



HAL
open science

Modelling in Demography: From Statistics to Simulations

Jakub Bijak, Daniel Courgeau, Robert Franck, Eric Silverman

► **To cite this version:**

Jakub Bijak, Daniel Courgeau, Robert Franck, Eric Silverman. Modelling in Demography: From Statistics to Simulations. Eric Silverman. Methodological Investigations in Agent-Based Modelling, 13, Springer, pp.167-187, 2018, Methodos series, 10.1007/978-3-319-72408-9_9 . hal-02079052

HAL Id: hal-02079052

<https://hal.science/hal-02079052>

Submitted on 29 Jun 2023

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

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

Chapter 9

Modelling in Demography: From Statistics to Simulations

Jakub Bijak, Daniel Courceau, Robert Franck, and Eric Silverman

9.1 An Introduction to Demography

Demography, the study of population change, is a discipline with a lengthy and storied history. Most consider demography to have begun approximately 350 years ago, through the work of Graunt (1662), and the field continues to evolve today and incorporate new methods (Courceau et al. 2017). As discussed in Parts I and II, the advent of agent-based modelling has brought with it the possibility of applying simulation methodologies to the social sciences, and in this respect demography is no exception.

Recent work in demography has identified agent-based modelling in particular as a method with particular relevance for demographers; Billari et al. suggest the incorporation of these methods could result in the development of a new subfield of *agent-based computational demography*, or ABCD (Billari and Prskawetz 2003). ABMs have also been cited as a way to increase the theoretical relevance of demography, by inspiring demographers to delve deeper into social theory in their quest to design and parameterise more sophisticated models (Silverman et al. 2011; Burch 2002, 2003).

In this chapter we will go further and propose that ABMs, beyond just heralding the birth of a new subfield, have the potential to push the demographic research agenda in a new direction.¹ We will summarise the historical development of demography from Graunt's time until the present day, and demonstrate the demography has displayed a penchant for incorporating new methodologies and

¹This chapter is based upon ideas from Courceau et al. (2017), please see the original paper for some additional detail of some finer points related to demographic history and knowledge.

building upon the work of years past. We will propose that ABMs and related simulation methods can form the foundations for a new *model-based demography*, a step forward for the progression of demography in the face of an ever more complex and changeable world.

In Sect. 9.2 below, we will discuss the historical development of demographic methods. We will follow this in Sect. 9.3 with a description of the challenges faced by demography due to uncertainty and complexity in human populations. Section 9.4 will present the means by which demographers can incorporate simulation methods into the scientific programme, and Sect. 9.5 will bring this all together into a proposal for a model-based demography.

9.2 The Historical Development of Demography

Demography, as mentioned above, is commonly thought to have begun with the work of Graunt in the seventeenth century (Graunt 1662). While the discipline itself may be ancient, over time the methods used in demography have evolved continuously, and demographers have incorporated a wide variety of approaches into their methodological toolbox.

Note that throughout this chapter, we will refer to these methodological changes as ‘paradigms’, but we are using the term somewhat differently from the modern-day Kuhnian interpretation (Kuhn 1962). Here we use ‘paradigm’ to refer to the relationship between observed phenomena within a population and the key factors of mortality, fertility and migration which are used to explain population change in demography. We will identify four main paradigms over the course of demography’s history, and outline the primary differences between each and how these approaches have built on one another cumulatively over the centuries.

Following Courgeau et al. (2017), we will recall Bacon’s seventeenth-century elaboration of the inductive method as we begin to discuss the origins of demographic thought (Bacon 1863)[aphorism 19]:

There are and can be only two ways of searching into and discovering truth. The one flies from the senses and particulars to the most general axioms, and from these principles, the truth of which it takes for settled and immovable, proceeds to judgement and to the discovery of middle axioms. And this way is now in fashion. The other derives from the senses and particulars, rising by a gradual and unbroken ascent, so that it arrives at the most general axioms at last of all. This is the true way, but as yet untried.

Here Bacon sets up a contrast between the dominant methods in his time, and a second approach which became a foundational principle for modern scientific thought. In the former case, axioms are derived from human notions and intuitions, rather than observation of nature itself. Courgeau et al. argue that the ‘idols’ which derive from such an approach are still a problem in certain areas of science, even in the modern era (Courgeau et al. 2014).

Bacon's alternative proposal requires detailed observation of the object of study and the development of axioms by induction. Through observation and experimentation, we can discover the principles governing the natural or social properties we seek to study. As argued by Franck (2002b), induction in the Baconian method requires that these principles are such that if they were not present, ultimately the properties we observe would take an entirely different form.

These foundational ideas were essential to the development of the first era of demography. Before Graunt, human events like births and deaths were not something to be studied or predicted, but were instead within the strict province of God's plans for humanity. Graunt instead brought us the concept of the *statistical individual* – an important concept which we will revisit later in Part III – which would experience abstract events of fertility and mortality (Graunt 1662; Courgeau 2007). This key conceptual revolution allowed for the development of a science of human population, and thus to the advent of demography, epidemiology, and other related fields.

Graunt also demonstrated the importance of probability in population studies, based upon the work of Huyghens (1657). He was able to use an estimation of the probability of death to estimate the population of the city of London, and in so doing was the first to use the concept of a statistical individual to examine population change scientifically. This inspired the work of contemporaries, such as Sir William Petty and his *Political Arithmetick* (Petty 1690), and was significant for the development of political economics in general. In the ensuing decades, and the following century, European thinkers continued to develop this school of thought.

9.2.1 *Early Statistical Tools in Demography*

The concept of epistemic probability brought forth by Bayes (1763) and Laplace (1774, 1812), had a significant impact on demography. We will not go into a great deal of detail on the specifics here; for that, please turn to Courgeau (2012) and Courgeau et al. (2017) for a more in-depth summary. Speaking broadly, the advent of these techniques allowed demographers to answer more salient questions about human events through the use of prior probabilities, which had notable benefits for the growing field of population science. Similarly, the least-squares method drawn from astronomy began to be applied to demography as well, and over the course of the nineteenth century this method became quite widely used (Courgeau 2012). Censuses becoming more widely used meant that extensive data was also much more accessible.

However, these early statistical tools assumed that the variables under investigation displayed a certain mathematical structure, and these structures are not necessarily evident in the real world. This can lead to the *ecological fallacy*, meaning that aggregate, population-level data cannot be applied to the study of individual-level behaviour. The data being collected during this period was also entirely period-based or cross-sectional (Courgeau 2007). This *cross-sectional*

paradigm implied that social factors influencing individuals are a result of aspects of the society surrounding them (i.e., political, economic, or social characteristics). As we will soon see, this separation of individuals and their social realities does not hold up under further scrutiny in many cases.

9.2.2 Cohort Analysis

The development of the *cohort analysis paradigm* happened in the United States following World War II. Ryder (1951) was one of the early proponents, but Henry (1959) further developed and formalised the cohort approach. Courgeau summarises the core concept of the cohort analysis paradigm (Courgeau 2007, p. 36):

The demographer can study the occurrence of only a single event, during the life of a generation or cohort, in a population that preserves all its characteristics and the same characteristics for as long as the phenomenon manifests itself.

In order for these analyses to work well, however, we must assume that the population is homogeneous and that any interfering phenomena must be independent. Such restrictions meant that demographers quickly sought out methods to study heterogeneous cohorts and interdependent phenomena. Thus Aalen developed a demographic application of a general theory of stochastic processes (Aalen 1975).

9.2.3 Event-History Analysis

The resultant *event-history paradigm* built upon these foundations allowed us to study the complex life histories of individuals (Courgeau and Lelièvre 1992). We are able to identify how both demographic and non-demographic factors affect individual behaviour. Of course, these analyses require extensive data; we need to follow individuals through their lives and collect information on their individual characteristics and the events which befall them. This means that longitudinal surveys become highly important in this type of demographic research.

The event-history paradigm enforces a collective point of view, in which we estimate the parameters of a random process that affects all individuals, their trajectories through life, via analysis of a sample of individuals and their characteristics. This is perhaps conceptually difficult, but in essence we are seeking understanding of a process underlying all of these individual trajectories, rather than insight into the individuals themselves. Again we are studying *statistical individuals*, not real, observed individuals.

However, in contrast to the ecological fallacy of the cross-sectional paradigm, here we may fall afoul of the *atomistic fallacy*, in which our focus on individual characteristics leads us to ignore the broader, societal context in which individual

behaviours develop. As described in Part II, individual behaviours are inextricably tied to the complex, multi-layered society in which they live, so isolating these processes can lead to misleading results and incorrect conclusions.

9.2.4 *Multilevel Approaches*

In order to surpass these limitations, we must introduce the concept of social groupings, which can include families, social networks, workplaces, political affiliations, and many more besides. Alongside these concepts demographers developed multilevel analyses to better link the individual with the social context around them (Mason et al. 1983; Goldstein 1987). This *multilevel paradigm* can resolve the gap between population-level models and individual event-histories:

The new paradigm will therefore continue to regard a person's behaviour as dependent on his or her past history, viewed in its full complexity, but this behaviour can also depend on external constraints on the individual, whether he or she is aware of them or not. (Courgeau 2007, pp. 79–80)

Both the ecological and atomistic fallacies fade away under this new approach, given the direct links that are now possible between individuals and their social context:

The ecological fallacy is eliminated, since aggregate characteristics are no longer regarded as substitutes for individual characteristics, but as characteristics of the sub-population in which individuals live and as external factors that will affect their behaviour. At the same time, we eliminate the atomistic fallacy provided that we incorporate correctly into the model the context in which individuals live. (Courgeau 2007, pp. 79–80)

9.2.5 *Cumulativity*

As demonstrated in the brief historical summary above, demography has advanced over the centuries due to a steady process of advancement through a series of paradigms. Each new paradigm has taken previous approaches as a starting point, identified their shortcomings and offered a means to overcome them. Having said that, the new paradigms have not eliminated the old; period, cross-sectional and cohort analyses remain relevant today, and are still used when the research question being posed would be suitably answered by one of those approaches. This is reminiscent of the situation in physics, where Newtonian physics is still perfectly relevant and useful in situations where relativistic effects have little or no impact.

In demography, following Courgeau et al. (2017), we describe this as a *cumulativity* of knowledge, in which new paradigms have added new perspectives while retaining the insights provided by previous ones:

Table 9.1 Four successive paradigms in demography

| Paradigm | Period | Key focus |
|---------------|--------|---|
| Period | 1662– | Macro-level phenomena |
| Cohort | 1950s– | Macro-level phenomena, measured along the cohort dimension |
| Event history | 1980s– | Micro-level phenomena |
| Multilevel | 1980s– | Macro-, micro-, and meso-level phenomena, measured from multiple perspectives |

Cumulativeness of knowledge seems self-evident throughout the history of population sciences: the shift from regularity of rates to their variation; the shift from independent phenomena and homogeneous populations to interdependent phenomena and heterogeneous populations; the shift from dependence on society to dependence on the individual, ending in a fully multilevel approach. Each new stage incorporates some elements of the previous one and rejects others. The discipline has thus effectively advanced thanks to the introduction of successive paradigms. (Courgeau 2012, p. 239)

Table 9.1 provides a summary of the four paradigms described above. Each of the four seeks to examine the key processes in population sciences – mortality, fertility and migration. However, each paradigm characterises the relationships between these processes differently.

These paradigms are not static, either; each is undergoing continuous refinement over time. In addition, these four current paradigms still struggle to resolve certain types of research questions. For example, examining interactions between different elements of complex populations remains beyond the scope of even the cutting-edge event-history approach. This difficult problem, often referred to as the *micro-macro link* in demographic circles, is a familiar one from our perspective, forming one a significant motivation for deploying agent-based models and social simulations in general. Conte et al. offer a succinct description of the micro-macro link²:

...the loop process by which behaviour at the individual level generates higher-level structures (bottom-up process), which feedback to the lower level (top-down), sometimes reinforcing the producing behaviour either directly or indirectly. (Conte et al. 2012, p. 336)

In addition, the micro-macro link is not necessarily uni-directional; higher-level actions, for example political decisions, can affect individual behaviours, which might then necessitate additional policy measures, and so forth. Multilevel approaches cannot cope with this kind of bidirectional effect. As we will see, a model-based demography may be better-placed to help demography cope with this complex aspects of population change.

²We would go further, however, and note that sometimes the higher-level behaviours can go in the opposite direction from those at the lower level, producing what Boudon (1977) refers to as “perverse effects”.

9.3 Uncertainty, Complexity and Interactions in Population Systems

As described in Part II, studying the complexities of human social interaction introduces a host of challenges for the modeller. These challenges are worth re-examining in the specific context of demography, as here they take on a somewhat different character than in other social sciences. Here we will identify the three primary epistemological challenges facing demography today, which will further inform our development of a model-based research agenda.

Demography, while incorporating aspects of both individual- and population-level behaviour and their attendant complexities, benefits from having frequent access to very rich datasets, due to the inherent usefulness of those datasets for governments throughout the world. Demographic data also displays strong and persistent relationships, and much critical information on future population dynamics is already embedded in a population's age structures. The long-term empirical focus of demography has allowed for these relationships to be examined in significant detail (Xie 2000; Morgan and Lynch 2001).

9.3.1 *Uncertainty in Demographic Forecasts*

The three main processes of population change – mortality, fertility, and migration – all display significant amounts of uncertainty (Hajnal 1955; Orrell 2007). However, the relative levels of uncertainty differ between them; mortality is generally considered the least uncertain, and migration the most uncertain (National Research Council 2000). As demographers have come to accept the significant challenge posed by uncertainty, statistical demography has grown significantly in recent decades. Courgeau refers to this as the “return of the variance” to demography (Courgeau 2012).

The limits of predictability in demographic forecasting has been a topic of significant discussion within the demographic community (see Keyfitz 1981; Willekens 1990; Bijak 2010). Demographers have argued that forecasts should move from deterministic to probabilistic approaches, for example Alho and Spencer (2005). The field also acknowledges that predictions beyond a relatively short time horizon – a generation or so at most – have such high levels of uncertainty that scenario-based approaches to forecasting should be favoured (Orrell and McSharry 2009; Wright and Goodwin 2009).

9.3.2 *The Problem of Aggregation*

Here we refer once again to the ecological and atomistic fallacies described above. While the advent of the event-history approach and related methodologies like

microsimulation has moved demography away from focusing exclusively on either the individual or the population, these methods are still relatively new (Willekens 2005; Courgeau 2007; Zinn et al. 2009). Microsimulation models are both multi-level and multi-state, meaning that individuals can move between states (such as health status, age group, socioeconomic class, etc.) according to transition probabilities estimated from survey data or census data.

However, while these methods are certainly powerful, the challenge for demographers has been their ever-increasing data requirements. The parameter space explodes in size as the ambition of these models grows, and thus demographers find themselves at the mercy of either data that is too limited to accommodate their research questions, or are simply unable to collect sufficient data in the first place due to prohibitive cost or organisational difficulties. We will examine this particular point in more detail in Sect. 9.8 when we discuss ‘feeding the beast’.

9.3.3 *Complexity vs. Simplicity*

The third main epistemological issue for demographers is a direct consequence of the challenges of uncertainty and aggregation. While the temptation in demography today is to tend toward ever more complex and sophisticated models, whether these models are actually more powerful than their simpler neighbours is still an open question. Demography is fundamentally a discipline focused on predicting population change, and in that respect, there is no evidence to suggest that complex models outperform simpler ones (Ahlburg 1995; Smith 1997).

Having said that, however, if we were to react too strongly to this revelation and throw aside complex models in favour of simpler ones, we may not achieve the results we desire. Developing detailed understanding of demographic processes may require coping with highly complex datasets and interactions. In those instances, simplicity may abstract away too many of the relevant factors for us to identify the key elements in the processes we wish to study. Prediction, after all, is a key goal in demography, but is far from the *only* goal; understanding and explanation are just as valid goals for us to pursue, and – unfortunately for us – often we cannot escape the impact of complexity in that context.

9.3.4 *Addressing the Challenges*

While these three key challenges are quite significant, demography has moved forward in partnership with statistical innovations to develop techniques that can help us cope with these new realities. For example, recent developments in uncertainty quantification for complex models has made clear that models are themselves a source of uncertainty, right alongside the factors mentioned above

(Raftery 1995). Bayesian statistics has presented us with several approaches to dealing with this aspect, such as including a term for model error or code error while building a model (Kennedy and O’Hagan 2001).

The Bayesian perspective has also informed new approaches to mapping the relationships between model parameters and model outputs, even in highly complex computational models. Perhaps most accessible among these has been the Gaussian process emulator, a method for analysing the impact of model parameters on the final output variance (Kennedy and O’Hagan 2001; Oakley and O’Hagan 2002). While this approach has been most commonly used in highly complex computational simulations like global climate models, Gaussian process emulators have also been put to use in demographic projects³ in recent years (Bijak et al. 2013; Silverman et al. 2013a,b).

Thus, as demography has continued to advance to cope with the challenges wrought by complexity, it has moved toward methods and perspectives more commonly associated with complexity science and related disciplines. The prospect of incorporating more exploratory modelling practices within the discipline has led some to seek a movement toward demography as a *model-based science* (Burch 2003; Courgeau et al. 2017), much like in population biology (Godfrey-Smith 2006; Levins 1966).

9.4 Moving Toward a Model-Based Demography

Having established that demography, and the population sciences more broadly, have begun to move toward a model-based paradigm and incorporate insights from disciplines already inclined in that direction, we will revisit some concepts from Part II in order to start to bring together a coherent framework.

In Chap. 5, we outlined a key distinction between two streams of modelling for the social sciences: social simulation and systems sociology. Systems sociology is fundamentally a more explanatory, and exploratory, form of modelling in which we focus on understanding foundational social theories that lead to the development and evolution of society. Demography, generally speaking, clearly leans more toward the social simulation stream, in which the focus is on modelling specific populations and developing powerful links with empirical data. Microsimulations, for example, fall under this category, given their dependence on transition probabilities derived from empirical population data (Zinn et al. 2009; Willekens 2005).

Huneman (2014) suggests that we can further distinguish simulation approaches within that social simulation branch, between *weak* and *strong simulations*. The former aims for a scientific approach, looking to test a hypothesis even when data is hard to come by. The latter lies more in ‘opaque thought experiment’ territory,

³We will examine the application of Gaussian process emulators to demography in significant detail later in Part III.

looking to explore simple models without being dependent on a specific theoretical basis. In the context of modelling for the social sciences there are strong similarities here of course to the systems sociology approach, in that in both cases we seek to step away from strong empirical ties and examine theories at a more foundational level.

However we may characterise these more abstract models – as systems sociology models or strong simulations – demography has at its core a strong commitment to population data and empirically-relevant results. In that context, demographers seek empirically-derived foundations for modelling social mechanisms within a model; the exploratory, generative approach does not provide the kind of insight demographers find most valuable. In order to develop these foundations in a model-based demographic approach, we can bear in mind the suggestion of Conte et al. that:

... simulations must be accompanied by micro-macro-loop theories, i.e., theories of mechanisms at the individual level that affect the global behavior, and theories of loop-closing downward effects or second-order emergence. (Conte et al. 2012, p. 342)

Thus, a critical component of this modelling enterprise must be developing an understanding of this micro-macro link, and in the context of a simulation approach that suggests we must remain committed to a multilevel approach. This provides certain advantages as well, in that powerful tools already exist for multilevel modelling within demography; in a sense, then, we are simply updating the way in which these levels are being represented by putting them into simulation.

9.4.1 The Explanatory Capacity of Simulation

As we have discussed at length previously, agent-based models are uniquely positioned to provide greater explanatory power when applied to complex adaptive systems. This is just as attractive within a demographic context as it is for other social sciences (Burch 2003; Silverman et al. 2011). By allowing the modeller to represent the interactions between individuals and macro-level processes, agent-based models can grant us greater insight into how these different levels of activity influence one another (Chattoe 2003). However, taking advantage of this aspect requires that we develop a more sophisticated understanding of these interactions themselves; in the empirically-focused demographic context, simply creating behavioural rules for these interactions out of best-guess intuitions is not sufficiently rigorous.

In order to delve more deeply into interactions between these levels of analysis, we may situate these interactions themselves as objects of scientific enquiry. By explicitly modelling these interactions in simulation, we can better represent the role of multiple, interacting systems in the final demographic outcomes we see in our empirical observations. This would shift demography more toward a model-based framework, and in so doing allow demographers to contribute more to theoretical

advancements in the study of population change. To an extent this shift has already begun, as the incidence of demographic agent-based models influenced by theories of social complexity has increased since the turn of the century (Kniveton et al. 2011; Bijak et al. 2013; Silverman et al. 2013a; Willekens 2012; Geard et al. 2013).

9.4.2 *The Difficulties of Demographic Simulation*

While the prospect of a model-based demography offers many advantages, no approach comes free of drawbacks. As discussed in Part II, demography – and social science more generally – presents a difficult target for simulation modellers given the need for robust social theories to underpin their simulation efforts. Social theories are not difficult to find, but they are difficult to validate (Moss and Edmonds 2005). While demography differs from other social sciences in its applied focus and the rich population data from which it draws its insights (Xie 2000; Hirschman 2008), demographers interested in simulation must still rely on a solid theoretical backdrop in order to justify the conclusions drawn from their models.

For demography to move forward as a model-based discipline, particularly with agent-based models, the discipline's practical focus must be maintained. This means that simulations demographers build must be underpinned by population data, and, crucially, they must be constructed *inductively*. To do otherwise would be to construct social simulations that, while perhaps enlightening in terms of testing social theories, would have little to say about the core questions that have motivated demographers for these last 350 years.

As Courgeau et al. suggest (2017), these tensions between the expansive explanatory power of simulation and the focused empirical character of demography are not necessarily unresolvable. Following the example set by the historically cumulative progression of demographic knowledge outlined earlier, a model-based demography can build upon the power of the multilevel paradigm, incorporating the capabilities afforded by simulation approaches. In this way we establish a true model-based demography which retains the core empirical character of the discipline, while using simulation to enhance the explanatory power of demographic research.

9.5 Demography and the Classical Scientific Programme

Returning once again to the pioneering, foundational work of Bacon, Graunt and others, we can revisit the classical scientific programme of research and illustrate how a model-based demography enhances this approach. In the natural sciences this approach is very much still in evidence, but in the social sciences we see it less frequently.

In essence, in the classical scientific approach we use observations of some natural property, and from those observations attempt to determine the axioms or

laws driving the process behind this property (Courgeau et al. 2017; Franck 2002a). The classical programme is thus *functional-mechanistic*, in that we are investigating the process generating the property we are interested in by modelling the functions underlying that process. In the context of demography, for example, we set out to understand the process generating patterns of population change, which at the core are controlled by the three functions of mortality, fertility and migration. Courgeau et al. (2017, p. 42, original emphasis) identify some key examples of this functional-mechanistic approach to social sciences in the past:

The ‘law’ of *supply and demand*, as another example, is the ‘first’ structure of functions which was inferred (*induced*) by Adam Smith from the observation of markets: it rules the process of social exchanges generating the market. Karl Marx inferred the general structure of *functions* ruling the process that generates industrial production from a thorough historical study of the technical and social organisation: this ‘first’ principle consists of *separating labour and capital*. Finally, Durkheim inferred the *integration theory* from a sustained statistical analysis of the differences in suicide rates between several social milieus: the social process which generates suicides, whichever their causes, is ruled by the *integration* of the individual agents. The application of the classical programme led to these prominent *theoretical* results at the height of *social sciences*.

In these examples we see that significant theoretical advances in social sciences have come about thanks to the considered application of the classical scientific programme. Smith, Marx and Durkheim chose a social property to focus on – the market, industrial production, and suicide, respectively – and in each case used thorough observations to infer the functional structure underlying these social properties. Of course the impact of these inductive scientific efforts should not be underestimated; Marx’s *Capital*, for example, remains perhaps the most influential critique of capitalist modes of production ever written, while Adam Smith is memorialised in the names of free-market thinktanks the world over.

Demography, as pointed out above, has adhered to a largely similar programme over its history. The observations of populations over the centuries from Graunt onwards has identified mortality, fertility and migration as the primary functions ruling the process of demographic change. Identifying these core functions has helped in turn to focus demographers on those social factors which contribute to these functions, which in turn helps identify those specific demographic variables which may be of greatest interest for further refinement of our understanding of those three functions.

However, in recent years some have proposed that demography has strayed from its scientific lineage (Courgeau and Franck 2007). The power of demographic methods, and their widespread acceptance amongst policy-makers, has led to a reduced focus on theoretical innovation in the discipline (Silverman et al. 2011). Yet we cannot declare an ‘end of history’ in demography; the surge in interest in complexity and agent-based approaches since the early 2000s makes that clear. Demography should continue to evolve cumulatively to adapt to new challenges, as it has done in the past, and here we suggest that a model-based demography rooted in that classical scientific programme should be the next step in that evolution.

9.6 Stages of Model-Based Demographic Research

Figure 9.1 below illustrates the key stages of a functional-mechanistic approach to demography. We start, as has been the case historically, with observations of the properties of demographic processes (mortality, fertility and migration). Through detailed analysis of this data, we proceed to the second step, and attempt to infer the functional structure underlying those processes. From there, we seek to identify the social factors which contribute to those functions, which then leads us to be able to model that structure in detail.

Model-based demography then uses this process as a basis for the next two stages. Conceptual modelling allows us to develop and construct simulations of the interactions at play in the demographic system of interest. The results produced by these simulations can help us to identify areas in which further data collection would be advantageous, and at that point we can start the cycle again. Thus we see model-based demography as an iterative process, in which each trip through this cycle allows us to further refine both the empirical processes and the simulation design and development.

Courgeau et al. (2017, p. 44) characterise this new research programme as consisting of three main pillars:

1. Adherence to the classical programme of scientific enquiry
2. Enhancement of the ways in which demographic phenomena are measured and interpreted
3. The use of formal models, based on the functional-mechanistic principles, as fully-fledged tools of population enquiries.

Thus the focus is on integrating functional-mechanistic models directly into the practice of demography, as a cumulative enhancement of previous paradigms. These models are not intended to become a replacement for previous methods,

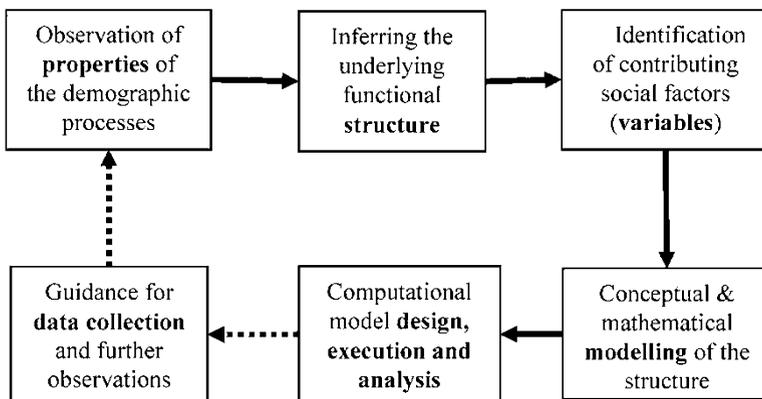


Fig. 9.1 The inductive functional-mechanistic approach in model-based demography (See Courgeau et al. 2017, p. 43)

nor an object of interest in and of themselves; instead, they are part and parcel of the demographic research process, both informing and being informed by the observations that form the empirical heart of the discipline.

Courgeau et al. further suggest that demography cannot rely on other social science disciplines to provide key innovations (Courgeau et al. 2017). Indeed, as we have seen in Part II, the difficulties we encounter in the simulation of social systems are common to the field at large. Demography has a certain empirical advantage over most other social science disciplines, as well, so taking on board theories and methods from less empirically-focused social sciences could instead reduce demographers' ability to benefit from the data-rich nature of population data.

The discipline thus presents an intriguing example of the challenges we must face when developing a model-based approach that is amenable to simulation. While agent-based models can offer substantial power and flexibility to answer appropriately-posed research questions, there is a balance to be struck between embracing that power and ensuring that the core empirical basis and theoretical backdrop of a given discipline are maintained. We need to consider carefully how a model-based approach may move us closer to solving the core epistemological difficulties in demography; transforming demography into a social simulation discipline wholesale might help us shift away from those responsibilities, but it may not help us actually address them.

9.7 Overcoming the Limitations of Demographic Knowledge

As discussed in Sect. 9.3 above, demography faces some key limitations in its ability to explain demographic phenomena. One measure of the success of our proposed model-based demography is whether it could allow us to overcome these limitations, and bolster both the theoretical and explanatory capacity of demography beyond the limits of its current statistically-focused methodological foundation.

The problem of uncertainty in demography has led to the emergence of new statistical methods within the discipline, and a general agreement that demographic predictions become too uncertain to be useful beyond a generation or so (Keyfitz 1981). The use of simulation within a model-based demography could help us to circumvent this limitation by facilitating the use of computational models for *scenario generation*. Simulations can be used to explore the parameter space in which they operate, investigating how different scenarios might affect the behaviour of the simulated population at both the individual and population levels. While these scenarios would not magically present us with enhanced predictive power, they would enable us to present possible ways in which populations may change beyond the one-generation time horizon, given certain assumptions about which parameters are most susceptible to variation.

Model-based demography may also be able to help demographers cope with the aggregation problem. While representing both the micro and macro levels of analysis within a simulation is far from simple, some simulation projects have

allowed for feedbacks between these levels (Billari and Prskawetz 2003; Murphy 2003; Silverman and Bryden 2007). Representing these multiple levels in a single simulation, as well as the interactions between those levels, allows us to avoid the aggregation problem. However, in this context the question of which observations are most useful in such a complex model becomes more critical; we will revisit this issue in more detail in the next section.

Finally, the problem of simplicity can also be addressed by well-considered simulation methods, particularly agent-based modelling. Statistical demographic models easily provoke a tendency toward the inclusion of ever-increasing amounts of data. However, agent-based simulations exhibit behaviour that is more driven by the choice and values of simulation parameters rather than the data which is fed into them. As we have discussed in Part II, in some social simulations data is of little or no importance (as in the case of Schelling's residential segregation model Schelling 1978). Demography by its nature and its empirical focus requires more data input than most areas of social science, but the widespread use of agent-based approaches would necessitate a more careful approach to the integration and use of data. Failure to do so would see us struggling with issues of tractability as models became increasingly unwieldy and difficult to analyse; here we may wish to use the insights drawn in Part II from Lars-Erik Cederman (2002) and Levins (1966) to consider the appropriate balance of tractability versus the levels of generality, precision and realism required for the research question at hand.

9.8 The Pragmatic Benefits of Model-Based Demography

- But you are paying a lot of money for the dragon!
- And what, should we just give it to the citizens instead? [...] I see you know nothing about the principles of economics! Export credit warms up the economy and increases the global turnover.
- But it also increases the dragon as such – I stopped him. – The more intensely you feed him, the bigger he gets; and the bigger he gets, the higher his appetite. What kind of a calculation is it? He will finally devour you all!

Stanisław Lem, *Pożytek ze smoka* [The Use of a Dragon] (1983/2008: 186)

As alluded to above, a common problem faced by statistical demographers is the pressure to bolster the empirical power of demography, or perhaps more properly the *perceived* empirical power of demography, by including ever-larger amounts of population data (Silverman et al. 2011). The rise of multilevel modelling and microsimulation approaches has made the problem even more evident, as the laudable goals of reducing uncertainty and unravelling the micro-macro link leads to an explosive growth in data requirements.

This tendency has effects beyond just creating large and unwieldy models. The process of population data collection itself is both time-consuming and expensive, requiring the design, distribution and collection and increasingly complex surveys. As these surveys grow more complicated, so does the data analysis process, and

designing the subsequent statistical or microsimulation models becomes ever more difficult.

Silverman, Bijak and Noble call this process ‘feeding the beast’ (Silverman et al. 2011), in which demographers get caught in a vicious cycle of sorts, attempting to feed data-hungry models with increasing amounts of data, only to feel pressure to further ‘improve’ these models next time around with yet another injection of observations. While this process is a result of fundamentally positive motivations, evidence suggests that complex models do not necessarily demonstrate better predictive capacity than their simpler fellows, though complex models do require more costly data collection and would tend to have a longer turn-around time between new versions.

Silverman and Bijak cite a couple of examples of this phenomenon:

- Weidlich and Haag (1988) developed an ambitious system dynamics model of migration which attempted to address the micro-macro link; however, the model had very significant data requirements and did not fully address some of the complexities of the migration process itself due to the lack of individual agency in the model.
- The MicMac project (Willekens 2005; Zinn et al. 2009) proposed a new method of dynamic microsimulation which consists of a macro portion (‘Mac’) and a micro-level model (‘Mic’). However, this modelling method is likewise very hungry for data; the ‘Mac’ portion needs detailed data on transition rates, while the ‘Mic’ portion requires a number of variables to be specified at the individual level.

With this in mind, our proposed model-based demography should proceed with awareness of the problem posed by the data-hungry ‘beast’, and offer solutions that protect the empirical focus of demography while helping us to build models that – in the words of John Hajnal – “involve less computation and more cognition than has generally been applied” (Hajnal 1955, p. 321). In Chap. 10, we will begin to present some demographic models which attempt to apply these principles, avoiding ‘the beast’ while maintaining the empirical focus expected in the discipline.

9.9 Benefits of the Classical Scientific Approach

Even if we accept that simulation approaches to demography can provide us significant benefits, both theoretical and more pragmatic, there is a danger that we may exchange some strengths of demography for weaknesses of simulation. We propose that a fully fleshed-out model-based paradigm connected to the classical scientific programme in demography would alleviate at least some of these concerns.

A significant problem in the simulation approach, as outlined in Part II, is the complexity of social systems and thus the inherent difficulty in selecting which components of those systems are to be represented in simulation. Selecting these components generally comes about through the selection of a favoured social theory,

or a core set of assumptions about the functioning of the social processes under examination.

The classical scientific programme helps in our drive to select the most relevant structures and functions which should be replicated in simulation. When undertaking an examination of some social property under the classical programme, we narrow our focus to those processes which generate that property. Our observations focus on that one particular social property, and using those observations we then seek to infer (induce) those functions which generate that property. In this way proceeding via the classical scientific programme helps reduce complexity in our modelling enterprise; our observations focus our enquiries on processes and functions plausibly connected to the property of interest, and this in turn provides clearer guidance on which particular variables must be parameterised and instantiated in simulation. This inductive process thus helps us to avoid the problem of complexity in demographic research.

Another advantage of the classical scientific programme is its ability to generalise social models. The classical scientific focus on functional-mechanistic explanations means that we are examining social systems in an analogous way to natural and biological systems, in an attempt to *reverse engineer* the means by which a social process is generated (Franck 2002a). From this viewpoint, when we see a social process replicated in another population, for example, we can reasonably posit that the same generative process, and thus the same functional structure, should be present. In this way we are developing functional-mechanistic explanations for social processes which can be validated in the real world – due to the inductive process underlying these explanations which relies upon empirical observations in the first place – and which can be generalised, assuming that the iterative process of model refinement and data collection confirms our conclusions about the generative process we identified.

9.10 Conclusions

This chapter has been, in a sense, a whirlwind tour of the discipline of demography, its strengths and weaknesses, and its prospects for the future. As we have seen, demography is a storied discipline, centuries old and tied deeply into local, national and global institutions of politics and policy-making. Understanding and forecasting human population change is of vital relevance to any modern society, after all; without a clear picture of our society and where it is headed, planning for social policies, immigration, health services, tax structures, and so many more aspects of our governance become far more difficult.

That real-world, empirical focus in demography is clearly its greatest strength; the rich nature of population data has allowed demography to develop into a methodical, statistically advanced discipline quite unlike most social sciences. However, these strengths have brought their own challenges, and in particular the three epistemological limitations in demography of uncertainty, aggregation and

complexity have led to significant debate within the field about the constraints of demographic enquiry and how to proceed (Keyfitz 1981; Xie 2000; Silverman et al. 2011).

Here we have outlined a model-based demographic research programme, taking inspiration once again from population biology developing the social simulation stream of research into a form that maintains the empirical richness of demography. The model-based programme builds upon the four existing methodological paradigms in demography, enhancing the power and flexibility of multilevel modelling approaches. The model-based programme is intended to be an integrated part of the demographic research process, allowing models to influence and in turn be influenced by developments in data collection and analysis.

Model-based demography also allows us to extend our predictive horizon in demography, using scenario-based simulation approaches to explore areas of the parameter space beyond the notional one-generation time horizon. As we will see in subsequent chapters, exploring this parameter space in detail using methods like Gaussian process emulators further enables us to understand the behaviour of our simulations, and identify scenarios that may be of particular interest to policy-makers looking to plan for policy spillover effects or unexpected shifts in the population (Silverman et al. 2013a). The incorporation of multiple, interacting levels of social processes in our models can allow us to avoid the ecological and atomistic fallacies (Silverman et al. 2011), and better understand the interactions between social processes that generate key effects at the population level. The ability of agent-based simulations to incorporate individual-level behaviours means that we can also incorporate qualitative data into our models in both the design and implementation phases (Polhill et al. 2010), adding another avenue of empirical relevance to our arsenal.

Next we will analyse some examples of simulation modelling in the demographic context, in order to further develop the model-based programme and identify productive avenues for simulation approaches to population change. In so doing we will discuss aspects of model analysis and uncertainty quantification and how they can help us avoid the problem of complex social models becoming intractable and impenetrable. Ultimately, demography gives us an exciting example of how a fundamentally classical, empirical discipline can use those strengths to its advantage when adopting a methodology most commonly associated with generative, theoretical explanations of social processes. This should serve as a useful model for other disciplines wishing to expand into the simulation arena while maintaining a focus on empirically-driven, policy-relevant research.

Acknowledgements Jakub Bijak and Eric Silverman wish to thank the Engineering and Physical Sciences Research Council (EPSRC) grant EP/H021698/1 “Care Life Cycle” for supporting this research. We also thank Frans Willekens, Anna Klabunde and Jason Hilton for valuable input and discussion.

References

- Aalen, O. O. (1975). Statistical inference for a family of counting processes. PhD thesis. University of California, Berkeley
- Ahlburg, D. A. (1995). Simple versus complex models: Evaluation, accuracy and combining. *Mathematical Population Studies*, 5(3), 281–290.
- Alho, J. M., & Spencer, B. D. (2005). *Statistical demography and forecasting*. Berlin/Heidelberg: Springer.
- Bacon, F. (1863). *Novum organum* (The works, Vol. VIII). Boston: Taggard and Thompson. English translation by Spedding, J., Ellis, R. L., & Heath, D. D. (1863)
- Bayes, T. R. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53, 370–418.
- Bijak, J. (2010). *Forecasting international migration in Europe: A Bayesian view* (Springer series on demographic methods and population analysis, Vol. 24). Dordrecht: Springer.
- Bijak, J., Hilton, J., Silverman, E., & Cao, V. (2013). Reforging the wedding ring: Exploring a semi-artificial model of population for the united kingdom with Gaussian process emulators. *Demographic Research*, 29(27), 729–766.
- Billari, F., & Prskawetz, A. (2003) *Agent-based computational demography: Using simulation to improve our understanding of demographic behaviour*. Heidelberg: Physica Verlag.
- Boudon, R. (1977). *Effet pervers et ordre social*. Paris: Presses Universitaires de France.
- Burch, T. K. (2002). Computer modelling of theory: Explanation for the 21st century. In R. Franck (Ed.), *The explanatory power of models* (pp. 245–265). Amsterdam: Kluwer Academic.
- Burch, T. (2003). Demography in a new key: A theory of population theory. *Demographic Research*, 9(11), 263–284.
- Burch, T. K. (2003). Data, models, theory and reality: The structure of demographic knowledge. In F. Billari & A. Prskawetz (Eds.), *Agent-based computational demography: Using simulation to improve our understanding of demographic behaviour* (pp. 19–40). Heidelberg/New York: Physica Verlag.
- Cederman, L.-E. (2002). Endogenizing geopolitical boundaries with agent-based modeling. *Proceedings of the National Academy of Sciences*, 99(suppl 3), 7296–7303.
- Chattoe, E. (2003). The role of agent-based models in demographic explanation. In F. Billari & A. Prskawetz (Eds.), *Agent-based computational demography: Using simulation to improve our understanding of demographic behaviour* (pp. 41–54). Heidelberg/New York: Physica Verlag.
- Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, V., Loreto, V., Moat, S., Nadal, J.-P., Sanchez, A., Nowak, A., Flache, A., San Miguel, M., & Helbing, D. (2012). Manifesto of computational social science. *European Physical Journal Special Topics*, 214, 325–346.
- Courgeau, D. (2007). *Multilevel synthesis: From the group to the individual*. Dordrecht: Springer.
- Courgeau, D. (2012). *Probability and social science: Methodological relationships between the two approaches* (Methodos series, Vol. 10). Dordrecht: Springer.
- Courgeau, D., & Franck, R. (2007). Demography, a fully formed science or a science in the making. *Population-E*, 62(1), 39–46.
- Courgeau, D., & Lelièvre, E. (1992). *Event history analysis in demography*. Oxford: Clarendon Press.
- Courgeau, D., Bijak, J., Franck, R., & Silverman, E. (2014). Are the four Baconian idols still alive in demography? *Revue Quetelet/Quetelet Journal*, 2(2), 31–59.
- Courgeau, D., Bijak, J., Franck, R., & Silverman, E. (2017). Model-based demography: Towards a research agenda. In A. Grow & J. Van Bavel (Eds.), *Agent-based modelling and population studies* (Springer series on demographic methods and population analysis, pp. 29–51). Berlin/Heidelberg: Springer.
- Franck, R. (2002a). Computer simulation and the reverse engineering method: Conclusions of part II. In R. Franck (Ed.), *The explanatory power of models* (Methodos series, Vol. 1, pp. 141–146). Dordrecht/Boston/London: Kluwer Academic.

- Franck, R. (Ed.). (2002b). *The explanatory power of models: Bridging the gap between empirical and theoretical research in the social sciences* (Methodos series 1). Boston/Dordrecht/London: Kluwer Academic.
- Geard, N., McCaw, J. M., Dorin, A., Korb, K. B., & McVernon, J. (2013). Synthetic population dynamics: A model of household demography. *Journal of Artificial Societies and Social Simulation*, 16(1), article 8.
- Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy*, 21(5), 725–740.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Arnold.
- Graunt, J. (1662). *Natural and political observations mentioned in a following index, and made upon the bills of mortality*. London: Tho. Roycroft.
- Hajnal, J. (1955). The prospects for population forecasts. *Journal of the American Statistical Association*, 50(270), 309–322.
- Henry, L. (1959). D'un problème fondamental de l'analyse démographique. *Population*, 14(1), 9–32.
- Hirschman, Ch. (2008). The future of demography. *Asian Population Studies*, 4(3), 233–234.
- Huneman, P. (2014). Mapping an expanding territory: Computer simulations in evolutionary biology. *History and Philosophy of the Life Sciences*, 36(1), 60–89.
- Huyghens, C. (1657). *De ratiociniis in ludo aleae*. Leyde: Elsevier.
- Kennedy, M., & O'Hagan, T. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B*, 63(3), 425–464.
- Keyfitz, N. (1981). The limits of population forecasting. *Population and Development Review*, 7(4), 579–593.
- Kniveton, D., Smith, Ch., & Wood, S. (2011). Agent-based model simulations of future changes in migration flows for Burkina Faso. *Global Environmental Change*, 21(Suppl. 1), S34–S40.
- Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago/London: The University of Chicago Press.
- Laplace, P. S. (1774). Mémoire sur la probabilité des causes par les événements. *Mémoires de l'Académie Royale des Sciences de Paris* (Tome. VI, pp. 621–656).
- Laplace, P. S. (1812). *Théorie analytique des Probabilités* (2 Vols.). Paris: Courcier Imprimeur.
- Levins, R. (1966). The strategy of model-building in population biology. *American Scientist*, 54, 421–431.
- Mason, W. M., Wong, G. W., & Entwistle, B. (1983). Contextual analysis through the multilevel linear model. In S. Leinhardt (Ed.), *Sociological methodology 1983–1984* (pp. 72–103). San Francisco: Jossey-Bass.
- Morgan, S. P., & Lynch, S. M. (2001). Success and future of demography. The role of data and methods. *Annals of the New York Academy of Sciences*, 954, 35–51.
- Moss, S., & Edmonds, B. (2005). Towards good social science. *Journal of Artificial Societies and Social Simulation*, 8(4), article 13.
- Murphy, M. (2003). Bringing behavior back into micro-simulation: Feedback mechanisms in demographic models. In F. Billari & A. Prskawetz (Eds.), *Agent-based computational demography: Using simulation to improve our understanding of demographic behaviour* (pp. 159–174). Heidelberg/New York: Physica Verlag.
- National Research Council. (2000). *Beyond six billion: Forecasting the world's population*. Washington, DC: National Academies Press.
- Oakley, J., & O'Hagan, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89(4), 769–784.
- Orrell, D. (2007). *The future of everything: The science of prediction*. New York: Thunders Mouth Press.
- Orrell, D., & McSharry, P. (2009). System economics: Overcoming the pitfalls of forecasting models via a multidisciplinary approach. *International Journal of Forecasting*, 25(4):734–743.
- Petty, W. (1690). *Political arithmetick*. London: Robert Clavel and Hen. Mortlock.
- Polhill, J. G., Sutherland, L.-A., & Gotts, N. M. (2010). Using qualitative evidence to enhance an agent-based modelling system for studying land use change. *Journal of Artificial Societies and Social Simulation*, 13(2), article 10.

- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111–163.
- Ryder, N. B. (1951). The cohort approach: Essays in the measurement of temporal variations in demographic behaviour. PhD thesis, Princeton University.
- Schelling, T. C. (1978). *Micromotives and macrobehavior*. New York: W.W. Norton.
- Silverman, E., & Bryden, J. (2007). From artificial societies to new social science theory. In *To appear in the Proceedings of ECAL 2007*.
- Silverman, E., Bijak, J., & Noble, J. (2011). Feeding the beast: Can computational demographic models free us from the tyranny of data? In G. Kampis, I. Karsai, & E. Szathmáry (Eds.), *Advances in artificial life: ECAL 2011* (pp. 747–754). Cambridge, MA: MIT Press.
- Silverman, E., Bijak, J., Hilton, J., Cao, V., & Noble, J. (2013). When demography met social simulation: A tale of two modelling approaches. *Journal of Artificial Societies and Social Simulation*, 16(4), article 9.
- Silverman, E., Hilton, J., Noble, J., & Bijak, J. (2013). Simulating the cost of social care in an ageing society. In W. Rekdalsbakken, R. T. Bye, & H. Zhang (Eds.), *Proceedings of the 27th European Conference on Modelling and Simulation* (pp. 689–695). Dudweiler: Digitaldruck Pirrot.
- Smith, S. K. (1997). Further thoughts on simplicity and complexity in population projection models. *International Journal of Forecasting*, 13(4), 557–565.
- Weidlich, W., & Haag, G. (1988). *Interregional migration: Dynamic theory and comparative analysis*. Berlin/Heidelberg: Springer.
- Willekens, F. (1990). Demographic forecasting; state-of-the-art and research needs. In C. A. Hazeu & G. A. B. Frinking (Eds.), *Emerging issues in demographic research* (pp. 9–66). Amsterdam: Elsevier.
- Willekens, F. (2005). Biographic forecasting: Bridging the micro–macro gap in population forecasting. *New Zealand Population Review*, 31(1), 77–124.
- Willekens, F. (2012). Migration: A perspective from complexity science. *Paper for the Migration Workshop of the Complexity Science for the Real World (CSRW) Network*, Chilworth, 16(2012), Feb 2012.
- Wright, G., & Goodwin, P. (2009). Decision making and planning under low levels of predictability: Enhancing the scenario method. *International Journal of Forecasting*, 25(4), 813–825.
- Xie, Y. (2000). Demography: Past, present and future. *Journal of the American Statistical Association*, 95(450), 670–673.
- Zinn, S., Gampe, J., Himmelspace, J., & Uhrmacher, A. M. (2009). MIC-CORE: A tool for microsimulation. In M. D. Rosetti, R. R. Hill, B. Johansson, A. Dunkin, & R. G. Ingalls (Eds.), *Proceedings of the 2009 Winter Simulation Conference* (pp. 992–1002). IEEE.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

