Population Modeling By Examples III

|  |
| --- |
|  |
| Population Modeling Working Group |
|  |
|  |
| popmodwkgrpimag-news@simtk.org  |

ABSTRACT

This review paper contains examples of population modeling that were collected through self introductions sent to the population modeling mailing list. It is the third review this group has composed collaboratively. The paper forms a definition of a complex field spanning many disciplines by examples of research. The purpose is to further map the field to support future collaboration, cross over, and synergy between population modelers.

**Keywords:** Population Modeling, Definition, Multi Disciplinary, Classification.

# INTRODUCTION

Modelers are capable of modeling many phenomena to great accuracy. Examples of such models can be found in many technical engineering fields where models are very predictive. However, when moving towards more complex systems such as biological systems or behavior, we still have not perfected tools to model phenomena to great accuracy. Specifically, our capabilities to model populations are still limited.

Tools to better model populations of many types have been suggested through the years. In the distant past differential equations were used. However, with recent advances in computing, other computational techniques such as microsimulation and agent based modeling have been suggested. Techniques continue to improve and can be applicable to many types of population modeling.

To help bring together population modelers from multiple disciplines the Interagency Modeling and Analysis Group (IMAG) (IMAG Online) created the population modeling working group that is composed of researchers worldwide. This group has a project site in SmiTK (SimTk Online) and a mailing list (PopModWkGrpIMAG-news Online). In the last few years the group recruited many modelers to join the mailing list and introduce their work. The initial thought was to define a complex field that spans disciplines such as healthcare, transportation, emergency response, and many others. Towards that end the working group stated assembling review papers that edited self introductions of the authors and formed a map of the field by examples.

The first review paper (Population Modeling Working Group 2015) was focused on establishing a definition of population modeling that was defined as “Modeling a collection of entities with different levels of heterogeneity”. The variety of modeling types and techniques was noticeable. In the second review paper (Smith? et. al. 2016) the group extended the work to bring more examples and to start forming a map of the field. This third paper extends this map further by providing additional examples, it will continue by listing members that introduced their work publicly to the mailing list. The introduction text segments were edited for brevity and format. The order of introduction groups contributors by common categories as shown in Table 1.

Table 1: two-dimensional view of the organizational structure of this paper

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Epidemiology and public health** | **Managing disease spread** | **Resource planning & allocation, economics** | **Predicting drug effects** | **Risk assessment** | **Ecosystem management** | **Testing theory** | **Behavior modeling** | **Tools** | **Summary of methods** |
| Nathan Geffen | √ | √ |  |  |  |  |  |  |  | Agent based modeling. matching algorithms, equation based models, microsimulation |
| Christopher Fonnesbeck | √ | √ |  |  |  |  |  |  | √ | MCMC, Baysian models, meta analysis, reinforcement learning |
| Dan Yamin | √ | √ | √ | √ |  |  |  |  |  | Cost effectiveness, Markov chains, differential equations, game theory |
| Katherine Ogurtsova | √ |  |  | √ |  |  |  |  |  | Cost effectiveness analysis |
| Jeff Shrager | √ |  |  | √ |  |  |  | √ | √ | Machine learning, Bayesian methods. |
| Feilim Mac Gabhann | √ |  |  | √ |  |  | √ |  |  | Differential equations, optimization, population generation |
| Carl Asche | √ |  | √ |  | √ |  |  |  |  | Cost effectiveness analysis |
| Michael Thomas | √ |  |  |  |  |  |  | √ |  | Machine learning, Genetic Algorithms |
| Marco Ajelli | √ |  |  |  |  |  |  |  |  | Agent based models, synthetic populations |
| Amit Huppert |  | √ |  |  |  | √ |  | √ |  | Predator prey models, Differential equations |
| Ram Pendyala  |  |  | √ |  |  |  |  | √ | √ | Population generation, microsimulation |
| Bishal Paudel  |  |  |  | √ |  |  |  |  |  | Differential equations, MCMC  |
| Resit Akcakaya |  |  |  |  | √ | √ |  |  |  | Coupled niche-demographic models, matrix population models, metapopulation models with dynamic spatial structure |
| Pawel Topa |  |  |  |  |  | √ | √ |  |  | Agent Based Modeling, Evolutionary Computations |
| Vivek Balaraman  |  |  |  |  |  |  |  | √ |  | Agent Based Modeling, surveys, serious games |
| Matthias Templ  |  |  |  |  |  |  |  |  | √ | Population generation, iterative proportional fitting |
| Leandro Watanabe |  |  |  |  |  |  |  |  | √ | SBML arrays, stochastic simulation |

# Examples

## Nathan Geffen, University of Cape Town, South Africa

This research is dealing with different outcomes modeling of various Sexually Transmitted Infections (STIs) in South Africa using equation-based (Frequency-dependent) models versus microsimulation (network models) (Johnson & Geffen 2016).

In simulations of sexually transmitted infections if there is information (or assumptions) about which people are actually more likely to partner with, it may result in more accurate or realistic models and better insights into how infections spread.

The problem can be stated like this: given a set of agents, seeking a (sexual) partnership, find a set of partnerships such that every agent is paired with one and only one other agent. Also for any two agents, a and b, we have a distance function that calculates how realistic it is that a and b can become sexual partners. This distance function creates an ordering such that for any two potential partners, b and c, of a, it is possible to calculate whether b or c is the more likely partner (perhaps with some arbitrary tie-breaking method). For example, the distance function might calculate that a 25 year-old heterosexual male living in Berlin is more likely to partner with a 25-year-old heterosexual female living in Berlin than a 45 year-old gay male living in Munich. However, the latter is more likely to partner with a 40-year-old gay male living in Munich than either of the first two individuals. A matching algorithm was developed that tries to balance speed and quality (Gefen Online). Current studies are exploring how this affects the outcomes of a simulation.

## Christopher Fonnesbeck, Vanderbilt University. USA

This research applies Bayesian computational methods to address epidemiological problems and tools development. The first of these is in the modeling of the effectiveness of interventions in the control of infectious disease outbreaks (Probert et. al. 2016). In particular, the goal is to estimate optimal policies for controlling outbreaks under uncertainty, and how information collected during the outbreak can be used to update the information state to improve future decisions, with an aim towards reducing the severity or duration of the epidemic. Our approach allows for the recreation of past epidemics, and the exploration of the likely effects of alternative intervention strategies, as well as the value of reducing uncertainty in relevant parameters of the disease system.

Another research focus is meta-analytic modeling for evidence-based medicine (Fonnesbeck et. al. 2012). Bayesian hierarchical models are applied for evaluating comparative effectiveness across studies, because it allows information to be shared in a single analysis without completely pooling the information and ignoring heterogeneity. This paradigm allows multiple sources of independent information to be combined in a single analytic framework to provide information regarding a common set of parameters of interest.

An important aspect of this research involves the development of computational tools for Bayesian modeling. The PyMC project (Salvatier et. al. 2016) was created in 2003. It is a Python library for probabilistic programming. PyMC implements modern algorithms for fitting Baysian models, including Markov Chain Monte Carlo (MCMC) and variational inference, which makes it easy for applied statisticians and modelers to implement arbitrary models, fit them, and analyze their outputs without having to hand-code algorithms.

## Dan Yamin, Tel Aviv University, Israel

The Laboratory for Epidemic Modeling and Analysis is focused on healthcare, predicting the spread of infectious diseases, as well as analyzing population-level effectiveness and the cost-effectiveness of potential intervention programs. In a broader context, there is interest in topics that can be modeled in the same manner as the spread of infectious diseases. These topics include cyber security and computer viruses, viral marketing, information spread, and even behaviors with social-contagion aspects such as imitation of facial expression, smoking habits and risk to become obese.

The core research explores the dynamic between individuals in the population by using tools such as Markov chains, statistics, differential equations modeling, game theory and network science, in an interdisciplinary way. In one example, age-dependent transmission model was developed (Yamin et al. 2016) to evaluate the population effectiveness of an RSV vaccination program in the United States. Results showed that vaccinating children younger than five years of age would be the most efficient and effective way to prevent RSV infection in both children and older adults. When using theoretical epidemiological game model to find the optimal incentive for influenza vaccination, the findings suggested that for the benefit of the elderly greater incentives should be administered to the non-elderly rather than targeting the elderly themselves. Using contact network simulations, results showed that individuals infected with influenza in the previous year, were substantially more likely to have higher connectivity in the network. Consequently, previously infected were at an elevated risk to become infected in subsequent years, and thus should be prioritized for influenza vaccination. Results obtained by contact network simulations were also validated by analysis of medical records. In light of these conclusions, the Israeli Ministry of Health updated the policy from 2016, and the largest HMO in Israel applied various intervention programs to promote vaccination among the targeted population. Another example is the development of a population model for Ebola disease based on methods from scheduling processes (Yamin et. al. 2014). Results from this research demonstrated that Ebola can be eliminated if the World Health Organization allocates resources by focusing on the isolation of infected individuals in critical condition within 4 days from symptom onset rather than isolating any case based on hospital capacity.

## Katherine Ogurtsova, DDZ, German Diabetes Centre, Düsseldorf, Germany

The center is building a diabetes model which will be suited to German settings, reflects German healthcare system structure, and is established primarily on German data where it is possible. The plan is to use this model for cost-effectiveness analysis of population-based prevention programs. Currently the focus is on systematic literature review with the topic: External validation of Type 2 diabetes models: definitions, approaches, implications and room for improvement. The objective is to identify and critically appraise approaches that are used for external validation of existing models of development and of progression of type 2 diabetes. The scope of the review includes: models of type 2 diabetes incidence and/or progression, with or without complications, which are built on simulation techniques on organ/systems, individual or cohort level.

## Jeff Shrager, Stanford, USA

Relevant past and present population modeling research is mostly in the healthcare domain, with some tools development. Past work included multi-agent search problem focused on the question of pre-vs.-post publication peer review (Shrager 2010). Other past work created what was among the earliest, and almost certainly the most advanced, through-the-web programmable biocomputing engines (now called software as a service). Current work is focused on creating (and simulating) a new kind of clinical trial, called Global Cumulative Treatment Analysis (GCTA) (Shrager 2013), which amounts to operating an air traffic control system over the whole biomedical system. The GCTA method uses Bayesian methods to adaptively (and so, theoretically, very efficiently) search the huge combinatoric space, while treating each patient with the best validated knowledge to that moment.

## Feilim Mac Gabhann, Johns Hopkins University, USA

The Institute for Computational Medicine focuses on epidemiology and public health. Examples of modeling used in this context include: 1) Vascular endothelial growth factor (VEGF) / Sema network in cancer. 2) Personalized HIV time courses for stem cell transplant.

Research on population-level differences in cancer with an impact on drug treatment have largely focused on two components: pharmacokinetics (the disposition of the drug being acted upon by metabolic enzymes, renal clearance, and other processes); and genetics (where mutations may render drugs ineffectual, or may even be required for function). Less well studied is the impact of the expression of the drug's target protein, as well as the expression of proteins that interact with that target. Researchers at the institute were able to identify the relative levels of: the 'accelerator' of blood vessel growth - VEGFR2 signaling; and a 'brake' on blood vessel growth - Plexin signaling. We identified both group-wide differences (primary prostate tumors had both the accelerator and brake on, while prostate metastases had the accelerator and no brake), and individual differences within those subpopulations, which enabled us to identify optimal treatments for each case.

Developing a model of the disease course of HIV has enabled the simulation of complex therapeutic interventions, including a potentially curative bone marrow transplant – the introduction of donor stem cells that have been genetically modified to be HIV-resistant. By using longitudinal data from hundreds of HIV patients, we were able to create a virtual patient population that could each be tested with these different interventions. The result is a 'virtual clinical trial', and an estimate of the likelihood of a cure across the population for a given treatment. In addition, insight was gained into the most potent levers or indicators of treatment success, to better identify who would be ideal recipients of the treatment.

## Carl Asche, Center for Outcomes Research, University of Illinois at Peoria, USA

The Center for Outcomes Research (COR) focuses their efforts on the use of comparative effectiveness research and cost-effectiveness analysis in health care decision making. Specifically the COR is interested in utilizing predictive modeling techniques to help reduce hospital no-shows and readmissions (Asche et. al. 2016).

The COR is also interested in determining the value of diabetes therapies which includes development of new techniques for the measurement of cost-effectiveness evaluations of the impact of these therapies on populations (i.e. mortality and morbidity) and healthcare costs. Towards this end, research was conducted towards mapping models or oral treatment of type 2 diabetes (Asche Hippler & Eurich 2014).

## Michael Thomas, Birkbeck University of London, UK

The interest is to use population modeling methods to understand the causes of variations in trajectories of cognitive development between children, from developmental disorders to giftedness. Research is focused on using multi-disciplinary methods to understand the brain and cognitive bases of cognitive variability, including behavioral, brain imaging, computational, and genetic methods (the developmental neurocognition lab Online).

A recent computational approach is to simulate cognitive development in large populations of children and to include intrinsic (neurocomputational) and extrinsic (environmental) factors whose interaction produces variability in developmental trajectories across the whole population. This approach is used within several contexts, including: studying the causes of delay in developmental trajectories of language (Thomas & Knowland 2014), investigating how socio-economic status effects may influence language development (Thomas, Forrester & Ronald 2013), and investigating how population-wide individual differences can interact within pathological mechanisms in neurogenetic disorders such as autism.

Within a machine learning framework, artificial neural network architectures can be complemented with genetic algorithms, to stipulate a genetic level of description. The framework permits a consideration of the relationship between population variability at multiple levels of description, including genes, brain structure, brain function, and behavior. Examples of this work include simulation of genome-wide association analyses and simulation of developmental changes in the heritability of intelligence with age and their relation to differential rates of change of brain structure. This multi-scale framework was used within a purely machine learning context to evaluate whether it may offer new methods for transfer learning (Stamate, Magoulas & Thomas 2015).

## Marco Ajelli, Northeastern University, USA & Bruno Kessler Foundation, Italy

Work is focused on epidemiology and population generation methods. Social contacts are essential for the spread of human-to-human transmissible infectious diseases. In the last decade a lot of attention has been posed to quantify social mixing patterns – age mixing patterns in particular – and a variety of approaches have been used. These approaches include: diary-based surveys, use of wearable sensors, analysis of time-use data, and the development of synthetic populations of agents (Iozzi et. al. 2010).

Our work falls in the latter category (Fumanelli et. al. 2012), where we estimated age-mixing patterns in 26 European countries through the simulation of a synthetic population of agents mimicking real-world contacts between individuals. To build the synthetic populations, publicly available census and survey data is used as well as the main socio-demographic characteristics of each country (e.g., household size, age distribution by household size, schooling and employment rates by age). The resulting contact matrices describe the average frequency of “adequate” contacts that an individual of age i has with individuals aged j. The inferred contact matrices are validated by comparison with matrices obtained in six European countries using a diary-based study on mixing patterns and against epidemiological influenza data.

This methodology allows a large-scale comparison of mixing patterns in Europe, highlighting general common features as well as country-specific differences. Clear relations were found between epidemiologically relevant quantities (such as reproduction number and epidemic size) and socio-demographic characteristics of the populations (e.g., average age of the population). This work provides a numerical approach for the generation of human mixing patterns, which is straightforward to apply to other countries as it is entirely based on routinely collected socio-demographic statistics. This approach could be instrumental for improving model accuracy, especially in the absence of specific empirical data on human mixing patterns.

## Amit Huppert, Gertner institute, Israel

Two classical population dynamics examples of interest include: the unfolding of epidemics and predator prey interactions. In the case of infectious disease outbreaks the goal is to utilize data in order to first estimate the model parameters and conduct model selection. In the second phase, the selected model can be used to study different control methods with the aim of reducing and/or curtailing the outbreak in an optimal way (Yaari et. al. 2016). When studying ecological interactions, there are many theoretical studies that explore the role of spatial structures on predation and/or competition among different organisms. However, there are only few field studies that have validated, and quantified their predictions. In our work (Dattner et. al.. 2017), we combine experimental results, which observed the temporal dynamic of the predatory bacterium, and its prey, in a structured environment composed of sand under various regimes of water content. The ability of the model to fit the data combined with realistic parameter estimates indicate that bacterial predation in the sand can be described by a relatively simple model, and stress the importance of prey refuge on predation dynamics in heterogeneous environments.

We developed and analyzed a set of novel dynamic models (Miller et. al. 2016) for multi-host transmission. The model analysis reveals a new mechanism for disease amplification. In this mechanism the maximum disease risk is obtained when both host species are present in the community. The model expands on the previous understanding about the relationship between host diversity and disease risk by formulating the exact conditions under which diversity amplification, or dilution, would occur. Such formulation is able to account for the different and contradictory patterns often observed in nature. We have also extended the model to look at personal protection techniques, such as insecticide treated nets, repellents, and medications. These include some of the most important and common ways used today to protect individuals from vector borne infectious diseases.

## Ram Pendyala, Arizona State University, USA

Research is focused on population generation tools and modeling for transportation. PopGen (Synthetic Population Generator PopGen Online) is a synthetic population generator that was developed in 2008 to support the development and implementation of activity-based microsimulation models for travel demand forecasting. The advent of activity-based microsimulation models of travel demand provided the ability to simulate activity-travel patterns of individual travelers in time and space. The application of activity-based travel demand systems requires the generation of synthetic populations with attributes that match those of the general population within small geographies (say, a traffic analysis zone, census tract, block group, or block). In (Ye et. al. 2009), we developed and introduced a heuristic iterative procedure known as the IPU (Iterative Proportional Updating) algorithm to facilitate the generation of a synthetic population that matches census data with respect to both household and person attributes. This algorithm has been implemented in PopGen, which is an open source synthetic population generator. In PopGen, the iterative proportional fitting (IPF) procedure is applied first to both household- and person-level control variables of interest to obtain the number of households and persons in each cell of the respective joint distributions. Appropriate rounding procedures are applied to obtain cell "constraints" that must be matched through the population synthesis process.

In addition to developing the IPU algorithm, we also developed and implemented an entropy-maximization based approach to derive weights for sample records, so that both household and person attributes of interest are matched very closely. A relaxed formulation of this algorithm accommodates for data inconsistencies, which are often encountered when dealing with input data used for transportation demand forecasting. The PopGen software package incorporates both the IPU and entropy-maximization methodologies and the user can choose the method to be applied in any specific PopGen run. We have since enhanced the PopGen package to accommodate control variables at multiple geographic resolutions through an enhancement of the IPU algorithm. Current work focuses on the development of a full-fledged demographic and socio-economic evolution model system so that a base (current) year synthetic population can be aged through various lifecycle processes over time, thus enabling the generation of a future year synthetic population. In addition, we are implementing a cloud-version of PopGen to enable ease of use.

## Bishal Paudel, Vanderbilt University, USA

The Quaranta Lab is modeling cell populations, where systems biology approaches are used to understand how cancer cells respond to therapies. Much of our knowledge still comes from analyzing post-resistant tumors or cells, and little is known about what happens to tumors in the early phase leading up to resistance. Experimentation and mathematical modeling are used to track and quantify drug response outcomes and dynamics at the cell population, clonal and single-cell level. One of the projects studies how melanoma, an aggressive form of skin cancer, responds to various targeted and chemotherapeutic agents. Quaranta lab has developed high-throughput assays (Frick, et. al. 2015) to examine the clonal fitness that exist within cell lines and how that fitness changes over time. The recently published paper (Harris et al. 2016), proposed using drug-induced proliferation (DIP) rate which accurately captures drug sensitivities missed out by traditional end-point assays. To formalize differential drug responses of cancer cell population, a simple model of cell proliferation was constructed to describe the experimental data. Three cellular states were defined: a regressing state, a stable state, and an expanding state with negative, zero, and positive net proliferation rates in drug respectively. The assumption is that cells in each state can die, divide or (reversibly) transition into “adjacent” states, thus changing the proportion of cells in each state over time. The proliferation kinetics of cells in each state and total population is defined by coupled ordinary differential equations that take into account the net-proliferation rates of each state, and difference of transition rates. Monte Carlo Markov Chain (MCMC) sampling was used to measure uncertainties in transition rates, and are in the process of identifying molecular signatures of such states, with a hope to calibrate the model. Future work will involve expanding the model to include multiple distinct states, and predict cancer cells response dynamics before resistance develops.

## Resit Akcakaya, Stony Brook University, USA

In the Akcakaya lab, population models are used with dynamic spatial structure to study the effects of climate and landscape changes on species. The approach links climate change models, ecological niche (species distribution or habitat suitability) models, and demographic models to predict the extinction risk of species under global change. The development of this coupled niche-population modeling approach was based on our previous work on meta-population models with dynamic spatial structure (Akçakaya et. al. 2013), which were applied to simulate the effects of landscape changes resulting from timber harvest and fires. Later applications analyzed the effectiveness of conservation measures for the world's most endangered cat, which is impacted by climate change, and for one of the most threatened North American mammals, which is impacted by plague. These studies demonstrated that spatially structured prey-predator models allow realistic conservation planning that takes into account emerging threats such as climate change and infectious diseases.

One goal has been to contribute to the development of red list guidelines for identifying species threatened by climate change. To test the effectiveness of the red list criteria as an early-warning system for climate change-related extinctions, we adapted the coupled niche-population modeling approach to simulate generic life-history types instead of particular species. The resulting analysis found that climate change causes high, but predictable, extinction risks, and that the red list system would provide several decades of warning time for species that might go extinct because of climate change. In a related study, we combined the coupled niche-demographic models with recent advances in geochronological dating, palaeoclimate reconstructions and molecular techniques for retrospective modeling (Fordham et. al. 2016). Our preliminary results suggest that such long-term retrospective analyses will improve efforts to predict the likely effects of future climate and other environmental change on biodiversity, and target conservation management resources most effectively.

## Pawel Topa, AGH University of Science and Technology, Poland

This research is focused on ecological and evolutionary population modeling. The "eVOLUTUS: project (Topa et. al. 2015) is a simulator of multiscale evolutionary processes tested on Foraminifera. The goal is to design a new algorithmic framework for testing and simulating evolutionary principles and their consequences in defined environments at various spatiotemporal scales. Due to biological complexity of organisms and their interactions with the environment, there has never been an attempt to design such a numerical framework for the simulation of evolutionary patterns derived from the fossil record of real organisms. Foraminifera are single-celled eukaryotes that occupy marine benthic and pelagic zones throughout the world and have an extraordinary fossil record throughout the Phanerozoic. This makes them an ideal model often used for testing general evolutionary hypotheses. Recent studies have shown that shell architecture and most prominent morphogenetic trends in shell development distinctly follow molecular phylogenetic patterns. Individual based modeling is used, in which the modeler tries to present population of living organisms as being composed of discrete individual organisms. The model that can be implemented using Agent-based Modeling computational paradigm (Topa & Faber 2017). In order to model evolutionary processes we use the well known genetic operators: mutation and crossing over. These are applied to the virtual genome consisting of parameters that govern agent (foraminifers) behavior. Selection is simply made by adaptation to environmental conditions in virtual habitat.

## Vivek Balaraman, Human Centric Systems Research Group in TCS Research, India

The research group is conducting work in the area of grounded fine grained behavioral models of human populations, with current focus on modeling work teams and groups within organizations (Singh et. al. 2016). Grounded means that ideally the model elements have to be grounded in observations and studies, and not assumed. Fine grained means that multiple behavioral dimensions are investigated such as personality or affect influence decision making. This is done using 3 techniques: 1) conducting studies on the ground using surveys as well as other data collection techniques, 2) mining past literature for useful empirical results in a context, 3) working towards using serious games to capture behavior. Through these processes a behavior patterns repository is populated with the intention to provide in time foundational elements to build behavioral models of entities in a context (Patel et. al. 2016). The business applications of this work will help better understand existing behavior of organizations, allow what-if analysis of hypothetical situations as well as a sandbox to test out behavior change interventions.

## Matthias Templ, Vienna University of Technology, Austria

This work is focused on population generation and population modeling tools. The R package simPop (SimPop Online) can be used to create synthetic population data needed for microsimulation tasks. It includes methods such as combinatorial optimization, iterative proportional fitting and updating, model-based regression methods to simulate data as well as some tools that are useful for demographers and economists that carry out microsimulations with data including age or income heaping problems. The aim is to simulate a population for all individuals, the data required for time T\_0 where microsimulation methods are applied to carry on to T\_1, T\_2, ... The package is specialized for complex data from official statistics but is not limited to these data sets.

## Leandro Watanabe, The University of Utah, USA

The Myers research group is active in standard development and tool development that can be applied towards population modeling. Standards are important for the ability to share and reuse models. This importance has risen due to the increasing number of computer-aided design tools being developed, and such standards improve interoperability between tools. In the systems/synthetic biology fields, the most widely used modeling language for simulation is the Systems Biology Markup Language (SBML Online), which allows the representation of quantitative models. SBML has many features that allow the representation of chemical reaction networks, discrete event models, and ordinary differential equations models. SBML supports extensions that increases the usability of the standard for different applications. The arrays package has been proposed to allow the representation of regular structures more concisely. The arrays package in SBML has been used for the representation of population models in the systems/synthetic biology domains within the iBioSim tool (iBioSim Online). In particular, the arrays package was used to construct a lattice-based model of cellular populations to identify the behavior of genetic circuits in different cells when placed in an environment that allows cell-cell interactions (Watanabe & Myers 2016). This is just one of the many applications of the arrays package, which could potentially be used for many other applications such as disease models.

# Discussion

This is the third review paper that the working group produces. Therefore it is possible to deduce trends in field development over a few years. In the first review paper, the main discussion was regarding definition of the field and whether cohort modeling or individual based modeling should be both included in the same group. By this third review paper it is clear that many population modelers are now leaning towards individual based models such as microsimulation or agent based models. This is not surprising considering advances in computing in the last several decades.

Some editorial freedoms were taken while writing this paper. For example virtual clinical trials were considered as part of the epidemiology category and the testing theory category. Also, modifications were made to category titles used in Table 1: The word economics was added to the resource planning and allocation category since this is where health economists belong. Yet more importantly two new categories were added: behavior modeling, and tools categories. The last category seems to be very active and many tools are reported in this paper that either span fields such as healthcare and transportation, or are generic and non affiliated to a specific field of use. This is a promising development that may help cross over and fertilization between fields which may lead to further standardization.

As a consequence of the review of this paper, working group members started creating project sites on the SimTK repository to create live versions of this paper to enhance communication of modeling knowledge.

# Acknowledgments

Jacob Barhak edited this paper from contributions provided on the mailing list. Chad Ridgley helped with grammar. We thank all contributors to the discussions on the mailing list. We thank the reviewers Thomas Woolley, Jane Heffernan, and Robert Smith?. Their reviews can be accessed through: <https://groups.google.com/forum/#!topic/public-scientific-reviews/7lr3pCUgZv4>

references

Akçakaya Lab, Online. <http://life.bio.sunysb.edu/ee/akcakayalab/>

Akçakaya, H.R. & Root, W.T. 2013. “RAMAS GIS: Linking Spatial Data with Population Viability Analysis. Version 6.0.” In *Applied Biomathematics*, Setauket, New York.

Asche, C.V. Ren, J. Kim, M. Kirkness, C. Dong, Y. & Hippler, S. 2016. “A prediction model to identify acute myocardial infarction (AMI) patients at risk for 30-day readmission”. In *SummerSim 2016*. <https://doi.org/10.22360/SummerSim.2016.SCSC.005>

Asche C.V., Hippler S.E. & Eurich, D.T. 2014. “ Review of models used in economic analyses of new oral treatments for type 2 diabetes mellitus”. In *Pharmacoeconomics*. 32(1), pp. 15-27. PMID: 24357160 <https://doi.org/10.1007/s40273-013-0117-7>

Bender, R.J. Mac Gabhann, F. 2015. “Dysregulation of the vascular endothelial growth factor and semaphorin ligand-receptor families in prostate cancer metastasis” In *BMC Syst Biol.* 9 pp. 55. <https://doi.org/10.1186/s12918-015-0201-z>

Dattner, I. Miller, E. Petrenko, M. Kadouri, D. E. Jurkevitch, E.. & Huppert, A. 2017. “Modelling and parameter inference of predator–prey dynamics in heterogeneous environments using the direct integral approach”. In *Journal of The Royal Society,* 14(126).

Deutsches Diabetes Zentrum. Online, <http://ddz.uni-duesseldorf.de/en/>

Fonnesbeck, C. J. McPheeters, M. L. Krishnaswami, S. Lindegren, M. L. & Reimschisel, T. 2012. “Estimating the probability of IQ impairment from blood phenylalanine for phenylketonuria patients: a hierarchical meta-analysis”. In *Journal of Inherited Metabolic Disease*, 36(5), pp. 757–766. [http://doi.org/10.1007/s10545–012–9564–0](http://doi.org/10.1007/s10545%E2%80%93012%E2%80%939564%E2%80%930)

Fordham, D.A., H.R. Akçakaya, J. Alroy, F. Saltré, T.M. Wigley, B.W. Brook. 2016. “Predicting and mitigating future biodiversity loss using long-term ecological proxies”. In *Nature Climate Change,* 6, pp. 909-916.

Frick, P.L. Paudel, B.B. Tyson, D.R. Quaranta, V. 2015. “Quantifying heterogeneity and dynamics of clonal fitness in response to perturbation”. In *J Cell Physiol*.;230(7) pp. 1403-12. <http://dx.doi.org/10.1002/jcp.24888>

Fumanelli, L. Ajelli M. Manfredi, P. Vespignani, A. Merler, S. 2012. “Inferring the Structure of Social Contacts from Demographic Data in the Analysis of Infectious Diseases Spread”. In *PLoS Comput Biol* 8(9). <https://doi.org/10.1371/journal.pcbi.1002673>

Geffen N., Online, “Optimising partner matching for microsimulations of the HIV epidemic” Online: [http://nathangeffen.webfactional.com/partnermatching/partnermatching.html accessed 4/15/2017](http://nathangeffen.webfactional.com/partnermatching/partnermatching.html%20accessed%204/15/2017)

Harris, L. A., Frick, P. L., Garbett, S. P., Hardeman, K. N., Paudel, B. B., Lopez, C. F., Quaranta V. & Tyson, D. R. (2016). “An unbiased metric of antiproliferative drug effect in vitro”. *Nature Methods*, 13, pp. 1–6. <http://doi.org/10.1038/nmeth.3852>

Hosseini, I. & Mac Gabhann, F. 2016. “Mechanistic Models Predict Efficacy of CCR5-Deficient Stem Cell Transplants in HIV Patient Populations” In *CPT Pharmacometrics Syst Pharmacol.* 5(2) pp. 82-90. <https://doi.org/10.1002/psp4.12059>

iBioSim, Online <http://www.async.ece.utah.edu/ibiosim>

IMAG: Population Modeling Working Group, Online: <http://www.imagwiki.nibib.nih.gov/content/population-modeling-working-group> accessed 4/15/2017

Iozzi, F. Trusiano, F. Chinazzi, M. Billari, F.C. Zagheni, E. Merler, S. Ajelli, M. Del Fava, E. & Manfredi P. 2010. “Little Italy: an agent-based approach to the estimation of contact patterns- fitting predicted matrices to serological data”. In *PLOS Comput Biol*. 6(12). <https://doi.org/10.1371/journal.pcbi.1001021>

Shrager, J. 2010. “The Promise and Perils of Pre-Publication Review: A Multi-Agent Simulation of Biomedical Discovery Under Varying Levels of Review Stringency”. PLoS One. 5(5): [https://dx.doi.org/10.1371%2Fjournal.pone.0010782](https://dx.doi.org/10.1371/journal.pone.0010782)

Shrager, J. 2013. “Theoretical Issues for Global Cumulative Treatment Analysis (GCTA)”. ArXiv.org. <https://arxiv.org/abs/1308.1066>

Johnson, L.F. Geffen, N. 2016. “A Comparison of Two Mathematical Modeling Frameworks for Evaluating Sexually Transmitted Infection Epidemiology”. In *Sex Transm Dis.* 43(3) pp. 139-46. <http://dx.doi.org/10.1097/OLQ.0000000000000412>

Miller, E. Dushoff, J. Huppert, A. 2016. “The risk of incomplete personal protection in vector borne disease”. In *Journal of The Royal Society Interface,* 13. <http://dx.doi.org/10.1098/rsif.2015.0666>

Patel, S. Mahamuni, M. Singh, M. Clarance, D. Duggirala, M. Sharma, S. Katiyar, K. Deshpande, G. Deshmukh, Vaibhav, & Balaraman, V. 2016. “Mining multi-source data to study workplace activity patterns”, In *Conference on Data Science CoDS*

PopModWkGrpIMAG-news - Population Modeling mailing list, Online: <https://simtk.org/mailman/listinfo/popmodwkgrpimag-news>

Population Modeling Workgroup, 2015. "Population Modeling by Examples (WIP)" SpringSim 2015 , April 12 - 15, Alexandria, VA, USA . <http://dl.acm.org/citation.cfm?id=2887741>

Probert, W. J. M., Shea, K., Fonnesbeck, C. J., Runge, M. C., Carpenter, T. E., Dürr, S., Garner, M. G. Harvey, N. Stevenson, M. A. Webb, C. T. Werkman, M. Tildesley, M. J. & Ferrari M.J. 2016. “Decision-making for foot-and-mouth disease control: Objectives matter”. In *Epidemics*, 15, pp. 10–19. <http://doi.org/10.1016/j.epidem.2015.11.002>

Salvatier, J. Wiecki, T. V. & Fonnesbeck, C. 2016. “Probabilistic programming in Python using PyMC3". In *PeerJ Computer Science*, 2(2), e55. <http://doi.org/10.7717/peerj-cs.55>

SBML.org, Online. <http://sbml.org/Main_Page>

simPop, Online, <https://cran.r-project.org/web/packages/simPop/index.html>

SimTk: Population Modeling Workgroup Project, Online: <https://simtk.org/home/popmodwkgrpimag>

Singh, M. Duggirala, M. Hayatnagarkar, H. Balaraman, V. A. 2016. “Multi-Agent Model of Workgroup Behaviour in an Enterprise using a Compositional Approach”., *ModSym Workshop*

Smith?, R. Lee, B.Y. Moustakas, A. Zeigler, A., Prague, M. Santos, R. Chung, M. Gras, R. Forbes, V. Borg, S. Comans, T. Ma, Y. Punt, N. Jusko, W. Brotz, L. & Hyder, A. 2016. “Population Modeling by Examples II”. In *SummerSim 2016*, Montreal, CA. <https://doi.org/10.22360/SummerSim.2016.SCSC.060>

Stamate, C. Magoulas, G. D. & Thomas, M. S. C. 2015. “Transfer learning approaches for financial applications”. In R. Everson, E. Keedwell & D. Walker (Eds.), In *Proceedings of the UK Workshop on Computational Intelligence*, University of Exeter.

Synthetic Population Generator PopGen, Online, <http://www.mobilityanalytics.org/popgen.html>

The Developmental neurocognition lab Online: <http://www.bbk.ac.uk/psychology/dnl//>

The PopModWkGrpIMAG-news Archives, Online: <https://simtk.org/pipermail/popmodwkgrpimag-news/>

Thomas, M. S. C. & Knowland, V. C. P. 2014. “Modelling mechanisms of persisting and resolving delay in language development”. In *Journal of Speech, Language, and Hearing Research*, 57(2), pp. 467-483. <http://dx.doi.org/10.1044/2013_JSLHR-L-12-0254>

Thomas, M. S. C., Forrester, N. A., & Ronald, A. 2013. “Modeling socioeconomic status effects on language development”. *Developmental Psychology*, 49(12), pp. 2325-43. <http://dx.doi.org/10.1037/a0032301>

Topa, P. Faber, Ł. Tyszka, J. Komosinski, M. 2017. “Modelling ecology and evolution of Foraminifera in the agent-oriented distributed platform”, In *Journal of Computational Science*, 18, pp. 69-84, ISSN 1877-7503, <http://dx.doi.org/10.1016/j.jocs.2016.07.009>

Topa, P. Komosinski, M. Bassara, M. Tyszka, J. 2015. “eVolutus: A Configurable Platform Designed for Ecological and Evolutionary Experiments Tested on Foraminifera”. In *Man–Machine Interactions* 4 pp. 269-278. <http://dx.doi.org/10.1007/978-3-319-23437-3_23>

Watanabe, L. Myers, C.J. 2016. "Efficient Analysis of Systems Biology Markup Language Models of Cellular Populations Using Arrays". In *ACS Synth. Biol.* 5(8), pp. 835–841 <https://doi.org/10.1021/acssynbio.5b00242>

Yaari, R. Katriel, G. Stone, L. Mendelson, E. Mandelboim, M. & Huppert, A. 2016. “Model-based reconstruction of an epidemic using multiple datasets: understanding influenza A/H1N1 pandemic dynamics in Israel”. In *Journal of The Royal Society Interface,* 13. <http://dx.doi.org/10.1098/rsif.2016.0099>

Yamin, D. Jones, F.K. DeVincenzo, J.P. Gertler, S. Kobiler, O. Townsend, J.P. Galvani, A.P. 2016 “Vaccination strategies against respiratory syncytial virus “. In *PNAS 2016,* 113 (46) pp. 13239-13244 . <https://doi.org/10.1073/pnas.1522597113>

Yamin, D. Gertler, S. Ndeffo-Mbah, M.L. Skrip, L.A. Fallah, M. Nyenswah, T.G. Altice, F.L. Galvani, A.P. "Effect of Ebola Progression on Transmission and Control in Liberia". In *Ann Intern Med.* 162 pp.11-17. <https://doi.org/10.7326/M14-2255>

Ye, X. Konduri, K. Pendyala, R.M. Sana, B. & P. Waddell. 2009. “A Methodology to Match Distributions of Both Household and Person Attributes in the Generation of Synthetic Populations”. In *Proceedings of 88th Annual Meeting of the Transportation Research Board*, National Research Council, Washington, D.C.

AUTHOR BIOGRAPHIES

**THE POPULATION MODELING WORKING GROUP** is composed of researchers that gather together in the population modeling mailing list. The purpose of the group is to promote population modeling in multiple fields. The activities of the working group are reported annually at the Inter Agency Modeling and Analysis Group (IMAG) and MultiScale Modeling Consortium (MSM) meeting. IMAG is composed of government officials, yet the working group is international and composed of academics and industry members. The mailing list address is popmodwkgrpimag-news@simtk.org