QUANTICOL

A Quantitative Approach to Management and Design of Collective and Adaptive Behaviours

http://www.quanticol.eu

D2.1

A preliminary investigation of capturing spatial information for CAS

Revision: 1.0; Mar 31, 2014

Author(s): Vashti Galpin (UEDIN), Luca Bortolussi (CNR), Vincenzo Ciancia (CNR), Allan Clark (UEDIN), Rocco De Nicola (IMT), Cheng Feng (UEDIN), Stephen Gilmore (UEDIN), Nicolas Gast (EPFL/Inria), Jane Hillston (UEDIN), Alberto Lluch-Lafuente (IMT), Michele Loreti (IMT), Mieke Massink (CNR), Laura Nenzi (IMT), Daniël Reijsbergen (UEDIN), Valerio Senni (IMT), Francesco Tiezzi (IMT), Mirco Tribastone (SOTON), Max Tschaikowski (SOTON)

Due date of deliverable: Month 12 (March 2014)

Actual submission date: Mar 31, 2014

Nature: R. Dissemination level: PU

Funding Scheme: Small or medium scale focused research project (STREP)
Topic: ICT-2011 9.10: FET-Proactive 'Fundamentals of Collective Adaptive Systems' (FOCAS)
Project number: 600708

Coordinator: Jane Hillston (UEDIN)
e-mail: `Jane.Hillston@ed.ac.uk`
Fax: +44 131 651 1426

Part. no.	Participant organisation name	Acronym	Country
1 (Coord.)	University of Edinburgh	UEDIN	UK
2	Consiglio Nazionale delle Ricerche – Istituto di Scienza e Tecnologie della Informazione "A. Faedo"	CNR	Italy
3	Ludwig-Maximilians-Universität München	LMU	Germany
4	Ecole Polytechnique Fédérale de Lausanne	EPFL	Switzerland
5	IMT Lucca	IMT	Italy
6	University of Southampton	SOTON	UK
7	Institut National de Recherche en Informatique et en Automatique	INRIA	France

COOPERATION

Executive summary

Space is important in the QUANTICOL project because the project case studies include smart transport, and quantitative modelling of transport has inherent spatial aspects. This deliverable presents a review of the literature about spatial modelling within and beyond computer science, and a classification of the different approaches reviewed. The objective of the classification is to make clear what approaches are available and how they differ from each other. This will be used to guide future work on spatial approaches within the project. Furthermore, the classification enables the identification of the approaches that have been used in the initial work on case studies in the realm of smart transport.

This deliverable first identifies the aspects of non-spatial modelling that are important in the context of the QUANTICOL project. Time can be modelled in a discrete or continuous manner. States can be discrete, representing attributes of an individual. For example, when considering bike sharing, inUse (busy), onStand (idle) or atWorkshop (under repair) might be appropriate states for a bike. Alternatively, states can be continuous representing an attribute (for example, seat height). In the case of discrete states, it is possible to perform aggregation by considering populations, namely how many individuals are in each state, and to acquire an understanding of the overall behaviour of the population, rather than of individuals. Mean-field techniques can be employed to transform a discrete population approach to one that considers continuous populations that approximate the discrete approach.

To this context, space is introduced. Space can be discrete and described by a graph of locations. Depending on the structure of the graph and the parameters associated with locations and movement between locations, discrete space can be classified as regular or homogeneous. Space can be seen as continuous: as Euclidean space in one, two or three dimensions. Space can also be considered abstractly as topological space, whether discrete or continuous and this approach allows for reasoning about concepts such as adjacency and neighbourhoods.

This deliverable describes the modelling techniques that are currently available for the different combinations of time, state, aggregation and space, giving both a tabular classification as well as high-level and formal descriptions of the techniques. For each representation, examples are given of its use in different disciplines, including ecology, biology, epidemiology and computer science. In particular, the modelling goals are considered for these techniques, and compared with the goals of the QUANTICOL project. This deliverable also has the aim of identifying disparate uses of terminology in various approaches.

Both current and future case studies relevant to the project are classified in terms of how they use time, state, aggregation and space and finally conclusions are presented taking into account the literature reviewed, what has been modelled and what the goals of the project are for modelling of smart transport.

The preliminary guidelines arising from the review and classification are to focus on patch models and associated techniques although continuous space models of individuals may be important in certain cases. Items proposed for further research are understanding and developing mean-field techniques, spatial and non-spatial moment closure methods and hybrid spatial approaches. The document closes with a future work plan for Work Package 2.

Contents

1 Introduction

In the QUANTICOL project, we focus on the quantitative modelling of collective adaptive systems (CAS), looking particularly at smart transport systems such as bus services and bicycle sharing systems. An important aspect of both of these systems is location in and movement through space, and we wish to include this feature in our models to ensure that our modelling techniques allow for a sufficient level of accuracy to achieve our modelling goals. An appropriate starting point is to understand how space has been modelled both in the discipline of computer science and in other disciplines. Additionally, modelling of the QUANTICOL case studies is already ongoing using existing modelling techniques as the project is strongly driven by its domain of application and assessment of the way space has been modelled in these examples gives further guidance for how to proceed.

This deliverable presents the results of our investigation into the mathematical modelling techniques that have been used to represent space and their match to the objectives of the QUANTICOL project. Our general approach focusses on the behaviour over time of groups of individual entities, and understanding from the possible behaviour of these entities what the overall behaviour of populations is, considering both stochastic interpretations and deterministic approximations of this time-varying behaviour. To investigate the introduction of space into this approach, we have reviewed the literature to understand the modelling techniques that consider individuals and their populations in space. To clarify terminology, the terms *modelling technique* and *modelling approach* will be used to describe a general mathematical method used to construct models, and the term *model* will be used to describe a specific representation of a particular system using a modelling technique or approach. In this deliverable, we assume that the modelling approaches have associated analysis techniques: either in the form of simulation or mathematical techniques that lead to an analytic result, but we do not describe these techniques here.

This deliverable does not consider modelling languages such as process algebras as that is the focus of Task 2.2, and process algebraic features for modelling space will be considered in the Internal Report of Task 2.2 (due in month 33, December 2015). The deliverable also does not consider spatial programming languages such as Proto [VBU13] and TOTA [MVZ05]. The features of these languages will be considered later in this work package. Furthermore, this deliverable does not cover costs of analysis, transformations, convergence results and abstractions of and between spatial representations as these will be reported in Deliverable 2.3 (due in month 42, October 2016).

Hence this document reports on what representations are present in the modelling literature, drawing from a number of disciplines, and developing a classification of the techniques involved. These techniques will be described abstractly and concretely, through being related both to examples in the literature, and to the smart transport case studies that are currently being investigated in the QUANTICOL project. An aim of this document is understanding the goals of the existing techniques so that they can be compared to the goals of the project and so that preliminary guidelines can be provided. A secondary aim of the document is to identify disparate uses of terminology for the same technique across different disciplines (or the use of the same terms for different techniques) to provide clarification for the spatial modelling aspects of this project. Any review of this type has the potential to be rather long, and hence a separate technical report has been written [GFH$^+$14] for a discussion of details that are not directly relevant for this higher-level overview document.

This document is organised as follows. First, a high-level classification of space is presented, and some example modelling techniques discussed. After that there are three chapters, each relating to a different approach to space modelling. In each of these chapters, the techniques are described, examples are given of existing use and there is a discussion of how these approaches can be applied to our smart transport case studies, based on the work already done with respect to these case studies and looking forward to future work. Finally the document presents conclusions, preliminary guidelines, the relationship of this deliverable to other parts of the project and a work plan.

In terms of authorship, Vashti Galpin has written the bulk of the document with others contributing short pieces of text and descriptions of their modelling of various case studies.

2 A classification of space and movement modelling techniques

This section provides a framework in which to understand the various choices that can be made in terms of spatial modelling techniques. There exists a rich literature about these techniques and it is necessary to understand this before making decisions relating to the spatial aspects of the languages and their semantics in the QUANTICOL project.

We are interested in modelling the collective behaviour of many individuals using a stochastic approach (based on Markov chains that are discrete time or continuous time[1] and we use mean-field techniques[2] to obtain a fluid or deterministic approximation of their behaviour [Kur81]. We start from information about how each individual changes state over time, and we are interested in how this affects the overall behaviour of the system over time, either obtaining an exact stochastic analysis of the system, or a deterministic approximation of its behaviour.

Our specific modelling goals are determined by the case studies in the QUANTICOL project which are smart transport and smart grids. Choice of modelling technique should be informed by the sort of questions for which answers are required. The requirements section in Deliverable 5.1 [TCG$^+$14] considers some of these questions and others will be revealed as the project continues and properties for verification by spatial model-checking are considered. Furthermore, when existing models of case studies are discussed in Sections 3.8, 4.4 and 5.2, the goal of each model is identified. We return to consideration of this issue in the final section of this document.

2.1 Notation

To set the scene, some notation will be useful. For illustrative purposes in this document, we consider two *populations* P_A and P_B. At each point in time, each individual in P_A is in exactly one of the *states* $A_1, \dots, A_n$ and each individual in P_B is in exactly one of the states $B_1, \dots, B_m$. We can consider the populations in terms of individuals (we may need a naming convention to be able to refer to each individual) or in aggregation by counting the number of individuals in each state. This is called a *state-based aggregation* and it provides a view of the number of individuals in a population in a particular state, for each population and each state allowed for that population. Thus, let $N_{A_i}(t)$ refer to the number of individuals in population P_A that are in state A_i at time t and let $N_{B_j}(t)$ refer to the number of individuals in population P_B that are in state B_j at time t. These will be called *subpopulations*. The total number of individuals in the two populations at time t can be expressed as $N_A(t) = \sum_{i=1}^{n} N_{A_i}(t)$ and $N_B(t) = \sum_{j=1}^{m} N_{B_j}(t)$ respectively. Clearly, these counts are in $\mathbb{N}$ (which includes zero). Furthermore, if no births or deaths are assumed, and an individual must be in one of the available states[3], then $N_A(t_1) = N_A(t_2)$ for all times t_1 and t_2 and the size of P_A is a constant N_A; similarly the size of P_B is N_B.

To complete the notation required for populations, we use $X_{A_i}(t) \in \mathbb{R}_{\geq 0}$ to represent a non-negative real-valued description of the population P_A which in certain approaches is an approximation to $N_{A_i}(t)$ and similarly for $X_{B_i}(t) \in \mathbb{R}_{\geq 0}$ for P_B.

2.2 Non-spatial modelling dimensions

Before space is considered, there are already a number of choices that lead to different approaches to modelling dynamic systems in a quantified manner. We now consider the dimensions and the choices on each dimension as informed by the prior research of members of the project into dynamic modelling of systems [BHLM13]. For example, the time dimension considers how time is treated in different types of Markov chains. There are other aspects of time such as non-determinism and causality, but these are not a strong focus of our general modelling approach, and so are not included in the classification.

[1] See the Appendix of this document for a brief introduction to these concepts and population Markov chains.

[2] See the tutorial [BHLM13] and the references therein for more details of the mean-field approach.

[3] In some models, births and deaths can be included for a fixed size population by introducing a "dead" state. However, this requires that there is a finite maximum population size.

TIME	discrete			
AGGR	none		state	
STATE	discrete	continuous	discrete	continuous
	DTMC (discrete-time Markov chain) [Nor98]	LMP (labelled Markov process) [Pan09]	population DTMC [BHLM13]	difference equations, ordinary differential equations (ODEs) [BHLM13, MNS11]

TIME	continuous			
AGGR	none		state	
STATE	discrete	continuous	discrete	continuous
	CTMC (continuous-time Markov chain) [Nor98]	CTMP (continuous-time Markov process) [DP03]	population CTMC [BHLM13, Kur81]	population ODEs [BHLM13, Kur81]

Figure 1: Classification of mathematical models in terms of time, aggregation and state

Time: Time is usually non-negative, strictly increasing and infinite, and can either be a non-negative real or integer. In some models, a finite end-point may be used to delimit the period of interest.

***discrete*:** In the context of this research, discrete time is used in those modelling approaches where choices are probabilistic. At each clock tick (which could be associated with an integer if useful for the specific model), each individual chooses probabilistically its next state. For example, discrete time Markov chains (DTMCs) use this approach [KS76, Nor98].

***continuous*:** Here, time is continuous and this is captured by the fact that actions such as changing state have a duration associated with them. In the case of continuous time Markov chains (CTMCs), stochasticity is introduced by having random durations that are drawn from exponential distributions [Nor98].

State: States can be viewed as capturing a quality or attribute of an individual. As described above, an individual is assumed to be in a single state at each point in time[4].

***discrete*:** Usually when the states associated with an individual are discrete, there are a finite number of them. However, in the case of an attribute like *year-of-birth*, there may be a countably infinite number of values.

***continuous*:** A continuous-valued state can be interpreted as measurement of some quantity associated with the individual. An example of this would be *temperature* or *height*.

Aggregation: As discussed previously, individuals can be considered separately, or the focus can be on the number of individuals in each state. This is more relevant to discrete state approaches than continuous state. In the continuous case, aggregation can be described by a function, or discretisation can be applied to obtain frequency data.

[4] An individual could have more than one attribute, and then the individual's state is multidimensional with a value for each attribute. In this case, the individual's state can be seen as a tuple of values.

***none*:** Behaviour of each individual is considered separately.

***state-based*:** The behaviour of groups of individuals is considered by counting the number of individuals in each state over time (giving a non-negative integer value), or by having a non-negative real-valued approximation to this number. This approach appears under a number of different names in the literature including *population-based*, *state frequency data*, *numerical vector form*, and *counting abstraction*. The term *occupancy measure* is used when counts are normalised by the population size.

These possibilities can be expressed in a table, which can then be populated with mathematical modelling techniques from the literature. Figure 1 illustrates this and describes the modelling techniques that fit each combination of each element of each dimension.

2.3 Time-based non-spatial modelling techniques

An important aspect of our prior research is the application of the mean-field technique where the analysis of a population CTMC or DTMC can be approximated by an analysis using ordinary differential equations (ODEs) [Kur81, BHLM13]. As the number of states of a Markov chain increases (the "state-space explosion" problem), the analysis of the Markov chain becomes intractable. Modelling a large number of individuals can lead to a very large Markov chain. This can be mitigated by using a population Markov chain where behaviour is considered at a population level rather than at an individual level. The choice of a population Markov chain means we are interested in how many individuals are in each state i, given by N_{A_i}, and the states in the Markov chain have the form $(N_{A_1}, \ldots, N_{A_n})$. However, for large systems this may still not be sufficient to obtain reasonable analysis times, and an approximation using ODEs obtained from the population Markov chain can be used. This gives a system of ODEs for the variables $(X_{A_1}, \ldots, X_{A_n})$. This population Markov chain considers non-negative integer-valued population counts whereas the ODEs take a fluid approach and population quantities are non-negative real values. Considering the modelling techniques in Figure 1 for both discrete time and continuous time, the Markov chain obtained by considering many individuals (in the first column) can be transformed to a smaller Markov chain (in the third column) which can then be approximated by ODEs (in the fourth column).

This last transformation uses the mean-field approximation technique which comes originally from physics, where it refers to the approach where the movement of an individual particle is considered in the field generated by other particles rather than trying to solve the more complex problem of many particles interacting [CL07, MP12]. In modelling of systems, it has come to mean an approach where it is assumed that the number of individuals in a stochastic system becomes very large so that the population-level behaviour of the system can be expressed as ODEs which provide an "average" behaviour of the system. Results such as those proved by Kurtz [Kur81] demonstrate that under certain conditions, convergence occurs, namely as the number of individuals tends to infinity, the difference between the stochastic trajectories of the subpopulation sizes and the deterministic trajectories of the subpopulation sizes tends to zero. Practically, in many cases, good approximations using the ODE approach over the stochastic approach can be achieved at relatively low numbers of individuals [TGH12] and there are error bounds on the approximations [DN08].

Markov processes (in the second column) do not fit into this work flow and seem different from the other modelling techniques, as they are characterised by a continuous state space (which can also be interpreted as any continuous aspect of a model, including space).

2.4 Introducing space

Space can be considered in different ways.

***continuous*:** Here, space is represented by real values in the case of one-dimensional space, pairs of real values in the two-dimensional case and triples of real values in the three-dimensional case.

It is always (uncountably) infinite but may be bounded in extent. Continuous space used in this way can be seen as an exact representation of actual physical space.

***discrete*:** Approaches that use discrete space assume a number (usually finite) of distinct locations where connectivity between locations is described by an adjacency relation. At each location, there can be multiple individuals, although in some cases, such as cellular automata [Ila01], this may be restricted to a single individual. A location may be an abstraction or aggregation of actual space.

topological: This approach to space considers the relationships between points in space and contains no notion of time. It can be applied to both discrete and continuous space. Topological spaces consider space in an abstract manner using open sets from which concepts of continuity, adjacency and neighbourhoods are defined. Metric spaces have a notion of distance which is used to define these concepts.

As before, we introduce notation for the remainder of this document. For the purposes of this section, we only consider 2-dimensional space when we consider continuous space, either $\mathbb{R} \times \mathbb{R}$ or a bounded contiguous subset of $\mathbb{R} \times \mathbb{R}$. Since this provides a uncountably infinite set of points, the most straightforward way to refer to each point is by its coordinates (x, y).

In the case of discrete space, we assume a finite (or at most countably infinite) set of points $\mathcal{L}$ with some naming convention. There can be a relationship between points in 2-dimensional continuous space and locations in discrete space. If there is a partition of continuous space into regions then each location can represent a region in continuous space.

Discrete space also requires notions of adjacency and neighbourhood. In the general case, the set of locations $\mathcal{L}$ can be taken as the vertices of a graph, and the connections between locations (the adjacency relation) can be defined as edges in that graph. Then the edges of the graph $E_{\mathcal{L}}$ are drawn from the subsets of size two of the location set $\mathcal{P}_2(\mathcal{L})$, so $E_{\mathcal{L}} \subseteq \mathcal{P}_2(\mathcal{L})$. Each edge has the form $\{l_1, l_2\}$, and edges of the form $\{l, l\}$ are permitted. For reasons we describe in Section 3, we have chosen to use an undirected graph which is to be understood as allowing movement or interaction in at least one direction between the two locations.

By adding restrictions to the general case, subclasses of discrete space can be obtained. For example, *regular discrete space* modelling techniques assume that there is a *regularity* in defining the neighbours of a location but not necessarily in parameters [DL94b, OS71]. In contrast, *homogeneous space* techniques take a different approach and assume *full connectivity* between all regions (giving a complete graph) and equality of parameters between locations and at locations [Che81]. Both of these issues will be discussed in more detail and formality in the section on discrete space (Section 3).

We now address an issue of terminology. The term *map* will be used to represent a two-dimensional continuous representation of space that is to scale, meaning that the ratio between distances is preserved. On the other hand, the terms *location graph* or simply *graph* will be used to denote a representation that is not to scale but indicates the connections between locations. In between these two, is the category of *topological map* which is

> *"a type of diagram that has been simplified so that only vital information remains and unnecessary detail has been removed. These maps lack scale, and distance and direction are subject to change and variation, but the relationship between points is maintained. A good example is the tube map of the London Underground."* [Wik13b]

The distinction between a topological map and a graph is that although both represent relationships between points, the topological map is a continuous deformation of the original map, and it also is a two-dimensional representation of a graph. A graph is more abstract and has no specific two-dimensional representation. Note that planarity (the lack of overlapping edges when embedded in the plane) is not required for topological maps or graphs. Any topological map can be abstracted to a graph of locations, hence defining discrete space.

TIME	continuous			
AGGR	none		state and/or space	
STATE	discrete	continuous	discrete	continuous
SPACE				
discrete				
general	multiple individuals CTMC	TDSHA, PDMP [BP10, Dav93]	patch CTMC [CLBR09]	patch ODEs [CLBR09]
regular	interacting particle system (IPS) [DL94b]		foraging model [MSH05]	foraging model [MSH05]
homogeneous			bike sharing [FG12]	bike sharing [FG12]
continuous	molecular dynamics [CPB08] agents	CTMP [DP03]	spatio-temporal point processes [SBG02]	partial differential equations (PDES) [HLBV94]

Figure 2: Classification of mathematical models in terms of time, aggregation, state and space

A table has been constructed to identify mathematical models for the different combinations of time, aggregation, state and space (see Figure 2). Here, we have chosen to focus on continuous time models; however there are discrete time models of various approaches, for example, some variants of interacting particle systems (IPSs) use probabilities [DL94b].

All the models appearing in the table consider changing behaviour over time. Characteristics of space may or may not change as time passes[5]. In the case where there are no changes, space can be considered independently of time and represented as a topological space. When there are changes in the characteristics of space over time, the characteristics of space at a specific point in time can be considered topologically.

The next section considers each entry of the table in Figure 2 and illustrates the ideas using a consistent diagrammatic framework.

2.5 Spatial modelling techniques

The techniques described in this section are mainly continuous time, although some have discrete time analogues, as mentioned above.

Discrete space, no aggregation, discrete state: The techniques in this category consider space to consist of a (finite) number of locations that have connections between them. The most straightforward way is to consider these models as graphs with the locations as nodes and the links as edges. This type of model is illustrated in Figure 3(1). Here, and throughout this subsection, we assume individuals from the two populations already defined. The first, P_A consists of red and white tokens, and has states A_1 and A_2, using our previously introduced notation. The second, P_B consists of blue and white tokens with states B_1, B_2 and B_3. The current state of an individual is indicated on the

[5] An example of this would be off-peak road closures for the painting of road markings.

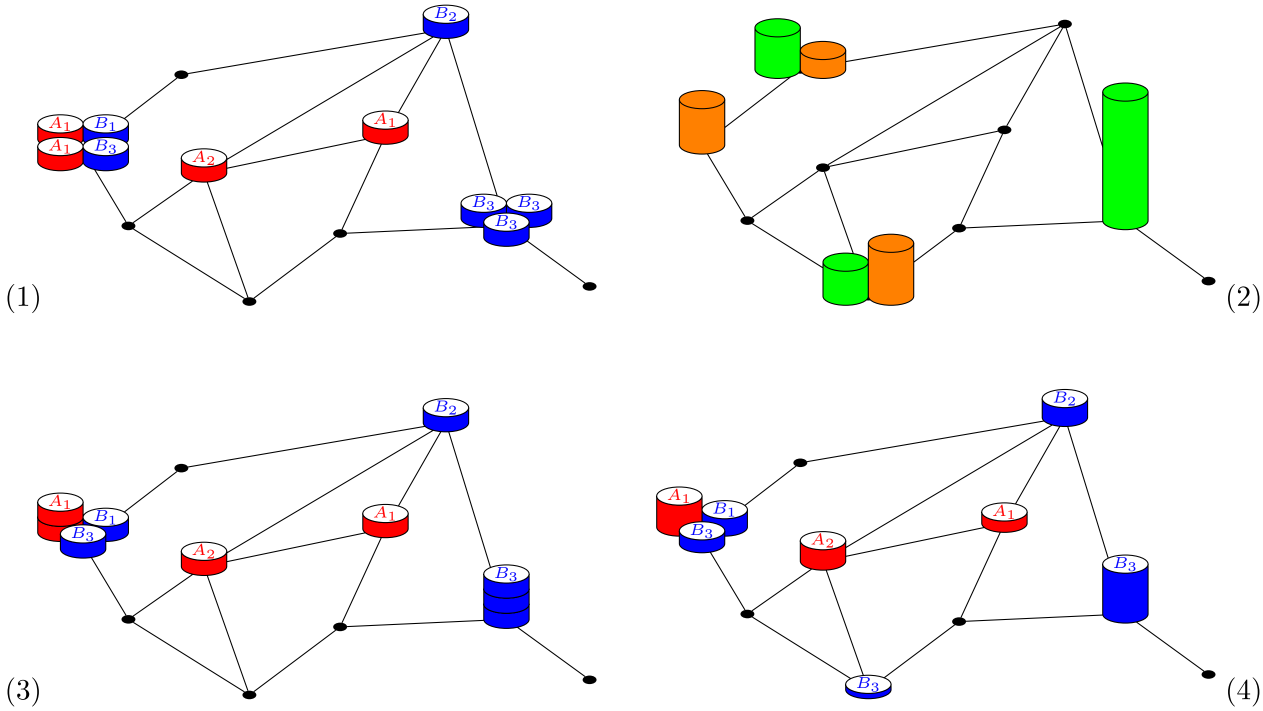

Figure 3: Discrete space: (1) no aggregation, discrete state; (2) no aggregation, continuous state; (3) aggregation of state, possible aggregation of space, discrete state (4) aggregation of state, possible aggregation of space, continuous state

top of the token. Note that the four diagrams in Figure 3 represent four single points in time and do not show change over time (and similarly for subsequent diagrams). For two-dimensional and three-dimensional space, the best visualisation method for change over time is video. For one-dimensional space, a graph with two axes can be used.

Regular space models in this category have a regular pattern of locations [DL94a, DL94b]. For example, the locations could be laid out in the rectangular grid, or a hexagonal tiling. The locations that represent space can be placed at the nodes of the regular graphs or in the spaces (faces) created by the regular graph as shown in Figure 4(1). Some models only allow one individual in each location, such as interacting particle systems (IPSs) [DL94b] and cellular automata (CA) [Ila01], but others may allow multiple individuals. There is no aggregation of individuals. In Section 3, regular space will be formally defined.

Discrete space, no aggregation, continuous state: These techniques differ from those above in the fact that the state is continuous. This is indicated by a solid token where the height indicates the value of a single continuous state. This is an inherently continuous value rather than the notion of approximation by continuous values described earlier in this section. This could be viewed as a measurement such as strength of radio signal or length of battery life. In Figure 3(2), there is an assumption of at most one individual per node, and two values associated with that individual. Different colours have been used in the diagram to make it clear that the values are continuous.

The major difference between this category and the previous one when regular space is considered is the fact that instead of having discrete states, there is one or more non-negative real values associated with each individual as shown in Figure 4(1).

Discrete space, aggregation, discrete state: These techniques differ from those in the first category above in the fact that there is aggregation [MP12]. This means that instead of each individual

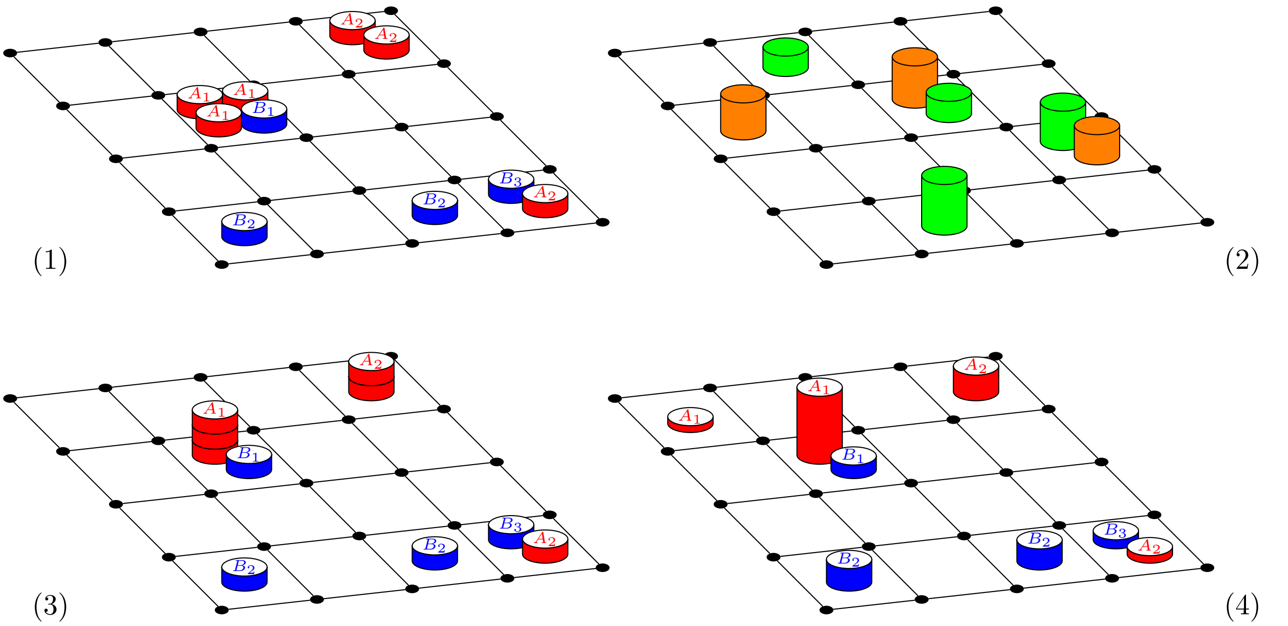

Figure 4: Regular discrete space: (1) no aggregation, discrete state; (2) no aggregation, continuous state; (3) aggregation of state, possible aggregation of space, discrete state (4) aggregation of state, possible aggregation of space, continuous state

being treated separately, individuals in the same state are considered as populations. This is illustrated in Figure 3(3) by the fact that individual tokens are grouped into stacks at nodes in the network. Figure 4(3) shows the regular space case where populations are aggregated at each location [EE04].

Discrete space, aggregation, continuous state: Here each region or point is associated with approximations to the discrete population approach mentioned in the previous item [MP12], as shown in Figure 3(4). At each node, for each state in each population, there is a real number that approximates the number of individuals in that state. This is illustrated in the graph by a column with a real-valued height for each state in each population. Note that in Figure 3(4), the lowest node has a non-zero value for blue tokens in state B_3 although there were none in the CTMC model in Figure 3(3), illustrating that approximation can occur. The case of regular space [LD96] is illustrated in Figure 4(4).

Continuous space, no aggregation, discrete state: These are approaches where each individual's location and state are modelled separately from those of other individuals. An example of this type of model is where the movement and interaction of each molecule is modelled individually in molecular dynamics [BU10]. Agent-based models take a similar approach. Figure 5(1) illustrates this. The continuous space is indicated by a bounded area and each individual is shown at its own location. These models are typically computationally expensive to simulate.

Continuous space, no aggregation, continuous state: In contrast with the previous category, the state is now continuous rather than discrete [DP03]. Since there is no aggregation, this approach models individuals rather than populations. The continuous space is indicated by a bounded area and each individual in shown at its own location. The continuous state is indicated by the varying heights of the tokens, and in Figure 5(2), it is assumed that there is only one (non-spatial) measurement per individual, although two different qualities may be measured.

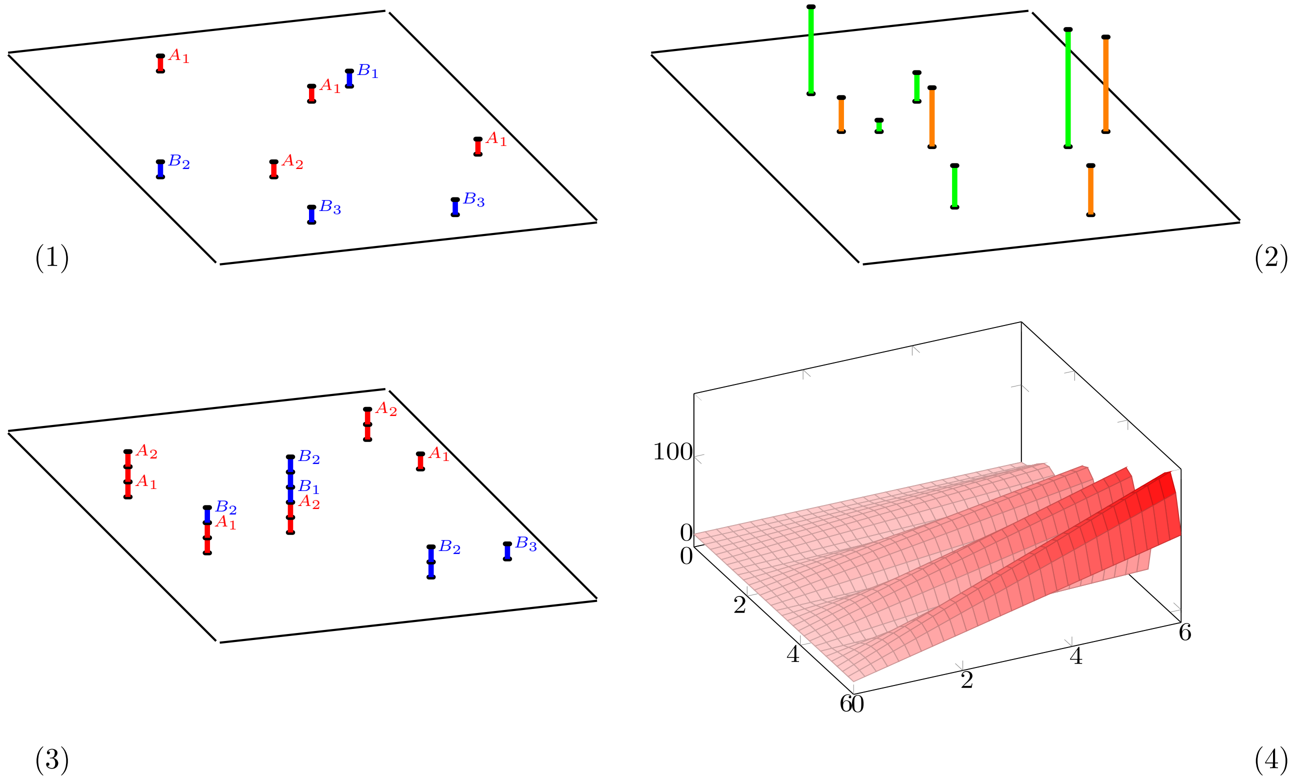

Figure 5: Continuous space: (1) no aggregation, discrete state; (2) no aggregation, continuous state; (3) aggregation, discrete state; (4) aggregation, continuous state.

Continuous space, aggregation, discrete state: In these techniques, each point in space can be filled by one or more individuals [SBG02]. Hence for each point in space, it is possible to aggregate the number of individuals in each state. Figure 5(c) show a fairly sparse number of individuals but much denser arrangements are also possible.

Continuous space, aggregation, continuous state: At each point in space, there is a real value describing an approximation to the number of individuals at that point [OL01, CPB08]. In the case of two-dimensional space, the population of each state can be represented in three-dimensions by surfaces. Figure 5(d) illustrates a surface describing the number of individuals at each point for state A_1. In contrast to Figure 5(c), this figure illustrates a very dense situation.

As is the case with techniques that do not include space, presented in Figure 1, the techniques using continuous state without aggregation (the second column of models in Figure 2) seem distinctly different to the other approaches. The techniques that can be applied to models without space described early in this document (approximation by ODEs of a population DTMC or CTMC) can be applied to discrete space since the Markov chain involved is a population Markopv chain that takes location into account. Furthermore, taking the hydrodynamic limit of IPS models provides PDEs [DMP91]. This subsection has not considered movement which will be examined now.

2.6 Movement

In all of the models described in the previous section, there may be interaction between individuals (even if this interaction is expressed at the population level). Opportunity for interaction is often related to colocation or proximity (which requires some notion of neighbourhood or distance). Many models capture movement of individuals explicitly and then use colocation or proximity to determine the possibility of interaction, although there are some models that only use proximity without movement such as IPSs and CA. In these two modelling techniques, space is regular and discrete and at most one individual is present at each location.

Because of the importance of movement in the modelling of smart transport, we must consider the choices that can be made, and they are now discussed for the two different types of time-based spatial modelling techniques, described in Section 2.4. However, in the case of continuous space, this discussion is split into modelling techniques where there is aggregation and those where there is none.

2.6.1 Discrete space

Assuming an undirected graph of locations, the presence of an edge between two locations describes the fact that movement or interaction along that edge is possible in at least one direction. The absence of an edge can be interpreted as meaning that movement and interaction can never take place, in either direction. As we will see in Section 3, parameters associated with an edge express (possibly in a time-varying manner) the propensity for movement or interaction in either direction. If it is zero at a particular time for a particular direction, it means that no active interaction or movement can take place at that time point. Hence, the graph of locations provides a skeleton for describing what movement or interaction is possible.

The adjacencies created in a location graph of regular space can capture where movement or interaction may occur (possibly with some weighting to capture likelihood) or what the neighbourhood of a location is.

2.6.2 Continuous space, no aggregation

In the case of continuous space where individuals are not aggregated, there are many different models of movement through two-dimensional space, such as models of animal movement and models for ad hoc and opportunistic networks [CBD02]. These are often random and capture the probability of movement in a particular direction at a certain speed. An additional concern is to determine what happens at the boundary of the space. This concern can be avoided by assuming the space is the surface of a torus and hence has no boundaries – this is more common than assuming the surface of a sphere. There are also models to describe the movement of a related group of individuals through the space [CBD02]. Connectivity models on the other hand, describe interaction (for example, contact duration and time between contacts) rather than location [KP07, CFB09, CMRM07]. Interaction can be interpreted as dynamic graphs with the individuals as the nodes. Connectivity models are more abstract than movement models.

2.6.3 Continuous space, aggregation

Here, movement is expressed through the form of the PDE. Diffusion-reaction PDEs are used since they can express movement as diffusion and interaction as reactions [CPB08, HLBV94, OS71, Tur53]. The diffusion terms can capture drift which is caused by obstacles or external stimuli such as wind, the likelihood of continuing in the same direction, the effect of the density of other individuals, and the impact of environmental characteristics. The reaction term describes interactions between individuals.

2.7 Conclusion

Figure 2 provides a classification of different spatial models taking into account aspects of time, aggregation and state. Movement models can be associated with particular aspects of space. This section has outlined some basic ideas, and the next three sections consider three different approaches to space. Two are time-based, namely discrete space and continuous space, and one is static, topological space. Each section will provide further details about the different approaches to modelling space, including a discussion of existing models and the areas in which they have been applied, and how these models are related to the smart transport case studies considered in the project.

3 Discrete space modelling techniques

We will focus here on the continuous time models, with pointers to the discrete time models where appropriate. As mentioned in the previous section, we assume a set of locations $\mathcal{L}$ and an undirected graph over locations $(\mathcal{L}, E_{\mathcal{L}})$ with $E_{\mathcal{L}} \subseteq \mathcal{P}_2(\mathcal{L})$. Each edge has the form $\{l_1, l_2\}$, and loops such as $\{l, l\}$ are allowed.

Locations in discrete space models can have two main sources, either they are essentially locations on a map, such as bike stations or bus stops, or alternatively each location represents a region on a two-dimensional map, and space is aggregated. The edges of the graph can be determined by various factors. Adjacency of regions is an obvious choice, but there may be other context-specific elements such as presence of connections between regions such as railway lines or similar.

If each location is used to represent a distinct region of continuous 2-dimensional space, a function $f : \mathbb{R} \times \mathbb{R} \to \mathcal{L}$ can be defined with the requirement that $f^{-1}(l_1) \cap f^{-1}(l_2) = \emptyset$ for locations $l_1 \neq l_2$, namely that regions are disjoint. Depending on the context, the union of $f^{-1}(l)$ for all l may be the whole continuous space under consideration so f defines a partition of the space, or this union may be a subset of the continuous space, only representing regions of interest. In some cases, where not all of continuous space is of interest, a single location can be used to represent the uninteresting regions, thus ensuring a partition.

A modelling technique with discrete space will have parameters that depend on locations, or links between locations. We can consider two groups of parameters; those that are associated with locations, namely with vertices of the graph and those that are associated with interaction or movement, namely the edges of the graph, and we define two functions to describe these parameter sets as follows

- $\lambda(l)$ for $l \in \mathcal{L}$, and
- $\eta(l_1, l_2)$ and $\eta(l_2, l_1)$ for $\{l_1, l_2\} \in E_{\mathcal{L}}$.

The range of these functions will remain abstract for the purposes of this discussion. Note that although the edges of the graph are not directed, the function η is sensitive to direction. We have chosen to use undirected graphs to give a basic skeleton to possible movement and interaction between locations, and to use the parameters to capture the directionality of that movement or interaction. Movement is obviously directional. Interaction can be undirected when considering an abstract view of effect or communication. Alternatively, it can be directed if one party is the sender and the other the recipient. This could be synchronous when the sending and receipt happen simultaneously, or asynchronous when the receipt happens later. Our choice of an undirected graph allows these details to be expressed in parameters. This separation of concerns is also useful when different subpopulations have different forms of interaction. In the rest of this document, the term *transfer* will be used to refer to both movement and interaction.

A topic whose exploration is beyond the scope of the current document but should be mentioned is that of how to divide a map in regions. A simple approach is to base it on a tiling of the plane using triangles, quadrilaterals or hexagons. More complex approaches involve taking local information into account and creating irregular patches. This is a topic for further research within QUANTICOL.

The discrete space approach as described above is very general as it allows arbitrary graphs over locations, as well as heterogeneity for parameters. In the literature there are modelling techniques that are defined for specific graph subclasses and we now identify two important spatial subclasses: spatially homogeneous and regular.

3.1 Spatial homogeneity

To be able to discuss formally aspects of discrete space, we develop the following definitions, leading to a definition of spatial homogeneity (a term which is used in the literature but not formally defined), by considering the location-related parameters. A spatial model is

- *location homogeneous* if $\lambda(l_i) = \lambda(l_j)$ for all locations $l_i, l_j \in \mathcal{L}$.
- *transfer homogeneous* if $\eta(l_i, l_j) = \eta(l_j, l_i) = \eta(l_{i'}, l_{j'}) = \eta(l_{j'}, l_{i'})$ for all edges $\{l_i, l_j\}, \{l_{i'}, l_{j'}\} \in E_{\mathcal{L}}$.
- *(spatially) parameter homogeneous* if it is both location and transfer homogeneous.
- *spatially homogeneous*[6] if it is parameter homogeneous, and its location graph is complete[7]. Regular connections between locations which do not give total connectivity are discussed in the next section on regular space.

Models with spatial homogeneity have a symmetry that can allow for analyses that are not possible for more complex models. Examples are the bike sharing system considered in [FG12] where the metrics of interest are the number of empty and full bike stations.

Spatial inhomogeneity can be introduced in two ways: the first involves connectivity where equal accessibility is no longer assumed, and the second where all locations are still accessible from all other locations, but parameters vary between locations. These are not necessarily distinct concepts. Consider the case where there is a parameter $\rho_{i,j} \in \eta(l_i, l_j)$ which describes the rate of movement from location i to location j. If $\rho_{i,j}$ is the same for all i and j and no other parameters vary by location then the model is spatially homogeneous. However, if $\rho_{i,j}$ can vary and possibly be zero then not only does a specific parameter vary by location but additionally, equal accessibility no longer holds (either because on average it takes longer depending on the rate, or if the rate is zero there is no accessibility). However, if $\rho_{i,j}$ is constant for all i and j but other parameters vary by locations, then the model is spatially inhomogeneous.

3.2 Spatial regularity

The category of regular discrete space covers those spatially inhomogeneous models where the organisation of space is regular (rather than an arbitrary graph where each vertex may have an arbitrary number of edges) but parameters can vary for each location in space.

In contrast to spatial homogeneity, regularity of space is more difficult to define formally when starting from a graph, although it is very straightforward to identify visually [OS71]. There are three possible approaches to describing regularity in two dimensions and these are discussed in more detail in the associated technical report [GFH$^+$14].

1. A lattice, grid or mesh graph is defined as "a graph whose drawing, embedded in some Euclidean space $\mathbb{R}^n$, forms a regular tiling" [Wik13a]. Since we focus on two-dimensional space, we only consider planar graphs, and we obtain those graphs obtained from regular tilings by equilateral triangles, squares and regular hexagons.
2. Another approach to specifying graphs of regular space is to specify how many edges each face[8] of the graph has and what the degree of the vertices are. For example, a regular location graph with triangular faces is a planar graph in which each face has three edges and each vertex has degree six.
3. Finally, a graph with regular structure can be constructed by identifying points in $\mathbb{Z} \times \mathbb{Z}$ or $\mathbb{R} \times \mathbb{R}$, and adding links between these points.

There are other divisions of two-dimensional space that can be viewed as regular such as that provided by a dartboard but we will not attempt that level of generality for discrete space beyond saying that

[6] This is a different notion to the graph theoretic definition of homogeneous graph which is a condition on isomorphic subgraphs [Gar76] and to homogeneous Markov graphs which are a specific class of random graphs [FS86].

[7] A complete undirected graph has an edge $\{l, l'\}$ between each pair of vertices l and l'.

[8] In a planar graph, a face is a region bounded by edges.

regular space should have the property that at each location (except possibly at boundary locations) there is a uniform way to determine the neighbours. However, we exclude from this definition n-hop neighbours in an arbitrary graph (see definition of n-hop in the next subsection).

One-dimensional regular space can be represented simply as a undirected path. We do not tackle the definition of three-dimensional regular space.

Note that the complete graph requirement for spatial homogeneity means that a location graph cannot be both regular and homogeneous[9]. However, a spatially regular location graph can be parameter homogeneous. We will use the term *irregular space* whenever the usage of space is not regular or homogeneous. To distinguish irregularity from regularity, we can consider the pattern of connectivity between neighbours and this is discussed in the next subsection.

3.3 Neighbours and neighbourhoods

In an undirected graph of locations representing discrete space, the links between locations are used to define neighbours. Given a location l, its *immediate neighbours* are those vertices l' such that $\{l, l'\}$ is an edge in the graph. Its *n-hop neighbours* are those that can be reached through a path in the location graph of at most n steps (but excluding the location l itself). In the case of a regular grid graph, the immediate neighbours (west, north, east and south) are referred to as the Von Neumann neighbourhood. The larger neighbourhood that includes the northwest, northeast, southeast and southwest points as well as the immediate neighbourhood is known as the Moore neighbourhood. Both types of neighbourhoods can be extended to n-hop neighbours and also applied to hexagonal and triangular regular location graphs.

This is a purely spatial approach to defining neighbourhoods. However, in some cases, it can be the entity or process itself that defines its neighbourhood depending on its capability. Other approaches use a (perception) function that determines the neighbours of an individual by specifying the other individuals with which it can interact.

To distinguish irregular location graphs from regular location graphs, one can say that if it is possible to define the 1-hop neighbours of a location in a regular fashion except at boundaries, then this will be considered a regular discrete space model even if the parameters can vary by location. To make this meaningful, it is necessary to exclude the "regular" notion of 1-hop neighbours in a general graph structure from this definition of regularity. An example of this is given in the approach taken to modelling fluid limits for stochastic mobile networks [TT13]).

3.4 Boundary conditions

An issue for discrete space (and continuous space) is determining what happens at the boundaries of the space. One approach is to ensure there are none by working with infinite structures such as infinite graphs, or alternatively boundaryless structures such as tori. A rectangular region can be transformed into a torus by joining the top and bottom edges (to form a cyclinder) and then joining the left and right ends (by curving the tube). Other approaches work with boundaries and either choose to keep individuals inside the region (by reflection or other techniques) or to treat boundary locations as sources and/or sinks.

Furthermore, when solving ODEs, there may be boundary conditions that constrain the ODEs by specifying the values that must occur in boundary regions [AMR95]. The conditions could constrain the value of the solution (Dirichlet, first-type) or the value of the derivative (Neumann, second-type). Cauchy boundary conditions provide a curve or surface that constrains both solution and derivative.

[9] The complete graph with three vertices is regular and can be spatially homogeneous. However, since we are considering tilings/graphs with multiple faces, this graph is not included in our definitions.

3.5 Discrete space without state-based aggregation

We now consider the different modelling techniques that have been applied to discrete space starting with those that do not involve aggregation. When there is no aggregation and state is discrete, the technique models individuals and can be seen as an agent-based system where space is discrete. Hence each individual has some state and is located at exactly one location. Multiple individuals can be located at a single location. The movement of individuals between locations can be determined by a rate in the case of continuous time. To describe these in their most general form, we assume that each individual I (where I is a unique name for the individual) has associated time-based information:

- $\mathsf{loc}(I,t) \in \mathcal{L}$ which is its location at time t
- $\mathsf{state}(I,t) = A_i$ which is its state at time t

Moreover, there are rules that describe how an individual can change location or change state. Since this is a continuous time model, these rules may specify rate constants (each rate defining an exponential distribution) to describe how long it takes for the changes to occur, or the rates may be functional rates (but still defining an exponential distribution) that take into account the presence of others at that location, the characteristics of the location or even the current time (thus introducing time inhomogeneity). The behaviour of the agents in this modelling technique is thus described as they individually change state and/or location. Assuming a fixed population size, we can model this system as a CTMC, where each state in the CTMC is a tuple consisting of information about each individual in the system. If we assume N individuals then a state has the following form

$$\big((\mathsf{loc}(I_1,t),\mathsf{state}(I_1,t)),\ldots,(\mathsf{loc}(I_N,t),\mathsf{state}(I_N,t))\big)$$

There are $(L \times n)^N$ states in this Markov chain if there are L locations and n states and if it is possible for all individuals to be in all possible combinations of location and state.

Simulation suits this type of model, and techniques for simulating systems where behaviour is based on functional exponential rates are well understood [Gil97]. Typically in the case of general discrete space, movement is assumed to be single-hop so that movement is only possible to an immediate neighbour. In regular space, movement is often possible to an n-hop neighbourhood, and in spatially homogeneous models, movement is possible to any other location as a result of the completeness of the location graph.

Next, considering discrete space modelling techniques without aggregation where the state is continuous, instead of having a rule describing what the next state is, there needs to be a rule describing how this continuous value changes over time. A good candidate for this type of rule is an ODE. These techniques are hybrid in that they exhibit both continuous and stochastic behaviour. Additionally, they may also have instantaneous behaviour. Transition-driven stochastic hybrid automata (TDSHAs) [BP10] and piecewise deterministic Markov chains (PDMPs) [Dav93] are suitable modelling techniques.

3.6 Discrete space with state-based aggregation

In the two approaches that are described next, aggregation of state occurs. It is assumed that we have many individuals to whom the same set of rules apply with the same parameters, and we choose to view them as a population, P_A and to reason about them as a population. To extend the notation introduced earlier, assume we have a fixed number of locations, $l_1,\ldots,l_L$. We can now consider the counts of subpopulations at each location. So for P_A, we have a value $N_{A_i}^{(k)}$ which is the number of individuals at location k in state i. Additionally

$$N_{A_i} = \sum_{k=1}^{L} N_{A_i}^{(k)} \qquad \text{and} \qquad N_A^{(k)} = \sum_{i=1}^{n} N_{A_i}^{(k)} \qquad \text{and} \qquad N_A = \sum_{i=1}^{n} N_{A_i} = \sum_{k=1}^{L} N_A^{(k)}$$

We can create a continuous time Markov chain smaller than that of the previous section consisting of at most $(N_A + 1)^{L\times n}$ states where each state has the form

$$\big(N_{A_1}^{(1)}, \ldots, N_{A_n}^{(1)}, \ldots, N_{A_1}^{(k)}, \ldots, N_{A_n}^{(k)}, \ldots, N_{A_1}^{(L)}, \ldots, N_{A_n}^{(L)}\big).$$

This provides a discrete aggregated representation of space where for each location, we know how many individuals are in each state without knowing exactly which individual at that location is in which state. As in the case without space, this is a population CTMC. This model is amenable to state-based analysis techniques (assuming a small enough state space) and stochastic simulation. There appears to be no real difference in the size of the Markov chain when using regular space in this technique. In the case of spatial homogeneity, the fact that parameters are identical may make the model amenable to an analytic approach, rather than requiring simulation [FG12].

In the continuous state variant of this technique, the notation $X_{A_i}^{(k)}$ is used for the real value that describes the quantity of individuals in state i at location k. Since this can be a non-integer value, it is an approximation to the actual count $N_{A_i}^{(k)}$. Since the subpopulation sizes are treated as continuous values, a standard modelling technique is to express the change in this quantity in terms of a population ODE which tracks the changes in subpopulation size over time. There are $L \times n$ variables in total; one for each combination of state and location. This ODE often has the following form

$$\begin{aligned}\frac{dX_{A_i}^{(k)}}{dt} &= f_{i,k}\big(X_{A_1}^{(k)}, \ldots, X_{A_n}^{(k)}\big) + \\ &\sum_{j=1, j\neq k}^{L} \big(g_{i,k,j}\big(X_{A_1}^{(k)}, \ldots, X_{A_n}^{(k)}, X_{A_1}^{(j)}, \ldots, X_{A_n}^{(j)}\big) - h_{i,k,j}\big(X_{A_1}^{(k)}, \ldots, X_{A_n}^{(k)}, X_{A_1}^{(j)}, \ldots, X_{A_n}^{(j)}\big)\big)\end{aligned}$$

where $f_{i,k}$ captures the local behaviour which only depends on the subpopulation sizes locally, $g_{i,k,j}$ describes the inflow of population from location j to location k, $h_{i,k,j}$ describes the outflow of population from location k to location j, and these flows depend only on the subpopulation sizes in location k and location j. This is a time-homogeneous ODE since change over time is only dependent on subpopulation sizes (that are dependent on time) rather on time directly.

For both the general and regular space cases and assuming only movement/interaction between 1-hop neighbours, then a term $X_{A_i}^{(j)}$ should only appear in the right hand side of the ODE if $\{l_k, l_j\}$ is an edge in the location graph. A spatially homogeneous model would require terms from all locations.

The modelling techniques in the categories of discrete state (with or without aggregation) and continuous state with aggregation (these are the categories in the first, third and fourth columns for general discrete space in Figure 2) are very similar to those in the same categories (and the same columns) in Figure 1. The discrete state approaches without aggregation (first column) are characterised by CTMCs, discrete state approaches with aggregation (third column) by population CTMCs which abstract from individuals, and continuous state approaches (fourth column) by ODEs. This similarity is not surprising, as in the general case of discrete space, location is essentially another attribute and hence the same techniques apply. Similarly to before, the technique for continuous state modelling without aggregation (second column) is out of step with the other techniques since it requires a hybrid approach when discrete space is introduced.

Introducing locations while keeping the population size fixed can result in a decrease in accuracy of the approximations as now the number of variables (and ODEs) has increased by a factor of L and hence each ODE refers to a smaller number of individuals thus leading to population fragmentation and less justification for applying mean-field techniques at a location.

3.7 Examples of existing use

This section reviews the literature of modelling with discrete space techniques, considering different approaches in different disciplines. It is selective and illustrative rather than exhaustive.

Before going further, we diverge into a discussion of terminology as it is not consistent between different disciplines or even within disciplines. Morozov and Poggiale [MP12] identify five different uses of the term "mean-field" in ecology. Four of these usages explicitly include the use of space and hence differ from the meaning used in this document (as presented in Section 2.3) which does not include any spatial aspects. Another confusing term is "patch models". Durrett and Levin [DL94a] describe a patch model (Chesson [Che81]) as one in which all locations are connected to each other. In later papers, the terminology refers to discrete space models with arbitrary graphs, and this is the sense in which we will use "patch". Hence, both the discrete and continuous population models defined over general discrete space in this section are patch models.

3.7.1 Ecology

Space plays an crucial role in ecological models but the modelling goals are often qualitative (such as persistence, coexistence, stationarity, oscillatory behaviour, chaos or multistability [MP12]) or quantitative at the global level (proportion of sites occupied, for example). Hence, there is a focus on global behaviour. Berec [Ber02] classifies spatial models according to their time, space and population dimensions. This last dimension differs from our classification, as it only considers aggregation.

discrete population, continuous time: In interacting particle systems, behaviour is determined by a set of rules. These systems use a regular discrete space approach often with parameter homogeneity and an assumption of at most a single individual at each location [DL94b].

continuous population, continuous time: Reaction-dispersal networks describe change over time by a system of ODEs over species in locations. These are also called metapopulation models [Lev69] and patch models [Lev74, Tak96] and they are the same as patch ODE models in our terminology. The ODEs used in these models often focus on the probability of each patch being in one of a number of states (such as uninhabitable, habitable but unoccupied, occupied) [XFAS06] so as to determine the proportion of occupied patches and to identify equilibria. Colonisation rates can reflect distance between patches, so these models are not often transfer homogeneous. In hierarchical patch models, different spatial and time scales are taken into account [WL95].

discrete population, discrete time: These are regular discrete space individual-based models where the behaviour of an individual is given by a list of rules which are applied at each time tick. Probabilistic cellular automata fall into this class as do interacting particle systems which use probabilities rather than rates [DL94b].

continuous population, discrete time: Coupled-map lattices are a type of discrete time equation where coupling terms capture the effect of neighbours. They can be defined by difference equations [HCM91, Kan98]. They use regular space and continuous subpopulation sizes.

We now consider some relevant references in detail. Durrett and Levin [DL94a] compare a non-spatial mean-field model, a PDE model that uses continuous space, a spatially homogeneous model by Chesson [Che81] and a grid based IPS model. The authors identify different scenarios in which there are different outcomes from the models due to stochastic or spatial features. This illustrates that the choice of model is important and can affect the global results.

Morozov and Poggiale [MP12] start with a discrete space, discrete aggregation model (such as a patch CTMC) and consider the various ways to obtain mean-field models. The first four are essentially moment closure models that provide mean-field models of global subpopulation sizes by introducing spatial terms into the ODEs, namely spatial moment closure [MSH05], pairwise approximation approach [WKB07], modified mean-field technique [PRL11] and scale transition theory [Che12]. The fifth technique considers the structure of the model and differences in dispersal rates [APS12].

As can be seen, there are multiple approaches to mean-field models for patch-based models in ecology. Most consider global outcomes and qualitative assessment with few focussing on local details.

3.7.2 Biology

Bittig and Urmacher [BU10] identify five distinct methods for spatial modelling in cell biology that offer different granularities in their approximation of physical reality. Two of these are continuous space approaches and will be discussed in the next section.

compartments: In this model, space is divided into compartments which may be contained in other compartments, giving nesting of space, although adjacency is also used. Compartments represent the physical reality of cells. Within compartments, a non-spatial approach is used and transfer rates are defined between compartments. This technique is an irregular discrete space patch-based population CTMC approach or ODE approach, A spatial moment approximation using the log-normal distributions has been applied in this context [MMRL02].

discrete space, lattice: The size of the lattice is chosen to reflect molecule distance and size, and at most one molecule can occupy a lattice block, requiring a way in which to deal with collisions. This is similar to cellular automata [Ila01] and the cellular Potts model [GG92, CHK$^+$05] based on the Ising model [Bru67] where a potential energy function is used to determine updates.

discrete space, subvolumes: Here each block in the lattice can contain multiple molecules. If individual molecules are modelled then this is a regular discrete space approach without state-based aggregation. If aggregation is used, then it is a regular patch-based population model. Subvolume approaches use the Reaction Diffusion Master Equation (RDME) [GMWM76, Isa08].

Pattern formation and stability is also important in biology. Turing's paper gave an initial insight into this process [Tur53] and later work has considered this further [OS71] both in terms of regular discrete space models that describe cells in a matrix and partial differential equations.

3.7.3 Epidemiology

Riley [Ril07] identifies four distinct approaches to disease spread modelling. In **patch-based transmission**, individuals within a patch have the same risk of infection determined by the current infection rate over all patches, together with distance-based likelihood of infection from other patches [vdD08]. **Distance transmission** models have pairwise transmission and the risk is determined by distance between pairs. **Multigroup transmission** considers households and other groups with group membership ensuring a higher risk of infection whereas **network transmission models** consider individuals and their interactions. Many of these techniques assume that individuals do not move between patches, and that the disease spreads within patches and between patches.

Patch-based or metapopulation models[10] are used extensively in modelling of epidemics both generally [AP02, AvdD03, vdD08, ADH$^+$05, Mol77] and for various specific diseases such as rabies [Mur03, Chapter 13] and measles [BG95, Gre92]. These models often focus on the basic reproduction number, R_0, which describes the average number of infecteds each infected generates.

Regular discrete space has been used to study the effect of vaccination of disease spread [RA97] and the investigation of pairwise approximation techniques [LD96]. Cholera models can include water-borne transmission as well as human contact [BCG$^+$09, GMB$^+$12, MBR$^+$12].

Arrigoni and Pugliese [AP02] investigate patch models when both the number of patches and the number of individuals increase, providing PDEs, where the nth PDE describes the change in the number of patches with n infecteds. Hence, a discrete space model can be made more fluid without moving to continuous space.

[10] The basic epidemiological model is called the compartment model [Bra08] which consists of a single population where each individual has a state (susceptible, infected, recovered, immune, for example). It has no spatial aspects. This should not be confused with compartment models in biology which are patch-based models.

3.7.4 Other

Chaintreau *et al* [CLBR09] develop a mean-field model of movement and data ageing using real data collected from cabs in the San Francisco Bay area. The model is expressed as a PDE (due to two variables, time and age of data), but the spatial aspect is treated discretely. This research provides some ideas for parameterisation when detailed GPS data is available.

Propagation of forest fires is investigated in the context of Multi-class Multi-type Markovian Agent Model (M^2MAM) [CGB+10]. The approach models individual agents in discrete space which may be regular and from this, a patch ODE model is derived.

Other domains in which irregular discrete space models have been developed by QUANTICOL project members include emergency egress, swarm robotics, delay tolerant networking and crowd movement. The modelling of emergency egress from a multi-story building included human factors characteristics of average speed of walking, size of doors and corridors and is irregular discrete space [MLB+12]. The robotics case study consists of a swarm of robots that have to collectively identify a shortest path [MBL+13]. ZebraNet is an delay-tolerant network for data collection. From a model of individual zebra movement in continuous space, a general discrete space model is developed based on ODEs [Fen14]. Investigation of emergence (in the mathematical sense) of spontaneous drinking parties in Spanish cities [RG03] shows that the introduction of small variations that break symmetry, both in space and in the degree of connectivity between locations, can lead to new behaviour [BLM13].

3.8 Application to smart transport case studies

In the QUANTICOL project, discrete space has been considered when modelling various case studies. Each case study will be given a short name (based on the process algebra or formalism in which they were developed or the site name where they were developed). These names will then be used in classifying the case studies spatially in the final section of this document (see Table 6 on page 28). All the models are **continuous time** models except for one.

3.8.1 Bus models

PEPA bus: This is an **irregular discrete space** model where buses can be treated individually or aggregated. The state of a bus is its current location. A bus route consists of a number of contiguous locations, each representing a portion of the city. To travel a route, a bus must interact with an agent representing each location. The number of agents for a location limit how many buses can be in it, and if no agent is available for a location, any bus wishing to traverse it must wait. Parameters are associated with the location agents and the model is **location inhomogeneous**. The goal of the model is to assess the impact of contention for physical space on the quality of service.

HYPE bus: This is a model of individual buses travelling through different regions of a city where the time that a bus spends in a region is determined by the length of the route as well as modifiers that speed up or slow down the bus. It is a **irregular discrete space** model that is **location inhomogeneous**. Buses are not allowed to proceed to the next region if they are ahead of schedule. The model is extended to determine how many buses should be scheduled in a period to ensure a regular bus service under different modifiers such as those representing peak traffic where speeds are slower and night traffic when speeds are faster.

CTMC route: This is a **regular discrete** one-dimensional **space** model of individual bus movement along a route. Each stage on the route has different parameters and hence the model is **transfer inhomogeneous**. A goal of the model is to assess how changes in speed by the driver impact quality of service. The model raises issues of **time-inhomogeneity** because the parameters could vary with time of day. An extension of this model may include height above sea level to enable modelling of hybrid buses (those that use both hydrocarbon fuels and electricity) and each stop could have an associated elevation that would determine parameters. This model is concerned with the regularity of a frequent bus service and hence spatial aspects can be abstracted. When modelling a less frequent bus service, a more fine-grained idea of space may be necessary and this is discussed in the section on continuous space.

3.8.2 Bike sharing models

Bike sharing has been a popular case study for the application of our existing and new formalisms, and hence there are multiple examples in this category. The models differ in their focus.

StocS bike [BFN+14]: This is an **irregular discrete space** model where neighbourhoods can be based on a distance or user-preference function as well as adjacency. Bikes are aggregated (discretely) and users are treated individually with two states: on foot or using a bike. Locations are collections of bike stations (giving aggregation of space) and for each location there is a count of the number of available bikes and available slots for returns. Parameters associated with the edges of the location graph can differ as they reflect distance, giving **transfer inhomogeneity** and are used to determine which stations individuals will use. The goal of the model is to see the effect of a local rebalancing mechanism on global performance of service delivery. This model has been expressed in different process algebra approaches, specifically PEPA-S, PALOMA and stochastic HYPE, and are described in Deliverable 4.1 [BFN+14].

PALOMA bike: There are two PALOMA models of bike sharing that have been developed. One is mentioned above and the second is now described. A set of locations is specified as in the Multi-class Multi-type Markovian Agent Model (M^2MAM) [CGB+10]. A perception function is used that does not give a regular pattern of neighbours, therefore it is an **irregular discrete space** model. Each bike station is a distinct location and bikes are aggregated by location (with a distinct abstract location representing that a bike is in use and not at a particular station). The level of occupancy of the vehicle responsible for redistributing bikes is also a variable in the model. The goal of the model is to understand user dissatisfaction (when they are unable to take a bike or return a bike) across different redistribution policies. The model is amenable to both a discrete aggregated state and a continuous aggregated state interpretation.

PEPA-S bike: This model is similar to the PALOMA one, since it is based on a spatial extension of PEPA integrating ideas from Markovian Agents (M^2MAM) [CGB+10]. It uses a perception function to encode a probability routing matrix, so it uses **irregular discrete space**. Each location corresponds to a bike station. There are two classes of agents: bike racks (with states full and empty) and users (with states pedestrian and biker). Users in both states move between locations as determined by the perception function. A truck can be modelled to capture redistribution of bikes. Dependency of the perception function on populations in each location can be exploited to define redistribution policies. The goal is to minimise user dissatisfaction.

HYPE1 bike: This is an **irregular discrete space** model based on a complete graph. Bikes are treated individually, and can either be at a specific station or in transit between two specific stations. The model has **transfer inhomogeneity** because the rate between stations reflects the distance between them. Additionally the rate for taking bikes from a station is specific to that station, hence it also has **location inhomogeneity**. The model also considers redistribution of bikes by a vehicle. The goal of the model is to consider how many stations are full or empty and how rebalancing can help reduce these numbers.

Bio-PEPA bike: In this model, regular space based on a grid is used, and an additional abstract location for when a bike is in use is required, meaning that this is an **irregular discrete space** model. The focus is user preferences for stations and how long adjustment of this preference takes when a preferred station cannot supply a bike, and is influenced by the foraging model of [MSH05]. Users are aggregated, and each location in the grid represents a number of bike stations with similar characteristics. Parameter homogeneity is not required hence the model may have **transfer inhomogeneity** and **location inhomogeneity**.

StoKlaim bike: This model has a grid representation of space but also an abstract location to represent when bikes are in use (the same as in the Bio-PEPA model above), hence it is not regular space. It is **parameter inhomogeneous** because movement is determined by a non-uniform probability distribution. It models bikes and users individually, and includes redistribution. The goal of this model is to investigate the effect of unbalanced redistribution of bikes.

CTMC bike: This model takes a queueing theory approach and models stations as servers. Bikes are aggregated at and between pairs of stations. All bike stations are connected giving a complete graph of locations. Only some stations are treated explicitly, and the transfer and location parameters for these are inhomogeneous. There are exogenous arrivals at each station which represent the effect of the other stations. The structure of the model has been determined in part by the real data that is available. Due to the **parameter inhomogeneity**, it is an **irregular discrete space model**. The goal here is to build a realistic model using the data that is available for a specific bike sharing scheme.

EPFL1 bike [FG12]: Bike sharing has been modelled with **homogeneous discrete space** in a continuous time setting. The goal is to understand the proportion of problematic stations (full or empty) for a fixed total number of bikes and the number of slots at each station. The model has **parameter homogeneity** and this regularity of parameters enables the model to be interpreted as a $M/M/1/K$ queue where K is the station capacity and using a parametric curve it is possible to determine the optimal fleet size. The model is further extended to a grid where a local search may be necessary to find a station with a slot for a return, or there is a choice of two neighbours for return. In this case, analytical solutions are not possible. Redistribution of bikes is also investigated.

EPFL2 bike [FGM12]: A different approach is taken to **spatial inhomogeneity** in [FGM12] where there are clusters of locations with the same parameters and mean-field limits are obtained for two models. One model only allows arrival at non-empty stations and returns at non-full stations.

DTMC bike [LLM13]: This model is a **discrete time, homogeneous discrete space** revisitation of the system presented in [FG12]. The model is a DTMC population model consisting of bike stations which have slots for bikes, and this number is fixed across all stations. In every time step, each station has the same probability that a user retrieves a bike. The probability that a bike is returned to a station depends on the number of bikes that are in circulation (i.e. not parked). As in [FG12], one goal is to study problematic stations (full or empty) and the analysis has been performed using the FlyFast on-the-fly fast mean-field model checker [MBC+14]; both global aspects of the system as well as properties of an individual station have been analyzed. Results are compatible with those in [FG12].

4 Continuous space modelling techniques

Continuous space is more straightforward to define than discrete space. We focus on two-dimensional space; however, both one- and three-dimensional space may be useful for modelling transport. Continuous space can either be the Euclidean plane extending infinitely in all directions or it can be a bounded connected (contiguous) subset of this plane. Points in the plane can be referred to by their coordinates $(x, y) \in \mathbb{R} \times \mathbb{R}$. As with discrete space, we can consider two cases.

4.1 Continuous space without state-based aggregation

In these models, we consider identifiable individuals. If I is an individual, then it has associated information, similar to the discrete state case.

- $\mathsf{loc}(I, t) \in \mathbb{R} \times \mathbb{R}$ which is its location at time t
- $\mathsf{state}(I, t) = A_i$ which is its state at time t

There are rules which describe how the individual changes state that may take into account the individual's current location, and rules that describe an individual's movement through space which may take into account the individual's state. In the networking literature, there are a number of movement models and these will be discussed later in this section. As with discrete space, the rates for state change can be functional and exponential. Unlike with discrete space, it is not useful to construct a Markov chain whose states are obtained from the locations and states of each individual.

In the case that the state is continuous, then

- $\mathsf{state}(I, t) = Y$ which is a continuous variable representing its state at time t.

As with the discrete space case, some way is required that describes the change of state over time, and an ODE can be used for this. Some models require both discrete and continuous non-aggregated states and this requires a hybrid solution.

A different approach to modelling continuous state with continuous time is that of continuous time Markov processes (CTMP) [DP03]. A CTMP is a tuple (S, Σ, R, L) where (S, Σ) forms a specific type of topological manifold and $R : S \times \Sigma \rightarrow \mathbb{R}_{\geq 0}$ is a rate function which is measurable in its first coordinate and a measure on its second coordinate. L is a state labelling function. Applying this in the context of space, the manifold is $(\mathbb{R} \times \mathbb{R}, \Sigma)$ where Σ consists of the open sets of $\mathbb{R} \times \mathbb{R}$, hence defining a σ-algebra. A notion of path through this space can be defined describing the behaviour of an individual. Furthermore, if there are additional continuous quantities associated with the individual then additional dimensions of $\mathbb{R}$ can be used.

4.2 Continuous space with state-based aggregation

When individuals are aggregated, there is no need to keep track of them individually and *densities* become more important. In spatio-temporal point processes[11], each point in space (x, y) has an associated integral count for a state in a population at a specific point in time t. We can denote this as $N_{A_i}((x, y), t)$ and its behaviour is described by a function $\lambda((x, y), t)$. In general, λ can depend on all preceding events, but in the case of a Poisson process, it only depends on (x, y) and t [SBG02].

For continuous aggregation of populations, we consider the classical model of continuous space, partial differential equations. For populations described by $X_{A_i}((x, y), t)$, the general form is

$$F_i\big(x, y, t, X_{A_1}, \ldots, X_{A_n}, \frac{\partial X_{A_i}}{\partial x}, \frac{\partial X_{A_i}}{\partial y}, \frac{\partial X_{A_i}}{\partial t}, \frac{\partial^2 X_{A_i}}{\partial x^2}, \frac{\partial^2 X_{A_i}}{\partial xy}, \frac{\partial^2 X_{A_i}}{\partial y^2}\big) = 0$$

if we assume that we are interested in second order partial derivatives over space only for the population $X_{A_i}((x, y), t)$. Note that writing the PDE in this form simply allows it to be described as a function over all the derivatives of interest rather than as a single partial derivative being equal to a function of other derivatives. There are various techniques for solving PDES, many of which involve discretising the area into a mesh [SSML03].

4.3 Examples of existing use

PDEs of reaction-diffusion type are very well understood in many disciplines, such as ecology [OL01], biology [Mur02], and chemistry [Van07]. We focus on a few articles to illustrate the techniques used.

[11] In contrast to spatio-temporal point processes, spatial point processes describe distributions in space, and do not include a notion of change over time [BBS07] and hence are unsuitable for our purposes.

4.3.1 Ecology

Spatio-temporal point processes have been used to model plant growth and dispersal [BP97, BP99]. The local density of a point is defined in terms of a competition kernel. Integro-differential moment equations using moment closure describe behaviour over time and involve average density, competition kernel and spatial autocovariance that measures association between two points.

Holmes *et al* [HLBV94] review the use of PDEs in ecological applications. In the least spatially heterogeneous case, the PDE expressing interaction and Brownian random motion where movement rate is independent of time and space, is defined as follows

$$\frac{\partial X_{A_i}((x,y),t)}{\partial t} = D\left(\frac{\partial^2 X_{A_i}}{\partial x^2} + \frac{\partial^2 X_{A_i}}{\partial y^2}\right) + f_i(X_{A_1},\ldots,X_{A_n}) = D \triangle X_{A_i} + f_i(X_{A_1},\ldots,X_{A_n})$$

where D is the *diffusion constant* and $\triangle f$ is the *Laplacian* of a function f. The functions f_i capture the interaction of the various populations at a specific point and time. A more general reaction-diffusion PDE has the form

$$\frac{\partial X_{A_i}}{\partial t} = \frac{\partial}{\partial x}\left(D(X_{A_i},(x,y))\frac{\partial X_{A_i}}{\partial x}\right) + \frac{\partial}{\partial y}\left(D(X_{A_i},(x,y))\frac{\partial X_{A_i}}{\partial y}\right) + f_i(X_{A_1},\ldots,X_{A_n}).$$

In this equation, the function D only depends on X_{A_i} and (x, y) but a function that also depends on $X_{A_1},\ldots,X_{A_n}$ could be used, thus including effects from other subpopulations.

4.3.2 Biology

Bittig and Uhrmacher [BU10] describe two continuous space approaches for cellular modelling.

particle space: In the case of *molecular* dynamics each particle is modelled individually and movement is done by rules, leading to some form of random walk. These models can be made more efficient by assuming that each particle is only affected by nearby events, so that the effect of other events can be excluded from consideration.

gradients/PDEs: These have been discussed above. Often in biomolecular modelling, only simple diffusion is required. The link between these models and those based on discrete regular space is an area of ongoing research.

Fange *et al* [FBSE10] describe techniques for spatially heterogeneous kinetics as *microscopic* when each individual particle is considered in terms of its position (continuous space), as *mesoscopic* when the Reaction Diffusion Master Equation (RDME) is used (discrete space) and as *macroscopic* when using PDEs (continuous space). The relationship between the RDME and PDEs can be established by means of a moment closure procedure, rather than a more rigorous limiting procedure in the sense of Kurtz [Kur81]. PDEs can also be obtained by taking the hydrodynamic limit of IPSs [DMP91]. Chemical reactions have also been modelled as PDEs in a series of papers [AT80, Kot86, Blo91, Blo93, Blo96]. In these models, local reactions model only birth and death, and there are no interactions between different types of agents. Random walks on regular space models involving individuals have been surveyed in [CPB08] and [OH02]. Starting from continuous space or regular discrete space, movement is described by PDEs in the limit. Restrictions on the walk give rise to different behaviours.

4.3.3 Epidemiology

Kendall [Ken65] proposed the first spatial epidemic model based on the Kermack-McKendrick non-spatial compartment model. A general discussion of PDEs in epidemiological modelling can be found in [Wu08] and specific examples in [LZ09, MZ08, MM91]. An aspect of interest in spatial disease modelling is generation of waves or fronts as disease moves through the population which is distributed in space. Murray describes a two-dimensional continuous space model that takes into account local fox densities and carrying capacities of an outbreak of rabies in the south of England and extends it to consider both natural immunity of foxes to rabies and size of breaks to curb the spread of the disease [Mur03]. A slightly different approach is taken in [BCG+09] where PDEs have diffusion terms that take into account the concentration of the cholera in the water.

4.3.4 Networking

There is a substantial amount of work on mobility models, both at the analytical level and experimentally through traces [CBD02]. Owing to its analytical tractability, the random walk (RW) model has been extensively studied in networking research. Unbiased RWs are proposed in [AHL96] and [ALLC00] to study movements across cellular networks, and to study routing protocols [IM06, GV06] and performance characteristics in ad hoc networks [GMPS04]. Direction based on an individual's state is a feature of a discrete time Markovian model developed for the comparison of update strategies in cellular networks [BNKS94]. RW is also used in [DT05] as the basic mobility model to obtain a deterministic reaction-diffusion type equation for information propagation in ad hoc networks, without considering it as a limit of a stochastic model. In [GL07], the PDEs interpreted as the deterministic limit of node concentrations, by appealing to the strong law of large numbers.

4.4 Application to smart transport case studies

To date the examples considered within the QUANTICOL project, as described in the previous sections, have focussed on discrete space except for the stochastic HYPE model of bike sharing (**HYPE2 bike**) that uses continuous space, and a possible continuous space extension to the StocS bike sharing model (**StocS bike extension**). In the stochastic HYPE model of bike sharing (which is presented in Deliverable 4.1 [BFN+14]), space is a bounded continuous region. Users have an initial (x, y) location and choose a random location to move to. If the distance to the location is sufficiently large, they will decide to go by bike rather than walk. In each case, their speed and time taken is determined by the actual distance between points and their method of travel.

We can also consider the data we are working with. The location-at-time data we have obtained from Lothian buses is spatially continuous, in that the data gathered from each bus at regular intervals can be translated to 2-dimensional map coordinates (and then projected onto a 1-dimensional representation of route). Since this is not aggregated data, but data for individual buses, an appropriate spatial approach would be an agent-based system which simulates the movement of each bus.

The data is discrete time as it is not reported continuously and there are different approaches to developing a continuous time, discrete space model. In the first, the 2-dimensional data can be projected onto a 1-dimensional route map, stops are identified and then fine division of the distance between stops is used to determine from the non-stop data how much time is spent in each division, and then a realistic model of bus movement can be derived [YGW+13]. This is a very fine-grained approach. In the second approach, which is coarser, the area of interest is divided into patches, and the GPS data is analysed to determine rates of movements from one patch to another. This can be done in a regular way, whether the area of interest is divided up into same sized patches, giving a regular discrete space model, or alternatively using geographical information, such as locations of traffic lights or bus stops to give a irregular discrete space model. See Deliverable 5.1 [TCG+14] for further details.

5 Topological space models

Topological, *distance*, and *closure* spaces are suitable as the foundations of a simple, general mathematical theory for representing and reasoning about various kinds of spatial models, Space is represented as a set of *points* equipped with some additional structure. Functions that operate on space, or spatially distributed entities (for example, the dynamics of a system) are constrained to preserve the structure. For a technical introduction to the subject, the report [CLM14] is available; furthermore, a brief account of *spatial logics*, can be found in Deliverable D3.1 [MBC+14].

Topological spaces. Topological spaces are best suited to represent continuous space. A topological space is a set X equipped with a set of subsets of X, called *open sets*, which are closed under union and finite intersection. Open sets play the role of constraints to functions, ensuring that *continuity* is well-defined. For example, a function representing movement of entities in continuous space must be "smooth" with respect to the open sets, ensuring that no discontinuous "jump" is possible. An example is the Euclidean plane $\mathbb{R}^2$ where the open sets are generated by union and finite intersection of *open balls*, namely "circles without borders".

Closure spaces. Closure spaces are better suited than topological spaces to model discrete spatial structures. A closure space is a set X equipped with a *closure operator* subject to some axioms. The idea of a closure operator is to add to a set the "least possible enlargement" of it. Closure operators have been studied and applied in the context of computer graphics and image processing. A closure operator can be derived from the open sets of a topological space, making topological spaces a subclass of closure spaces. Closure spaces are also induced by regular grids, and more generally, by graphs.

Metric and distance spaces. Metric spaces have a binary, symmetric *distance* operator obeying the *triangular inequality.* Distance spaces relax any or all of these constraints. Distance spaces and their variants can be used to describe physical distances and they can be generalised so that *costs* can be associated with pairs of points in the space. This aspect is orthogonal to the fact that a model may use continuous or discrete space. A classical example of a metric space is the Euclidean plane equipped with the Euclidean distance. When dealing with distances, one may sometimes resort to *pseudo-metrics* where two different points may have null distance.

The abstract spaces that we introduced have in common a remarkable definitional simplicity. Nevertheless, the basic axioms result in a rich theory, finding applications in several, apparently unrelated domains.

5.1 Examples of existing use

Topological and metric spaces are widely known and used throughout mathematics, physics, engineering, and computer science. As is typical with abstractions, it is not obvious to point at specific applications, since the abstractions are mostly used to ease the development of general mathematical theories which are then instantiated to concrete examples. Relating to the QUANTICOL project, topology and distance are used to provide models of *spatial logics*. Spatial logics are used to describe properties of space [APHvB07] and selected topics have been summarised in the technical report [CLM14], and in Deliverable D3.1 [MBC+14]. The advantage of using abstract spaces as models of a logic is the ability to give a uniform definition of logical operators, that encompasses a wide range of cases; for example, one may define a logic that describes equally well objects moving in planes, three-dimensional spaces, space-time, discrete regular grids and arbitrary graphs. Moreover, one is able to separate concerns with respect to the features of a logical language. In particular, the notion of distance is orthogonal to that of continuous movement and therefore, one may adopt models that can be viewed both as metric and as topological spaces. By adhering to the existing mathematical tradition, we expect to be able to reuse such knowledge, in order to provide a general verification framework for the QUANTICOL case studies, without reinventing theory.

5.2 Application to smart transport case studies

In the smart transport case studies it is fundamental to take into account a *map* providing information on how movement of entities in space is constrained. A map may be described topologically and metrically. Topologically, the various interesting regions of a map (streets, parking lots, rivers, etc.) may be described as predicates over the points of a topological space (or a closure space, in the discrete case). Principles of *continuity of change* (that exist even in the discrete case) permit one to derive interesting conclusions about given configurations of a model, for example, reachability of a parking lot from relevant points of interest. Typically, we expect that models in the smart transport case studies are discrete. In some cases, when considering the graph obtained from a map or topological map of bus and bike stations of a town, space is discrete. In some other cases, discreteness merely comes as an approximation of continuity. As a simple example, consider *smart-lock* applications in bike sharing. By using a digitally enhanced lock, customers may leave a bike at arbitrary (continuous) coordinates on the map.

Furthermore, continuity may allow reasoning about spatial abstraction in the sense that a complex map may be continuously transformed to a less complex one that still provides important information about the ability to find a path from one point to another. This may allow for the construction of a topological map from a map, as described in Section 2.

Distances may be added on top of this basic framework, with different meanings. One may consider just physical distance between points. More generally, a distance may represent a *cost*. For example, in the bike sharing scenario, it is relevant to know whether a point of interest is on top of a hill; the cost of moving a bike may be different for different stations, even though the physical distance is the same. In a bus network, urban and extra-urban rides may have different costs. Another example of a distance is the time needed to move between points, which may be only loosely related to physical distance (consider for example, road traffic and peak hours). Costs may be combined when considering integrated bus and bike sharing scenarios, giving rise to complex use cases that require some abstraction to be treated satisfactorily. We expect that abstract modelling of space will allow us to extrapolate the interesting common features among the variety of cases that we depicted, and to be able to derive a reasoning and automated verification framework that is suitable to derive useful predictions in the smart transport case studies.

6 Conclusion

6.1 The classification revisited

We can now apply the classification to the QUANTICOL case studies and this is shown in Figure 6. As can be seen, there is a fair distribution across the table. As mentioned previously, the techniques that fall into column two differ from the usual approaches that have been taken in prior research by the members of this project, as it involves considering individuals and continuous quantities. We now consider what could fill the empty positions in the table.

Discrete homogeneous space, no aggregation: An extension of the CTMC route model could see quantities such as power consumption associated with individual buses which would fit into this category. If individual buses also have a discrete state (for example, whether they are on time or not), then this could result in a continuous time, homogeneous space model without state aggregation where there are both discrete state and continuous state variables. This would be a hybrid model.

Discrete space, continuous state, no aggregation: Similarly to the case above, one could extend any of the models under discrete state to include some continuous value, and this extension would be a hybrid model.

<table>
<tr><td>TIME</td><td colspan="4">continuous or discrete</td></tr>
<tr><td>AGGR</td><td colspan="2">none</td><td colspan="2">state and/or space</td></tr>
<tr><td>STATE</td><td>discrete</td><td>continuous</td><td>discrete</td><td>continuous</td></tr>
<tr><td>SPACE</td><td></td><td></td><td></td><td></td></tr>
<tr><td>discrete</td><td></td><td></td><td></td><td></td></tr>
<tr><td>general</td><td>PEPA bus
HYPE bus
HYPE1 bike
StoKlaim bike</td><td></td><td colspan="2">PEPA bus
PALOMA bike
EPFL2 bike</td></tr>
<tr><td></td><td></td><td></td><td>StocS bike
PEPA-S bike
Bio-PEPA bike
CTMC bike</td><td></td></tr>
<tr><td>regular</td><td>CTMC route</td><td></td><td colspan="2">EPFL1 bike</td></tr>
<tr><td>homo-geneous</td><td></td><td></td><td colspan="2">EPFL1 bike
DTMC bike</td></tr>
<tr><td>continuous</td><td></td><td></td><td>HYPE2 bike
StocS bike
extension</td><td></td></tr>
</table>

Figure 6: Classification of QUANTICOL case studies in terms of time, aggregation, state and space

Continuous space, no aggregation: Only one slot is filled; however, as the project progresses, we may find that there are some models without aggregation for which continuous space is important for reasons of accuracy.

Continuous space, continuous-state aggregation: It is unclear whether PDEs will play a role in this project, as it seems unlikely that the kind of diffusion and dispersal represented by this technique has a role in modelling smart transport, unless we model passenger movement both on and off buses.

Looking at Figure 6, the focus has been on individual models with discrete space (with continuous or discrete time), and population models with discrete space, both discrete and continuous populations. This suggests that these techniques are the appropriate techniques for modelling smart transport, and that we should not be trying to construct models in other categories just because space modelling techniques exist for these categories. However, our choice of modelling techniques must also be determined by our modelling goals and the questions we wish to answer through modelling. At this stage of the project, we have a partial list of these goals or these questions. Hence, choice of modelling technique (within reasonable bounds) may evolve as the project proceeds.

6.2 Conclusions

The following points summarise the main conclusions about the mathematical techniques for spatial modelling that have been obtained from the literature and relating to the QUANTICOL case studies.

- The techniques in the literature seem to use well-understood theory of difference and differential equations, in particular, those differential equations that are easy to solve numerically. For the smart transport case studies in QUANTICOL, we may need to use more complex ODEs.

- There are various techniques in ecological and epidemiological modelling for constructing mean-field patch models and these may be appropriate or it may be possible to modify these approaches to model smart transport. A general issue with patch models is that the number of individuals per state per patch may become too low to obtain reasonable approximations.
- Meanfield techniques in ecological and epidemiological modelling focus on global behaviour rather than local behaviour. This may be appropriate for some QUANTICOL models such as global proportion of empty and full bike stations but less appropriate if there is a need to understand what is happening more locally, or to understand causality within the system.
- Techniques in ecological and epidemiological modelling for obtaining mean-field models often use particular features of the modelling scenario to construct useful approximations and aggregations, and it is unclear at this stage whether the QUANTICOL case studies have these type of properties, or whether new approaches will need be developed. A particular example of this is differing rates between and within patches.
- The methods for state aggregation in continuous space do not appear to match with the objectives of the QUANTICOL project. PDEs capture continuous aggregation in continuous space, and it is not clear whether this type of smooth approach can be applied to smart transport case studies. Spatio-temporal point processes are analysed in terms of ODEs for average density and average covariance which are both global measures and hence may not be appropriate for our case studies.
- Most techniques in the literature are time homogeneous, in the sense that subpopulation sizes at later time points depend on subpopulation sizes at earlier time points but they do not depend on time itself.
- Time inhomogeneity of space (space varying over time) does not appear to feature in the mathematical space modelling literature. In bus modelling, it seems likely that we will want to describe space as having different qualities depending on the time of day. There are techniques that allow one to avoid the construction of such models. This issue has also been identified as a challenge for bike modelling in Deliverable 5.1 [TCG+14].
- Whether working stochastically or deterministically, there appears to be very little literature that allows one to reason generally about the time and space dependencies of a model. This has implications for numerical solution of these ODEs and it may be important to understand our models and modelling techniques in terms of this dependence.
- The literature on spatial modelling does not consider hybrid modelling where both discrete and continuous approaches are used. Hybrid approaches may be appropriate when subpopulation sizes are not sufficient for a mean-field approach. Hybrid treatment of space could be important to understand local details and hybrid treatment of populations could be important to ensure that relevant stochasticity is not lost.

To conclude, discrete space modelling techniques have been used for the majority of QUANTICOL smart transport case studies to date and spatial heterogeneity is a core feature of these case studies. For bus modelling, individual models may be most important, and here some continuous space models may be necessary for accuracy of representation. For bike sharing models, patch-based discrete and continuous populations seem most relevant.

Preliminary guidelines: We suggest that patch models seem most suitable for QUANTICOL although there may be a role for individual-based continuous space models, as well as transformation from continuous space models to discrete space models. Hence we propose further investigation of mean-field models to understand which techniques are useful and also to possibly develop new approaches, consideration of spatial moment closure techniques for both global and local measures, and exploration of hybrid approaches for space modelling.

Relationships with other deliverables and work packages:

WP1 Emergent Behaviour and Adaptivity**:** The work in Work Package 1 on multiscale modelling [BGH+14] may be relevant for spatial modelling as we may wish to model at different spatial scales and the work on temporal scales may also be applicable. This may be particularly important when rates within patches operate on a different timescale to those between patches. Task 1.3 looks at language and analysis techniques and space has a role to play in this task as well. Furthermore, mean-field techniques are important when working with patch models.

WP3 Logic and Scalable Verification**:** Work Package 3 includes spatial logics for model checking techniques that will exploit mean-field approximation, and hence Deliverable 2.1 is relevant because it contributes to finding suitable spatial representations.

WP4 Language and Design Methodology**:** Deliverable 2.1 may indirectly influence Work Package 4 as it will be used to determine process algebraic features that will be described in the internal report of Task 2.2 that will then become part of the CAS-SCEL language.

WP5 Model Validation and Tool Support**:** Deliverable 2.1 influences WP5 as it provides guidelines for spatial modelling, and it is also influenced by the requirements identified for the case studies in the project [TCG+14].

Work plan for the second reporting period: The DOW describes three tasks in Work Package 2.

T2.1 Scalable representations of space (started, ends month 42)

T2.2 Stochastic process algebras for spatial aspects (starts month 25, ends month 48)

T2.3 Parameter and model fitting from spatial data (starts month 19, ends month 36)

The second phase of Task 2.1 will now begin and this involves preparing the internal report titled "A unified view of spatial representation and analysis techniques". In this phase, the ways in which the different spatial modelling techniques can be analysed will be investigated further, providing an understanding of the costs of various methods, and allowing for appropriate choices for modelling the QUANTICOL case studies. Additionally the items proposed above will be investigated.

Task 2.3 starts in October 2014 and will consider parameterisation of space for spatial models. Deliverable 2.1 and its associated technical report will provide a starting point for this research, as a number of papers referenced consider parameterisation. Deliverable 2.2 will document this task and is due in month 36 (end of March 2016) at the end of the task.

Looking further forward, Task 2.2 will start in April 2015, and its internal report is due in month 30 (end of September 2015). This report will be an investigation of space modelling in stochastic process algebras.

Acknowledgements

We thank members of the advisory board of the QUANTICOL project for their helpful comments.

References

[ADH+05] J. Arino, J.R. Davis, D. Hartley, R. Jordan, J.M. Miller, and P. van den Driessche. A multi-species epidemic model with spatial dynamics. *Mathematical Medicine and Biology*, 22(2):129–142, 2005.

[AHL96] I.F. Akyildiz, J.S.M. Ho, and Y.-B. Lin. Movement-based location update and selective paging for PCS networks. *IEEE/ACM Transactions on Networking*, 4(4):629–638, 1996.

[ALLC00] I.F. Akyildiz, Y.-B. Lin, W.-R. Lai, and R.-J. Chen. A new random walk model for PCS networks. *IEEE Journal on Selected Areas in Communications*, 18(7):1254–1260, 2000.

[AMR95] U.M. Ascher, R.M.M. Mattheij, and R.D. Russell. *Numerical solution of boundary value problems for ordinary differential equations*. SIAM, 1995.

[AP02] F. Arrigoni and A. Pugliese. Limits of a multi-patch SIS epidemic model. *Journal of Mathematical Biology*, 45(5):419–440, November 2002.

[APHvB07] M. Aiello, I. Pratt-Hartmann, and J. van Benthem, editors. *Handbook of Spatial Logics*. Springer, 2007.

[APS12] P. Auger, J.C. Poggiale, and E. Sánchez. A review on spatial aggregation methods involving several time scales. *Ecological Complexity*, 10:12–25, June 2012.

[AT80] L. Arnold and M. Theodosopulu. Deterministic limit of the stochastic model of chemical reactions with diffusion. *Advances in Applied Probability*, 12(2):367–379, 1980.

[AvdD03] J. Arino and P. van den Driessche. A multi-city epidemic model. *Mathematical Population Studies*, 10(3):175–193, 2003.

[BBS07] A. Baddeley, I. Bárány, and R. Schneider. Spatial point processes and their applications. *Stochastic Geometry: Lectures given at the CIME Summer School 2004*, page 175, 2007.

[BCG+09] E. Bertuzzo, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. On spatially explicit models of cholera epidemics. *Journal of The Royal Society Interface*, 7(43):321–333, 2009.

[Ber02] L. Berec. Techniques of spatially explicit individual-based models: construction, simulation, and mean-field analysis. *Ecological Modelling*, 150(1):5581, 2002.

[BFN+14] L. Bortolussi, C. Feng, R. De Nicola, J. Hillston, V. Galpin, D. Latella, M. Loreti, M. Massink, and V. Senni. CAS-SCEL language design. QUANTICOL Deliverable 4.1, March 2014.

[BG95] B. Bolker and B. Grenfell. Space, persistence and dynamics of measles epidemics. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 348(1325):309–320, 1995.

[BGH+14] L. Bortolussi, N. Gast, J. Hillston, R. Paškauskas, and M. Tribastone. Multiscale modelling informed by smart grids. QUANTICOL Deliverable 1.1, March 2014.

[BHLM13] L. Bortolussi, J. Hillston, D. Latella, and M. Massink. Continuous approximation of collective systems behaviour: a tutorial. *Performance Evaluation*, 70(5):317–349, 2013.

[BKHW05] C. Baier, J.-P. Katoen, H. Hermanns, and V. Wolf. Comparative branching-time semantics for markov chains. *Information and Computation*, 200(2):149–214, 2005.

[BLM13] L. Bortolussi, D. Latella, and M. Massink. Stochastic process algebra and stability analysis of collective systems. In R. De Nicola and C. Julien, editors, *15th International Conference on Coordination Models and Languages (COORDINATION 2013)*, LNCS 7890, pages 1–15. Springer, 2013.

[Blo91] D. Blount. Comparison of stochastic and deterministic models of a linear chemical reaction with diffusion. *The Annals of Probability*, 19(4):1440–1462, 1991.

[Blo93] D. Blount. Limit theorems for a sequence of nonlinear reaction-diffusion systems. *Stochastic Processes and their Applications*, 45(2):193–207, 1993.

[Blo96] D. Blount. Diffusion limits for a nonlinear density dependent space-time population model. *The Annals of Probability*, 24(2):639–659, 1996.

[BNKS94] A. Bar-Noy, I. Kessler, and M. Sidi. Mobile users: To update or not to update? In *INFOCOM 1994*, pages 570–576, 1994.

[BP97] B. Bolker and S.W. Pacala. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical population biology*, 52(3):179–197, 1997.

[BP99] B.M. Bolker and S.W. Pacala. Spatial moment equations for plant competition: Understanding spatial strategies and the advantages of short dispersal. *The American Naturalist*, 153(6):575–602, 1999.

[BP10] L. Bortolussi and A. Policriti. Hybrid dynamics of stochastic programs. *Theoretical Computer Science*, 411:2052–2077, 2010.

[Bra08] F. Brauer. Compartmental models in epidemiology. In L.J.S. Allen, F. Brauer, P. van den Driessche, and J. Wu, editors, *Mathematical Epidemiology*, pages 19–80. Springer, 2008.

[Bru67] S.G. Brush. History of the Lenz-Ising model. *Reviews of Modern Physics*, 39(4):883, 1967.

[BU10] A.T. Bittig and A.M. Uhrmacher. Spatial modeling in cell biology at multiple levels. In *Winter Simulation Conference (WSC 2010)*, pages 608–619. IEEE, 2010.

[CBD02] T. Camp, J. Boleng, and V. Davies. A survey of mobility models for ad hoc network research. *Wireless Communications and Mobile Computing*, 2(5):483–502, 2002.

[CFB09] H. Conceicao, M. Ferreira, and J. Barros. A cautionary view of mobility and connectivity modeling in vehicular ad-hoc networks. In *IEEE 69th Vehicular Technology Conference (VTC 2009)*, pages 1–5. IEEE, 2009.

[CGB+10] D. Cerotti, M. Gribaudo, A. Bobbio, C.T. Calafate, and P. Manzoni. A Markovian agent model for fire propagation in outdoor environments. In *Seventh European Performance Engineering Workshop (EPEW 2010)*, LNCS 6342, pages 131–146. Springer, 2010.

[Che81] P.L. Chesson. Models for spatially distributed populations: the effect of within-patch variability. *Theoretical Population Biology*, 19(3):288–325, 1981.

[Che12] P. Chesson. Scale transition theory: its aims, motivations and predictions. *Ecological Complexity*, 10:52–68, 2012.

[CHK+05] R. Chaturvedi, C. Huang, B. Kazmierczak, T. Schneider, J.A. Izaguirre, T. Glimm, H.G.E. Hentschel, J.A. Glazier, S.A. Newman, and M.S. Alber. On multiscale approaches to three-dimensional modelling of morphogenesis. *Journal of the Royal Society Interface*, 2(3):237–253, 2005.

[CL07] P.M. Chaikin and T.C. Lubensky. *Principles of Condensed Matter Physics*. Cambridge University Press, 4th edition, 2007.

[CLBR09] A. Chaintreau, J.-Y. Le Boudec, and N. Ristanovic. The age of gossip: spatial mean field regime. In *Eleventh International Joint Conference on Measurement and Modeling of Computer Systems (SIGMETRICS/Performance 2009)*, pages 109–120. ACM, 2009.

[CLM14] V. Ciancia, D. Latella, and M. Massink. Logics of space and time. Technical Report TR-QC-01-2014, QUANTICOL, 2014.

[CMRM07] R. Calegari, M. Musolesi, F. Raimondi, and C. Mascolo. CTG: a connectivity trace generator for testing the performance of opportunistic mobile systems. In *6th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering*, pages 415–424. ACM, 2007.

[CPB08] E.A Codling, M.J Plank, and S. Benhamou. Random walk models in biology. *Journal of the Royal Society Interface*, 5(25):813–834, 2008.

[Dav93] M.H.A. Davis. *Markov Models and Optimization*. Chapman & Hall, 1993.

[DL94a] R. Durrett and S. Levin. The importance of being discrete (and spatial). *Theoretical Population Biology*, 46(3):363–394, 1994.

[DL94b] R. Durrett and S.A. Levin. Stochastic spatial models: a user's guide to ecological applications. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 343(1305):329–350, 1994.

[DMP91] A. De Masi and E. Presutti. *Mathematical methods for hydrodynamic limits*. Lecture Notes in Mathematics. Springer-Verlag, 1991.

[DN08] R.W.R. Darling and J.R. Norris. Differential equation approximations for Markov chains. *Probability surveys*, 5:37–79, 2008.

[DP03] J. Desharnais and P. Panangaden. Continuous stochastic logic characterizes bisimulation of continuous-time markov processes. *The Journal of Logic and Algebraic Programming*, 56(1-2):99–115, 2003.

[DT05] M. Durvy and P. Thiran. Reaction-diffusion based transmission patterns for ad hoc networks. In *INFOCOM 2005*, volume 3, pages 2195–2205, 2005.

[EE04] J. Elf and M. Ehrenberg. Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Systems Biology*, 1(2):230–236, 2004.

[FBSE10] D. Fange, O.G. Berg, P. Sjöberg, and J. Elf. Stochastic reaction-diffusion kinetics in the microscopic limit. *Proceedings of the National Academy of Sciences*, 107(46):19820–19825, 2010.

[Fen14] C. Feng. Patch-based hybrid modelling of spatially distributed systems by using stochastic HYPE – ZebraNet as an example. In *Twelve International Workshop on Quantitative Aspects of Programming Languages and Systems (QAPL 2014)*, 2014.

[FG12] C. Fricker and N. Gast. Incentives and regulations in bike-sharing systems with stations of finite capacity. arXiv preprint arXiv:1201.1178, 2012.

[FGM12] C. Fricker, N. Gast, and H. Mohamed. Mean field analysis for inhomogeneous bike sharing systems. *DMTCS Proceedings*, 01:365–376, 2012.

[FS86] O. Frank and D. Strauss. Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842, 1986.

[Gar76] A. Gardiner. Homogeneous graphs. *Journal of Combinatorial Theory, Series B*, 20(1):94–102, 1976.

[GFH+14] V. Galpin, C. Feng, J. Hillston, M. Massink, M. Tribastone, and M. Tschaikowski. Review of time-based techniques for modelling space. Technical Report TR-QC-05-2014, QUANTICOL, 2014.

[GG92] F. Graner and J.A. Glazier. Simulation of biological cell sorting using a two-dimensional extended potts model. *Physical Review Letters*, 69(13):2013, 1992.

[Gil97] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1997.

[GL07] M. Garetto and E. Leonardi. Analysis of random mobility models with partial differential equations. *IEEE Transactions on Mobile Computing*, 6(11):1204–1217, 2007.

[GMB+12] M. Gatto, L. Mari, E. Bertuzzo, R. Casagrandi, L. Righetto, I. Rodriguez-Iturbe, and A. Rinaldo. Generalized reproduction numbers and the prediction of patterns in waterborne disease. *Proceedings of the National Academy of Sciences*, 109(48):19703–19708, 2012.

[GMPS04] A.El. Gamal, J. Mammen, B. Prabhakar, and D. Shah. Throughput-delay trade-off in wireless networks. In *INFOCOM 2004*, 2004.

[GMWM76] C.W. Gardiner, K.J. McNeil, D.F. Walls, and I.S. Matheson. Correlations in stochastic theories of chemical reactions. *Journal of Statistical Physics*, 14(4):307–331, 1976.

[Gre92] B.T. Grenfell. Chance and chaos in measles dynamics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54(2):383–398, 1992.

[GV06] M. Grossglauser and M. Vetterli. Locating mobile nodes with EASE: learning efficient routes from encounter histories alone. *IEEE/ACM Transactions on Networks*, 14(3):457–469, 2006.

[HCM91] M.P. Hassell, H.N. Comins, and R.M. May. Spatial structure and chaos in insect population dynamics. *Nature*, 353(6341):255–258, 1991.

[HLBV94] E.E. Holmes, M.A. Lewis, J.E. Banks, and R.R. Veit. Partial differential equations in ecology: spatial interactions and population dynamics. *Ecology*, 75(1):17–29, 1994.

[Ila01] A. Ilachinski. *Cellular automata: a discrete universe*. World Scientific, 2001.

[IM06] S. Ioannidis and P. Marbach. A Brownian motion model for last encounter routing. In *INFOCOM 2006*, 2006.

[Isa08] S.A. Isaacson. Relationship between the reaction–diffusion master equation and particle tracking models. *Journal of Physics A: Mathematical and Theoretical*, 41(6):065003, 2008.

[Kan98] K. Kaneko. Diversity, stability, and metadynamics: Remarks from coupled map studies. In J. Bascompte and R.V. Solé, editors, *Modeling spatiotemporal dynamics in ecology*, pages 27–45. Springer-Verlag, 1998.

[Ken65] D.G. Kendall. Mathematical models of the spread of infection. In *Mathematics and Computer Science in Biology and Medicine*, pages 213–225. Medical Research Council London, 1965.

[Kot86] P. Kotelenez. Law of large numbers and central limit theorem for linear chemical reactions with diffusion. *The Annals of Probability*, 14(1):173–193, 1986.

[KP07] T. Kathiravelu and A. Pears. Reproducing opportunistic connectivity traces using connectivity models. In *2007 ACM CoNEXT conference*, page 34. ACM, 2007.

[KS76] J.G. Kemeny and J.L. Snell. *Finite Markov chains: with a new appendix: Generalization of a fundamental matrix*. Springer, New York, 1976.

[Kur81] T.G. Kurtz. *Approximation of population processes*. SIAM, 1981.

[LD96] S.A. Levin and R. Durrett. From individuals to epidemics. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1347):1615–1621, 1996.

[Lev69] R. Levins. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the ESA*, 15(3):237–240, 1969.

[Lev74] S.A. Levin. Dispersion and population interactions. *American Naturalist*, 108(960):207–228, 1974.

[LLM13] D. Latella, M. Loreti, and M. Massink. On-the-fly PCTL fast mean-field model-checking for self-organising coordination – preliminary version. Technical Report TR-QC-01-2013, QUANTICOL, 2013.

[LZ09] J. Li and X. Zou. Modeling spatial spread of infectious diseases with a fixed latent period in a spatially continuous domain. *Bulletin of mathematical biology*, 71(8):20482079, 2009.

[MBC+14] M. Massink, L. Bortolussi, V. Ciancia, J. Hillston, A. Lluch-Lafuente, D. Latella, M. Loreti, D. Reijsbergen, and A. Vandin. Foundations of scalable verification for stochastic logics. QUANTICOL Deliverable 3.1, March 2014.

[MBL+13] M. Massink, M. Brambilla, D. Latella, M. Dorigo, and M. Birattari. On the use of Bio-PEPA for modelling and analysing collective behaviors in swarm intelligence. *Swarm Intelligence*, 7(2-3):201–228, 2013.

[MBR+12] L. Mari, E. Bertuzzo, L. Righetto, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. Modelling cholera epidemics: the role of waterways, human mobility and sanitation. *Journal of The Royal Society Interface*, 9(67):376–388, 2012.

[MLB+12] M. Massink, D. Latella, A. Bracciali, M.D. Harrison, and J. Hillston. Scalable context-dependent analysis of emergency egress models. *Formal Aspects of Computing*, 24(2):267–302, 2012.

[MM91] D.B. Meade and F.A. Milner. An SIR model for epidemics with diffusion to avoid infection and overcrowding. In *13th IMACS World Congress on Computation and Applied Mathematics*, volume 3, pages 1444–1445, 1991.

[MMRL02] G. Marion, X. Mao, E. Renshaw, and J. Liu. Spatial heterogeneity and the stability of reaction states in autocatalysis. *Physical Review E*, 66(5):051915, November 2002.

[MNS11] C. McCaig, R. Norman, and C. Shankland. From individuals to populations: a mean field semantics for process algebra. *Theoretical Computer Science*, 412(17):1557–1580, 2011.

[Mol77] D. Mollison. Spatial contact models for ecological and epidemic spread. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(3):283–326, 1977.

[MP12] A. Morozov and J.-C. Poggiale. From spatially explicit ecological models to mean-field dynamics: the state of the art and perspectives. *Ecological Complexity*, 10:1–11, 2012.

[MSH05] G. Marion, D.L. Swain, and M.R. Hutchings. Understanding foraging behaviour in spatially heterogeneous environments. *Journal of Theoretical Biology*, 232(1):127–142, 2005.

[Mur02] J.D. Murray. *Mathematical Biology I: An Introduction*. Springer, 3rd edition, 2002.

[Mur03] J.D. Murray. *Mathematical biology II: Spatial Models and Biomedical Applications*. Springer, New York, 2003.

[MVZ05] M. Mamei, M. Vasirani, and F. Zambonelli. Self-organizing spatial shapes in mobile particles: the TOTA approach. In *Engineering Self-Organising Systems*, pages 138–153. Springer, 2005.

[MZ08] F.A. Milner and R. Zhao. SIR model with directed spatial diffusion. *Mathematical Population Studies*, 15(3):160–181, 2008.

[Nor98] J.R. Norris. *Markov chains*. Cambridge University Press, 1998.

[OH02] H.G. Othmer and T. Hillen. The diffusion limit of transport equations II: chemotaxis equations. *SIAM Journal on Applied Mathematics*, 62(4):1222–1250, 2002.

[OL01] A. Okubo and S. A. Levin. *Diffusion and ecological problems: modern perspectives*. Springer, 2001.

[OS71] H.G. Othmer and L.E. Scriven. Instability and dynamic pattern in cellular networks. *Journal of theoretical biology*, 32(3):507–537, 1971.

[Pan09] P. Panangaden. *Labelled Markov Processes*. Imperial College Press, 2009.

[PRL11] M. Pascual, M. Roy, and K. Laneri. Simple models for complex systems: exploiting the relationship between local and global densities. *Theoretical Ecology*, 4(2):211–222, 2011.

[RA97] C.J. Rhodes and R.M. Anderson. Epidemic thresholds and vaccination in a lattice model of disease spread. *Theoretical Population Biology*, 52(2):101–118, 1997.

[RG03] J. E. Rowe and R. Gomez. El Botellón: Modeling the movement of crowds in a city. *Complex Systems*, 14:363–370, 2003.

[Ril07] S. Riley. Large-scale spatial-transmission models of infectious disease. *Science*, 316(5829):1298–1301, 2007.

[SBG02] F.P. Schoenberg, D.R. Brillinger, and P. Guttorp. Point processes, spatial–temporal. In *Encyclopedia of environmetrics*, volume 3, pages 1573–1577. Wiley Online Library, 2002.

[SSML03] B.M. Slepchenko, J.C. Schaff, I. Macara, and L.M. Loew. Quantitative cell biology with the virtual cell. *Trends in cell biology*, 13(11):570–576, 2003.

[Tak96] Y. Takeuchi. *Global dynamical properties of Lotka-Volterra systems*. World Scientific, 1996.

[TCG+14] M. Tribastone, A. Clark, N. Gast, S. Gilmore, and D. Reijsbergen. Data validation and requirements for case studies. QUANTICOL Deliverable 5.1, March 2014.

[TGH12] M. Tribastone, S. Gilmore, and J. Hillston. Scalable Differential Analysis of Process Algebra Models. *IEEE Transactions on Software Engineering*, 38(1):205–219, 2012.

[TT13] M. Tschaikowski and M. Tribastone. Spatial fluid limits for stochastic mobile networks. *arXiv preprint arXiv:1307.4566*, 2013.

[Tur53] A.M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society (Part B)*, 237:37–72, 1953.

[Van07] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, 3rd edition, 2007.

[VBU13] M. Viroli, J. Beal, and K. Usbeck. Operational semantics of proto. *Science of Computer Programming*, 78(6):633–656, 2013.

[vdD08] P. van den Driessche. Spatial structure: Patch models. In L.J.S. Allen, F. Brauer, P. van den Driessche, and J. Wu, editors, *Mathematical Epidemiology*, pages 179–190. Springer, 2008.

[Wik13a] Wikipedia. Lattice graph, 2013. Last accessed: 4 Feb 2013, `http://en.wikipedia.org/wiki/Lattice_graph`.

[Wik13b] Wikipedia. Topological map, 2013. Last accessed: 4 Feb 2013, `http://en.wikipedia.org/wiki/Topological_map`.

[WKB07] S.D. Webb, M.J. Keeling, and M. Boots. Host-parasite interactions between the local and the mean-field: how and when does spatial population structure matter? *Journal of Theoretical Biology*, 249(1):140–152, 2007.

[WL95] J. Wu and O.L. Loucks. From balance of nature to hierarchical patch dynamics: a paradigm shift in ecology. *The Quarterly Review of Biology*, 70(4):439–466, 1995.

[Wu08] J. Wu. Spatial structure: partial differential equations models. In L.J.S. Allen, F. Brauer, P. van den Driessche, and J. Wu, editors, *Mathematical Epidemiology*, pages 191–204. Springer, 2008.

[XFAS06] D. Xu, Z. Feng, L.J.S. Allen, and R.K. Swihart. A spatially structured metapopulation model with patch dynamics. *Journal of Theoretical Biology*, 239(4):469–481, 2006.

[YGW+13] Y. Yang, D. Gerstle, P. Widhalm, D. Bauer, and M.C. Gonzalez. The potential of low-frequency AVL data for the monitoring and control of bus performance. In *Transportation Research Board 92nd Annual Meeting*, pages 13–4011, 2013.

Appendix: Discrete and continuous time Markov chains

This section briefly introduces these concepts, as they would be used in stochastic modelling both without aggregation of state and with aggregation of state (population-based Markov chains) [BKHW05, BHLM13].

Definition 1. *A* discrete time Markov chain (DTMC) *is a tuple* $\mathcal{M}_D = (\mathcal{S}, \mathbf{P})$ *where*

- $\mathcal{S}$ *is a finite set of* states*, and*
- $\mathbf{P} : \mathcal{S} \times \mathcal{S} \to [0, 1]$ *is a* probability matrix *satisfying* $\sum_{S' \in \mathcal{S}} \mathbf{P}(S, S') = 1$ *for all* $S \in \mathcal{S}$*.*

A DTMC is *time-abstract* [BKHW05] in the sense that time is viewed as a sequence of discrete steps or clock ticks. It describes behaviour as follows: if an entity or individual is currently in state $S \in \mathcal{S}$ then the probability of the entity being in state S' at the next time step is defined by $\mathbf{P}(S, S')$. Under certain conditions, the steady state of the DTMC can be determined and this describes when the DTMC is at equilibrium and gives the (unchanging) probability of being in any of the states of $\mathcal{S}$. By contrast, transient state probabilities can be determined at each point in time before steady state is achieved.

Definition 2. *A* continuous time Markov chain (CTMC) *is a tuple* $\mathcal{M}_C = (\mathcal{S}, \mathbf{R})$ *where*

- $\mathcal{S}$ *is a finite set of* states*, and*
- $\mathbf{R} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}_{\geq 0}$ *is a* rate matrix*.*

CTMCs are *time-aware* [BKHW05] since they use continuous time. If an entity is currently in state S, then $\mathbf{R}(S, S')$ is a non-negative number that defines an exponential distribution from which the duration of the time taken to transition from state S to state S' can be drawn. As with DTMCs and under certain conditions, transient and steady state probabilities can be calculated which describe the probability of being in each state at a particular time t or in the long run, respectively.

Let $E(S) = \sum_{S' \in \mathcal{S}} \mathbf{R}(S, S')$ be the exit rate of state S'. Then the embedded DTMC of a CTMC has entries in its probability matrix of the form $\mathbf{P}(S, S') = \mathbf{R}(S, S')/E(S)$ if $E(S) > 0$ and $\mathbf{P}(S, S') = 0$ otherwise. DTMCs and CTMCs can be state-labelled (usually with propositions) or transition-labelled (usually with actions). The research in QUANTICOL focusses on transition-labelled Markov chains. We next consider population Markov chains, both discrete time and continuous time. Instead of considering an entity with states, we now consider a vector of counts $\mathbf{X}$ that describes how many entities are in each state; thus it is a population view rather than an individual view.

Definition 3. *A* population discrete time Markov chain (PDTMC) *is a tuple* $\mathcal{X}_D = (\mathbf{X}, \mathcal{D}, \mathcal{T})$ *where*

- $\mathbf{X} = (X_1, \ldots, X_n)$ *is a vector of* variables
- $\mathcal{D}$ *is a countable set of* states *defined as* $\mathcal{D} = \mathcal{D}_1 \times \ldots \times \mathcal{D}_n$ *where each* $\mathcal{D}_i \subseteq \mathbb{N}$ *represents the domain of* X_i
- $\mathcal{T} = \{\tau_1, \ldots \tau_m\}$ *is the set of* transitions *of the form* $\tau_j = (\mathbf{v}, p)$ *where*
  - $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{N}^n$ *is the* state change or update vector *where* v_i *describes the change in number of units of* X_i *caused by transition* τ_j
  - $p : \mathcal{D} \to \mathbb{R}_{\geq 0}$ *is the* probability function *of transition* τ_j *that defines a sub-probability distribution, namely* $\sum_{\tau \in \mathcal{T}} p_\tau(\mathbf{d}) \leq 1$ *for all* $\mathbf{d} \in \mathcal{D}$*, such that* $p(\mathbf{d}) = 0$ *whenever* $\mathbf{d} + \mathbf{v} \notin \mathcal{D}$

Definition 4. *A* population continuous time Markov chain (PCTMC) *is a tuple* $\mathcal{X}_C = (\mathbf{X}, \mathcal{D}, \mathcal{T})$ *where*

- $\mathbf{X}$ *and* $\mathcal{D}$ *are defined as in the previous definition,*
- $\mathcal{T} = \{\tau_1, \ldots \tau_m\}$ *is the set of transitions of the form* $\tau_j = (\mathbf{v}, r)$ *where*
  - $\mathbf{v}$ *is defined as in the previous definition,*
  - $r : \mathcal{D} \to \mathbb{R}_{\geq 0}$ *is the* rate function *of transition* τ_j *with* $r(\mathbf{d}) = 0$ *whenever* $\mathbf{d} + \mathbf{v} \notin \mathcal{D}$*.*

In both types of population Markov chain, the associated Markov chain can be obtained. In both cases, $\mathcal{D}$ is the state space $\mathcal{S}$. For the population DTMC, the probability matrix of its associated DTMC is defined as

$$\mathbf{P}(\mathbf{d}, \mathbf{d}') = \sum_{\tau \in \mathcal{T}, \mathbf{v}_\tau = \mathbf{d}' - \mathbf{d}} p_\tau(\mathbf{d}) \text{ whenever } \mathbf{d} \neq \mathbf{d}'$$

and since probability functions define sub-probabilities then the rest of the probability mass must be accounted for by defining

$$\mathbf{P}(\mathbf{d}, \mathbf{d}) = 1 - \sum_{\tau \in \mathcal{T}, \mathbf{v}_\tau \neq 0} p_\tau(\mathbf{d}).$$

For the population CTMC, the rate matrix of its associated CTMC is

$$\mathbf{R}(\mathbf{d}, \mathbf{d}') = \sum_{\tau \in \mathcal{T}, \mathbf{v}_\tau = \mathbf{d}' - \mathbf{d}} r_\tau(\mathbf{d}) \text{ whenever } \mathbf{d} \neq \mathbf{d}'$$

and if the summation is empty, then $\mathbf{R}(\mathbf{d}, \mathbf{d}') = 0$.

As the size of the population increases, it has been shown [Kur81] under specific conditions that cover a large range of models that the behaviour of an (appropriately normalised) population CTMC at time t is very close to the solution of a set of ODEs, expressed in the form $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))$ defining a trajectory over time. The ODEs can be expressed in terms of a single vector ODE as

$$\dot{\mathbf{X}} = \frac{d\mathbf{X}}{dt} = \mathbf{f}(\mathbf{X}).$$