

# Specifying and Monitoring Properties of Stochastic Spatio-Temporal Systems in Signal Temporal Logic\*

Luca Bortolussi  
Modelling and Simulation Group  
University of Saarland, Germany  
DMG, University of Trieste, Italy  
ISTI-CNR, Pisa, Italy  
luca@dmi.units.it

Laura Nenzi  
IMT, Lucca, Italy  
laura.nenzi@imtlucca.it

## ABSTRACT

We present an extension of the linear time, time-bounded, *Signal Temporal Logic* to describe spatio-temporal properties. We consider a discrete location/patch-based representation of space, with a population of interacting agents evolving in each location and with agents migrating from one patch to another one. We provide both a boolean and a quantitative semantics to this logic. We then present *monitoring algorithms* to check the validity of a formula, or to compute its satisfaction (robustness) score, over a spatio-temporal trace, exploiting these routines to do *statistical model checking* of stochastic models. We illustrate the logic at work on an epidemic example, looking at the diffusion of a cholera infection among communities living along a river.

**Keywords:** Monitoring, Spatial Logic, Signal Temporal Logic, Spatio-Temporal Modelling.

## 1. INTRODUCTION

Many of the scientific and technological challenges we are currently facing deal with the complexity arising from the non-linear interactions between heterogeneous components. Examples span from biological systems, from bacteria up in the ladder of life, to ecological interactions, to socio-technical and cyberphysical systems, in which digital sensors and devices interact with the natural and social environment. Our ability to understand, control and design such systems requires refined mathematical and computational modelling tools.

Formal methods offer a powerful framework in this sense, providing modelling languages, requirements and properties specification languages, and efficient analysis routines/algorithms. In particular, logical methods based on *temporal logics* have proved to be very effective both in studying properties of existing systems, like bi-

\*Work partially funded by the EU-FET project QUANTICOL (nr. 600708), by the German Research Council (DFG) as part of the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University and the Transregional Collaborative Research Center SFB/TR 14 AVACS.

ological ones [24], and in the design of new systems [5]. This is due to their ability to easily express complex temporal behavioural patterns, and the availability of efficient model checking and monitoring tools, for many classes of mathematical models, ranging from ODEs [28] to stochastic processes [4, 5].

The literature of applications of formal methods to complex systems, of temporal logic in particular, has been mainly focused on well-mixed systems, where the spatial nature of interactions is abstracted for the sake of simplicity. However, *spatial information* is often crucial to properly understand their emerging dynamics. For instance, in bacteria chemotaxis, the motility of cells is realised by the spatial polarisation of small-G-protease signalling proteins [26]. Epidemic spreading at the national or worldwide scale depends crucially on the asymmetric/inhomogeneous population distribution and on the mobility patterns and available air routes [32]. The design and deployment of a bike sharing system cannot neglect the spatial distribution of bike stations, reflecting into different time-inhomogeneous request trends [21]. All these systems have inherent spatial aspects: they involve a large number of heterogeneous spatial entities that are located and can move on a physical space.

While from the point of view of modelling languages and simulation, some relevant work in spatial modelling has been done in recent years [10,20], the use of modal logic to describe and especially validate spatio-temporal properties is much less developed. While spatial and spatio-temporal logics have been proposed and theoretically investigated in literature [1], model checking routines have a much more recent history. Relevant works are those on spatial logics for process algebra with locations as [12,30], or spatial logic for rewrite theories [3]. Other relevant literature on purely spatial logic is [13,23]. In particular, we are not aware of any linear-time spatio-temporal logic with monitoring or model checking routines for checking properties of spatially located differential equations (PDE or patch ODEs) or of spatially located stochastic processes (patch CTMC).

Here we begin tackling this problem from a practical perspective. We start by introducing *Signal Spatio-Temporal Logic* (SSTL), a spatial extension of *Signal Temporal Logic* (STL) [15, 28]. STL is a temporal logic suitable to specify behaviours of real-valued time series generated during the simulation of a (stochastic) dynamical system. It extends the dense-time semantics of Metric Interval Temporal Logic [2] (MITL), parameterising it by a set of numerical predicates playing the role of atomic propositions. A special characteristic of STL is the definition of a quantitative semantics, allowing us to quantify the satisfiability of a property, i.e. to give a

value expressing how robustly a property is satisfied.

We extend STL by introducing simple spatial operators, inspired from the modal spatial operators of the *Multiprocess Network Logic* [31]. They permit to describe properties such as “there exists a location  $\ell_j$  at a certain distance from location  $\ell_i$  where the property  $\varphi$  is satisfied”, or “all the locations with a distance no more than  $d$  from  $\ell$  satisfy the property  $\varphi$ ”. The semantics of such spatial operators is strictly related to the description of space in which the dynamics takes place. Space can be logical, i.e. a set of locations, it can be discrete, i.e. a grid or a more general graph, or it can be continuous, for instance the euclidian space. The first option that we decided to investigate are systems with a discrete space, structured as a weighted graph, called *patch-based quantitative population models*. These models describe the number of entities in each node, with agents located in nodes and moving from a node to another one.<sup>1</sup>

The contribution of this paper are: (1) a formal description of patch-based population models; (2) the definition of SSTL, introducing new spatial operators, giving in detail the boolean and quantitative semantics, extending the monitoring algorithms to compute the satisfaction over boolean or real-valued signals, and providing also a stochastic version of this logic, leveraging the monitoring algorithms within a stochastic model checking routine; and (3) a discussion of the logic at work on a model of the spreading of cholera, a waterborne disease, considering the effect of rivers in the diffusion of the epidemics [8, 29].

The paper is structured as follows: in Section 2 we introduce the population model, while in section 3 we define the patch-based population model. In Section 4 we describe the syntax and the semantics of the Signal Spatio-Temporal Logic (SSTL), sketching the monitoring algorithms in Section 5. In Section 6 we present the stochastic extension of SSTL and in Section 7 we discuss the case study and some experimental results. Finally, the conclusions and future works are drawn in Section 8.

## 2. POPULATION MODELS

A *population model* intuitively is a system in which a large number of different agents or components interact together and take, through local transitions, a number of different states. The transitions can be seen as descriptions of events changing the global state of the system. There are many example of population processes, like social systems, biochemical networks, ecological systems and computer networks.

**DEFINITION 2.1 (Population model).** A *population model*  $\mathcal{M}$  is a tuple  $\mathcal{M} = (\mathbb{S}, \mathbf{X}, \mathcal{T})$ , where:

- $\mathbb{S} = \{1, \dots, n\}$  is the set of states of the agents in the population.
- $\mathbf{X} = (X_1, \dots, X_n)$  is the state vector, describing the state of the population model. The variable  $X_i \in \mathbb{R}_{\geq 0}$  represents the density, concentration, or number (in which case it takes integer values) of components in the  $i^{\text{th}}$  state. The domain of  $\mathbf{X}$ , i.e. the state space of the system, is a subset of  $\mathbb{R}^n$  and is denoted by  $\mathbb{D}$ .
- $\mathcal{T} = \{\tau_1, \dots, \tau_m\}$  is the set of global transitions of the form  $\tau_l = (a_l, \mathbf{v}_l, f_l)$  where
  - $a_l$  is the label of the transition,
  - $\mathbf{v}_l \in \mathbb{R}^n$  is the update vector, giving the net change of each counting variable due to the transition,

<sup>1</sup>The term patch is borrowed from ecological modelling.

- $f : \mathbb{D} \rightarrow \mathbb{R}_{\geq 0}$  is the rate function, giving the rate of the transition as a function of the global state of the system.

Each transition  $\tau_l$  can be seen as a rule of the form

$$q_1 s_{i_1} + \dots + q_k s_{i_k} \rightarrow r_1 s_{j_1} + \dots + r_h s_{i_h}$$

where  $s_{i_a}, s_{j_b} \in \mathbb{S}$  are states of the system, and  $q_i, r_j$  are the stoichiometric coefficients, i.e. the amount of components/entities consumed or produced by the transition. The update vector  $\mathbf{v}_l$  condenses the stoichiometric information as  $\mathbf{v}_l = \sum_{b \leq h} r_b \vec{e}_{j_b} - \sum_{a \leq k} q_a \vec{e}_{i_a}$ , where  $\vec{e}_j$  equals one in position  $j$  and zero elsewhere.

The dynamical evolution of these models can be described in different ways: we can interpret them stochastically as a *Markov chains* or deterministically as a system of Ordinary Differential Equations (ODEs).

**Stochastic dynamics.** The most important class of stochastic processes we will consider are Continuous Time Markov Chains (CTMC) [17] that describe population processes (PCTMC). A PCTMC can be represented as a time-indexed family of vectors:

$$(\mathbf{X}(t))_{t \in \mathbb{T}} = (X_1(t), \dots, X_n(t))_{t \in \mathbb{T}}$$

where the vector  $\mathbf{X}(t)$  corresponds to the state of the system at time  $t \in \mathbb{T}$ , with  $\mathbb{T} \subseteq \mathbb{R}_{\geq 0}$ , with each variable  $X_i$  counting the number of entities in the  $i^{\text{th}}$  state at time  $t$ . We assume  $\mathbf{X}(t) \in \mathbb{D}$ , where  $\mathbb{D} \subseteq \mathbb{N}^n$  is the state space of the system.

Given a population model  $\mathcal{M} = (\mathbb{S}, \mathbf{X}, \mathcal{T})$  from the set of transitions  $\mathcal{T}$ , we can easily derive the formal representation of the PCTMC [11] in terms of its infinitesimal generator matrix as:

$$\mathbf{Q}(\mathbf{d}_i, \mathbf{d}_j) = \sum_{\tau \in \mathcal{T} | \mathbf{v}_\tau = \mathbf{d}_j - \mathbf{d}_i} f_\tau(\mathbf{d}_i), \quad i \neq j \quad (1)$$

For each  $\mathbf{d}_i, \mathbf{d}_j \in \mathbb{D}$ , with  $i \neq j$ ,  $\mathbf{Q}(\mathbf{d}_i, \mathbf{d}_j)$  represents the rate of an exponential distribution, namely the distribution of a random variable modelling the time needed to go from state  $\mathbf{d}_i$  to state  $\mathbf{d}_j$ . The diagonal elements of the matrix are equal to  $\mathbf{Q}(\mathbf{d}_i, \mathbf{d}_j) = -\sum_{i \neq j} \mathbf{Q}(\mathbf{d}_i, \mathbf{d}_j)$  and represent the opposite of the *exit* rate from the state  $\mathbf{d}_i$ . For all  $i, j$  s.t.  $\nexists \mathbf{v}_\tau = \mathbf{d}_i - \mathbf{d}_j$ ,  $\mathbf{Q}(\mathbf{d}_i, \mathbf{d}_j) = 0$ .

**Deterministic dynamics.** Due to state space explosion, population models are often intractable when interpreted as CTMC, but they can frequently be approximated by a set of non-linear Ordinary Differential Equations (ODEs). This is the so called fluid approximation [11]. The fluid ODEs can be seen as a macroscopic description of the microscopic CTMC dynamics, and, when populations are large, it is a good description of the system behaviour [11].

From a Population model  $\mathcal{M} = (\mathbb{S}, \mathbf{X}, \mathcal{T})$ , we can construct a set of ODEs, assuming variables  $\mathbf{X}$  to be continuous and interpreting each rate as a flow, thus obtaining the vector field

$$F(\mathbf{X}) = \sum_{\tau \in \mathcal{T}} \mathbf{v}_\tau f_\tau(\mathbf{X}), \quad (2)$$

defining the ODEs  $\frac{d\mathbf{X}}{dt} = F(\mathbf{X})$ . These equations, called also *rate equations*, can be shown to be a first order approximation of the average transient behaviour of the PCTMC, and, under a suitable rescaling of the variables (dividing by the system size, often the total population), the convergence of the PCTMC to the solution of the ODEs (see [11]) as populations and system size go to infinity can be proved.

**EXAMPLE 2.1 (SIR MODEL).** We will illustrate the previous definition by a simple epidemic scenario involving  $N$  individuals. Each individual can be in three different states: susceptible to infection ( $S$ ), infected ( $I$ ), and recovered and immune to infection ( $R$ ). The set of agent's states is then  $\mathbb{S} = \{S, I, R\}$  and the counting variables are  $\mathbf{X} = (X_S, X_I, X_R)$ , where  $X_i \in \{0, \dots, N\}$ . The state space  $\mathbb{D}$  of the system is a subset of  $[0, N]^3$ . There are 3 different transitions:

- \* a susceptible individual can be infected by getting in contact with an infected individual, *inf*:  $I + S \rightarrow 2I$  with  $v_{inf} = (-1, 1, 0)$  and rate function  $f_{inf}(\mathbf{X}) = k_{inf} \cdot X_S \cdot X_I$ ,
- \* an infected individual can recover, *rec*:  $I \rightarrow R$  with  $v_{rec} = (0, -1, 1)$  and rate function  $f_{rec}(\mathbf{X}) = k_{rec} \cdot X_I$ ,
- \* a recovered individual can lose its immunity, *loss*:  $R \rightarrow S$  with  $v_{loss} = (1, 0, -1)$  and rate function  $f_{loss}(\mathbf{X}) = k_{loss} \cdot X_R$ .

The population model is then  $\mathcal{M} = (\{S, I, R\}, \mathbf{X}, \mathcal{T})$  with  $\mathcal{T} = \{(inf, v_{inf}, f_{inf}), (rec, v_{rec}, f_{rec}), (loss, v_{loss}, f_{loss})\}$ . The stochastic process for this model can be represented by the family of vectors  $(\mathbf{X}(t))_{t \in \mathbb{R}_{\geq 0}} = (X_S(t), X_I(t), X_R(t))_{t \in \mathbb{R}_{\geq 0}}$  where each  $X_i(t) \in \mathbb{N}$  counts the number of entities in the state  $i$  at time  $t$ . The fluid approximation of this model, instead, is obtained applying (2) and assuming  $\mathbf{X} \in \mathbb{R}^3$ , and it corresponds to the system of ODEs:

$$\begin{cases} \frac{dX_S}{dt} = -X_I \cdot X_S \cdot k_{inf} + X_R k_{loss}, \\ \frac{dX_I}{dt} = X_I \cdot X_S \cdot k_{inf} - X_I k_{rec}, \\ \frac{dX_R}{dt} = X_I k_{rec} - X_R \cdot k_{loss}. \end{cases}$$

### 3. PATCH-BASED POPULATION MODEL

We now adapt the definition and the dynamics of population models to systems embedded in a physical space.

The definition of a spatial population model is strictly related to the choice of the space in which the model is embedded. We decided to work with a discrete space, in particular with weighted graphs (i.e. graphs where each edge is labeled with a value indicating the cost/weight/distance to pass from a node to the next one). Each node represents a different spatial location (*patch*), each patch contains a population of agents, described by indicating the number of individuals in each state as a classic population process, as explained in Section 2. The weight of edges has to be interpreted in a more broader sense than a spatial distance, and as such we will refer to it by the word ‘‘cost’’. The idea is that the cost of an edge can represent more complex information than the sheer distance between two nodes. For instance, in a traffic model, we may want to distinguish two streets with the same distance but different travel times due to the presence of traffic lights or congestion.

**DEFINITION 3.1 (Patch-based population model).** A Patch-based population model is a tuple  $(\mathcal{M}, G, \mathcal{V})$  where:

- $\mathcal{M} = (\mathbb{S}, \mathbf{X}, \mathcal{T})$  is a population model, satisfying Definition 2.1,
- $G = (L, E, w)$  is a weighted graph where:
  - $L$  is the finite set of locations (nodes),  $L \neq \emptyset$
  - $E \subseteq L \times L$  is the set of connections (edges),
  - $w : E \rightarrow \mathbb{R}_{>0}$  is the function that returns the cost of each edge.

Furthermore, we denote by  $E^*$  the set containing all the pairs of connected locations, i.e. the transitive closure of  $E$ , extending  $w$  to the domain  $E^*$  as the sum of costs of a shortest path between two different locations<sup>2</sup>,

<sup>2</sup>The shortest path is the path that minimizes the sum of costs.

–  $\mathcal{V} = \{\nu_1, \dots, \nu_k\}$  is the set of inter-patch transitions, i.e. the transitions that describe the migration of entities between patches, each transition is of the form  $\nu_i = (a_i, s, g_i)$ , where:

- $a_i$  is the label of the transition,
- $s \in \mathbb{S}$  is the state of the entity that migrates
- $g_i : \mathbb{D} \times L \times L \rightarrow \mathbb{R}_{\geq 0}$  is the rate function, where  $\mathbb{D}$  is the state space of the system;  $g_i(\mathbf{X}, \ell_i, \ell_j)$  is the rate for the migration of a component in state  $s$  from location  $\ell_i$  to location  $\ell_j$  when the global state of the system is  $\mathbf{X}$ .

The dynamics of these kind of models is given by intra-patch interactions and inter-patch migration of agents resulting in a PCTMC or in a ODEs system where the variables represent the number/density of each agent state in each location. In the following, we refer to the description of such a process by a family of vectors  $\mathbf{X}(t, \ell)_{t \in \mathcal{T}, \ell \in L} = (X_1(t, \ell), X_2(t, \ell), \dots, X_n(t, \ell))_{t \in \mathcal{T}, \ell \in L}$ , indexed by time and space. However, for practical purposes, e.g. to numerically solve ODEs, we consider one (time-dependent) variable  $X_{i,\ell}$  for each state-location pair.

### 4. SIGNAL SPATIO-TEMPORAL LOGIC

Signal Spatio-Temporal Logic (SSTL) is an extension of Signal Temporal Logic (STL) [28] with a spatial operator suitable to specify spatial behaviour of complex systems. The new spatial operator is interpreted on patch-based quantitative population models with a discrete space structured as a graph, as defined in the previous section.

We recall that the state of the system under study is described by the state vector  $\mathbf{X} = (X_1, \dots, X_n)$ , with state space  $\mathbb{D} = \mathbb{D}_1 \times \dots \times \mathbb{D}_n \subseteq \mathbb{R}^n$ , where each variable  $X_i \in \mathbb{D}_i \subseteq \mathbb{R}$ . Given, a model  $(\mathcal{M}, G, \mathcal{V})$ , SSTL formulae are evaluated over real valued traces  $\bar{x} : \mathbb{T} \times L \rightarrow \mathbb{D}$ , generated by simulating the model<sup>3</sup>, where  $\mathbb{T}$  is the time set and  $L$  is the set of locations. We write  $\bar{x}(t, \ell) = (x_1(t, \ell), \dots, x_n(t, \ell)) \in \mathbb{D}$ , where each  $x_i : \mathbb{T} \times L \rightarrow \mathbb{D}_i$ , for  $i = 1, \dots, n$ , is the projection on the  $i^{th}$  coordinate/variable and it is called the *primary signal*.

#### 4.1 Syntax

The syntax of SSTL is given by

$$\varphi := \mu \mid \neg \varphi \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \mathcal{U}_{[t_1, t_2]} \varphi_2 \mid \diamond_{[w_1, w_2]} \varphi$$

where conjunction and negation are the standard boolean connectives,  $[t_1, t_2]$  and  $[w_1, w_2]$  are real positive dense intervals with  $t_1 < t_2$  and  $w_1 < w_2$ ,  $\mathcal{U}_{[t_1, t_2]}$  is the *until* operator and  $\diamond_{[w_1, w_2]}$  is the spatial *somewhere* operator. The atomic predicate  $\mu : \mathbb{D} \rightarrow \mathbb{B}$  is of the form  $\mu \equiv (f \geq 0)$ , where  $f : \mathbb{D} \rightarrow \mathbb{R}$  is a (possibly non-linear) real-valued function, and  $\mathbb{B} = \{0, 1\}$  are boolean values. The function  $y : \mathbb{T} \times L \rightarrow \mathbb{R}$  defined from the image of the trace  $\bar{x}(t, \ell)$  by  $f$ ,  $y(t, \ell) := f(\bar{x}(t, \ell))$ , is called the *secondary signal*.

The (bounded) until operator  $\varphi_1 \mathcal{U}_{[a, b]} \varphi_2$  requires  $\varphi_1$  to hold from now until, in a time between  $a$  and  $b$  time units in the future,  $\varphi_2$  becomes true. The spatial somewhere operator  $\diamond_{[w_1, w_2]} \varphi$  requires  $\varphi$  to hold in a location reachable from the current one with a total cost greater than or equal to  $w_1$  and less than or equal to  $w_2$ . The *eventually* operator  $F_{[a, b]} \varphi$  and the *always* operator  $G_{[a, b]} \varphi$  can be defined as usual:  $F_{[a, b]} \varphi := \top \mathcal{U}_{[a, b]} \varphi$ ,  $G_{[a, b]} \varphi := \neg F_{[a, b]} \neg \varphi$ . In a similar way, we derive the *everywhere* spatial operator  $\boxplus_{[w_1, w_2]} \varphi :=$

<sup>3</sup>With simulation we mean stochastic simulation in case of a stochastic dynamics and numerical integration in case of a deterministic dynamics.

$\neg \diamond_{[w_1, w_2]} \neg \varphi$  which requires  $\varphi$  to hold in all the locations reachable from the current one with a total cost greater than or equal to  $w_1$  and less than or equal to  $w_2$ .

## 4.2 Semantics

STL has two different semantics: a boolean semantics that returns yes/no depending if the observed trace satisfies or not the STL specification and a quantitative semantics that extends the boolean one by returning a measure of robustness of the specification. We now define both semantics for the SSTL logic.

**DEFINITION 4.1 (SSTL Boolean Semantics).** *The boolean satisfaction relation for an SSTL formula  $\varphi$  over a spatio-temporal trace  $\bar{x}$  is given by:*

$$\begin{aligned} (\bar{x}, t, \ell) \models \mu & \Leftrightarrow \mu(\bar{x}(t, \ell)) = 1 \\ (\bar{x}, t, \ell) \models \neg \varphi & \Leftrightarrow (\bar{x}, t, \ell) \not\models \varphi \\ (\bar{x}, t, \ell) \models \varphi_1 \wedge \varphi_2 & \Leftrightarrow (\bar{x}, t, \ell) \models \varphi_1 \text{ and } (\bar{x}, t, \ell) \models \varphi_2 \\ (\bar{x}, t, \ell) \models \varphi_1 \mathcal{U}_{[a, b]} \varphi_2 & \Leftrightarrow \exists t' \in [t + a, t + b] \text{ s.t. } (\bar{x}, t', \ell) \models \varphi_2 \\ & \text{and } \forall t'' \in [t, t'], (\bar{x}, t'', \ell) \models \varphi_1 \\ (\bar{x}, t, \ell) \models \diamond_{[w_1, w_2]} \varphi & \Leftrightarrow \exists \ell' \in L \text{ s.t. } (\ell', \ell) \in E^*, \\ & w_1 \leq w(\ell', \ell) \leq w_2 \text{ and } (\bar{x}, t, \ell') \models \varphi \end{aligned}$$

A trace  $\bar{x}$  satisfies  $\varphi$  in location  $\ell$ , denoted by  $(\bar{x}, \ell) \models \varphi$ , if and only if  $(\bar{x}, 0, \ell) \models \varphi$ .

**DEFINITION 4.2 (SSTL Quantitative Semantics).** *The quantitative satisfaction relation for an SSTL formula  $\varphi$  over a spatio-temporal trace  $\bar{x}$  is given by:*

$$\begin{aligned} \rho(\mu, \bar{x}, t, \ell) & = f(\bar{x}(t, \ell)) \quad \text{where } \mu \equiv (f \geq 0) \\ \rho(\neg \varphi, \bar{x}, t, \ell) & = -\rho(\varphi, \bar{x}, t, \ell) \\ \rho(\varphi_1 \wedge \varphi_2, \bar{x}, t, \ell) & = \min(\rho(\varphi_1, \bar{x}, t, \ell), \rho(\varphi_2, \bar{x}, t, \ell)) \\ \rho(\varphi_1 \mathcal{U}_{[a, b]} \varphi_2, \bar{x}, t, \ell) & = \sup_{t' \in [t+a, t+b]} (\min\{\rho(\varphi_2, \bar{x}, t', \ell), \\ & \quad \inf_{t'' \in [t, t']} (\rho(\varphi_1, \bar{x}, t'', \ell))\}) \\ \rho(\diamond_{[w_1, w_2]} \varphi, \bar{x}, t, \ell) & = \max\{\rho(\varphi, \bar{x}, t, \ell') \mid \ell' \in L, (\ell', \ell) \in E^* \\ & \quad \text{and } w_1 \leq w(\ell', \ell) \leq w_2\} \end{aligned}$$

where  $\rho$  is the quantitative satisfaction function, returning a real number  $\rho(\varphi, \bar{x}, t)$  quantifying the degree of satisfaction of the property by the trace  $\bar{x}$  at time  $t$ . Moreover,  $\rho(\varphi, \bar{x}, \ell) := \rho(\varphi, \bar{x}, 0, \ell)$ .

The definition of this quantitative measure is a reformulation of the robustness degree of [19]. The idea is that the sign of the satisfaction degree corresponds to the truth value of the formula (with positive standing for true), while the absolute value of the score returns the maximum perturbation sustainable by the secondary signals such that the truth value of the formula does not change. Note that, indeed, the robustness of an atomic proposition  $\mu$  is equal to the secondary signal  $y(t, \ell) = f((\bar{x}, t, \ell))$ . The choice of the secondary signals  $y$  is an essential part of the definition of the SSTL formula. Different choices of  $y$  result in different formulae, hence in different robustness measures. We also remark that this robustness notion deviates from the more usual sensitivity-based measure of robustness, like [27]. Despite this fact, the application of the sensitivity analysis and its related techniques to the robustness score of Definition 4.2 is still possible; for more details we refer to [15].

## 5. MONITORING ALGORITHMS

This section is concerned with the *monitoring* of an SSTL formula; a *property monitor* is an algorithm for deciding whether a given behaviour or trace  $\bar{x}$  satisfies a property  $\varphi$ .

We need to provide an algorithm for each semantics that we have defined. We will start from the property monitors introduced in [28] for the boolean semantics and in [15] for the quantitative semantics, extending them with a procedure to check the spatial properties.

### 5.1 Boolean semantics

The algorithm proceeds inductively bottom-up on the parse tree of the formula. Given a formula  $\varphi$ , to determine if  $(\bar{x}, \ell) \models \varphi$ , we construct, for every sub formula  $\psi$  and every location  $\ell \in L$ , a *boolean signal*  $s_{\psi, \ell}(t)$ , i.e. a function from  $[0, T]$ ,  $T \in \mathbb{R}_{>0}$ , to  $\{0, 1\}$  s.t.  $s_{\psi, \ell}(t) = 1$  if  $(\bar{x}, t, \ell) \models \psi$  and 0 otherwise. At the termination of the algorithm, we have the signal  $s_{\varphi, \ell}(t)$  whose value at  $t = 0$  determines if the trace  $\bar{x}$  satisfies  $\varphi$ . Note that we are fixing in the signal both the location and the subformula, because the properties can be verified over each location independently, due to Definition 4.1:  $(\bar{x}, \ell) \models \varphi$  means “the trace  $\bar{x}$  in location  $\ell$  satisfies the property  $\varphi$ ”.

A central notion in the monitoring algorithm is that of the *minimal interval covering*  $\mathcal{I}_{s_{\psi, \ell}}$  consistent with the signal  $s_{\psi, \ell}$ . It corresponds to the shortest finite sequence of left-closed right open intervals  $I_1, \dots, I_h$  such that  $\bigcup_j I_j$  is an interval,  $I_i \cap I_j = \emptyset$ ,  $\forall i \neq j$ , and  $s_{\psi, \ell}(t) = s_{\psi, \ell}(t')$ ,  $\forall t, t'$  belonging to the same interval.<sup>4</sup> We can derive then the set of positive and negative intervals of  $s_{\psi, \ell}$  as:

$$\mathcal{I}_{s_{\psi, \ell}}^+ = \{I \in \mathcal{I}_{s_{\psi, \ell}} \mid \forall t \in I : s_{\psi, \ell}(t) = \psi\},$$

$$\mathcal{I}_{s_{\psi, \ell}}^- = \{I \in \mathcal{I}_{s_{\psi, \ell}} \mid \forall t \in I : s_{\psi, \ell}(t) \neq \psi\}.$$

The positive interval  $\mathcal{I}_{s_{\psi, \ell}}^+$  corresponds to the *satisfaction set* of the formula over the signal  $s_{\psi, \ell}$ . Note that any signal can be written as  $s = s_1 \vee s_2 \vee \dots \vee s_k$  where each  $s_i$  is an *unitary signal*, meaning that it has a singleton positive interval, i.e.  $\mathcal{I}_{s_i}^+ = \{[t_1, t_2]\}$  for some  $t_1 < t_2 \in \mathbb{R}_{>0}$ . Furthermore,  $\mathcal{I}_{s_{\psi, \ell}} = \mathcal{I}_{s_{\psi, \ell}}^+ \cup \mathcal{I}_{s_{\psi, \ell}}^-$  and  $\mathcal{I}_{s_{\psi, \ell}}^+ \cap \mathcal{I}_{s_{\psi, \ell}}^- = \emptyset$ .

Using these definitions of signals, interval coverings, and satisfaction set, the procedure for the classic operators of STL is equivalent to the one described in paper [28]. We briefly recall it in the following.

**Atomic Predicates:**  $\psi = \mu$ . The computation of the boolean signal associate with an atomic predicate is a direct application of Definition 4.1:  $s_{\mu, \ell}(t) = \mu(\bar{x}(t, \ell))$ .

**Negation:**  $\psi = \neg \varphi$ , then  $\mathcal{I}_{s_{\neg \varphi, \ell}}^+ = \mathcal{I}_{s_{\varphi, \ell}}^-$ ,

**Disjunction:**  $\psi = \varphi_1 \vee \varphi_2$ , then, given  $s_{\varphi_1, \ell}, s_{\varphi_2, \ell}$ , let be  $\mathcal{I}$  the minimal interval covering consistent with *both* signals. For each  $I_i \in \mathcal{I}$ , we construct the signal  $s_{\psi, \ell}(I_i) = s_{\varphi_1, \ell}(I_i) \vee s_{\varphi_2, \ell}(I_i)$  and we merge adjacent positive intervals to obtain  $\mathcal{I}_{s_{\psi, \ell}}^+$ .

**Until:**  $\psi = \varphi_1 \mathcal{U}_{[a, b]} \varphi_2$ . As we are working with future temporal modalities, we need to shift intervals *backwards*. This has to be done independently for each unitary signal, then taking the union of the so obtained satisfaction sets. Given two unitary signals  $p$  and  $q$ , the signal  $\psi = p \mathcal{U}_{[a, b]} q$  is the unitary signal such that  $\mathcal{I}_{\psi}^+ = \{((I_p \cap I_q) \ominus [a, b]) \cap I_p\}$ , where  $[m, n] \ominus [a, b] =$

<sup>4</sup>The fact that we can always obtain a finite interval covering is a consequence of the restriction to closed intervals  $[t_1, t_2]$ ,  $t_1 < t_2$ , in STL. For more detail about signals and interval covering see [28].

$[m - b, n - a] \cap [0, T]$ . In the general case, let  $s_{\varphi_1, \ell} = p_1 \vee \dots \vee p_n$  and  $s_{\varphi_2, \ell} = q_1 \vee \dots \vee q_m$  be signals written as union of unitary signals, then  $\psi = s_{\varphi_1, \ell} \mathcal{U}_{[a, b]} s_{\varphi_2, \ell} = \bigvee_{i=1}^n \bigvee_{j=1}^m p_i \mathcal{U}_{[a, b]} q_j$ . The proof of this result can be found in [28].

We now treat the spatial *somewhere* operator  $\psi = \diamond_{[w_1, w_2]} \varphi$ . As remarked at the beginning of this section, and relying on the fact that we have a finite number of locations, we can process independently each location in the signal. Given the signal  $s_{\psi, \ell}$ , for a fixed location  $\ell$ , we can rewrite the spatial operator as a disjunction between all signals in locations  $\ell'$  s.t.  $(\ell', \ell) \in E^*$  and  $w_1 \leq w(\ell', \ell) \leq w_2$ . This allows us to use the monitoring procedure for disjunction, constructing the minimal interval covering  $\mathcal{I}$  consistent with all  $s_{\varphi, \ell'}$  signals s.t.  $(\ell', \ell) \in E^*$  and  $w_1 \leq w(\ell', \ell) \leq w_2$ , and defining, for each  $I_i \in \mathcal{I}$ :

$$s_{\psi, \ell}(I_i) = \bigvee_{w_1 \leq w(\ell', \ell) \leq w_2} s_{\varphi, \ell'}(I_i).$$

The satisfaction set  $\mathcal{I}_{s_{\psi, \ell}}^+$  is then the union of the positive  $I_i$  (i.e.  $I_i$  s.t.  $s_{\psi, \ell}(I_i) = 1$ ), merging adjacent positive intervals.

We stress here that the introduction of the spatial somewhere operator is not merely syntactic sugar, for two reasons. First, its definition can be applied also to countable discrete spaces, and it can be easily generalised to continuous spaces. Secondly, even assuming a finite discrete space, expanding it as a disjunction would produce a blowup of the formula size *exponential* in the nesting level of spatial operators, and hence an exponential increase in the complexity of the monitoring procedure.

## 5.2 Quantitative semantics

Given a formula  $\varphi$ , this algorithm computes the quantitative satisfaction function  $\rho(\varphi, \vec{x}, t, \ell)$ . Similarly to the Boolean semantics,  $\rho$  is computed inductively on the formula structure, proceeding bottom-up along the parse tree of the formula. For each node of the tree, the algorithm maintains a real-valued satisfaction signal associated with the corresponding subformula, processing the real-valued signals of the children nodes.

We recall that the primary signals  $x_i(t, \ell)$ ,  $i = 0, \dots, n$ , come from the simulation of the model, via numerical integration or stochastic simulation, returning a finite set of sample points. Note that the number of primary signals  $x_{i, \ell}(t) = x_i(t, \ell)$  is finite, as so is the set of locations  $L$ . Furthermore,  $x_{i, \ell}(t)$  are assumed to be *piecewise-affine* functions, obtained by linear interpolation of the finite set of sample points. Similarly, secondary signals  $y_\ell(t) = y(t, \ell)$ ,  $y(t, \ell) = f(\vec{x}(t, \ell))$  with  $\mu \equiv (f \geq 0)$ , can be also seen as piecewise-affine functions. We need to assume that  $y_\ell$  is right-continuous and admits everywhere a right-derivative

$$\frac{dy_{\ell+}}{dt}(t) := \lim_{\epsilon \rightarrow 0} \frac{y_\ell(t + \epsilon) - y_\ell(t)}{\epsilon}.$$

These secondary signals are then completely represented by the sequence  $(t_i, y_\ell(t_i), dy_\ell(t_i))_{i < n_{y_\ell}}$ , with a final or cut-off time  $t_{n_{y_\ell}}$ . Furthermore, they can be extended to  $t \in [0, \infty)$  by assuming  $y_\ell(t) = y_\ell(t_1)$  for all  $t < t_1$  and  $y_\ell(t) = y_\ell(t_{n_{y_\ell}})$  for all  $t > t_{n_{y_\ell}}$ .

Treating also in this case the location as a fixed coordinate, the computation of the signals associated with the negation ( $\neg$ ), disjunction ( $\vee$ ), and until ( $\mathcal{U}_{[a, b]}$ ) will be the same as in the monitoring algorithm described in [15]. Due to the complexity of such algorithm, especially for the temporal operator, we refer the reader to [15] for a comprehensive description.

Here we limit ourselves to describe the monitoring procedure for the spatial somewhere operator  $\diamond_{[w_1, w_2]}$ . From Definition 4.2,

$$\rho(\diamond_{[w_1, w_2]} \varphi, \vec{x}, t, \ell) = \max_{\ell' \in B_\ell} (\rho(\varphi, \vec{x}, t, \ell')),$$

where  $B_\ell = \{\ell' \in L \mid (\ell', \ell) \in E^*, \ell' \neq \ell, w_1 \leq w(\ell', \ell) \leq w_2\}$  is a finite subset of locations. Let us fix two locations  $\ell_1$  and  $\ell_2$  in  $B_\ell$ , and denote by  $(t_i, y_{\ell_1}(t_i), dy_{\ell_1}(t_i))_{i < n_{y_{\ell_1}}}$  and  $(t_i, y_{\ell_2}(t_i), dy_{\ell_2}(t_i))_{i < n_{y_{\ell_2}}}$  the sequences representing the signals  $\rho(\varphi, \vec{x}, \ell_1)$  and  $\rho(\varphi, \vec{x}, \ell_2)$ , which are assumed known by induction. We now build the sequence of time points  $(r_i)_{i < n_z}$ , containing the sorted sampling times of  $y_{\ell_1}$  and  $y_{\ell_2}$  and the time points in which the corresponding piecewise affine functions intersect. Note that  $n_z \leq 4 \times \max\{n_{y_{\ell_1}}, n_{y_{\ell_2}}\}$  because there can be at most  $n_{\ell_1} + n_{\ell_2}$  intersections and no more than  $n_{\ell_1} + n_{\ell_2}$  unique sampling instants. For all  $i < n_z$ , we let

$$z(r_i) = \max(y_{\ell_1}(r_i), y_{\ell_2}(r_i)).$$

and

$$dz(r_i) = \max(dy_{\ell_1}(r_i), dy_{\ell_2}(r_i)).$$

Now, if  $\exists \ell' \in B_\ell$  s.t.  $\ell' \neq \ell_1$  and  $\ell' \neq \ell_2$ , we apply the same algorithm to  $(t_i, y_{\ell'}(t_i), dy_{\ell'}(t_i))_{i < n_{y_{\ell'}}}$ , the sequences representing the robustness signal  $\rho(\varphi, \vec{x}, \ell')$  and the sequence  $(r_i, z(r_i), dz(r_i))_{i < n_z}$  obtained from the previous computation, iterating this process until all the locations in  $B_\ell$  have been considered. The resulting sequence  $(r_i, z_\ell(r_i), dz(r_i))_{i < n_{z_\ell}}$  represents  $\rho(\diamond_{[w_1, w_2]} \varphi, \vec{x}, t, \ell)$  as a piecewise affine function.

## 6. STOCHASTIC SEMANTICS

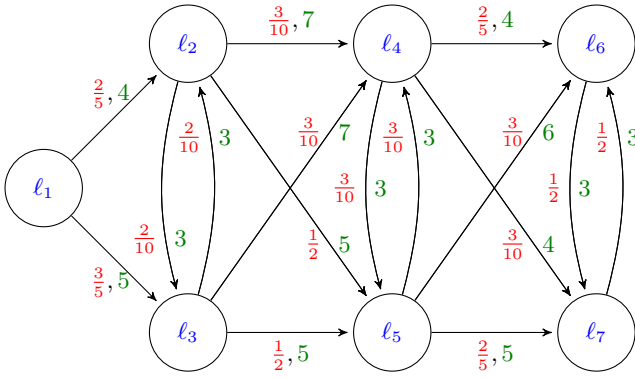
We discuss now how to extend the definition of the semantics of SSTL to a stochastic system, following [6]. Each SSTL formula  $\varphi$  is interpreted over a trajectory  $\vec{x} = \vec{x}(t, \ell)$ . Let us denote by  $\mathcal{D}$  the space of all possible trajectories of a CTMC  $\vec{X}(t, \ell)$  associated with a patch-based population model  $(\mathcal{M}, G, \mathcal{V})$ . The standard cylindrical construction [9] turns this set into a metric measurable space. The CTMC  $\vec{X}(t, \ell)$ , in turn, induces a probability measure on  $\mathcal{D}$ . Furthermore, we can interpret the boolean and the quantitative semantics of  $\varphi$  as functionals on the space  $\mathcal{D}$ , assigning to each trajectory  $\vec{x}$  its corresponding truth value according to the boolean semantics or its satisfaction degree according to the quantitative semantics. In the boolean case, such a functional is  $I_\varphi : \mathcal{D} \rightarrow \{0, 1\}$ , such that  $I_\varphi(\vec{x}) = 1$  if and only if  $\vec{x} \models \varphi$ , hence it identifies the subset of trajectories of  $\mathcal{D}$  that satisfy the formula  $\varphi$ . It follows that

$$P(\varphi) = \mathbb{P}\{\vec{x} \in \mathcal{D} \mid I_\varphi(\vec{x}) = 1\} = \mathbb{P}\{\vec{x} \in \mathcal{D} \mid \vec{x} \models \varphi\}.$$

An analogous construction allows us to extend the robustness of SSTL to PCTMC models: given a trajectory  $\vec{x}(t, \ell)$  of the PCTMC, its robustness  $\rho(\varphi, \vec{x}, 0)$  can be seen as a functional  $R_\varphi$  from the trajectories in  $\mathcal{D}$  to  $\mathbb{R}$ . It can be proved that  $R_\varphi$  is measurable (with respect to the Borel  $\sigma$ -algebra of the standard topology of  $\mathcal{D}$ ) [6]. From this assumption, we can define the real-valued random variable  $R_\varphi = R_\varphi(\vec{X})$  with probability distribution:

$$\mathbb{P}(R_\varphi(\vec{X}) \in [a, b]) = \mathbb{P}(\vec{X} \in \{\vec{x} \in \mathcal{D} \mid \rho(\varphi, \vec{x}, 0) \in [a, b]\}).$$

More details about this definition can be found in [6]. Applying this definition to a stochastic model, we obtain a distribution of robustness degrees which can be summarised by the average robustness degree of this distribution,  $\mathbb{E}(R_\varphi)$ , giving a measure of how strongly a formula is satisfied. This means that the satisfaction is more robust when this value is higher.



**Figure 1: The graph of the population spatial distribution /locations distribution. Note that this is not the graph of the patch-PCTMC model but just the graph of the locations. The nodes  $\ell_1, \dots, \ell_7$  represent the different communities; the edges represent the connection between the communities through the water basin, with the verse of the rows chosen in agreement with the river flow. The red numbers correspond to the probability  $p_{i,j}$  to go from  $\ell_i$  to  $\ell_j$  and the green numbers correspond to the values  $w$  of the distance between locations.**

Numerical methods to compute the satisfaction probability or the average robustness of a STL formula are not available, hence we resort, as in [6], to statistical estimation, leveraging the monitoring routines presented in the previous section. In particular, we use a *Bayesian statistical model checking* approach to estimate the satisfaction probability [25], and classic statistical tools for the average robustness degree [6].

## 7. CASE STUDY

We consider a model of an epidemic of cholera, an infection of the intestine caused by the bacterium *Vibrio cholerae* and a prominent example of a waterborne disease [8, 29]. Typically, the infection is transmitted by contaminated food or water. An explicit modelling of the hydrological space where the infection spreads is therefore a crucial aspect to understand and analyse this kind of disease. For this reason this example is very suitable to show the potentiality of this logic, in particular we will see how our logic can easily describe elaborated spatio-temporal behaviours by means of the spatial operators. We represent space as a weighted oriented graph, shown in Figure 1. The nodes represent the human communities and the edges describe the links between water basins of different areas. The idea is to analyse the diffusion of an epidemic along the communities that live close to a river, so that edge directions are chosen according to the direction of the water flow. There are two agent classes: the bacteria and the individuals. The bacteria have only one state ( $B$ ) but they can be transported to different nodes via the river. An individual, instead, can be in three different states: susceptible ( $S$ ) infected ( $I$ ) and recovered ( $R$ ), but cannot change location. Ignoring human mobility is a simplification justified by the fact that our focus here is to illustrate the logic at work, rather than presenting a fully realistic model. Extension to more complex scenarios, however, are relatively straightforward [29], and can be easily described within our formal framework, so that analogous or more complex spatio-temporal logical properties can still be checked.

The model of this system has state variables  $X_S, X_I, X_R, X_B$ , counting the number of susceptible, infected, and recovered indi-

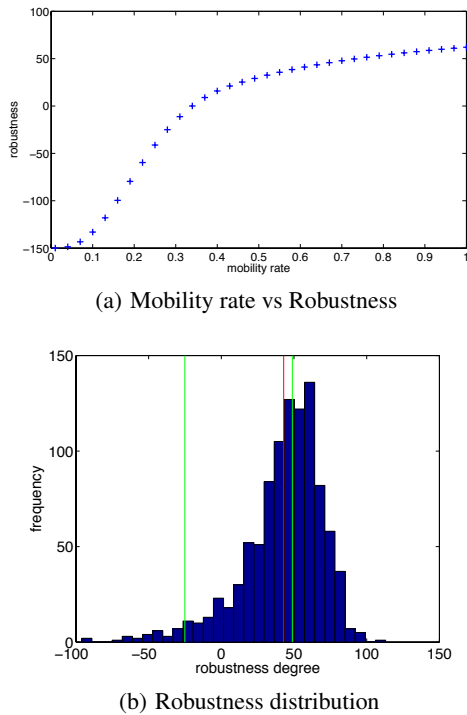
viduals, and the concentration of the bacteria in each location. Discrete space is described by the graph in Figure 1, which is equipped with a weight function  $w : E \rightarrow \mathbb{R}$  that returns the cost of each edge. In this specific case, we will interpret the cost  $w(\ell_i, \ell_j)$  of an edge between nodes  $\ell_i$  and  $\ell_j$  as the distance between them. Recall that if two locations are not directly connected,  $w$  is equal to the sum of costs of a shortest path between the two locations; for example, in Figure 1,  $w(\ell_1, \ell_5) = w(\ell_1, \ell_2) + w(\ell_2, \ell_5) = 9$ . The movement of bacteria from a location  $\ell_i$  to a directly connected location  $\ell_j$  is specified by the inter-patch transitions  $\nu_{mov} = (mov, B, g_{mov})$ . Here  $g_{mov}(\mathbf{X}(t, \ell), \ell_i, \ell_j) = lp_{ij}X_B(t, \ell_i)$  is the rate function, where  $l$  is the total water flow (assumed equal for all nodes) and  $p_{ij}$  is the fraction of water flowing out of node  $\ell_i$  which reaches node  $\ell_j$ . Such a rate function describes the bacteria mobility as a passive mobility due to the water flow. The sum of the values  $p_{ij}$  associated with the outgoing edges of each location has is equal to one, i.e  $\sum_k |(\ell_i, \ell_k) \in E| p_{ik} = 1$ , except for the locations  $\ell_6$  and  $\ell_7$  where is less than one to capture the continuation of the river. The other transitions describe the change of the individual state, according to a model similar to the SIR epidemics introduced in Example 2.1, with the notable difference of the infection rate, which now depends on the concentration of bacteria in water and not directly on the number of infected individuals. We also consider transitions describing decrease or growth of the bacterium population.

The patch population model is thus  $((S, \mathbf{X}, \mathcal{T}), G, \mathcal{V})$ , where  $S = \{S, I, R, B\}$  is the set of states,  $\mathbf{X} = (X_S, X_I, X_R, X_B)$  is the state vector,  $G = (\{\ell_1, \dots, \ell_n\}, E, w)$  is the graph as in Figure 1,  $\mathcal{V} = \{(mob, B, g_{mob})\}$  is the set of inter-patch transition, as described above, and  $\mathcal{T}$  is the set containing the following intra-patch transitions. More specifically:

- \*  $\tau_{inf}$  is the *infection* transition:  $S+B \rightarrow I+B$  with rate  $f_{inf}(\mathbf{X}) = \beta(\ell) \frac{X_B}{K+X_B} X_S$ , where  $K$  is the half saturation constant,  $\beta(\ell)$  represents the site-dependent rate of exposure to contaminated water,
- \*  $\tau_{Snat}$  is the *natality* transition:  $\emptyset \rightarrow S$ , with rate  $f_{Snat}(\mathbf{X}) = \mu H(\ell)$ , where  $\mu$  is the human mortality rate and  $H(\ell)$  is the size of the community in location  $\ell$ ,
- \*  $\tau_{Smort}$  is the *mortality* of a susceptible individual:  $S \rightarrow \emptyset$ , with rate  $\mu$ ,
- \*  $\tau_{Imort}$  is the *mortality* of an infected individual:  $I \rightarrow \emptyset$ , with rate  $\mu + \alpha$ , where  $\alpha$  is the mortality rate due to cholera,
- \*  $\tau_{rec}$  is the *recovery* transition:  $I \rightarrow R$ , with rate  $\gamma$ ,
- \*  $\tau_{Bdeg}$  is the *degradation* transition:  $B \rightarrow \emptyset$ , with rate  $\mu_B$ ,
- \*  $\tau_{Bgrowth}$  is the *bacterial growth* transition:  $\emptyset \rightarrow B$ , with rate function  $f_{Bgrowth}(\mathbf{X}) = \frac{p}{W} X_I$ , where  $p$  is the rate at which bacteria are produced by one infected person and  $W$  is the volume of the contaminated water.

We implemented this model in `Matlab`, both as a set of differential equations and as a stochastic process. In order to compute the boolean and the quantitative semantics of SCTL formulae, we exploited the routines provided by the `Breach` toolbox [14], adapting them to check SCTL formula. ODEs describing the dynamics of the systems, obtained from the prescriptions of Section 2, have been numerically integrated using a `Matlab` built-in solver. The stochastic interpretation of the population model, instead, has been analysed by a dedicated `Java` implementation integrated within `Breach`, combining standard Monte Carlo simulation (by the Gillespie algorithm [22]) and Bayesian statistical model checking [25].

The first spatio-temporal behaviour that we consider is how the epi-



**Figure 2: (a) Dependency of the robustness degree on the mobility rate (for the ODE interpretation of the model). (b) Empirical robustness distribution of the formula 3 with  $w_1 = 12$  and  $w_2 = 15$  and  $T_{inf} = 5$ , obtained from 1000 simulation runs.**

demographic propagates along the river. The idea is to consider a model that starts with the infection in just one location,  $\ell_1$ , and then to check if the infection has propagated at a certain distance from  $\ell_1$  after a certain time. This behaviour can be captured by the SSTL formula:

$$\varphi_1 = F_{[0, T_{inf}]} \diamond_{[w_1, w_2]} (X_I > c_{inf}), \quad (3)$$

verifying it in location  $\ell_1$ . The exact meaning of the formula is: eventually, in less than  $T_{inf}$  unit time, the number of infected individuals becomes more than  $c_{inf}$  in a location  $\ell$  with  $w(\ell_1, \ell) \in [w_1, w_2]$ , i.e. at a distance from location  $\ell_1$  equal or greater than  $w_1$  and equal or less than  $w_2$ .

We analyse a system with 7 locations, all with the same size, with distance between locations  $w(\ell_i, \ell_j)$ , in agreement with the green integer values labelling the edges in Figure 1. As initial variables we set, for all the locations apart  $\ell_1$ , the same number of susceptible individuals,  $X_S = 500$ , and we start with zero infected individuals,  $X_I = 0$ ; instead, for  $\ell_1$  we set 100 infected individuals and 400 susceptible. There are no recovered individuals at the beginning. We set also the bacterium concentration equal to zero in all the locations except  $\ell_1$ . The parameters of the model have been set as  $\mu = 0.0005$ ,  $H(\ell_i) = 500$ ,  $\beta = 1$ ,  $K = 5$ ,  $\gamma = 0.2$ ,  $\alpha = 0.0004$ ,  $\mu_B = 0.228$ ,  $p = 0.2$ ,  $w = 50$ ,  $l = 0.5$  while the matrix  $(p_{ij})_{i, j \leq 7}$  has as non-zero entries those specified in Figure 1. For the formula parameters, we set  $w_1 = 12$  and  $w_2 = 15$ ,  $c_{inf} = 150$  and  $T_{inf} = 5$  unit times.

In Figure 2, we show some results of the monitoring of this formula. In the top panel, we consider the ODE interpretation and show how the robustness score increases as a function of the mo-

bility rate. In the bottom panel, we show the empirical distribution from 1000 runs of the robustness degree from the stochastic model, with vertical lines denoting the average (red lines) and conditional averages on the formula being false/ true (green lines). Changing the mobility rate in the stochastic model from 0.3 to 0.6, the satisfaction probability varies from 0.397 to 0.975, while the average robustness score varies (monotonically) from -16.93 to 52.72. We stress here how the robustness score can be used for system design purposes, trying to robustly match spatio-temporal SSTL specifications, following e.g. [6].

In order to illustrate in more detail the expressive power of this logic, we discuss now two additional properties, using the following building blocks:

$$\begin{aligned} \psi_1 &= F_{[0, T_{start}]} (\diamond_{[0, w_{near}]} (X_I > c_{inf})) \\ \psi_2 &= (F_{[T_{start}, T_{start} + DT]} \diamond_{[w_{far}, d_{max}]} (X_I > c_{inf})) \end{aligned}$$

The first one is  $\boxplus_{[0, d_{max}]} (\psi_1 \rightarrow \psi_2)$ , stating that a large infection (with at least  $c_{inf}$  individuals) happening at some time  $t \in [0, T_{start}]$  and localised within distance  $w_{near}$  from a given reference point, will spread further away, at a location at distance between  $w_{far}$  and  $d_{max}$ , at some time  $t' \in [T_{start}, T_{start} + DT]$ . Furthermore, this is true for every reference point (say at distance at most  $d_{max}$  from  $\ell_1$ ). Checking this formula on 1000 runs of the stochastic model, with parameters  $w_{near} = 3$ ,  $w_{far} = 13$ ,  $d_{max} = 15$ ,  $T_{start} = 1$  and  $DT = 4$ , we obtain a satisfaction probability of  $0.9220 \pm 0.0085$  (all results reported at 95% confidence), and an average robustness degree of  $42 \pm 0.8771$ . The robustness score for the ODE interpretation, instead, is 30.26.

The other formula we consider is  $\psi_2 \rightarrow \psi_1$ , stating that a high infection level at a far-away location at a late time must have been high at a nearer location some time before, i.e. that the current reference point is close to the epicentre of the epidemic. In this case, with the same parameters of the previous formula, verifying it in location  $\ell_2$ , we obtain a satisfaction probability of  $0.997 \pm 0.0017$ , an average robustness score of  $55.48 \pm 0.5677$  and a robustness score of the ODE model of 58.08.

In general, we can observe how the ODE interpretation generates robustness scores that are in agreement to the ones of the stochastic model, at a much cheaper computational price. This may be the effect of some convergence theorem at work, and may be exploited for system design purposes, as numerically solving ODEs is in general computationally cheaper than analysing or simulating stochastic models.

## 8. DISCUSSION

In this paper we presented Signal Spatio-Temporal Logic, an extension of Signal Temporal Logic with spatial operators. First we formalized the definition of patch-based population model, embedding a “classic” homogeneous population model into a discrete space, represented as a weighted graph, and specifying movements by dedicated inter-patch transitions. Then, we introduced SSTL, gave it both a boolean and a quantitative semantics, and extended the definition and the monitoring algorithms presented in [15, 28]. As a case study, we discussed an epidemic scenario of the spreading of cholera among neighbouring communities, where space plays a crucial role, showing how our logic is suitable to describe spatio-temporal behaviour as the epidemic propagation along the river. In this specific example we worked only with few patches, but the same type of analysis can easily be applied to more complex spatial structures, even with thousands of nodes. Furthermore, the same

logics can be applied to more general notions of weighted graphs, not necessarily representing a physical space. For instance, the graph may represent sensors connected in a network, with weights representing the energy cost of transmission.

As for future work, we want to apply our approach to transport networks in the context of smart cities, in particular to bike sharing and bus transport networks. Another direction to investigate is the extension of the semantics over a continuous space. In this case, instead of Patch-CTMC or an ODE system we have to work with spatio-temporal point processes or PDE systems. The main challenge is the design of efficient monitoring algorithms, as we cannot rely anymore on simple extensions of those for STL. Another possible improvement is to consider a richer spatial operator like the spatial until, which describes bounded regions, extending its interpretation with metric constraints [13]. Finally, we plan to release an implementation of the monitoring algorithms, extending the system design procedures of [6] to spatio-temporal properties. We will also consider the problem of learning spatio-temporal formulae satisfied by a model with high probability, or effectively discriminating two models, combining ideas of [7] with the spatio-temporal machine learning tools of [33].

## 9. REFERENCES

- [1] Marco Aiello, Ian Pratt-Hartmann, and Johan van Benthem, editors. *Handbook of Spatial Logics*. Springer, 2007.
- [2] R. Alur, T. Feder, and T.A. Henzinger. The benefits of relaxing punctuality. *J. ACM*, 1996.
- [3] K. Bae and J. Meseguer. A rewriting-based model checker for the linear temporal logic of rewriting. *ENTCS*, 290:19–36, 2012.
- [4] C. Baier, B. Haverkort, H. Hermanns, and J.-P. Katoen. Model-checking algorithms for continuous-time markov chains. *IEEE Trans. Softw. Eng.*, 29(6):524–541, 2003.
- [5] C. Baier and J.-P. Katoen. *Principles of model checking*. MIT Press, 2008.
- [6] E. Bartocci, L. Bortolussi, L. Nenzi, and G. Sanguinetti. On the robustness of temporal properties for stochastic models. In *HSB 2013*.
- [7] E. Bartocci, L. Bortolussi, and G. Sanguinetti. Data-driven statistical learning of temporal logic properties. In *FORMATS*, 2014.
- [8] E. Bertuzzo, S. Azaele, A. Maritan, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. On the space-time evolution of a cholera epidemic. *Water Resources Research*, 44(1), 2008.
- [9] P. Billingsley. *Convergence of probability measures*. Wiley, 1999.
- [10] A. T. Bittig, F. Haack, C. Maus, and A. M. Uhrmacher. Adapting rule-based model descriptions for simulating in continuous and hybrid space. In *CMSB*, 2011.
- [11] L. Bortolussi, J. Hillston, D. Latella, and M. Massink. Continuous approximation of collective systems behaviour: a tutorial. *Performance Evaluation*, 70(5):317–349, 2013.
- [12] L. Cardelli and A. D. Gordon. Anytime, anywhere: Modal logics for mobile ambients. In *POPL*, 2000.
- [13] V. Ciancia, D. Latella, M. Loreti, and M. Massink. Specifying and verifying properties of space. In *IFIP-TCS*, 2014.
- [14] A. Donzé. Breach, a toolbox for verification and parameter synthesis of hybrid systems. In *CAV*, 2010.
- [15] A. Donzé, T. Ferrer, and O. Maler. Efficient robust monitoring for STL. In *CAV*, 2013.
- [16] A. Donzé and O. Maler. Robust satisfaction of temporal logic over real-valued signals. In *FORMATS*, 2010.
- [17] R. Durrett. *Essentials of stochastic processes*. Springer, 2012.
- [18] G. E. Fainekos and G. J. Pappas. Robustness of temporal logic specifications for continuous-time signals. *Theor. Comput. Sci.*, 410(42):4262–4291, 2009.
- [19] G. E. Fainekos and G. J. Pappas. Robustness of temporal logic specifications for continuous-time signals. *Theor. Comput. Sci.*, 410(42):4262–4291, 2009.
- [20] C. Feng and J. Hillston. PALOMA: A process algebra for located markovian agents. In *QEST*, 2014.
- [21] C. Fricker, N. Gast, and H. Mohamed. Mean field analysis for inhomogeneous bike sharing systems. *DMTCS*, 2012.
- [22] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. of Physical Chemistry*, 81(25), 1977.
- [23] R. Grosu, E. Bartocci, F. Corradini, E. Entcheva, S. A. Smolka, and A. Wasilewska. Learning and detecting emergent behavior in networks of cardiac myocytes. In *HSCC*, 2008.
- [24] J. Heath, M. Kwiatkowska, G. Norman, D. Parker, and O. Tymchyshyn. Probabilistic model checking of complex biological pathways. In *CMSB*, 2006.
- [25] S. K. Jha, E. M. Clarke, C. J. Langmead, A. Legay, A. Platzer, and P. Zuliani. Statistical model checking for complex stochastic models in systems biology. In *Proceedings of CMSB’09*, 2009.
- [26] A. Jilkine, A. F. M. Marée, and L. Edelstein-Keshet. Mathematical model for spatial segregation of the rho-family GTPases based on inhibitory crosstalk. *Bulletin of Mathematical Biology*, 69(6):1943–1978, 2007.
- [27] M. Komorowski, M. J. Costa, D. A. Rand, and M. PH Stumpf. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *PNAS USA*, 108(21):8645–8650, 2011.
- [28] O. Maler and D. Nickovic. Monitoring temporal properties of continuous signals. In *Proceedings FORMATS 2004*, volume 3253, pages 152–166, 2004.
- [29] 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, February 2012.
- [30] R. De Nicola, J.-P. Katoen, D. Latella, M. Loreti, and M. Massink. Model checking mobile stochastic logic. *Theor. Comput. Sci.*, 382, 2007.
- [31] John H. Reif and A.P. Sistla. A multiprocess network logic with temporal and spatial modalities. *Journal of Computer and System Sciences*, 30(1):41–53, 1985.
- [32] M. Tizzoni, P. Bajardi, C. Poletto, J. J. Ramasco, D. Balcan, B. Goncalves, N. Perra, V. Colizza, and A. Vespignani. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Medicine*, 10(1):165, December 2012.
- [33] A. Zammit-Mangion, M. Dewar, V. Kadiramanathan, and G. Sanguinetti. Point process modelling of the afghan war diary. *PNAS*, 109(31):12414–12419, 2012.