Chapter 4
An Incremental Multi-Modelling Method to Simulate Systems of Cities' Evolution

Abstract Explaining the evolution of urban systems at large spatio-temporal scales is uneasy. Processes are frequently unobserved empirically and equifinality is a challenge for any generative explanation models. We try to address the causation challenge in urban modelling by proposing a multi-modelling framework for the comparison of several model structures. Each structure represents a combination of mechanisms translating alternative or complementary hypotheses about the processes at play. This approach implies that the conception, implementation and evaluation of the model(s) integrate a diversity of mechanisms. Their contribution to the explanation of urbanization is evaluated in time and space by confronting models to empirical data through an interactive visualization platform. We argue that multi-modelling can provide an alternative way to account for the possible causes generating observed patterns, between traditional approaches such as 1/simple models focusing on a single cause (as is often the case for proving a theory) or 2/very complex models including all possible mechanisms at once (as it might prevent from distinguishing their individual contribution).

4.1 Introduction

Given the complexity of 'real' urban systems, our plea in the introduction was for parsimonious but fully explored models, and for multiple models which account for the equifinality playing in the model (several mechanisms can produce the observed pattern) as well as in reality (without historical and detailed information on the actual processes at work, we only have theoretical hypotheses on how to model cities). These reasons argue for a model-building framework which allows multiple modules to be assembled and combined, each of which represents a particular hypothesis as to the urban dynamics needed to model a particular system of cities at a particular time. This framework should allow the different structures of models to be evaluated in the same way and consequently compared in their ability to simulate an observed trajectory of cities' growth. Because of the diversity of possible factors and theories for explaining the diversity of cities' trajectories, we need a visualization that displays the same kind of information for different model structures and city attributes. The interactive

feature provided by an online (automated) application serves this purpose quite well, and makes it possible for the user to explore different aspects of model results with the same visualization design, thus easing the process of model comparison. Moreover, the exploration of a variety of results and model outputs is guided by a selection of representations used in geography to compare systems of cities, to analyze models and their residuals.

This chapter describes a multi-modelling method that was developed and applied to the system of Soviet and post-Soviet cities (but could be later transferred to any system of cities for which we have historical data). We first present the theoretical, methodological and technical framework of multi-modelling (Sect. 4.2), before detailing the resulting family of models that was developed to simulate Soviet and post-Soviet city growth (Sect. 4.3). Section 4.4 presents the results obtained and the geographical knowledge that one can draw from such a methodology. Section 4.5 describes the application built to explore and communicate these results interactively and online. Section 4.6 concludes.

4.2 Methodological and Technical Framework for Multi-modelling Systems of Cities

In this section, we review the theories competing for the explanation of the evolution of systems of cities (or their stylized facts). We also review previous attempts to build multi-modelling frameworks and expose our own approach.

4.2.1 Complementary and Competing Theories

As stated in the first chapter, systems of cities give rise to very robust regularities over time and space. For instance, the Zipf's distribution of city sizes has been described and studied for almost a century (Lotka 1925 and Nitsch 2005 for a meta-analysis). This 'mystery' (Krugman 1996a) has fostered a wide range of possible explanations, from random processes to economic, social and geographical rationality. If we focus on causal mechanisms (thus excluding random generative models), we can identify five broad categories of explanations, reflecting one or several theories to account for the evolving sizes, locations and functional specialization of cities within a given system (Pumain et al. 2006; Schmitt et al. 2015; Cottineau et al. 2015b):

1. Spatial interactions and the diffusion of innovations (Pumain 1997; Pumain et al. 2006; Pumain 2006) explain the stability of the distribution of city sizes and the functional differentiation by formalizing exchange mechanisms of competition and cooperation between cities that diffuse social, political and economic innovations in a way that gives an advantage to large cities, thus explaining their fastest growth, rank inertia and inner diversity on the long term.

2. Size effects comprise the theories of agglomeration economies and diseconomies (for a review, cf. Rosenthal and Strange 2001). They explain the existence of cities of different sizes by the different possible equilibria between centripetal and centrifugal economic forces. Centripetal forces refer to matching, sharing, learning and sorting advantages of large cities (Duranton and Puga 2004). Centrifugal forces usually refer to congestion and pollution externalities of population agglomerations (Krugman 1996b).

3. Site effects explain the location of cities, and the spatial distribution of growth which is due to an easy access to some localized resources, may they be natural deposits (oil, river, seaside, climate) or social amenities (patrimonial sites, creative atmosphere). The causal mechanism translating this principle is very simple: cities which are located near advantageous resources attract more people and create specific products at a lesser cost, and therefore tend to grow faster.

4. Situation effects such as the one used in location theories (Reynaud 1841; Christaller 1933; Ullman 1941) explain the regular spacing of cities, their size and specialization by looking at the relative accessibility in the system. For instance, hub locations on transportation networks provide advantageous locations for urban growth, as well as large cities because they provide a larger access to a larger pool of products.

5. Territorial effects finally differentiate cities according to the political territories they belong to and look at factors of common evolution enhanced by public policies (fiscal redistribution for example) and shared habits (with respect to natality, for example). It also explains the particular evolution of capital cities by their specific function in the system (Preston 1979; Brockerhoff 1999; Bretagnolle and Pumain 2010).

Theoretically, there is a simple reason why we should try and combine different theoretical (partial) explanations into a unique model: it is to evaluate the explaining power of different hypotheses and of their combination on an empirical case study (Martin 2015). By allowing different accounts to play in the same simulation, we can compare and order different theories, we identify equifinality for the ones performing equally, we spot areas or periods for which some theories work better than others—thus characterizing the genericity and specificity of different hypotheses— and finally we build a composite theory made of existing complementary mechanisms (Thiele 2015).

4.2.2 A Methodology for Implementing Multi-models

Methodologically, there are examples of complexification of the models proposed by agent-based modellers. The pioneers (Epstein and Axtell 1996) indeed proposed a modelling framework of the Anasazis that started from a simple model and added supplementary mechanisms of individual interactions (trade, reproduction, etc.). This incremental approach has been applied later in geographical (Conte et al. 2012) and ecological (Grimm and Railsback 2012) models. At earlier stages of the modelling

process, we also have examples of organized reviews of the literature aimed at formalizing the pool of competing theories to account for the pattern to simulate, in organization science (Contractor et al. 2000) and health studies (Auchincloss and Roux 2008; Galster 2012).

However, we do not know of many attempts to combine model structures in the same framework and thorough explorations of multiple model structures against empirical data. Indeed, this is recurrent plea in the literature (cf. Batty and Torrens 2005) for which we provide a proposition. The only example we know of this kind of approach is the pioneer one of S. Openshaw (1983; 1988). His 'model-crunching' method produced a way to select efficient model structures of spatial interactions. However, his pool of alternatives was restricted to different mathematical forms of relating spatial interactions to masses and distances between geographical zones, and led to model structures that were not always meaningful and interpretable. We propose a framework which builds on a consistent set of causal mechanisms drawn from the theoretical literature on systems of cities and which will produce models that we can interpret and use for understanding, explaining and predicting urban systems dynamics.

An initial set of mechanisms is implemented in a programming language that enables their combination (in our case, Scala). A model structure is a certain combination of mechanisms, i.e. the core mechanisms plus additional activated mechanisms. All models in the same family are initialized with the same empirical data and are evaluated with the same measures. Those measures can refer to stylized facts to reproduce (for example, a rank-size distribution of city populations) or to empirical patterns (the actual growth of cities and their hierarchical differentiation).

4.2.3   Exploiting the Results of a Family of Models

To assess the characteristics (performance vs. data, equifinality property, genericity or specificity degree) of the mechanisms of our composite theory, we need to calibrate all model structures with the same criteria—or objective function—(Sect. 4.2.3.1). This requires to identify measures of what a good simulation is with respect to its distance to the observed patterns and empirical data, but also to control for unrealistic dynamics (Sect. 4.2.3.2). The different model structures are then compared according to this measure, and analyzed according to the values of parameters for which the best simulation is obtained. The systematic combination of mechanisms allows to estimate the explaining power of a single mechanism (everything else being equal), this explaining power being measured as how much it reduces the distance to the empirical pattern (Sect. 4.2.3.3).

4.2.3.1 Quantitative Measures to Define a Good Simulation at the Micro-Geographical Level

In order to compare simulated systems with empirical systems of cities with respect to the spatial and hierarchical distribution of growth over time, we compute a measure which sums the distance between the simulated population and the observed population for each city of the system. We sum this distance for each time steps for which we can compare simulated populations with empirical ones (typically, a census year):

$$\delta = \sum_t \left(\sum_i (log(P_{o,i,t}) - log(P_{s,i,t}))^2 \right) \tag{4.1}$$

We use logarithms to compare the impact of relative differences in small cities with differences in large cities, and use the power 2 to give a larger weight to large discrepancies in the sum of distances. Finally, we normalize this index by t the number of time steps for which we can compare simulated populations with empirical ones and by n the number of cities simulated, in order to compare systems with different sizes and simulations of different historical lengths.

We assess the quality of a simulation by looking at how small δ is, considering it has passed micro-behaviour validity tests. We control for unrealistic micro-geographical dynamics by checking for each simulation that there is no city with no wealth and that no city produces and consumes more during a step than the wealth it accumulated over time (for more details, cf. Cottineau et al. 2015a). Taking these three criteria into consideration during the calibration process filters the parameters space of a given model structure, excluding portions which lead to unrealistic behaviours of the model during simulation. The minimization of the distance δ (given the two boolean controls) represents the objective function of the (multi-) calibration.

4.2.3.2 The Multicalibration Procedure

The different modules of the model were combined and calibrated using mixin methods (Steyaert et al. 1993, Lucas and Steyaert 1994, Prehofer 1997) in the object-oriented programming framework of the Scala language. These methods allow the implementation of different alternatives for a single trait (in our case: a mechanism of city interaction or growth) and the generation of a source code containing all the possible combinations (and their dependencies in terms of parameters and variables). To run one of the possible implementations of the model, one has to specify an index referring to the corresponding combination, and a vector containing values for all the possible parameters, even when the given mechanism combination does not make use of some of them. Given this functional way of implementation, the multicalibration thus corresponds to the calibration procedure described in Chap. 3, with an additional parameter corresponding to the model index: therefore the genome of a model defined as a combination of mechanisms contains the vector of all parame-

ter values and the index driving the composition of the model. Models are run on the European Grid Infrastructure and evaluated with respect to the fitness function described in Chap. 3. The only difference with the single-model calibration is that we want results for each possible value of the model index parameter. Therefore, the elitism specification of the calibration algorithm has been transformed to keep the best individuals of each subpopulation (models run with a specific index). The top 50 best performing sets of parameters were kept. The mutation specification of the calibration algorithm has also been tuned to favour a fast convergence: the model index has a 10% chance of mutation. This feature facilitates the exchange of efficient solutions between the different model combinations.

Combining the baseline model with five additional mechanisms for two different time periods (1959–1989 and 1989–2010) resulted in 64 different model implementations (64 values for the model index), approximately 72 million evaluations of which were drawn the best 3200 parameter sets evaluated during the multicalibration (50 for each model instantiation). This database is the one we use to analyze the family of models in the next section.

4.2.3.3   Analyzing the Calibrated Models of the Family

There are three types of analyses that can be drawn from the multicalibrated family of models.

- First, we propose to interpret the overall **performance of the different model structures**, by looking at the shortest distance to the observed pattern obtained for each parsimonious model (the core model plus one additional mechanism). This performance can also be measured as the average distance reduction reached by any model that contains this mechanism compared to models which do not. It means that mechanisms and the theories they formalize can be compared and ordered according to these two criteria, for each spatio-temporal simulation and between territories and time periods. For example, if site effects produce systematically better simulations for a time period, but not in the next one, this process can be said a good candidate for explaining the empirical urban dynamics in the first time span, but another range of explanation might be more relevant to understand the subsequent period.
- Second, for a given structure of model, we propose to interpret the meaning of the **calibrated values of parameters** that give the best simulation. That way, we gain an insight into the strength of different processes and can compare them in different systems (in time or in space).
- Finally, we propose to study the **residuals**, i.e. the cities that cannot be modelled in a satisfactory way with the given structure of model. The magnitude of deviation and the location of those cities tell us about the singularity of their trajectories, that we can try to explain further with by historical events or supplementary explanations. This last analysis is of crucial importance for the geographer as it reveals the areas

of the observed urban evolution that 'resist modelling' (Durand-Dastès 2001) and that suggest the singularity of the realized trajectory of the system.

As we will see in the following sections, such residual trajectories are of particular importance in the Soviet urbanization. However, some cities of this system can be simulated with generic mechanisms and thus the modelling process helps us disentangle the dynamics of cities that are common to other systems and the trajectories of cities that one can only understand if one knows about the history of the Soviet Union.

4.3 A Family of Models of (Post-) Soviet Cities: MARIUS

The application of our multi-modelling framework on the case of (post-) Soviet cities relies on the Simpop principles for modelling cities (Bura et al. 1996; Sanders 2005; Bretagnolle and Pumain 2010) and on a harmonized urban database of 1929 urban agglomerations and their populations over the twentieth and twenty-first centuries (Cottineau 2014a, b). As in the Simpop models, we consider cities as collective agents and model time with 1-year steps. The MARIUS contribution brings up a new way to categorize mechanisms, to order them ex ante given their specificity to the case study (Sect. 4.3.1), as well as a reusable open-source modular implementation (Sect. 4.3.2).

4.3.1 Ordering Possible Causes of Evolution from the Most Generic to the Most Specific

We reviewed five classes of explanation that could account for the regular features of systems of cities in Sect. 4.2.1. They describe systems of cities in general. In the study of a particular system of cities, we expect the realization of general processes to take a particular twist, but we can also expect: 1/other processes to take place, for example political and economic processes shaping the overall geography and affecting cities and 2/the different theoretical processes to appear at different levels of importance in the empirical mix. Indeed, the Soviet and post-Soviet cities exhibit some of the general features of systems cities: a hierarchy of city sizes that follows a power law, the increase of size inequality between cities over time, the spacing of cities in the inhabited space, a specialization of functions and economic interactions. However, we identified empirically the territorial immensity, the importance of subsurface resources and the planned nature of some of the economic interactions during the Soviet Union to be particular and singular features affecting the location and growth rates of specific cities (compared to the generic structure predicted, cf. Cottineau (2014b). With this particular knowledge in mind, we identified and ordered the mechanisms that we think are at play in the evolution of Soviet and post-Soviet cities. We also distinguished between mechanisms as to those which imply interac-
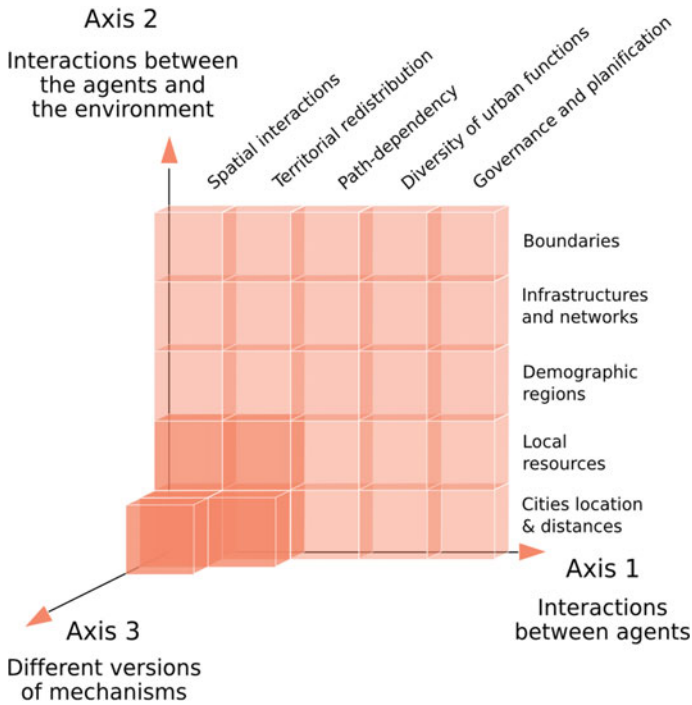
Fig. 4.1 Ordering possible causes of urbanization in the (post-) Soviet case

tions between cities and those which include interactions between cities and their
environment. There might be several ways of implementing each of them, so we
organized our path of model particularization into three axes (Fig. 4.1).

Axis 1 comprises mechanisms of interurban interactions. The first of this kind,
that we think the most generic yet important to model cities in the post-Soviet space,
corresponds to the theory of spatial interactions. The second one refers to territor-
ial effects and consists in a fiscal redistribution between cities of the same political
region. Other mechanisms, more and more specific to the system under study, include
the path dependency and lock-in of interactions' networks, the economic specializa-
tion (and monopoles for example) and the planning policy.

We place on axis 2 the mechanisms that formalize rules of interactions between
city agents and their broader geographical environment, such as the specification of
spatial interactions by actual distances, site effects encompassed in the extraction of
localized resources (a general mechanism, yet of particular importance for under-
standing the contemporary economy and location of growth in Russia and Central
Asia). The imperial construction of the Soviet Union makes us consider that the
different demographic regions and their differentiated paces of urbanization played
a particular role of situation and territorial effects in the trajectories of Soviet and
post-Soviet cities. Situation effect mechanisms such as the accessibility by trans-

portation networks might be of singular relevance to the huge territory of the Soviet Union. Finally, the role of open and closed boundaries of the system appears very singular to this case, but a strong amplifier of territorial effects during the Soviet period, compared to other systems of cities.

The third axis of alternative implementations of the same conceptual processes has been exploited here by allowing different mechanisms to represent spatial interactions, with different levels of complexity.

4.3.2   Implementing Modular Mechanisms

The different implementations of mechanisms have been described in detail in (Cottineau et al. 2015b). In this chapter, we will only outline the main features of the mechanisms actually implemented and evaluated as part of the multicalibration.

4.3.2.1   Size Effects and Spatial Interactions: The Baseline Model

The baseline model includes basic features of cities: their population initialized at the empirical starting point of the simulation, and a wealth estimated as a power law of this population, with a parameter populationToWealth ranging from 1 (no economic size effect) to 2 (larger cities are wealthier). In this baseline model, each city produces and consumes as a power law function of its population at each time steps, with two parameters sizeEffectOnSupply and sizeEffectOnDemand ranging from 1 (no productive/consumptive size effect) to 2 (larger cities are increasingly productive/consumptive per capita) and a normalizing parameter economicMultiplier. Each city then proceeds to an estimation of potential exchanges of value with other cities based on their respective size and distance, following a gravity model of distance exponent distanceDecay ranging from 0 (no distance effect) to 2 (the interaction between cities decreases faster than proportionately with the distance between them). It then shares its supply (/demand) between potential city clients (/city providers) and updates its wealth by adding the amount produced during the current time step, subtracting the total demand, adding unsatisfied demand and subtracting the unsold supplies during the external exchange round. The conclusive operation of a simulation step involves translating the wealth differential into a population gain (or loss), using a power law of exponent wealthToPopulation between 0 and 2.

This simple baseline model has proved unsatisfactory by itself to model the evolution of Soviet cities but other implementations of spatial interactions were shown necessary and sufficient to do so (Cottineau et al. 2015b). Such implementations included a transactional bonus mechanism and a fixed cost of transaction mechanism.

4.3.2.2   Spatial Interactions: The Bonus Mechanism

The bonus mechanism models positive externalities[1] of external exchanges of cities (compared to internal production for internal consumption within a city). It simply adds to the wealth update a term B_i, which is a positive function of the volume traded by a city i to all its urban partners, and the number of cities with which it interacted (relatively to the total number of cities n). When this mechanism is activated, it creates an alternative implementation of the spatial interactions baseline model (cf. axis 3 of Fig. 4.1).

4.3.2.3   Spatial Interactions and Situation Effects: The Fixed Cost Mechanism

The fixed cost mechanism complements the spatial interactions baseline model by including a condition on the realization of exchanges between cities after the computation of interaction potentials. The new rule states that this potential needs to exceed a value fixedCost because each exchange generates transaction costs (Spulber 2007). If the trading potential between two cities is under this value, because of their small size and/or large distance, they will not interact. Otherwise, they will share their supply and demand over the remaining set of potential partners as in the baseline model. During the wealth updating step, each city will subtract the value of fixedCost as many times as the number of transactions it was involved in. When this mechanisms is activated (and when it is activated along with the bonus mechanism), it creates an alternative implementation of the spatial interactions baseline model (cf. axis 3 of Fig. 4.1).

4.3.2.4   Site Effects: The Resource Mechanism

Site effects in MARIUS are understood as subsurface resources. Natural deposits are long known to be favoured locations of growth (Reynaud 1841), but their abundance in the Soviet area makes it a relevant choice for explaining the spatial distribution of growth. Resources can be of two types: coal and hydrocarbons. The location of deposits is initialized empirically from observed patterns, and cities with access to each of these resources are given an extracting advantage that depends on their total wealth (a proxy for the capital they can invest in extracting the resource locally). This mechanism thus has two parameters: coalEffect translates the percentage of wealth added at each time step for cities located on coal deposits (by comparison with cities located elsewhere), and oilAndGasEffect translates the percentage of wealth added at each time step for cities located on oil and gas deposits (by comparison with cities located elsewhere). Both range from -1 (the site has negative externalities on cities' wealth) to 1 (the site has positive externalities on cities' wealth), with 0 corresponding

---

[1]Accounting for knowledge spillovers for example.

to the absence of site effects. This mechanism represents the first increment that we think specific to the Soviet system with respect to the interactions between cities and their environment (cf. axis 2 of Fig. 4.1).

4.3.2.5   Territorial Effects: The Redistribution Mechanism

Territorial effects are implemented in MARIUS as a redistribution of wealth within regions and within countries. At the beginning of a simulation step, cities of the same territory mutualize a share territorialTaxes (from 0 to 1) of their wealth. From this amount, the capital city raises a share capitalShareOfTaxes (from 0 to 1) to sustain its administrative duty. The remaining amount of money is redistributed to every city according to its size (in population). The balance of this redistribution is included in the update of wealth at the end of the simulation step. This mechanism represents the first increment that we think specific to the Soviet system with respect to the interactions between cities (cf. axis 1 of Fig. 4.1).

4.3.2.6   Territorial and Situation Effects: The Urban Transition Mechanism

In this second increment relating to the interaction between cities and their environment, we formalize uneven opportunities of rural immigration for cities of different regions by modelling a logistic curve of the urbanization rate and locating each region on this curve given its level of urbanization at the initial date of the simulation.[2] At each time step, the region moves from one unit on the relative urbanization time, and reach a higher urbanization rate. The cities which belong to each region have an extra growth of population due to rural migration that is a negative function of the urbanization rate. This function is normalized by a parameter ruralMultiplier which possibly ranges from 0 (no migration) to 1 (the population is doubled by rural migrants).

All these increments are combined into 64 model structures that we have calibrated over two periods of time: the Soviet stable era (1959–1989) and the post-Soviet transition (1989–2010).

4.4   Geographical Insights on (Post-) Soviet City Growth from Multi-modelling

By looking at mechanisms' performance, corresponding parameters and residual trajectories, we hope to understand better the probable drivers of urbanization before

---

[2]The logistic curve was estimated on empirical urban and regional demographic data between 1959 and 2010 for 108 regions of the Former Soviet Union (cf. Cottineau 2014b).

and after the crash of the USSR, and to compare the power of different theories in this explanation. Simulations of the Soviet period correspond to models of 30 steps from 1959 to 1989, with an initialization of 1145 cities at their empirical population in 1959 and 3 census check-up dates (1970, 1979 and 1989) for the evaluation. In other words, in Eq. (4.1), $t = 3$ and $n = 1145$. Simulations of the post-Soviet period correspond to models of 21 steps from 1989 to 2010, with an initialization of 1822 cities at their empirical population in 1989 and 2 census check-up dates (2002 and 2010) for the evaluation. In other words, in Eq. (4.1), $t = 2$ and $n = 1822$.

All the results presented come from the multicalibration of the 64 model structures, evaluated with the open database DARIUS on post-Soviet agglomerations,[3] after 400,000 generations of a generic algorithm which objective function was to minimize the distance δ while meeting the microdynamics controls, using parallel computing through OpenMOLE.[4] These results can be explored and replicated within an online application called VARIUS.[5] The point of this section is more about the geographical insights and knowledge that are gained through multi-modelling.

4.4.1 Mechanisms' Performance

From the pool of 64 model structures calibrated for each time period, we first look at the best performance achievable (the controls for realistic dynamics being met) and the corresponding model structure for the given period. Between 1959 and 1989, this best performing model corresponds to a complete model (all mechanisms are active), minus the resource mechanism. The normalized distance to empirical data amounts to 0.0123. Between 1989 and 2010, the best performing model corresponds to a complete model (all mechanisms are active), minus the bonus mechanism. It amounts to a normalized distance to data of 0.0041. These results confirm the intuition that the differentiated urbanization process might more probably be the consequence of a **mix of effects (or partial explanations)**—site, situation, size, territory and interaction—than the result of a single mechanism. Thus, more complete models simulate better the trajectory of all cities in the system (they also have more degrees of freedom during the calibration, some mechanisms balancing others). These first results also show that the **dynamics of post-Soviet cities are on average three times easier to model than the trajectory of Soviet cities**. Is this evidence of some 'normalization' of the economic and political system or does it only attest the low population growth (and even demographic shrinkage on most of the post-Soviet territory) of the last 20 years? We cannot say at this point. However, we can observe that the location of resources and the way we modelled site effects do not help reproduce cities' trajectories before the transition (as the best model performs without this mechanism). This reinforces the empirical impression of diversity of trajectories in neighbouring locations at this

---

[3]http://dx.doi.org/10.6084/m9.figshare.1108081.

[4]http://openmole.org/.

[5]http://shiny.parisgeo.cnrs.fr/VARIUS.

period Cottineau 2014b. However, this mechanism of resource extraction seems to be an important candidate for explaining trajectories during the next period. However, bonified interactions between cities tend to increase simulated deviations from the observed urban trajectories. We see two interpretations to this result: 1/ technically, the bonus parameter permits to model larger demographic growth (cf. Sect. 4.3.2.2), and is thus not required at a time of demographic shrinkage) and 2/ the absence of spillover 'bonuses' from interurban exchange might mirror the new localism of post-Soviet urban strategies, which are less prone to deal with distant, uncertain and costly suppliers from within the Former Soviet Union, but also rely more on the wealth from subsurface resources and/or international partners (Europe, China, Middle East, etc.).

Another way to look at the **performance of single mechanisms** is to compare model structures composed of the baseline model plus a single additional mechanism. For the first period, the best performing parsimonious such model structure involves the **Urban Transition** mechanism. In this case, the normalized distance to empirical δ amounts to 0.0142, which is just over the best performing complete model ($\delta =$ 0.0123), but in the same range of performance. For the post-Soviet period, the best parsimonious model performs 25% worse ($delta = 0.0052$) than a more complete model ($delta = 0.0041$), with the **Resource** mechanism only. This indicates a possible shift in the main drivers of differentiation of urban trajectories before and after the transition. During the late Soviet Union, difference in rural migrant potentials would be the most important criterion to distinguish fast growing cities from more steady relative trajectories. Territorial and temporal lags in the urban transition would have been the important determinants of the evolution of cities in the different parts of the former empire. In the post-Soviet New Independent States, on the other hand, the access to important resources such as oil and gas would explain much better the contrasted destinies of population growth and economic dynamics of cities.

Overall, we then see on Fig. 4.2 that the alternative implementations of the spatial interaction mechanism (Bonus and Fixed Cost, appearing as the Bonus_true and Cost_true in the bars of the Fig. 4.2.) contribute significantly to the reduction of the distance to observed trends, as well as the mechanisms complexifying the environment with which cities interact: Resources and Urban Transition (although differently for the two periods). The Redistributive mechanism is not significant (at a threshold of 0.5% of statistical error) in this average reduction. Finally, everything being equal with respect to the structure of the model, applying it to the latest period gives much better simulations.

We cannot make further comments on the modelled dynamics that simulate (post-) Soviet cities most satisfactorily without looking at the values of the parameters calibrated for a given model structure.

4.4.2 Parameter Values

To simplify the analysis, let us focus on the best performing models for each period, with a mix of four additional mechanisms each. The study of their calibrated parame-
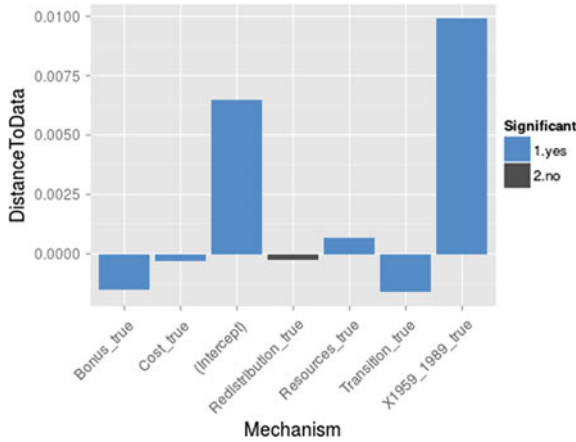
Fig. 4.2 Estimating the contribution of each mechanism to a 'good simulation'. Each bar plots the coefficient estimated during a regression of the normalized distance delta to empirical delta. The intercept gives the average delta for the baseline model during the period 1989–2010. The other bars correspond to the specific contribution of each mechanism, or the average surplus of distance for the period 1959–1989. The *colour* of the bar indicates if the coefficient estimated is significant or not in the regression (*pvalue* < 0.005 in *blue*)

Parameter	Associated Mechanism	1959-1989	1989-2010
populationToWealthExponent	Baseline Model	1.009626528	1.045500
economicMultiplier	Baseline Model	0.459641548	0.5981786
sizeEffectOnSupply	Baseline Model	1.000997749	1.008078
sizeEffectOnDemand	Baseline Model	1.073854144	1.000000
distanceDecay	Baseline Model	0.139061393	0.3090559
wealthToPopulationExponent	Baseline Model	0.388314458	0.6477288
bonusMultiplier	Bonus	79.713516669	-
fixedCost	Fixed Cost	0.370626676	100.0000
oilAndGazEffect	Resource	-	0.03939965
coalEffect	Resource	-	0.001998449
territorialTaxes	Redistribution	0.008163108	0.09916132
capitalShareOfTaxes	Redistribution	0.993792729	0.2643902
ruralMultiplier	Urban Transition	0.019761433	0.001722130

Fig. 4.3 Calibrated parameters of best performing model structures for two periods

ters (Fig. 4.3) reveals insightful variations of the effect of the different mechanisms needed to simulate two sets of very different historical urban dynamics.

The higher value of the parameter populationToWealthExponent for the initialization of the second period (i.e. superlinear scaling of wealth with population compared to the linear relation of the precedent period) indicates a higher economic inequality between cities with respect to their size at the beginning the post-Soviet era, which is necessary to simulate observed trajectories under the modelled assumptions.

Size effects on yearly production and consumption behaviours are almost insignificant for the two periods (sizeEffects ∼1). The exception relates to consumption dur-

ing the Soviet Union, which appears superlinear with population: large cities generate a higher demand per capita during this period under the modelled assumptions.

The reducing effect of distance on potential interactions (distanceDecay) is low in this huge country, compared to empirical estimations on France and the UK (Fotheringham 1981; Baccaïni and Pumain 1998) but doubles over time, from 0.14 to 0.31, suggesting a decrease of large distance transactions under market conditions. This hypothesis is strengthened by the disproportionate increase of the value of the fixed-Cost parameter. As it represents the threshold under which potential interactions are not profitable, it renders a picture of exchanges limited to large volumes between large and neighbouring cities. This framework fits with the descriptions of metropolization and localism within the New Independent States after the transition and under globalization processes.

The resource effect which is significant in post-Soviet urbanization patterns is due to oil and gas deposits, generating a surplus of growth equivalent to 4 points in percentage to cities located there, every year between 1989 and 2010. By comparison, coal cities benefit from twenty times less boosting effect, a consequence of the obsolete economic cycle of this resource.

Redistribution is almost absent from the first period's model (<1% of wealth is taxed, transferred directly to the capital city) but constitutes an important factor of equalization thereafter.

Finally, rural migration appears ten times less important to explain urban trajectories in the second period.

4.4.3 Residual Trajectories

We end the analysis of the best performing models at each time period by looking at the cities that resist modelling, i.e. the urban trajectories which the implemented mechanisms do not succeed in simulating. In particular, we look at the global disper-
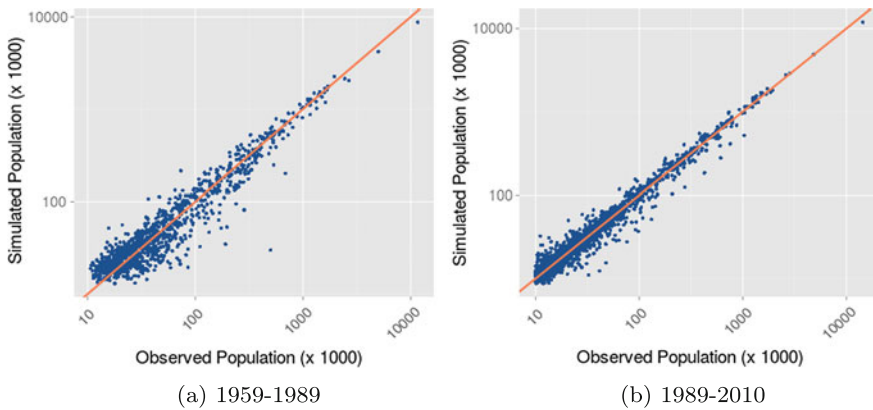

(a) 1959-1989   (b) 1989-2010

Fig. 4.4 Global dispersion of residuals for each period

(a) 1959-1989

(b) 1989-2010

Fig. 4.5 Spatial distribution for maximum residuals for each period

sion of residuals (Fig. 4.4), their spatial distribution (Fig. 4.5) and singular trajectories (Fig. 4.6).

   The global dispersion of residuals indicate that the most recent period is best simulated by the model, and that in general, the most striking outliers correspond

1959-1989			
City	Republic	Observed Population	Simulated Population
Naberezhnye Tchelny	Russia	500,000	30,000
Volgodonsk	Russia	191,000	35,000
Chajkovskij	Russia	86,000	20,000
Balakovo	Russia	197,000	53,000
Bratsk	Russia	285,000	82,000
...	...	...	...
Zaozernyj	Russia	16,000	52,000
Kizel	Russia	37,000	108,000
Cheremhovo	Russia	74,000	215,000
Gremjachnsk	Russia	21,000	56,000
Gornoaltajsk	Russia	46,000	113,000
1989-2010			
City	Republic	Observed Population	Simulated Population
Mirnyja	Russia	41,000	12,000
Sertolovo	Russia	48,000	16,000
Beineu	Kazakhstan	32,000	11,000
Govurdak	Turkmenistan	76,000	28,000
Serdar / Gyzylarbat	Turkmenistan	98,000	37,000
...	...	...	...
Sovetabad	Uzbekistan	11,000	32,000
Zhanatas	Kazakhstan	21,000	50,000
Krasnozavodsk	Russia	13,000	31,000
Gagra	Georgia	11,000	25,000
Nevelsk	Russia	12,000	25,000

Fig. 4.6 Cities that deviate most from their simulated trajectories

to cities which are much more populated 'in real life' than what the model predicts. In the model for 1959–1989 for example, such cities include Naberezhnye Chelny or Volgodonsk, respectively, 16 and 5 times bigger than expected. These cities were indeed flagship industrial projects of the Stalin era, in the automobile and power industries.

On the other hand, negative residuals like Sovetabad or Zhanatas in the most recent model correspond to cities which shrank or grew less than expected given their locations, attributes and predicted interactions. In the post-Soviet Uzbekistan or Kazakhstan, they can be cities deserted by Russian migrants after the crash of the USSR. The models of urban interactions of the MARIUS family are thus not designed to simulate such paths and historical events.

4.5   VARIUS: A Visual Aid to Model Composition and Interpretation

The challenge of analyzing and communicating processes and results of geographical modelling, especially in the context of multi-modelling, calls for effective methods

of visual representation (Batty et al. 2011). Indeed, we want to describe the urban evolution and to explore the adequacy of several simulation models to reproduce this evolution. These aims imply handling large spatio-temporal quantitative datasets and comparing their features with the realized trajectory of the empirical system of Soviet and post-Soviet cities. Visual representations, such as graphics and maps, seem to be the simplest and most powerful way to do so (Tufte 2001), and the ideal basis for argumentation and geographical interpretation. Indeed, the visual representation of models performance and the distribution of spatial residuals provide supplementary elements for face validation of the simulation models (Hermann 1967), beyond quantitative measures used in the automated calibration process.

VARIUS (http://shiny.parisgeo.cnrs.fr/VARIUS/) provides a platform for interactive exploration of models which complements MARIUS model building, and allows to share and open the exploration of simulated urban trajectories online. Indeed, any user can run the combination of mechanisms and parameterization of their choices, and visualize the resulting urban evolution. Opening the black box of model building is necessary for collective model validation and can be eased by the provision of predesigned tools for exploration, besides the open-licencing of data, models and codes[6].

4.5.1 Building the Model Online

The first part of VARIUS application ('What happened?/Census data') consists in a quick presentation of quantitative evidence about the system to simulate that will help selecting the most relevant set of mechanisms to analyze. Basically, it represents the content of the DARIUS database, and represents urban demographic structures in time and space.

- The first interactive map shows the population of all cities in the Former Soviet Union at the chosen date (left chooser, the right slider adjusting the size of circles, Fig. 4.7). This interactive map therefore shows the spatial and hierarchical distributions of cities in the post-Soviet space from the first Russian census in 1897 to the last in 2010 censuses are like transversal photographs of urbanization taken at irregular points in time. To study an evolution between these photographs, the second map provides a more dynamic approach.
- The second interactive map proposes two choosers representing the starting (T) and the stopping date ($T + P$) of a period P under investigation. It relies on the computation of the average annual population growth rate g of cities i during this period P (Eq. 4.2). Average annual growth rates ease the interpretation and comparison of growth and shrinkage trends across irregular periods (typically, intercensus intervals). As a result, maps produced in this section reveal the spatial and hierarchical distribution of growth (red) and shrinkage (blue).

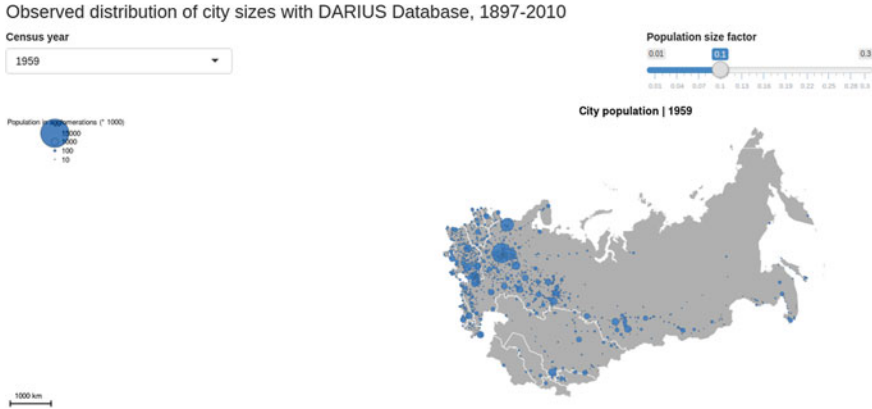---

[6]https://github.com/ClementineCttn/VARIUS.

Fig. 4.7 Exploration of the empirical data. Size of *circles* can be adjusted (*top right slider*) and the year of census for which the data are displayed is selected via the *top left* menu

$$g_{i,T,P} = \left(\sqrt[p]{\frac{Population_{i,T+P}}{Population_{i,T}}} \right) * 100 \qquad (4.2)$$

- A third interactive map represents the spatial distribution of fixed categories of cities (statuses, access to resources, locations). Seven attributes are available to display: the status of national and regional capital, 342 mono-industry towns as defined by the Russian Federation in 2013[7]; the location in areas of coal and hydrocarbon extraction; accessibility by rail and by air, and absolute east/west location.

In the part 'How to Simulate it', VARIUS provides tools to analyze the modular structure of the MARIUS family of models, and estimates the contribution of different mechanisms and their combination to the reduction of discrepancies between observed and simulated urban trajectories. Two approaches are offered to this estimation. First, a linear regression of the fitness measure performed on all calibrated models allows the users to identify by themselves the type of model that they want to run, activating or deactivating the mechanisms that they find interesting for their performance (or underperformance). The second approach yields an optimal combination of mechanisms for simulating the observed evolution in the FSU, given the number of mechanisms to combine (if one seeks parcimony above all). The platform then goes on to providing a parameterization board to run the model online.
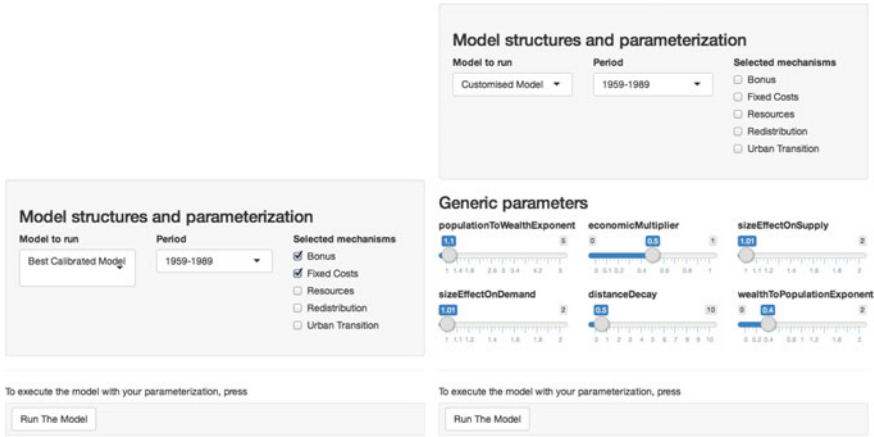
---

[7]http://www.veb.ru/common/upload/files/veb/br/mono/list342.pdf.

Fig. 4.8 Parameterization board for running any MARIUS model

4.5.2   Running the Model Online

The central part of the application gives access to the model itself, by making it possible to run it online, given a particular set of mechanisms and parameters. The mechanism structure and parameter values necessary to execute the model are defined in the tab 'Run a MARIUS model', and can be set automatically as a result of previous calibration, or explored manually by the user. By choosing the option 'Best calibrated model' (Fig. 4.8) in this tab, the user focuses on best performing models only, and explores their performance according to time periods (before or after the dissolution of the USSR) and different mechanisms' combinations. If one does not focus on precalibrated models but seeks to explore the effect of single parameters on the model's behaviour and simulated patterns, it is possible to run a 'Customized model' instead and to define manually its parametrization. With this option, the user can choose a value for each parameter in the intervals considered realistic and interesting (the ones used for the calibration process). For the baseline model, six parameters need to be defined, whereas additional mechanisms include one or two parameters each.

4.5.3   Analyzing Results Online or 'How Close Are We?'

The analyzis of model simulations is proposed at three scales: the macro geographic level of urban hierarchy, the micro level of cities, and a meso level of categories of cities, based on their function or status.

- At the macroscale, two visualizations help to explore the quality of a simulation of city sizes. In the first one, cities are ordered by population and plotted against their ranks in the system in a Zipf tradition. The simulated hierarchy of cities (described by the slope of the curve and its deviation from a straight line) can be compared with empirical observation of over time (in grey). Being able to reproduce this pattern is a basic requirement for any model to achieve. It means that the distribution of city sizes match the observed one at the last date, but also the evolution towards an increasing unevenness over time. The second representation displays the dispersion of residuals (as in Fig. 4.4), that is the distribution of simulation errors for each city at the last step of simulation. A model with a perfect fit would be characterized by a distribution of blue dots along the orange line, meaning that every city's simulated population equals its observed population. This limit case is not one we hope to achieve with parsimonious models. However, models are considered good enough if they exhibit a small and symmetrical dispersion around this line. This representation helps spotting outlier cities that the model is not suited to simulate (the dots significantly away from the identity line).
- Outliers are the object of representation of the next tab in the application, at the micro level of cities. Indeed, the maps plotted here show cities whose simulated trajectories deviate most from the empirical one. Using appropriate thresholds and looking at the spatial distribution of these cities at the different points in time, the user is given a glimpse of the spatial and hierarchical pattern of residuals. This sometimes gives way to hypotheses (e.g. the growth of large cities is underestimated by the simulation) that can be tested in the final section.
- At a meso level of groups of cities, this multiple regression aims to profile residuals according to the attributes of cities. We consider this last section as the beginning of a new reflection to complement the model with new mechanisms. The intuition for these new mechanisms would come from the observation of a semi-general feature non-included in the current modules but shared by a large group of cities displaying the highest residuals.

This application is therefore a communication tool for the work done on MARIUS as much as a basis for a work-in-progress regarding model building.

4.6 Conclusion

Because the family of models is designed as a modular framework and because methods were developed to handle modular models in the evaluation processes, the expansion of the model via new mechanisms or its transfer to different urban contexts is made easier and more straightforward, reducing the development cost to the new mechanisms to implement only.

This is a great step forward in the conception of the family of simulation models for several reasons.

In terms of model design, the contribution of a component can be soundly confronted to the others: the introduction of a new component in the family of models leads to new explorations of models dynamics, producing measures that assess both the performance of the new mechanism, but also the performance of the rest of the mechanisms when they interact with it. Each mechanism addition thus reinforces the confidence we have in the mechanisms of the model family, refining the conditions under which they perform best, in an entirely tractable and comparable way.

Furthermore, exploring the models structures allows orienting the model design either towards parsimony or specificity, which is a great support when it comes to strengthening a composite theory where several stylized facts may interfere depending on the chosen level of granularity.

Finally, the fact that a family of models produces comparable and validated trajectories (with respect to their mechanisms, parameterization, and the data against which they are calibrated) would make prospective outcomes, such as territorial policies scenarii, more directly interpretable. This knowledge thus enables to develop policies that are especially adapted to local situations.

References

Auchincloss, A.H., Roux, A.V.D.: A New Tool for Epidemiology: The Usefulness of Dynamic-Agent Models in Understanding Place Effects on Health. Am. J. of epidem. **168**(1), 1–8 (2008)

Baccaïni, B., Pumain, D.: Les migrations dans le systéme des villes françaises de 1982 á 1990. Population **53**(5), 947–977 (1998)

Batty, M., Vargas-Ruiz, C., Smith, D., Serras, J., Reades, J., Johansson, A.: Visually intelligible land use transportation models for the rapid assessment of urban futures. CASA Working Paper vol. **163** (2011)

Batty, M., Torrens, P.M.: Modelling and prediction in a complex world. Futures **37**(7), 745–766 (2005)

Bretagnolle,.A., Pumain, D.: Simulating urban networks through multiscalar space-time dynamics: Europe and the united states, 17–20th centuries. Urban Stud. **47**(13), 2819–2839 (2010). 00043

Brockerhoff, M.: Urban growth in developing countries: a review of projections and predictions. Popul. Dev. Rev. **25**(4), 757–778 (1999). 00067

Bura, S., Guérin-Pace, F., Mathian, H., Pumain, D., Sanders, L.: Multiagent systems and the dynamics of a settlement system. Geogr. Anal. **28**(2), 161–178 (1996). 00160

Christaller, W. (1933). *Die zentralen Orte in Süddeutschland: eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*. Jena, Fischer. 00316

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., Loreto, V., Moat, S., Nadal, J.P., Sanchez, A., Nowak, A., Flache, A., San Miguel, M., Helbing, D.: Manifesto of computational social science. Eur. Phys. J.: Spec. Topics **214**, 325–346 (2012)

Contractor, N., Whitbred, R., Fonti, F., Hyatt, A., O'Keefe, B., Jones, P.: Structuration theory and self-organizing networks. In: Organization Science Winter Conference, Keystone, CO (2000) 00017

Cottineau, C.: DARIUS Database (2014a)

Cottineau, C.: L'évolution des villes dans l'espace post-soviétique. Observation et modélisations. PhD thesis, Université Paris 1 Panthéon-Sorbonne (2014b) 00001

Cottineau, C., Chapron, P., Reuillon, R.: Growing models from the bottom up. an evaluation-based incremental modelling method (ebimm) applied to the simulation of systems of cities. Journal of Artificial Societies and Social. Simulation **18**(4), 9 (2015a)

Cottineau, C., Reuillon, R., Chapron, P., Rey-Coyrehourcq, S., Pumain, D.: A modular modelling framework for hypotheses testing in the simulation of urbanisation. Systems **3**(4), 348 (2015b)

Durand-Dastès, F.: *Les concepts de la modélisation en analyse spatiale*. Paris, Hermés, Lavoisier (2001) 00000

Duranton, G., Puga, D.: Micro-foundations of urban agglomeration economies. Handbook Reg. Urban Econ. **4**(2063–2117), 01533 (2004)

Epstein, J.M., Axtell, R.L.: Growing Artificial Societies: Social Science from the Bottom Up. MIT press, Cambridge (1996)

Galster, G.C.: The mechanism(s) of neighbourhood effects: Theory, evidence, and policy implications. In Neighbourhood effects research: New perspectives. Springer pp. 23–56 (2012)

Fotheringham, A.S.: Spatial structure and distance-decay parameters. Ann. Assoc. Am. Geogr. **71**(3), 425–436 (1981). 00302

Grimm, V., Railsback, S.F.: Pattern-oriented modelling: a multi-scope for predictive systems ecology. Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci. **367**(1586), 298–310 (2012)

Hermann, C.F.: Validation problems in games and simulations with special reference to models of international politics. Behav. Sci. **12**(3), 216–231 (1967)

Krugman, P.: Confronting the mystery of urban hierarchy. J. Japn. Int. Econ. **10**(4), 399–418 (1996a). 00228

Krugman, P.: Urban concentration: the role of increasing returns and transport costs. Int. Reg.Sci. Rev. **19**(1–2), 5–30 (1996b). 00333

Lotka, A.J.: Elements of Physical Biology. Williams & Wilkins Company, Baltimore (1925)

Lucas, C. Steyaert, P.: Modular inheritance of objects through mixin-methods. In Proceedings of the 1994 Joint Modular Languages Conference (JMLC); Ulm, German, 28–30 September, pp. 273–282 (1994)

Martin, R.: Rebalancing the spatial economy: the challenge for regional theory. Terri. Polit. Gov. **3**(3), 235–272 (2015). 00001

Nitsch, V.: Zipf zipped. J. Urban Econ. **57**(1), 86–100 (2005)

Openshaw, S.: From data crunching to model crunching: the dawn of a new era. Environ. Plann. A **15**(8), 1011–1012 (1983)

Openshaw, S.: Building an Automated Modeling System to Explore a Universe of Spatial Interaction Models. Geographical Analysis **20**(1) (1988)

Prehofer, C.: Feature-oriented programming: A fresh look at objects. In ECOOP'97 Object-Oriented Programming; Springer: Berlin and Heidelberg, German, pp. 419–443 (1997)

Preston, S.H.: Urban growth in developing countries: a demographic reappraisal. Popul. Dev. Rev. **5**(2), 195–215 (1979). 00231

Pumain, D.: Pour une théorie évolutive des villes. Espace géographique **26**(2), 119–134 (1997). 00138

Pumain, D. (ed.): Hierarchy in Natural and Social Sciences. Methodos Series. Springer, Berlin (2006)

Pumain, D., Paulus, F., Vacchiani-Marcuzzo, C., Lobo, J.: An evolutionary theory for interpreting urban scaling laws. Cybergeo: Eur. J. Geogr. (2006) 00000

Reynaud, J.: Villes. Encyclopédie nouvelle **8**, 670–687 (1841). 00034

Rosenthal, S.S., Strange, W.C.: The determinants of agglomeration. J. Urban Econ. **50**(2), 191–229 (2001)

Sanders, L.: Intelligence artificielle et agents collectifs: le modèle EUROSIM. In: SAGEO 2005, Colloque International de Góomatique et d'Analyse Spatiale, pp. XX. Collection, Actes Avignon, CD-ROM (2005). ISBN 2-910545-067

Schmitt, C., Rey-Coyrehourcq, S., Reuillon, R., Pumain, D.: Half a billion simulations: evolutionary algorithms and distributed computing for calibrating the SimpopLocal geographical model (preprint). Planning and Design, Environment and Planning B (2015)

Spulber, D.F.: Global Competitive Strategy, vol. 00091. Cambridge University Press, Cambridge (2007)

Steyaert, P., Codenie, W., D'Hondt, T., de Hondt, K., Lucas, C., van Limberghan, M.: Nested mixin-methods in Agora. In ECOOP'93?Object-Oriented Programming; Springer: Berlin and Heidelberg, German, pp. 197–219 (1993)

Thiele, J.C. (2015). Replicating and breaking models: good for you and good for ecology. *Oikos*, (January):1–6

Tufte, E.R.: The Visual Display of Quantitative Information. Graphics Pr, 2nd edn (2001)

Ullman, E.L.: A theory of location for cities. Am. J. Soc. **46**(6), 853–864 (1941)