see, as we have hypothesized, that the polarization increases when  $u_s$  increases until it reaches a plateau. However, as previously noticed for fluctuations, we observe extreme polarization for some  $u_s \leq \delta$  while our hypothesis predicts less polarization at this stage.

To sum up, our hypothesis is confirmed for  $u_s > \delta$  but not for  $u_s \leq \delta$ . Indeed, even if the observed fluctuations for  $u_s \leq \delta$  are on average significantly lower, the final states are not static for at least some replicas. How to explain these intra-group fluctuations? The dynamics on the main dimension is ruled by the classical bounded confidence model [8]. We know, from the study of this model done by [35] that some minor clusters appear sometimes between the major cluster and on the border of the attitude space. Their masses are around  $3 \cdot 10^{-4}$ . The hypothesis we can make is that the presence of such minor cluster is responsible for these fluctuations that we observe for  $u_s \leq \delta$ . Indeed, even if it includes a single individual, a cluster can strongly modify the global configuration of the population in the dynamics we are studying.

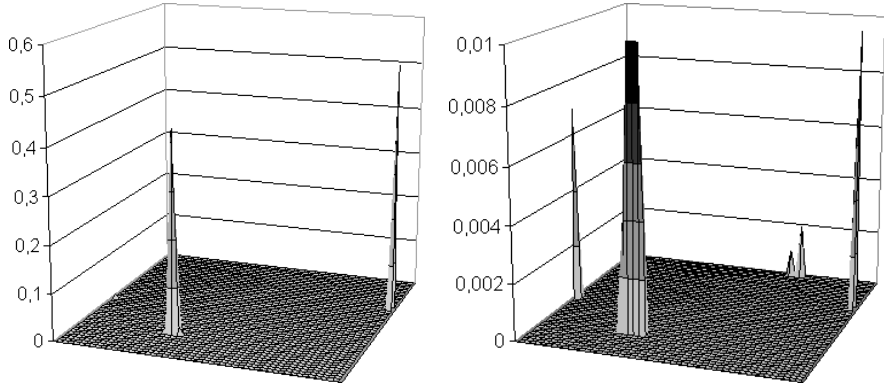


Fig. 11. Density of individuals on the attitude space at the equilibrium for a 1000-individual population for  $u_m = 0.4$ ,  $u_s = 0.8$ : on the left at the normal size; zoomed on the lower part of the graph on the right to be able to observe the minority clusters.

This is illustrated by Fig. 11, which shows a final state for  $u_m = 0.4$ ,  $u_s = 0.8$  on the left. On the right, there is a focus on the lower part of the diagram presented on the left. We can see on the right the presence of much smaller clusters. These small clusters reject the big ones, and increase the expected polarization. In other cases, their presence can generate fluctuations.

In a closer analysis of the simulations, we found a second important difference with the general hypothesis which led to the computation of  $\delta$ : a theoretical stable final state can be very difficult to reach in practise, and requires an extremely long time, during which fluctuations are observed. For instance, Fig. 12 presents the opinion clusters for a 1000-individual population after 7,500,000 iteration for  $u_m = 0.4$ ,  $u_s = 0.5$ . Each dot represents a cluster and the label indicates its size. We count four clusters on the figure. If we follow our first analysis, this configuration should be stable because  $u_s \leq (2/(4 - 1)) = \delta$ ). However, one can see on the figure that the two minor clusters are not stable because they are too close on the secondary dimension. In this configuration, the necessary space between the clusters on the  $u_s$  axis should be obtained by the progressive drift of the big clusters under the influence of the small ones: the cluster having size 1 pushes the one of size 11, which in turn, pushes the cluster of size 456, which itself pushes the cluster having a size of 532. The time required to get the stability is so long that it is difficult to observe it practically. Therefore, in some configurations, the minor clusters can induce fluctuations that remain for such a long time that we did not observe the final stable state in our experiments.

Figure 13 shows another example of this very long fluctuating transitory state for  $u_m = 0.1$ ,  $u_s = 0.1$ . We notice on the right that after 4,005,000,000 iterations, the clusters are far from occupying the whole space of the secondary dimension and remain in a conflicting situation. On the right, we see how the average opinion of the most extreme negative cluster slowly changes to find a stable position. It gives an idea of the time required to lead the stable state.



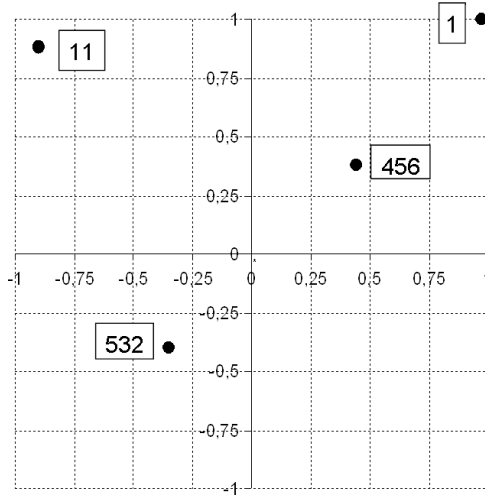


Fig. 12. Clusters of individuals in their attitude space for a 1000-individual population after 7,500,000 iterations for  $u_m = 0.4$ ,  $u_s = 0.5$ . Each dot represent a cluster and the labels indicate their size. The  $x$ -axis represents the main attitude, the  $y$ -axis the secondary attitude.

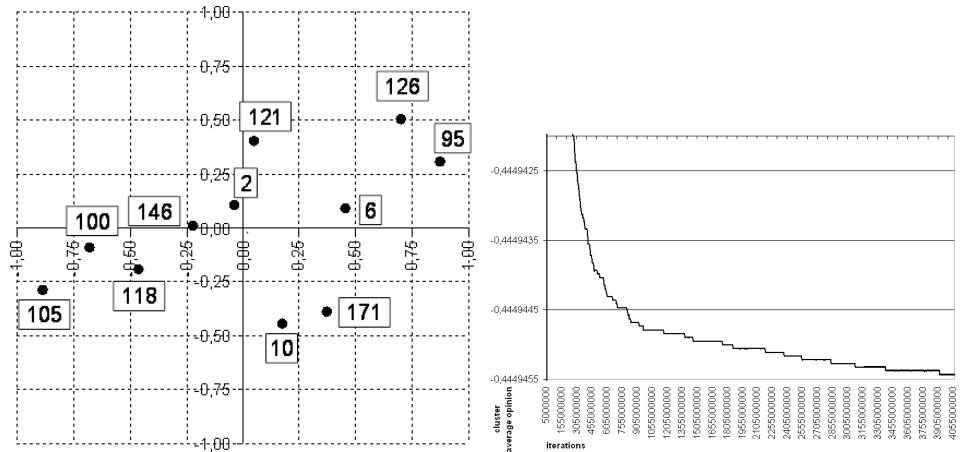


Fig. 13. On the left: Clusters of individuals in their attitude space for a 1000-individual population after 4,005,000,000 iterations for  $u_m = 0.1$ ,  $u_s = 0.1$ . Each dot represents a cluster and the labels indicate their sizes. The main attitude is represented in abscissa while the secondary is on the  $y$ -axis. On the right: Evolution of the average opinion on the least important dimension ( $y$ ) of the more extreme negative cluster regarding this dimension (it has the size 10).

#### 4.4. Size of the major clusters

Figure 14 shows on abscissa the tested values of the attraction and rejection thresholds. It presents the average size of the biggest and the smallest cluster (expressed in percentage of the population size). The graph also shows what would be the size of the clusters if this size is equal for all the clusters. This size is

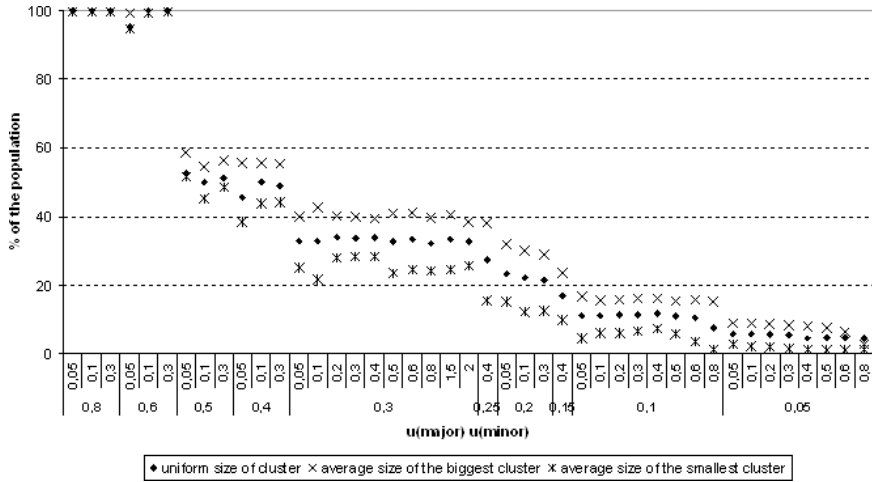


Fig. 14. Average size of the biggest and the smallest cluster and size of clusters if all clusters have equal size (it means that the size of a cluster corresponds to the size of the population divided by the average measured number of clusters larger than 1% of the population) for various values of  $u_m$  and  $u_s$  (all expressed in percentage of the population composed from 1000 individuals).

calculated considering the population size divided by the measured average number of clusters larger than 1% of the population. All are expressed in percentage of the population.

We observe that, for a given value of  $u_m$  and  $u_s$ , the major clusters are approximately all of the same size in the population. Indeed the “identical for all clusters” size is always at most more or less 4% of the size of the smallest and the largest cluster.

### 5. Discussion and Conclusion

The experimental results are in good accordance with our hypothesis. The number and the size of clusters are ruled by the bounded confidence dynamics on the main dimension. This behavior is not modified by the population size, as for the bounded confidence model. The measures on fluctuations and polarization confirmed our hypothesis for  $u_s > \frac{2}{(\frac{1}{u_m}-1)}$  and disconfirmed, in a finite time, for the other values of  $u_s$ . While we predict static and low polarized clusters for these latter values, we observe quite highly polarized clusters and intra-group fluctuations. This is due to the presence of minor clusters which imply a very long continuous rejection making the stable state reachable in a very, very long time, sometimes too long in practice. The model is ruled by the bounded confidence model on the main dimension and, for this model, minor clusters regularly appear [35,36] between the big clusters and on the border of the attitude space, as pointed out by [35, 36].

Now, the main discussion is about the interpretative potential of this model. Does its typical opinion evolutions fit observed stylized facts? The first typical

opinion evolution is obtained for  $u_s \leq \frac{2}{(\frac{1}{u_m} - 1)}$ , meaning for these values, the simulations are mainly ruled by the attraction process (we suppose here that the width of the attitudinal dimension is 2 because the attitudes take values between  $-1$  and  $+1$ ). Individuals begin to discuss and they quickly agree on the secondary attitudinal dimension: they all join the mid-position on this dimension. They act exactly as people who easily agree on details. Then, individuals form clusters on the main dimension. The most important aspects define each group as a unique entity. When the groups are sufficiently formed, individuals begin to reject each other on the secondary dimension due to their high distance on the main dimension. We get a behavior which reminds the results of experiments: when they belong to different groups (defined on the main attitude), individuals having the same attitude on secondary aspects reject each other. It is also very close to the process of group formation and the increase of the cohesion described by Turner in 1984 [34].

For these opinion evolutions, we can also observe a great effect of small minorities. They almost always exist even if they generally represent less than 1% of the population. They maintain some intra-group fluctuations in the major groups due to their rejection for a very long time. Depending on their attitude values, they can also push the major groups to slowly polarize more than they would do without these minor clusters. For a long time in social science, minorities had reputed having no effect on majority groups. It has now changed; they appear as a source of creativity and interrogation, a sort of openness or alternative. Indeed, they avoid too much stability and often shake the public debate. Our model may account for such dynamics.

The second typical opinion evolution occurs when the rejection threshold is significantly higher than the attraction threshold. For these values, the rejection process is dominant. The initial attraction on the secondary dimension is weak. Indeed, as the attraction on the main dimension is low, many individuals stay far from each other. The polarization on this dimension begins before the groups have been really formed. In this case, people are very narrow-minded about what is important for them. Thus they form a lot of groups. Moreover, individuals want to be very different from individuals of other groups on the secondary dimension. They socially define themselves by differentiation to others. It results on constant fluctuations on the secondary dimension.

These fluctuations remain in-group fluctuations on the secondary dimension if the rejection threshold is not too high. Groups are less cohesive on the least important dimension. Individuals continuously define themselves on this dimension by differentiation to the other groups. However, they define themselves as a member of their group on the important dimension.

When the rejection threshold becomes even larger, individuals remain in a continuous indecision and always fluctuate without being able to form a group on this secondary dimension. This can remind political regimes with a lot of small parties which are subject to frequent tactical changes of positions to differentiate from each other. However, this particular opinion evolution does not fit any observation from

the experiments we took as a source of inspiration. A deeper investigation in the socio-psychological work would help to determine if it can be related to precise observations.

Another effect of a very large rejection threshold is the creation of extremist groups. The cohesiveness and the stability of these groups depend on the attraction threshold. Individuals composing these groups fluctuate a lot on the secondary dimension when the attraction threshold is low, as mentioned previously. However, when the attraction threshold is large, these extremist groups are stable and it is the centrist groups which are less cohesive. This latter situation sounds more realistic. In the political domain, the groups of extremists are generally cohesive even on a question which does not define their groups, whereas, the more centrist groups are more likely to vary on questions which are not group-relevant.

Complementary investigations would be useful to check the robustness of these conclusions:

- The initial distribution of attitudes has an impact on the stationary state configuration, and using the uniform distribution is not the most realistic hypothesis. Testing other rules for initialization could be useful.
- We should also vary the speed of the attitude move (parameter  $\mu$ ), since we noticed that this parameter can have a strong impact on our first model with rejection [34] but also because it is probably able to suppress the minor clusters as suggested by [37].
- It would be worth considering a distribution of values for the thresholds instead of considering that all individual share the same values.
- We should study the model with more than two attitudes and determine the impact of a selective discussion (an individual has to choose what it wants to discuss).

## References

- [1] Axelrod, R., The Dissemination of Culture — A model with local convergence and global polarization, *Journal of Conflict Resolution* **41** (1997) 203–226.
- [2] Brehm, J. W., *A Theory of Psychological Reactance* (New York Academic Press, 1966).
- [3] Brewer, M. B., The Social Self: On Being the Same and Different at the Same Time, *Pers. Soc. Psychol. Bulletin* **17** (1991) 475–482.
- [4] Cacioppo, J. T. and Petty, R. E., The need for cognition, *Journal of Personality and Social Psychology* **42** (1982) 116–131.
- [5] Deffuant, G., Amblard, F., Weisbuch, G. and Faure, T., How can extremism prevail? A study based on the relative agreement interaction model, *Journal of Artificial Societies and Social Simulation* **5** (2002).
- [6] Deffuant, G. and Huet, S., Propagation effects of filtering incongruent information, *Journal of Business Research* **60**(8) (2006) 816–825.
- [7] Deffuant, G., Huet, S. and Amblard, F., An individual-based model of innovation diffusion mixing social value and individual payoff dynamics, *American Journal of Sociology* **110** (2005) 1041–1069.

- [8] Deffuant, G., Neau, D., Amblard, F. and Weisbuch, G., Mixing beliefs among interacting agents, *Advances in Complex Systems* (2001) 87–98.
- [9] Eagly, A. H. and Chaiken, S., *The Psychology of Attitudes* (Fort Worth, TX: Harcourt Brace Jovanovich, Thomson/Wadsworth ed, 1993).
- [10] Festinger, L., *A Theory of Cognitive Dissonance* (Stanford University Press ed, 1957).
- [11] Fortunato, S., Latora, V., Pluchino, A. and Rapisarda A., Vector Opinion Dynamics in a Bounded Confidence Consensus Model, *International Journal of Modern Physics* **16** (2005) 1535–1551.
- [12] Galam, S. and Moscovici, S., Towards a theory of collective phenomena: Consensus and attitude changes in groups, *J. Soc. Psychology* (1991) 49–74.
- [13] Hegselmann, R. and Krause, U., Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation, *Journal of Artificial Societies and Social Simulation* **5** (2002).
- [14] Heider, F., Attitudes and Cognitive Organization, *Journal of Psychology* (1946) 107–112.
- [15] Huet, S., Expérimentations en psychologie sociale et modèle dynamique de simulation. Réflexions ontologiques sur un exemple portant sur l'identification et la différenciation, presented at Ontologies et dynamiques des systèmes complexes, perspectives interdisciplinaires, Rochebrune 2009.
- [16] Huet, S. and Deffuant, G., When do Interactions Increase or Decrease Primacy Effect? presented at Model-2-Model, Marseille, 2007.
- [17] Huet, S., Deffuant, G. and Jager, W., Rejection Mechanism in 2D Bounded Confidence Provides More Conformity, *Advances in Complex Systems* **11** (2008) 1–21.
- [18] Jager, W. and Amblard, F., Uniformity, Bipolarization and Pluriformity Captured as Generic Stylized Behavior with an Agent-Base Simulation Model of Attitude Change, *Computation & Mathematical Organization Theory* (2004) 295–303.
- [19] Jager, W. and Amblard, F., Multiple Attitude Dynamics in Large Populations. Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms, 2005.
- [20] Jager, W. and Amblard, F., Guess you're right on this one too: Central and Peripheral processing in attitude changes in large populations. 1st World Conference of Social Simulation, Kyoto, 2006.
- [21] Latané, B., Dynamic Social Impact: The Creation of Culture by Communication, *Journal of Communication* **46** (1996) 13–25.
- [22] Lorenz, J., About the power to Enforce and Prevent Consensus by Manipulating Communication, Rules *Advances in Complex Systems* (2007) 19.
- [23] Mason, W. A., Conrey, F. R. and Smith, E. R., Situating Social Influence Processes: Dynamic, Multidimensional Flows of Influence Within Social Networks, *Personality and Social Psychology Review* (2007) 279–300.
- [24] Moschis, G. P., Social Comparison and Informal Group Influence, *Journal of Marketing Research* **13** (1976) 237–244.
- [25] Price, L. L. and Feick, L. F., The role of interpersonal sources in external search: An informational perspective, *Advances in Consumer Research* (1984) 250–253.
- [26] Rogers, R. W. and Prentice-Dunn, S., Protection motivation theory, in *Handbook of Health Behavior Research*, Vol. 1, D. Gochman, Ed.: New York:Plenum, (1997) 113–132.
- [27] Salzarulo, L., A Continuous Opinion Dynamics Model Based on the Principle of Meta-Contrast, *Journal of Artificial Societies and Social Simulation* **9** (2006).
- [28] Sherif, M. and Hovland, C. I., *Social Judgement* (Yale University Press, 1961).
- [29] Tajfel, H., *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations* (London: Academic Press, 1978).

- [30] Turner, J. C., Hogg, M., Oakes, P. J., Reicher, S. D. and Wetherell, M. S., *Rediscovering the Social Groups: A Self-Categorization Theory* (Cambridge, MA: Basil Blackwell, 1987).
- [31] Urbig, D. and Malitz, R., Dynamics of structured attitudes and opinions. Third Conference of the European Social Simulation Association, September 5–8 (2005) Koblenz, Germany.
- [32] Urbig, D. and Malitz, R., Drifting to more extreme but balanced attitudes: Multidimensional attitudes and selective exposure, ESSA, Toulouse, 2007.
- [33] Wood, W., Pool, G. J., Leck, K. and Purvis, D., Self-definition, defensive processing, and influence: the normative impact of majority and minority groups, *Jour. of Personality and Social Psychology* (1996) 1181–1193.
- [34] Turner, J.C., Social identification and psychological group formation. In H. Tajfel (Eds). *The Social Dimension*, Vol. 2 (Cambridge: University Press, 1984).
- [35] Ben-Naim, E., Krapivsky, P. L. and Redner, S., Bifurcations and Patterns in Compromise Processes, *Physica D: Nonlinear Phenomena* **183**(3–4) (2003) 190–204.
- [36] Lorenz, J., Continuous Opinion Dynamics Under Bounded Confidence: A Survey, *International Journal of Modern Physics C* **18**(12) (2007) 1819–1838.
- [37] Laguna, M. F., Abramson, G. and Zanette, D. H., Minorities in a Model for Opinion Formation, *Complexity* **9**(4) (2004) 31–36.



# Conclusions and perspectives

## A methodological point of view on data and design

The present conclusion proposes a slightly different view of my work, transversal to the two types of modelling composing the main parts of this document. This is a much more methodological view pointing out some particular issues I found interesting about data and design.

We distinguish the issue of using data for the design of a data-driven modelling from the one for theoretical modelling. Various difficulties and cautions are pointed out for each type of model.

## Data and design in data-driven modelling

### *Using quantitative and qualitative data*

If data is often considered as a measure in laboratory or through surveys, qualitative data can also be useful to design a model if their description is precise enough.

A description of a theory, the hypotheses of an experiment as well as the questions of a survey can be considered as qualitative data.

Indeed we can design starting from theories as we did in chapter 2.3 trying to build dynamics from their common elements. We can also, as in chapter 2.5, start from the confirmed hypotheses of an experiment in social psychology to design the individual dynamics. In a data-driven model such the one presented in part 1, surveys are used as the main inspiration for designing. Quantitative data from a survey as the census contributes largely to the design of the model. That is the case for example with the generation of the initial population presented in chapter 1.2.

Qualitative data contained in surveys are also useful. Indeed, not only the resulting data from the answer to the questions, but also the questions themselves are useful to design a dynamics model. They give indications about the underlying hypothesis about dynamics made by the designer of the survey. That is for example the case in France for the survey carried out by the Farming Accountancy Network of Information, which mainly contains questions about the farming practices and the monetary equilibrium of the farm of the interviewed farmers. On the other hand, a survey such as the one realised by (Cautres and Mallein 1993) is much more interested by the farmer characteristics than by the farm: it aims at designing the values of the farmers from three French regions. From both of these surveys, typologies have been extracted and compared considering the net outcome of the farms. (Cautres and Mallein 1993) show the classes of their typology mainly based on the farmer's characteristics are much more correlated to the net outcome of the farms than the typology based on the characteristics of the farm (the one of the Farming Accountancy



Network of Information). However, it is the Farming Accountancy Network of Information which is used to assess impact of new policies in agriculture. This example clearly shows that some hypothesis about drivers underlies the surveys and the questions of the survey: some can consider geographical constraints and type of farming as responsible for the outcome while others can consider the characteristics of the farmers and his management choices are responsible.

Then, choosing one or another survey as a source of data, also choosing a particular set of questions in a survey instead another set correspond to a design issue for the modeller. This is the case for example with the Labour Force Survey containing many questions related to the individual status and dynamics on the labour market. Using this survey to design the working activities of individuals from SimMunicipalities, we choose in the chapter 1.3 not to retain questions regarding salaries. Then, while many labour dynamics model in economy use the salary and suppose people aim at increasing their revenue, we prefer using the simple question “do you search for a job?” considering people looks at first for a job and not necessary for a greater salary. This hypothesis appears relevant in France where there is a minimum wages. Moreover, wages in rural areas are known as lower than in urban areas and our model is interested only in rural areas.

In data-driven modelling, the design issues embedded in choice of surveys and questions are important to identify. Indeed, contrary to the theories and experiments in which hypotheses are explicit, in surveys, hypotheses on drivers are often more or less hidden.

### *Integrating the data availability at the early stage and ensuring the consistency*

We have seen in the first part the importance of data in data driven models. The conclusion of chapter 1.2 stresses out the necessity to collect various data sources at an early stage, to integrate knowledge about data in the design of the model. Indeed, all the main design choices (e.g. main objects, calibration process, parameterization for a particular application) should take into account the availability, the definition, the representativeness of the data ... It is not only to ensure that some data are available for the validation, it also ensures a consistency of the chosen concepts and their link to data through the different phases of the development of the model. As an example, the initial controlled<sup>19</sup> attributes of the population have to be the same as the ones impacted by the chosen dynamics or the ones chosen to calibrate the model. Also, the attributes of individuals should be consistent in all their uses in the model; for example, the profession of an individual should be defined consistently with the job offers of municipalities. The modeller has to design the whole having in mind what data can be used and sometimes constrain the design to their availability.

---

<sup>19</sup> A controlled attribute means the value of the attribute is designed to be closer as possible from the reference value measured through survey. That is for example the number of people being a given age or the number of inhabitants of a given municipality.

## *Difficulties of the individual level coupled to a large scale approach*

From another point of view, we want to stress out a particular difficulty linked to the individual-based approach associated to a large scale spatial approach. The chapter 1.3, dedicated to the activity dynamics of the individual and the labour market gives an example. Indeed a shortcoming of our approach is probably linked to the level of decomposition in jobs (i.e. the number of jobs considered) compared to the geographical level we implement. To be clearer, we have to consider 24 jobs in order to have a sufficiently precise image of the various professions and activity sectors of the implemented region, in our case study. At the same time, the LFS is initially designed to be representative at the national level. Therefore, to extract transition probabilities over the different statuses and professions considering a sufficiently large number of individuals concerned by the transition, we have to extract the probability transitions from the French regions. But it is possible that these probabilities are different in Cantal. Cantal is a *department* where farming activities remain high and even if the number of farmers decreases overall, the probability to become a farmer seems to be higher on average than in the other *départements*. Moreover, it seems that a large number of farmers have a double activity and it cannot be ruled from the LFS because the sample for each couple of double activities is not sufficient to extract transition probabilities.

The theoretical modelling approach generally does not suffer from such a problem. Indeed, the chosen concept, objects and dynamics are generally too abstract or generic to relate to a particular measurable data. However, it is possible as presented in the next section to use qualitative data to guide the design of the model.

## **Data and design in theoretical modelling**

### *A rich openness*

The model of chapter 2.3 is inspired from psycho-sociology theories whereas the model of chapter 2.5 is inspired from the results of some experiments. This kind of approach is encouraging to build models that incorporate the detailed, micro level understanding of influence processes derived from focused laboratory studies in social science (Mason, Conrey et al. 2007). Similarly (Sobkowicz 2009) argues that it would be a way to improve the value of the research in opinion dynamics.

An investigation in the social psychology studies is a source of inspiration, yet at the same time it constrains creation. It is a useful work to go beyond the intuitive psychology which is said to be naïve and sometimes misleading by social scientists. It is also a source of innovation. This is the case for example in the model presented in chapter 2.5, inspired from experiments and which considers in some particular case that the attraction between two individuals is unbounded. Such an unbounded attraction has not been envisaged in the model presented in chapter 2.3 since it is not expressed in theories: only experimental descriptions allow for identifying such a dynamic.

An interdisciplinary approach shows however some difficulties, because the different sciences (social psychology and research on social individual-based modelling) have different purposes. For example, social psychology often aims to test the effect of a particular global

context<sup>20</sup> on the individual attitude, while the social dynamics modeller aims at determining the effect that a particular individual attitudinal dynamics has on the population collective behaviour (which can be measured for example in terms of number of groups of opinion).

### *Many interpretations and simplifications*

Trying to design models remaining close to social psychological experiments outlines how much modelling implies interpretation and simplification. Let's take the model presented in chapter 2.5 as an example.

Firstly, we study group formation from initially scattered opinions, whereas the experiments took place in the context of already existing groups. Thus, in our model, we consider pair interactions only (we do not consider the group directly). We reinterpret the experiment by considering two individuals who can share or not fundamental values, and influence each other on less important (secondary) opinions.

Secondly, in the experiment, the individual is told that a given group (which is highly positively or negatively self-relevant for him) said something with which the individual is in agreement or disagreement. The high self-relevance means the individual defines himself through being a member of this group (positive self-relevance) or through not being a member of this group (negative self-relevance). In the model we consider two attitudinal dimensions. One is for fundamental values (called main dimension) which are supposed to be the values of the highly self-relevant group. The other attitudinal dimension is the secondary one which is supposed to represent the issue on which the group communicates in the experiment.

On each dimension in the model, we suppose that the opinions can take continuous values, between -1 and +1. That is not so far from the experiment in which the attitude is measured through a scale going from negative to positive discrete values.

Then, we introduce influence dynamics:

- In the experiment, a disagreement is a large difference of attitude between the implicit attitude diffused by a message delivered by a group source and the known attitude of the participant; an agreement is a small difference or no difference. In the model, the disagreement also corresponds to a large difference between a source (an interlocutor) and an individual during a pair meeting; a threshold is assumed to define if the difference is large enough to be considered as a disagreement. If the difference is lower than the threshold, an agreement is diagnosed by the virtual individual.
- In the experiment, a highly positively self-relevant group source is a very important group for the participant: the participant is a member of this group and strongly socially defines himself as a member of this group; a negatively self-relevant group

---

<sup>20</sup> Indeed, the context is often manipulated in experiments in social psychology. That can be for example the fact the participant is a member or not of a given group and that the group has a particular message defending a position pro or con a particular issue. This group can be a major group or a minor group, ...

source is a group which is also very important but the participant defines socially himself as not a member of this group. We assume in the model the individual makes the values of his group his own. We assume these values can be embedded in an attitudinal dimension called main dimension for the individual.

- The message delivered by the group source in the experiment is supposed about a discussed issue. The attitude about this issue is supposed to be the secondary attitudinal dimension of our virtual individual. Then, during a pair meeting, an individual tells about her secondary attitudinal dimension (the discussed issue) and about her main attitudinal dimension (her membership).
- In the model, when two individuals are close of the main dimension, they are supposed to be members of the same group and, as in the experiment with the group source, they become closer in any case of the secondary dimension. The model also assumes if they are closer than the threshold on the main dimension, they become closer on this main dimension, as well as on the secondary dimension. This is not in the experiment but we have to remind groups don't exist at the initial time in our model while they exist in the experiment.
- In the model, when two individuals are further than the threshold on the main dimension, they are supposed to be members of an hated group and, as in the experiment, shift away from each other.

We can see from this example that using experiments as a source of inspiration for design requires a lot of hypothesis.

Data, whatever its nature, virtual, qualitative or quantitative, measured during experiments in laboratory or via a survey, can also be used as a basis for studying the model by comparison.

## Perspectives

Two different sets of perspectives motivate me. A first set aims at filling some gaps between the different modelling approaches. I present some ideas in this direction in the next section. The second set is more prospective and aims at maintaining and developing my knowledge in social dynamics. It consists in developing collaborations with social psychologists and investigating new virtual social dynamics. These two sets appear to me as complementary since they allow me to look at the social system from two different viewpoints.

## Developing the loop between data-driven and theoretical modelling

We have seen in the literature that the residential mobility dynamics remains poorly known. The model we designed and selected for its ability to produce results close to the reference, is very simple. It considers a global bounded diffusion of individuals over the space at the same time some particular locations partially reject people. This dynamics seems to have an important impact on the evolution of the spatial distribution of individuals. Despite this importance, we finally know very few on this dynamics and this impact. Then, two complementary studies should be relevant:

- A theoretical one considering this dynamics in isolation from the others with which it interacts; it aims at collecting the asymptotic behaviour and trajectories of the dynamics in the most simple and generic case ;
- A more practical one, done in the data-driven model, considering how this dynamics interacts with others without reducing the study to the case where the model produces results close to the reference. In practice, the aim is to understand how the asymptotic behaviours of the model can be affected by its interaction with other dynamics.

Then, both the asymptotic behaviours and their sensitivity to other dynamics can be used as criteria to choose a dynamics to model an unknown law of another case study. For example, if we want to implement another case study region with SimMunicipalities, and if the spatial pattern of the population evolution is the same, we can certainly use the same chosen laws as the one chosen for Cantal in chapter 1.4. But if we imagine the spatial pattern of inhabitant evolution is different, we can't only decide from the knowledge obtained from Cantal implementation. On the contrary, a theoretical study of trajectories and final states of the moving laws chosen for Cantal would allow the modeller to conclude if the spatial pattern of the new case study can be probably model with these same moving laws. Indeed, the modeller knows if a state (or trajectory) observed from his new case study are close to one state of the collection of equilibrium states (or trajectories) given by the theoretical study. For sure, the choice should be checked for its ability to produce a result close to a reference of the new case study. It is especially true when the implemented region remains very particular compared to what is already known about this dynamics. If the model does fit the reference, the whole process (i.e. proposing and checking a model variant – starting from the existing one – and study it theoretically) can be done again. In any case, enlarging the criteria of choice would increase its relevance since it is not limited to the particular implemented application.

We notice from this perspective that the gap between theoretical and data-driven modelling is not so big and both models are complementary. But going from data-driven modelling to theoretical then back to data-driven modelling is not the only interesting loop. To start from theoretical modelling in order to gain knowledge from a data-driven approach appears also very relevant to me.

It was for example the case of the work we did on the filtering or the rejection of a particular attitude. We now know well the asymptotic behaviour of these models. On the other hand, we have injected in our data-driven model of municipality evolution a very simple and debatable model for the decision to change residence based on a satisfaction given by the average age of the adults of the household and the size of the residence compared to the size of the household. It would be particularly interesting to consider a satisfaction based on attitudes. Several dimensions can be considered such as the social attitude about the distance to the job, to the school but also global comfort of living in this particular place. It is possible that a survey such as the European Household Panel gives indications and data on the possible dimensions to take into account.

## Social dynamics

### *Develop the interdisciplinary work*

We have seen previously that, even if encouraged, collaboration between modellers and social scientists is not so simple, particularly because of the different purposes of these research domains. However, the collaboration could be certainly organised and I propose a first loop for a multidisciplinary study between individual-based social dynamics modellers and social psychologists.

The modeller often uses the research results from the social scientists as an assumption in his models. He checks that he obtains, for some parameter values, the same emerging collective states as the ones assumed by the social scientist. Then, after having studied the model, the modeller can give a complete collection of collective states, some of them that the social scientist might not have envisaged. When submitted to experiments by the social scientist, do these collective states always lead to conclude about the same individual dynamics? If not, a new dynamics has probably to be investigated by the modeller, if possible designed in common by the modeller and the social scientist.

Of course, as usual, this type of collaboration must overcome the difference of culture and languages between the scientific disciplines. For example, the modeller is concerned by the interactions between individuals while the social scientist generally tells the participant of an experiment that a given group have said something in a given situation.

### *Investigating new dynamics*

Regarding theoretical modelling of social dynamics, I would like to investigate two questions which seem complementary to me: how do groups change over the time, how do they disappear, appear, increase or decrease? And how is the self-esteem of interacting individuals responsible for a human organisation?

In part 2 of this document, I study two mechanisms allowing the individuals to resist social influence in order to preserve an internal stability. We have seen that this desire for stability can lead to a global instability and more or less large fluctuations inside groups when we add a rejection mechanism and when the attitudinal space is bounded. It is due to interactions, which can also be responsible for an increase or a decrease of the rationality of the individuals. It seems that we capture a part of the whole system dynamics: it is an assumption on how an individual can make the whole system instable and perhaps close to change since more sensitive to new events. However we have not proposed the counterpart allowing the system to become stable again, probably without passing through the totally instable state. Indeed, we can suppose some individual dynamics allow a transition occur when the system tend to become instable, in order to make it stable again, possibly in another equilibrium state. Such an individual dynamics remains to define.

Thus, in the future, it would be interesting to follow up this work by developing a more general model considering the various actions an individual use to protect herself from an

internal discomfort. Following (Matz and Wood 2005), three strategies can be chosen to reduce the individuals' discomfort created by the heterogeneous opinions inside its group:

1. changing one's own opinion to agree with others in the group,
2. influencing others to change their opinions,
3. or joining a different, attitudinally more congenial group.

The first two relate to the classical tendency to conformity expressed by modellers in many opinion dynamics models (such as the bounded confidence). The third has not been envisaged in the presented works but can also appear as a solution leading to a well organised social space as we have shown (Gargiulo and Huet 2010; Gargiulo and Huet 2012). It seems to us that a model including these three mechanisms would be very interesting.

Confronting the results of such an attitude dynamics model with the typologies of values established by sociologists would also be very interesting. Indeed, values, which are considered as our most stable attitudes (Rokeach 1968) have been collected and measured by various researchers (Schwartz 1999; Dietz, Fitzgerald et al. 2005) to define specific culture of nations or continents. It is in my view a promising way to use data for modelling social dynamics.

The value comes logically to me talking about the second question I want to investigate: the self probably embedded in our more stable attitudes.

From a more individual point of view, we have seen that the social psychology literature points out that some specific behaviours are adopted when someone feels her identity attacked. It is the case with rejection occurring only if the individual is highly involved (Pool, Wood et al. 1998). It is a central point of the minority theory of Moscovici but it also motivates a large part of the great interest for the self in social science (Moscovici 1979; Hoorens 1993; Monteil 1993; Hoorens 2011).

As a first step in this direction, altogether with Guillaume Deffuant and Timoteo Carletti, I am involved in a new opinion dynamics adventure considering a population of individuals talking in pairs about themselves. An individual has an attitude about every other and is more influenced by those she respects (meaning her attitude about them is higher than her attitude about herself). This first mechanism defines how the self allows individuals ranking their peers. The second mechanism, called vanity but which can be seen as a protection-of-the-self mechanism, also influences this ranking. Indeed, if the individual is undervalued by her interlocutor (meaning that the interlocutor's attitude about him is lower than her attitude about herself), she decreases her attitude about her interlocutor. If her interlocutor overvalues her, then she increases her attitude about her interlocutor. This model produces several patterns. As far as I know, it is the only model in which the leadership really emerges; it is not at all exogenously defined (Deffuant, Carletti et al. 2012).

From a more practical point of view, the self-esteem has been identified as responsible for a lot of individual behaviours. For example the self-esteem has been proved a good predictor of job search outcomes and job search motivation and satisfaction (Ellis and Taylor 1983). It is also very probable for instance that unemployment has a particular impact on the self-

esteem. Then, taking into account such a loop to determine the level of effort of unemployed individuals of our SimMunicipalities model to search for a job would be interesting. This is another example of possible synergies between theoretical and data driven models.





## References

- Abelson, R. P. and J. C. Miller (1967). "Negative persuasion via personal insult." *Journal of Experimental Social Psychology* 3(4): 321-333.
- Anderson, N. H. (1971) "Integration theory and attitude change". *Psychological Review*, 78, 171–206.
- Asch, E. S. (1946). "Forming Impressions of Personality." *Journal of Abnormal and Social Psychology*(41): 258-290.
- Aubert, F., J. C. Dissart, et al. (2009). "Facteurs de localisation de l'emploi résidentiel en France." XLVIème Colloque de l'Association de Science Régionale de Langue Française (ASRDLF) 6-8 juillet, Clermont-Ferrand, France: 27.
- Bacon, B. and S. Pennecc (2007). APPSIM - Modelling Family Formation and Dissolution. N. N. C. f. S. a. E. Modelling, Universit of Canberra: 30.
- Balcan, D., H. Hu, et al. (2009). "Seasonal transmission potential and activity peaks on the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility." *BMP Medicine* 2009: 7-45.
- Ballas, D., G. P. Clarke, et al. (2007). "Using SimBritain to Model the Geographical Impact of National Government Policies." *Geographical Analysis*(39): 44-77.
- Ballas, D., G. P. Clarke, et al. (2005). "Building a Dynamic Spatial Microsimulation Model for Ireland." *Population, Space and Place* 11: 157-172.
- Ballas, D., G. P. Clarke, et al. (2006). "Spatial Microsimulation for Rural Policy Analysis in Ireland: The Implications of CAP Reforms for the National Spatial Strategy." *Journal of Rural Studies* 22: 367-378.
- Ballas, D., R. Kingston, et al. (2007b). "Building a spatial microsimulation-based planning support system for local policy making." *Environment and Planning A* 39: 2482-2499.
- Baqueiro Espinosa, O., I. Unay-Gailhard, et al. (2011). Two adaptations of a Micro-Simulation Model to Study the Impacts of Policies at the Municipality Level. Working paper. P. E. project. IAMO, Newcastle University, Cemagref: 62.
- Barabasi, A. L. and R. Albert (1999). "Emergence of Scaling in Random Networks." *Science* 286(5439): 509-512
- Ben-Naim, E., P. L. Krapivsky, et al. (2003). "Bifurcations and Patterns in Compromise Processes." *Physica D: Non linear phenomena* 183(3-4): 190-204.
- Berger, T. and P. Schreinemachers (2006). "Creating Agents and Landscapes for Multiagent Systems from Random Sample." *Ecology and Society* 11(2): 19.
- Birkin, M. and M. Clarke (2011). Spatial Microsimulation Models: A Review and a Glimpse into the Future. *Population Dynamics and Projection Methods*. J. Stillwell and M. Clarke, Springer Netherlands. 4: 193-208.
- Birkin, M., A. Turner, et al. (2006). A synthetic demographic model of the UK population: methods, progress and problems. Second International Conference on e-Social Science.
- Birkin, M. and B. Wu (2012). A review of Microsimulation and Hybrid Agent-Based Approaches. *Agent-Based Models of Geographical Systems*. A. J. Heppenstall, A. T. Crooks, L. M. See and M. Batty, Springer: 51-68.
- Blanc, M. and B. Schmitt (2007). "Orientation économique et croissance locale de l'emploi dans les bassins de vie des bourgs et petites villes." *Economie et Statistique* 402: 57-74.
- Blanchet, D., S. Buffeteau, et al. (2011). "Le modèle de microsimulation Destinie 2 : principales caractéristiques et premeirs résultats." *Economie et Statistique* 441-442: 101-122.
- Blanchet, D. and J. M. Chanut (1998). "Projeter les retraites à long terme. Résultats d'un modèle de microsimulation." *Economie et Statistique* 315(95-106).
- Boman, M. and E. Holm (2005). *Multi-Agent Systems, Time Geography, and Microsimulations*

- Systems Approaches and Their Application. M.-O. Olsson and G. Sjöstedt, Springer Netherlands: 95-118.
- Bonny, S. (2003). "Why are most Europeans opposed to GMOs? Factors explaining rejection in France and Europe." *Electronic Journal of Biotechnology* 6(1 - April, 15): 50-71.
- Bonny, S. (2008). "How have opinions about GMOs changed over time? The situation in the European Union and the USA." *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* 3(93): 3-17.
- Bousquet, F. and C. Le Page (2004). "Multi-Agent Simulations and Ecosystem Management: a review." *Ecological Modelling* 176: 313-332.
- Bozon, M. and F. Héran (1987). "La découverte du conjoint : I. Evolution et morphologie des scènes de rencontre." *Population* 42(6): 943-985.
- Bozon, M. and F. Héran (1988). "La découverte du conjoint : II. Les scènes de rencontre dans l'espace social." *Population* 43(1): 121-150.
- Brehm, J. W. (1966). *A theory of psychological reactance*, New York Academic Press.
- Brewer, M. B. (1991). "The Social Self: On Being the Same and Different at the Same Time." *Pers. Soc. Psychol. Bulletin* 17(5): 475-482.
- Brown, D. G., R. Aspinall, et al. (2006). "Landscape Models and Explanation in Landscape Ecology - A Space for Generative Landscape Science." *The Professional Geographer* 58(4): 369-382.
- Castellano, C., S. Fortunato, et al. (2009). "Review of Modern Physics." <http://arxiv.org/abs/0710.3256> 81(April-June): 591-645.
- Cautres, B. and P. Mallein (1993). *Les identités professionnelles des agriculteurs : les "profils de valeurs"*. Tele Promotion Rurale Rhône-Alpes Massif Central, Université des Sciences Sociales: 189.
- Champetier, Y. (2000). "The (re)population of rural areas." *Leader Magazine* 23(Spring. Special Issue).
- Ciofi degli Atti, M. L., S. Merler, et al. (2008). "Mitigation measures for pandemic influenza in Italy: an individual based model considering different scenarios." *PLoS One* 2008(3(3)): e1790.
- Colizza, V., A. Barrat, et al. (2007b). "Modeling the world-wide spread of pandemic influenza: baseline case and containment interventions." *PLoS Med* 4(1): e13.
- Colizza, V., A. Barrat, et al. (2007a). "Predictability and epidemic pathways in global outbreak of infectious diseases: the SARS case study." *BMC Medicine* 2007: 5-34.
- Communities), E. S. O. o. t. E. (1999). "Demography, fecundity 1999." [epp.eurostat.ec.europa.eu/portal/page/portal/population/data/database](http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/database).
- Coulombel, N. (2010). *Residential Choice and Household Behavior: State of the Art*, Ecole Normale Supérieure de Cachan. Working Paper 2.2a: 69.
- Davezies, L. (2009). "L'économie locale "résidentielle"." *Géographie, économie, société* 11(1): 47-53.
- David, B. and J. C. Turner (1996). "Studies in self-categorization and minority conversion: Is being a member of the out-group an advantage?" *British Journal of Social Psychology*(35): 179-199.
- De Angelis, D. L. and L. J. Gross (1992). *Individual-based models and approaches in ecology*, Chapman & Hall.
- Debrand, T. and C. Taffin (2005). "Les facteurs structurels et conjoncturels de la mobilité résidentielle depuis 20 ans." *Economie et Statistique* 381-382: 125-146.
- Debrand, T. and C. Taffin (2006). "Les changements de résidence: entre contraintes familiales et professionnelles." *Données sociales. La Société Française*: 505-513.
- Deffuant, G. (2004). *Modéliser la complexité: quelques pistes pour relever le défi*. M. d. H. à. D. d. R. d. I. E. d. SPI. Clermont-Ferrand, France, Université Blaise Pascal: 133.
- Deffuant, G. (2006). "Comparing Extremism Propagation in Continuous Opinion Models." *Journal of Artificial Societies and Social Simulation* 9(3).
- Deffuant, G. and e. al. (2001). "Rapport final du projet FAIR 3 2092 IMAGES : Modélisation de la diffusion de l'adoption de mesures agri-environnementales par les agriculteurs (1997-2001).".

- Deffuant, G., F. Amblard, et al. (2000). Simulation de l'évolution des conversions à l'agriculture biologique dans le département de l'Allier entre 1994 et 1999 par un modèle multi-agents de la population d'agriculteurs, Cemagref. Rapport intermédiaire du projet Européen IMAGES: 21.
- Deffuant, G., F. Amblard, et al. (2002). "How can extremism prevail ? A study based on the relative agreement interaction model." *Journal of Artificial Societies and Social Simulation* 5(4).
- Deffuant, G., T. Carletti, et al. (2012). "The Leviathan model: Absolute dominance, generalised distrust and other patterns emerging from combining vanity with opinion propagation." *Journal of Artificial Societies and Social Simulation*: 23.
- Deffuant, G. and S. Huet (2007). "Propagation effects of filtering incongruent information." *Journal of Business Research* 60(8): 816-825.
- Deffuant, G. and S. Huet (2009). "Collective increase of first impression bias." *Complexity* 9999(9999): 1-9.
- Deffuant, G., S. Huet, et al. (2005). "An individual-based model of innovation diffusion mixing social value and individual payoff dynamics." *American Journal of Sociology* 110 (January)(4): 1041-1069.
- Deffuant, G., S. Huet, et al. (2002). Agent-based simulation of organic farming conversion in Allier département. Agent-based simulation of organic farming conversion in Allier département. M. A. Janssen. Cheltenham, UK - Northampton, USA, Edward Elgar Publishing Limited: 158-187.
- Deffuant, G., D. Neau, et al. (2001). "Mixing beliefs among interacting agents." *Advances in Complex Systems*(3): 87-98.
- Deffuant, G., S. Skerrat, et al. (2008). An agent based model of agri-environmental measure diffusion: what for? *Agent Based Modelling in Natural Resource Management*. A. Lopez Paredes and C. Hernandez Iglesias. Universidad de Valladolid, INSISOC: 55 - 73.
- Deffuant, G. and G. Weisbuch (2007). Probability distribution dynamics explaining agent model convergence to extremism. *Social Simulation. Technologies, Advances, and New Discoveries*. B. Edmonds, C. Hernandez and K. G. Troitzsch, Information Science Reference, Hershey, USA: 43-60.
- Deming, W. E. and F. F. Stephan (1940). "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known." *Annals of Mathematical Statistics* 11(4): 427-444.
- Dietz, T., A. Fitzgerald, et al. (2005). "Environmental values." *Annual Reviews of Environment and Resources*(30): 335-372.
- Djefal, S. and E. S. (2004). "Etre propriétaire de sa maison, un rêve largement partagé, quelques risques ressentis." *CREDOC - Consommation et modes de vie* 177: 4.
- Duboz, R., E. Ramat, et al. (2003). Scale transfer modelling : using emergent computation for coupling an ordinary differential equation system with a reactive individual model. *Systems Analysis Modelling Simulation*. T. a. F. eds. 43: 793-814.
- Dubuc, S. (2004). "Dynamisme rural: l'effet des petites villes." *L'Espace Géographique* 1: 69-85.
- Duée, M. (2005). La modélisation des comportements démographiques dans le modèle de microsimulation Destinie. INSEE - Direction des Etudes et Synthèses Economiques. Malakoff (France), INSEE: 37.
- Eagly, A. H. and S. Chaiken (1993). *The Psychology of Attitudes*.
- Edwards, M. (2004). Intérêt d'un modèle agrégé pour étudier le comportement et simplifier la simulation d'un modèle individu-centré de consommation couplé à un modèle de ressource en eau Informatique, Université Paris VI Pierre et Marie Curie.
- Edwards, M., N. Ferrand, et al. (2005). "The relevance of aggregating a water consumption model cannot be disconnected from the choice of information available on the resource." *Simulation Modelling Practice and Theory* 13(4): 287-307.

- Edwards, M., S. Huet, et al. (2003). "Comparing individual-based model of behaviour diffusion with its mean-field aggregated approximation." *Journal of Artificial Societies and Social Simulation* 6(4).
- Ellis, R. A. and M. S. Taylor (1983). "Role of self-esteem within the job search process." *Journal of Applied Psychology* 68(4): 632-640.
- Erdős, P. and A. Rényi (1960). *Random Graphs*. London 1985, B. Bollobas, Academic Press.
- Eubank, S., H. Guclu, et al. (2004). "Modeling disease outbreaks in realistic urban social networks." *Nature* 429: 180-184.
- Fahse, L., C. Wissel, et al. (1998). "Reconciling Classical and Individual-Based Approaches in Theoretical Population Ecology: A Protocol for Extracting Population Parameters from Individual-Based Models." *The American Naturalist* 152(6): 838-852.
- Felemou, M. (2011). *Analyse de données de flux de navetteurs. Extractions de modèles*. Aubière, Cemagref: 31.
- Fernandez, L. E., D. G. Brown, et al. (2005). "Characterizing location preferences in an exurban population: implications for agent-based modeling." *Environment and Planning B: Planning and Design* 32(6): 21.
- Festinger, L. (1954). "A theory of social comparison processes." *Human Relations* 7(2): 117-140.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*.
- Fishbein, M. and I. Ajzen (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*, Reading MA: Addison-Wesley.
- Flache, A. and M. Mäs (2008a). "Why do faultlines matter? A computational model of how strong demographic faultlines undermine team cohesion." *Simulation Modelling Practice and Theory* 16(2): 175-191.
- Flache, A. and M. Mäs (2008b). "How to get the timing right. A computational model of the effects of the timing of contacts on team cohesion in demographically diverse teams." *Computational & Mathematical Organization Theory* 14(1): 23-51.
- Fontaine, C. and M. Rounsevell (2009). "An agent-based approach to model future residential pressure on a regional landscape." *Landscape Ecology* 24(9): 1237-1254.
- Forgas, J. P. and K. D. Williams (2001). *Social Influence: Introduction and Overview*. Social Influence - Direct and Indirect Processes. J. P. Forgas and K. D. Williams, Psychology Press - Taylor & Francis Group. Part I. Social Influence: Fundamental Processes and Theories: 3-24.
- Galam, S. (2004). "Contrarian deterministic effects on opinion dynamics: "the hung elections scenario"." *Physica A* 333: 453-460.
- Gargiulo, F. and S. Huet (2010). "Opinion dynamics in a group-based society." *Europhys. Lett.* 91(5).
- Gargiulo, F. and S. Huet (2012). "New discussions challenge the organization of societies." *Advances in Complex Systems* 0(0): 1250033.
- Gargiulo, F., M. Lenormand, et al. (2011). "Commuting network model: going to the bulk." *JASSS*: 15.
- Gargiulo, F., S. Ternès, et al. (2010). "An Iterative Approach for Generating Statistically Realistic Populations of Households." *PLoS One* 5(1): 9.
- Gilbert, N. and K. G. Troitzsch (2005). *Simulation for the social scientist, Maidenhead [etc.]* : Open University Press.
- Givord, P. (2003). "Une nouvelle enquête Emploi." *Economie et Statistiques* 362: 59-66.
- Gobillon, L. (2001). "Emploi, logement et mobilité résidentielle." *Economie et Statistique* 349-350(9/10): 77-98.
- Gotts, N., J. G. Polhill, et al. (2003). "Aspiration levels in a land use simulation." *Cybernetics and Systems* 34: 663-683.
- Goux, D. (2003). "Une histoire de l'enquête Emploi." *Economie et Statistiques* 362: 41-57.
- Grimm, V. (1999). "Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future?" *Ecological Modelling* 115: 129-148.
- Grimm, V., U. Berger, et al. (2006). "A standard protocol for describing individual-based and agent-based models." *Ecological Modelling*(198): 115-126.

- Grimm, V., U. Berger, et al. (2010). "The ODD protocol: A review and first update." *Ecological Modelling* 221: 2760-2768.
- Haugtvedt, C. P. and D. T. Wegener (1994). "Message Order Effects in Persuasion: An Attitude Strength Perspective." *Journal of Consumer Research* 21: 205-218.
- Hegselmann, R. and U. Krause (2002). "Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation." *Journal of Artificial Societies and Social Simulation* 5(3).
- Heider, F. (1946). "Attitudes and Cognitive Organization." *Journal of Psychology*(21): 107-112.
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hill CA, Sage Publication.
- Hogg, M. A., J. C. Turner, et al. (1990). "Polarized Norms and Social Frames of Reference: A Test of the Self-Categorization Theory of Group Polarization." *Basic and Applied Social Psychology* 11(1): 77-100.
- Holme, E., K. Holme, et al. (2004). *The SVERIGE spatial microsimulation model: content, validation and example applications*. , Spatial Modelling Centre, Sweden. Technical report: 55 p.
- Hoorens, V. (1993). "Self-enhancement and superiority biases in social comparison." *European Review of Social Psychology* 4: 113-139.
- Hoorens, V. (2011). *The Social Consequences of Self-Enhancement and Self-Protection*. Handbook of Self-Enhancement and Self-Protection. M. D. Alicke and C. Sedikides. New York, London, The Guilford Press: 235-257.
- Huet, S. and G. Deffuant (2001). *Calcul de l'impact économique de la conversion à l'agriculture biologique : projet IMAGES, modèle Allier Conversion à l'agriculture biologique, Cemagref. Rapport du projet Européen IMAGES: 77*.
- Huet, S. and G. Deffuant (2008 a). "Differential equation models derived from an individual-based model can help to understand emergent effects." *Journal of Artificial Societies and Social Simulation* 11(2): 21.
- Huet, S. and G. Deffuant (2008 b). *Bounded Confidence with Rejection: The Infinite Population Limit With Perfect Uniform Initial Density*. MASHS, Modèles et Apprentissages en Sciences Humaines et Sociales. Créteil, France: 9.
- Huet, S. and G. Deffuant (2008 c). *Bounded Confidence with Rejection: Clusters or Scattered Opinions?* ESSA, European Social Simulation Association. Brescia, Italy: 12.
- Huet, S. and G. Deffuant (2010). *Common Framework for Micro-simulation Model in PRIMA project, Cemagref - PRIMA working paper: 23 p*.
- Huet, S. and G. Deffuant (2010). "Openness leads to opinion stability and narrowness to volatility." *Advances in Complex Systems* 13(3): 405-423.
- Huet, S., G. Deffuant, et al. (2007). *Rejection Mechanism in 2D Bounded Confidence Provides More Conformity*. ECCS, European Conference on Complex Systems, Dresden, October 1-5.
- Huet, S., G. Deffuant, et al. (2008 d). "Rejection Mechanism in 2D Bounded Confidence Provides More Conformity." *Advances in Complex Systems* 11(4): 1-21.
- Huet, S., N. Dumoulin, et al. (2012). *Micro-simulation model of municipality network in the Auvergne case study*. P. W. paper. Aubière, France, IRSTEA: 63.
- Huet, S., M. Edwards, et al. (2007). "Taking into account the variations of social network in the mean-field approximation of the threshold behaviour diffusion model." *Journal of Artificial Societies and Social Simulation* 10(1).
- Huet, S., M. Lenormand, et al. (2012). *Parameterisation of individual working dynamics*. Empirical Agent-Based Modeling: Parameterization Techniques in Social Simulations. A. Smajgl and O. Barreteau, Springer: 22.
- INSEE (1990). "French population census of 1990." Data distributed by the Centre Maurice Halbwachs, 48 bd Jourdan, Paris.
- INSEE (1990-2002 ; 2003-2007). "Enquête Emploi: questionnaire individuel."
- INSEE (1999). "Enquête sur l'histoire familiale." Data distributed by the Centre Maurice Halbwachs, 48 bd Jourdan, Paris.

- INSEE, d. R. e. P. S. (1999). "Le modèle de simulation dynamique DESTINIE." Série des documents de travail de la Direction des Etudes et Synthèses Economiques: 124.
- Isenberg, D. (1986). "Group Polarization: A critical Review and Meta-Analysis." *Journal of Personality and Social Psychology* 50(6): 1141-1151.
- Jager, W. and F. Amblard (2004). "Uniformity, Bipolarization and Pluriformity Captured as Generic Stylized Behavior with an Agent-Based Simulation Model of Attitude Change." *Computational & Mathematical Organization Theory*(10): 295-303.
- Jager, W. and F. Amblard (2006). Guess you're right on this too: Central and Peripheral processing in attitude changes in large populations. . 1st World Conference of Social Simulation. Kyoto: 12.
- Johansson, M. and D. Rauhut (2007). The Spatial Effects of Demographic Trends and Migration. E. P. 1.1.4., European Observation Network for Territorial Development and Cohesion (ESPON): 65.
- King, A., H. Baekgaard, et al., Eds. (1999). DYNAMOD2: an overview. University of Canberra, NATSEM.
- Krizan, Z. and R. S. Baron (2007). "Group polarization and choice-dilemmas: how important is self-categorization?" *European Journal of Social Psychology* 37(1): 191-201.
- Krosnick, J. A. (1988). "Attitude Importance and Attitude Change." *Journal of Experimental Social Psychology* 24(3): 240-255.
- Krosnick, J. A., L. Holbrook, et al. (2000). "The impact of the fall 1997 debate about global warming on American public opinion." *Public Understand. Sci.*(9): 239-260.
- Lenormand, M. and G. Deffuant (2012). "Generating a synthetic population of individuals in households: Sample-free versus sample-based methods." *Journal of Artificial Societies and Social Simulation* submitted: 15.
- Lenormand, M., S. Huet, et al. (2012b). "Generating French Virtual Commuting Network at Municipality Level." *Journal of Transport and Land Use* submitted.
- Lenormand, M., S. Huet, et al. (2012a). "A Universal Model of Commuting Networks." *PLoS One* 7(10).
- Lenormand, M., S. Huet, et al. (2011). A commuting generation model requiring only aggregated data. ESSA 2011. E. S. S. Association. Montpellier: 16.
- Li, J. and C. O'Donoghue (2012). A methodological survey of dynamic microsimulation models. U.-M. W. P. Series, Maastricht University, Maastricht Graduate School of Governance, The Netherlands. 2012: 34.
- Lorenz, J. (2007). "Continuous Opinion Dynamics Under Bounded Confidence: A Survey." *International Journal of Modern Physics C* 18(12): 1819-1838.
- Mabrouk, N. (2010). Analyzing individual-based models of microbial systems. Doctorate, Blaise Pascal University.
- Mahdavi, B., D. O'Sullivan, et al. (2007). An agent-based microsimulation framework for investigating residential segregation using census data. MODSIM 2007 International Congress on Modelling and Simulation. L. Oxley and D. Kulasiri, Modelling and Simulation Society of Australia and New Zealand. December, 2007: 365-371.
- Mark, N. P. (2003). "Culture and Competition: Homophily and Distancing Explanations for Cultural Niches." *American Sociological Review* 68(3): 319-345.
- Martin, S., G. Deffuant, et al. (2004). Comparing the Binary Opinion Vector Version of the BC Model with Different Approximating Aggregated Dynamics. ESSA 2004. Valladolid, Spain.
- Mäs, M. (2010). The diversity puzzle - Explaining clustering and polarization of opinions. Doctorate, Groningen University.
- Mason, W. A., F. R. Conrey, et al. (2007). "Situating Social Influence Processes: Dynamic, Multidimensional Flows of Influence Within Social Networks." *Personality and Social Psychology Review*(11): 279-300.
- Mathias, J. D. (2011). De la double modélisation à la validation expérimentale par des méthodes de mesure de champs sans contact du comportement des matériaux. Habilitation à Diriger des Recherches, Blaise Pascal.

- Matthews, R., N. Gilbert, et al. (2007). "Agent-based land-use models: a review of applications." *Landscape Ecology* 22(10): 1447-1459.
- Matz, D. and W. Wood (2005). "Cognitive dissonance in groups: The consequences of disagreement." *Journal of Personality and Social Psychology* 88(1): 22-37.
- Miller, N. and D. T. Campbell (1959). "Recency and primacy in persuasion as a function of the timing of speeches and measurements." *Journal of Abnormal and Social Psychology*(59): 1-9.
- Minodier, C. (2006). "Changer de logement dans le même environnement." *Données sociales. La Société Française*: 515-523.
- Moeckel, R., K. Spiekermann, et al. (2003). "Microsimulation of land use." *International Journal of Urban Sciences* 71(1): 14-31.
- Monteil, J. M. (1993). *Soi & le contexte*, Armand Colin.
- Morand, E., L. Toulemon, et al. (2010). *Demographic modelling: the state of the art*. Paris, INED. FP7-244557 *Projet SustainCity*: 39.
- Moscovici, S. (1979). *Psychologie des minorités actives*. France, PUF.
- Moscovici, S. (2000). *Psychologie sociale des relations à autrui*. France, Nathan Université.
- Moscovici, S. and M. Zavalloni (1969). "The Group as a Polarizer Of Attitudes." *Journal of Personality and Social Psychology* 12(2): 125-135.
- Müller, K. and A. K.W. (2011). *Population synthesis for microsimulation: State of the art*. 90th Annual Meeting of the Transportation Research Board, Washington D.C.
- Nagel, K., R. L. Beckman, et al. (1999). *TRANSIMS for transportation planning*. 6th Int. Conf. on Computers in Urban Planning and Urban Management. Venice, Italy.
- Norman, P. (1999). *Putting iterative Proportional Fitting on the researchers' desk*. S. o. Geography, University of Leeds, United Kingdom. Working paper 99/03: 36.
- O'Donoghue, C. (2001). "Dynamic microsimulation: a methodological survey." *Brazilian Electronic Journal of Economics* 4(2): 77 p.
- Parker, D. C., S. M. Manson, et al. (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.
- Perrier-Cornet, P. (2001). "La dynamique des espaces ruraux dans la société française : un cadre d'analyse." *Territoires 2020 3*(Etudes et prospectives - Data): 61-74.
- Petty, R. E., Z. L. Tormala, et al. (2001). "Motivation to Think and Order Effects in Persuasion: The Moderating Role of Chunking." *Personality and Social Psychology Bulletin* 27(3): 332-344.
- Picard, N. and A. Franc (2001). "Aggregation of an individual-based space-dependent model of forest dynamics into distribution-based and space-independent models." *Ecological Modelling* 145(1): 69-84.
- Polhill, J. G., D. C. Parker, et al. (2008). "Using the ODD Protocol For Describing Three Agent-Based Social Simulation Models of Land-Use Change." *Journal of Artificial Societies and Social Simulation* 11(2 3): 30.
- Poubelle, V. (2006). "Prisme, le modèle de la Cnav." *Retraite et Société* 48: 202-215.
- Radillo-Díaz, A., L. A. Pérez, et al. (2009). "Axelrod models of social influence with cultural repulsion." *Physical Review E* 80(6): 066107.
- Rindfuss, R., S. Walsh, et al. (2004). "Developing a science of land change: Challenges and methodological issues." *PNAS* 101(39): 13976-13981.
- Robert-Bobée, I. (2001). *Modelling demographic behaviours in the French microsimulation model Destinie: An analysis of future change in completed fertility*. D. d. E. e. S. E. Institut National de la Statistique et des Etudes Economiques. Malakoff, France, INSEE: 36.
- Rogers, R. W. and S. Prentice-Dunn (1997). *Protection motivation theory*. Handbook of health behavior research. D. Gochman, New York:Plenum. 1: 113-132.
- Rokeach, M. (1968). *Beliefs, attitudes and values*. San Francisco - London, Jossey-Bass Publishers.
- Salzarulo, L. (2006). "A continuous Opinion Dynamics Model Based on the Principle of Meta-Contrast." *Journal of Artificial Societies and Social Simulation* 9(1).



- Sherman, D. K. and G. L. Cohen (2006). *The Psychology of Self-defense: Self-Affirmation Theory*. *Advances in Experimental Social Psychology*. P. Z. Mark, Academic Press. Volume 38: 183-242.
- Schultz, P. W. and L. Zelezny (2003). "Reframing Environmental Messages to be Congruent with American Values." *Human Ecology Review* 10(2): 126-136.
- Schwartz, S. H. (1999). "A Theory of Cultural Values and Some Implications for Work." *Applied Psychology: An International Review* 48: 23-47.
- Sherif, M. and C. I. Hovland (1961). *Social Judgement*, New Haven, CT: Yale University Press.
- Smith, D. M., G. P. Clarke, et al. (2009). "Improving the synthetic data generation process in spatial microsimulation models." *Environment and Planning A* 41: 1251-1268.
- Sobkowicz, P. (2009). "Modelling Opinion Formation with Physics Tools: Call for Closer Link with Reality." *Journal of Artificial Societies and Social Simulation* 12(1 11): 14.
- Soumagne, J. (2003). "Les services en milieu rural, enjeu d'aménagement territorial." *Revista da Faculdade de Letras - Geografia I série XIX*.
- Tajfel, H. (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations*, London: Academic Press.
- Trompenaars, F. and C. Hampden-Turner (1997). *Riding The Waves Of Culture*. London, Nicholas Brealey Publishing.
- Turci, L., A. Bringé, et al. (2010). *Provisional demographic outline*. Paris, INED. FP7-244557 Projet SustainCity: 24.
- Turner, J. C. (1984). *Social identification and psychological group formation. The social dimension*. H. Tajfel. Cambridge, Cambridge University Press: 518-538.
- Turner, J. C., M. Hogg, et al. (1987). *Rediscovering the social groups: A self-categorization theory*, Cambridge, MA: Basil Blackwell.
- Tversky, A. and D. Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases." *Science* 185(4157): 1124-1131.
- Ullman, E. and M. Dacey (1960). "The minimum requirement approach to the urban economic base." *Papers and Proceedings of the Regional Science Association* 6(192).
- Valente, T. W. (1995). *Networks Models of the Diffusion of Innovations*, Hampton Press, Inc 23 Broadway, Cresskill, NJ 07626.
- Vanderschelden, M. (2006). "Les ruptures d'unions : plus fréquentes mais pas plus précoces." *Insee Première novembre*(1107): 4.
- Vaz Martins, T., M. Pineda, et al. (2010). "Mass media and repulsive interactions in continuous-opinion dynamics." *EPL (Europhysics Letters)* 91(4): 48003.
- Verburg, P. H., P. P. Schot, et al. (2004). "Land Use Modelling: Current Practice and Research Priorities." *GeoJournal* 61: 309-324.
- Verburg, P. H., C. J. E. Schulp, et al. (2006). "Downscaling of Land Use Change Scenarios to Assess the Dynamics of European Landscapes." *Agriculture, Ecosystems and Environment* 114: 39-56.
- Verburg, P. H., W. Soepboer, et al. (2002). "Modelling the Spatial Dynamics of Regional Land Use: The CLUE-S Model." *Environmental Management* 30(3): 391-405.
- Vinokur, A. and E. Burnstein (1978 a). "Novel argumentation and attitude change: The case of polarization following group discussion." *European Journal of Social Psychology* 8: 335-348.
- Vinokur, A. and E. Burnstein (1978 b). "Depolarization of attitudes in groups." *Journal of Personality and Social Psychology* 36(8 August): 872-885.
- Visser, P.S., Bizer, G.Y. and Krosnick, J.A. (2006) "Exploring the Latent Structure of Strength-Related Attitude Attributes". *Advances in Experimental Social Psychology*, vol. 38, 1-76.
- Waddell, P., A. Borning, et al. (2003). "Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim." *Networks and Spatial Economics* 3(1): 43-67.
- Weidlich, W. (2002). "Sociodynamics - A systematic Approach to Mathematical Modelling in the Social Sciences." *Nonlinear Phenomena in Complex Systems* 5(4): 479-487.

- Williamson, P., M. Birkin, et al. (1998). "The estimation of population microdata by using data from small area statistics and samples of anonymised records." *Environment and Planning A* 30: 785-816.
- Wood, W., G. J. Pool, et al. (1996). "Self-definition, defensive processing, and influence: the normative impact of majority and minority groups." *Journal of Personality and Social Psychology*(71): 1181-1193.





## Abstract:

This thesis is dedicated to individual-based modeling of social systems. While the first part is very practical, decision-support oriented, presenting a model which studies the evolution of a rural population, the second part is more theoretical, interested in various mechanisms allowing individual to accept or resist to social influence.

Firstly, we propose an individual-based model of the European rural municipalities implemented for the French Cantal *département*. We use a new sample-free algorithm for generating the initial population, while classical methods require an initial sample. We design and parameterize the individual activity dynamics with data from the European Labour Force Survey database. The individual dynamics includes an original heuristic for labour statuses and employments changes, based on individual age, profession and activity sector when she is occupied. The last part of the model deals with dynamics that we have not been able to derive from data, mainly the demographic dynamics. Based on the Occam razor principle, we test very simple dynamics and choose them on their capacity to lead to model results close to reference data. In particular, we propose a simple residential mobility model, partly ruling the emigration, which integrates decision to move and location choice.

Secondly, we study the collective effects of various mechanisms leading individuals to resist or accept social influence. A first mechanism leads individuals to neglect some features of an object if they are not important enough or incongruent. These individuals exhibit the primacy bias because their attitudes are determined by the first accepted feature. We show that this bias increases when individuals directly exchange about features compared to when they only get the features from the media, in a random order. The second mechanism is a rejection reaction that we suppose occurring because of the discomfort taking place when individuals are close on one dimension of attitude and far on another dimension. The main effect of this rejection mechanism is to lead to a lower number of clusters than with the attraction mechanism alone.

## Résumé :

Cette thèse a pour objet la modélisation individu-centrée des systèmes sociaux. Une première partie orientée aide à la décision présente un modèle d'évolution des populations rurales fortement inspiré des données. Une seconde partie, plus théorique, étudie divers mécanismes permettant à un individu d'accepter ou de résister à une influence sociale.

Nous proposons tout d'abord un modèle individu-centré de la dynamique des municipalités rurales européennes, implémenté pour le département du Cantal. Nous proposons un nouvel algorithme de génération des populations initiales ne nécessitant pas d'échantillon de population (approche classique). Nous concevons et paramétrons un modèle de la dynamique de l'individu face au marché du travail basé sur les données de la « European Labour Force Survey ». Il inclut des heuristiques originales de transition d'états tel qu'actifs, inactifs, chômeurs, ... prenant en compte l'âge, la profession et le secteur d'activité de l'individu. Nous déterminons les dynamiques non fondés sur des données individuelles en testant la capacité de dynamiques simples à produire des résultats proches des données agrégées disponibles. Est ainsi conçu un modèle de mobilité résidentielle, expliquant partiellement la migration et intégrant décision de déménager et choix d'une nouvelle résidence.

La seconde partie de la thèse étudie les effets collectifs de différents mécanismes permettant aux individus de résister à ou d'accepter une influence sociale. Un premier mécanisme étudié est un filtre cognitif impliquant qu'un individu ne reçoit pas une information incongruente ou peu importante. Les individus « filtreurs » exhibent le biais de primauté car leur attitude n'est déterminée que par les premiers éléments reçus et se montrent négatifs alors que le message diffusé par un media est neutre. Le taux d'individus négatifs dans la population est accru ou diminuer par l'échange direct d'information entre les individus. Un second mécanisme est un rejet de la tentative d'influence qui mène l'individu à différencier davantage son attitude de celle de son interlocuteur. Il intervient lorsque l'individu éprouve un inconfort lié au fait qu'il est à la fois en accord et en désaccord avec son interlocuteur. Le couplage de ce rejet à un mécanisme d'attraction entre individus en accord entraîne un nombre moindre de groupes d'opinion différentes à l'échelle de la population (ie par rapport au nombre de groupes obtenus avec le seul mécanisme d'attraction).

**Keywords:** modèle dynamique individu-centré; dynamique d'opinion; dynamique de population