





# Understanding Social Feedback in Biological Collectives with Smoothed Model Checking

Julia Klein<sup>1,2</sup>  and Tatjana Petrov<sup>1,2</sup> 

<sup>1</sup> Department of Computer and Information Sciences, University of Konstanz,  
Konstanz, Germany

[julia.klein@uni-konstanz.de](mailto:julia.klein@uni-konstanz.de)

<sup>2</sup> Centre for the Advanced Study of Collective Behaviour, University of Konstanz,  
78464 Konstanz, Germany

**Abstract.** Biological groups exhibit fascinating collective dynamics without centralised control, through only local interactions between individuals. Desirable group behaviours are typically linked to a certain fitness function, which the group robustly performs under different perturbations in, for instance, group structure, group size, noise, or environmental factors. Deriving this fitness function is an important step towards understanding the collective response, yet it easily becomes non-trivial in the context of complex collective dynamics. In particular, understanding the social feedback - how the collective behaviour adapts to changes in the group size - requires dealing with complex models and limited experimental data. In this work, we assume that the collective response is experimentally observed for a chosen, finite set of group sizes. Based on such data, we propose a framework which allows to: (i) predict the collective response for any given group size, and (ii) automatically propose a fitness function. We use Smoothed Model Checking, an approach based on Gaussian Process Classification, to develop a methodology that is scalable, flexible, and data-efficient; We specify the fitness function as a template temporal logic formula with unknown parameters, and we automatically infer the missing quantities from data. We evaluate the framework over a case study of a collective stinging defence mechanism in honeybee colonies.

**Keywords:** Social feedback · Gaussian processes · Biological collectives · Smoothed model checking

## 1 Introduction

Biological groups exhibit fascinating collective dynamics without centralised control, through only local interactions between individuals. Quantitative models of the mechanisms underlying biological grouping can directly serve important

societal concerns (for example, prediction of seismic activity [22]), inspire the design of distributed algorithms (for example, ant colony algorithm [9]), or aid robust design and engineering of collective, adaptive systems under given functionality and resources, which is increasingly gaining attention in vision of smart cities [17, 21]. Quantitative prediction of the behaviour of a population of agents over time and space, each having several behavioural modes, results in a high-dimensional, non-linear, and stochastic system [12]. Computational modelling with population models is therefore challenging, especially since model parameters are often unknown and repeated experiments are costly and time-consuming.

In this paper, we focus on the phenomenon of collective social feedback in biological groups, that is, how the collective behaviour adapts to changes in the group size. Examples of social adaptation include the emergence of sensing abilities through interactions and only exist at the group level [3], or, colony defence [19] or thermoregulation [8] in social insects (as altruistic behaviours that do not occur in isolated individuals), to name but a few. Understanding such social adaptation cannot be done by extrapolating from observing individuals in isolation. Computationally, the challenge of understanding how social context shapes group behaviours emerges at two levels. First, models of group-behaviours enumerating each possible social context of an individual suffer from the combinatorial explosion of states, but also from a prohibitive number of model parameters. With no simplifying assumptions, an individual within a group of size  $n$  adapts to at least  $n$  different social contexts that need to be parametrized [14, 27]. While simplifying assumptions are justified for some experimental systems, they generally need to be validated for each experimental system at hand. For instance, in modelling molecular dynamics with chemical reaction networks, mass-action law assumes a linear influence of reactants' counts to reaction propensities, but this is not justified in case of animal collectives, due to a richer cognitive aspect of individuals. Second, while experimentally measuring the overall group response is significantly simpler than measuring the response of each individual within a group via continuous tracking, it still remains impossible to measure the group response for *each* group size; Instead, one must choose a set of representative group sizes. In other words, in order to find a general pattern of behaviours, it is necessary to analyse groups of many different sizes, both of small and large scale.

For the above reasons, it becomes important to develop methods that are scalable with respect to growing group size, flexible in terms of model size and parameters, and data-efficient - producing reliable results for scarce data sets. Our methodology relies on *Gaussian Processes*, a powerful Bayesian approach in Machine Learning, to learn unknown functions from data. Gaussian Processes, considered a “desired meta-model in various applications” [7], fulfil our requirements of scalability, flexibility, and data-efficiency. In addition, in contrast to other Machine Learning models, Gaussian Processes deal not only with uncertainty in the training data, but also provide guarantees of the predictions in form of credible intervals.

The contributions of this work are as follows. We assume that the collective response is experimentally observed for a chosen, finite set of group sizes. Based on such data, we propose a framework which allows to: (i) predict the collective response for any given group size, and (ii) automatically propose a fitness function that is robustly preserved under perturbations in group size. We use Gaussian Process Regression for task (i), allowing to overcome the need of conducting new experiments and analysing many large models, but still having an informed estimate of the group response. Second, we apply Smoothed Model Checking [5], a novel technique based on Gaussian Process Classification, for task (ii) to derive the fitness function a collective robustly performs by setting up a template formula and inferring the missing quantity from data to understand the social feedback mechanism of the collective. An illustrative example of the developed methods in context of elucidating social feedback in collectives is provided in Sect. 1.1. Finally, we test and evaluate the proposed methods over a real-world case study with social honeybees.

**Related Work.** The framework we present here is specifically inspired by the application of collective defence in honeybee colonies. Honeybees protect their colonies against vertebrates by releasing an alarm pheromone to recruit a large number of defenders into a massive stinging response [25]. However, these workers will then die from abdominal damage caused by the sting tearing loose [30]. In order to achieve a balanced trade-off towards efficient defence, yet no critical worker loss, each bee’s response to the same amount of pheromone may vary greatly, depending on its social context. Our own related works [14,27] focus on extracting individual behaviours from group-level data, by hypothesising a mechanistic behavioural model and developing suitable methods for parameter inference. Here, instead, we also assume that group-level data is available, but we provide a model-free methodology with different aims - predicting the group response, and inferring the group-level fitness function. To the best of our knowledge, our method is the first application of Smoothed Model Checking towards understanding collective animal behaviour.

Methodologically, our work is inspired by the general technique of Smoothed Model Checking [5] (SMMC), implemented in the software tool U-Check [4]. SMMC was used for several applications in systems and synthetic biology. Bartocci et al. [1] propose a notion of robustness degree for stochastic biochemical circuits; They furthermore show how such robustness measure can be used for designing a biochemical circuit robustly exhibiting a certain temporal property; Specifically, the design goal is that a specific behaviour of a biological process is maintained despite uncertainties from noise or uncertain model parameters. Instead of computing the robustness degree of each sample trajectory of a system, in this work we focus on measuring the satisfaction only on steady state data and evaluate the robustness over the satisfaction distribution across different group sizes. In [2], the proposed notion of robustness is used so to optimise certain control parameters of a stochastic model to maximise the robustness of desired specifications. In [6], the authors show how to learn and design continuous-time

Markov chains (CTMCs) from observations formulated in terms of temporal logic formulae; They maximise the likelihood of parameters with the Gaussian Process Upper Confidence Bound (GP-UCB) algorithm. In contrast to the previously mentioned works, we consider a model-free approach and aim to infer a general description of the collective response based only on experimental data. Hence, we do not analyse varying model parameters, but different group sizes. In reference to inferring the fitness function, [18] propose how to infer parameters given a requirement template expressed in signal temporal logic (STL) and simulation traces of the model; The approach is based on finding the tightest valuation function to prevent getting overly conservative parameters. Our method of finding the parameter of a given template differs in selecting the value not according to the tightest valuation function, but according to a measure of variation.

### 1.1 Illustrative Example

We next illustrate how the methodology we develop can be used to study social influence in groups, i.e., how the group size affects the behaviour of individuals. Assume a group of  $n$  identical agents in which each individual is confronted with a task and either solves it successfully or fails, with certain probability. We further assume that, each time an agent in a group succeeds, other individuals in the group are more likely to succeed. Specifically, if the baseline probability of success is  $p_0$ , assume that the probability of succeeding with  $i$  already successful agents in the system grows with the number of successful agents according to a function  $p_i = f(p_0, \alpha, i)$  (simple examples could be  $p_i = p_0 + \alpha \cdot (i > 0)$ , where the probability increases by  $\alpha$  if at least one other agent in the group succeeded, or  $p_i = p_0 + \alpha \cdot i$ , where the probability increases linearly with the number of other successful agents). Now, if measurements are available for groups of size 1, 2, and 10 ( $n \in \{1, 2, 10\}$ ), inferring parameters  $p_0$  and  $\alpha$  clearly becomes possible from only measurements over isolated individuals (groups of size  $n = 1$ ), and pairs of interacting individuals (groups of size  $n = 2$ ). These two parameters, coupled with an underlying mechanistic model of interaction, would allow to predict the outcome for  $n = 10$ . Finally, if model-based predictions for  $n = 10$  significantly differ from the experimental data for  $n = 10$  – the increment parameter  $\alpha$  differs significantly for groups of size 2 and 10 – we can conclude that the agent is aware of its social context and there is a feedback mechanism from the group that influences the individual’s behaviour. Otherwise, if there is no significant difference, one may conclude that there is no influence of group size to the problem solving efficiency.

The methodology we develop here allows to predict group outcomes for any group size, with uncertainty quantification, when group measurements are available for only certain group sizes. In the context of our illustrating example, this means that one could predict measurements for groups of e.g. size  $n = 3$ , from measurements for  $n \in \{1, 2, 10\}$ . Then, the hypothesis of social feedback can be accessed by only making predictions for the model for group with  $n = 3$  agents,

that is dramatically smaller than the model for  $n = 10$  agents (due to combinatorial explosion of the state space, models for  $n$  agents would be described by  $O(2^n)$  states).

Furthermore, assume that the group aims to satisfy a certain group outcome, independently of its size. In the above example, such function may be that ‘eventually, between 40% and 60% of group members succeed at solving the task’. Inferring such *fitness* function - a high-level behavioural outcome that tends to be robustly preserved under environmental perturbations - is of high importance for a biological understanding of grouping. While the qualitative form of the fitness function is often assumed by experts, quantitative parameters (e.g. the range from 40% to 60%) are typically not explored in an automated way. To this end, our second methodological contribution is automatising the search for such a fitness function from only available data measurements and a template logical formula.

## 2 Methods

In this section, we present the methodology based on Gaussian Processes to understand social feedback mechanisms in biological collectives. First, we describe the theoretical and mathematical background of the methods and subsequently we demonstrate how to apply these existing techniques in our framework to address the previously stated research problems. All definitions of Gaussian Processes and the corresponding regression and classification follow closely the description by Rasmussen and Williams [29].

### 2.1 Gaussian Process

A Gaussian Process (GP) is a generalisation of a multivariate Gaussian distribution to infinite dimension. As a non-parametric distribution over a space of functions, a GP is designed to solve regression and classification problems by approximating unknown functions. Since a GP model is data-driven, applicable without specifying the underlying distribution beforehand, and powerful even for little data, it surpasses many of the traditional regression and classification methods. The predictions of a GP are probabilistic, such that the results provide not only an estimate, but additionally a quantification of uncertainty in form of credible intervals.

Mathematically, we define a prior probability distribution directly over functions, from which the posterior distribution can be inferred when data is observed. A kernel-based probabilistic model is set up to learn relationships between observed data and make predictions about new data points. In general, a GP’s characteristics are completely specified by a mean function  $m(x)$  and a positive definite kernel function  $k(x, x')$  for input values  $x, x' \in \mathbb{R}$ . Kernels are used as a similarity measure between data points and generate the covariance matrix  $\Sigma$  by evaluating the kernel function at all pairs of input points.

We denote matrices by capitalised letters and vectors in bold type. A subscript asterisk refers to a test set quantity.

## 2.2 Gaussian Process Regression

We define a GP prior  $f(x) \sim \mathcal{GP}(m(x), k(x, x'))$  independent of any training data that specifies some properties of the unknown functions through the choice of the kernel function. Three of the most common kernel functions are implemented in our framework:

- *Linear kernel*:  $k_{lin}(x, x') = \sigma_b^2 + \sigma^2(x-c)(x'-c)$  with variance  $\sigma_b$ , scale factor  $\sigma^2$  and offset  $c$ ,
- *Radial Basis Function (RBF)*:  $k_{rbf}(x, x') = \sigma^2 \exp(-\frac{\|x-x'\|^2}{2\ell^2})$  with scale factor  $\sigma^2$  and lengthscale  $\ell$ , and
- *Periodic kernel*:  $k_{per}(x, x') = \sigma^2 \exp(-\frac{2 \sin^2(\frac{\pi\|x-x'\|}{\ell p})}{\ell^2})$  with scale factor  $\sigma^2$ , periodicity parameter  $p$  and lengthscale  $\ell$ .

Beyond that, all kernels are pairwise combined by addition and multiplication to achieve higher-level structures [13].

Let  $X$  be the training data set with observed function values  $\mathbf{f}$ , and  $X_*$  the test data set for which we want to predict the corresponding function outputs  $\mathbf{f}_*$ . The joint distribution of training and test outputs is given by

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} \Sigma & \Sigma_* \\ \Sigma_*^T & \Sigma_{**} \end{bmatrix} \right), \quad (1)$$

where  $\mu = m(x_i)$ ,  $i = 1, \dots, n$ , denotes the training mean values and analogously  $\mu_*$  the test mean values. The covariance matrix  $\Sigma$  is evaluated at all pairs of training points,  $\Sigma_*$  at training and test points, and  $\Sigma_{**}$  at test points. The posterior distribution is obtained by conditioning the joint Gaussian prior distributions on the observations:

$$\mathbf{f}_* | X_*, X, \mathbf{f} \sim \mathcal{N}(\Sigma_*^T \Sigma^{-1} \mathbf{f}, \Sigma_{**} - \Sigma_*^T \Sigma^{-1} \Sigma_*). \quad (2)$$

By evaluating the mean and covariance we derive the function values  $\mathbf{f}_*$  from the posterior distribution. Computing two times the standard deviation of each test point around the mean generates 95% credible regions.

Normally distributed observational noise can be considered in the training data,  $y = f(x) + \epsilon$  with  $f \sim \mathcal{GP}(0, \Sigma)$  and  $\epsilon \sim \mathcal{N}(0, \sigma_f^2 I)$ . The noise variance  $\sigma_f^2$  is independently added to each observation,  $p(y|f) = \mathcal{N}(y|f, \sigma_f^2 I)$ , what changes the joint distribution of training and test values to

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left( 0, \begin{bmatrix} \Sigma_y & \Sigma_* \\ \Sigma_*^T & \Sigma_{**} \end{bmatrix} \right) \quad (3)$$

with  $\Sigma_y := \Sigma + \sigma_f^2 I$ . Deriving the posterior distribution results in:

$$\mathbf{f}_* | X_*, X, \mathbf{y} \sim \mathcal{N}(\Sigma_*^T \Sigma_y^{-1} \mathbf{y}, \Sigma_{**} - \Sigma_*^T \Sigma_y^{-1} \Sigma_*). \quad (4)$$

Each kernel has a number of hyperparameters that specify the precise shape of the covariance function. Optimising the kernels' hyperparameters increases the

accuracy of predictions. As standard practice, we follow an empirical Bayesian approach to maximise the log marginal likelihood

$$\log p(y|X) = \log \mathcal{N}(y|0, \Sigma_y) = -\frac{1}{2}y\Sigma_y^{-1}y - \frac{1}{2}\log |\Sigma_y| - \frac{1}{2}N\log (2\pi), \quad (5)$$

where the first term is a data fit term, the second term a model complexity term, and the third term a constant. Minimising the negative log marginal likelihood with respect to the hyperparameters of a kernel gives us an optimised posterior distribution [24].

### 2.3 Gaussian Process Classification

Gaussian Process Classification (GPC) is applied to binary classification problems, where class labels  $\mathbf{y} \in [0, 1]$  are observed for input values  $X$ . After defining a GP Prior over a suitable function space, the functional form of the likelihood is determined to approximate the posterior distribution. The goal is to get an estimate of a class probability for unknown data points from Boolean observations. The probability of belonging to a certain class at an input value  $x$  is related to the value of some latent function  $f(x)$  at this location. In the first step, a GP prior is placed over the latent function  $f(x)$ . As we apply GPC only for multi-dimensional input, we implement the *RBF-ARD* kernel for all data sets:  $k_{rbf-ard}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp(-\frac{1}{2} \sum_{d=1}^D \frac{|x_d - x'_d|^2}{\ell_d^2})$  for  $d$  dimensions with scale factor  $\sigma^2$  and  $d$  different lengthscales  $\ell_i$ .

The prior is squashed with the inverse probit transformation to transform real values into probabilities,

$$\Phi(z) = \int_{-\infty}^z \mathcal{N}(x|0, 1)dx = \frac{1}{2} + \frac{1}{2} \cdot \text{erf} \left( \frac{z}{\sqrt{2}} \right), \quad (6)$$

where  $\text{erf}$  is the Gauss error function, defined as  $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$  [26]. Therefore, we obtain a prior on class probabilities  $\pi(x) \triangleq p(y = 1|x) = \Phi(f(x))$ . Then, the distribution of the latent variable corresponding to a test case is computed with

$$p(f_*|X, \mathbf{y}, x_*) = \int p(f_*|X, x_*, \mathbf{f})p(\mathbf{f}|X, \mathbf{y})d\mathbf{f}. \quad (7)$$

This distribution contains the posterior over the latent variables as the product of a normalisation term containing the marginal likelihood, the prior and the likelihood,

$$p(\mathbf{f}|X, \mathbf{y}) = \frac{1}{p(\mathbf{y}|X)}p(\mathbf{f}|X) \prod_{i=1}^n p(y_i|f_i). \quad (8)$$

To increase the accuracy of the approach, the kernel's hyperparameters are optimised by minimising the negative log marginal likelihood. For predictions for a class probability we have

$$\bar{\pi}_* \triangleq p(y_* = 1|X, \mathbf{y}, x_*) = \int \Phi(f_*)p(f_*|X, \mathbf{y}, x_*)df_*. \quad (9)$$

As the observations are Boolean and the probit function is used, the corresponding likelihood is non-Gaussian and consequently the integral of the posterior distribution in Eq. 7 is intractable. Therefore, we approximate the joint posterior by a Gaussian distribution using the popular analytic approximation *Expectation Propagation* [20].

## 2.4 Smoothed Model Checking

When modelling biological collectives, it is often necessary to analyse uncertain stochastic systems and infer missing parameters. A novel approach based on GPC is called Smoothed Model Checking [5] (SMMC) and aims to estimate the satisfaction function of an uncertain CTMC for a specific temporal logic property. Given is an uncertain CTMC  $\mathcal{M}_\theta$  with unknown parameters  $\theta$  and a temporal logic property  $\varphi$ . For a few fixed values of  $\theta$ , several trajectories of  $\mathcal{M}_\theta$  are simulated and the satisfactions of  $\varphi$  (i.e.  $\mathcal{M}_\theta \models \varphi$ ) collected. These observations follow a Binomial distribution and are the input to GPC. However, the algorithm of GPC is changed in such a way that it can deal with multiple observations per data point and make use of the exact statistical model. As a result, we get an accurate estimation of the satisfaction function  $f_\varphi(\theta) = P(\mathcal{M}_\theta \models \varphi)$  over varying parameters  $\theta$ .

In contrast to the original work, we use SMMC to estimate the satisfaction of a property not over varying model parameters, but over varying group sizes. Our application of SMMC aims to find the most plausible value of the missing quantity in a template formula to derive the fitness function. We explain the detailed workflow in Sect. 2.6.

## 2.5 Model Selection

Gaussian Process models are essentially defined by the chosen kernel function that determines the shape of the function to be estimated. Without prior knowledge about the shape, it is recommended to test different kernels and afterwards select the best fit. Because of only few data available, we apply Leave-One-Out Cross-Validation (LOOCV) to estimate the expected prediction error and decide for the best model. LOOCV provides reliable and unbiased estimates of the model's performance, even for small data sets [24]. In particular, the training data is split into  $K = n$  folds, where  $n$  is the size of the data set. Then, for each fold  $k \in \{1, \dots, K\}$ , the model is trained on all folds but the  $k$ -th one, and tested on the remaining ones [16].

The summary statistics of the test set gives an overall evaluation of the goodness of the model. Here, we compute the amount of error for each kernel with the Mean Squared Error (MSE),

$$MSE = \frac{\sum_i^n (y_i - f_{*i})^2}{n} \quad (10)$$

with  $y_i$  being the observations,  $f_{*i}$  the predictions, and  $n$  the size of the test set [24].



For GPR, we consider multiple kernel functions and use LOOCV to automatically select the best kernel. However, for GPC we always use the RBF-ARD kernel and thus only compute the MSE to evaluate the quality of the model and perform no model selection.

## 2.6 Problem Statement

**Predict the Collective Response with GPR.** In our two-folded approach, we first use GPR to predict the collective response for varying group sizes to obtain more information about the social influence within the collective. We assume to have only data about the collective response of a few group sizes available that consist of the final states of the agents. That means we can present the data as histograms counting the frequencies of different outcomes for each available group size. We extract the mean and variance of the histogram and use it as a measure of collective response, e.g. how many agents have successfully solved the task on average. Then, we apply GPR on mean values and variances with different kernels for which we optimise the hyperparameters. We select the best kernel using LOOCV, where we compare the MSE of each model. As a result, we get a prediction of the mean collective response (and variance) within a 95% credible interval for different group sizes and thus gain a better understanding of the general collective behaviour without making any previous assumptions.

**Inferring the Fitness Function with SMMC.** After the first part helps to understand the trend of the collective response over varying group sizes, we aim to find out how the social context influences the individual response in the second part. More precisely, we propose a general fitness function that is likely to explain the collective behaviour but contains an unknown parameter  $t \in \mathbb{R}$  that specifies the exact mechanism. The fitness function is defined as a template temporal logic formula  $\varphi_t$  in the language Probabilistic Computation Tree Logic [15] for discrete systems, and in Signal Temporal Logic [23] for continuous systems. We expect the fitness function to describe a behaviour that is robustly performed across all group sizes  $n$ , which relates to  $\varphi_t$  being robustly satisfied over all  $n$ . Specifically, we set up the template formula as an atomic proposition with one unknown parameter  $t$ . To finally infer the value of  $t$  that best describes the behaviour, we choose a few equally-spaced values of  $t$  and for each of these collect the number of satisfactions of  $\varphi_t$  for given group sizes. Then, we run SMMC for each  $t$  individually to obtain a smooth satisfaction function of  $\varphi_t$  over all varying group sizes.

The resulting posterior distributions are then compared with respect to their shapes. A high mean value indicates a high satisfaction probability of the property, while a low standard deviation implies small variation across different group sizes, and therefore a robust behaviour. For our specific scenario, we compute the coefficient of variation for the posterior distribution of each  $t$  as the fraction of mean and standard deviation,  $c_v(t) = \frac{\mu}{sd}$ . According to the literature (e.g., [10, 28]), a  $c_v < 0.1$  is considered low and indicates a distribution with our

desired properties, specific to the previously defined fitness function template. Finally, we select the largest value of  $t$  with  $c_v(t) < 0.1$  to get the most plausible quantity of the formula and a valid fitness function the collective robustly performs. This fitness function helps to describe how the social context influences the individual's response.

### 3 Results

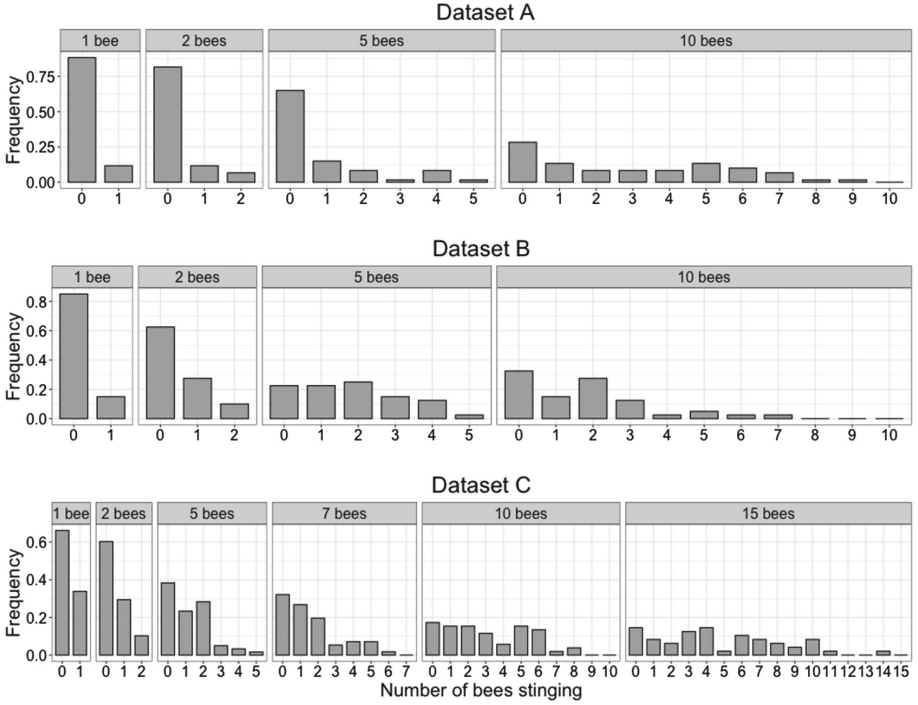
In this section we present the results of our framework on a case study discussing a social feedback mechanism found in honeybees. Biological collectives of honeybees protect their colonies against a threat by releasing an alarm pheromone that warns the other bees and leads to a recruitment of a large number of further defenders. More specifically, to successfully defend their territory, some of the bees become aggressive, decide to sting, and consequently die. During stinging, the alarm pheromone  $P$  is released and increases the aggressiveness of other bees in the colony. However, if the aggressiveness increased endlessly, eventually all bees of the population would die, which is an unreasonable assumption. Therefore, there needs to be some regulatory mechanism that prevents the colony from becoming extinct, while still being able to effectively defend against the threat. From this follows the hypothesis that the bee is socially influenced by its colony and aware of its social context, i.e., the group size. See [14] for a more detailed description of the case study and the assumptions for the associated stochastic model. To better understand the exact underlying mechanism of social feedback, we apply our methods on experimental data of this phenomenon for a few group sizes. We use Gaussian Process Regression to predict the collective response over all intermediate group sizes and learn about the trend of how the context regulates the bees' behaviour. We then aim to derive the non-trivial fitness function by setting up a plausible template formula and applying Smoothed Model Checking to automatically infer the missing quantity that explains the collective dynamics.

#### 3.1 Data

To test our framework on real-world observations, we make use of experimental data collected at the University of Konstanz (Germany) [27]. In three experiments, groups of 1, 2, 5, 7, 10, or 15 bees were put into an arena and exposed to a threat. After a certain time, the number of stinging bees was counted which provides a measure of the collective response. This procedure was repeated several times for each population size within each experiment. Hence, we get three histograms with the frequencies of stinging bees of each population size. See Fig. 1 for the result distributions of all data sets.

#### 3.2 Predict the Collective Response

Our data contains information about the collective response of a few selected group sizes. However, to get predictions for all other intermediate group sizes,

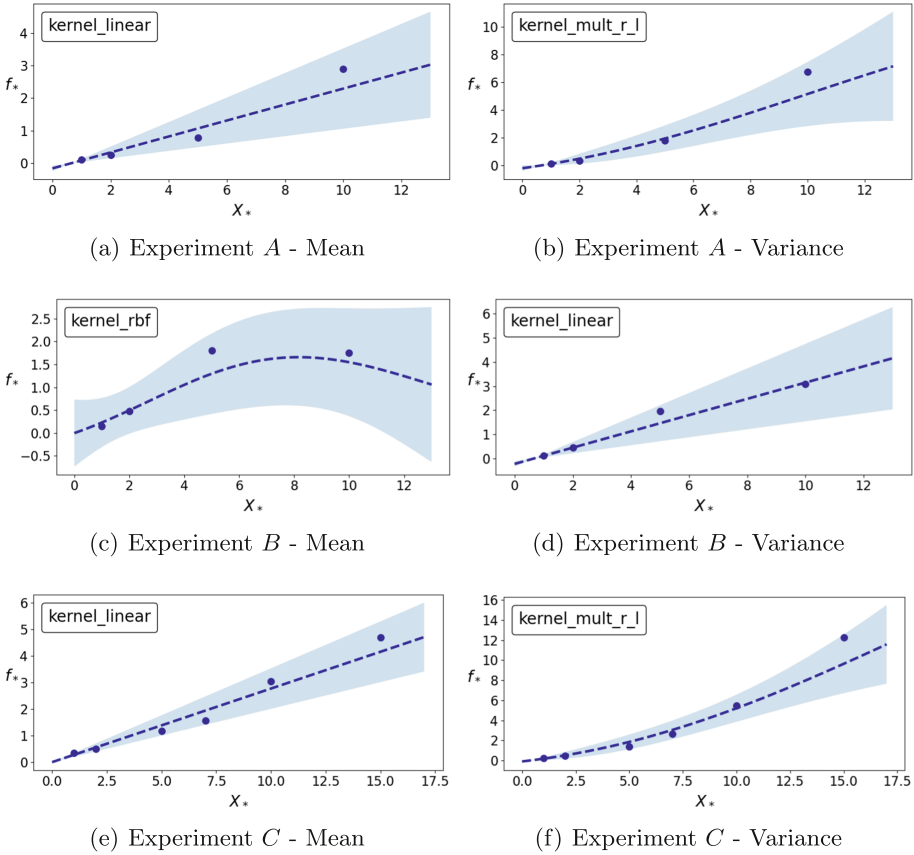


**Fig. 1.** Overview of experimental data from three data sets showing the frequencies of the number of stinging bees. Experiments were repeated with sample sizes  $N_A = [60, 60, 60, 60]$ ,  $N_B = [40, 40, 40, 40]$  and  $N_C = [68, 68, 60, 56, 52, 48]$ .

we apply GPR on the three data sets. As mentioned above, we consider the number of stinging bees as the collective response to defend the territory. We compute the mean and variance of each histogram, corresponding to the mean and variance of stinging bees for each population size, and use these values as input to the algorithm. Noise, computed as the 95% credible interval, is added to each data point to account for observation errors and limited sample sizes [31]. Then, we run GPR for each implemented kernel and combination of two kernels with optimised hyperparameters. As a result, we get the posterior predictive distribution of the collective response for different population sizes. The best model is selected with the lowest MSE according to LOOCV and shown in Fig. 2.

We observe that the uncertainty increases for larger group sizes due to the small sample size. For a group size of one bee, there are only two outcomes of the experiment: either no bees sting, or one bee stings. In contrast, for a group size of ten bees, there are eleven possible outcomes of the experiment. Since the sample size remains the same, the uncertainty increases.

The results show that we can model different trends of collective response with the same algorithm and without specifying any previous assumptions



**Fig. 2.** Posterior distributions for mean and variance of histograms for experimental data. Points are training data points, dashed lines are predictive means and shaded areas are 95% credible regions. Best kernel according to LOOCV is written in the left upper corner with following MSEs: Experiment A - Mean: linear kernel,  $MSE = 0.8882$  - Variance: multiplication of RBF and linear kernel,  $MSE = 4.4019$ . Experiment B - Mean: RBF,  $MSE = 0.5536$  - Variance: linear kernel,  $MSE = 0.1219$ . Experiment C - Mean: linear kernel,  $MSE = 0.2157$  - Variance: multiplication of RBF and linear kernel,  $MSE = 7.3824$ .

beforehand. The linear trend of the number of stinging bees in Experiment A and C is well captured, as well as the non-trivial trend in Experiment B. From these distributions we can easily infer the collective response of all other group sizes. In the case of having social feedback in a colony, we are therefore still able to get reliable estimates of the behaviour of the colony without the need of conducting new experiments. Note that predictions for group sizes outside the range of available data points are also possible, but introduce even larger uncertainties.

### 3.3 Inferring the Fitness Function

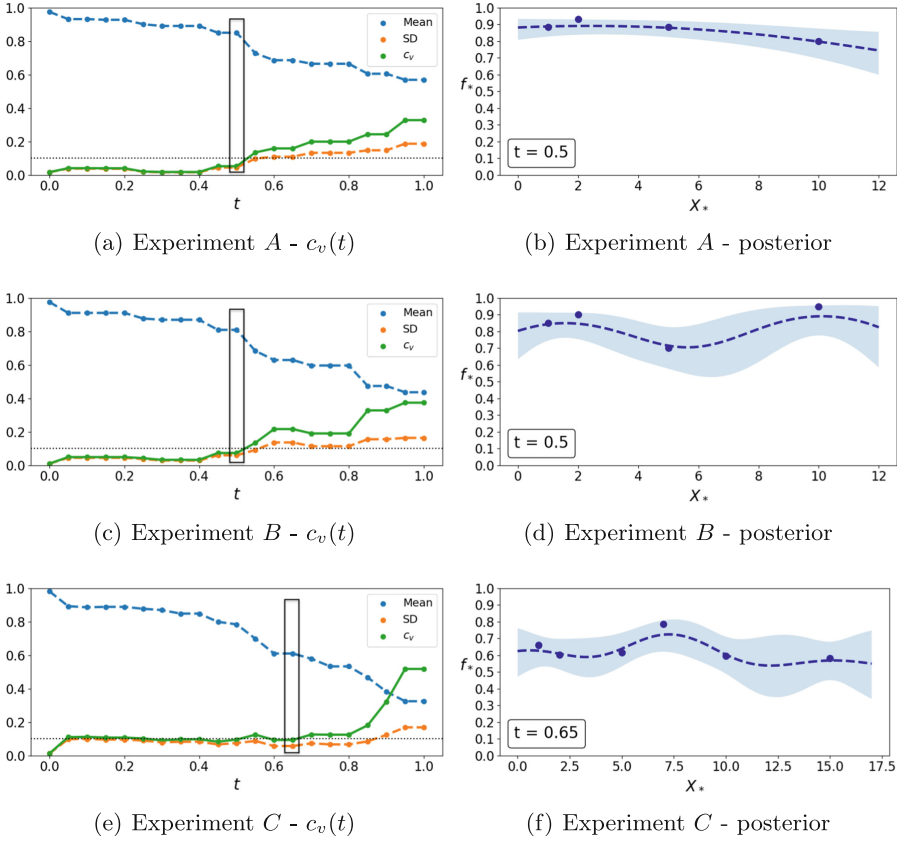
In the second step, we aim to get a better understanding of social feedback and how the collective behaviour adapts to changes in the group sizes. For this case study, we want to investigate if there is always a certain proportion of the population defending the whole collective. Therefore, we want to derive the most plausible quantity for the fitness function ‘at least  $(100 \cdot t)\%$  of the colony survives’ and set up the corresponding template temporal logic formula  $\varphi : \mathbf{F}(X \geq t)$  with  $X$  being the number of surviving bees and  $t \in [0, 1]$  the unknown threshold.

We select 21 equally-spaced thresholds  $t$  and analyse the satisfaction function of  $\varphi$  for different group sizes with SMMC with optimised hyperparameters. For each  $t$ , the inputs to SMMC are the number of observations for which the property is satisfied in each group size, read from the respective histogram. Again, we get a posterior distribution with 95% credible regions for different group sizes. Computing the coefficient of variation of the posterior distribution for all values of  $t$  gives us an estimate about the most plausible distribution with respect to the previously defined fitness function. We select the largest  $t$  with a low  $c_v < 0.1$  to obtain a distribution with high mean values and little variation. This gives us the property that defines a behaviour that is robustly satisfied by the collective over all group sizes. In Fig. 3, we show on the left side the coefficients of variation  $c_v$  for different thresholds  $t$  together with the mean and standard deviation of the posterior distributions for each experiment. On the right side, we visualise the SMMC posterior of the selected value of  $t$ .

The obtained results indicate the most plausible values to be  $t = 0.5$  for experiment  $A$  and  $B$ , and  $t = 0.65$  for experiment  $C$ . The biological interpretation of the analysis is that, on average, 50 – 65% of a honeybee colony survives when being exposed to a threat. Put differently, at least 50 – 65% of a colony needs to perform a stinging response to successfully defend the territory. With this method we were able to automatically quantify the high level behavioural outcome of the collective that is robustly performed under perturbations. Furthermore, we observe that all posterior distributions capture the data well according to the low MSEs.

## 4 Conclusion

In this paper, we presented a framework based on Gaussian Processes to better understand the phenomenon of social feedback in biological collectives. Our contribution is two-fold: first, we predict the collective response for any given group size from only limited experimental data; Second, we derive a fitness function that is robustly preserved under perturbations in group size. On the one hand, the application of our methods helps to test the hypothesis of social feedback in a collective, when only measurements of few group sizes are available. The resulting predictions of collective response eliminate the need of conducting new experiments and analysing combinatorially large stochastic models. Still, we get reliable results for any group size, together with a quantification of uncertainty.



**Fig. 3.** SMMC results of experimental data for the property  $\varphi$  : at least  $(100 \cdot t)\%$  of the colony survives. Left: mean (blue), standard deviation (orange), and coefficient of variation (green) of posterior distributions over varying  $t$  are shown. Black dotted line shows the threshold  $c_v = 0.1$ . Black rectangle shows the values for the largest  $t$  with  $c_v(t) < 0.1$ . Right: SMMC posterior for selected  $t$ , points are training data points, dashed lines are predictive posterior means and shaded areas are 95% credible regions. Experiment *A* -  $t = 0.5$  with  $c_v(t) = 0.052$ ,  $MSE = 0.0038$ . Experiment *B* -  $t = 0.5$  with  $c_v(t) = 0.0731$ ,  $MSE = 0.0136$ . Experiment *C* -  $t = 0.65$  with  $c_v(t) = 0.093$ ,  $MSE = 0.0072$ . (Color figure online)

On the other hand, our framework can be used to assess the trend of social feedback in the sense of how social context influences the collective response. The missing quantities in a template logical formula (usually proposed by experts), is automatically inferred to derive a fitness function that describes the collective behaviour under group-size perturbations.

Both applications are based on Gaussian Processes, which has several key advantages compared to traditional methods. Usually, the analysis of models for larger group sizes becomes computationally infeasible due to state explo-

sion. Gaussian Processes, as a non-parametric and model-agnostic approach, are instead scalable for any given group size and therefore particularly useful to measure the collective response with respect to growing group size. Beyond that, the analyses of our proposed methods are data-efficient and produce reliable results with uncertainty quantification even for scarce data sets. Especially when working with real-world experimental data, there are often not enough resources available to collect large amounts of data. Therefore, we decided to use Gaussian Processes that are able to find the underlying relationships between only few available data points and also provide statistical guarantees. Last, we want to emphasise the flexibility of this framework, where we not only discard any previous assumptions on the underlying model and its parameters, but further are able to use it on any related application. While we focus on understanding social feedback in honeybees in this work, other use cases of analysing collective behaviour are possible. Instead of predicting the steady state for any group size, the method could also be applied to any quantitative measurement of the collective. Accordingly, the template fitness function can be exchanged by any temporal logic formula for which we can assess the satisfaction probability, and the coefficient of variation by a different measure of robustness to suit the particular case study and research question.

Despite highlighting the power of the proposed methods, we also want to point out possible limitations. One major drawback of Gaussian Processes is the computational complexity of  $O(N^3)$  [24]. In our work, we implemented all functions by hand, in order to have full control over the computations. However, using available libraries like GPyTorch [11], could speed up the calculations. Another limitation of Gaussian Processes is the extrapolation needed for larger or smaller group sizes than those available in the data set. In this case, the uncertainty quickly becomes large and the predictions imprecise. In practice, one would encourage to conduct a new experiment for much smaller/larger group sizes to counteract these high uncertainties and focus on interpolation of intermediate group sizes.

Future work will focus on exploring the full potential of the presented techniques in terms of automatically learning unknown parameters of a model or even the entire mechanisms. In general, the approach could be automatised and integrated into a probabilistic reasoning framework.

**Acknowledgements.** JK’s research was supported by the AFF (Der Ausschuss für Forschungsfragen, EU-Anschubfinanzierung, Univ. of Konstanz), TP’s research is supported by the Ministry of Science, Research and the Arts of the state of Baden-Württemberg. TP and JK were further funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy - EXC 2117 - 422037984. The authors would like to thank Francesca Cairoli, Laura Nenzi, and Wolfram Barfuss for the fruitful and inspiring discussions on the topic.

## References

1. Bartocci, E., Bortolussi, L., Nenzi, L., Sanguinetti, G.: On the robustness of temporal properties for stochastic models. *Electron. Proc. Theor. Comput. Sci.* **125**, 3–19 (2013) 10.4204/EPTCS.125.1, <https://arxiv.org/abs/1309.0866v1>
2. Bartocci, E., Bortolussi, L., Nenzi, L., Sanguinetti, G.: System design of stochastic models using robustness of temporal properties. *Theor. Comput. Sci.* **587**, 3–25 (2015). <https://doi.org/10.1016/j.tcs.2015.02.046>
3. Berdahl, A., Torney, C.J., Ioannou, C.C., Faria, J.J., Couzin, I.D.: Emergent sensing of complex environments by mobile animal groups. *Science* **339**(6119), 574–6 (2013). <https://doi.org/10.1126/science.1225883>, <https://www.ncbi.nlm.nih.gov/pubmed/23372013>
4. Bortolussi, L., Milios, D., Sanguinetti, G.: U-Check: model checking and parameter synthesis under uncertainty. In: Campos, J., Haverkort, B.R. (eds.) *QEST 2015*. LNCS, vol. 9259, pp. 89–104. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-22264-6\\_6](https://doi.org/10.1007/978-3-319-22264-6_6)
5. Bortolussi, L., Milios, D., Sanguinetti, G.: Smoothed model checking for uncertain continuous-time Markov chains. *Inf. Comput.* **247**, 235–253 (2016)
6. Bortolussi, L., Sanguinetti, G.: Learning and designing stochastic processes from logical constraints. In: Joshi, K., Siegle, M., Stoelinga, M., D’Argenio, P.R. (eds.) *QEST 2013*. LNCS, vol. 8054, pp. 89–105. Springer, Heidelberg (2013). [https://doi.org/10.1007/978-3-642-40196-1\\_7](https://doi.org/10.1007/978-3-642-40196-1_7)
7. Chen, T., Hadinoto, K., Yan, W., Ma, Y.: Efficient meta-modelling of complex process simulations with time-space-dependent outputs. *Comput. Chem. Eng.* **35**(3), 502–509 (2011). <https://doi.org/10.1016/j.compchemeng.2010.05.013>, <https://linkinghub.elsevier.com/retrieve/pii/S009813541000195X>
8. Cook, C.N., Breed, M.D.: Social context influences the initiation and threshold of thermoregulatory behaviour in honeybees. *Anim. Behav.* **86**(2), 323–329 (2013). <https://doi.org/10.1016/j.anbehav.2013.05.021>
9. Dorigo, M., Birattari, M., Blum, C., Clerc, M., Stützle, T., Winfield, A.F.T. (eds.): *ANTS 2008*. LNCS, vol. 5217. Springer, Heidelberg (2008). <https://doi.org/10.1007/978-3-540-87527-7>
10. Faria Filho, D., et al.: Classification of coefficients of variation in experiments with commercial layers. *Revista Brasileira de Ciência Avícola* **12**(4), 255–257 (2010). <https://doi.org/10.1590/S1516-635X2010000400006>
11. Gardner, J.R., Pleiss, G., Bindel, D., Weinberger, K.Q., Wilson, A.G.: GPyTorch: blackbox matrix-matrix gaussian process inference with GPU acceleration. In: *Advances in Neural Information Processing Systems* (2018)
12. Giardina, I.: Collective behavior in animal groups: theoretical models and empirical studies. *HFSP J.* **2**(4), 205–219 (2008)
13. Görtler, J., Kehlbeck, R., Deussen, O.: A visual exploration of gaussian processes. *Distill* **4**(4), e17 (2019)
14. Hajnal, M., Nouvian, M., Šafránek, D., Petrov, T.: Data-informed parameter synthesis for population Markov chains. In: Češka, M., Paoletti, N. (eds.) *HSB 2019*. LNCS, vol. 11705, pp. 147–164. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-28042-0\\_10](https://doi.org/10.1007/978-3-030-28042-0_10)
15. Hansson, H., Jonsson, B.: A logic for reasoning about time and reliability. *Formal Aspects Comput.* **6**(5), 512–535 (1994)
16. Hastie, T., Tibshirani, R., Friedman, J.H.: *The elements of statistical learning: data mining, inference, and prediction*. Springer series in statistics, Springer, New York, 2nd edn. (2009)



17. Hillston, J.: Challenges for quantitative analysis of collective adaptive systems. In: Abadi, M., Lluch Lafuente, A. (eds.) TGC 2013. LNCS, vol. 8358, pp. 14–21. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-05119-2\\_2](https://doi.org/10.1007/978-3-319-05119-2_2)
18. Jin, X., Donzé, A., Deshmukh, J.V., Seshia, S.A.: Mining requirements from closed-loop control models. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **34**(11), 1704–1717 (2015)
19. Kleineidam, C.J., Heeb, E.L., Neupert, S.: Social interactions promote adaptive resource defense in ants. *PLoS One* **12**(9), e0183872 (2017). <https://doi.org/10.1371/journal.pone.0183872>, <https://www.ncbi.nlm.nih.gov/pubmed/28910322>
20. Kuß, M.: Gaussian Process Models. Ph.D. thesis (2006)
21. Loreti, M., Hillston, J.: Modelling and analysis of collective adaptive systems with CARMA and its tools. In: Bernardo, M., De Nicola, R., Hillston, J. (eds.) SFM 2016. LNCS, vol. 9700, pp. 83–119. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-34096-8\\_4](https://doi.org/10.1007/978-3-319-34096-8_4)
22. Mai, M., et al.: Monitoring pre-seismic activity changes in a domestic animal collective in central Italy. In: EGU General Assembly Conference Abstracts. vol. 20, p. 19348 (2018)
23. Maler, O., Nickovic, D.: Monitoring temporal properties of continuous signals. In: Lakhnech, Y., Yovine, S. (eds.) FORMATS/FTRTFT -2004. LNCS, vol. 3253, pp. 152–166. Springer, Heidelberg (2004). [https://doi.org/10.1007/978-3-540-30206-3\\_12](https://doi.org/10.1007/978-3-540-30206-3_12)
24. Murphy, K.P.: *Machine Learning: A Probabilistic Perspective*. MIT press, Cambridge (2012)
25. Nouvian, M., Reinhard, J., Giurfa, M.: The defensive response of the honeybee *Apis mellifera*. *J. Exp. Biol.* **219**(22), 3505–3517 (2016)
26. Patel, J.K., Read, C.B.: *Handbook of the Normal Distribution*, vol. 150. CRC Press, Boca Raton (1996)
27. Petrov, T., Hajnal, M., Klein, J., Šafránek, D., Nouvian, M.: Extracting individual characteristics from population data reveals a negative social effect during honeybee defence. *PLOS Comput. Biol.* **18**(9), e1010305 (2022). <https://doi.org/10.1371/journal.pcbi.1010305>
28. Pimentel Gomes, F.: *Curso de Estatística Experimental*, 13th edn. Nobel, São Paulo (2000)
29. Rasmussen, C.E., Williams, C.K.I.: *Gaussian processes for machine learning*, 3, vol. 2. MIT Press, Cambridge (2006)
30. Shorter, J.R., Rueppell, O.: A review on self-destructive defense behaviors in social insects. *Insectes Soc.* **59**(1), 1–10 (2012)
31. Zhang, J., Yin, J., Wang, R.: Basic framework and main methods of uncertainty quantification. *Math. Prob. Eng.* **2020** (2020)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

