

RESEARCH ARTICLE

A Bayesian approach to modelling the impact of hydrodynamic shear stress on biofilm deformation

Oluwole K. Oyebamiji^{1*}, Darren J. Wilkinson¹, Pahala Gedara Jayathilake², Steve P. Rushton³, Ben Bridgens², Bowen Li⁴, Paolo Zuliani⁴

1 School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom, **2** School of Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom, **3** School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom, **4** School of Computing Science, Newcastle University, Newcastle upon Tyne, NE4 5TG, United Kingdom

* wolemi2@yahoo.com



OPEN ACCESS

Citation: Oyebamiji OK, Wilkinson DJ, Jayathilake PG, Rushton SP, Bridgens B, Li B, et al. (2018) A Bayesian approach to modelling the impact of hydrodynamic shear stress on biofilm deformation. PLoS ONE 13(4): e0195484. <https://doi.org/10.1371/journal.pone.0195484>

Editor: Robert Nerenberg, University of Notre Dame, UNITED STATES

Received: September 29, 2017

Accepted: March 24, 2018

Published: April 12, 2018

Copyright: © 2018 Oyebamiji et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data supporting this publication are openly available under an 'Open Data Commons Open Database License'. Additional metadata are available at: <http://dx.doi.org/10.17634/123172-3> Please contact Newcastle Research Data Service at rdm@ncl.ac.uk for access instructions.

Funding: This work was funded by the EPSRC (grant No EP/K039083/1) under the Newcastle University Frontiers in Engineering Biology (NUFEB) project.

Abstract

We investigate the feasibility of using a surrogate-based method to emulate the deformation and detachment behaviour of a biofilm in response to hydrodynamic shear stress. The influence of shear force, growth rate and viscoelastic parameters on the patterns of growth, structure and resulting shape of microbial biofilms was examined. We develop a statistical modelling approach to this problem, using combination of Bayesian Poisson regression and dynamic linear models for the emulation. We observe that the hydrodynamic shear force affects biofilm deformation in line with some literature. Sensitivity results also showed that the expected number of shear events, shear flow, yield coefficient for heterotrophic bacteria and extracellular polymeric substance (EPS) stiffness per unit EPS mass are the four principal mechanisms governing the bacteria detachment in this study. The sensitivity of the model parameters is temporally dynamic, emphasising the significance of conducting the sensitivity analysis across multiple time points. The surrogate models are shown to perform well, and produced \approx 480 fold increase in computational efficiency. We conclude that a surrogate-based approach is effective, and resulting biofilm structure is determined primarily by a balance between bacteria growth, viscoelastic parameters and applied shear stress.

Introduction

Water is crucial for life on earth and is valuable also for its supporting role in ecosystem function. With growing global population and climate change induced water crises, there is an increase in the amount of wastewater for municipal uses and this might lead to a shortage of drinkable water. Biofilm technology is being deployed in the management and treatment of wastewater. A model is required that describes the individual processes in the wastewater treatment system. The simulation of microbial communities has important application in wastewater treatment studies. Wastewater treatment plants are open systems that depend on many

Competing interests: The authors have declared that no competing interests exist.

species of bacteria to form a microbial community for the transformation of waste into biomass and other substances. According to [1], biofilms are regarded as the commonest form of bacteria on earth.

It has been established that the growth, structure and performance of bacteria biofilms are strongly affected by the hydrodynamic shear force. It is increasingly being recognised that hydrodynamic shear stress has a significant role to play on the deformation of biofilms and detachment of bacteria. There has been a large number of research projects dealing with the assessment of the impact of hydrodynamic stress on biofilms deformation. [2] observed that steady state structures of biofilms are strongly affected by the hydrodynamic shear stress. The general understanding of the influence of bacteria detachment is documented in [3–5].

The detachment process is an essential mechanism for removal of biomass from biofilm thereby controlling the biofilm key processes like growth, development and performance [2, 6, 7]. Moreover, [8] and [4] identified five different mechanisms of biomass detachment in biofilm while recently [3] focus on just only three out of these processes. The first category is the shear detachment which occurs as a result of fluid flow in the bacteria compartment, and an erosion detachment which is breakage of small particles from the surface of biofilm into bulk fluid. The third type is the nutrient-limited detachment which is associated with insufficient nutrient effects. However, [6] and [9] limit their attention to erosion (small-particle loss) and sloughing detachment of relatively large portions of the biofilm. In a similar vein, [9] noted that the biofilm simulation subjected to an erosion type of detachment event has the potential for making the biofilm surface smoother while sloughing type detachment can cause an increase biofilm surface roughness. In other words, the detachment phenomenon is a significant determinant of the shape, composition and structure of emerging biofilm.

We know that biofilm growth simulation is computationally demanding, and most of the available studies on biomass detachment usually base their inference on a relatively small sample of simulation data. For instance, [3] consider only three parameter values for shear, nutrient-limited and erosion detachment coefficients in their experiments while the consideration of only six parameter values is used as detachment parameters in [5]. The limitation of these studies is their lack of sufficient data to make a rigorous validation for testing how shear forces influence bacteria detachment. Similarly, [5] and [3] use only the simple detachment rate and probabilities in their studies but fail to incorporate the knowledge of mechanical interactions among the particles in their simulations, while [10] and [11] completely neglect the effect of biomass deformation in their studies.

The simulation of the impacts of shear flow on the size and structure of biofilm is undertaken in this paper such that the shear force from the fluid flow on mature biofilm leads to deformation and eventual breakage as an emergent event. The approach is computationally demanding as noted in some literature eg [5] because of the partial differential equation (PDE) necessary to model the mechanical forces in the system.

To gain deeper insights into how the shear force affects the deformation of biofilm, we develop a novel surrogate-based technique for solving this problem. A surrogate model is a simplified statistical approximation to an expensive computer model, often referred to as a “statistical emulator” [12] for predicting the shearing behaviour of model systems without having to rely on an expensive simulator. We are interested in strengthening the study of the essential role of shear stress breakage of microbial aggregates. We note that the crude parameter scans used in [3, 5] is not an ideal approach to fully understand the emerging properties of the microbial organism, and could be improved using a properly designed experiment. Our approach is to use advanced statistical techniques based on a large ensemble of simulation data to make a rigorously tested and validated assessment of the effect of shear force on microbial

deformation. This approach will provide new insights into how quantitative statistical techniques can be used to simplify and study this complex problem.

The simulation data we analysed in this paper are from an expensive dynamic model. There have been a large number of studies that have examined data from a dynamic simulator. For instance, [13] emulated the characterized biofilms and floc outputs of an individual-based (IB) model of microbial communities using statistical principles of dynamic emulation while [14] focus on a low-order dynamic model that approximates the response of the high-order dynamic simulator at a low computational cost. [15] described a Bayesian method for quantification of uncertainty in complex computer models while [16] applied a Bayesian technique to calibrate computer models. Also, [17], focusing more on heterogeneous data, used a hybrid Markov chain Monte-Carlo method.

Bayesian computations have extended and broadened the scope of statistical models that can be handled in practice due to the development of Markov chain Monte Carlo (MCMC). However, if the model errors have a Gaussian distribution and a given form of the prior distribution is assumed, then the posterior distributions of the model's parameters can sometimes be obtained analytically, without MCMC. A major benefit of using Bayesian regression is the provision of a measure of uncertainty in its analysis. The disadvantage is that, computationally, it can be very demanding. This is not a serious drawback for the problems we addressed in this paper because the sample size is moderate and the simplifying assumption we made under the DLM implementation also reduces the computational expense.

The two outputs we consider to model in this paper are the expected number of detachment events (count data) and volume of detached clusters. The traditional approach for modelling event-count data is to use a Poisson model. In the earlier studies of [18], a Bayesian Poisson regression model based on the Gibbs sampler was used while [19, 20] apply a similar Bayesian Poisson regression for modelling the crowd counting and injury count data respectively. We use a 3D individual-based model simulation of microbial organisms incorporating a fluid flow that is based on LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator), a classical dynamical model for particle simulation. This simulator was enhanced to incorporate biological and physical processes to model bacterial growth, decay and mechanical interactions among bacteria cells [21].

We know from the available literature that the impact of hydrodynamic shearing force on the biofilm fragmentation has not been thoroughly studied using quantitative statistical techniques. The primary objective of this work is to investigate the effect of shearing force on the biofilm deformation and bacteria detachment using a surrogate-based method. Firstly, we assume that the biofilm fragmentation occurs as an event and we proceed by examining the extent to which shearing force impacts on the hazard of a bacteria detached from a parent biofilm.

Secondly, we quantify the relationship between the average number of shearing events per unit time and some covariates like total number of particles, shear rates, biofilm height, mass and extracellular polymeric substance (EPS) composition, using a novel combination of Bayesian Poisson regression and dynamic linear models to study the biofilm detachment problem. Bacteria are embedded in a sticky extracellular polymeric substance (EPS) produced by bacteria themselves and mostly composed of polysaccharide, proteins and nucleic acids etc. EPS helps for the structural integrity of the biofilm and hence biofilms can resist applied external forces. It also plays a significant role in microbial competitions in a multi-species biofilm [21–23].

We then predict the total volume of detached clusters per unit time as a continuous function of the predicted number of shearing events, shear rates and other covariates given in [Table 1](#) using a dynamic linear model (DLM). This modular approach will enable us to

Table 1. IB model parameters.

Index	Parameters	Symbol	Values	Units	References
1	Substrate affinity	$K_{s,HET}$	0.000035	kgm^{-3}	[26]
2	Max. specific growth rate	$\mu_{m,HET}$	1.000000	h^{-1}	[27]
3	Yield coefficient	Y_{HET}	0.610000	$\frac{gCOD}{gCOD}$	[28]
4	Shear rates	γ	0.250000	s^{-1}	chosen
5	Spring coef for collision	K_n	1×10^{-4}	Nm^{-1}	[29]
6	Viscous coef for collision	γ_n	1×10^{-5}	s^{-1}	Chosen
7	EPS stiffness	K_e	5×10^9	s^{-2}	[30]
	Dimension	$L_x \times L_y \times L_z$	$200 \times 40 \times 100$	μm^3	-
	Cartesian grid cells	$N_x \times N_y \times N_z$	$30 \times 12 \times 30$	-	

<https://doi.org/10.1371/journal.pone.0195484.t001>

predict the distribution of detached clusters over time. We describe the models and simulation data utilised for the analysis in Section 2. In Section 3, we describe the Bayesian methods including the dynamic linear models and Poisson regression. Section 4 provides the results of the analysis including the sensitivity analysis. Section 5 presents the discussion and concluding comments.

Materials and models

Simulation model

The present study models the biofilm that might be found in a wastewater treatment plant (WWTP) at the individual microbe level since pilot scale plants and laboratory scale experiments of wastewater treatment plants WWTP are expensive, cumbersome, non-invasive and often cannot provide information at the micro-scale, which is required for operational optimisation of WWTP. The mathematical models used for biological treatment can be mainly divided into two general classes according to the way the biomass is represented: continuous and discrete models. In the present work, a discrete individual-based (IB) model is used. Biofilms are the aggregated microbial communities attached to surfaces. We have one functional group of microorganisms and one dormant state as soft agents within the present model. The microorganisms are heterotrophs (HET) which consume organic carbon source and oxygen. The inert state is extracellular polymeric substance (EPS), secreted by some heterotrophs. In this agent-based model the EPS is modelled as discrete particles. The adhesive forces between EPS-EPS and EPS-Bacteria particles are modelled by springs in which the spring coefficient is a function of local EPS mass.

The dead agents are represented by soft spheres (labelled DEAD). Agents have four state variables: position, mass, radius, and type. This model consists of two principal submodels; one deals with the growth and behaviour of individual bacteria as autonomous agents (i.e., biological processes); the other deals with the substrate and product diffusion and reaction and fluid flow (i.e., physical processes). Each cell grows by consuming the substrate and divides when a certain mass is reached. When agents grow and split, the system deviates from its mechanical equilibrium due to some extra pressure built-up in the biomass.

Depending on the net force acting on each agent, resulting from its spatial interaction with other local agents, the position of each agent is updated using the Discrete Element Method (DEM). In DEM, contact, EPS adhesion and shear force are considered, and the position of agents are updated by solving Newton's second law equation. For the substrates, Chemical Oxygen Demand (COD) is considered. The diffusion-reaction equation governs the substrate

concentrations, and this transport equation is solved in a fixed Cartesian grid using a Finite Difference Method. This Newcastle University Frontiers in Engineering Biology (NUFEB) model extends the traditional IB model by incorporating mechanical interactions among bacteria. The NUFEB model is implemented in LAMMPS, an open-source C++ molecular dynamics code (<http://lammmps.sandia.gov/>) [24]. More details about the model can be found in [21].

NUFEB simulation data

The IB model was run with a series of growth parameters and shearing forces over an extended period using a Latin hypercube design to generate biofilms of various size. The LHD technique provides good coverage of the input space with a relatively small number of design points [25]. The parameters are varied within the range of $\pm 50\%$ of the standard values given in Table 1 for 140 training points and five replicates at each design point, due to the expense of this computer model. The parameters are $\mu_{m,HET}$ which is the maximum specific growth rate for HET, $K_{s,HET}$ is the substrate affinity for HET. The parameter $K_{s,HET}$ is an inverse measure of bacterium affinity to the organic substrate ie $1/K_{s,HET}$ can be considered as a representation of how easily nutrients are transported across a bacterial membrane. Other parameters are Y_{HET} which is the yield coefficient for HET growth, γ is the hydrodynamic shear rate, the spring coefficient for collision is denoted as K_n , viscous coefficient for collision γ_n and EPS stiffness per unit EPS mass is denoted as K_e ; see Table 1 below, other simulation parameters are held constant; see Table A in S1 Text.

The design matrix is denoted as $\mathbf{X}_{140 \times 7} = (\theta_p^i, p = 1, \dots, 7; i = 1, \dots, 140)$; where the subscript p represents the 7 model parameters that are varied in Table 1. The superscript i denotes the 140 different realisations. The simulator was run for 40000 s to grow the biofilm to a certain height without flow and then subjected the resulting biofilms to shear flow for an additional 200000 s where the biofilm detachment events occur. For each i , the output data is recorded at every 2000 s giving 120 time slices, $t = 1, \dots, 120$. The number of particles and volume of detached clusters lost from the parent biofilm was recorded for different shearing forces. We compute the biofilm height as given below. Other morphological characteristics like biofilm mass, total number of particles and EPS composition are also calculated for predicting expected number of shearing events over time.

The simulation box has dimension $\{0, L_x\} \times \{0, L_y\} \times \{0, L_z\}$, where $L_x = 200\mu m$, $L_y = 40\mu m$ and $L_z = 100\mu m$. To compute average biofilm height, the biofilms are partitioned into several smaller blocks or grids. The number of blocks in each direction are given as $N_x = 30$, $N_y = 12$ and $N_z = 30$. We compute the Euclidean distances between the center of each particle and the lattice blocks along the baseline (plane $z = 0$) to identify the occupied blocks. We, therefore, marked as “occupied” every block with one or more particle centers contained within it while the others are marked as “vacant”. The height $h_t(x, y)$ of the biofilm above each base block is defined as the maximum of the particle z - values of the occupied blocks. The biofilm mean height at time t is then given as $\bar{h}(t) = \frac{1}{L_x L_y} \int \int h_t(x, y) dx dy$.

The scope of the problem

The problem we are addressing in this paper is the simulation of biofilm under shear stress, where the fluid force of appropriate magnitude flow on individual biofilms leads to the weakening of biofilms. This can result in eventual deformation and bacteria loss from the surface. Our focus in this paper is to be able to treat each shear phenomenon as an event, and we are interested in testing the feasibility of using a surrogate-based model for predicting expected number of shear events and size of detached clusters.

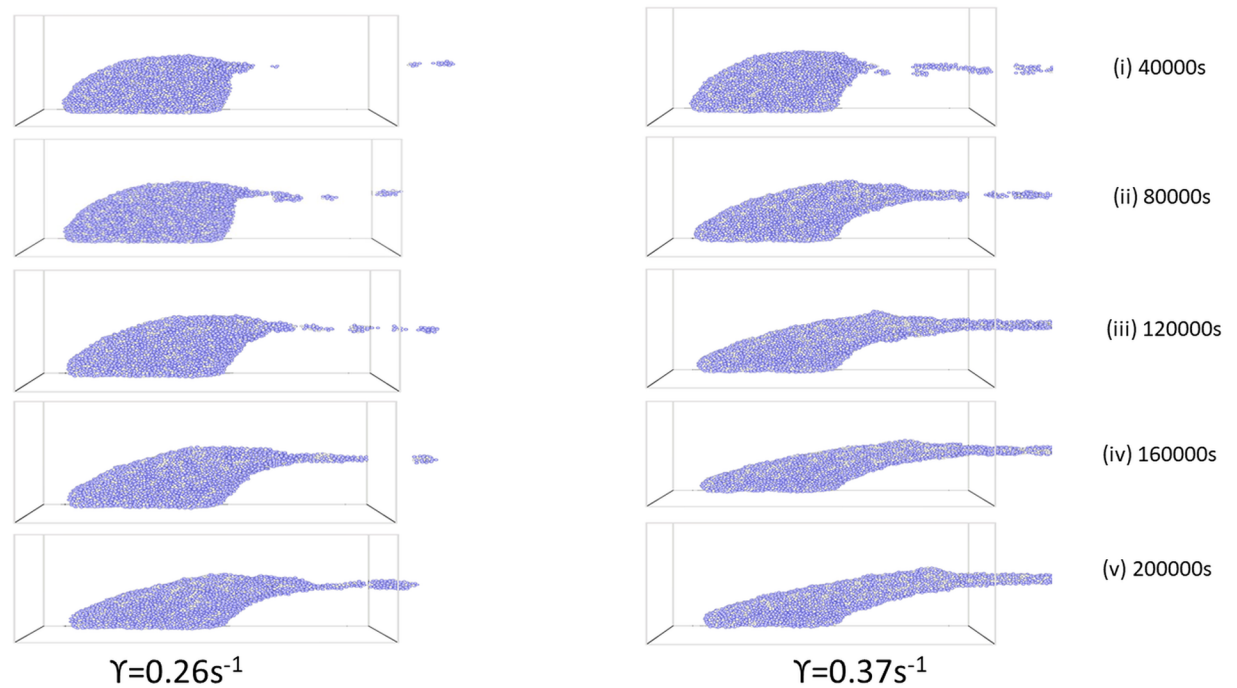


Fig 1. Biofilm structures showing temporal evolution of effect of shear rate on biofilm deformation and bacteria detachment for different shear rates for the period the flow is applied. Left-plot: $\gamma = 0.26s^{-1}$ and right-plot: $\gamma = 0.37s^{-1}$ for 40,000, 80,000, 120,000, 160,000 and 200,000 seconds respectively.

<https://doi.org/10.1371/journal.pone.0195484.g001>

The knowledge about the emerging composition and structure of biofilms subjected to shear flow is useful to improve the performance and stability of wastewater reactors. Fig 1 represents a typical simulation of biofilm under two different shear rates at $\gamma = 0.26s^{-1}$ and $0.37s^{-1}$ respectively, and we can see diverging temporal behaviours and structures when the shear flow is applied on a mature biofilm, the influence of cell detachment from the surface begins to emerge. The detachment phenomenon occurs when cohesive failures happen due to hydrodynamic shear force. At 40,000 s, the frequency of detachment events is higher for $0.37s^{-1}$ rate than for $0.26s^{-1}$. The detached clusters are moving out in the opposite direction to the bulk flow (shear flow from left-hand) as expected. The two shear rates give rise to different detachment patterns. We also see a gradual decreasing and flattening of the biofilms over time. In particular, the elongated filamentous cell clusters (streamers and clumps of cells) at a later time is obvious.

The density plots for the two outputs we are considering are given in Fig 2 (column 1). This enables us to look closely at the distribution of the two outputs to be analysed. It is apparent that each density plot has relatively different distribution under different shear rates and are highly skewed and nonnormal. The number of shear events is modelled using a Poisson distributed random variable because they are count data. We modelled the expected number of shear events function using a Poisson regression having an exponential-mean parameter that is a quadratic function of the explanatory variables. The inputs to the Poisson model are six variables namely biofilm height, mass, EPS composition, number of particles, shear rates and time. It is evident that density of detached clusters is left-skewed therefore logarithm transforming of these variables will reduce their skewness and make the data more interpretable to meet our dynamic linear model assumptions.

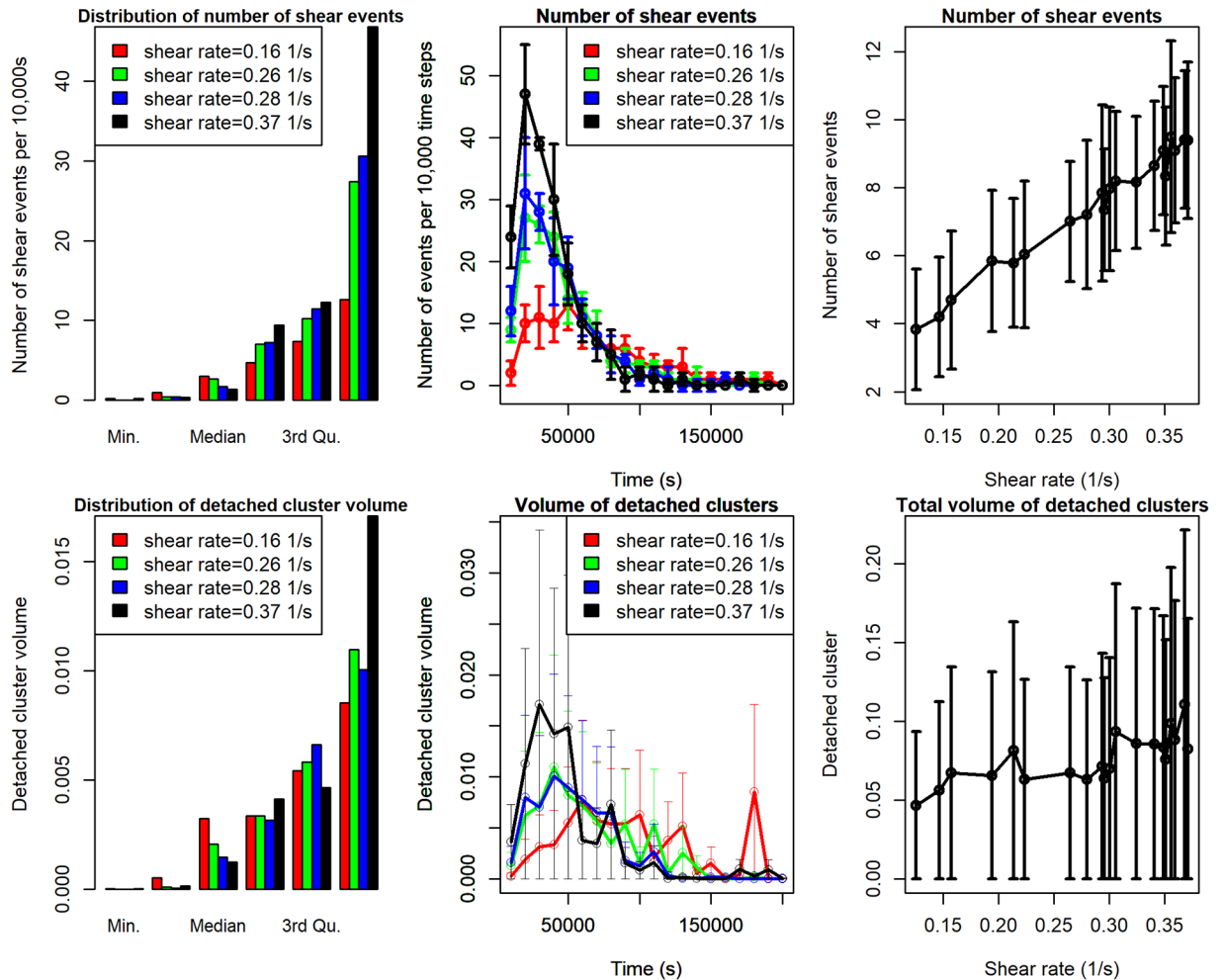


Fig 2. Density and time series plots for expected number of shear events and volume of detached clusters for different shear rates. The error bars show ± 1 standard deviation calculated from five replicates. Note: volumes are normalized by their initial biofilm volume.

<https://doi.org/10.1371/journal.pone.0195484.g002>

The corresponding time-series plots are displayed in Fig 2 (middle column) for different shear rates. There is a reduction in the number of shear events after reaching the maximum values for shear rates $0.37s^{-1}$, $0.28s^{-1}$ and $0.26s^{-1}$ respectively. A phenomenon which can be attributed to a more frequent detachment of smaller cells from biofilm surface at the beginning of experiment $< 30,000s$, often called erosion detachment. At a low shear rate of $0.16s^{-1}$, the number of shear events increases slowly and is relatively constant over time afterwards. We can conclude that the number of shear events under different shear rates has slightly varying patterns.

Similar to the number of events, the detached volume has a different trend under different shear rates which increase slowly over the time. For instance, at shear rates of $0.37s^{-1}$ and $0.28s^{-1}$ in Fig 2(middle column, bottom plot), there is a gradual increment in the volume of detached clusters due to top surface cells sheared off quite early after which there is a reduction and relatively constant detachment. This trend could be attributed to the particle at the top surface growing to a larger size because of better access to nutrients from the bulk medium. The effect of stochastic variation is pronounced because of large standard deviations around each output.

Fig 2 (column 3) shows the expected number of shear events averaged over all time and total volume of the detached cluster over all time. The growth in the number of events at higher shear rates agrees with the works of [31] and [32] who observed that doubling the shear stress frequency from 21.8 to 43.6mPa resulted in a multiple fold increase in detachment rate for both erosion and sloughing detachments.

It is apparent that while shear events increase linearly with an increase in shear rates, the total volume of detached clusters also increases nonlinearly with large stochastic variations around the mean values. We do not observe a reduction as reported in some literature, eg [32] recorded that there will be a decreasing in mean size of eroded clumps as shear rate increases.

To test the the influence of viscoelastic parameters, additional plots are given in the Supporting information (S4 Fig) where the time series of expected number of shear events and volume of detached clusters is examined for four different values of spring coefficients (K_n) ($0.0000503Nm^{-1}$, $0.0000831Nm^{-1}$, $0.0001387Nm^{-1}$, $0.0001501Nm^{-1}$). Similar to what we observe under the shear rates, higher spring coefficients gives larger number of shear events and volume of detached clusters for period $< 50,000s$. This pattern is expected because the repulsive forces increases as the spring coefficients increase making the particle to detach easily from the biofilm surface.

Other summary outputs for the simulation data which are used as explanatory variables in the Bayesian Poisson modelling are displayed in Fig 3. There is a general decreasing trend over the time, those with higher shear rates (eg $0.37s^{-1}$) decline more rapidly than those with lower shear rate (eg $0.16s^{-1}$). While the biofilm height increases slowly between the first 30,000s, the EPS composition, biofilm mass and total number of particle plots are relatively constant within this corresponding period before a gradual decrease.

The EPS composition denotes the number of EPS particles in the simulation. EPS is a gel-like material that keeps bacteria together in the biofilms. Therefore high EPS composition

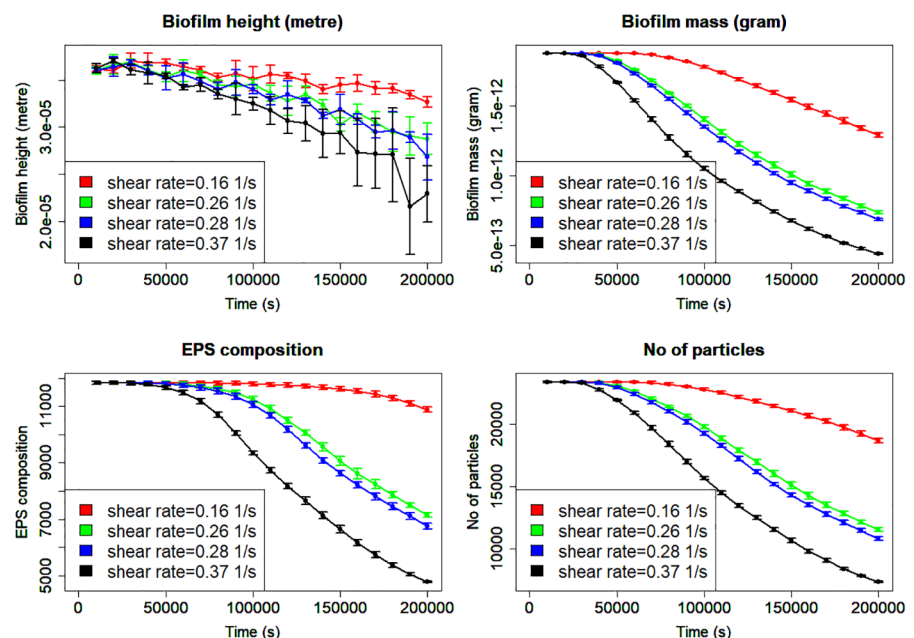


Fig 3. Dynamics of characterized outputs under different shear rates for biofilm height, mass, EPS composition and total number of particles at every 10,000s respectively. The error bars show ± 1 standard deviation calculated from five replicates. These summary outputs are used as explanatory variables for predicting the expected number of shear events using Poisson regression. Note: EPS composition is the number of EPS particles.

<https://doi.org/10.1371/journal.pone.0195484.g003>

rate will favour attachment of bacteria. EPS composition declines at a faster rate than biofilm height. The biofilm mass, EPS composition and total number of particles have a relatively similar trend which declines rapidly as expected because the clusters are being continuously detached from the surface. [33] also observed an exponential and asymptotic decrease of the biofilm thickness and mass when exposed to high shear stress. The effect of stochastic variation is considerably larger for biofilm height at the shear rate of $0.37s^{-1}$ and increases with time.

Methods

The Bayesian framework involves combining observed data with a likelihood, and a prior distribution on unknown parameters to obtain the posterior distribution of the parameter given the data. It is often difficult to derive the posterior distribution in a closed form in most applications, and this often results in the use of Markov chain Monte Carlo simulation methods as an alternative. Markov chain Monte Carlo has been widely used in many complex applications for parameter estimation. [34] describes MCMC as a general tool for simulation of complex stochastic processes useful for making statistical inference. MCMC produces a sequence of random variables which can be used to approximate the true posterior distribution.

Gibbs sampling, for instance, is based on the principle that the knowledge of the conditional distributions is often sufficient to determine a joint distribution [35]. On the other hand, the Metropolis-Hastings algorithm allows one to make random draws from such a non-standard posterior distribution using proposal distributions. Moreover, [35] gives a detailed explanation of the theory behind Gibbs sampling while [36] highlights the theoretical background behind the Metropolis-Hastings algorithm.

Dynamic linear models (DLMs)

Let us consider bacterial biofilms transported in a fluid flow with the following morphological characteristics measured on them. The volume of the detached cluster (Y_t), number of shear-ing events (noe) and time to a detachment event (t). We are interested in developing a surrogate model for predicting the volume of the detached cluster. The input variables for modelling this output are the 7 parameters listed in Table 1 ($\mu_{m,HET}$, $K_{s,HET}$, Y_{HET} , γ , K_n , γ_n and K_e) and expected number of shear events (noe). We propose to use a dynamic linear model for modelling the log-transformed volume of the detached cluster because of the time series nature of our data. The dynamic linear model is an extension of standard linear regression models with incorporation of time-varying regression coefficients [37]. Therefore, a dynamic estimation of our model parameters will enable us to have a better understanding of the complex problem we are addressing. A dynamic model is usually given as a pair of equations such that for $t > 0$, we have

$$\begin{cases} Y_t = F_t \beta_t + v_t, & v_t \sim N(0, V_t), & (a) \\ \beta_t = G_t \beta_{t-1} + w_t, & w_t \sim N(0, W_t), & (b) \end{cases} \quad (1)$$

where F_t is an $m \times p$ dynamic regression matrix (explanatory variables such that $F_t = \{\mu_{m,HET}, K_{s,HET}, Y_{HET}, \gamma, K_n, \gamma_n, K_e, noe\}$) and G_t is an $p \times p$ state evolution matrix. v_t and w_t are two independent Gaussian random vectors with mean 0 and variances V_t and W_t , respectively. W_t is the evolution variance matrix for β_t and V_t is the observation variance matrix while β_t is an $p \times 1$ vector of regression parameters. We assume that matrices of unknown parameters are time-invariant, $G_t = G$, $V_t = V$ and $W_t = W$. Suppose further that the matrix of the explanatory variable is also time-invariant, then we have $F_t = F$. Eq 1(a) and 1(b) are usually called

observation and state equations, respectively. Let

$$\beta_0 \sim N(m_0, C_0), \tag{2}$$

where m_0 and C_0 are known constants fixed in this analysis. Combining these two equations above, we can easily infer that $Y_t|\beta_t \sim N(F_t\beta_t, V_t)$ and $\beta_t|\beta_{t-1} \sim N(G_t\beta_{t-1}, W_t)$. The two equations above can be applied together for making one-step ahead predictions of our output of interest, ie volume of the detached particle. We can sequentially estimate the dynamic state β_t given the data, using a recursive pair of matrix equations, often referred to as the Kalman filter. For instance, to predict observations Y_{t+1} based on data $Y_{1:t}$ we can first estimate β_{t+1} of the state vector and use the estimated values for making predictions Y_{t+1} . In other words, we can obtain the one-step ahead observation predictive density $\pi(Y_{t+1}|Y_{1:t})$, from one-step ahead state predictive density $\pi(\beta_{t+1}|Y_{1:t})$. See Supporting information [S1 Text](#) for further details.

Here, we describe a Bayesian method where the unknown parameters are treated as random variables. In our case, the unknown parameters that are required to be estimated are the state evolution matrix G_t and the evolution and observation variance matrices W_t and V_t . To simplify our approach and make the problem identifiable, we assume they are diagonal matrices and constant over time. Therefore, we define the matrices of unknown parameters as below

$$\begin{cases} G_t = G = \text{diag}(\psi_1, \dots, \psi_k), \\ W_t = W = \text{diag}(\phi_{y,1}^{-1}, \dots, \phi_{y,m}^{-1}), \\ V_t = V = \text{diag}(\phi_{\beta,1}^{-1}, \dots, \phi_{\beta,p}^{-1}), \end{cases} \tag{3}$$

and take the parameters of the evolution matrix G to be normally independent and identically distributed such that

$$\psi_j \sim N(\psi_0, \tau_0), \quad j = 1, \dots, k, \tag{4}$$

and $\phi_{y,i}^{-1}$ and $\phi_{\beta,j}^{-1}$ to have independent gamma distributions

$$\begin{cases} \phi_{y,i}^{-1} = Ga(\alpha_{y,i}, b_{y,i}), & i = 1, \dots, m, \\ \phi_{\beta,j}^{-1} = Ga(\alpha_{\beta,j}, b_{\beta,j}), & j = 1, \dots, p. \end{cases} \tag{5}$$

Further derivation and a summary of the Gibbs sampling algorithm we use are given in [S1 Text](#) of the Supporting information.

Poisson regression

The Poisson model is employed for modelling event count data, eg the number of shearing events or organisms in an experiment. One of the key properties of count data is that they must be non-negative integers. A Poisson regression model expresses the logarithm of a response or dependent variable (count or rate data) as a linear function of a set of predictor variables. Such a log-linear Poisson model is often adopted to describe a time series of counts or rates. The model assumes the outcome y_k to be Poisson with mean λ_k , so that for a univariate predictor variable x_i (eg biofilm height), the model is

$$y_k \sim \text{Poisson}(\lambda_k), \tag{6}$$

$$\lambda_k(x) = \exp\left(\sum_{i=1}^p x_{ik}\beta_i\right), \tag{7}$$

where $k = 1, \dots, n$, $\lambda_k(x)$ is the exponential-mean function and $\mathbf{B} = (\beta_1, \dots, \beta_p)$ is the vector of unknown parameters and x_{ik} are the explanatory variables. A discrete random variable \mathbf{Y} with the probability mass function of \mathbf{Y} given as

$$f(k; \lambda) = \Pr(\mathbf{Y} = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \tag{8}$$

with parameter $\lambda > 0$, for $k = 0, 1, 2, \dots$ is regarded as a Poisson distribution. The mean and variance of a Poisson-distributed random variable are both equal to λ . Parameters \mathbf{B} are unknown and need to be estimated.

We have seen earlier in Fig 3 (middle column, top-plot) that the number of shear events has different temporal patterns for different shear rates, and also large stochastic variations (third column, top-plot). We apply a Bayesian MCMC algorithm to efficiently estimate our parameters and make reliable predictions, including a measure of uncertainty. Adopting a fully Bayesian approach, the Poisson likelihood function is given by

$$L(\mathbf{B}) = \frac{1}{\prod_{k=1}^n y_k!} \exp \left[- \sum_{k=1}^n \exp \left(\sum_{i=1}^p x_{ik} \beta_i \right) + \sum_{i=1}^p \beta_i \sum_{k=1}^n x_{ik} y_k \right]. \tag{9}$$

Let the prior π for parameter β_i be given as an independent normal distribution with mean m_i and variance v_i , ie $\beta_i \sim N(m_i, v_i)$. Under this procedure we will have a joint density of \mathbf{B} given as

$$\pi(\beta_1, \dots, \beta_p) = \prod_{i=1}^p \frac{1}{\sqrt{(2\pi v_i)}} \exp \left(- \frac{1}{2v_i} (\beta_i - m_i)^2 \right). \tag{10}$$

Using Bayes' theorem, the posterior distribution is proportional to the product of the likelihood function and the joint prior of all parameters. Here, the posterior distribution of β_i conditioning on the given data can be obtained by combining Eqs 9 and 10 above as

$$\pi_y(\beta_1, \dots, \beta_p) \propto \exp \left[- \sum_{i=1}^p \frac{\beta_i^2}{2v_i} - \sum_{i=1}^p \alpha_i \beta_i - \sum_{k=1}^n \exp \left(\sum_{i=1}^p x_{ik} \beta_i \right) \right], \tag{11}$$

where $\alpha_i = \frac{m_i}{v_i} + \sum_{k=1}^n x_{ik} y_k$. We note that there is no conjugacy between the Poisson likelihood and normal prior distribution which makes exact inference analytically infeasible [19]. We use Poisson regression to model expected number of shear events per unit time and apply Bayesian MCMC to estimate the parameters of the model by assigning a prior distribution on the regression parameters. The expected value for the Poisson model can be derived from the posterior draws of \mathbf{B} based on the MCMC iterations and is given by

$$E(\mathbf{y}|\mathbf{X}) = \lambda_k = \exp(\mathbf{x}_k \mathbf{B}), \tag{12}$$

while the predictions are draws from the Poisson distribution with parameter λ_k [38].

Results

Procedure for modelling outputs

We use data from the LAMMPS model simulation output. We consider two different simulation datasets in this paper. The first dataset is the expected number of shearing events per unit time. The second dataset is the volume of detached biofilm clusters per unit time. The input variables to the simulator are the seven parameters listed in Table 1. They are $K_{s,HET}$, $\mu_{m,HET}$, Y_{HET} , shear rate γ , spring coefficient for collision K_n , viscous coefficient γ_n and EPS stiffness

K_e . These seven parameters in addition to the number of shear events noe are used for predicting the volume of detached clusters. The four auxiliary variables of total number of particles, EPS composition, biofilm height and mass (Fig 3) are computed summary statistics. These four variables including shear rates and time are used for predicting the expected number of events.

Here, we present the results of our analysis. We based our analysis on the last 200,000 s, corresponding only to the period when the shear flow was applied. We chose to further reduce the dimension of our data by averaging at every 10,000 s which made handling and processing of the data much easier. In other words, our outputs of interest are given as the number of events per 10,000s and detached volume per 10,000s. We have 140 simulations with five replicates for each of them. Our data are averaged and taken to be deterministic. We subdivided our data into two groups. We use 130 data points as a training dataset and use the remaining 10 data points as the test data to verify the performance of our surrogate models.

Bayesian Poisson results

To proceed with our analyses, we first fitted a Bayesian Poisson regression to the number of shear events as a quadratic function of time, number of particles, shear rates, EPS composition, biofilm height and mass. We used a quadratic model of the form

$$\sum_{i=0}^p \sum_{j=0}^p \beta_{i,j} \mathbf{x}_{ik} \mathbf{x}_{jk}, \tag{13}$$

where each $\mathbf{x}_{,k}$ for $k = 1, \dots, n$ represents an explanatory variable ($p = 6$) in this subsection. In the Bayesian context, the data are augmented by a prior distribution. This prior information given over the parameters is then combined with the likelihood function using Bayes theorem to give the posterior distribution of the parameters. We used MCMC to estimate the unknown β parameters. Our prior for each variable is taken as a normal distribution. To initialize our algorithm we used the maximum likelihood estimates of β as the starting values. The prior mean m_i is taken as 0.5 for all the six β values and prior precision $v_i^{-1} = 0$.

We generate samples from the posterior distribution of Poisson regression given in Eq 11 using a Metropolis algorithm. We run the algorithm for 6,000,000 MCMC iterations, with a burn-in of 1000 samples discarded to remove the influence of the starting point from the estimate. We kept every 1000 iterations (thin = 1000) to reduce the autocorrelation in the saved MCMC samples. Fig 4 shows the diagnostic plots for examination of MCMC samples for convergence. We provide the estimates for the shear rate, biofilm height and EPS composition. The posterior density plots provide information about the shape of posterior distributions of parameters. The ergodic means (middle column) from MCMC samples are relatively stable after 1000 simulations. The ergodic means are computed by a batch mean technique where the stationary Markov chains are divided into different batches after removing the burn-in. The mean and standard error based on the average of the batch means are then calculated. The trace plots indicated a well-mixed chain. We can conclude that the convergence of the MCMC has been reached to estimate the posterior means of unknown parameters. The plots for other parameters are not shown in this paper but also indicated that those parameters have converged.

Now, we test the performance of fitted Poisson regression models on the 10 left-out observations. Fig 5 is the cross-validation plot that compares the expected number of shear events under four different shear rates from simulation and the predictive model. Each output is plotted against time. As we earlier observed in Fig 2, there is a moderate linear increase in the number of shear events until a threshold value is reached and then gradually declines

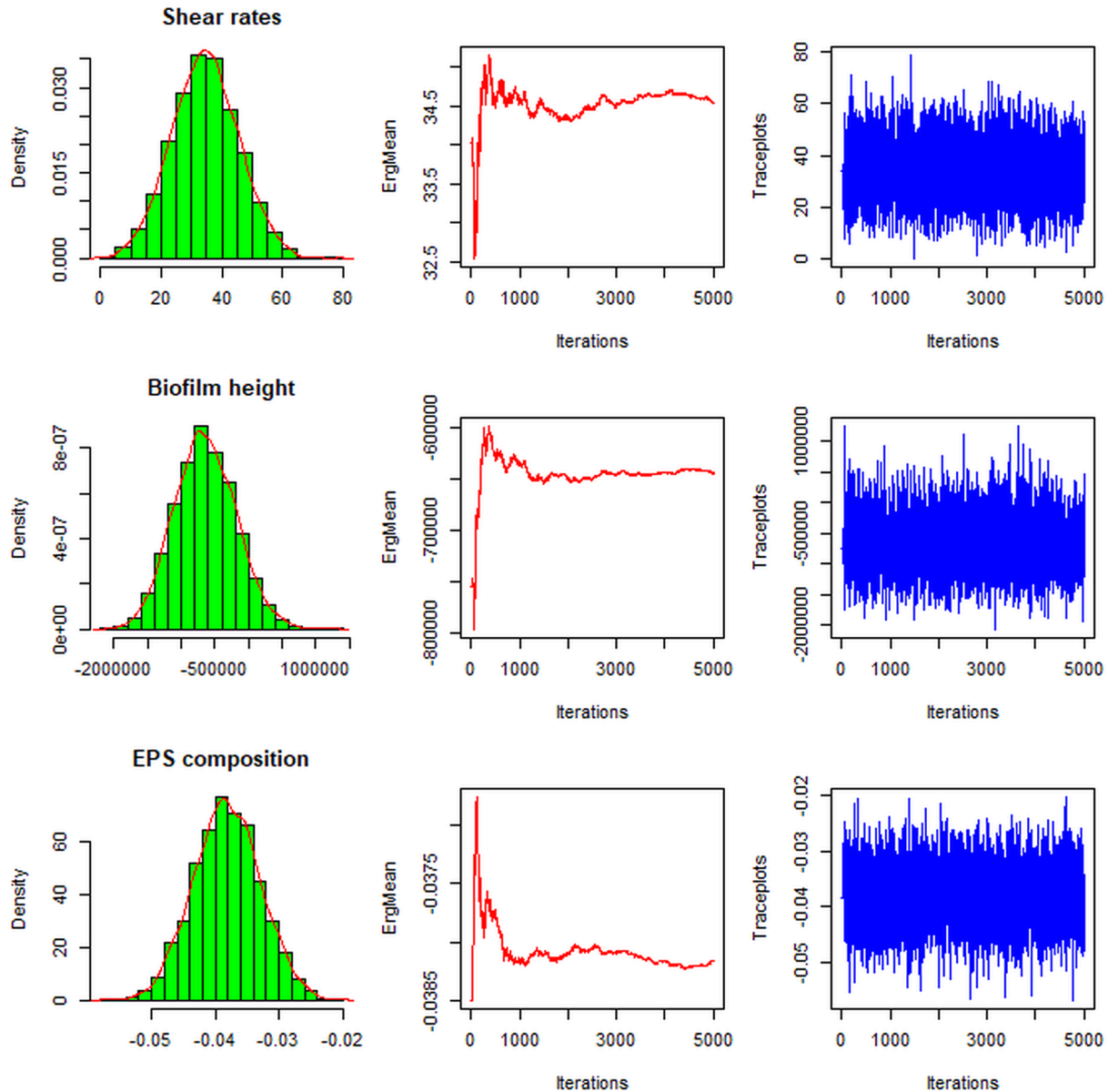


Fig 4. Diagnostic plots showing the convergence of some of the estimated θ parameters of Bayesian Poisson model for expected number of events. The first column shows the posterior density of the state variables. The middle column is the running ergodic means of MCMC samples. The third column is the trace plots for the MCMC samples.

<https://doi.org/10.1371/journal.pone.0195484.g004>

afterwards. The patterns are consistent with different shear rates. Overall, the four results in this Fig 5 are well predicted as most of the simulated values lie within the 95% probability intervals. We assess the overall performance of Poisson model by computing the root mean squared value (RMSE = 3.24) and percentage of variance explained for the left out data points ($\rho = 91\%$).

Dynamic linear model results

We next fit the dynamic linear model to the volume of detached clusters using Eq 1 where $m = 130$, $p = 8$ here and $k = p$. We standardize our input data to range over $[0, 1]$. This

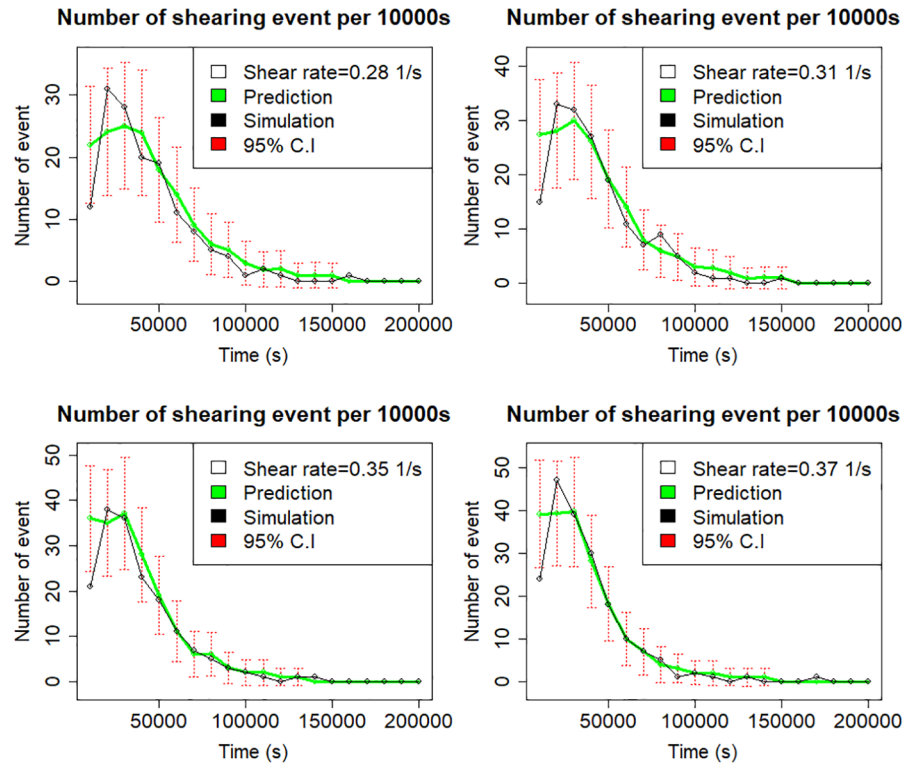


Fig 5. Comparison between the simulation and prediction for expected number of shearing events per 10000s for different shear forces.

<https://doi.org/10.1371/journal.pone.0195484.g005>

transformation will eliminate the effect of different measurement units and will enable us to get better parameter estimates. The data is normalized by centering the column with their respective minimum values and divided by their range such that $x' = (x - x_{min}) / (x_{max} - x_{min})$.

We initialized the Markov chain sampler with the following prior hyperparameters; $\psi_0 = 0$; $\tau_0 = 1$; $\alpha_{y,i} = 3$; $b_{y,i} = 0.01$; $\alpha_{\beta,j} = 3$; $b_{\beta,j} = 1$, $m_0 = 0$, $C_0 = I_p$, where I_p is a diagonal matrix of ones. We run the DLM algorithm for 1,000,000 MCMC iterations, keeping every 100 iterations (thin = 100) in order to reduce the autocorrelation in the saved MCMC samples where a burn of 5000 samples is removed before making the diagnostic plots.

Fig 6 shows the diagnostic plots obtained from the MCMC outputs for the three randomly selected regression parameters. The posterior density plots provide information about the shape of posterior distributions of parameters. The ergodic means (middle column) from MCMC samples are relatively stable after 1000 iterations. We can conclude that the convergence of the MCMC has been reached to estimate the posterior means. Diagnostic plots for three randomly selected values from the state and observation variance (V and W) parameters are displayed in the S1 and S2 Figs of the Supporting information. These plots are similar in pattern to Fig 4 in term of convergence while the diagnostic plots for evolution matrix G are not shown but also indicated convergence.

Now, we test the performance of fitted models on the left-out observations. S3 Fig of the Supporting information is the cross-validation plot that compares the detached volume under four different shear rates for the simulator and emulator predictions. The plot for each shear rate has a relatively similar pattern. Similar to what we earlier saw in Fig 2 (middle panel), the detached volume grows linearly over time < 40,000s for all shear rates and a rapid increase

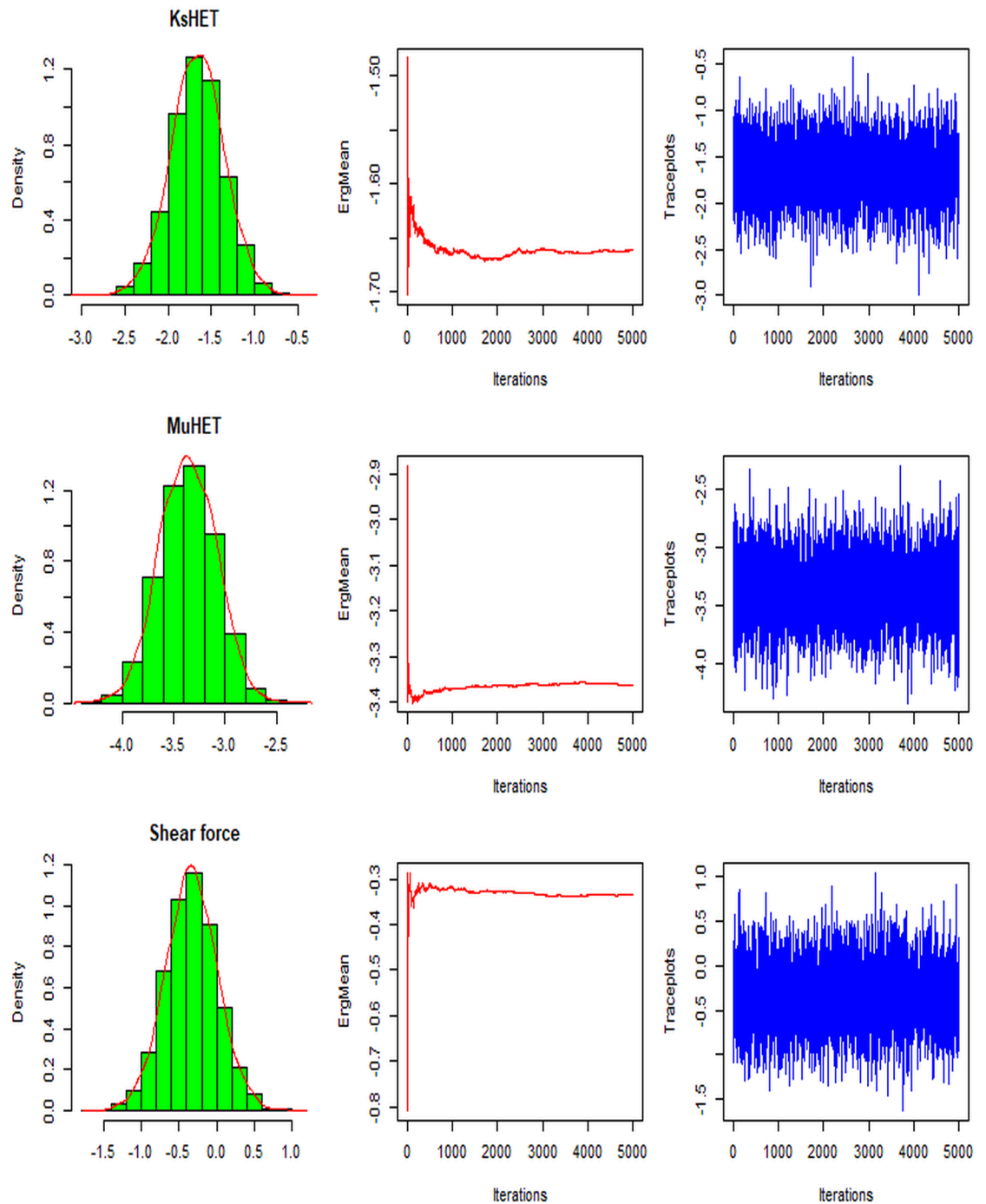


Fig 6. Diagnostics plots showing the convergence of the three randomly chosen regression parameters θ of the Bayesian dynamic linear model at time 10000s. The first column shows the posterior density of the state variables. The middle column is the running ergodic means of MCMC samples. The third column is the trace plots for the MCMC samples.

<https://doi.org/10.1371/journal.pone.0195484.g006>

followed by a moderately decreasing trend. There is a consistency in the pattern observed for the four selected shear rates. Overall, the simulated output values and that of predictions are relatively close. The degree of closeness reflects the accuracy of our DLM model. The uncertainty levels are a little bit higher for the first time point compared to the remaining time

points. The percentage of variance explained and root mean squared error (RMSE) for this model are 81.5% and 1.023 respectively.

Sensitivity analysis

To further understand the dynamics of the system we are modelling, the relative contribution of each variable to the total output variance is explored. We perform dynamic sensitivity analysis because of time-dependent nature of our data. We examine how sensitive the log-transformed volume of detached clusters are to changes in parameters over time. We use the Sobol method which calculates indices by variance decomposition. We compute the first order and total indices. Suppose our model is represented by $y_t = f(\mathbf{x}_{1,t}, \dots, \mathbf{x}_{p,t})$. The first order index is given as $S_{i,t} = \frac{\text{Var}[E(y|x_{i,t})]}{\text{Var}(y_t)}$, where $\text{Var}[E(y|x_{i,t})]$ is partial variance or the main effect of variable x_i , and $\text{Var}(y)$ is the total variance of the response y [39].

We sampled 10,000 observations from a uniform distribution for each of the eight input variables. It is also possible to sample directly from the DLM Markov chain results. The relative importance of each parameter is shown in Fig 7. We observed that detached cluster volume is mostly sensitive to the number of shear events (noe), shear rates (γ), yield coefficient for heterotrophic bacteria (Y_{HET}) and EPS stiffness (K_e). At earlier times, the sensitivity of EPS stiffness is high and gradually becomes less sensitive at the later time. On the contrary, the sensitivity of the number of shear events grows over time.

While the sensitivity of viscous coefficient is relatively moderate and constant in the first few time points, its index reduces at a later time. The effect of shear rates, on the contrary, is pronounced at later times. The parameters $K_{s,HET}$, μ_{HET} and spring coefficient have very low indices, an indication that the volume of the detached cluster does not react greatly to a change in these parameter values. It is obvious that sensitivity of the model parameters is temporally dynamic, emphasising the significance of conducting the sensitivity analysis across multiple time points. Overall, the noe , γ , Y_{HET} , K_e are the four principal determinants of the volume of detached clusters.

Discussion and conclusion

There is a significant change in the morphology and dynamics of biofilm formation when a shear flow is applied on a mature biofilm as seen in Fig 1. Also, it is obvious that shear force affects the biofilm structure in line with [2] observations. Moreover, at higher shear rates, a more dense and stable biofilm is likely to be produced because of stronger adherence from EPS matrix than those subjected to lower shear forces.

The role of interactive effects of shear force and other factors like pH and temperature on the biofilm fragmentation should be explored with the surrogate model but we do not currently have access to this simulation data in our study. We also remark that biofilms belong to viscoelastic materials and this property is also a significant determinant of the deformation behaviour of biofilm growing under shear flow as seen in the sensitivity results. In this study, a shear flow is applied to a pre-grown biofilm of certain height to explore the detachment event. It is also possible to simultaneously model both the attachment and detachment, but only the dominant process will be explicitly modelled as noted in [26]. This implies that if the detached velocity is greater the attached velocity only the net detachment will be modelled and there will be no particle attachment.

In summary, the influence of hydrodynamic shear force on biofilm fragmentation has been examined. We have developed a surrogate-based model for quantifying the effect of shear stress on the volume of detached clusters and number of shear events. This paper provides new insights on how advanced statistical techniques can be used to simplify and study biofilm

Sensitivity analysis

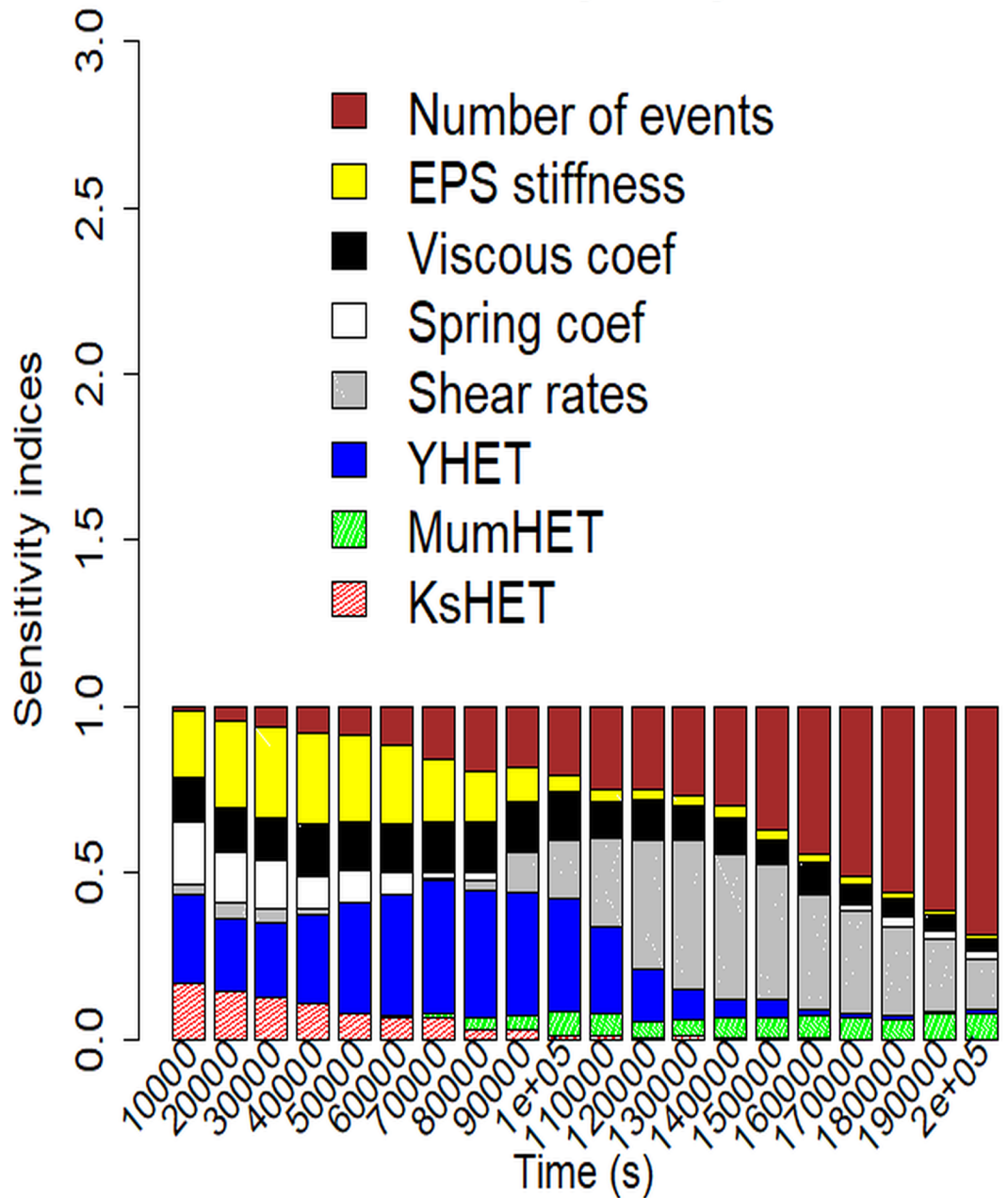


Fig 7. Barplots showing the sensitivity indices of the volume of detached clusters for the eight input variables over time.

<https://doi.org/10.1371/journal.pone.0195484.g007>

deformation and bacteria detachment. We note that it is essential to develop a cheaper predictive model of biofilm deformation and bacteria detachment in response to mechanical forces and growth parameters because the knowledge can advance the performance and operational stability of wastewater reactors. For instance, the surrogate model can be incorporated into the NUFEB model at mesoscale to produce more refined NUFEB models that are computationally efficient for providing information on a large scale such as WWTP.

The biofilm simulation was initialized and grown for 40000 s without flow then subjected to shear flow to reduce the biofilms size because of the predominance of the breakup process. This results in biofilm of smaller size than the original size due to the shearing event. The volume of biofilm that gets sheared-off and the number of shear event over time are recorded for different shear rates. This study examines the extent to which the shear force affect the number of shear events and volume of detached clusters using a cheaper surrogate model. The joint impact of shear stress and other covariates are examined on biofilm of different sizes. We assume that each occurrence of shearing can be modelled in terms of an event.

We used a 10,000s averaging as a strategy to condense the time series data. It will be interesting to assess the effect of this averaging on our predictive models. In our analyses, we have used normal and gamma priors because they are flexible and widely employed in various applications for modelling with Bayesian MCMC. The limitation of the MCMC algorithm is that the computational cost of a large parameter space is high. We compute the average number of shear events and volume of detached clusters that occur over time. We observe that the number of shear events increases until maximum values after which there is a gradual reduction. We used a Bayesian Poisson log-linear model to relate the expected number of shear events to characterize output summaries from the simulation.

The sensitivity analysis indicated that the number of shear events noe , shear rates γ , yield coefficient Y_{HET} and EPS stiffness K_e are the four primary variables for predicting the volume of detached clusters and are less affected by $K_{s,HET}$ and μ_{HET} . We can conclude that the growth, structure and performance of bacteria biofilms are highly related to the hydrodynamic shear force. The IB model simulation implemented within LAMMPS is computationally expensive, and our surrogate models are much faster to run than the simulator. Under different parameter combinations, it takes an average of between 8-11 hours to simulate both the growth and detachment patterns for about 3 days at 2000s timestep on a Linux cluster machine. Apart from the computational time required to estimate the necessary parameters, the emulator produces the required outputs within $\approx 60s$.

This approximately 480-fold increase in computational efficiency is particularly useful as a computational tool for the simulation and analysis of multiscale biological systems. This novel combination of advanced statistical techniques for modelling biofilm detachment behaviour using a surrogate-based approach is capable of greatly reducing the computational cost of modelling across large spatial and temporal scales. This study provides a significant step towards improving the performance, robustness and stability of biofilm-based wastewater treatment plant by helping to scale-up agent based models to reactor scale.

Supporting information

S1 Text. Supporting text. Derivation of DLMs, summary of Gibbs sampling algorithm and additional Table referenced in the original article.
(PDF)

S1 Fig. Diagnostic plots. Plots showing the convergence of the three V_1 , V_2 and V_3 randomly chosen observation variance parameters of the Bayesian dynamic linear model. The first column shows the posterior density of the observation variances. The middle column is the running ergodic means of MCMC samples. The third column is the traceplots for the MCMC samples.
(TIF)

S2 Fig. Diagnostic plots. Plots showing the convergence of the three randomly chosen W_1 , W_2 and W_3 state variance parameters of the Bayesian dynamic linear model. The first column

shows the posterior density of the state variances. The middle column is the running ergodic means of MCMC samples. The third column is the trace plots for the MCMC samples. (TIF)

S3 Fig. Model comparison. Comparison between the simulation and prediction for log-transformed detached volume over time for different shear forces. The results are normalized by initial biofilm volume. (TIF)

S4 Fig. Time series plots. Expected number of shear events and volume of detached clusters for different spring coefficients for elastic collision. (TIF)

Acknowledgments

We thank the NUFEB modelling team for their useful comments that have helped improve this paper.

Author Contributions

Conceptualization: Oluwole K. Oyebamiji, Steve P. Rushton, Ben Bridgens, Bowen Li, Paolo Zuliani.

Data curation: Oluwole K. Oyebamiji, Pahala Gedara Jayathilake.

Formal analysis: Oluwole K. Oyebamiji, Pahala Gedara Jayathilake, Steve P. Rushton.

Funding acquisition: Darren J. Wilkinson.

Investigation: Oluwole K. Oyebamiji, Pahala Gedara Jayathilake, Ben Bridgens.

Methodology: Oluwole K. Oyebamiji, Paolo Zuliani.

Project administration: Darren J. Wilkinson.

Resources: Pahala Gedara Jayathilake, Steve P. Rushton.

Software: Bowen Li, Paolo Zuliani.

Supervision: Darren J. Wilkinson.

Validation: Oluwole K. Oyebamiji, Darren J. Wilkinson, Bowen Li.

Writing – original draft: Oluwole K. Oyebamiji.

Writing – review & editing: Darren J. Wilkinson, Steve P. Rushton, Ben Bridgens.

References

1. Merkey BV, Lardon LA, Seoane JM, Kreft JU, Smets BF. Growth dependence of conjugation explains limited plasmid invasion in biofilms: an individual-based modelling study. *Environmental microbiology*. 2011; 13(9):2435–2452. <https://doi.org/10.1111/j.1462-2920.2011.02535.x> PMID: 21906217
2. Liu Y, Tay JH. The essential role of hydrodynamic shear force in the formation of biofilm and granular sludge. *Water Research*. 2002; 36(7):1653–1665. [https://doi.org/10.1016/S0043-1354\(01\)00379-7](https://doi.org/10.1016/S0043-1354(01)00379-7) PMID: 12044065
3. Li C, Zhang Y, Yehuda C. Individual based modeling of *Pseudomonas aeruginosa* biofilm with three detachment mechanisms. *RSC Advances*. 2015; 5(96):79001–79010. <https://doi.org/10.1039/C5RA11041F>
4. Bryers J. Modeling biofilm accumulation. *Physiological models in microbiology*. 1988; 2:109–144.

5. Xavier JdB, Picioreanu C, van Loosdrecht M. A general description of detachment for multidimensional modelling of biofilms. *Biotechnology and bioengineering*. 2005; 91(6):651–669. <https://doi.org/10.1002/bit.20544>
6. Choi Y, Morgenroth E. Monitoring biofilm detachment under dynamic changes in shear stress using laser-based particle size analysis and mass fractionation. *Water Science and Technology*. 2003; 47(5):69–76. PMID: [12701909](https://pubmed.ncbi.nlm.nih.gov/12701909/)
7. Rittmann B, Trinet F, Amar D, Chang H. Measurement of the activity of a biofilm: Effects of surface loading and detachment on a three-phase, liquid-fluidized-bed reactor. *Water Science and Technology*. 1992; 26(3-4):585–594.
8. Kommedal R, Bakke R. Modeling *Pseudomonas aeruginosa* biofilm detachment. 2003; p. 3.
9. Picioreanu C, Van Loosdrecht MC, Heijnen JJ. Two-dimensional model of biofilm detachment caused by internal stress from liquid flow. *Biotechnology & Bioengineering*. 2001; 72(2):205–218. [https://doi.org/10.1002/1097-0290\(20000120\)72:2<205::AID-BIT9>3.0.CO;2-L](https://doi.org/10.1002/1097-0290(20000120)72:2<205::AID-BIT9>3.0.CO;2-L)
10. Picioreanu C, Van Loosdrecht MC, Heijnen JJ. Effect of diffusive and convective substrate transport on biofilm structure formation: a two-dimensional modeling study. *Biotechnology and bioengineering*. 2000; 69(5):504–515. [https://doi.org/10.1002/1097-0290\(20000905\)69:5<504::AID-BIT5>3.0.CO;2-S](https://doi.org/10.1002/1097-0290(20000905)69:5<504::AID-BIT5>3.0.CO;2-S) PMID: [10898860](https://pubmed.ncbi.nlm.nih.gov/10898860/)
11. Kreft JU, Booth G, Wimpenny JW. BacSim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology*. 1998; 144(12):3275–3287. <https://doi.org/10.1099/00221287-144-12-3275> PMID: [9884219](https://pubmed.ncbi.nlm.nih.gov/9884219/)
12. Oyebamiji OK, Edwards NR, Holden PB, Garthwaite PH, Schaphoff S, Gerten D. Emulating global climate change impacts on crop yields. *Statistical Modelling*. 2015; 15(6):499–525. <https://doi.org/10.1177/1471082X14568248>
13. Oyebamiji O, Wilkinson D, Jayathilake P, Curtis T, Rushton S, Li B, et al. Gaussian process emulation of an individual-based model simulation of microbial communities. *Journal of Computational Science*. 2017; 22:69–84. <https://doi.org/10.1016/j.jocs.2017.08.006>
14. Young PC, Ratto M. Statistical Emulation of Large Linear Dynamic Models. *Technometrics*. 2011; 53(1):29–43. <https://doi.org/10.1198/TECH.2010.07151>
15. Oakley JE, O'Hagan A. Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2004; 66(3):751–769. <https://doi.org/10.1111/j.1467-9868.2004.05304.x>
16. Kennedy MC, O'Hagan A. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2001; 63(3):425–464. <https://doi.org/10.1111/1467-9868.00294>
17. Shi JQ, Murray-Smith R, Titterton D. Bayesian regression and classification using mixtures of Gaussian processes. *International Journal of Adaptive Control and Signal Processing*. 2003; 17(2):149–161. <https://doi.org/10.1002/acs.744>
18. Doss H, Narasimhan B. Bayesian Poisson regression using the Gibbs sampler: Sensitivity analysis through dynamic graphics. Technical report, Citeseer; 1994.
19. Chan AB, Vasconcelos N. Bayesian poisson regression for crowd counting. In: *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE; 2009. p. 545–551.
20. Ma J, Kockelman K. Bayesian multivariate Poisson regression for models of injury count, by severity. *Transportation Research Record: Journal of the Transportation Research Board*. 2006;(1950):24–34. <https://doi.org/10.3141/1950-04>
21. Jayathilake PG, Gupta P, Li B, Madsen C, Oyebamiji O, González-Cabaleiro R, et al. A mechanistic Individual-based Model of microbial communities. *PloS one*. 2017; 12(8):e0181965. <https://doi.org/10.1371/journal.pone.0181965> PMID: [28771505](https://pubmed.ncbi.nlm.nih.gov/28771505/)
22. Jayathilake PG, Jana S, Rushton S, Swailes D, Bridgens B, Curtis T, et al. Extracellular Polymeric Substance Production and Aggregated Bacteria Colonization Influence the Competition of Microbes in Biofilms. *Frontiers in microbiology*. 2017; 8:1865. <https://doi.org/10.3389/fmicb.2017.01865> PMID: [29021783](https://pubmed.ncbi.nlm.nih.gov/29021783/)
23. Wingender J, Neu TR, Flemming HC. What are bacterial extracellular polymeric substances? In: *Microbial extracellular polymeric substances*. Springer; 1999. p. 1–19.
24. Plimpton S. Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics*. 1995; 117(1):1–19. <https://doi.org/10.1006/jcph.1995.1039>
25. Santner TJ, Williams BJ, Notz WI. The design and analysis of computer experiments. Springer Science & Business Media; 2013.

26. Wanner O, Reichert P. Mathematical modeling of mixed-culture biofilms. *Biotechnology and bioengineering*. 1996; 49(2):172–184. [https://doi.org/10.1002/\(SICI\)1097-0290\(19960120\)49:2<172::AID-BIT6>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1097-0290(19960120)49:2<172::AID-BIT6>3.0.CO;2-N) PMID: 18623567
27. Schluter J, Nadell CD, Bassler BL, Foster KR. Adhesion as a weapon in microbial competition. *The ISME journal*. 2015; 9(1):139. <https://doi.org/10.1038/ismej.2014.174> PMID: 25290505
28. Ni BJ, Fang F, Xie WM, Sun M, Sheng GP, Li WH, et al. Characterization of extracellular polymeric substances produced by mixed microorganisms in activated sludge with gel-permeating chromatography, excitation—emission matrix fluorescence spectroscopy measurement and kinetic modeling. *Water Research*. 2009; 43(5):1350–1358. <https://doi.org/10.1016/j.watres.2008.12.004> PMID: 19215955
29. Celler K, Hödl I, Simone A, Battin T, Picioreanu C. A mass-spring model unveils the morphogenesis of phototrophic *Diatoma* biofilms. *Scientific reports*. 2014; 4. <https://doi.org/10.1038/srep03649> PMID: 24413376
30. Head D. Linear surface roughness growth and flow smoothing in a three-dimensional biofilm model. *Physical Review E*. 2013; 88(3):032702. <https://doi.org/10.1103/PhysRevE.88.032702>
31. Stoodley P, Cargo R, Rupp C, Wilson S, Klapper I. Biofilm material properties as related to shear-induced deformation and detachment phenomena. *Journal of Industrial Microbiology and Biotechnology*. 2002; 29(6):361–367. <https://doi.org/10.1038/sj.jim.7000282> PMID: 12483479
32. Walter M, Safari A, Ivankovic A, Casey E. Detachment characteristics of a mixed culture biofilm using particle size analysis. *Chemical engineering journal*. 2013; 228:1140–1147. <https://doi.org/10.1016/j.cej.2013.05.071>
33. Paul E, Ochoa JC, Pechaud Y, Liu Y, Liné A. Effect of shear stress and growth conditions on detachment and physical properties of biofilms. *Water Research*. 2012; 46(17):5499–5508. <https://doi.org/10.1016/j.watres.2012.07.029> PMID: 22898671
34. Geyer CJ. Markov chain Monte Carlo maximum likelihood. 1991; p. 156–163.
35. Casella G, George EI. Explaining the Gibbs sampler. *The American Statistician*. 1992; 46(3):167–174. <https://doi.org/10.2307/2685208>
36. Chib S, Greenberg E. Understanding the Metropolis-Hastings algorithm. *The American Statistician*. 1995; 49(4):327–335. <https://doi.org/10.1080/00031305.1995.10476177>
37. Petris G, Petrone S, Campagnoli P. *Dynamic linear models*. In: *Dynamic Linear Models with R*. Springer; 2009. p. 31–84.
38. Martin AD, Quinn KM, Park JH. Markov Chain Monte Carlo (MCMC) Package. <http://mcmcpack.wustl.edu>. 2005;.
39. Saltelli A. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*. 2002; 145(2):280–297. [https://doi.org/10.1016/S0010-4655\(02\)00280-1](https://doi.org/10.1016/S0010-4655(02)00280-1)