GENSTAR

Generation of spatialized and socially structured synthetic populations for social simulation

GENERATE SYNTHETIC POPULATIONS FOR SOCIAL SIMULATION IN THE MANAGEMENT OF SOCIETAL AND SOCIO-ENVIRONMENTAL ISSUES

Institut de Recherche pour le Développement FRANCE

IDEES

AIRBUS DEFENCE & SPACE

IRIT Institut de Recherche en Informatique de Toulouse — CNRS INPT UPS UT1

UMMISCO Unité Mixte Internationale

Project leader: Alexis Drogoul, UMMISCO, IRD, 32 av. Henri Varagnat, 93140 Bondy, France; alexis.drogoul@ird.fr
Project duration: 42 months
Starting date : September 2013
Total budget : 561 K€
4 partners : IRD/ UMMISCO, Bondy, France; University of Toulouse 1 Capitole / IRIT, France; University of Rouen / IDEES, France; Cassidian, France.

CONTEXT

Agent-based social simulation is gaining ground as a candidate of choice for building decision-support tools in the management of complex socio-environmental systems, and the resulting models are therefore being driven to produce more realistic outcomes. Hence, integrating large data corpuses (demographical, environmental, geographical...) as an input to these models becomes more critical.

In such models, the evolution of simulations is partly being driven by social agents, which may represent individuals, households or institutions. The behaviors of these agents are determined by their attributes, their connections with other agents, but also their location in the artificial worlds they populate. As large volumes of relevant data become available, generating synthetic populations of agents that conform to the data available on real populations becomes a necessity and a concern for most social modelers, and although several approaches have been tested in recent works, it constitutes a significant scientific and methodological challenge, which is addressed in this project.

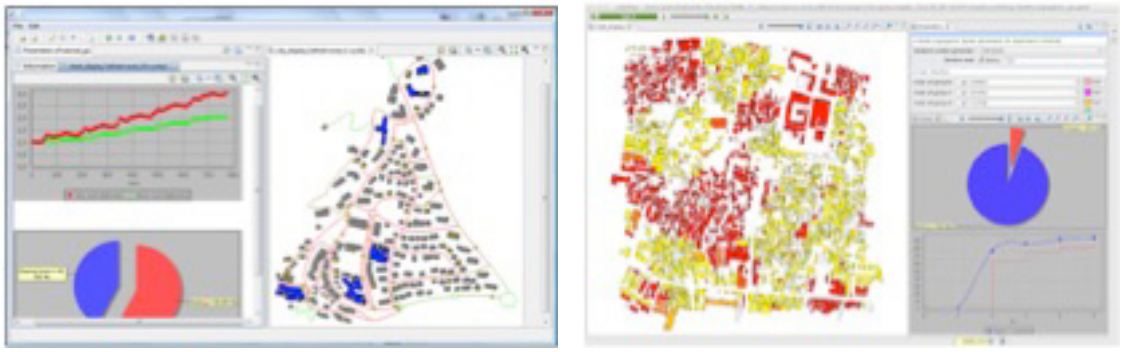


Figure 1 : examples of spatialized simulations using synthetic populations

OBJECTIVES

The aim of the project is to build a set of methods and open source software libraries, for generating synthetic populations. These synthetic populations are sets of individuals, described by variables (relevant for the use case), localized (at a scale depending on the use case) and socially structured as social networks. It is important to note, however, that, although we target agent-based models, we are not aiming at producing realistic behavioral models of agents, but statistically realistic datasets in order to generate their individual attributes, and these synthetic populations are designed to feed agent-based social simulations following a data-driven approach. The case studies that will be selected to validate our approach in the course of the project will cover both civil (epidemiology, natural risks, management of renewable resources) and military issues (decision-support, planning in a non-kinetic intervention framework).

The software produced within the project is tested on various popular agent-based platforms (NetLogo, RePast), and is integrated into GAMA, result of two decades of joint development effort by several teams around IRD, enriching on the long term GAMA functionalities.

At a higher level, the project aims at popularizing agent-based social simulation as a concrete and efficient way to build decision-support systems in the management of societal and socio-environmental issues.

FOCUS ON GAMA

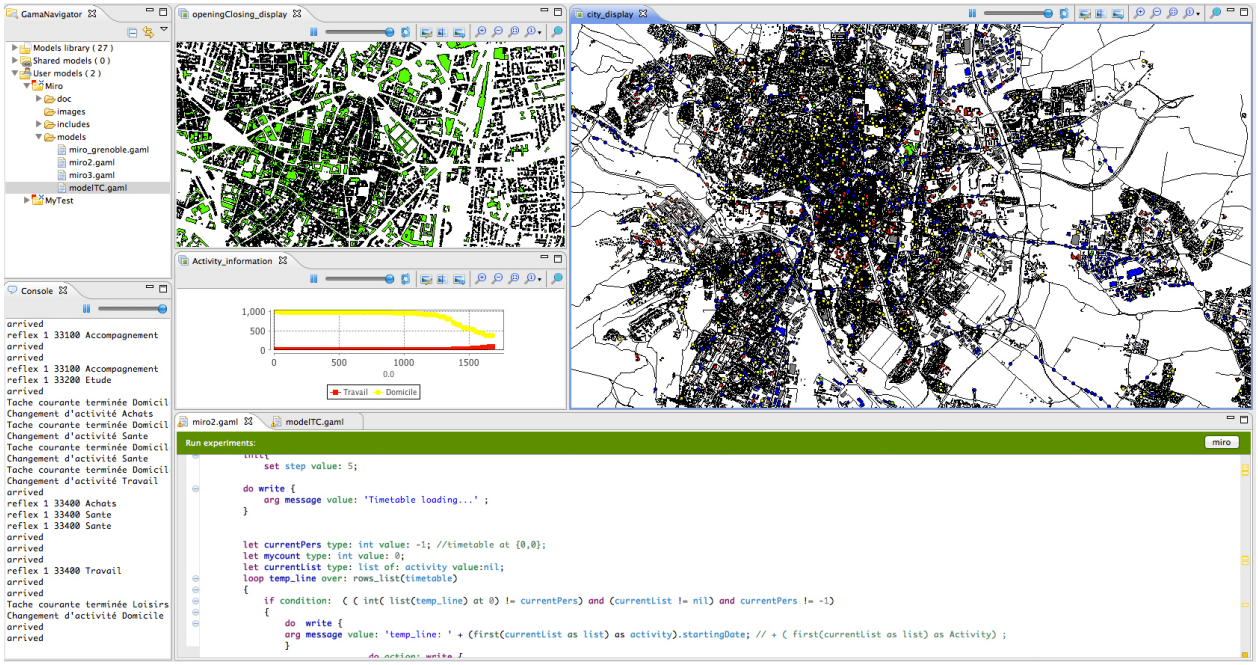
GAMA is a modeling and simulation development environment for building spatially explicit agent-based simulations.

• Multiple application domains: use GAMA for whatever application domain you want.

• High-level and Intuitive Agent-based language: write your models easily using GAML, a high-level and intuitive agent-based language.

• GIS and Data-Driven models: instantiate agents from any dataset, including GIS data, and execute large-scale simulations (up to millions of agents).

• Declarative user interface: declare interfaces supporting deep inspections on agents, user-controlled action panels, multi-layer 2D/3D displays & agent aspects.



SPECIFIC CHALLENGE.

The main bottlenecks identified are:

• Couple existing algorithms and approaches, which are nowadays scattered among several disciplines, application domains and dedicated tools, for generating agents populations in various generation scenarios;

• Develop generators of social networks that enable to create realistic networks between artificial agents;

• Establish a methodological correspondence between the problems raised by the case studies and the solutions of population generation proposed by applied mathematics and computer science;

• Identify a meta-model of population data satisfying the needs of modelers and that could conform to the standards of the Open Data Foundation;

• Formalize the population generation tasks as a (re)usable workflow that could be ported to different case studies.

RESULTS

The main results obtained are as follows:

• Methods for the co-generation of realistic spatialized populations and social networks, coupling applied mathematics (statistics, data analysis) and evolutionary approaches (genetic algorithms, Bayesian networks, classifier systems, multi-agent systems);

• Documented libraries in R and Java implementing these methods, tested on several concrete case studies. The integration tests have been achieved on different simulation platforms (firstly GAMA, as two of the partners are developing it, and then the most popular ones like NetLogo and RePast);

• Software components enabling the integration with open-data resources and the most commonly used data formats used in such context (e.g. GIS);

• Methodological materials (tutorials and methodological guidelines) for supporting end-users in understanding general population generation issues and for training them to the methods proposed in the Genstar libraries;

• A web-interfaced standalone platform supporting modelers in building workflows representing the coupling between data sources, data converters, generation and spatialization methods, social networks generation; offering a way to execute these workflows (possibly on remote High Performance Computing environments); and saving both their outputs and their structures for later use in an Open Data compatible format;

• Also, a community (mixing both academics and industrials, of developers, modelers and users) built around the issues of synthetic population generation and the software libraries developed in the project.