

Chapter 2

The SimpopLocal Model

Abstract The model is described in detail in order to explain how to implement the stylized facts that are theoretically essential for representing the emergence of cities into a multi-agent system. The model activates rather simple feedback loops between settlement size, innovation and resources enabling settlement growth from local innovation creation and inter-settlement diffusion. The necessary attributes, processes and parameters of the model, that were identified according to a rule of parsimony, are described and their estimated values after simulations are presented.

2.1 Introduction

From the stylized facts summarizing the historical process of the emergence of cities which we have briefly recalled in Chap. 1, we create SimpopLocal to simulate the growth dynamics of agrarian settlements and their possible evolution towards urban settlements under strong environmental constraints that are progressively overcome by successive innovations. In order to ensure the replicability of the model, the source code of SimpopLocal is filed in a public repository (<http://iscpif.github.io/simpoplocal-epb/>).

2.2 Purpose of SimpopLocal

This exploratory model aims at reproducing a remarkable aspect of the spatial structure of settlements systems, defined in the literature as a major stylized fact: in any system already studied, in any places and in any period of history or prehistory times, the distribution of size (population or spatial extent) is strongly differentiated, including many very small settlements and only few very large settlements according to a rather regular distribution of the Zipf or log-normal type (Fletcher 1986; Liu 1996).

This hierarchical pattern is a structural property (order in the size of entities) at macroscopic level that is particularly resilient over time, whatever the local fluctuations which take place at entity level. The SimpopLocal model is designed for testing

the hypothesis enunciated in the evolutionary theory of urban systems (Pumain 1997) which explains this structural feature from the urban growth process sustained by all kinds of technological and societal innovations and their spatial diffusion among connected settlements.

This model adds to the usual stochastic model of urban growth that is simply proportional to city size and leading to a log-normal distribution (Gibrat 1931) the effect of the spatial interaction which amplifies the growing hierarchical differentiation among settlement sizes that is observed over time in geographical urban systems (Favaro and Pumain 2011).

SimpopLocal is part of the Simpop agent based model family. In comparison of models already developed, SimpopLocal adopts some new original paths. First, it simplifies the way that was used until now to qualitatively discriminate the successive innovation waves that were represented by various urban functions, it captures all of them in a single abstract innovation object.

Second, SimpopLocal makes the process of innovation creation endogenous by linking it with the size of settlement. This more parsimonious version of model building enables the development of better and more systematic exploration and evaluation of ABM. SimpopLocal was initially developed using Netlogo language, and later redeveloped using Scala programming language.

We describe the model following the ODD standard principles (Grimm et al. 2010), in a slightly different order, and without describing the ‘design concepts’, whose categories are not relevant here.

2.3 Entities, State Variables and Scales

The model represents the evolution of settlement units that are dispersed in an area large enough for sustaining a few thousands population but limited enough in surface for ensuring the possible connection between settlements according to the transportation means that are available at the time. Typically, it could be a region as antique Mesopotamia or Meso America. The landscape of the simulation space is composed of hundreds of settlements. Each settlement is considered as a fixed agent and is described by three attributes: the location of its permanent habitat (x, y) , the size of its population P , and the available resources in its local environment.

The amount of available resources R is quantified in units of inhabitants and can be understood as the carrying capacity of the local environment for sustaining a population which depends on the resource exploitation skills that the local population has acquired from inventing or acquiring innovation. This resource exploitation is done locally and sharing or trade is not represented explicitly in the model. Each new innovation created or acquired by a settlement develops its exploitation skills. Contrary to previous more ‘realistic’ models of the Simpop family, we do not want to consider the nature of innovation by identifying each significant innovation wave as a new urban function. We simplify the model by retaining only the processes of emergence of innovation and their effect on urban growth. The innovation entity is understood

here as a large and abstract invention socially accepted which could represent a technical invention, a discovery, a social organization, some new habits or practices ... Each acquisition of innovation by a settlement brings there the possibility to surpass its capacity threshold, and by consequence authorizes a demographic growth. The state variables defined at macro-level are the size distribution of settlements and the slope of the rank size distribution.

2.4 Processes Overview and Scheduling

The scheduling of a simulation of the model is presented on Fig. 2.1 and will be further detailed for each part of the process linking the evolution of innovation, resource and population growth in the settlements.

After the initialization of the settlements, the interaction network is created. Then, at each simulation step, the mechanisms of population growth (*grow population*) and innovation diffusion (*diffuse innovation*) are applied. According to the number of innovation, the impact of these innovations is applied on the settlement's resource extraction efficiency (*apply innovations*). Then, the innovation creation mechanism (*create innovation*) is applied, with the same effect on resource extraction efficiency. This loop is iterated until the stopping criterion is reached: in this case after 4000 steps or if the maximum number of innovation has been reached. We now present each of these mechanisms in detail. Regarding the ODD protocol, these mechanisms would be labelled as the submodels of SimpopLocal.

2.4.1 Population Growth Mechanism

The growth dynamics of a settlement are modelled according to the assumption that its size is dependent on the amount of available resources in the local environment and is inspired by the Verhulst model (Verhulst 1845) or logistic growth.

For this experiment, we assume a continuous general growth trend for population—this may be different in another application of the model. The growth factor r is expressed on an annual basis; thus, one iteration or step of the model simulates one year of demographic growth. The limiting factor of growth R_M^i is the amount of available resource that depends on the number M of innovations the settlement i has acquired by the end of the simulation step t .

P_t^i is the population of the settlement i at the time t :

$$P_{t+1}^i = P_t^i \left[1 + r \left(1 - \frac{P_t^i}{R_m^i} \right) \right] \quad (2.1)$$

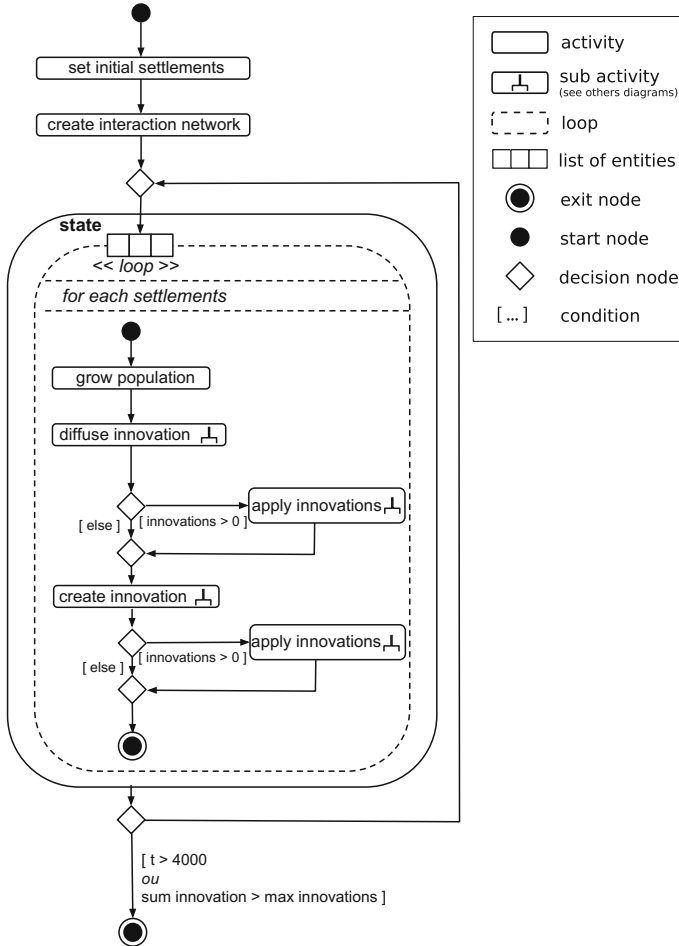


Fig. 2.1 SimpopLocal activity diagramm

2.4.2 Apply Innovation Mechanism

The acquisition of a new innovation by a settlement allows it to overtake its previous growth limitation by enabling a more efficient extraction of resources and thus a gain in population-size sustainability. With the acquisition of innovations, the amount of available resources tends to the maximal carrying capacity R_{max} of the simulation environment:

$$R_M^i \xrightarrow{\text{innovations acquisition}} R_{max} \quad (2.2)$$

The mechanism of this impact follows the Ricardo model of diminishing returns (which also is a logistic model). The *InnovationImpact* represents the impact of

the acquisition of an innovation and has a decreasing effect on the amount of available resources R_{M+1}^i with the acquisition of innovations:

$$R_{M+1}^i = R_M^i \left[1 + InnovationImpact \left(1 - \frac{R_M^i}{R_{max}} \right) \right] \quad (2.3)$$

2.4.3 Create and Diffuse Innovation Mechanisms

Acquisition of innovations can occur in two ways, either by the emergence of innovation within a settlement or by its diffusion through the settlement system. In both cases, interaction between people inside a settlement or between settlements is the driving force of the dynamics of the settlement system. It is a probabilistic mechanism, depending on the size of the settlement. Indeed, innovation scales superlinearly: the larger the number of innovations acquired the larger the settlement and the higher the probability of innovation. To model the superlinearity of the emergence of innovation within a settlement, we model its probability to be created by a binomial law.

If $P_{creation}$ is the probability that the interaction between two individuals of the same settlement is fruitful, that is, leads to the creation of an innovation, and N the number of possible interactions, then, by the binomial law, the probability of the emergence of at least one innovation $P(m_{creation} > 0)$ can be calculated and then used in a random drawing:

$$\begin{aligned} P(m_{creation} > 0) &= 1 - P(m_{creation}=0), \\ &= 1 - \left[\frac{N!}{0!(N-0)!} * P_{creation}^0 * (1 - P_{creation})^{N-0} \right], \quad (2.4) \\ &= 1 - (1 - P_{creation})^N \end{aligned}$$

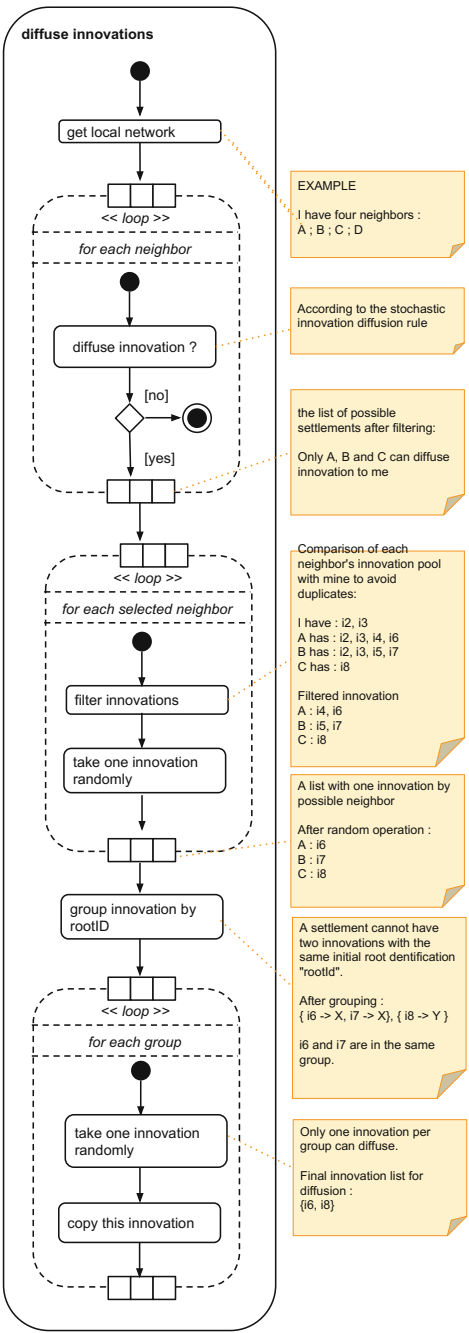
If the size of the settlement is P_t^i then the number N of possible interactions between individuals of that settlement is:

$$N = \frac{P_t^i (P_t^i - 1)}{2} \quad (2.5)$$

The diffusion of an innovation between two settlements depends on both the size of populations and the distance between them.

If $P_{diffusion}$ is the probability that the interaction of two individuals of two different settlements is fruitful—that is, leads to the transmission of the innovation—and K is the number of possible interactions, then, by the binomial law, the probability of diffusion of at least one innovation $P(m_{diffusion} > 0)$ can be calculated and used in a random drawing:

Fig. 2.2 Diffuse innovation mechanism activity diagram



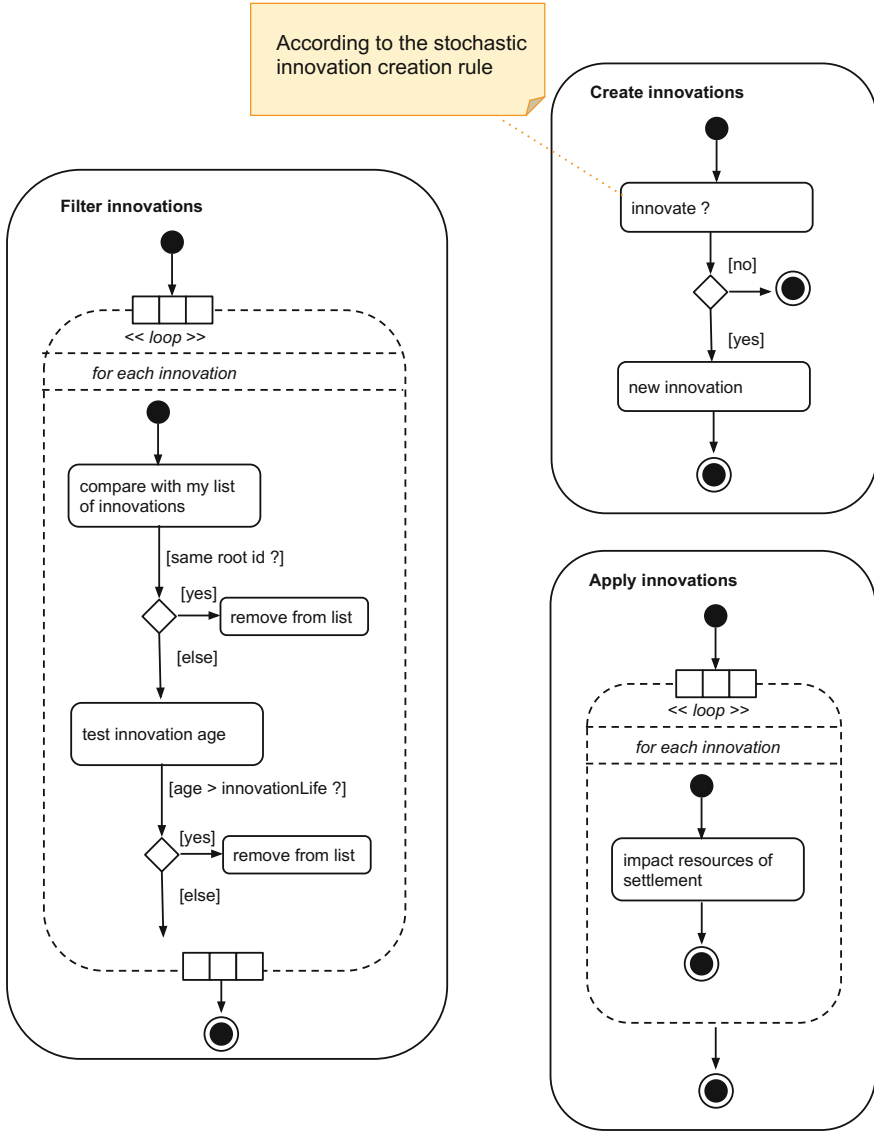


Fig. 2.3 Creation, filter and apply innovation mechanisms activity diagrams

$$P(m_{diffusion} > 0) = 1 - (1 - P_{diffusion})^K \quad (2.6)$$

But in this case, the size K of the total population interacting is a fraction of the population of the two settlements i and j which is decreasing by a factor

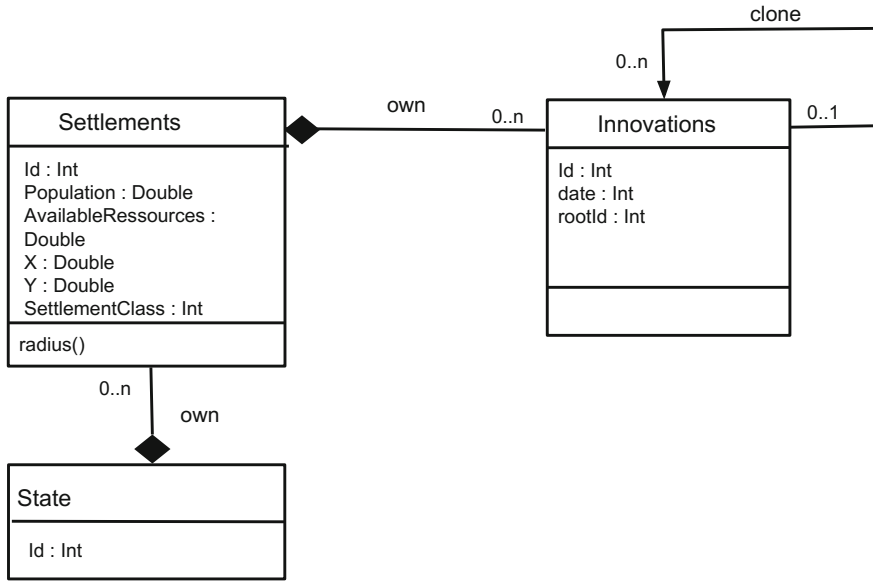


Fig. 2.4 UML class diagram of SimpopLocal

DistanceDecay with the distance D_{ij} between the settlements, as in the gravity model:

$$K = \frac{P_t^i P_t^j}{2D_{ij}^{DistanceDecay}} \quad (2.7)$$

The process of population growth and the process of innovation creation and diffusion are reiterated throughout the simulation (Figs. 2.2 and 2.3). Because of the two positive feedbacks that operate on resource and population growth through the creation of innovation, the model is able to generate a very rapid expansion of settlements: that is, an escalation of settlement growth. The simplest way to avoid situations where too many innovations are created, which would lead to huge time-consuming simulations, is to decide to stop the simulation when it reaches an arbitrary number of, say, 10,000 innovations. Finally, on Fig. 2.4 the UML class diagram of the model is illustrated.

2.5 Initial Conditions

The initial configuration we have chosen to keep in every experiment has therefore been defined the following rules to ensure a good representation of common structures of settlement systems: The size of settlements follows a log-normal distribution.

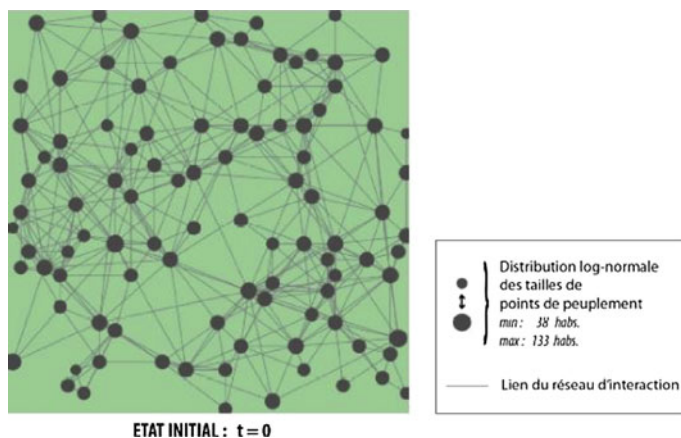


Fig. 2.5 Agents (settlements) and attributes at initial configuration in SimpopLocal

100 settlements size are initialized with a random demographic size vary from 38 to 133 inhabitants. The spatial repartition of the settlements assumes a Christallerian pattern (according to Christaller,¹ a theoretical pattern of central places has regular spacing distances between settlement nodes of the same size category, the largest settlement nodes having larger spacing distances than the smaller) (Fig. 2.5).

The network interlinking the settlements enables spatial interactions according to the same Christallerian logic: small settlement nodes establish less remote connections than the largest. Furthermore, at initial state, population of settlements is considered at equilibrium regarding the available resources: the initial amount of resource of each settlement is considered equal to its initial population.

2.6 Input

Although relatively parsimonious as a multi-agent system, SimpopLocal has a dozen of parameters that have to be estimated for calibrating the model. Some can be empirically evaluated with the help of historical data and knowledge, while it is very difficult to give values to others (Table 2.1). Those regarding the initial spatial distribution and organization of settlements in the landscape can be approximated. The log-normal distribution of the settlement sizes and the central place theory of Christaller for the geographical distribution of locations are models that are widely used by

¹Christaller's central place theory (1933) considers cities as centres serving services of different levels of rarity to a regional resident population according to a hierarchy of range and size of the centres. As residents minimize the cost of access to services centres would exhibit regular patterns on an homogenous plain. The simplest pattern is made of nodes located at the summits of hexagons that are embedded in hexagons of larger areas designed for larger centres of the next upper level.

Table 2.1 Parameters of the SimpopLocal model (Source: Schmitt (2014), p. 167)

| Parameters | Description |
|--------------------|---|
| R_{max} | Maximum carrying capacity of a settlement (measured in number of residents) |
| r_{growth} | Mean growth rate as in verhulst model |
| $InnovationImpact$ | Impact of any innovation on available resources |
| $P_{Creation}$ | Probability of creating an innovation in one settlement |
| $P_{Diffusion}$ | Probability of diffusing innovation between settlements |
| $DistanceDecay$ | Deterrent effect of distance on innovation diffusion |
| $InnovationLife$ | Time during which an innovation may diffuse |
| $MaxInnovation$ | Total number of innovation generated before end of simulation |

archaeologists to describe their spatial data (Archaeomedes 1998; Johnson 1977; Sanders 2012) including Neolithic archaeological sites (Liu 1996).

In SimpopLocal, the mean density of that landscape and the average size of each settlement are representative of the usual orders of magnitude presented in these works. A hundred settlements are distributed according to these two theories and each settlement is initially composed of some 80–400 inhabitants. Several scholars agree that an average annual growth of 0.02% is representative of the growth of agrarian settlements in the Neolithic times (Bairoch 1985; Renfrew and Poston 1979). The length of time required for a transition from agrarian to urban settlements is estimated according to (Bairoch 1985; Marcus and Sabloff 2008) to about three thousand years. We choose to operate our simulations on a four thousand years time period for settlements ranging from one hundred inhabitants up to about ten thousand inhabitants.

Because of a lack of empirical data, five parameters cannot be empirically approximated and have to be estimated through simulation:

- $P_{creation}$, the probability that an innovation emerges from the interaction between two individuals of a same settlement.
- $P_{diffusion}$, the probability that an innovation is transmitted between two individuals of different settlements. We consider that the probability of diffusion is greater than the probability of creation, which means that copying is easier than inventing (Pennisi 2010) in the model.
- $InnovationImpact$, the impact of the acquisition of innovation on the growth of settlements.
- $DistanceDecay$, the deterrent effect of distance on diffusion.
- R_{max} , the maximum carrying capacity of the landscape of each settlement (measured in number of inhabitants).

2.7 Running the Model for Parameter Estimates: Calibration

The principle of parsimony that led the development of SimpopLocal was applied as well in designing a way for estimating possible values for each parameter of the model. This original estimation process that leads as well to a huge qualitative improvement in the validation process of the hypothesis of the model will be examined in detail in Chap. 3. We only mention here which general line was followed in order to make understandable the results of simulation that are recalled below. As we lack of observed measurements for determining the possible values of most of parameters of the model, our method of estimation is not exactly a ‘calibration’ exercise. It consists in determining which global subsets of parameters values are leading to an emergence of a system of cities whose characteristics match at best the stylized facts that were identified in Chap. 1. A kind of machine learning method is necessary in order to identify through many possible behaviours of the model, the one which is able to correctly reproduce the expected results of simulation. The values we will get for the parameters are thus not measured in absolute units, they are abstract estimations and their significance relies on these measurements taken as a whole, each parameter remaining associated to the others having to be considered in relative terms. As two parameters involve probability distributions, the model is stochastic, therefore the two techniques (by trial and error and by full plan) usually used to calibrate a model are not suited to calibrate the SimpopLocal model (or any multi-agents model in general). We adopted innovative methods of automatic exploration of the patterns in the behavioural space of parameters which are developed on the simulation platform OpenMole. In Chap. 3 we shall explain in detail how genetic algorithms and grid computing are used to explore in a comprehensive way the parameter space, as it was roughly defined at first by a plausible but large enough variation domain for each parameter (Table 2.2).

We briefly retrace here in a vocabulary that is accessible to non-specialists of computing science how the method is working. An important first step in calibration is to define an objective function which is identified from the stylized facts describing the period. It includes three quantifiable elements that must be obtained at the end of simulation for the simulated system of cities:

- A log-normal distribution of settlement size
- The size of the largest aggregate settlement of about 10,000 inhabitants
- A total of 4,000 simulation time steps (equivalent to some 4,000 years).

The first requirement of the objective function reflects the essential hierarchical property of any system of cities; the second acknowledges that in the political and technological conditions of the time, groups of resident population over 2,000 inhabitants were very rare and a concentration of 10,000 could represent a major political and economic capital of a kingdom or empire; the third condition is constraining the model to be contained in a domain of growth regime for settlements that is plausible in demographic terms for the post-Neolithic period: at that time, rapid urban growth

Table 2.2 Variation domain for parameters after 500 millions simulations with SimpopLocal and their precision with calibration profile algorithm (Source: Schmitt (2014), p. 203)

| Parameters | Initially assumed variation domain | Variation domain inside Pareto Front | Calibration validated domain |
|--------------------|------------------------------------|--------------------------------------|------------------------------|
| R_{max} | [1; 40000] | [9500; 11500] | [10090; 10465] |
| $InnovationImpact$ | [0; 2] | $[6.10.10^{-3}; .10^{-2}]$ | $[7.7.10^{-3}; 8.4.10^{-3}]$ |
| $P_{Creation}$ | [0; 1] | $[4.0.10^{-7}; 2.1.10^{-6}]$ | $[1.1.10^{-6}; 1.3.10^{-6}]$ |
| $P_{Diffusion}$ | [0; 1] | $[3.10^{-7}; 1.8.10^{-6}]$ | $[6.7.10^{-7}; 6.9.10^{-7}]$ |
| $DistanceDecay$ | [0; 4] | [0.2; 1.1] | [0.66; 0.75] |

rates were hampered by insufficient resource, high mortality rates due to bad sanitary conditions and frequent catastrophes due to natural hazards or devastation caused by war. Meeting these objectives entails contradictory dynamic trends. Moreover, the need of (at least) a hundred replications of the simulations using the same set of parameter values to handle the stochasticity and the wide range of variation attributed a priori to five unknown parameters led us to use an evolutionary algorithm to solve this multi-objective optimization problem as well as massively distributed computing for the exploration of the entire parameter space.

2.8 Simulation Results and Return on Observations

In total, 500 million of model runs were conducted to achieve the calibration of parameters presented in Table 2.2. This table shows next to each parameter in a first column the hypothetical variation domain initially assumed, which was designed deliberately very wide, and in a second column the possible interval of values as it was reduced from the simulations to calibrate the model. This result does not lead to a single value but provides a range of possible values for each parameter, because a 'Pareto front' establishes compromise between values that ensures the multi-objective optimization function (that is explained with more details in Chap. 3). A third column shows which precision gain was realized for each parameter by using a more powerful algorithm that calculates the model's sensitivity to variations of a parameter at a time, all things being equal as to changes in the others. In addition, we must remember that the values presented in Table 2.2 are not independent measurements, they are connected and so it is their entire configuration that must be adapted when the model will be applied for a calibration on empirical, historical or archaeological situation (Figs. 2.6 and 2.7).

Another novelty of this experience is that the method for exploring the behaviour of the model is also a validation: we can establish to what extent the assumptions chosen to implement the mechanisms of the model are both necessary and sufficient to achieve the desired result—of course within the framework of the description

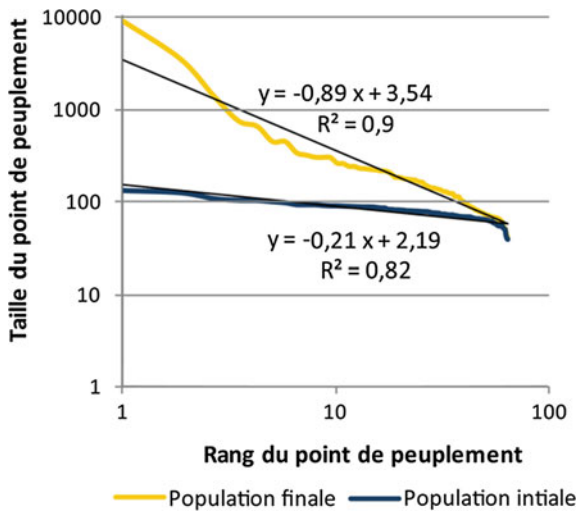


Fig. 2.6 Initial and simulated distribution of city sizes. Source Schmitt 2014

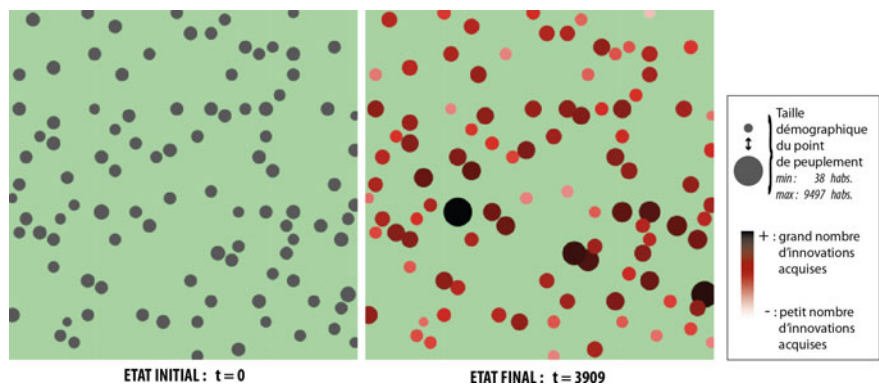


Fig. 2.7 Comparing initial and simulated settlement distributions. Source Schmitt 2014

chosen for the selected model. This second result comes from the development of a method called ‘calibration profiles’ (described in detail in Chap. 3) that calculates for each parameter the effect of its variation on the quality of the model, as fitted by the objective function, ceteris paribus about the changes of other parameters. That particular method led to the rejection of *InnovationLife* parameter (lifetime of innovation) which was not sufficiently constraining the development of the model. The method also contributed to clarify the role of parameters through reducing their domain of variation to a more precise interval that is both necessary and sufficient to achieve the desired change (last column of Table 2.2).

These results provided by the analysis of calibration profiles provide valuable feedback on the modelling assumptions and urban theory that oversees the develop-

ment of the model. The urban evolutionary theory that guided the construction of SimpopLocal model and more generally all models of the Simpop family insists on the concept of system, that is to say, relationship and interaction between the elementary entities (cities, villages or settlements points) that are components of the system. Yet, this is the first time that the exploration of the simulation model demonstrates that the mechanisms of interaction between the entities of the system are essential to the production of an evolution similar to that of real systems, in the context of the Simpop family of modelling (Pumain and Robic 2012).

Without these mechanisms describing spatial interaction, which in SimpopLocal are controlled by *Pdiffusion* and *DistanceDecay*, so without diffusion of innovation between cities, according to a gravity principle, it is not possible to generate urban growth dynamics that are representative of dynamics actually observed in real systems. These results also show the importance of the role of space in structuring and organizing the settlement system: without the effect of *DistanceDecay* parameter, which reduces the frequency of interaction with distance, changes in the simulated system are no longer representative of actual system developments. These first evaluations of mechanisms will also be useful for the next versions and applications of the SimpopLocal model. If this model is too simplified to be fully compliant with current or former real settlement systems in its first abstract and parsimonious version, the concepts generating the simulated processes can be reflected in certain contexts (i.e. the proto-historic cities, for example) or archaeological theories such as ‘peer polity interaction’ (Renfrew 1975).

Christopher Renfrew noticed how frequently the first small states were not born in isolation but in cluster, with strong similarities in terms of size, social structure, material culture, etc. He also observed that political entities comparable in size and organization (as the first forms of state organization) tended to emerge in the same areas and evolve simultaneously. Moreover, archaeological evidence suggests that these changes did not emanate from a single source of innovation, but emerged contemporaneously in several interacting units. According to these remarkable observations, our theory do confirm the central explaining role of the mechanism of exchange between settlement sites and describe the interaction processes as essential in urban development and social change.

The SimpopLocal model, whose dynamic is grounded in social and spatial interaction, could be used as a core model for testing this theory by simulation.

Perspectives of the application of complexity theory and methodological means to construct models (agent-based modeling) underline possible implications for the study of some theoretical issues of scientific research in archeology. The process of model building on the basis of theoretical concepts itself reveals gaps in our data. Within the archaeological record we lack data for some processes which must be supplemented with estimates (Turchin and Gravilets 2009).

References

- Archaeomedes.: Des oppida aux métropoles. Anthropos, Paris (1998)
- Bairoch, P.: De Jéricho à Mexico: villes et économie dans l'histoire. Gallimard, Paris (1985)
- Favaro, J.-M., Pumain, D.: Gibrat revisited: an urban growth model including spatial interaction and innovation cycles. *Geogr. Anal.* **43**, 261–286 (2011)
- Fletcher, R.: Settlement Archaeology: World-Wide Comparisons. *World Archaeol.* **18**(1), 59–83 (1986)
- Gibrat, R.: Les inégalités économiques. Librairie du Recueil Sirey, Paris (1931)
- Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J.G., Giske, J., Railsback, S.F.: The ODD protocol: A review and first update. *Ecol. Model.* **221**(23), 2760–2768 (2010)
- Johnson, G.A.: Aspects of regional analysis in archaeology. *Annu. Rev. Anthropol.* **6**, 479–508 (1977)
- Liu, L.: Settlement patterns, chiefdom variability, and the development of early states in North China. *J. Anthropol. Archaeol.* **15**(3), 237–288 (1996)
- Marcus, J., Sabloff, J. A.: The ancient city: New perspectives on urbanism in the old and new world. Resident Scholar book, vol. 2005. School for Advanced Research, Santa Fe (2008)
- Pennisi, E.: Conquering by Copying. *Science* **328**(5975), 165 (2010)
- Pumain, D.: Pour une théorie évolutive des villes. *Espace géogr.* **26**(2), 119–134 (1997)
- Pumain, D., Robic, M.-C.: Les Quarante Glorieuses de l'Espace géographique. *L'Espace géogr.* **1**(41), 1–3 (2012)
- Renfrew, C.: Trade as action at a distance: questions of integration and communication. In: Sabloff, J.A., Lamberg-Karlovsky, C.C. (eds) *A School of American Research Book*, 25 cm. Papers from a seminar held in Santa Fe, Oct. 28–Nov. 2, 1973, pp. 3–60. School of American Research. University of New Mexico Press, Albuquerque (1975)
- Renfrew, C., Poston, T.: Discontinuities in the endogenous change of settlement pattern. In: Renfrew, C., Cooke, K.L. (eds.) *Transformations: Mathematical Approaches to Culture Change*, pp. 437–461. Academic Press, New York (1979)
- Sanders, L.: Regards croisés de géographes, économistes et archéologues sur la hiérarchie des systèmes de peuplement: de l'empirie aux systèmes complexes. *Région et Développement*, 36 (2012)
- Schmitt, C.: Modélisation de la dynamique des systèmes de peuplement: de SimpopLocal à SimpopNet. Ph.D. thesis, Université Paris 1 Panthéon-Sorbonne (2014)
- Turchin, P., Gavrillets, S.: Evolution of complex hierarchical societies. *Social Evolution & History* **8**(2), 167–198 (2009)
- Verhulst, P.F.: Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles* **18**, 1–41 (1845)

Urban Dynamics and Simulation Models

Pumain, D.; Reuillon, R.

2017, XXII, 123 p. 40 illus., 27 illus. in color., Hardcover

ISBN: 978-3-319-46495-4