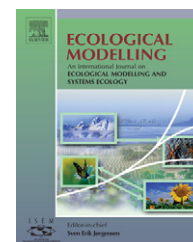


available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/ecolmodel

Review

A bunch of tiny individuals—Individual-based modeling for microbes

Ferdi L. Hellweger*, Vanni Bucci

Civil and Environmental Engineering Department, 400 Snell Engineering Center,
Northeastern University, Boston, MA 02115, USA

ARTICLE INFO

Article history:

Received 28 May 2008

Received in revised form

26 August 2008

Accepted 5 September 2008

Published on line 14 October 2008

Keywords:

Agent-based modeling

Individual-based modeling

Systems ecology

Intra-population variability

Population heterogeneity

Emergence

Continuum

Microorganisms

Microbes

Bacteria

Phytoplankton

ABSTRACT

The individual-based (aka agent-based) approach is now well established in ecological modeling. Traditionally, most applications have been to organisms at higher trophic levels, where the importance of population heterogeneity (intra-population variability), complete life cycles and behavior adapted to internal and external conditions has been recognized for some time. However, advances in molecular biology and biochemistry have brought about an increase in the application of individual-based modeling (IBM) to microbes as well. This literature review summarizes 46 IBM papers for bacteria in wastewater treatment plants, phytoplankton in ocean and inland waters, bacteria in biofilms, bacteria in food and other environs, and “digital organisms” and “domesticated computer viruses” *in silico*. The use of IBM in these applications was motivated by population heterogeneity (45%), emergence (24%), absence of a continuum (5%), and other unknown reasons (26%). In general, the challenges and concepts of IBM modeling for microbes and higher trophic levels are similar. However, there are differences in the microbe population dynamics and their environment that create somewhat different challenges, which have led to somewhat different modeling concepts. Several topics are discussed, including producing, maintaining and changing population heterogeneity (different life histories, internal variability, positive feedback, inter-generation memory), dealing with very large numbers of individuals (different up-scaling methods, including representative space vs. super-individual, number vs. biomass based, discrete vs. continuous kinetics, various agent accounting methods), handling space, simulating interactions with the extracellular environment (hybrid Eulerian–Lagrangian approach), modeling agent–agent interaction (self-shading, predation, shoving) and passive transport (random walk with spatially variable diffusivity, well-mixed reactors). Overall, the literature indicates that the application of IBM to microbes is developing into a mature field. However, several challenges remain, including simulating various types of agent–agent interactions (formation and function of colonies or filaments, sexual reproduction) and even smaller individuals (viruses, genes). Further increases in intracellular detail and complexity in microbe IBMs may be considered the combination of systems biology and systems ecology, or the new field of systems bioecology.

© 2008 Elsevier B.V. All rights reserved.

* Corresponding author. Tel.: +1 617 373 3992; fax: +1 617 373 4419.

E-mail address: ferdi@coe.neu.edu (F.L. Hellweger).

0304-3800/\$ – see front matter © 2008 Elsevier B.V. All rights reserved.

doi:10.1016/j.ecolmodel.2008.09.004

Contents

1. Introduction.....	9
2. Review of existing models.....	9
2.1. Bacteria in wastewater treatment plants.....	11
2.2. Phytoplankton in ocean and inland waters.....	12
2.3. Bacteria in biofilms.....	13
2.4. Bacteria in food and other environs.....	13
2.5. “Digital organisms” and “domesticated computer viruses” <i>in silico</i>	13
3. Why IBM for microbes?.....	14
3.1. Population heterogeneity.....	14
3.2. Emergence.....	15
3.3. Absence of a continuum.....	16
4. Challenges and concepts for microbe IBMs.....	16
4.1. Producing, maintaining and changing population heterogeneity.....	16
4.2. Dealing with very large numbers of individuals.....	17
4.2.1. Number (N)- vs. biomass (B)-based up-scaling.....	17
4.2.2. Discrete (D) vs. continuous (C) kinetics.....	17
4.2.3. Agent accounting methods.....	17
4.3. Handling space.....	18
4.4. Simulating interactions with the extracellular environment.....	19
4.5. Modeling agent–agent interaction.....	19
4.5.1. Self-shading.....	19
4.5.2. Predation.....	19
4.5.3. Shoving.....	19
4.6. Modeling passive transport.....	20
5. Outlook.....	20
Acknowledgements.....	21
References.....	21

1. Introduction

Individual-based models (IBMs) (aka agent-based models, ABMs) explicitly simulate individuals, and the population-level behavior emerges from their cumulative behavior and interaction. It is a “systems” approach, where the system-level behavior emerges from that of low-level components. In ecological modeling, IBM constitutes an alternative to the traditional population-level (aka state variable) approach, where population parameters are modified directly using model equations. For the wolves and moose of an island, for example, a population-level model (PLM) would take differential equations for the density of wolves and moose and solve them for various times (Lotka-Volterra equations). An IBM would simulate individual wolves and moose, and their behavior (e.g. wolf eats moose), and the population densities would emerge from their cumulative behavior and interaction. In ecological modeling, the individual-based approach has been applied for about 20 years and reviews are presented by [Lomnicki \(1999\)](#), [Grimm \(1999\)](#) and [Grimm and Railsback \(2005\)](#). Most applications of IBM have been to higher trophic levels (e.g. animals). However, advances in molecular biology and biochemistry have spurred an increase in the application of IBM to microbes as well.

This review focuses on the application of IBM to microbes. Microbes are different from organisms at higher trophic levels in more than just size (e.g. very large numbers, fast turnover rate, passive transport). Are the motivations for using IBM for

microbes the same as for higher trophic levels? Do the behaviors of microbes or their environment pose new or different challenges? Have IBM modeling concepts been developed that are new or different? To answer these questions, we start by surveying the literature on microbe IBMs with a focus on the technological modeling aspects (Section 2). Then the literature is synthesized and discussed with an emphasis on the motivations for using IBM for microbes (Section 3) and IBM modeling concepts that are new or different for microbes (Section 4). The paper ends with an outlook section that highlights future challenges and directions for microbe IBMs (Section 5).

2. Review of existing models

This section summarizes IBMs for microbes from the literature. [Grimm \(1999\)](#) defines IBMs as “simulation models that treat individuals as unique and discrete entities which have at least one property in addition to age that changes during the life cycle”. That definition is generally adopted here, but since microbe models often do not include age (rather size or cell cycle stage) or an explicit representation of the life cycle, the definition is relaxed to include at least two independent properties (not counting position). Models of single individuals that are not up-scaled to the population level are not included, and to make it manageable, the scope of the review is restricted to the lowest trophic level, bacteria and phytoplankton.

Table 1 – Individual-based modeling frameworks (codes)^a

State variables			Scales		Process overview		Up-scaling ^b	logN	Platform(s)/ Model	Reference(s)
Organism(s)	Properties		Time Δ / maximum	Space maximum	Behavior					
	Transport ^c	Other ^d			Transport ^e	Other ^f				
Top predators, carnivores, herbivores, phytoplankton, bacteria, viruses	1D	S, N, E, CC, P	3–30 months/34+ years	500 m	ADV, D, SET, SWI	U, E, G, X, D	SI/B/D and C	~4	C++/LE	Woods and Onken (1982), Woods (2005), Woods et al. (2005), Nogueira et al. (2006) ^g
Diatoms, Dinoflagellates, Flagellates	3D	2N	30 months/1 year	200 km	ADV, D, RW	U, G, X	? ^h	1.3–2	? ^h	Dippner (1993,1998)
Phytoplankton	2D	S, N	1 h/20 days	120 km	ADV, D/RW, SET	G	SI/B/C	5.2	?/- ^h	Gabric et al. (1996)
Bacteria	3D	S, P, DNA	0.1–1 months/10 weeks	6 cm	SHO, RT, CT, D, CMR	E, G, D, X, EPS	- ^h	- ^h	Swarm/Gecko/ BacSim	Kreft et al. (1998, 2001), Kreft and Wimpenny (2001), Kreft (2004), Standaert et al. (2004), Dens et al. (2005), Grijspeerdt et al. (2005)
Phytoplankton	1D	PIG	1 days/2 years	37 m	SET, ADD, REM	X, B	- ^h	3	?/- ^h	Cuddington and Leavitt (1999)
Diatoms, Dinoflagellates	3D	S, 2N	15 months/3 years	400 km	ADV, D, SET, SWI	U, E, G, D, X	SI/N/D and C	5.3	C/B#-LE/B#- LAE	Broekhuizen (1999), Broekhuizen et al. (2003)
Bacteria	K	N	0.5 months/3 days	N = 12	CMR	U	SI/B/C	4–5	Berkeley Madonna/- ^h	Gujer (2002)
Bacteria	3D	S, CC, N	? ^h /240 h	N = 10 ⁴	RW	U, E, G, D, X	- ^h	<4.8	? ^h /INDISIM	Ginovart et al. (2002)
Digital organisms	- ^h	E, G	N = 10 ⁶	- ^h	CMR	G, D, X, Z, C	- ^h	3–3.6	? ^h /Avida 1.6&1.99	Lenski et al. (2003), Chow et al. (2004)
Bacteria, 2 functional groups	K	3N	0.001 days/9 days	N = 13	CMR	U, G	SI/B/C/	3–3.3	MATLAB/ DisSimulator 1.0	Schuler (2004, 2005, 2006a,b), Schuler and Jassby (2007)
Bacteria, mult. species	3D, K	S, N, P	3.6 months/600 days	1 mm	SHO, TM, DA	U, E, G, D, X, EPS	RS and SI/B/C and D	0.6–3.2	C++&JAVA/- ^h	Picioreanu et al. (2004, 2005, 2007a,b), Xavier et al. (2005a,b, 2007), Batstone et al. (2006), Xavier and Foster (2007)
E. coli	3D, O	S, O	0.01 s/1000 s	1.5 mm	RT, CT	- ^h	- ^h	2.7–3.1	RePast/ AgentCell	Emonet et al. (2005)

Phytoplankton	3D	S, 2N, CC, O, P	0.001–0.02/3 years	80 km	CMR, RW, ADV, SET, SWI	U, G, D, M, O	SI/N/D and C	2.69–3	XLS/VBA/iAlgae 1.00 FOR/RCA/iAlgae rca1.0 XLS/VBA/iAlgae 1.0. lk XLS/VBA/iAlgae 2.65diel	Hellweger and Kianirad (2007a,b), Hellweger (2007, 2008a,b), Hellweger et al. (2008)
Phytoplankton	1/2D	O	1 days/100 days	N = 656	D, A	D, X	– ^h	2–3	SmallTalk/Visual Works Env./– ^h	El Saadi and Bah (2007)
<i>E. coli</i> , mult. genotypes	2D	O	0.02 s/20 months	900 µm	RT, CT	– ^h	– ^h	3	? ^h /E. pluribus/BCT	Bray et al. (2007)
Digital Organisms	K	S, P, N	N = 300,000	N = 10	CMR	U, E, G, D, T, X, Z	– ^h	2+	Flask	Williams and Lenton (2007a,b, 2008)

^a Characteristics presented generally apply to latest version of model.
^b Up-scaling approaches—general approach: RS, reduced space; SI, super-individual. Basis: B, biomass; N, number. Kinetics: D, discrete; C, continuous. See text and Figs. 6 and 7.
^c Transport properties—1, 2, 3D: 1-, 2-, 3-dimensional position; K, discrete (i.e. reactor no.); O, orientation.
^d Other properties: S, size (including biomass, e.g. g dry, g C, mol C or volume); E, energy (including light history parameter); N, nutrient; CC, cell, growth or life cycle position; G, genetic information; DNA, DNA replication forks; PIG, pigments; O, others: P, parameter (i.e. process coefficient rather than state variable). Only individual properties that vary at the individual level are listed.
^e Transport behavior: CMR, completely mixed reactors; RW, random walk; RT, run and tumble; D, diffusion/dispersion; ADV, advection; SHO, shoving; CT, chemotaxis; TM, twitching motility; SET, settling/sinking; SWI, swimming; DA, detachment and attachment (for biofilms); ADD, addition; REM, removal; A, aggregation.
^f Other behavior: B, bleaching; U, uptake; E, excretion; G, growth (including photosynthesis, metabolism, respiration, maintenance); M, maturation; D, division; X, death; T, transformation of extracellular constituents; Z, mutation; EPS, extracellular polymer substance production; C, computation; O, others. Monod uptake and growth is designated as G.
^g A list of 20 LE publications is provided by Woods (2005).
^h (?) Unknown; (–) not present.

In the following, models are grouped by problem area (e.g. bacteria in wastewater treatment plants, phytoplankton in ocean and inland waters, etc.). The development and application of most models is generally motivated by specific questions (e.g. what limits phytoplankton productivity in the Arabian Sea?). However, the focus of this review is on the motivation for using the IBM approach and any challenges, concepts and findings related to the use of IBM (e.g. the effect of population heterogeneity on the simulation results). Problem- or application-specific findings and conclusions that are not directly relevant to IBM modeling are not discussed. Table 1 lists the models and summarizes pertinent characteristics of the modeling frameworks. The presentation is structured consistent with the ODD (Overview, Design concepts, and Details) standard protocol for describing agent-based models (Grimm et al., 2006). The ODD protocol consists of seven elements, grouped into overview (purpose, state variables and scales, process overview and scheduling), design concepts and details (initialization, input, submodels). The overview\purpose information is presented in the following discussion. Information falling under the overview\state variables and scales and overview\process overview (scheduling is not included) categories is included in Table 1. Design concepts and details, where relevant to this review, are also presented in the discussion.

2.1. Bacteria in wastewater treatment plants

In biological wastewater treatment, bacteria grow on and consume various constituents (e.g. organic carbon, nitrogen, phosphorus) in the water and are subsequently removed as sludge. The bacteria can grow suspended in the water and/or attached to solid media. In continuous flow systems (e.g. conventional activated sludge), the bacteria grow in “mixed liquor” reactor(s) and are settled out in subsequent clarifier(s), and a portion of the settled biomass is recycled back to the biological reactor(s). In batch systems (e.g. sequencing batch reactor, SBR), the various processes are carried out in a single reactor by altering the conditions (i.e. mixing, aeration). Traditional models, designed to simulate bulk organic matter (biochemical oxygen demand, BOD) degradation, were based on relatively simple Monod-type growth equations and yield coefficients, and they were implemented using the PLM approach. However, modern models, like those for simulating enhanced biological phosphorus removal (EBPR), require accounting for internal microbial storage products (MSPs) and growth is simulated as a non-linear function of those MSPs. In continuous flow systems, variable life histories (i.e. hydraulic residence time distribution) of individuals lead to population heterogeneity in MSPs. The combination of population heterogeneity and non-linear model equations is one motivation for using IBM for wastewater treatment.

Gujer (2002) developed a relatively simple IBM for continuous and batch suspended growth systems, based on the Activated Sludge Model (ASM), that accounts for one bacteria functional group and MSP. Schuler (2004, 2005, 2006a,b) and Schuler and Jassby (2007) developed the DisSimulator IBM, a more complex and realistic model for the continuous treatment scheme also based on ASM, which accounts for multiple functional groups and MSPs, as well as variable biomass. Gui-

ger and Schuler both performed side-by-side comparisons of PLM and IBM models with identical parameter values and found the results to be significantly different. Schuler further analyzed the effect of this problem on the design of treatment plants and concluded that the assumption of lumped states could result in the undersizing of plant components (i.e. reactors). Gujer and Schuler raise an important point about the meaning and use of parameter values obtained by calibration of PLMs (see Section 3.1).

In the attached growth processes, bacteria can interact spatially by setting up zones with different environmental conditions. For example, nitrifying bacteria may grow on the aerobic surface and denitrifying bacteria in the anoxic interior of a biofilm. IBMs can be used to simulate the emergence of such structures under different conditions. Biofilm IBMs are reviewed in Section 2.3, which includes several models for wastewater treatment plants.

2.2. Phytoplankton in ocean and inland waters

Phytoplankton models are evolving in a similar manner as those for wastewater treatment plants. Traditionally, bulk growth was simulated based on external nutrient levels, but now models are starting to use non-linear formulations based on internal states (e.g. Droop kinetics). Natural waterbodies are rarely truly well mixed, which means the phytoplankton population at any given location is comprised of individuals with different life histories, and consequently different properties (e.g. stored light energy, nutrient). One specific problem that has been studied extensively in oceanography is the effect of variable light exposure on photosynthesis. For that problem, various Lagrangian particle-tracking models that integrate the light history of individuals have been developed (e.g. Falkowski and Wirick, 1981; Nagai et al., 2003, and references therein). Many of these models account for only one individual property (e.g. stored light energy) and do not include cell growth and/or division, and they therefore do not meet the definition of IBM adopted here. However, several of these models were subsequently expanded to include more properties and mechanisms, so they constitute the starting point of IBM in oceanography.

The use of IBM in oceanography was pioneered by Woods and Onken (1982) and Woods (2005) who developed the Lagrangian Ensemble (LE) model. The present version of the LE model is “the Cadillac” of IBMs, and it includes a complete food chain consisting of a number of functional groups (e.g. phytoplankton, bacteria), each subdivided into species and varieties of species (e.g. mutants). The model framework is operational and workshops can be attended for training. The literature contains numerous applications of the LE model (see Woods, 2005), and reviewing each one is not necessary in this context. Recent applications of the LE model examine the problems of stability and predictability (chaos and the butterfly effect; Woods et al., 2005) and species co-existence (paradox of the plankton; Nogueira et al., 2006). The LE model is the most widely used IBM in oceanography, but there are others. Gabric et al. (1996) developed an IBM to simulate the phytoplankton dynamics in a coastal upwelling zone, which includes phytoplankton and nutrient particles. Broekhuizen (1999) and Broekhuizen et al. (2003) used IBM to simulate

the effect of vertical migration on the fitness of diatoms and dinoflagellates. They also performed a side-by-side comparison between PLM and IBM models and found differences in excess of 30% for standing crop and productivity. El Saadi and Bah (2007) used an IBM to simulate phytoplankton aggregation behavior by excretion of chemical signals and chemotaxis towards them. They calculated a drift velocity that is a function of the location of the surrounding individuals. Like the shoving mechanism in biofilms (see Section 2.3), this is a form of direct agent-agent interaction (even though the process of chemical excretion and chemotaxis is strictly speaking indirect). They used the IBM to evaluate the relative contribution of this process and division (which also leads to aggregation because mother and daughter cells are initially in the same location). Dippner (1993, 1998) developed an IBM for diatoms and flagellates. An interesting feature of that model is that it considers whole parcels of water as agents, meaning the two phytoplankton species and extracellular constituents are all moved together. Their model also produced “intrinsic bifurcation” that resulted in two stable states, a process that can change population heterogeneity.

IBMs have also found application in inland waters. Cuddington and Leavitt (1999) developed an IBM to study the effect of various processes (production, bleaching, grazing, etc.) on vertical photosynthetic pigment flux in lakes. In that model, phytoplankton cells were introduced into the lake at a rate dependent on the amount of individuals in the simulation and followed until deposition to the sediment, so the life cycle is not explicitly considered. Hellweger and Kianirad (2007a,b), Hellweger (2007, 2008a,b) and Hellweger et al. (2008) developed the iAlgae IBM framework and used it to analyze the effect of population heterogeneity on simulation results for an idealized river and Lake Biwa, in a manner similar to the studies of Gujer and Schuler for wastewater. Applications of parallel IBM and PLM models demonstrated that accounting for heterogeneity can be significant. The Lake Biwa application (Hellweger, 2007) also included transformation of arsenic by phytoplankton, which is even more non-linear than the Droop-type dependence of growth on internal nutrient, and consequently resulted in an even larger discrepancy between the PLM and IBM approaches. That model also includes a positive feedback mechanism related to luxury uptake. This causes a cell with higher nutrient quota to increase its uptake, leading to even higher nutrient quota and so on. This positive feedback cannot produce population heterogeneity (like different life histories or internal variability), but it can maintain or change it. Mechanisms for creating and maintaining heterogeneity are discussed in Section 4.1. Hellweger et al. (2008) developed an IBM for the cyanobacterium *Anabaena*, which has a complex life cycle that includes resting stage cells. Depending on the environmental conditions, normal (vegetative) cells can differentiate to resting stage (akinetes) cells, and germinate back.

IBM has also been used to study the fundamental problem of division rhythms in phytoplankton. Phytoplankton exhibit diel rhythms (variation with a daily period) in just about everything they do (e.g. nutrient uptake, biomass growth, cell division), which is controlled by sunlight and an internal circadian clock. A population growing with a generation time of 24 h is synchronized and thus homogeneous (e.g. all individuals

divide at 7 p.m. today). However, a population growing with a generation time greater than 24 h, which is more common, is phased (rather than synchronized) and heterogeneous (e.g. only some individuals divide at 7 p.m. today). This population heterogeneity motivates the use of IBM in this context. While studying this problem, Chisholm et al. (1980) recognized that “considering the distributional properties of cells comprising phytoplankton populations rather than simply considering the average behavior, could be a key to understanding many unexplained phenomena in phytoplankton ecology”. Hellweger (2008b) developed an IBM for diel division patterns, which includes a deterministic inter-generation memory or biomass inheritance mechanism (i.e. small mother cell → small daughter cell), and demonstrated the role of this mechanism in diel division patterns. As with the positive feedback mechanism discussed above, inter-generation memory cannot produce heterogeneity, but it can maintain it.

2.3. Bacteria in biofilms

Biofilms are aggregations of microbes attached to surfaces and held together by the excretion of a protective or adhesive matrix. They form, for example, on the surface of trickling filters used for wastewater treatment and on particles in the sediment bed of a river. Traditionally, PLMs were used to model biofilms. The use of IBMs for bacteria was pioneered by Kreft et al. (1998, 2001), Kreft and Wimpenny (2001) and Kreft (2004), who developed the BacSim framework, which was motivated primarily as a more realistic approach that quantitatively incorporates the physiology of individual cells. Consequently, those studies have focused on the emergence of biofilm structure and how it relates to individual behavior. A side-by-side comparison of PLM and IBM models is presented by Kreft et al. (2001). They found that the results were similar in principle, because the models described the same processes. However, significant differences occur in the biofilm shape and growth of minority species, and they concluded that the IBM approach is advantageous when seeking higher resolution and detail. Similar conclusions were reached in a comparison study by Picioreanu et al. (2004).

The biofilm environment poses some interesting challenges for IBMs. First, space is a limiting resource and bacteria shove each other, as discussed by Kreft et al. (2001) (see Section 4.5). Another interesting feature is that individual cells can be held together by a capsule (protective layer) formed by an extracellular polymeric substance (EPS) produced and excreted by the cells. This feature was first modeled by Kreft and Wimpenny (2001). Xavier and Foster (2007) studied cooperation and conflict of bacteria strains that differ in level of EPS production. They concluded that EPS secretion gives bacteria a competitive advantage, because it pushes their progenies up towards higher oxygen concentrations while suffocating non-EPS producing bacteria. Finally, bacteria can move around the biofilm surface using twitching motility, detach from the biofilm and re-attach to it, which was simulated by Xavier et al. (2005a,b) and Picioreanu et al. (2007a).

Several IBMs for biofilms in wastewater treatment plants have been developed, including those for growth on granules in anaerobic (Picioreanu et al., 2005; Batstone et al., 2006) and

aerobic (Xavier et al., 2007) sludge digestion. Picioreanu et al. (2007b) developed an IBM for biofilms in microbial fuel cells.

2.4. Bacteria in food and other environs

Many problems are concerned with the dynamics of bacteria populations in food and other environments, which has traditionally been modeled using the PLM approach. This approach is adequate for simulating growth as a function of external nutrient concentration (i.e. Monod model), but the past decades have resulted in tremendous advances in molecular biology and biochemistry. IBMs constitute a tool for linking these intracellular mechanics and processes to population-level behavior (emergence). Ginovart et al. (2002) presented an IBM for bacteria and a number of case studies investigating biomass distributions, nutrient and temperature effects on colony growth and metabolic oscillations. The model predicted population-level properties, which emerged from the behavior of individuals, were in good agreement with experimental data. One important practical problem in food microbiology is the prediction of the lag phase of bacterial growth. Transition from the lag to the exponential phase is not uniform across the population, so population heterogeneity plays a role. Further, it is difficult to relate the intracellular mechanisms to population-level behavior using the PLM approach. Standaert et al. (2004) used the BacSim model to simulate the lag phase in a batch culture. The lag phase of the population emerges as a result of a specified transition time from birth to the exponential phase for individuals. Dens et al. (2005) expanded this approach by incorporating more intracellular mechanisms, including multiple DNA replication forks and temperature dependence. Contamination of eggs by *Salmonella* requires penetration of the shell and membrane, migration through the growth-unfavorable albumen (egg white), and passage through another membrane into the yolk, where growth conditions are favorable. For un-cracked eggshells, this is a rare event and infection can occur by a single bacterium. This motivated Grijspeerdt et al. (2005) to use IBM to study this phenomenon. Chemotaxis is the active movement along chemical gradients (e.g. move towards nutrients), which in bacteria is done using a “run and tumble” pattern. Individuals “run” for as long as they experience improving chemical condition, and “tumble” (randomly change direction) when they do not. This behavior can be simulated using a PLM if significant simplifying assumptions are made (e.g. random walk coupled with net advection along chemical gradient), but as the behavior becomes more complex (e.g. memory) this becomes increasingly difficult and motivates the use of IBM. Emonet et al. (2005) and Bray et al. (2007) developed IBMs for *E. coli* that explicitly accounts for many of the intracellular mechanisms responsible for chemotaxis, including signaling pathways and explicit representation of the flagella and motors.

2.5. “Digital organisms” and “domesticated computer viruses” in silico

The final category of this review includes hypothetical organisms in hypothetical environments, in the context of answering basic or general question related to evolution and

ecology. Evolution of species results from discrete mutation and/or horizontal gene transfer (e.g. transduction by viruses) events. As with food contamination, the behavior of a single individual can be important. This is difficult to reconcile with the continuum hypothesis inherent in the PLM approach and motivates the use of IBM. Lenski et al. (2003) used the Avida model – an IBM of “domesticated computer viruses” that have a genome, self-replicate, spontaneously mutate, compete and evolve, which affects their growth rate – to study the evolution of complex features (e.g. the eye). Chow et al. (2004) used this model to study a fundamental problem in evolution, the fact that species richness peaks at intermediate productivity. In a similar manner, Williams and Lenton (2007a,b, 2008) developed an IBM for evolving “digital organisms”. They used the IBM to explore fundamental ecological problems, like the role of “rebel” organisms and if, or under what conditions, species selection occurs at the level of the individual or the ecosystem (due to interactions among species).

3. Why IBM for microbes?

We now address the question of what motivates the use of IBMs for microbes. The choice of basic modeling approach, population-level or individual-based, is an important decision that needs to be addressed at the beginning of a modeling project. There are arguments for and against either approach, and the right choice depends on project-specific aspects, the characteristics of the system to be simulated and the questions to be asked of the model. In general, arguments for the population-level approach are its simplicity, computational efficiency, that it has been used and tested widely, the availability of established and peer-reviewed modeling frameworks, and the ready relation to ecological theory. Arguments for the individual-based approach include the ability to simulate intra-population variability, complete life cycles and behavior adapted to internal and external conditions (Grimm et al., 2006). For the microbe IBMs reviewed above, the main motivations have been the ability to resolve population heterogeneity (intra-population variability), the ability to link mechanisms at the individual level to population level behavior (emergence), and inapplicability of the continuum hypothesis (see Fig. 1). Those arguments are discussed further below.

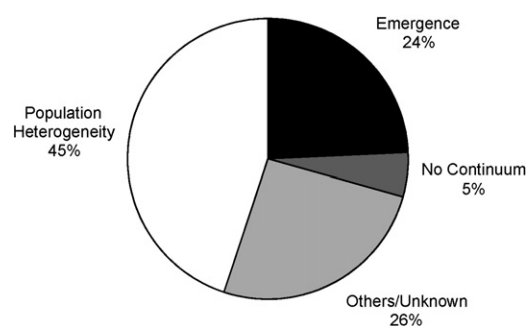


Fig. 1 – Breakdown of motivation for using individual-based modeling (IBM) for microbe models. Statistics based on the papers listed in Table 1.

3.1. Population heterogeneity

The existence and importance of population heterogeneity or intra-population variability is increasingly being recognized. One reason for this is the increased availability of IBOs, individual-based observations. Traditionally, microbes were observed and quantified at the population level. The phytoplankton species of a lake were quantified using Chlorophyll *a*. The bacteria in a wastewater treatment plant were quantified using volatile suspended solids (VSS). However, our observational techniques are changing. A recent review of individual-based observation technology by Brehm-Stecher and Johnson (2004) describes a number of fluorescence, cytometry, scanning probe microscopy and microspectroscopic techniques (see Fig. 2c). Those techniques, especially when used in combination, are very powerful and can be used to size and identify cells and measure respiratory activity, protein expression and behavior, substrate uptake, intracellular pH, nutrient and toxicant content and bacterivory at the individual cell level. One simple way of looking at the situation is that PLMs simply do not produce outputs that can be compared to these new IBOs.

Besides the very practical need to have the model produce outputs that can be compared to observations, there is a more fundamental problem related to population heterogeneity and non-linear models. The most widely used rate equations have a rectangular hyperbolic form (i.e. Monod, Michaelis-Menten, Droop) that give rise to saturating behavior, as illustrated in Fig. 3. This combination, heterogeneity and non-linear models, introduces a significant fundamental problem. Consider a population of phytoplankton cells with *P* cell quotas (*q*) uniformly distributed between 1.0 and 3.0 mmol P mol C⁻¹ (Fig. 3). The growth rate is computed using the Droop's Cell Quota model ($\mu(q) = \mu_{\text{MAX}}(1 - q_0/q)$, $\mu_{\text{MAX}} = 2.0 \text{ day}^{-1}$, $q_0 = 1.0 \text{ mmol P mol C}^{-1}$). The population-average growth rate for this case is 0.89 day^{-1} . However, the average cell quota is $2.0 \text{ mmol P mol C}^{-1}$, and the corresponding growth rate is 1.0 day^{-1} , which is higher than the population average. This is an example of Jensen's Inequality, which states that, for a concave function, the expected value of the function is less than the function of the expected value, or $\text{ave}[\mu(q)] \leq \mu(\text{ave}[q])$. This specific problem has motivated the development of several microbe IBMs.

In many practical applications, the error introduced by Jensen's Inequality may be small compared to uncertainties in coefficients and forcing functions, and it could be lost in the calibration. However, when parameters are adjusted to correct for this problem, it compromises their transferability. Specifically, parameter values obtained from the calibration of PLM models have to be considered “effective parameter values” (Beven and Binley, 1992) that are a function of two things. First, they naturally depend on the “true” value of the parameter, as it applies to an individual microbe. Second, because the PLM model ignores population heterogeneity, the calibrated parameter value incorporates a fudge factor to correct for this error. Consider the above example. To obtain the same population-level growth rate in the PLM as in the IBM (0.89 day^{-1}), the maximum growth rate (μ_{MAX}) would have to be decreased to 1.8 day^{-1} . This correction is a function of the population heterogeneity, which is a function of the

	OBSERVATION		MODELING
POPULATION LEVEL	•Density (10^6 cells mL^{-1}), •Biomass (mg dry L^{-1}), •Optical density (OD), •Chlorophyll <i>a</i> , •Nutrient ratio (e.g. C:N)	(A)	•Chemistry-type models, •Macroscopic simulation, •Advection-dispersion-reaction eq. •Logistic equation, •Gompertz equation, •Lotka-Volterra <i>Examples</i> WASP, AquaTox, RCA, OldLace, QUAL2K, PIM, BLOOM, ASM
	•Fluorescent dyes & stains •Fluorescent semiconductor nanocrystals (quantum dots) •Fluorescence in situ hybridization (FISH) •Immunofluorescence •Green fluorescent protein (GFP) •Flow cytometry •Laser scanning cytometry (LSC) •Image cytometry •Atomic force microscopy (AFM) •Scanning electrochemical microscopy (SECM) •Raman microspectroscopy •Microbeam analysis •Electrorotation	(C)	•Individual-based modeling, •Agent-based modeling, •Microscopic simulation <i>Examples</i> LEM, DisSimulator, BacSim, iAlgae, AgentCell

Fig. 2 – Techniques for observation and modeling at the population and individual scales. Individual observation techniques based on review by Brehm-Stecher and Johnson (2004). See text for specific models.

transport characteristics, which is a function of the specific system (e.g. Deer Island wastewater treatment plant, Lake Biwa) and condition (e.g. high flow, summer). The PLM parameters incorporate site- and condition-specific information, which limits their transferability. This “parameter transferability” problem applies to models of bacteria in engineered wastewater systems (Gujer, 2002; Schuler, 2004, 2005, 2006a,b), phytoplankton in natural systems (Broekhuizen et al., 2003; Hellweger, 2007; Hellweger and Kianirad, 2007b), and other areas.

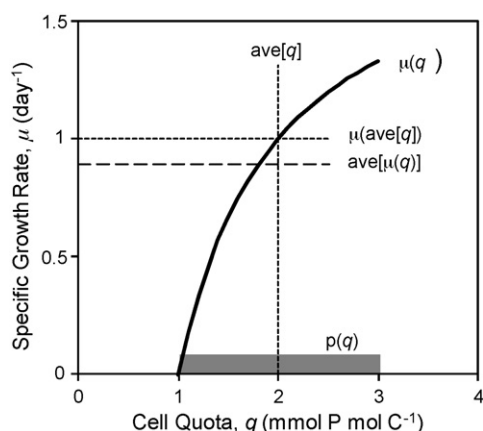


Fig. 3 – Jensen's Inequality. See text.

3.2. Emergence

Early models were concerned primarily with predicting growth, and not much was known or cared about the inner workings of microbes. Consequently, microbes were considered to be functionally equivalent to chemical molecules. A bacteria molecule reacts with dissolved oxygen and organic carbon to form more bacteria and CO_2 . A phytoplankton molecule reacts with light and various inorganic nutrients (CO_2 , NH_4 , PO_4) to form more phytoplankton and dissolved oxygen. These processes can be easily and correctly represented using the PLM approach. However, advances in molecular biology and biochemistry have increased our knowledge of the inner workings of microbes. Contemporary models cannot simply ignore this information, but they have to incorporate it in order to remain useful as research and management tools. When this happens, the degree of non-linearity and associated error introduced by the averaging assumption of the PLM approach increases (see previous section). Also, as the complexity increases, it becomes more difficult to develop PLM equations incorporating the various processes and behaviors of individuals. Can the process of DNA replication, including multiple replication forks and mutations, be incorporated in a population-level biofilm model? Can a population-level differential equation describe the induction of a metal detoxification gene (e.g. *arsA*), the transcription of the corresponding mRNA, and the synthesis

and action of the protein? It may be possible, but it would likely require significant simplifications that would compromise the utility of the model. However, it is straightforward to code these processes into the rules of an individual agent, and then let the population behavior emerge. The use of IBM for many of the microbe modeling studies reviewed above was motivated by the ability to link individual and population behaviors.

3.3. Absence of a continuum

The PLM approach is based on the continuum hypothesis, which approximates a population of discrete individuals using a continuous density. For microbe populations that can reach very high densities, this is typically a good assumption (probably more so than for higher trophic levels). However, because of high growth rates, there are also cases where the population-level response is affected by the fate of a single individual. That one individual can be important may seem incredible given the size of microbial populations. There are, for example, about 10^{14} enteric bacteria in the gut of a single human, but it may only take one for colonization by a different strain (traveler's diarrhea, TD). The marine cyanobacterium *Prochlorococcus* (Partensky et al., 1999) is the most abundant phytoplankton species on Earth with about 10^{29} individuals. Discrete mutation events in a single cell may have shaped the genome of all these cells, and future mutations may again eventually affect the whole population. This means the continuum assumption may not be applicable and can lead to an unacceptable error. It is however straightforward to incorporate this into an IBM, and the use of IBM for some of the microbe modeling studies reviewed above was motivated by this.

4. Challenges and concepts for microbe IBMs

4.1. Producing, maintaining and changing population heterogeneity

Microbe populations have high growth or turnover rates, which means that mechanisms that produce or change heterogeneity produce effects rapidly. Heterogeneity can be produced, maintained and/or changed by a number of different mechanisms, as illustrated in Fig. 4. Because the mechanisms are fundamentally different they can produce

different emerging population structures. Stochastically varying the cell parameters on division by drawing from a smooth probability distribution will likely result in a population structure with similar smooth characteristics. A positive feedback mechanism may result in chaotic behavior (at the individual level) and a population structure that is more polarized. The inter-generation memory mechanism may serve to maintain any structure that is present. To have the right population structure emerge in the model will require including the correct mechanism(s) for producing, maintaining and/or changing population heterogeneity. One problem area, where the specific mechanism responsible for heterogeneity was discussed, is diel division patterns in phytoplankton (see Section 2.2).

One common mechanism for producing population heterogeneity in microbe IBMs is internal variability introduced at birth. Most microbes reproduce by binary fission, meaning one mother cell splits into two daughter cells. Exceptions include the phytoplankton species *Chlamydomonas*, which divides by a multiple fission process and can produce eight daughter cells. There are some cases where the daughter cells are functionally different because of mutation or non-genetic cell differentiation. However, the most common case is division into two functionally equivalent cells. Nevertheless, small imperfections can lead to unequal or asymmetric cell division. This can be simulated by varying the size of daughter cells, but if the imperfection extends to other cell components (e.g. nutrient uptake proteins) then the resulting properties (i.e. maximum uptake rate) can vary as well. To account for this biological variability, Kreft et al. (1998) presented a method of introducing variability by randomly varying several individual model parameters (maximum uptake rate, vol. at division) upon division by drawing from a normal distribution truncated to $\pm 2\sigma$ and avoiding changes of sign (to avoid unrealistic values, e.g. negative maximum uptake rate). Values chosen are independent of those of the mother cell to avoid artificial evolution effects. Kreft et al. discuss literature on observed desynchronization rates and various methods for desynchronization and they concluded that a CV of 0.1 is reasonable for bacteria. This general method of introducing randomness was adopted by other IBMs (e.g. Hellweger and Kianirad, 2007b). In some cases, the randomness introduced by different life histories related to transport processes (e.g. random walk) may be much larger than any internal biological variability, and then internal variability may not affect the simulation results significantly.

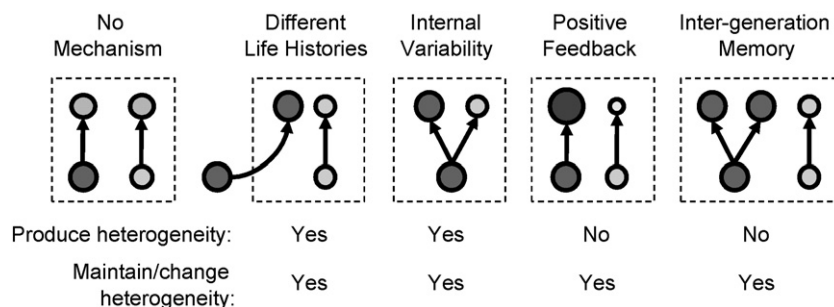


Fig. 4 – Different mechanisms that can cause, maintain and/or change population heterogeneity.

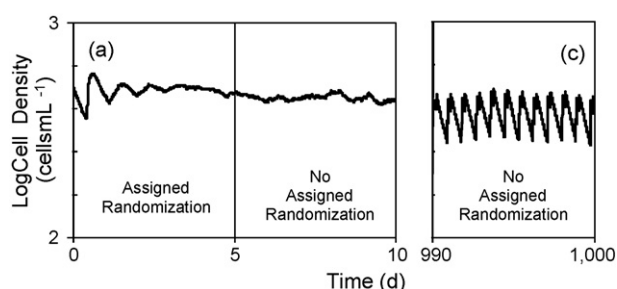


Fig. 5 – Effect of assigned internal variability in a chemostat reactor without any externally variable forcing functions.

Because microbes have a high turnover rate, population dynamics, including synchrony affects occur rapidly. For IBMs that simulate complete lifecycles (i.e. the cell cycle) an artifact termed “numerical synchronization” can occur when no heterogeneity is introduced by the above or some other method (e.g. random walk transport) (Hellweger and Kianirad, 2007b; Hellweger, 2008b). Without any variability, in an open system (chemostat reactor), one lineage will eventually outlast all others and then all individuals will be synchronous to one ancestor. Fig. 5 shows how a population is de-synchronized using assigned randomization in about 5 days, but then re-synchronizes completely in 1000 days in the absence of assigned randomization (from Hellweger, 2008b).

4.2. Dealing with very large numbers of individuals

Dealing with large numbers of individuals is one of the hallmark problems of individual-based modeling. Microbe populations can reach extremely large numbers, are turned over rapidly, and may be effected by the behavior of a single individual. These population dynamics mean that this problem is amplified for microbe IBMs.

There are two basic approaches for reducing the number of individuals explicitly simulated in the model (see Fig. 6). First, the spatial extent of the model can be reduced to a representative space (RS approach). For example, a small representative volume of a lake or an area of a biofilm can be simulated. This approach is most representative for that space, but it is difficult to up-scale to larger spaces when there is heterogeneity at a larger scale. Small simulation spaces are subject to “edge effects”, but that problem can be addressed by wrapping the boundaries (e.g. Kreft et al., 2001). Second, a smaller number of “super-individuals” (SI approach), each representative of a larger number of real individuals can be simulated (Scheffer et al., 1995). Each model super *E. coli* cell may, for example, be representative of 2500 real *E. coli* cells. This allows for simulating large spaces, but complicates direct and indirect agent–agent interaction (discussed further below). The two approaches are not mutually exclusive and can be combined (Picioreanu et al., 2004).

There are a number of variants to the SI approach and they can be categorized as number (N)- vs. biomass (B)-based up-scaling, and discrete (D) vs. continuous (C) kinetics, which are illustrated in Fig. 6 and further described below. The approach used by each model reviewed in Section 2 is included in Table 1 (e.g. the LE model uses SI/B/D and C, or super-individual,

biomass-based, discrete and continuous kinetics). In addition, there are a number of agent accounting methods, which are also reviewed below.

4.2.1. Number (N)- vs. biomass (B)-based up-scaling

Up-scaling from super to real individuals is typically number-based (e.g. each super-individual represents X real individuals). However, microbe modeling has historically been based on biomass, and because some IBMs were developed from existing population-level models (e.g. ASM > DisSimulator), they sometimes use a biomass-based approach (e.g. each super-individual represents Y kg dry weight bacteria). The difference between the two approaches is subtle, but amounts to more than a mere units conversion. When the size of individuals is not explicitly considered the biomass-based approach is more efficient, because it lumps state variables for biomass and up-scaling, whereas the number-based approach considers them separately. However, when the size of individuals is explicitly considered (e.g. the cell cycle is resolved), biomass has to be separated from up-scaling, which is more consistent with the number-based approach.

4.2.2. Discrete (D) vs. continuous (C) kinetics

With the super-individual approach there are two options for simulating kinetic processes. Consider, for example, outwash of phytoplankton from a well-mixed lake. The discrete (D) approach involves calculating the probability of being washed out, randomly selecting the corresponding number of individuals and eliminating them. The alternative approach, referred to as continuous (C), is to reduce the up-scaling factor of all super-individuals in the lake by an appropriate amount (Scheffer et al., 1995). The continuous approach is easier to implement and computationally more efficient (no agents are created/destroyed). Several kinetic processes can be applied using either approach, but there are some processes that need to be simulated discretely. When cell division is associated with mutation or non-genetic cell differentiation (e.g. formation of survivor or persister cells) or when starvation is due to low biomass or energy, the discrete approach needs to be used.

4.2.3. Agent accounting methods

For the super-individual approach, a number of agent accounting methods are available. The simplest one is to maintain a fixed up-scaling factor. However, often the number of individuals changes significantly, as during the course of a phytoplankton bloom. When a fixed up-scaling factor is used, the number of super-individuals fluctuates with the number of real individuals. This can lead to memory problems when the up-scaling factor is small, or large error and even artificial numerical extinction when it is large, as illustrated in Fig. 7 (dashed lines). Further, the method is inherently inefficient because the number of super-individuals required is a function of the population heterogeneity and not the number of real individuals. It is essentially a sampling problem, and if all individuals are identical only one super-individual is needed, regardless of the population size. Because of this problem, most IBMs use an accounting method that fixes the number of super-individuals and therefore lets the up-scaling factor vary (see Fig. 7, solid lines). When the super-individual number is

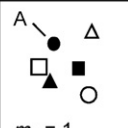
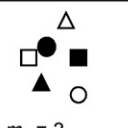
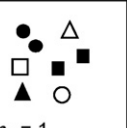
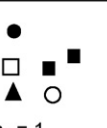








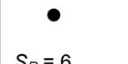
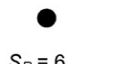
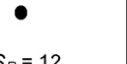
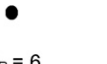
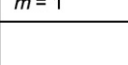
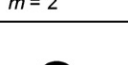
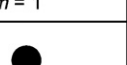
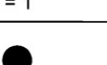




Up-Scaling Approach				Event				
				Start	Growth	Division	Outflow	
Super-Individual (SI)	Rep. Space (RS)	Real	 $m_A = 1$	 $m_A = 2$	 $m_A = 1$	 $m_A = 1$		
			 $m_A = 1$	 $m_A = 2$	 $m_A = 1$	 $m_A = 1$		
	Number(N)	Discr.(D)	 $S_R = 6$ $m = 1$	 $S_R = 6$ $m = 2$	 $S_R = 6$ $m = 1$	 $S_R = 6$ $m = 1$		
		Cont.(C)	 $S_R = 6$ $m = 1$	 $S_R = 6$ $m = 2$	 $S_R = 12$ $m = 1$	 $S_R = 6$ $m = 1$		
	Biomass(B)	Discr.(D)	 $m = 6$	 $m = 12$	 $m = 6$	 $m = 6$		
		Cont.(C)	 $m = 6$	 $m = 12$	 $m = 12$	 $m = 6$		

Fig. 6 – Representative space (RS) and super-individual (SI) up-scaling approaches: number (N)- and biomass (B)-based. Discrete (D) and continuous (C) kinetics. Symbol size indicates individual size.

maintained, the model population has to be adjusted when the agent number grows or shrinks outside user-specified bounds. The most common method is referred to as continuous split/combine (Rose et al., 1993). If the agent number needs to be increased, the largest (based on up-scaling factor) agent is split, and if it needs to be reduced, the two smallest agents are combined. Another method is based on periodic (based on a threshold) random re-sampling of the population (e.g. Hellweger and Kianirad, 2007a).

The above discussion is concerned with temporal changes in individual numbers, but the same argument can be applied to the spatial dimension. This is important, because microbe populations often exhibit significant spatial heterogeneity (e.g. phytoplankton “patchiness”). To address this problem, methods that maintain a fixed number of super-individuals in space, referred to as “fixed super-individual density” methods, have been developed (Woods, 2005; Hellweger, 2008a). Such methods are useful when (a) there are significant spatial gradients in individual density and (b) behavior at low individual densities is important. Hellweger et al. (2008) further extended this approach by introducing “accounting groups” based on lifecycle stages. That approach, not only maintains the total number of super-individuals in a certain spatial compartment, but also considers life cycle stage. The method, for example,

maintains 100 “normal” and 100 “resting” cells in the water column.

4.3. Handling space

The spatial dimension of the microbe IBMs reviewed above varies from none (i.e. space is not considered) to three. One way of handling space is to use discrete locations, like reactors or grid cells. Wastewater treatment plants, for example, consist of multiple well-mixed compartments. Which reactor an individual resides in is important, but the exact position within the reactor may be irrelevant, so the position is defined by reactor ID. The discrete approach facilitates interaction with the extracellular environment (e.g. nutrient uptake, see Section 4.4) and other agents (e.g. predation, see Section 4.5), because spatial relations are easily identified. More commonly however, space is handled in the more realistic continuous manner where an individual's position is defined by coordinates. This approach also allows for interaction with a gridded environment (e.g. discrete nutrient map), but requires calculation of the grid cell ID based on the agent's coordinates. Agent-agent interaction in continuous space, like shoving, is discussed in Section 4.5. When active transport (e.g. chemotaxis) is considered, the orientation in space is also important.

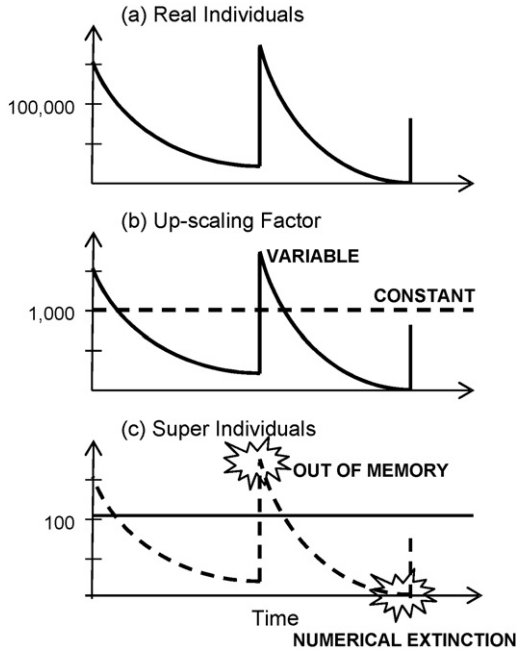


Fig. 7 – Fixed up-scaling factor (dashed line) vs. fixed super-individual number (solids line) agent accounting approaches.

4.4. Simulating interactions with the extracellular environment

Microbes interact with their extracellular environment, and many model applications are concerned specifically with this interaction (i.e. the field of biogeochemistry). The depletion of dissolved oxygen by respiring bacteria and the stimulation of phytoplankton growth by phosphate are two examples. It is possible to use an agent-based approach for extracellular chemicals as well, by using independent chemical agents (e.g. nitrate nutrient particles, Gabric et al., 1996) or by moving the extracellular chemical with the phytoplankton (e.g., Dippner, 1998). This may be advantageous, because it eliminates numerical dispersion effects. However, typically the traditional “concentration” approach is used. The combined “concentration” (Eulerian) approach for extracellular chemicals and agent-based (Lagrangian) approach for microbes is referred to as “hybrid Eulerian–Lagrangian approach”. The transfer of mass between microbe and extracellular pools is based on mass balance, as is illustrated by the following equations for chemical X in a microbe and the extracellular environment (a one-segment lake):

$$\Delta q = (V - W - \mu q)\Delta t \quad (1)$$

$$\frac{\Delta C}{\Delta t} = \left(\sum_{\text{all cells}} [(W - V)mS_R] - kC + QC_{IN} - QC \right) \Delta t + \sum_{\text{died cells}} [qmS_R] \quad (2)$$

where q (mol mol⁻¹) is the cell quota, V (mol mol⁻¹ day⁻¹) is the specific uptake rate, W (mol mol⁻¹ day⁻¹) is the spe-

cific excretion rate, and μ (day⁻¹) is the specific growth rate, t (day) is time, V (m³) is the segment volume, C (mol m⁻³) is the extracellular concentration, m (mol C cell⁻¹) is the cell size, S_R is the super-individual up-scaling factor (see Section 4.2), k (day⁻¹) is decay rate constant, Q (m³ day⁻¹) is the flow rate, and C_{IN} (mol m⁻³) is the extracellular inflow concentration. The uptake and excretion fluxes are typically parameterized based on the internal (q) and/or external (C) concentrations (e.g. Michaelis-Menten). The first summation term accounts for the uptake and excretion by all microbes within the lake. The second summation term accounts for the mass moved from microbes to the extracellular pool by death (e.g. viral lysis). The equations are presented in discrete form (rather than the more common continuous form), because death is a discrete event and equations are commonly integrated numerically. Depending on the application, the model can contain additional chemical species (e.g. nitrate, ammonia), sources and sinks (e.g. wastewater discharge), extracellular and intracellular reactions (e.g. nitrification) and transport mechanisms (3D hydrodynamics, settling), which will further complicate the equations. Nevertheless, the above equations illustrate the concept of the hybrid Eulerian–Lagrangian approach. Picioreanu et al. (2007b) present details for simulating fate and transport of extracellular substances within biofilms and exchange with the bulk water.

4.5. Modeling agent–agent interaction

Microbes interact in a number of different ways. Indirect interactions occur via competition for a nutrient, which can be simulated using the hybrid Eulerian–Lagrangian approach described above. Other ways of interaction are discussed below.

4.5.1. Self-shading

Light is often a limiting resource for phytoplankton and can be affected by the phytoplankton themselves, a process called self-shading. This has been incorporated into phytoplankton IBMs in a simplified manner by computing a light extinction coefficient as a function of phytoplankton (and sometimes other constituents) using empirical formulations (Woods, 2005; Gabric et al., 1996). Computationally this is straightforward and involves computing the phytoplankton density as a statistic of the model population, for use in the light extinction formulation.

4.5.2. Predation

Several phytoplankton IBMs include interaction among multiple trophic levels (i.e. zooplankton grazing) (Woods, 2005). In IBMs, because of computational reasons, this type of interaction is generally simulated using a population-level approach (e.g. Rose et al., 1999), where the grazing rate is computed based on the phytoplankton and zooplankton densities (and sometimes other things). From the rate, a number of phytoplankton cells can be selected at random and ingested by the zooplankton.

4.5.3. Shoving

In biofilms, bacteria compete for space and shove each other. This is a form of direct agent–agent interaction, which can be

computationally very demanding. Processing 1000 bacteria is one thing, but checking all for potential contact with all other 999 bacteria increases the computational burden dramatically. If the interaction is spatially controlled and restricted to the agents in the nearby vicinity, a tree data structure can be used to efficiently locate neighboring agents, as discussed by Kreft et al. (2001).

4.6. Modeling passive transport

Microbes can be motile. Chemotaxis by bacteria using flagella and diel vertical migration by phytoplankton using buoyancy regulation are two examples. These active transport behaviors are straightforward to implement in an IBM. However, microbes are also often transported passively by the ambient hydrodynamics. Advection is simulated by simply applying the velocity to the individual's position, and diffusion can be simulated using a random walk scheme. For a constant diffusivity, the step size is drawn from a normal distribution with standard deviation based on the diffusivity:

$$\Delta z = R\sqrt{2K\Delta t} \quad (3)$$

where z (m) is the position (e.g. depth), R is a random number from a standard normal distribution, K ($\text{m}^2 \text{s}^{-1}$) is the diffusivity and t (s) is time. However, the diffusivity often varies in space. Mixing in the ocean is driven by wind stress at the water surface and the diffusivity generally decreases with depth. In the metalimnion (thermocline) of a lake, density stratification has a stabilizing effect that leads to lower diffusivity. Application of the simple random walk model to this situation leads to an unrealistic net transport towards areas of lower diffusivity. This can be explained theoretically and numerically, or conceptually by the fact that “over a given time interval, particles at a particular location are influenced by slightly more energetic eddies originating in the area of high diffusion compared with somewhat less energetic eddies from the low diffusion region” (Visser, 1997). The problem can be avoided by transforming the coordinate system so that the diffusivity is constant or by using a different formulation for the step size (Hunter et al., 1993; Visser, 1997; Nagai et al., 2003):

$$\Delta z = K'(z)\Delta t + R\sqrt{2K(z + 0.5K'(z)\Delta t)\Delta t} \quad (4)$$

where $K'(z)$ is the derivative of $K(z)$. This formulation is different from the simple random walk (Eq. (3)) in that it includes a corrective advection term and the diffusivity is taken at an offset location. Nagai et al. (2003) found that the correction scheme does not work for zero diffusivity and solved this problem by incorporating background diffusivity. No-pass boundaries (e.g. water surface) are typically simulated as reflecting.

In some cases, individuals are transported by flow between multiple well-mixed compartments. Activated sludge wastewater treatment plants, for example, contain one or more completely stirred tank reactors (CSTRs). In that case, the location is defined by the reactor number (rather than coordinates x , y , z) and transport is done stochastically by defining a probability of being transported in a given time

step (P):

$$P = \frac{Q}{V}\Delta t \quad (5)$$

where Q ($\text{m}^3 \text{s}^{-1}$) is the flow rate, V (m^3) is the reactor volume and Δt (s) is the time step. Schuler (2005) discusses the mass balance error due to limited number of agents and presents a correction method. This type of transport scheme was also used by Hellweger (2007) to simulate transport between two well-mixed lake compartments (epilimnion and hypolimnion).

5. Outlook

Based on the literature reviewed above, IBM is a technology that is already well established for microbe modeling. We can expect that the motivations for using IBM for microbes will become stronger with continued advances in molecular biology and biochemistry, and the hurdles weaker with further advances in computing hardware and software. This means the use of IBM for microbes should continue to increase. Many microbe IBM concepts have been developed that should facilitate the development of new IBMs and ease the transition for newcomers into the field. However, there are many processes that have not been incorporated into IBMs and constitute future challenges. This includes many forms of direct agent-agent interactions, like the formation and function of cell colonies. The cyanobacteria *Anabaena*, for example, forms linear colonies (filaments) with a minority of cells (called heterocysts) that perform nitrogen fixation rather than photosynthesis. They then pass the fixed nitrogen to the photosynthesis performing cells of the filament. Another example is fusion and sexual reproduction. Dinoflagellates, for example, can reproduce by simple fission or sexually by fusion and formation of a fertilized cell (zygote) that can then change into a resting cell (dinocyst). In many cases, these processes can be handled using the same population-level approximations used for predator-prey interaction. However, what if the interaction between two individuals is not random, but consists of chemotactic attraction between two specific cells? What if the purpose of the model is to understand that process? There are also organisms that have yet to be simulated using IBM. The size of organisms simulated cannot decrease much further, since we are already simulating bacteria. However, IBMs for viruses have not yet been documented in the literature, which constitutes uncharted territory. Simulating the complex interaction of viruses and hosts (infection, latent period, burst) will likely pose new challenges to IBM. Also, an emerging topic is the lateral transfer of genes (e.g. by viruses, transduction), which leads into the area of “gene ecology”. That field is not concerned with the fitness of a certain species or the effect a certain gene has on the fitness of the species, but the fitness of individual genes. How did this gene survive eons of evolutionary change and remain to be so popular in many species? The use of IBM for genes may be the next frontier for the technology.

With further advances in molecular biology and biochemistry, we may expect the intracellular detail and complexity

in microbe IBMs to increase. This trend will be supported by models and modeling approaches developed as part of the rapidly growing field of systems biology (Kitano, 2002). Individual-based modeling is a systems ecology simulation approach. Therefore, the development and application of IBMs with increased intracellular detail and complexity constitutes the combination of systems biology and systems ecology. It may be considered the new field of “systems bioecology”. Systems bioecology is the combination of systems biology and systems ecology. Intracellular states and mechanisms of microbes are explicitly resolved and the microbe behavior emerges (systems biology). Individual microbes are simulated using agent-based methods and the population and ecosystem behaviors emerge (systems ecology).

Acknowledgements

We would like to thank V. Grimm and one anonymous reviewer for their constructive criticism.

REFERENCES

- Batstone, D.J., Picioreanu, C., Van Loosdrecht, M.C.M., 2006. Multidimensional modelling to investigate interspecies hydrogen transfer in anaerobic biofilms. *Water Res.* 40, 3099–3181.
- Beven, K., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process.* 6, 279–298.
- Bray, D., Levin, M.D., Lipkow, K., 2007. The chemotactic behavior of computer-based surrogate bacteria. *Curr. Biol.* 17, 12–19.
- Brehm-Stecher, B.F., Johnson, E.A., 2004. Single-cell microbiology: tools, technologies and applications. *Microbiol. Mol. Biol. Rev.* 68, 538–559.
- Broekhuizen, N., 1999. Simulating motile algae using a mixed Eulerian–Lagrangian approach: does motility promote dinoflagellate persistence or co-existence with diatoms? *J. Plankt. Res.* 21, 1191–1216.
- Broekhuizen, N., Oldham, J., Zeldis, J., 2003. Sub-grid-scale differences between individuals influence simulated phytoplankton production and biomass in a shelf-sea system. *Mar. Ecol. Prog. Ser.* 252, 61–76.
- Chisholm, S.W., Morel, F.M.M., Slocum, W.S., 1980. The phasing and distribution of cell division. In: Falkowski, P. (Ed.), *Primary Productivity in the Sea*. Plenum, New York, p. 281.
- Chow, S.S., Wilke, C.O., Ofria, C., Lenski, R.E., Adami, C., 2004. Adaptive radiation from resource competition in digital organisms. *Science* 305, 84–86.
- Cuddington, K., Leavitt, P.R., 1999. An individual-based model of pigment flux in lakes: implications for organic biogeochemistry and paleoecology. *Can. J. Fish. Aquat. Sci.* 56, 1964–1977.
- Dens, E.J., Bernaerts, K., Standaert, A.R., Kreft, J.U., Van Impe, J.F., 2005. Cell division theory and individual-based modeling of microbial lag. Part II. Modeling lag phenomena induced by temperature shifts. *Int. J. Food Microbiol.* 101, 319–332.
- Dippner, J.W., 1993. A Lagrangian model of phytoplankton growth dynamics for the Northern Adriatic Sea. *Cont. Shelf Res.* 13, 331–335.
- Dippner, J.W., 1998. Competition between different groups of phytoplankton for nutrients in the southern North Sea. *J. Marine Syst.* 14, 181–198.
- El Saadi, N., Bah, A., 2007. An individual-based model for studying the aggregation behavior of phytoplankton. *Ecol. Model.* 204, 193–212.
- Emonet, T., Macal, C.M., North, M.J., Wickersham, C.E., Cluzel, P., 2005. AgentCell: a digital single-cell assay for bacterial chemotaxis. *Bioinformatics* 21, 2714–2721.
- Falkowski, P.G., Wirick, C.D., 1981. A simulation model of the effects of vertical mixing on primary productivity. *Mar. Biol.* 65, 69–75.
- Gabric, A.J., Eifler, W., Schrimpf, W., 1996. A Lagrangian model of phytoplankton dynamics in the Northwest African coastal upwelling zone. *Adv. Space Res.* 18, 99–115.
- Ginovart, M., Lopez, D., Valls, J., 2002. INDISIM. An individual-based discrete simulation model to study bacterial cultures. *J. Theor. Biol.* 214, 305–319.
- Grijpspeerdt, K., Kreft, J.U., Messens, W., 2005. Individual-based modeling of growth and migration of *Salmonella enteritidis* in hens’ eggs. *Int. J. Food Microbiol.* 100, 323–333.
- Grimm, V., 1999. Ten years of individual-based modeling in ecology: what have we learned and what could we learn in the future? *Ecol. Model.* 115, 129–148.
- Grimm, V., Railsback, S.F., 2005. *Individual-based Modeling and Ecology*. Princeton University Press, Princeton, NJ.
- Grimm, V., Berger, U., Bastianen, F., Sgrun, E., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S.K., Huse, G., Huth, A., Jepsen, J.U., Jorgensen, C., Mooij, W.M., Müller, B., Pe’er, G., Piou, C., Railsback, S.F., Robbins, A.M., Robbins, M.M., Rossmanith, E., Rüger, N., Souissi, S., Stillman, R.A., Vabø, R., Visser, U., DeAngelis, D.L., 2006. A standard protocol for describing individual-based and agent-based models. *Ecol. Model.* 198, 115–126.
- Gujer, W., 2002. Microscopic versus macroscopic biomass models in activated sludge systems. *Water Sci. Technol.* 45, 1–11.
- Hellweger, F.L., 2007. Is it time to abandon the chemistry approach to biogeochemistry? (agent-based water quality modeling). In: *Proceedings of WEFTEC.07*. Water Environment Federation (WEF), Alexandria, Virginia, pp. 5646–5665.
- Hellweger, F.L., Kianirad, E., 2007a. Individual-based modeling of phytoplankton: evaluating approaches for applying the cell quota model. *J. Theor. Biol.* 249, 554–565.
- Hellweger, F.L., Kianirad, E., 2007b. Accounting for intra-population variability in biogeochemical models using agent-based methods. *Environ. Sci. Technol.* 41, 2855–2860.
- Hellweger, F.L., 2008a. Spatially explicit individual-based modeling using a fixed super-individual density. *Comput. Geosci.* 34, 144–152.
- Hellweger, F.L., 2008b. The role of inter-generation memory in diel phytoplankton division patterns. *Ecol. Model.* 212, 382–396.
- Hellweger, F.L., Kravchuk, E.S., Novotny, V., Gladyshev, M.I., 2008. Agent-based modeling of the complex lifecycle of a cyanobacterium (*Anabaena*) in a shallow reservoir. *Limnol. Oceanogr.* 53, 1227–1241.
- Hunter, J.R., Craig, P.D., Phillips, H.E., 1993. On the use of random walk models with spatially variable diffusivity. *J. Comput. Phys.* 106, 366–376.
- Kitano, H., 2002. Systems biology: a brief overview. *Science* 295, 1662–1664.
- Kreft, J.U., Booth, G., Wimpenny, J.W.T., 1998. BacSim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology* 144, 3275–3287.
- Kreft, J.U., Picioreanu, C., Wimpenny, J.W.T., Van Loosdrecht, M.C.M., 2001. Individual-based modelling of biofilms. *Microbiology* 147, 2897–2912.
- Kreft, J.U., Wimpenny, J.W.T., 2001. Effect of EPS on biofilm structure and function as revealed by an individual-based model of biofilm growth. *Water Sci. Technol.* 43, 135–141.

- Kreft, J.U., 2004. Biofilms promote altruism. *Microbiology* 150, 2751–2760.
- Lenski, R.E., Ofria, C., Pennock, R.T., Adami, C., 2003. The evolutionary origin of complex features. *Nature* 423, 139–144.
- Lomnicki, A., 1999. Individual-based models and the individual-based approach to population ecology. *Ecol. Model.* 115, 191–198.
- Nagai, T., Yamazaki, H., Kamykowski, D., 2003. A Lagrangian photoresponse model coupled with 2nd-order turbulence closure. *Mar. Ecol. Prog. Ser.* 265, 17–30.
- Nogueira, E., Woods, J.D., Harris, C., Field, A.J., Talbot, S., 2006. Phytoplankton co-existence: results from an individual-based simulation model. *Ecol. Model.* 198, 1–22.
- Partensky, F., Hess, W.R., Vault, D., 1999. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* 63, 106–127.
- Picioreanu, C., Kreft, J.U., Van Loosdrecht, M.C.M., 2004. Particle-based multidimensional multispecies biofilm model. *Appl. Environ. Microbiol.* 70, 3024–3040.
- Picioreanu, C., Batstone, D.J., Van Loosdrecht, M.C.M., 2005. Multidimensional modelling of anaerobic granules. *Water Sci. Technol.* 53, 501–507.
- Picioreanu, C., Kreft, J.U., Klausen, M., Haagensen, J.A.J., Tolker-Nielsen, T., Molin, S., 2007a. Microbial motility involvement in biofilm structure formation—a 3D modeling study. *Water Sci. Technol.* 55, 337–343.
- Picioreanu, C., Head, I.M., Katuri, K.P., Van Loosdrecht, M.C.M., Scott, K., 2007b. A computational model for biofilm-based microbial fuel cells. *Water Res.* 41, 2921–2940.
- Rose, K.A., Christensen, S.W., DeAngelis, D.L., 1993. Individual-based modeling of populations with high mortality: a new method for following a fixed number of model individuals. *Ecol. Model.* 68, 273–292.
- Rose, K.A., Rutherford, E.S., McDermot, D.S., Forney, J.L., Mills, E.L., 1999. Individual-based model of yellow perch and walleye populations in Oneida Lake. *Ecol. Monogr.* 69 (2), 127–154.
- Scheffer, M., Baveco, J.M., DeAngelis, D.L., Rose, K.A., Van Nes, E.H., 1995. Super-individuals a simple solution for modelling large populations on an individual basis. *Ecol. Model.* 80, 161–170.
- Schuler, A.J., 2004. Distributed microbial states in nutrient removing activated sludge systems: simulation of an important but neglected phenomenon. In: *Proceedings of the Water Environment Federation Annual Conference (WEFTEC)*, New Orleans, October 2–6, 2004.
- Schuler, A.J., 2005. Diversity matters: dynamic simulation of distributed bacterial states in suspended growth biological wastewater treatment systems. *Biotechnol. Bioeng.* 91, 62–74.
- Schuler, A.J., 2006a. Process hydraulics, distributed bacterial states, and biological phosphorus removal from wastewater. *Biotechnol. Bioeng.* 94, 909–920.
- Schuler, A.J., 2006b. Distributed microbial state effects on competition in enhanced biological phosphorus removal systems. *Water Sci. Technol.* 54, 199–207.
- Schuler, A.J., Jassby, D., 2007. Distributed state simulation of endogenous processes in biological wastewater treatment. *Bioeng. Biotech.* 97, 1087–1097.
- Standaert, A.R., Poschet, F., Geeraerd, A.H., Uylbak, F.V., Kreft, J.U., Van Impe, J.F., 2004. A novel class of predictive microbial growth models: implementation in an individual-based framework. In: Pons, M.N., Van Impe, J.F. (Eds.), *Contributions made at the Ninth International Symposium on Computer Applications in Biotechnology (CAB9: Computer Applications in Biotechnology 2004)*. Nancy, France, March 28–31, 2004. Elsevier, Amsterdam, pp. 183–188.
- Visser, A.W., 1997. Using random walk models to simulate the vertical distribution of particles in a turbulent water column. *Mar. Ecol. Prog. Ser.* 158, 275–281.
- Williams, H.T.P., Lenton, T.M., 2007a. Artificial selection of simulated microbial ecosystems. *PNAS* 104, 8918–8923.
- Williams, H.T.P., Lenton, T.M., 2007b. The Flask model: emergence of nutrient recycling microbial ecosystems and their disruption by environment-altering ‘rebel’ organisms. *Oikos* 116, 1087–1105.
- Williams, H.T.P., Lenton, T.M., 2008. Environmental regulation in a network of simulated microbial ecosystems. *PNAS* 105, 10432–10437.
- Woods, J.D., Onken, R., 1982. Diurnal variation and primary production in the ocean preliminary results of a Lagrangian ensemble model. *J. Plankt. Res.* 4, 735–756.
- Woods, J.D., 2005. The Lagrangian Ensemble metamodel for simulating plankton ecosystems. *Progr. Ocean* 67, 84–159.
- Woods, J., Perilli, A., Barkmann, W., 2005. Stability and predictability of a virtual plankton ecosystem created with an individual-based model. *Progr. Ocean* 67, 43–83.
- Xavier, J.B., Picioreanu, C., Van Loosdrecht, M.C.M., 2005a. A framework for multidimensional modeling of activity and structure of multispecies biofilms. *Environ. Microbiol.* 7, 1085–1103.
- Xavier, J.B., Picioreanu, C., Van Loosdrecht, M.C.M., 2005b. A general description of detachment for multidimensional modeling of biofilms. *Biotechnol. Bioeng.* 91, 651–669.
- Xavier, J.B., De Kreuck, M.K., Picioreanu, C., Van Loosdrecht, M.C.M., 2007. Multi-scale individual-based model of microbial and bioconversion dynamics in aerobic granular sludge. *Environ. Sci. Technol.* 41, 6410–6417.
- Xavier, J.B., Foster, K.R., 2007. Cooperation and conflict in microbial biofilms. *PNAS* 104, 881–886.