

# Studying Emergent Behaviours in Morphogenesis using Signal Spatio-Temporal Logic

Ezio Bartocci<sup>1</sup>, Luca Bortolussi<sup>2,3,4</sup>, Dimitrios Miliotis<sup>5</sup>,  
Laura Nenzi<sup>6</sup>, Guido Sanguinetti<sup>5,7</sup>

<sup>1</sup> Faculty of Informatics, Vienna University of Technology, Austria

<sup>2</sup> Dept. of Maths and Geosciences, University of Trieste, Italy

<sup>3</sup> CNR/ISTI, Pisa, Italy

<sup>4</sup> Modelling and Simulation Group, Saarland University, Germany

<sup>5</sup> School of Informatics, University of Edinburgh, UK

<sup>6</sup> IMT Lucca, Lucca, Italy

<sup>7</sup> SynthSys, Centre for Synthetic and Systems Biology, University of Edinburgh, UK

**Abstract.** Pattern formation is an important spatio-temporal emergent behaviour in biology. Mathematical models of pattern formation in the stochastic setting are extremely challenging to execute and analyse. Here we propose a formal analysis of the emergent behaviour of stochastic reaction diffusion systems in terms of Signal Spatio-Temporal Logic, a recently proposed logic for reasoning on spatio-temporal systems. We present a formal analysis of the spatio-temporal dynamics of the Bicoid morphogen in *Drosophila melanogaster*, one of the most important proteins in the formation of the horizontal segmentation in the development of the fly embryo. We use a recently proposed framework for statistical model checking of stochastic systems with uncertainty on parameters to characterise the parametric dependence and robustness of the French Flag pattern, highlighting non-trivial correlations between the parameter values and the emergence of the patterning.

## 1 Introduction

One of the most fascinating questions in biology is how regular patterns can emerge from biochemical processes acting at the cellular level, a process known as *morphogenesis* in developmental biology. Some evident examples of these patterns can be observed in the stripes of a zebra, the spots on a leopard, the filament structure of the cyanobacteria *Anabaena* or the square pattern of the sulfur bacteria *T. rosea*. Mathematical and computational methods hold enormous promise in the quest to unveil the underlying mechanisms of morphogenesis and reproducing, using computer-based simulations, the patterns observed in nature. Alan Turing, mostly known as the father of computer science, was also a pioneer in developing a first mathematical model [28] that provides *the chemical basis of morphogenesis*. This model, also referred as the *Turing's reaction-diffusion system*, is able to reproduce the formation of some complex patterns in nature such as the stripes seen in the animal skin.

Formal analysis of how patterns arise from mathematical models is however challenging due to the high computational burden of spatio-temporal modelling, as well as the intrinsic difficulty of defining spatio-temporal patterns in a suitable language.

Pattern recognition is generally considered as a branch of machine learning [6], where patterns are classified according statistical descriptors (or features) [21] or the structural relationship among them [25]. This approach, despite its success and popularity, lacks of a rigorous foundation to specify such patterns and to reason about them in a systematic way. On the other end, formal methods provide logic-based languages [2, 10, 17, 18] with a well-defined syntax and semantics to specify in a precise and concise way emergent behaviours and the necessary techniques to automatically detect them.

**Related work.** In the last year, two novel spatio-temporal logics, SpaTeL [18] and SSTL [10, 24] have made their appearance almost at the same time in the realm of formal methods to specify the emergence of spatio-temporal patterns.

The *Spatial-Temporal Logic* (SpaTeL) in [18] is the unification of Signal Temporal Logic [23] (STL) and Tree-Spatial-Superposition-Logic (TSSL) introduced in [2] to classify and detect spatial patterns. TSSL reasons over quad trees, spatial data structures that are constructed by recursively partitioning the space into uniform quadrants. TSSL is derived from Linear Spatial-Superposition-Logic (LSSL) [17], where the notion of superposition provides a way to describe statistically the distribution of discrete states in a particular partition of the space and the spatial operators correspond to *zooming in and out* of particular areas. In [17] the authors show also that by nesting these operators they are able to specify self-similar and fractal-like structures that generally characterize the patterns emerging in nature. SpaTeL is equipped with a qualitative (yes/no answer) and a quantitative semantics that provide a measure or robustness of how much the property is satisfied or violated. In [18] this measure of robustness is used as a fitness function to guide the parameter synthesis process for a deterministic reaction diffusion system using particle swarm optimisation (PSO) algorithms. However, the authors do not consider stochastic reaction-diffusion systems and PSO techniques generally do not provide any guarantee for reaching the global optimum. Hence, in this paper we will adopt a method with proved convergence guarantees, introduced previously in [3, 4] for the system design of stochastic processes using the robustness of temporal properties.

The *Signal Spatio-Temporal Logic* (SSTL) [10, 24] is the extension of STL [23] with three spatial modalities, *somewhere*, *everywhere* and *surround*, which can be nested arbitrarily with the original STL temporal operator. In [10, 24], the authors provide a qualitative and quantitative semantics of SSTL and efficient monitoring algorithms for both semantics. A more detailed description of SSTL is provided in Section 3. While in this paper we adopt SSTL to specify spatio-temporal patterns, the overall method for robust parameter synthesis for stochastic reaction diffusion systems presented here can be performed also using SpaTeL.

**Contribution.** In this work, we combine formal methods with statistical machine learning by presenting a novel analysis of a stochastic model of the *spatio-temporal behaviour* of the Bicoid protein in the *Drosophila's* Embryo. The spatial gradient of this molecule has been shown to be at the basis of the subdivision of the embryo along its main axis, as specific concentration thresholds in its gradient are detected by cells and lead to the expression of distinct set of target genes.

The main technical contribution of the paper is the combination of SSTL within the statistical machine learning framework of [4, 7–9], in order to efficiently perform parameter space exploration and system design of spatio-temporal properties.

From a system biology perspective, instead, we present a detailed spatio-temporal analysis of the French Flag pattern on the gradient of the Bicoid protein. This analysis permits novel insights as to how the various model parameters interact to give rise to the patterning behaviour.

**Paper structure.** The rest of the paper is organised as follows. In Section 2 we discuss the spatial pattern formation in the *Drosophila* embryos. In Section 3 we first recall the syntax and semantics of SSDL and then use it to specify the French Flag Property. The smoothed model checking and the parameter estimation is presented in Section 4. In Section 5, we present the results and we conclude with final remarks and directions for future work in Section 6.

## 2 Spatial Pattern Formation and the French Flag Model

In this section, we describe a model of segmentation in *Drosophila melanogaster* and the spatio-temporal pattern characterising it, known as the French Flag model.

### 2.1 Pattern formation and reaction-diffusion systems

Patterning is a ubiquitous feature of biological organisms, and the presence of regular geometric motifs on many organisms has long fascinated scientists. Pattern formation is also the subject of one of the earliest, and most influential, computational systems biology works, Alan Turing’s pioneering work on morphogenesis [28]. Turing’s insight was that biological patterns can be viewed as emergent behaviour (in modern terminology) arising from local interactions of microscopic agents. More precisely, Turing considered spatially distributed systems whose local concentration vector  $\mathbf{u}$  obeys a *reaction-diffusion* partial differential equation (PDE)

$$\frac{\partial \mathbf{u}}{\partial t} = D \nabla^2 \mathbf{u} + f(\mathbf{u}). \quad (2.1)$$

Equation (2.1) defines the time evolution of the local concentration  $\mathbf{u}$  as the sum of two terms: a dispersal or diffusion term  $D \nabla^2 \mathbf{u}$ , which globally drives the system towards a uniform equilibrium, and a reaction term  $f(\mathbf{u})$ , which accounts for local interactions of the chemicals. Turing then proved that, under certain conditions on the reaction/ diffusion parameters, these two counteracting processes could give rise to regular patterns of concentration, providing a plausible mechanistic model of biological pattern formation.

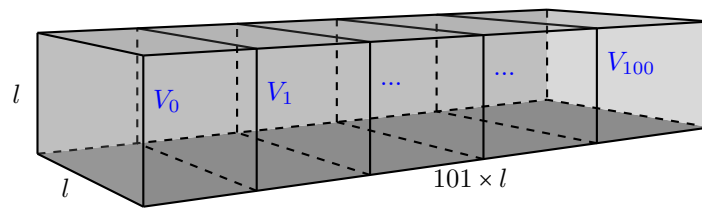
Turing’s ideas have been empirically demonstrated in many areas of biochemistry (see [22] for a recent review), and are still influential in particular in the field of developmental biology (see e.g. [16] for a recent paper building on these ideas). The crucial idea in the application of reaction-diffusion systems to development is that these mechanisms would underpin the local concentration patterns of regulatory proteins, which would instruct different genetic programs to be executed at different spatial locations. These special regulatory proteins are called *morphogens* in developmental biology, as they are believed to be responsible for the establishment of the shape of an organism in higher organisms. One of the most widely studied models of morphogenesis is the

establishment of spatial patterning (stripes) along the body of the fruit fly *Drosophila melanogaster*. Several morphogens are known in *Drosophila*; mostly, these are maternal proteins that are produced in a localised area of the embryo (in correspondence to a maternal deposit of messenger RNA), and then establish a concentration gradient during development, effectively providing cells within an embryo with a spatial reference. An important morphogen is the protein *Bicoid*, which is the central object of study in this paper and is described in detail in the next subsection.

Before closing this whirlwind review of developmental biology, it is worth remarking on a fundamental shift of perspective that has happened since Turing's pioneering work, the realisation of the importance of stochasticity in biology. Numerous lines of evidence indicate that biology at the single cell level is intrinsically stochastic. Stochasticity cannot be ignored when modelling early embryogenesis, when only a handful of cells are present. Morphogenetic reaction-diffusion models can therefore be modified to account for the intrinsic discreteness of biology at the microscopic level. The natural analogue, systems of agents moving in continuous space, is however prohibitively expensive computationally; an approach that is more amenable to analysis is to discretise space into a number of cells (voxels) which are assumed to be spatially homogenous, and to replace spatial diffusion with transitions between different cells. Morphogenetic systems, and in particular the *Bicoid* system, have already been analysed from a simulation perspective in [31] and from a statistical perspective in [12]. In this paper, we present a first analysis of this system from the point of view of (spatio) temporal logic, to analyse directly the system's behaviour at the level of the emergent properties of the trajectories.

## 2.2 The Bicoid Gradient

The *Bicoid* (Bcd) molecule was the first protein to be identified among the morphogens. In the *Drosophila* embryos, the Bcd protein is distributed along the Anterior-Posterior axis (A-P axis). The Bcd mRNA is translated at the anterior pole of the embryo, and the synthesised protein spreads through the A-P axis by *diffusion* accompanied by *decay*.

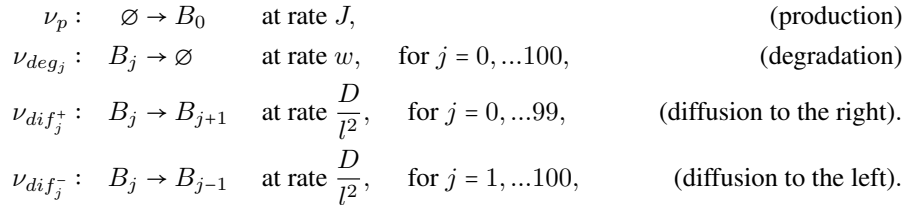


**Fig. 1.** A schematisation of the *Drosophila* embryo volume. The volume is divided in 101 cubic subvolumes,  $V_0, \dots, V_{100}$ , with side  $l = 5\mu m$ .

We will describe the dynamics of the Bcd protein by a stochastic *reaction-diffusion* system, as reported in [31]. Given a certain volume where the Bcd protein is distributed,

we can divide it into a series of subvolumes or voxels that are small enough to be regarded as well mixed. Then, we can consider the *decay* reaction as a transition that happens inside the subvolumes and the *diffusion* as exchange of molecules between neighbouring voxels. In particular, we consider 101 homogeneous cubic subvolumes with side  $l = 5\mu m$  that comprise the entire volume as in Fig. 1. The length of the side  $l$  and the number of subvolumes were chosen in light of those of actual *Drosophila* embryos, which are  $500\mu m$  long. The first subvolume ( $j = 0$ ), corresponds to the anterior pole of the embryo and it is the only subvolume where the Bcd protein is synthesised.

We can describe the set  $\mathcal{R}$  of reactions governing the stochastic dynamics of Bcd as:



where  $B_j$  is a Bcd protein in the  $j$ th subvolume.

The state vector of the system is then  $\mathbf{x}_B = (x_{B_0}, \dots, x_{B_{100}})$  where  $x_{B_j}$  is the number of Bcd molecules in the  $j$ th subvolume. From the set  $\mathcal{R}$  we can derive the infinitesimal generator matrix of the CTMC that formally represents the dynamics of the system. The CTMC can then be simulated with a standard algorithm, like SSA or tau-leaping.

Note that, from the set of reactions  $\mathcal{R}$ , we can easily revert the discretisation process and obtain a semantics in terms of Reaction-Diffusion Rate Equation (RDRE). This is obtained by converting variables into concentrations, taking the length of voxels to zero, and interpreting each rate as a flow, both in the degradation and in the diffusion reactions. In this way, we can define the system

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial y^2} - wu, \tag{2.2}$$

where  $u(y, t)$  is the concentration of Bcd at time  $t$  in position  $y$ , measured in  $\mu m$ ,  $y \in [0, 500]$ , giving the boundary conditions  $\frac{\partial u}{\partial y}|_{y=0} = -\frac{J}{\Delta}$  and  $\frac{\partial u}{\partial y}|_{y=500} = 0$ , where  $\Delta = l^3$ .

### 2.3 Segmentation and the French Flag model

The spatial distribution of the Bicoid protein has a crucial role in the formation of the horizontal segmentation in the development of the *Drosophila*'s embryo. One of the most important interpretations of this distribution is given by the *French Flag model* [29], and more generally by the theory of *gap genes* [19, 30]. The body of the fruit fly *Drosophila melanogaster*, as in most arthropods, exhibits a particular type of spatial patterning called *segmentation*, whereby the main body is composed of several segments. Gap genes were discovered and named following mutagenetic experiments, whereby biologists observed that deletion of certain genes resulted in the omission of a segment

in the fly’s body, as if the mutant organism had a gap. This observation implies that gap genes must be expressed in a precisely spatially co-ordinated manner, i.e., the biochemistry of the fruit fly must possess a way of measuring distances.

The French Flag model is a simplified model of gap gene regulation in early embryogenesis involving only four genes, the Bicoid morphogen protein and three target genes. The underlying assumption is that the spatial distribution of Bicoid protein, which as we have seen tends to decrease along the A-P axis (see Fig. 2), provides the ruler with which the *Drosophila* embryo measures distances. Gap genes are activated in a concentration dependent manner by Bicoid, so that a set of genes are activated at the high concentrations near the anterior part of the embryo (the blue in the French Flag), a different set of genes is activated in the central part (the white) and a third set is activated a low concentrations near the posterior end (red). This model has survived with some modifications [20] until this day, its beauty providing a paradigm for pattern development in many areas of biology. From our point of view, this model is particularly interesting because it refocuses attention from local intensive quantities (local concentrations) towards the importance of a global emergent property of the system (the establishment of a gradient), which is ideally suited for reasoning upon in terms of spatio-temporal logics. We will see in the next section this how to describe the French Flag pattern using a spatio-temporal logic.

### 3 Formula specification of Spatio-Temporal Behaviour

In this section, we describe *Signal Spatio-Temporal Logic* (SSTL) which will then be used to specify the spatio-temporal behaviour of the French Flag pattern.

#### 3.1 Signal Spatio-Temporal Logic

The *Signal Spatio-Temporal Logic* (SSTL) [10,24] is a linear time logic suitable to specify spatio-temporal behaviours of traces generated from simulations. It is an extension of *Signal Temporal Logic* (STL) [23] with two spatial modalities.

The space is described as a weighted graph  $G = (L, E, w)$  where  $L$  is a set of locations,  $E$  is a set of edges and  $w : E \rightarrow \mathbb{R}_{\geq 0}$  is the function that returns the cost/weight of each edge, typically encoding the distance between two nearby locations.

The syntax of SSTL is given by

$$\varphi := \text{true} \mid \mu \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \mathcal{U}_I \varphi_2 \mid \diamond_{[w_1, w_2]} \varphi \mid \varphi_1 \mathcal{S}_{[w_1, w_2]} \varphi_2,$$

where the STL operators are the atomic proposition  $\mu$ , the standard boolean connectives conjunction and negation and the *bounded until* operator  $\mathcal{U}_I$ , with  $I$  a dense-real interval. The new spatial operators are the *somewhere* operator,  $\diamond_{[w_1, w_2]}$ , and the *bounded surround* operator  $\mathcal{S}_{[w_1, w_2]}$ , where  $[w_1, w_2]$  is a closed real interval with  $w_1 < w_2$ . 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 surround formula  $\varphi_1 \mathcal{S}_{[w_1, w_2]} \varphi_2$ , instead, is true in a location  $\ell$  when  $\ell$  belongs to a subset of locations  $A$ , a region, satisfying  $\varphi_1$ , such that its

external boundary  $B^+(A)$  (i.e., all the nearest neighbours of locations in  $A$ ) contains only locations satisfying  $\varphi_2$ . Furthermore, locations in  $B^+(A)$  must be reached from  $\ell$  by a shortest path of cost between  $w_1$  and  $w_2$ , i.e. they have to be at distance between  $w_1$  and  $w_2$  from  $\ell$ . There are also three derivable operators: the *eventually* operator  $\mathcal{F}_I \varphi := \text{true} \mathcal{U}_I \varphi$ , the *always* operator  $\mathcal{G}_I \varphi := \neg \mathcal{F}_I \neg \varphi$  and the *everywhere* operator  $\boxplus_{[w_1, w_2]} \varphi := \neg \boxtimes_{[w_1, w_2]} \neg \varphi$  that requires  $\varphi$  to hold in all the locations reachable from the current one with a total cost between  $w_1$  and  $w_2$ .

SSTL is interpreted on spatio-temporal traces  $\mathbf{x} : \mathbb{T} \times L \rightarrow \mathbb{R}^n$ , where  $\mathbb{T}$  is the time domain, usually a real interval  $[0, T]$ , with  $T > 0$ ; we can write the trace as  $\mathbf{x}(t, \ell) = (x_1(t, \ell), \dots, x_n(t, \ell))$ , where each  $x_i : \mathbb{T} \times L \rightarrow \mathbb{R}$ , for  $i = 1, \dots, n$ , is the projection on the  $i^{\text{th}}$  coordinate/variable.

Similarly to STL, SSTL has two semantics, the classical *boolean* semantics and a *quantitative* semantics.

The boolean semantics returns true or false depending on whether the trace satisfies the SSTL property, i.e.  $(\mathbf{x}, t, \ell) \models \varphi$  is true if and only if the trace  $\mathbf{x}(t, \ell)$  satisfies  $\varphi$ . By convention, the whole trace satisfies a property in location  $\ell$  iff it satisfies the property at time zero, i.e.  $(\mathbf{x}, \ell) \models \varphi \Leftrightarrow (\mathbf{x}, 0, \ell) \models \varphi$ .

The quantitative semantics, instead, returns a real value  $\rho(\varphi, \mathbf{x}, t, \ell)$  that quantifies the level of satisfaction of the formula by the trajectory  $\mathbf{x}$  at time  $t$  in location  $\ell$ . The absolute value  $|\rho(\varphi, \mathbf{x}, t, \ell)|$  can be interpreted as measure of the robustness of the satisfaction or dissatisfaction. Furthermore, the sign of  $\rho(\varphi, \mathbf{x}, t, \ell)$  is related to the truth of the formula: if  $\rho(\varphi, \mathbf{x}, t, \ell) > 0$ , then  $(\mathbf{x}, t, \ell) \models \varphi$ , and similarly if  $\rho(\varphi, \mathbf{x}, t, \ell) < 0$ , then  $(\mathbf{x}, t, \ell) \not\models \varphi$ . The definition of this quantitative measure  $\rho$  is based on [13, 14], and it is a reformulation of the robustness degree of [15]. In accordance with the boolean semantics, the quantitative value of the whole trace in location  $\ell$  is given by its value at time zero, i.e.  $\rho(\mathbf{x}, \ell) = \rho(\mathbf{x}, 0, \ell)$ .

SSTL is equipped with efficient monitoring algorithms for both the boolean and the quantitative semantics, whose description, together with a formalisation of the semantics, can be found in [10, 24].

### 3.2 The French Flag property

To describe the French Flag pattern we have first to define the trajectories that we want to characterise and its related graph.

Let consider a trace (a simulation)  $(\mathbf{x}_B(t))_{t \in [0, T]} = (x_{B_0}(t), \dots, x_{B_{100}}(t))_{t \in [0, T]}$  of the Bicoid model described in the previous section, where  $[0, T]$  is the time domain, with  $T > 0$ . We can transform the temporal trace in a spatio-temporal trajectory defining  $x_B : L \times [0, T] \rightarrow \mathbb{R}$  s.t.  $x_B(V_i, t) := x_{B_i}(t)$ , where  $L = \{V_0, \dots, V_{100}\}$  is the set of locations. The graph  $G = (L, E, w)$  of the system is a one-dimensional graph where each  $V_i$  is connected only to  $V_{i-1}$  and  $V_{i+1}$ , with  $w(V_i, V_{i+1}) = 1$ , i.e. all the edges have weight equal to 1. The weight between two arbitrary locations is given by the weight of the shortest path connecting them.

We can now use the logic to specify the French Flag model. As we described in Section 2, this pattern is used to represent the effect of a morphogen in the expression of different genes, i.e. to represent the correlation between the concentration of the morphogen and the activation or repression of other genes. In particular, the spatial

distribution of the morphogen, at the steady state, is divided in three regions: a blue, a white and a red region, as shown in Fig.2 (left), that activate different target genes.

We can describe this behaviour with the property

$$\psi_{flag} := \varphi_{blue} \wedge \varphi_{white} \wedge \varphi_{red} \quad (3.3)$$

$$\begin{aligned} \varphi_{blue} &:= \boxplus_{[0, w_{blue}]}(x_B > K_{blue} - h_{bw}) \\ \varphi_{white} &:= \boxplus_{[w_{blue}, w_{white}]}((x_B < K_{blue} + h_{bw}) \wedge (x_B > K_{white} - h_{wr})) \\ \varphi_{red} &:= \boxplus_{[w_{white}, w_{max}]}(x_B < K_{white} + h_{wr}) \end{aligned} \quad (3.4)$$

The verification of the formula is done in the location  $V_0$ .  $(x, V_0) \models \psi_{flag}$  iff it satisfies each subformulae  $\varphi_{blue}, \varphi_{white}, \varphi_{red}$ ;  $(x, V_0) \models \varphi_{blue}$  iff, in all the locations  $V_i$  s.t.  $w(V_0, V_i) \leq w_{blue}$ , the number of Bicoid molecules is higher than  $K_{blue} - h_{bw}$ , i.e.  $x_B > K_{blue} - h_{bw}$ . In a similar way we can describe  $\varphi_{white}$  and  $\varphi_{red}$ . The meaning of the property is that the spatial distribution of the Bicoid protein is divided in three regions, the blue, where the  $x_B > K_{blue} - h_{bw}$ , the white, where  $K_{blue} + h_{bw} > x_B > K_{white} - h_{wr}$ , and the red, where  $x_B < K_{white} + h_{wr}$ . Note that  $h_{bw}$  and  $h_{wr}$  parameters have the role to relax the thresholds that define different regions, to properly deal with noise in Bcd expression, we will discuss this point more in detail in the Section 5.1.

At steady state, the concentration of the Bicoid protein is exponentially distributed along the anterior-posterior (A-P) axis, with higher concentrations towards the anterior. We can identify the insurgence time of this pattern, and if it remains stable, combining the spatial property with temporal operators as follows:

$$\psi_{stableflag} := \mathcal{F}_{[T_{flag}, T_{flag} + \delta]}(\mathcal{G}_{[0, T_{end}]} \psi_{flag}) \quad (3.5)$$

$\psi_{stableflag}$  means that eventually, in a time between  $T_{flag}$  and  $T_{flag} + \delta$ , the property  $\psi_{flag}$  remains true for at least  $T_{end}$  time units.

## 4 Methodologies

The main objective of this work is to study the effects of the Bicoid parameters on the satisfaction of the French Flag property. Exhaustive parameter exploration is particularly expensive for the model in question, due to the high cost of stochastic simulation. In this section, we briefly introduce the methodologies that we use to perform parameter synthesis and model checking in presence of parametric uncertainty.

### 4.1 Smoothed Model Checking

The Smoothed Model Checking algorithm [7] relies on the characterisation of the satisfaction probability of a formula  $\varphi$  as a function of the parameters. Given a CTMC  $\mathcal{M}_\theta$ , whose transition rates depend on a set of parameters  $\theta$ , the satisfaction function of  $\varphi$  is defined as follows:

$$f(\theta) \equiv p(\varphi = \text{true} | \mathcal{M}_\theta)$$



It has been proven in [7] that, if the transition rates of  $\mathcal{M}_\theta$  depends smoothly on the parameters  $\theta$  and polynomially on the state of the system, then the satisfaction function of  $\varphi$  is a smooth function of the parameters.

The smoothed model checking approach leverages of the smoothness of the satisfaction function and transfers information across nearby parameter values. More specifically, we place a Gaussian Process (GP) prior over the space of possible functions, and we evaluate the satisfaction function for a set of parameter values. We then calculate the GP posterior under the light of these observations, which constitutes analytical approximation to the satisfaction function. This implies that we can estimate the satisfaction probability at any point in the parameter space with no additional cost.

The premise is that fewer samples are required to achieve a given level of accuracy. In the experiments of [7], it has been possible to accurately approximate the satisfaction function over a wide range of parameters using less than 10% of the simulation runs required to obtain the same result with exhaustive parameter exploration. This resulted in a decrease of the total analysis time nearly by 90%.

## 4.2 Robust Parameter Synthesis

The problem of robust parameter synthesis constitutes of identifying the model parameters that maximise the robustness of some desired property. According to the quantitative semantics of SSTL, the robustness value  $\rho(\varphi, \mathbf{x}, t, \ell)$  expresses the level of satisfaction of  $\varphi$  by a trajectory  $\mathbf{x}$  at time  $t$  in location  $\ell$ . Since trajectories are random for a stochastic system, we designate the robustness of  $\varphi$  for a CTMC as a random variable  $R_\varphi$ . We are therefore interested in maximising the expected quantitative score:

$$E[R_\varphi] = \int \rho(\varphi, \mathbf{x}, t, \ell) p(\mathbf{x}) d\mathbf{x} \quad (4.6)$$

where  $p(\mathbf{x})$  is the probability density of trajectory  $\mathbf{x}$ . For a specified time  $t$  and location  $\ell$ , the expectation  $E[R_\varphi]$  constitutes an objective function, for which we can obtain noisy estimates by generating samples from the trajectory space via stochastic simulation.

Since evaluating the expected robustness is computationally expensive, we employ the Gaussian process optimisation algorithm described in [9]. In short, the objective function is approximated by a *Gaussian Process* (GP). The algorithm is initialised with a random grid of points, for each of which  $E[R_\varphi]$  is approximated via statistical means. Using these points as a training set, a GP is used to make predictions regarding the  $E[R_\varphi]$  value at different parts of the search space, without exhaustive exploration of the parameter space. We calculate the GP posterior for a set of test points; that involves calculating an estimate of the expected robustness and its associated variance. The GP optimisation algorithm dictates that the point that maximises the an upper quantile of the GP posterior is added to the training set, after being evaluated for its associated robustness via SMC. A high value for the upper quantile at any point in the parameter space indicates the possibility of an undiscovered maximum nearby. This feature allows us to direct the search towards areas of the parameter space that appear to be more promising. This process is repeated for a number of iterations, and the training set is progressively updated with new potential maxima. For a smooth objective function, the algorithm is proved to converge to the global optimum in [27].

## 5 Results

In this section, we perform a series of experiments to explore the sensitivity and robustness of the French Flag property w.r.t. changes in the rates of production  $J$  and degradation  $w$ , and the diffusion rate parameter  $D$ . The size of the cubic subvolumes is known, that is  $l = 5\mu m$ , as it is one of the main modelling assumptions.

### 5.1 Experimental Data

Following [26, 31], we chose as parameters of the  $\psi_{stableflag}$  property (3.5), specified in Section 3.2,  $T_{flag} = 3950$ ,  $\delta = 10$ ,  $T_{end} = 1000$ ,  $w_{blue} = 35.5$ ,  $w_{white} = 67.5$  and  $w_{max} = 101$ . The  $w_{blue}$  and  $w_{white}$  parameters mean that the blue area involves the subvolumes between  $V_0$  and  $V_{35}$ , the white area extends from volume  $V_{36}$  to  $V_{67}$ , and finally the red one from  $V_{68}$  to  $V_{100}$ ; the time is in terms of seconds.

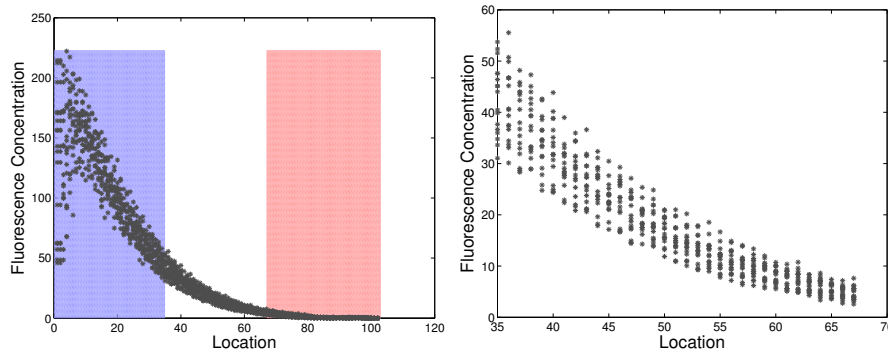
In order to fix the thresholds parameters  $K_{blue}$ ,  $K_{white}$  and  $h_{bw}, h_{wr}$  we use the Bicoid fluorescence concentration at cycle 13 (where the gradient is considered to be in the steady state) downloaded from the FlyEx database [1]. The choice of the data follows the analysis doing in [31]. To the best of our knowledge, all the quantifications of the Bicoid protein in the Drosophila embryo refers to the measurements of fluorescence concentrations, rather than direct observations of the Bicoid molecular population. From [31], we define the fluorescence concentration  $I = m \times x_B$ , where  $m$  is a scaling factor that denotes the fluorescence-to-molecule ratio. Our approach is to rescale the thresholds reported in terms of fluorescence concentrations with the  $m$  factor.

The data has been given originally in the form of two-dimensional coordinates paired, the A-P and D-V coordinate, from the central 10% strip. As in [31], we choose the embryos where the variation inside each spatial subregions is low, in particular in these embryos the inverse of the spatial exponential coefficient varied by less that 1%. We have transformed the data so that we have a single concentration value for each of the 101 discretised locations. Fig. 2 depicts the result. On the left-side figure, we see how the different locations lie within the areas prescribed by the French Flag property. Although the shape of the data is apparently negative exponential, there is a considerable amount of noise, which has to be taken into consideration in terms of the French Flag property. We therefore define the thresholds in the form regions, rather than strict values. On the right-side of Fig.2, we see a magnified version of the figure, where only the white area is depicted. The majority of the concentrations recorded for volumes from  $V_{36}$  to  $V_{67}$  are between 60 and 2. In the same way, we can empirically derive zones of desired concentration levels for the blue and read areas. Therefore we have  $K_{blue} = 45/m$ ,  $h_{bw} = 15/m$ ,  $K_{white} = 6/m$ , and  $h_{wr} = 4/m$ .

### 5.2 Optimisation of Expected Robustness

We now explore how the model parameters (including the scaling factor  $m$ ) can be tuned to increase the robustness of the French Flag pattern.

We applying the GP optimisation algorithm discussed in Section 4.2, for a four-dimensional space that involves the parameters:  $w \in [0.001, 0.01]$ ,  $J \in [10, 400]$ ,



**Fig. 2.** Left: Fluorescence concentrations of the Bicoid protein for 17 embryos during the cycle 13. Right: The same concentrations in the area between locations 35 and 67, which define the white area in the French flag property.

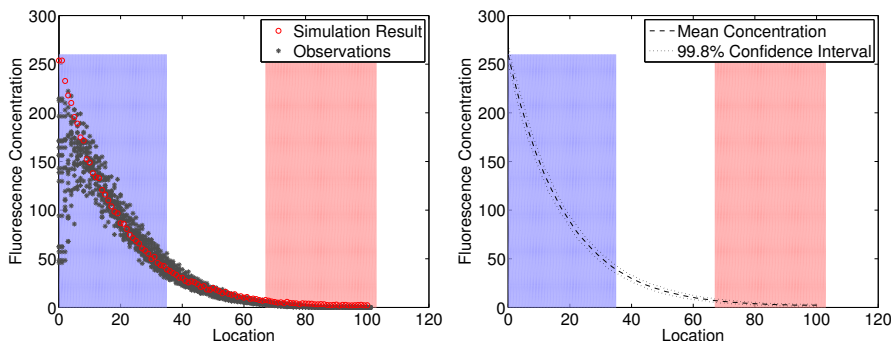
$D \in [1, 40]$ , and  $m \in [0.01, 1]$ . The parameter ranges have been selected so that the resulting space is a superset of the explored space in [31]. Regarding the fluorescence-to-molecule ratio in particular, we note that the extremes considered in [31] have been 0.07 and 0.7.

For each evaluation of the expected robustness, the system has been simulated up to time  $t = 4000$  sec, which is when the steady-state is approached according to [31]. The robustness expectation has been approximated statistically using 12 simulation runs for each parameter set. The algorithm has been initialised by 80 evaluations of the objective function at random points; a number of 282 evaluations were performed at points selected by the optimisation process, until convergence was detected. Convergence has been determined when no significant improvement of the expected robustness has been observed for 200 iterations. An improvement is considered significant, if it is more than 1% increase over the previously recorded maximum robustness.

In the end, a total of 362 function evaluations have been performed, which is arguably a small number of samples to explore a four-dimensional space. The execution times have been 85 minutes for the initial 80 evaluations, and 263 minutes for the actual optimisation process. Stochastic simulations have been performed in parallel using 12 threads. The experiments have been performed on an Intel® Xeon® CPU E5-2680 v3 2.50GHz. The majority of the computational effort was spent in simulation, despite the fact that only 12 trajectories have been generated for each parameter set considered. Therefore the idea of reducing the number of samples by exploiting the smoothness of the objective function has been a sensible practice.

The values returned by the optimisation process have been:  $w^* = 0.0038$ ,  $J^* = 390$ ,  $D^* = 32.5$ , and  $m^* = 0.048$ . The robustness of the optimum returned has been 2.99, implying that the property is robustly satisfied for the given solution. In Fig. 3, we present a sample trajectory for the given parameter configuration, and the average of 40 random trajectories, along with the associated 99.8% confidence bounds. The sample trajectory is plotted against the experimental data that were used to adjust the

threshold parameters of the French Flag property. We see that the optimised model has a behaviour very similar to the one observed in real-world experiments. However, it appears that the simulation results are much less noisy, when compared to the actual observations. This finding is in agreement with the result of [31], where it was argued that the intrinsic noise as modelled by the stochastic dynamics of the master equation is not sufficient to explain the variability in the data, i.e. the noise in the fluorescence measurement as a crucial role that has to be taken into account.



**Fig. 3.** Left: Sample trajectory for the parameter configuration that maximises the robustness of the French Flag property. Right: Average of 40 random trajectories; the dotted lines indicate the 99.8% confidence interval.

### 5.3 Parameter Exploration with Smoothed Model Checking

In this section, we perform a more thorough exploration of the parameter space. Our objective is to discover dependencies among the parameters, considering the satisfaction probability of the French Flag property. On that respect, the fluorescence-to-molecule ratio  $m$  is not significant, as this will have an obvious effect on the thresholds for the property. We fix the fluorescence-to-molecule ratio  $m$  to 0.048, which is the optimal value reported by the optimisation algorithm in the previous section. The rest of the model parameters,  $w \in [0.001, 0.01]$ ,  $J \in [10, 400]$ , and  $D \in [1, 40]$ , are explored via the smoothed model checking approach.

During the initialisation step of the algorithm, we have performed 216 evaluations of the satisfaction function of (3.3), for a regularly distributed set of values. As in the previous section, the satisfaction probability is approximated by statistical model checking using 12 simulation runs for each parameter configuration, where the system is simulated up to time  $t = 4000$  sec.

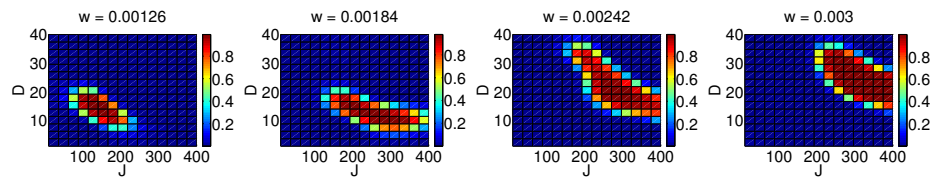
The duration of this initial statistical model checking process has been nearly 170 minutes, on an Intel® Xeon® CPU E5-2680 v3 2.50GHz, using 12 threads in parallel. The hyperparameter optimisation that is required to tune the GP probit regression model

subsequently required only 20 seconds, which is a trivial price to pay compared to the massive simulation cost. The final GP probit regression for a grid of 4096 points required only 1.2 seconds. Most importantly, it is only this last cost that we are required to pay to produce any further estimations of the satisfaction function.

Fig.4 depicts the satisfaction function for the French Flag property for parameters  $\theta = \{w, J, D\}$ , as this has been approximated by smoothed model checking. Each of the depicted subfigures shows the satisfaction probability as function of the production rate  $J$  and the diffusion parameter  $D$ , for a different value of the degradation rate  $w$ . Regarding the confidence of the estimated probabilities, we report that the 73.6% of the values are associated with 95% confidence intervals of width less than 0.2.

As a general remark, it appears that the manifestation of the gradient pattern, as this is captured by the French Flag property, is associated with a fine balance among the model parameters. There is a small area in the parameter space for which the property is satisfied with high probability. As we increase the decay parameter  $w$  however, we observe two behaviour regarding this area: its size is being increased, and its location is being shifted to the right. This implies that  $w$  is positively correlated with the production rate  $J$ . In other words, a particular ratio between protein production and decay is required for the formation of the particular pattern. At the same time, increasing the decay rate means that the formula may be satisfied for a wider range of the diffusion parameter.

It also appears that there is a negative correlation between the production rate  $J$  and the diffusion parameter  $D$ . This behaviour is present for the entire range of  $w$  examined, but it tends to become more obvious as  $w$  is increased. It is reasonable to conclude that a simultaneous increase of  $J$  and  $D$  would destroy the exponential shape of the Bicoid distribution across space.



**Fig. 4.** Emulated satisfaction probability of the French Flag property as function of  $\theta = \{w, J, D\}$ . Each subfigure has the  $w$  parameter fixed.

## 6 Conclusions

We present a framework for the formal analysis under parametric uncertainty and the robust parameter synthesis of spatio-temporal properties emerging in a stochastic reaction-diffusion system. These properties are specified using the spatio-temporal logic SSTL. The framework combines statistical machine learning techniques based on Gaussian processes with the algorithm for monitoring SSTL properties.

As a case study, we analyse the occurrence of the French Flag pattern in the Bicoid gradient, during the development of *Drosophila* embryo. Analysing how this property depends on the parameters of the model is challenging due to the very high computational cost of simulating a spatio-temporal model, and has only been possible by adopting recent efficient verification techniques that employ machine learning methodologies [8]. Furthermore, the combination of these new techniques with SSTL permits exploring behaviours that are extremely difficult to express (and monitor) with standard temporal logics, where each individual location would need to be accounted.

The natural extension of this work is the analysis of more complicated models and properties, for example adding to this model the proteins of the target genes related with the spatial distribution of the Bicoid protein, enabling the study of the spatial dependency between proteins. To be independent from the spatial approximation, we plan also to consider different discretisation of the *Drosophila*'s volume. Another future work could be the consideration of a model rescale with a random factor that mimics the extrinsic noise due to the fluorescence measurements. We plan also to extend our previous result in mining temporal logic properties [5, 11] for the spatio-temporal case. Finally, we are considering an extension of the logic to continuous spaces and we would like to compare the expressiveness of SSTL with SpaTeL.

*Acknowledgements.* L.B. acknowledges partial support from the EU-FET project QUANTICOL (nr. 600708) and by FRA-UniT.S. G.S. and D.M. acknowledge the support from the ERC under grant MLCS306999. E.B. acknowledges the partial support of the Austrian National Research Network S 11405-N23 (RiSE/SHiNE) of the Austrian Science Fund (FWF), the ICT COST Action IC1402 Runtime Verification beyond Monitoring (ARVI) and the IKT der Zukunft of Austrian FFG project HARMONIA (nr. 845631).

## References

1. Flyex database. <http://urchin.spbcas.ru/flyex/>
2. Aydin Gol, E., Bartocci, E., Belta, E.: A formal methods approach to pattern synthesis in reaction diffusion systems. In: Proc. of CDC 2014: the 53rd IEEE Conference on Decision and Control. pp. 108–113. IEEE (2014)
3. Bartocci, E., Bortolussi, L., Nenzi, L., Sanguinetti, G.: On the robustness of temporal properties for stochastic models. In: Proc. of HSB 2013: the Second International Workshop on Hybrid Systems and Biology. pp. 3–19 (2013)
4. Bartocci, E., Bortolussi, L., Nenzi, L., Sanguinetti, G.: System design of stochastic models using robustness of temporal properties. *Theoretical Computer Science*. pp. 3–25 (2015)
5. Bartocci, E., Bortolussi, L., Sanguinetti, G.: Data-driven statistical learning of temporal logic properties. In: Proc. of FORMATS 2014: the 12th International Conference on Formal Modeling and Analysis of Timed Systems. pp. 23–37 (2014)
6. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer (2006)
7. Bortolussi, L., Milios, D., Sanguinetti, G.: Smoothed model checking for uncertain continuous time Markov chains. CoRR ArXiv 1402.1450 (2014)
8. Bortolussi, L., Milios, D., Sanguinetti, G.: U-check: Model checking and parameter synthesis under uncertainty. In: Proc. of QEST 2015: 12th International Conference on Quantitative Evaluation of SysTems. LNCS, vol. 9259, pp. 89–104 (2015)

9. Bortolussi, L., Sanguinetti, G.: Learning and designing stochastic processes from logical constraints. In: Proc. of QEST 2013: 10th International Conference on Quantitative Evaluation of SysTems. LNCS, vol. 8054, pp. 89–105 (2013)
10. Bortolussi, L., Nenzi, L.: Specifying and monitoring properties of stochastic spatio-temporal systems in signal temporal logic. In: Proc. of VALUETOOLS 2014: the 8th International Conference on Performance Evaluation Methodologies and Tools. pp. 66–73. ICST (2014)
11. Bufo, S., Bartocci, E., Sanguinetti, G., Borelli, M., Lucangelo, U., Bortolussi, L.: Temporal logic based monitoring of assisted ventilation in intensive care patients. In: Proc. of ISOLA 2014: the 6th International Symposium on Leveraging Applications of Formal Methods, Verification and Validation. Part II. pp. 391–403 (2014)
12. Dewar, M.A., Kadirkamanathan, V., Opper, M., Sanguinetti, G.: Parameter estimation and inference for stochastic reaction-diffusion systems: application to morphogenesis in *d. melanogaster*. BMC Systems Biology. p. 21 (2010)
13. Donzé, A., Maler, O.: Robust satisfaction of temporal logic over real-valued signals. In: Proc. of FORMATS 2010: the 8th International Conference on Formal Modeling and Analysis of Timed Systems. LNCS, vol. 6246, pp. 92–106. Springer (2010)
14. Donzé, A., Ferrère, T., Maler, O.: Efficient robust monitoring for STL. In: Proc. of CAV 2013, the 25th International Conference on Computer Aided Verification. LNCS, vol. 8044, pp. 264–279. Springer (2013)
15. Fainekos, G.E., Pappas, G.J.: Robustness of temporal logic specifications for continuous-time signals. Theor. Comput. Sci. pp. 4262–4291 (2009)
16. Fried, P., Iber, D.: Dynamic scaling of morphogen gradients on growing domains. Nat. Comm. (2014)
17. Grosu, R., Smolka, S.A., Corradini, F., Wasilewska, A., Entcheva, E., Bartocci, E.: Learning and detecting emergent behavior in networks of cardiac myocytes. Commun. ACM pp. 97–105 (2009)
18. Haghghi, I., Jones, A., Kong, Z., Bartocci, E., Grosu, R., Belta, C.: SpaTeL: A novel spatial-temporal logic and its applications to networked systems. In: Proc. of HSCC '15: the 18th International Conference on Hybrid Systems: Computation and Control. pp. 189–198. ACM (2015)
19. Jaeger, J.: The gap gene network. Cellular and Molecular Life Sciences pp. 243–274 (2010)
20. Jaeger, J., Martinez-Arias, A.: Getting the Measure of Positional Information. PLoS Biology (2009)
21. Jain, A., Duin, R., Mao, J.: Statistical pattern recognition: A review. IEEE Transactions on pattern analysis and machine learning. pp. 4–37 (2000)
22. Maini, P.K., Woolley, T.E., Baker, R.E., Gaffney, E.A., Lee, S.S.: Turing’s model for biological pattern formation and the robustness problem. Interface focus (2012)
23. Maler, O., Nickovic, D.: Monitoring temporal properties of continuous signals. In: Proc. of FORMATS/FTRTFT 2004: joint International Conferences on Formal Modeling and Analysis of Timed Systems and Formal Techniques in Real-Time and Fault-Tolerant Systems. LNCS, vol. 3253, pp. 152–166 (2004)
24. Nenzi, L., Bortolussi, L., Ciancia, V., Loreti, M., Massink, M.: Qualitative and quantitative monitoring of spatio-temporal properties. In: Proc. of RV 2015: the 6th International Conference on Runtime verification. LNCS, vol. 9333, pp. 21–38 (2015)
25. Pavlidis, T.: Structural Pattern Recognition. Springer-Verlag (1980)
26. Phillips, R., Kondev, J., Theriot, J.: Physical biology of the cell. Garland Science (2009)
27. Srinivas, N., Krause, A., Kakade, S., Seeger, M.: Information-theoretic regret bounds for Gaussian process optimisation in the bandit setting. IEEE Trans. Inf. Th. pp. 3250–3265 (2012)
28. Turing, A.: The chemical basis of morphogenesis. Phil. Trans. R. Soc. Lond. pp. 37–72 (1952)

29. Wolpert, L.: The french flag problem: A contribution to the discussion on pattern development and regulation. *Towards a theoretical biology*. pp. 125–133 (1968)
30. Wolpert, L., Tickle, C., Arias, A.M.: *Principles of development*. Oxford university press (2015)
31. Wu, Y.F., Myasnikova, E., Reinitz, J.: Master equation simulation analysis of immunostained bicoid morphogen gradient. *BMC systems biology*. p. 52 (2007)