Inverse Uncertainty Quantification of a cell model using a Gaussian Process metamodel

de Vries, K.; Nikishova, A.; Czaja, B.; Závodszky, G.; Hoekstra, A.G.

DOI
10.1615/Int.J.UncertaintyQuantification.2020033186

Publication date
2020

Document Version
Final published version

Published in
International Journal for Uncertainty Quantification

International Journal for Uncertainty Quantification, 10(4):333–349 (2020)

INVERSE UNCERTAINTY QUANTIFICATION OF A CELL MODEL USING A GAUSSIAN PROCESS METAMODEL

Kevin de Vries, Anna Nikishova,* Benjamin Czaja, Gábor Závodszky, & Alfons G. Hoekstra

Computational Science Lab, Institute for Informatics, Faculty of Science, University of Amsterdam, The Netherlands, Amsterdam, 1098 XH

*Address all correspondence to: Anna Nikishova, Computational Science Lab, Institute for Informatics, Faculty of Science, University of Amsterdam, The Netherlands, Amsterdam, 1098 XH, E-mail: A.Nikishova@uva.nl

Original Manuscript Submitted: 12/6/2019; Final Draft Received: 6/7/2020

In order to accurately describe the mechanics of red blood cells (RBCs) and resulting fluid dynamics, a cell-resolved blood flow fluid solver is required. The parameters of the material model for the RBC membranes are carefully tuned to reproduce the behavior of real cells under various experimental conditions. In this work, uncertainty in the parameters of the material model for RBCs used in a model for RBC suspensions was estimated with Inverse Uncertainty Quantification (IUQ) using Bayesian Annealed Sequential Importance Sampling (BASIS). Due to the relatively high computational cost of the model, a Gaussian Process regression metamodel was trained in order to feasibly draw the large number of samples required to obtain an accurate posterior distribution estimate. Additionally, the identifiability of the model parameters was estimated using Sobol sensitivity indices. The elongation index of simulated RBCs in a perfect sheared environment was the model prediction used to calibrate model parameters. The results show good identifiability of the parameter defining the tensile properties of the cell membrane and viscosity ratio, and poor identifiability of the parameter defining the response of the cell surface while undergoing bending. This suggests that the latter should be identified using a different quantity of interest. Overall, the model outputs with the optimal values of the parameters obtained using the Gaussian Process metamodel match better or close to the measurements than the results with the parameters' values obtained with the original model. Therefore, we can conclude that it is a valid method to decrease the computational cost of IUQ of the model.

KEY WORDS: *inverse uncertainty quantification, Bayesian annealed sequential importance sampling, HemoCell, emulator, Markov chain Monte Carlo*

1. INTRODUCTION

Uncertainty quantification (UQ) is a topic of increasing importance in computational bio-medicine. Recent studies have applied UQ to computational models simulating the human heart [1], coronary artery in-stent restenosis [2], and ascending thoracic aortic aneurysms [3]. To advance the understanding of the physiology by computational modeling, it is important to highlight potential sources of uncertainty in those models. Therefore, validation, verification, and uncertainty quantification of the computational models should be performed to assess their fidelity.

HemoCell (High pErformance MicrOscopic CELlular Library) is an open-source framework for modeling blood flow at the cellular level [4–6]. The goal of this work is to identify the distributions of the input parameters of the material model for red blood cells (RBCs) in HemoCell. The posterior distributions of the model parameters are obtained by calibrating the parameters and quantifying uncertainty of the calibrated parameters using ektacytometry data for both healthy and stiffened red blood cells [7]. This is realized by applying both sensitivity analysis and Bayesian inverse uncertainty quantification (IUQ) to the mechanical model for red blood cells in HemoCell. More precisely, an analysis on how the force parameters impact the deformation of a single RBC in a uniform shear flow

environment is presented. The elongation index, a measure of the deformation of an RBC, is a quantity of interest (QoI) in this study, and the sensitivity of the outcome deformation given a range of force coefficient variations is investigated.

Usually, Uncertainty Quantication methods require a large number of model evaluations in order to receive reliable estimations of uncertainty. If the model is computationally expensive, like HemoCell, which requires approximately 10 min, or more, wall clock time to compute an average deformation of a single red blood cell, it is often replaced by surrogate models [8–11] which are a computationally cheap approximation of the model response. In this work, to make it feasible to obtain a sufficiently large ensemble of the model outputs, a surrogate for the deformation of the HemoCell RBC model was built using the Gaussian process (GP) regression. The surrogate can be trained, per shear rate, in approximately 1 min wall clock time, and evaluated thereafter on the order of micro seconds. Further development of such a surrogate for the deformation of a red blood cell could be applied in future large-scale blood flow models, which incorporate the mechanics of the red blood cell [12].

The novelty of this work identifies the parameters, through the Bayesian Annealed Sequential Importance Sampling method [13], needed to create a healthy and a stiffened HemoCell RBC numerical model, which is based on experimental data. It has recently been identified, through numerical simulation, that stiffened RBCs in the blood stream, as a result of disease [14,15], can decrease the localization of platelets to the vessel wall [16]. This is important as rigid RBCs may impede the body from naturally preventing and stopping bleeding during a vessel injury. Developing reliable numerical RBC models is crucial to study the influences of diseases that affect the deformability of RBCs on flowing whole blood, which may not immediately be accessible via experiment. Previous UQ work has studied bulk flow phenomena of blood such as particle adhesion [17] and diffusion [18] by inferring the mechanics of individual red blood cells. To the best of our knowledge this is the first study to perform IUQ on the input parameters of a red blood cell model to study the direct output of the mechanics of a single red blood cell, namely the deformation, and additionally proposes a surrogate model for the deformation.

Thus, the main results of this work include the probability distribution functions of the parameters and their maximum *a posteriori* estimation: the link force parameter has the value of 38.8 k_BT for healthy and 117.2 k_BT for treated cell; the obtained value of the bending force parameter is 293.2 k_BT for healthy and 371.4 k_BT for treated cell; and the viscosity ratio has the value equal to 4.9 for healthy and 2.5 for treated cell. From sensitivity analysis, it is concluded that the identifiability of the bending force coefficient is low, therefore, the obtained value can be corrected by applying the method with another QoI. Additionally, this work proposes a surrogate model, for the deformation of the RBC model, which is obtained using the GP regression for IUQ. The metamodel is much faster to execute whereby it was possible to draw a much large number of samples. Hence, the results of the experiments with the GP metamodel outperform or very close to the results with the HemoCell model.

The paper is structured as follows. In Section 2, a brief description of the HemoCell model and the ektacytometry measurements are given. Section 3 explains the methods for IUQ, sensitivity analysis, and uncertainty propagation as well as the surrogate model used in this work. The results are presented in Section 4. Section 5 summarizes and concludes the findings of this work.

2. HEMOCELL RED BLOOD CELL MODEL

Whole blood is a suspension of cells that exhibits complex non-Newtonian rheology, which is determined by its cellular components, mainly the red blood cells. In this work the free parameters of the HemoCell red blood cell material model [19] are the focus of the sensitivity analysis and Bayesian IUQ. HemoCell is a discrete element method used to describe the mechanical responses of a RBC membrane when undergoing deformation. The RBC membrane model is coupled, via the immersed boundary method, to the suspending blood plasma which is modeled with the lattice Boltzmann method. This model has been previously validated to replicate the deformation of a single RBC in pure shear flow, the membrane stretching in an optical tweezers experiment [19], and has been extended to model the viscosity contrast between the RBC cytoplasmic interior and the exterior blood plasma [20]. With the correct implementation of such a numerical RBC model, the non-Newtonian effects of flowing whole blood emerge as a result [19,21].

The free parameters of the HemoCell material model considered in this study are the link force coefficient κ_l, the bending force coefficient κ_b, and the interior viscosity contrast Λ. The link force coefficient scales the link force

which models the stretching and compression of the underlying spectrin-network, shown in Eq. (A1). The bending force coefficient κ_b scales the bending force, shown in Eq. (A2), which models the bending response of the RBC membrane arising from the non-zero thickness of the spectrin network. Both κ_l and κ_b are temperature dependent and are reported in terms of the Boltzmann constant k_b and T the temperature, which is taken at room temperature of $T = 23°\text{C}$. The force model of HemoCell is elaborated in the Appendix. As outlined in the validation of the numerical cell model [19], to reproduce the accurate mechanical behavior of the red blood cells we need to solve approximately 4000 equations per cell. The red cells are also coupled to the lattice Boltzmann fluid solver that solves the momentum exchange of the underlying blood plasma using a D3Q19 lattice [22], which results in $\approx 10^5$ elements that need to be computed in each step. This results in a typical wall time of 10 min for a simulation of sheared single cell over 5 milliseconds in low shear, running on one compute core. Bulk flow simulations, which typically include 10^3 cells, require many days of wall time on approximately 300 compute cores.

In this research the free parameters of the HemoCell red blood cell material model [19–21] are analyzed by matching the elongation index computed by a sheared single RBC in HemoCell to elongation index measured via ektacytometry from the experiments of Gutierrez et al. [7]. This is analyzed over multiple shear rates. Deformation of RBCs in this work is classified by the measured/computed elongation index (EL) which is defined as

$$\text{EL} = \frac{a-b}{a+b}, \tag{1}$$

where a is the major axis and b is the minor axis of an RBC membrane outline as it undergoes deformation due to shear. A schematic of a HemoCell RBC undergoing deformation due to shear is shown in Fig. 1. Presented in this paper is an extension of the HemoCell RBC model which aims to remain valid when undergoing much higher shear rates than previously studied, as well as proposing a new stiffened RBC model to represent RBCs with reduced deformability, such as those present in the case of diabetes.

2.1 Ektacytometry Elongation Index

The experimental elongation indices were measured from the deformation of real red blood cells using an ektacytometer. The experimental data, i.e., the elongation indices, were provided from a previous study by Gutierrez et al. [7] which measured the elongation index of healthy and stiffened RBCs across multiple shear stresses. Here collections of RBCs are sheared using the laser-assisted optical rotational cell analyzer (LORRCA; Mechatronics, Hoorn, The Netherlands) [23]. Elongation indices here are measured by the diffraction patterns of collections of RBCs undergoing increasing amount of applied shear. Elongation indices of healthy RBCs are shown as a blue line in Fig. 2 and stiffened RBCs treated with 0.5 mM tert-butyl hydroperoxide (TBHP) are shown as an orange line in Fig. 2. RBCs treated with TBHP results in a stiffened RBC membrane through oxidative stress, specifically causing spectrin and ankyrin degradation [24] of the cell membrane. This results in a chemically stiffened RBC model used to study the effects of decreased RBC deformability which is present in a variety of pathologies such as diabetes [14], sickle cell Anemia [15], malaria [25], and the human immunodeficiency virus (HIV) [26].

2.2 HemoCell Elongation Index

The numerical elongation indices from HemoCell are computed by fitting an ellipse to a single RBC in a perfect sheared environment. We follow previous single cell sheared simulations [19–21] where an RBC is placed in a

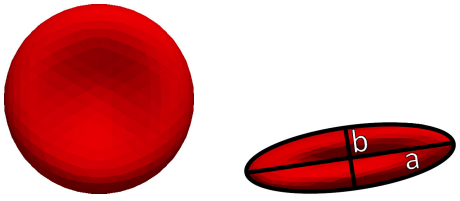

FIG. 1: The deformation of an RBC in a perfect sheared environment

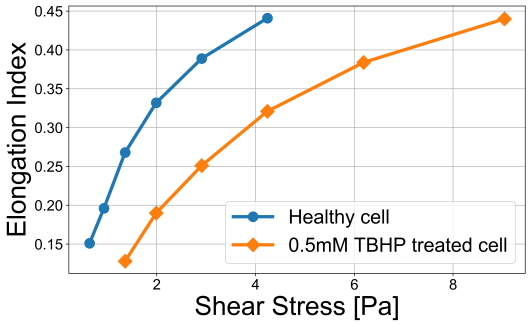

FIG. 2: Cell deformation for both healthy RBCs and RBCs treated with 0.5 mM TBHP

periodic (in the X direction) domain of dimensions $X = 20$, $Y = 20$, $Z = 10$ μm where a constant velocity V_x is provided at the top $+Y$ boundary, and $-V_x$ is provided at the bottom $-Y$ boundary of the simulation, resulting in a uniform shear along the Y axis. In order to convert to shear stress we follow the relation between shear rate and shear stress $\dot{\gamma} = \tau/\nu\rho$, where τ is the shear stress, ν is the kinematic viscosity of the blood plasma, and ρ is the density of the blood plasma. For the simulations we use the density of blood plasma $\rho = 1025$ kg/m^3 and we set the kinematic viscosity to an estimated value $\nu = 1.1 \times 10^{-6}$ m^2/s. Ektacytometry is performed on a population of RBCs, and as a result the elongation index is from an aggregate of sheared RBCs. In simulation, however, elongation index is computed from a single RBC in shear flow. The approximation that a single-cell elongation index matches with an aggregate elongation index is a source of discrepancy in the input parameters for the IUQ analysis as well as the surrogate model. This discrepancy between the data and model output is characterized using a prediction error model when performing Bayesian inference.

Red blood cells naturally exhibit a tank treading motion where the cell maintains a nearly constant angular inclination in the direction of flow while the membrane makes a continuous rotation, like the treads of a tank [27]. The HemoCell RBC also exhibits tank treading motion in shear flow; however, due to the discretization of the membrane mesh the computed elongation index will not exactly converge to a fixed value, but will oscillate around a mean value (Fig. 3). This transient behavior in the HemoCell model does not introduce any significant energy error into the system. We note that at $\tau > 2.8$ Pa the HemoCell model becomes numerically unstable. Due to the difference in force applied to the RBC at different shear rates, the lower shear stress simulations take longer wall time to converge to a stable oscillating elongation index than the higher shear stress simulations. Because of this reason we work in the characteristic timescale of strain $\dot{\gamma}t$, where each shear stress should exhibit the same convergence independent of shear rate. The evolution of the elongation measurements using the RBC model parameters determined in [19] and with the link force coefficient changed from 15 k_BT to 100 k_BT to simulate a stiffer RBC are shown in Fig. 3 in the left and right panels, respectively. We find that for the range of shear stresses used in the analysis the EL has

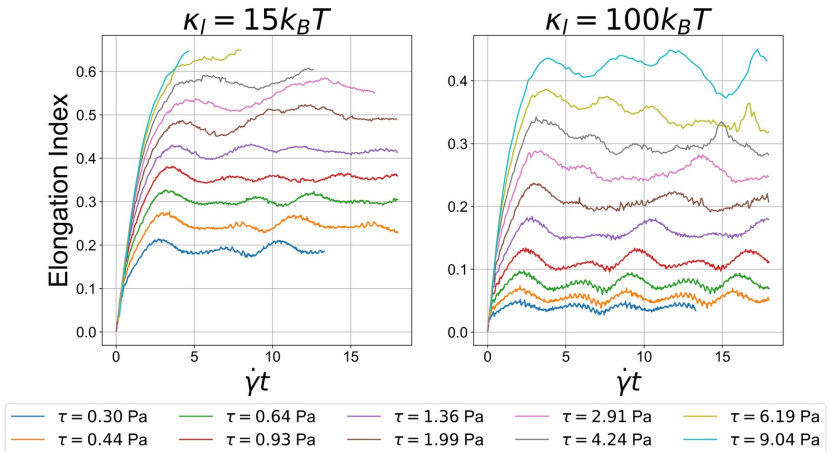

FIG. 3: Convergence of the EL while the RBC is being sheared using the default HemoCell RBC model parameters (left panel) and with κ_l increased to 100 k_BT (right panel). The parameter κ_l defines the tensile strength of the RBC membrane.

approximately converged around $\dot{\gamma}t = 7.5$ for all shear stress values, although lower shear stresses still seem to converge even before $\dot{\gamma}t = 5$. We shear the RBC until it has converged to a stable state at $\dot{\gamma}t_{conv} = 7.5$. After this the RBC is sheared until $t = 1.25 \times t_{conv}$, where the EL was measured at 10 evenly spaced time steps in the interval $[t_{conv}, 1.25 \times t_{conv}]$. We then take the arithmetic mean of the EL measurements as the Quantity of Interest (QoI) and the sample standard deviation of the EL measurements as the computational uncertainty.

3. UNCERTAINTY QUANTIFICATION AND PROPAGATION

3.1 Inverse Uncertainty Quantification

3.1.1 Bayesian Probability Modeling

The inverse Uncertainty Quantification was done using a Bayesian approach [28], where the parameters were calibrated using Bayesian updating. This was done by using Bayes' rule with the joint density $f(\theta, D)$, where θ and D are the modeling parameters and experimental dataset, respectively, to obtain

$$f(\theta|D) = \frac{f(D|\theta)\pi(\theta)}{f(D)}, \tag{2}$$

which is used to sample parameter values from the posterior $f(\theta|D)$ with the prior distribution $\pi(\theta)$ and the data likelihood given the parameters $f(D|\theta)$. The evidence $f(D)$ can be estimated from the prior and likelihood functions using

$$f(D) = \int_{\Theta} f(D|\theta)\pi(\theta)d\theta, \tag{3}$$

where we integrate over the full parameter space Θ.

3.1.2 Model Formulation

The inputs of the model are defined by $\theta_m \in \mathbb{R}^{N_m}$, where N_m is the number of model parameters. An experimental dataset is denoted by $D \equiv \{\hat{y}_j, j = 1, \ldots, N_{\hat{y}}\}$, where $N_{\hat{y}}$ is the number of observations. At the same time, the model predictions produced from simulations with particular values of the inputs θ_m are defined by $q(\theta_m|D) \in \mathbb{R}^{N_{\hat{y}}}$. Then, the measurements $\hat{y}$ can be written in terms of the model output and the discrepancy between the experimental data and the model response:

$$\hat{y} = q(\theta_m|D) + \varepsilon, \tag{4}$$

where ε is the prediction error, which can be decomposed as

$$\varepsilon = \varepsilon^d + \varepsilon^c + \varepsilon^m. \tag{5}$$

Here, ε^d, ε^c, and ε^m are given as the measurement, computational, and modeling uncertainties, respectively.

The measurement uncertainty ε^d describes the uncertainty of the experimental data caused by the often limited accuracy of the measurements from experimental equipment, which is characterized by the measurement error for each observation, respectively. The computational uncertainty ε^c describes the uncertainty from the variability of the model output. The modeling uncertainty ε^m describes the uncertainty on the model output itself. This uncertainty comes from the epistemic uncertainty of the model, such as the uncertainties in the underlying modeling assumptions and numerical approximation errors. It also partially characterizes the discrepancy between the experimental data and the model output. In contrast to the other uncertainty terms, the parameters for this uncertainty term are generally not known *a priori* and are thus estimated along with the modeling parameters [29,30].

3.1.3 Prediction Error Modeling

In order to model the exact posterior distribution the likelihood needs to be modeled, which is dependent on the prediction error model. Here $\theta = (\theta_m, \theta_e)$ is defined as the full set of input parameters, where θ_e is the set of prediction error parameters used for modeling the uncertainties in the prediction error ε given in Eq. (5). θ_m and θ_e are assumed to be independent, where the prior $\pi(\theta|M)$ is given by

$$\pi(\theta) = \pi(\theta_m) \times \pi(\theta_e). \tag{6}$$

In order to model the prediction error the maximum entropy principle is followed, which states that the distribution which best describes current knowledge is the distribution with the largest possible entropy given all known data [31]. When the interval of the possible output values is undefined for a continuous, real-valued function with a specified variance, the distribution with the largest entropy is a normal distribution. Assuming that the uncertainties are independent, the uncertainties can thus be modeled using zero-mean Gaussian random variables $\varepsilon^d \sim \mathcal{N}(0, \Sigma^d)$, $\varepsilon^c \sim \mathcal{N}(0, \Sigma^c)$, and $\varepsilon^m \sim \mathcal{N}(0, \Sigma^m)$. The individual observations are assumed to be i.i.d. and thus uncorrelated. This implies that the covariance matrices of the uncertainties are diagonal matrices, which contain only variances. For the measurement error $\Sigma^d_{ij} = \delta_{ij}(\sigma^d)^2_i$ is used, where σ^d_i is the measured error for the ith observation in $\hat{y}$ and δ_{ij} is the Kronecker delta. For the computational error $\Sigma^c_{ij} = \delta_{ij}(\sigma^c)^2_i\ \hat{y}^2_j$ is used, where σ^c is determined by the variability in the model predictions or simulations. For the modeling error $\Sigma^m_{ij} = \delta_{ij}(\sigma^m)^2_i\ \hat{y}^2_j$ is assumed. The set of prediction error parameters is given by $\theta_e = \{\sigma^m_1, \sigma^m_2, \ldots, \sigma^m_{N_{\hat{y}}}\}$. However, since the HemoCell RBC model is assumed to behave similarly over the full range of shear stresses, the modeling variance is assumed to be homoscedastic, meaning that σ^m_i is assumed to be the same for all indices i. The prediction error parameter set is thus simplified to $\theta_e = \{\sigma^m\}$, where σ^m is a scalar. The modeling standard deviation σ^m is determined during the Bayesian updating along with the modeling parameters. σ^m will tend to be relatively low when the discrepancy between the data and corresponding model output is small and relatively high when the discrepancy between the data and corresponding model output is large.

Since the uncertainties are assumed to be independent Gaussian random variables, the total uncertainty can be described using

$$\varepsilon \sim \mathcal{N}(0, \Sigma(\theta_e)), \tag{7}$$

with $\Sigma(\theta_e) = \Sigma^d + \Sigma^c + \Sigma^m$. Using this prediction error term, the expression for the data likelihood can be obtained, which is given by the multivariate Gaussian

$$f(D|\theta) = \mathcal{N}(\hat{y}|q(\theta_m|D), \Sigma(\theta_e)) = \frac{|\Sigma(\theta_e)|^{-1/2}}{(2\pi)^{N_{\hat{y}}/2}} \exp\left(-\frac{1}{2} J(\theta, \hat{y})\right), \tag{8}$$

where

$$J(\theta, \hat{y}) = (\hat{y} - q(\theta_m|D))^T \Sigma^{-1}(\theta_e)(\hat{y} - q(\theta_m|D)). \tag{9}$$

3.1.4 Bayesian Annealed Sequential Importance Sampling

In order to sample from the posterior distribution, Markov Chain Monte Carlo (MCMC) methods are often used. These MCMC methods are often variants of the Metropolis-Hastings (MH) algorithm. The details on derivation of the basic MH algorithm are described in [32]. There are many drawbacks to the MH algorithm. For instance, in order for the Markov chain to reach its stationary distribution, the MH algorithm often requires a long burn-in period where initial samples are discarded due to high correlations between these initial samples. Moreover, the algorithm is entirely sequential, which effectively limits any optimization efforts to the evaluation of the target distribution itself. A poorly chosen or tuned proposal distribution could also slow down the convergence to the stationary distribution or, in case of some multimodal or complicated distributions, cause the algorithm to get stuck on localized high-probability density regions. In order to address the problems with the convergence and the sequential nature of the MH algorithm, and to address the issue of choosing and tuning the proposal distribution, Bayesian Annealed Sequential Importance Sampling (BASIS) was used as the sampler for the inverse UQ.

BASIS is a method which uses a combination of Sequential Importance Sampling and MCMC to draw samples from multiple intermediate distributions. The intermediate distributions are given by

$$f_j(\theta) \propto [f(D|\theta)]^{p_j} \pi(\theta) \tag{10}$$

for $j = 1, \ldots, s$, where $0 = p_0 < p_1 < p_s = 1$. The exponent p_j is called the annealing parameter, since it is used to transition the intermediate distributions from the prior $f_0(\theta) = \pi(\theta)$ at $p_0 = 0$ to the posterior $f_s(\theta) \propto f(D|\theta)\pi(\theta) = p(\theta|D)$ at $p_s = 1$ in an annealed manner. The BASIS algorithm samples sequentially from each intermediate distribution through importance sampling, where every distribution is chosen adaptively using the annealing parameter p_j, where p_j is chosen such that the coefficient of variation (CV) of the importance weights is equal to or less than some tolerance value γ_{CV}, until the posterior distribution is reached at $p_s = 1$. The samples for every intermediate distribution are resampled from the previous distribution using the adaptively chosen importance weights. In order to reduce degeneracy resulting from repeated resampling, after every resampling step multiple parallel MCMC chains are constructed to generate unique samples from the resampled parameters. The details of the BASIS algorithm are described in [13].

3.2 Gaussian Process Regression Metamodeling

In order to improve the efficiency of the estimation of the posterior distribution for the parameters of the HemoCell model, a Gaussian Process (GP) Regression [33] model was trained as a surrogate model for the HemoCell model, which is much cheaper to evaluate and thus allows for more samples to be drawn from the posterior.

For the kernel of the GP model, we used the sum-kernel with the following covariance:

$$k(x_i, x_j) = \exp\left(-\frac{1}{2} d(x_i/l, x_j/l)^2\right) + c\mathbb{I}[x_i = x_j], \tag{11}$$

where x_i and x_j are two input vectors, l is a length-scale hyperparameter, c is a noise-level hyperparameter, and $d(\cdot, \cdot)$ is the Euclidean distance between the scaled input vectors. The first term is a squared exponential kernel and the second term is a noise-level estimation kernel. The length-scale hyperparameter l was chosen to be anisotropic, which means that the term x_i/l in Eq. (11) denotes an element-wise division of the elements of the input vector x_i by the corresponding elements of the length scale vector l. For the analysis the GP model implementation from scikit-learn [34] was used, which is based on Algorithm 2.1 in [33].

A full Bayesian analysis using a surrogate model is often computationally expensive and intractable, since in this case the posterior distribution would also be dependent on the hyperparameters of the surrogate model which subsequently need to be marginalized out to yield the calibrated model parameters. Therefore a modular Bayesian approach [30] with a pretrained surrogate model was used for the Inverse UQ analysis. In the modular approach the hyperparameters of the GP surrogate model are first separately estimated by training the model using, for example, maximum likelihood estimation. The estimated hyperparameters are then used to perform the Bayesian inference with the resulting GP metamodel.

The training data for the surrogate model were obtained by running the model 500 times with different parameter values for each shear rate used in the analysis. The parameters were sampled using a Maximin Latin-Hypercube design, implemented in the pyDOE (https://pythonhosted.org/pyDOE/) Python package for experimental design, separately for each shear rate within the limits of the priors used for the Inverse UQ analysis, which can be found in Table 1. This sampling method maximizes the distance between the points inside the parameter space and thus

TABLE 1: Prior distributions of the parameters used in the Inverse UQ analysis

Parameter	Symbol	Distribution	Lower Limit	Upper Limit
Link force coefficient	κ_l	Uniform	10 k_BT	300 k_BT
Bending force coefficient	κ_b	Uniform	50 k_BT	400 k_BT
Viscosity contrast	Λ	Uniform	1	15
Modeling uncertainty	σ^m	Uniform	0.001	0.1

increases the likelihood for any parameter set to be inside the convex hull of the training set. This prevents the GP model from making model predictions through extrapolation, given that GP models are better interpolators than they are extrapolators. Additionally, because the training samples were drawn only for the shear stresses used in the data set, a separate GP model was trained for each shear stress. This yielded an overall smaller number of parameters for the aggregate surrogate model and significantly reduced the cost of training and evaluation at a negligible cost in overall accuracy.

In order to estimate the predictivity of the metamodel the predictivity coefficient

$$Q_2 = 1 - \frac{\sum_{i=1}^{N_t} (q(\theta_i|D) - q_{\text{GP}}(\theta_i|D))^2}{\sum_{i=1}^{N_t} (q(\theta_i|D) - \bar{q})^2} \tag{12}$$

was calculated, where $q_{\text{GP}}(\theta|D)$ is a GP model QoI prediction and $\bar{q}$ is the sample mean of the HemoCell RBC model QoIs calculated from N_t test samples. The parameter samples used to estimate the sensitivity indices which do not result in failed simulations were reused as the test set, since these samples were generated using the low-discrepancy Sobol sequence, which provides test samples which are uniformly distributed over the explored parameter space. Using the trained GP model on this test set yielded a predictivity coefficient of $Q_2 \approx 0.998$. Since $Q_2 \geq 0.7$ is often considered a satisfactory approximation of the full model [35], the surrogate GP model was deemed to be a sufficiently accurate approximator to predict the HemoCell RBC model output.

3.3 Sensitivity Analysis

The results of sensitivity analysis were used to predict the identifiability of the parameter distributions with the given priors [36]. In order to estimate the sensitivity of the model output to the model parameters, first-order (S_i) and total (S_i^T) Sobol sensitivity indices were estimated [37,38]:

$$\begin{aligned} S_i &= \frac{\text{Var}_{\theta_i}(\text{E}_{\theta_{\sim i}}[q(\theta)|\theta_i])}{\text{Var}(q(\theta))}, \\ S_i^T &= \frac{\text{E}_{\theta_{\sim i}}(\text{Var}_{\theta_i}[q(\theta)|\theta_{\sim i}])}{\text{Var}(q(\theta))}, \end{aligned} \tag{13}$$

where $\text{Var}(q(\theta))$ is the variance of the model output, $\text{Var}_{\theta_i}(\text{E}_{\theta_{\sim i}}[q(\theta)|\theta_i])$ is the variance of the term in the analysis of variance (ANOVA) decomposition of the model function [39] that depends only on the parameter θ_i, and $\text{E}_{\theta_{\sim i}}(\text{Var}_{\theta_i}[q(\theta)|\theta_{\sim i}])$ is the variance of the sum of all the terms in the decomposition which contain θ_i. Here, θ_i denotes the ith element of θ and $\theta_{\sim i}$ the set of all elements of θ except the ith element.

The sensitivity indices were calculated by estimating the partial variances from Eq. (13) using the Monte Carlo estimators suggested in [38]:

$$\begin{aligned} \text{Var}_{\theta_i}(\text{E}_{\theta_{\sim i}}[q(\theta)|\theta_i]) &\approx \frac{1}{N} \sum_{j=1}^{N} q(\theta^{(j+N)}) \Big(q(\theta_{\sim i}^{(j)}, \theta_i^{(j+N)}) - q(\theta^{(j)}) \Big), \\ \text{E}_{\theta_{\sim i}}[\text{Var}_{\theta_i}(q(\theta)|\theta_{\sim i})] &\approx \frac{1}{2N} \sum_{j=1}^{N} \Big(q(\theta^{(j)}) - q(\theta_{\sim i}^{(j)}, \theta_i^{(j+N)}) \Big)^2, \end{aligned} \tag{14}$$

where $q(\theta^{(j)})$ and $q(\theta^{(j+N)})$ are two model runs with different values of the inputs sampled from their distributions, then, the model run $q(\theta_{\sim i}^{(j)}, \theta_i^{(j+N)})$ have the same values of $\theta_{\sim i}^{(j)}$ as for $q(\theta^{(j)})$ and the value of $\theta_i^{(j+N)}$ as in $q(\theta^{(j+N)})$, and $N(N_m + 2)$ is the total number of models run required to estimate the sensitivity indices for each of the input.

3.4 Uncertainty Propagation

The uncertainty on the parameters θ can be propagated to the model predictions of the QoI $q(\theta|D)$ using the posterior distribution of the parameters $f(\theta|D)$ [9]. If $q(\theta|D)$ is a probabilistic function with probability density function $f(q|\theta, D)$, the parameters sampled from the posterior can be used to compute the posterior robust distribution of the output $q(\theta|D)$

$$f(q|D) = \int_{\Theta} f(q|\theta, D) f(\theta|D) d\theta. \tag{15}$$

If the exact distribution $f(q|D)$ is unknown or complicated, $f(q|D)$ can also be estimated using Gaussian Kernel density estimation (KDE).

4. RESULTS

4.1 Setup

Since the BASIS algorithm is largely a tune-free algorithm, there are relatively few hyperparameters. The tolerance for the CV function and the scaling parameter for the covariance matrix of the proposal distribution were set to $\gamma_{\text{CV}} = 1$ and $\beta = 0.2$, respectively, which were suggested by [40]. Because of the long simulation time needed for the HemoCell RBC model, the maximum chain length and burn-in period were set to $l_{max} = 1$ and $n_{burn} = 0$, respectively, for the HemoCell RBC model for an optimal parallel work load balance. This should minimize the bias in the results due to the inhomogeneous Markov chain lengths, but could reduce the sampling efficiency if the posterior is very peaked. In order to ensure that the posterior for the GP model would sufficiently converge and to reduce the bias in the results, the maximum chain length and the number of burn-in MCMC steps for the BASIS algorithm were set to $l_{max} = 1$ and $n_{burn} = 10$, respectively, for the GP model.

The priors used for the Inverse UQ are shown in Table 1. Due to a lack of prior information about the optimal parameter range and high probability density regions in the parameter space, only uninformative priors were chosen in the form of uniform distributions. The limits were chosen such that the validated parameter values for the HemoCell RBC model were within the bounds of the prior by a relatively wide margin. The sensitivity indices were estimated for the HemoCell RBC parameters using the same limits as used by the priors.

Because of the high computational cost of the RBC shearing simulations only 1000 samples were drawn for the IUQ with the HemoCell RBC model. Since the GP model is much cheaper to evaluate than the HemoCell RBC model, 10,000 samples were drawn for the IUQ with the GP model.

4.2 Sensitivity Analysis

The sensitivity analysis was performed on the HemoCell RBC model to gauge for indications of possible unidentifiability for the RBC model parameters. The results of the sensitivity analysis are shown in Fig. 4. The SIs of the link force coefficient have values close to one. Hence, this parameter is an important model parameter and its identifiability in IUQ is expected to be high. On the contrary the bending force coefficient has low values of SIs and its poor identifiability is expected. The viscosity contrast variability slightly affects the model output with the SIs values up to 0.12 for high values of the shear stress and this parameter is expected to be identifiable.

4.3 Inverse Uncertainty Quantification

A comparison of the one-dimensional marginalized distributions of the parameters, approximated using 1D Gaussian Kernel Density Estimations (KDE), which were obtained with the HemoCell model and the GP surrogate model, is shown in Fig. 5. As was expected from sensitivity analysis, the link force coefficient and the viscosity contrast were identifiable and the results with the GP surrogate model approximate the HemoCell model distributions of these parameters well for the healthy cell. The GP model distributions of these parameters for the treated cell do not match the HemoCell model distributions very well, though. This could be explained by the fact that a significantly lower

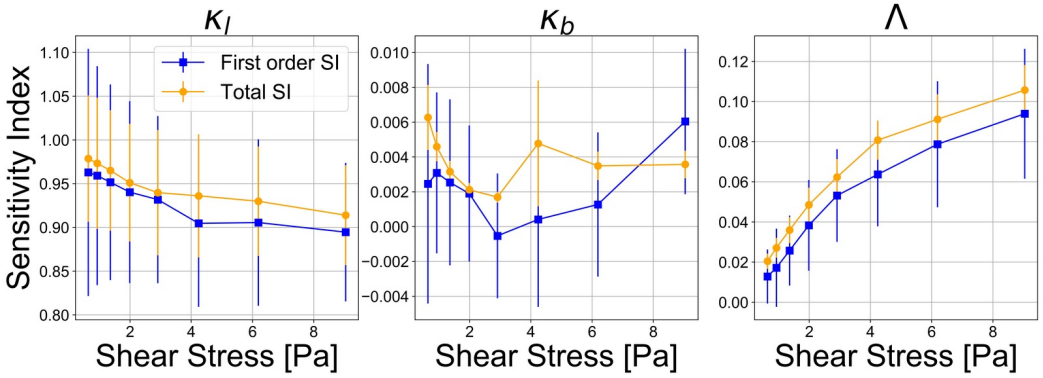

FIG. 4: SA results with the HemoCell RBC model, where κ_l, κ_b, and Λ are the link force coefficient, bending force coefficient, and viscosity ratio, respectively

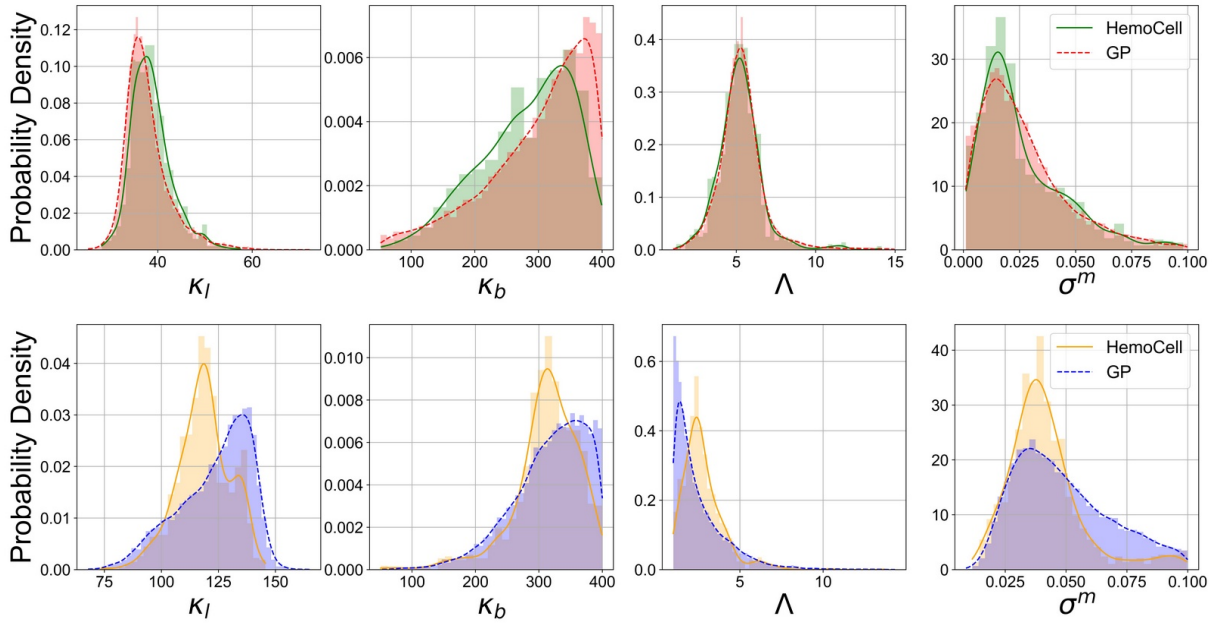

FIG. 5: 1D marginalized distributions obtained with the HemoCell and GP models for the healthy cell data (top row) and the stiffened cell data (bottom row)

number of samples was drawn with the HemoCell model and the convergence of these results was not reached. It could also be explained by the fact that the epistemic uncertainty arising from the small number of experimental data points used for the likelihood may add bias to the posterior of the GP model when combined with the errors in the GP model predictions. At the same time, the distribution of the bending force coefficient was not identified well and its approximation by the GP yielded a poor match.

The marginal distributions of σ^m for both the healthy and treated cells are also shown in Fig. 5. In both cases the modes of the posterior marginals are positioned closer to the lower bound of the prior of σ^m. This adds validity to the modeling assumption that the prediction errors are zero-mean Gaussian distributed, which suggests that the HemoCell model is correct on average.

The comparisons between the two-dimensional marginalized distributions obtained using the HemoCell and GP models are presented for the healthy RBC data in Fig. 6 and for the 0.5 mM TBHP treated RBC data in Fig. 7.

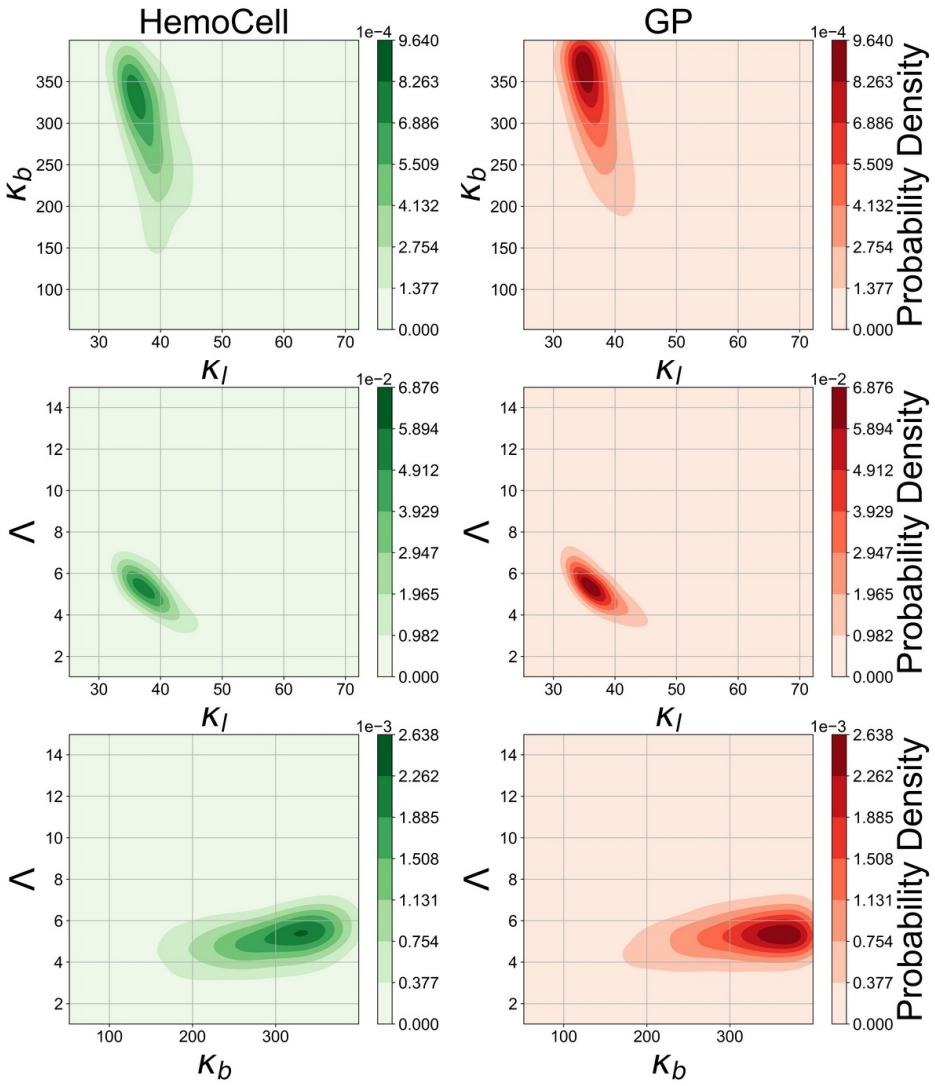

FIG. 6: 2D marginalized distributions obtained with the HemoCell (left column) and GP (right column) models for the healthy cell data

These results were approximated using 2D Gaussian KDEs. As with the one-dimensional results, the GP surrogate model approximates the distributions well in the direction of the link force coefficient and viscosity contrast, while yielding a slightly poorer match in the direction of the bending force coefficient. Additionally, a strong correlation is found between the link force coefficient and the viscosity contrast, which is also captured by the GP model.

The Maximum *a posteriori* (MAP) parameters and statistics from the estimated posteriors for both the HemoCell RBC model and the GP metamodel are shown in Table 2. The MAP parameters were found by computing the posterior density values obtained by multiplying the likelihood values with the prior values of the posterior samples and choosing the posterior sample which yields the highest posterior density value. Comparing the MAP parameters between both models showed the relative differences from about 13 to 58% between the optimal parameters. The relative differences in the mean value are significantly smaller for both healthy and treated cells and they are varying from approximately 1.6 to 7.1%. For the standard deviation, these values vary from 7.1 to 36.5%. In particular, as

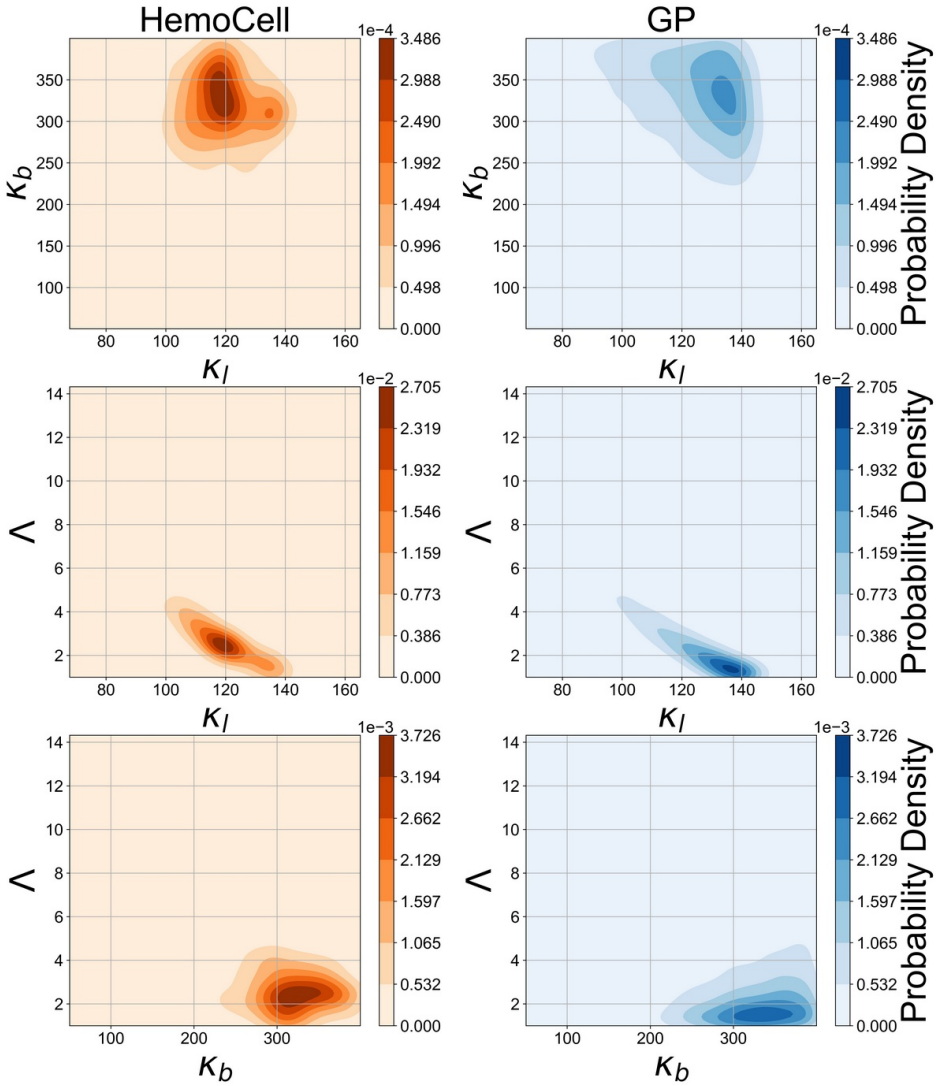

FIG. 7: 2D marginalized distributions obtained with the HemoCell (left column) and GP (right column) models for the stiffened cell data

it is reflected in the plots with 1D marginalised distributions, the healthy cell results for the link force coefficient and viscosity ratio obtained with the HemoCell model and with the GP metamodel match better than for the treated cell. This can mean that for the healthy cell, the convergence of the HemoCell results could be almost achieved, and which is probably not the case for the treated cell results. Moreover, as it is stated above, the difference between the HemoCell and GP model statistics is at least in part due to the low amount of experimental data points (epistemic uncertainty) combined with the error in the GP model predictions.

Additionally, the shift in the optimal values, for instance the shift in the MAP of κ_l from 38.8 in the healthy case to 117.2 in the diseased case, is physically expected. This is the stiffening of the RBC. Physiologically this is due to the oxidative stress induced by TBHP treatment, which causes spectrin and ankyrin degradation [24]. The spectrin network is modelled in the HemoCell model through the link force, and an increase in the link force coefficient κ_l results in an increase in the stiffness of the RBC. The result of the sensitivity analysis, given in this report, identified that the link force coefficient is this most important parameter for RBC stiffening.

TABLE 2: Optimal parameters and statistics obtained from the Inverse UQ. The parameters κ_l and κ_b are implicitly scaled with k_BT

	HemoCell			**Gaussian process regression**		
Parameter	**MAP HC**	$\mu_{\mathbf{HC}}$	$\sigma_{\mathbf{HC}}$	**MAP GP**	$\mu_{\mathbf{GP}}$	$\sigma_{\mathbf{GP}}$
Healthy RBC data						
κ_l	38.8	38.6	4.2	33.8	37.7	4.7
κ_b	293.2	282.2	71.9	395.8	300.5	79.2
Λ	4.9	5.2	1.4	5.64	5.3	1.5
σ^m	0.014	0.026	0.019	0.0021	0.027	0.019
Treated RBC data						
κ_l	117.2	118.5	11.5	142.2	123.4	15.7
κ_b	371.4	312.5	51.7	313.0	317.6	58.0
Λ	2.5	2.8	1.3	1.04	2.6	1.6
σ^m	0.027	0.041	0.016	0.026	0.049	0.020

4.3.1 Speed-Up

The evaluation times for the HemoCell model range from approximately 1 h and 10 min to 5 min wall clock time per core decreasing with the shear stress, while the GP surrogate model has an evaluation time of approximately 320 μs wall clock time per core regardless of the shear stress. Additionally, because separate models are trained for each shear stress, the GP surrogate model is trained within approximately 1 min wall clock time.

Whereas the surrogate model, once the training samples are obtained, can be trained and evaluated in a short amount of time on a single core, the large simulation times for the HemoCell RBC model require a highly parallel implementation of the BASIS algorithm. The BASIS algorithm was implemented in Python where multiple model evaluations were performed in parallel using concurrent processes executed by the Python subprocess module. The implementation was further parallelized for cluster computing by implementing communication between multiple multi-core processors on a supercomputer using a server-client model with network sockets. The analysis of the HemoCell RBC model using the parallel implementation of the BASIS algorithm was run on the Dutch National Supercomputer Cartesius with up to 720 concurrently running processes using multiple 24-core nodes. This implementation reduced the run time of the BASIS algorithm to up to approximately 7 h per 1000 parameter set evaluations.

4.4 Uncertainty Propagation

The results of the propagation of uncertainty with the parameters values obtained with the HemoCell and GP models are presented in Fig. 8 for both the healthy and the treated cells. The top plot includes the HemoCell model results run with the optimal values for different values of the shear stress where the solid lines are results with the HemoCell MAP values, and the dashed lines indicate the model output with the GP MAP values. These results show that both experiments produce a good approximation of the optimal values of the input parameters.

Uncertainty in the output of the HemoCell model estimated by the propagation of the obtained posterior distributions of the parameters is illustrated using the violin plots. This uncertainty must be taken into account in future simulations with the HemoCell model.

Additionally, for comparison, the measurement data are shown as a scatter plot (black dots). A more detailed comparison is shown in the plots below, where the absolute errors between the measurements and the model output with the MAP values of the parameters are presented. These plots for both healthy and treated cells show that for most values of the shear stress the GP results outperform or close to the results with the HemoCell estimators.

The bottom plot illustrates the converged healthy (shown in red) and treated (shown in blue) RBC shapes using the optimal parameters at each shear stress.

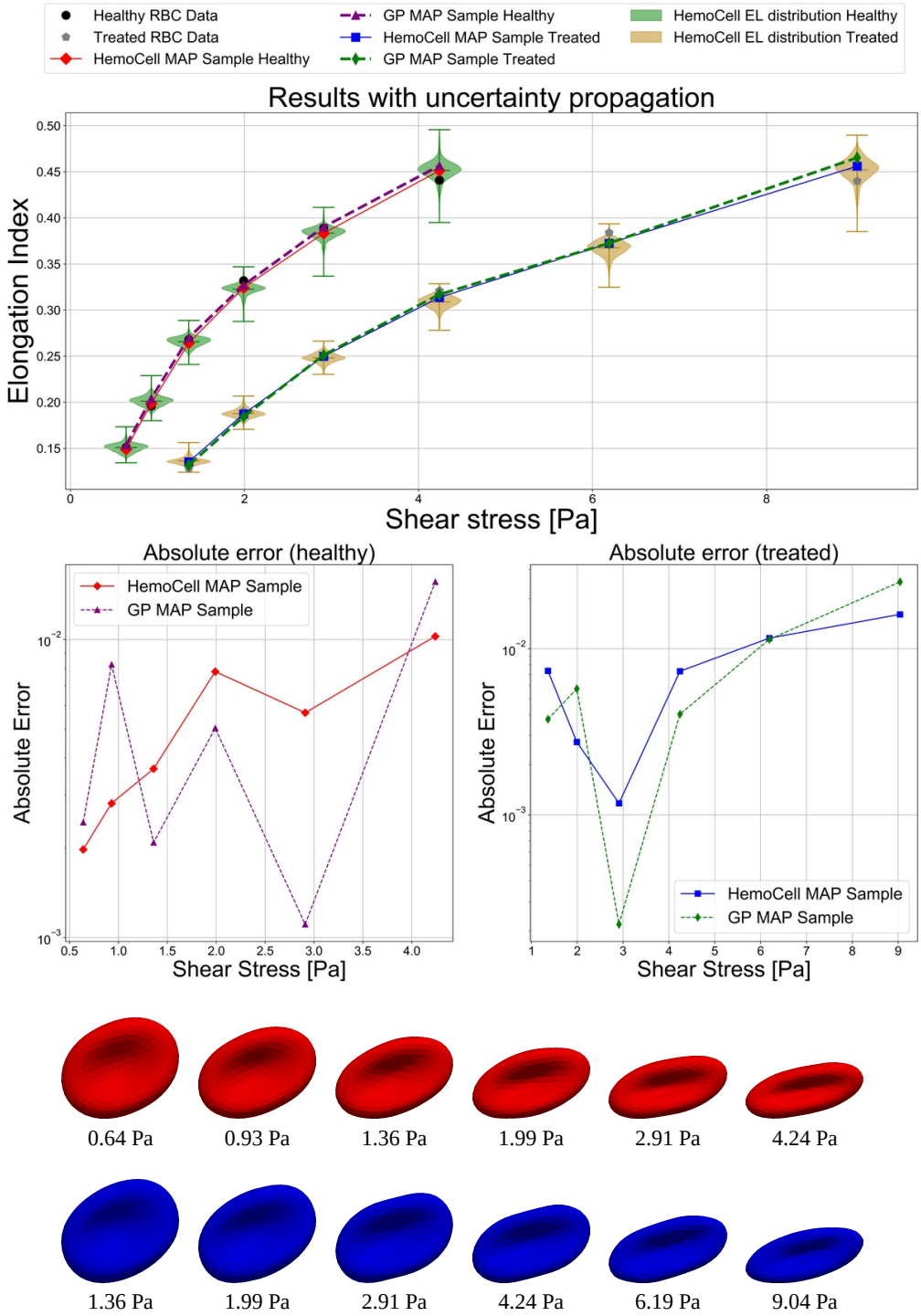

FIG. 8: Uncertainty Propagation results. Top plot: the results of uncertainty from the HemoCell model to the output are visualized using violin plots for the UQ analysis on the healthy and treated RBC data. The model outputs using the optimal MAP parameters for both models applied to both the healthy and treated RBC data are also plotted. Middle plots: the absolute error between the measurements data and the outputs of the HemoCell and GP models with the optimal parameters. Bottom plots: the converged healthy and treated RBC shapes with the optimal parameters at different values of the shear stress.

5. CONCLUSIONS

As predicted by the Sensitivity Analysis results, the link force coefficient and viscosity ratio were identified within reasonable uncertainty to influence the stiffening of the HemoCell RBC model. This is highlighted from the posterior marginals having a single clearly identifiable mode. The identifiability of these parameters were observed from the analysis of both healthy and treated cells using both the HemoCell model and the surrogate model. The poor identifiability of the bending force coefficient was also within expectations. The distribution of this parameter should be identified with another model output that depends heavily on the bending force coefficient.

Overall, a good match of the marginal distributions obtained with the HemoCell model and the metamodel is shown for the healthy cell. The results for the treated cell differ more visibly. This discrepancy might be because of a possible bias in the posterior marginals of the HemoCell model due to a relatively low number of samples. This can be suggested by the fact that, for the treated cell, the optimal values obtained with the GP model yield an output either closer to the experiment measurements or similar to the output yielded from the HemoCell optimal values. This is consistent across most values of shear stress.

This demonstrates that a Gaussian Process Regression model may provide a useful surrogate model for the HemoCell RBC model in the context of parameter calibration and uncertainty quantification. For the RBC shearing problem, replacing the HemoCell model with the GP surrogate model offered a trade-off of slightly lower accuracy in parameter estimation for a performance increase of five to seven orders of magnitude, which depends on the shear stress. This performance increase allowed a drastic increase in the number of samples to be drawn for the Inverse UQ analysis, while running on a single core. This yielded a more precise estimate for the exact posterior of the surrogate model.

In summary, we conclude that BASIS is an efficient method for Inverse UQ due to its parallel nature. Additionally, substituting the original model by a surrogate is a valid approach to reduce the computational burden of the method. In future work, the surrogate model could predict the size of the collision cross section of an RBC which will dependent on a local shear rate. The results presented in this paper could also be extended to more numerical RBC material validations. Namely the optical-tweezer experiment [41] as well as stiffer RBCs treated with higher concentrations of TBHP.

ACKNOWLEDGMENTS

We would like to thank Professor Omolola Eniola-Adefeso and her research group at the University of Michigan for providing the Ektacytometry elongation measurements for this work. Financial support was provided by the European Union Horizon 2020 research and innovation programme under Grant No. 675451 (CompBioMed project) and Grant No. 800925 (VECMA project). This work is also a part of the eMUSC (Enhancing Multiscale Computing with Sensitivity Analysis and Uncertainty Quantification) project. A.N. and A.H. gratefully acknowledge financial support from the Netherlands eScience Center. The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

1. Brault, A., Dumas, L., and Lucor, D., Uncertainty Quantification of Inflow Boundary Condition and Proximal Arterial Stiffness–Coupled Effect on Pulse Wave Propagation in a Vascular Network, *Int. J. Numer. Methods Biomed. Eng.*, **33**(10):e2859, 2017.
2. Nikishova, A., Veen, L., Zun, P., and Hoekstra, A., Semi-Intrusive Multiscale Metamodeling Uncertainty Quantification with Application to a Model of In-Stent Restenosis, *Philos. Trans. R. Soc. A*, **377**(2142):20180154, 2019.
3. Boccadifuoco, A., Mariotti, A., Celi, S., Martini, N., and Salvetti, M., Impact of Uncertainties in Outflow Boundary Conditions on the Predictions of Hemodynamic Simulations of Ascending Thoracic Aortic Aneurysms, *Comput. Fluids*, **165**:96–115, 2018.
4. Zavodszky, G., van Rooij, B., Azizi, V., Alowayyed, S., and Hoekstra, A., HemoCell: A High-Performance Microscopic Cellular Library, *Procedia Comput. Sci.*, **108**:159–165, 2017.

5. Alowayyed, S., Závodszky, G., Azizi, V., and Hoekstra, A.G., Load Balancing of Parallel Cell-Based Blood Flow Simulations, *J. Comput. Sci.*, **24**:1–7, 2018.
6. Tarksalooyeh, V.A., Závodszky, G., and Hoekstra, A.G., Optimizing Parallel Performance of the Cell based Blood Flow Simulation Software HemoCell, *Lecture Notes Comput. Sci.*, **11538**:537–547, 2019.
7. Gutierrez, M., Fish, M.B., Golinski, A.W., and Eniola-Adefeso, O., Presence of Rigid Red Blood Cells in Blood Flow Interferes with the Vascular Wall Adhesion of Leukocytes, *Langmuir*, **34**(6):2363–2372, 2018.
8. Wu, X., Kozlowski, T., Meidani, H., and Shirvan, K., Inverse Uncertainty Quantification Using the Modular Bayesian Approach based on Gaussian Process, Part 2: Application to TRACE, *Nucl. Eng. Design*, **335**:417–431, 2018.
9. Angelikopoulos, P., Papadimitriou, C., and Koumoutsakos, P., Bayesian Uncertainty Quantification and Propagation in Molecular Dynamics Simulations: A High Performance Computing Framework, *J. Chem. Phys.*, **137**(14):144103, 2012.
10. Diez, M., He, W., Campana, E.F., and Stern, F., Uncertainty Quantification of Delft Catamaran Resistance, Sinkage and Trim for Variable Froude Number and Geometry Using Metamodels, Quadrature and Karhunen–Loève Expansion, *J. Marine Sci. Technol.*, **19**(2):143–169, 2014.
11. Emmerich, M., Giotis, A., Özdemir, M., Bäck, T., and Giannakoglou, K., Metamodel-Assisted Evolution Strategies, *Lecture Notes Comput. Sci.*, **2439**:361–370, 2002.
12. Alowayyed, S.A., Vassaux, M., Czaja, B., Coveney, P.V., and Hoekstra, A.G., Towards Heterogeneous Multi-Scale Computing on Large Scale Parallel Supercomputers, *Supercomput. Front. Innov.*, **6**(4):20–43, 2020.
13. Wu, S., Angelikopoulos, P., Papadimitriou, C., and Koumoutsakos, P., Bayesian Annealed Sequential Importance Sampling (BASIS): An Unbiased Version of Transitional Markov Chain Monte Carlo, *ASCE-ASME J. Risk Uncert. Eng. Syst., Part B: Mech. Eng.*, **4**(1):011008, 2017.
14. Shin, S., Ku, Y.H., Ho, J.X., Kim, Y.K., Suh, J.S., and Singh, M., Progressive Impairment of Erythrocyte Deformability as Indicator of Microangiopathy in Type 2 Diabetes Mellitus, *Clin. Hemorheol. Microcirc.*, **36**(3):253–261, 2007.
15. Silva, D.G.H., Junior, E.B., De Almeida, E.A., and Bonini-Domingos, C.R., Oxidative Stress in Sickle Cell Disease: an Overview of Erythrocyte Redox Metabolism and Current Antioxidant Therapeutic Strategies, *Free Radical Biol. Med.*, **65**:1101–1109, 2013.
16. Czaja, B., Gutierrez, M., Závodszky, G., de Kanter, D., Hoekstra, A., and Eniola-Adefeso, O., The Influence of Red Blood Cell Deformability on Hematocrit Profiles and Platelet Margination, *PLOS Comput. Biol.*, **16**(3):1–18, 2020.
17. Li, Y., Stroberg, W., Lee, T.R., Kim, H.S., Man, H., Ho, D., Decuzzi, P., and Liu, W.K., Multiscale Modeling and Uncertainty Quantification in Nanoparticle-Mediated Drug/Gene Delivery, *Comput. Mech.*, **53**(3):511–537, 2014.
18. Lee, T.R., Greene, M.S., Jiang, Z., Kopacz, A.M., Decuzzi, P., Chen, W., and Liu, W.K., Quantifying Uncertainties in the Microvascular Transport of Nanoparticles, *Biomech. Model. Mechanobiol.*, **13**(3):515–526, 2014.
19. Závodszky, G., van Rooij, B., Azizi, V., and Hoekstra, A., Cellular Level In-Silico Modeling of Blood Rheology with an Improved Material Model for Red Blood Cells, *Front. Phys.*, **8**:563, 2017.
20. de Haan, M., Zavodszky, G., Azizi, V., and Hoekstra, A., Numerical Investigation of the Effects of Red Blood Cell Cytoplasmic Viscosity Contrasts on Single Cell and Bulk Transport Behavior, *Appl. Sci.*, **8**(9):1616, 2018.
21. Závodszky, G., van Rooij, B., Czaja, B., Azizi, V., de Kanter, D., and Hoekstra, A.G., Red Blood Cell and Platelet Diffusivity and Margination in the Presence of Cross-Stream Gradients in Blood Flows, *Phys. Fluids*, **31**(3):031903, 2019.
22. Timm, K., Kusumaatmaja, H., Kuzmin, A., Shardt, O., Silva, G., and Viggen, E., *The Lattice Boltzmann Method: Principles and Practice*, New York: Springer, pp. 1868–4521, 2016.
23. Hardeman, M.R., Besselink, G.A., Ebbing, I., De Korte, D., Ince, C., and Verhoeven, A.J., Laser-Assisted Optical Rotational Cell Analyzer Measurements Reveal Early Changes in Human RBC Deformability Induced by Photodynamic Treatment, *Transfusion*, **43**(11):1533–1537, 2003.
24. Caprari, P., Bozzi, A., Malorni, W., Bottini, A., Iosi, F., Santini, M.T., and Salvati, A.M., Junctional Sites of Erythrocyte Skeletal Proteins are Specific Targets of Tert-Butylhydroperoxide Oxidative Damage, *Chemico-Biol. Interact.*, **94**(3):243–258, 1995.
25. Katiyar, V.K. and Fisseha, D., Analysis of Mechanical Behavior of Red Blood Cell Membrane with Malaria Infection, *World J. Mech.*, **1**(03):100–108, 2011.
26. Athanassiou, G., Moutzouri, A., Gogos, C., and Skoutelis, A., Red Blood Cell Deformability in Patients with Human Immunodeficiency Virus Infection, *Eur. J. Clin. Microbiol. Infect. Dis.*, **29**(7):845–849, 2010.

27. Yazdani, A.Z., Kalluri, R.M., and Bagchi, P., Tank-Treading and Tumbling Frequencies of Capsules and Red Blood Cells, *Phys. Rev. E*, **83**(4):046305, 2011.
28. Wang, J. and Zabaras, N., A Bayesian Inference Approach to the Inverse Heat Conduction Problem, *Int. J. Heat Mass Transf.*, **47**(17-18):3927–3941, 2004.
29. Ling, Y., Mullins, J., and Mahadevan, S., Selection of Model Discrepancy Priors in Bayesian Calibration, *J. Comput. Phys.*, **276**:665–680, 2014.
30. Wu, X., Kozlowski, T., Meidani, H., and Shirvan, K., Inverse Uncertainty Quantification Using the Modular Bayesian Approach based on Gaussian Process, Part 1: Theory, *Nucl. Eng. Design*, **335**:339–355, 2018.
31. Jaynes, E.T., Information Theory and Statistical Mechanics, *Phys. Rev.*, **106**(4):620–630, 1957.
32. Chib, S. and Greenberg, E., Understanding the Metropolis-Hastings Algorithm, *Am. Stat.*, **49**(4):327–335, 1995.
33. Rasmussen, C.E., Gaussian Processes in Machine Learning, *Adv. Lectures Mach. Learning*, **3176**:63–71, 2004.
34. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., Scikit-Learn: Machine Learning in Python, *J. Mach. Learning Res.*, **12**:2825–2830, 2011.
35. Iooss, B., Boussouf, L., Feuillard, V., and Marrel, A., Numerical Studies of the Metamodel Fitting and Validation Processes, *Int. J. Adv. Syst. Measur.*, **3**(1-2):11–21, 2010.
36. Wu, X., Shirvan, K., and Kozlowski, T., Demonstration of the Relationship between Sensitivity and Identifiability for Inverse Uncertainty Quantification, *J. Comput. Phys.*, **396**:12–30, 2019.
37. Jansen, M.J., Analysis of Variance Designs for Model Output, *Comput. Phys. Commun.*, **117**(1-2):35–43, 1999.
38. Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S., Variance based Sensitivity Analysis of Model Output. Design and Estimator for the Total Sensitivity Index, *Comput. Phys. Commun.*, **181**(2):259–270, 2010.
39. Sobol', I.M., On Sensitivity Estimation for Nonlinear Mathematical Models, *Matematicheskoe Modelirovanie*, **2**(1):112–118, 1990.
40. Ching, J. and Chen, Y.C., Transitional Markov Chain Monte Carlo Method for Bayesian Model Updating, Model Class Selection, and Model Averaging, *J. Eng. Mech.*, **133**(7):816–832, 2007.
41. Mills, J., Qie, L., Dao, M., Lim, C., and Suresh, S., Nonlinear Elastic and Viscoelastic Deformation of the Human Red Blood Cell with Optical Tweezers, *Mol. Cell. Biomech.*, **1**(3):169–180, 2004.

APPENDIX. HEMOCELL MATERIAL MODEL

The HemoCell material model is a constitutive force model comprising four discrete forces that describe the mechanical properties of the RBC membrane. The two physical forces, the link and bending force, are elaborated here as they are used in this research. The area and volume force are left out of this description as they do not represent physical forces and are out of the scope of this research. More detailed description of the implantation and validation of the original model can be found in [19].

The link force captures the stretching and compression of the underlying spectrin network of the RBC and is written as follows:

$$F_{link} = -\frac{\kappa_l dL}{p}\left[1 + \frac{1}{\tau_l^2 - dL^2}\right]. \tag{A1}$$

Here $dL = (L_i - L_0)/L_0$ is the normal strain defined as the relative deviation from the equilibrium length L_0. τ_l is the relative expansion ratio, chosen to be 3, when the spectrin network reaches its persistence-length. The persistence length of a spectrin filament was chosen to be $p = 7.5$ nm. The link force coefficient is given by κ_l.

The bending force represents the bending response of the membrane arising primarily from the non-zero thickness of the spectrin network and is calculated between two neighboring surface elements:

$$F_{bend} = -\frac{K_b d\theta}{L_0}\left[1 + \frac{1}{\tau_b^2 + d\theta^2}\right]. \tag{A2}$$

Here $d\theta = \theta_i - \theta_0$ and since we chose a lattice fluid resolution of $L_0 = 0.5$ μm the limiting angle is chosen to be $\tau_b = \pi/6$. The bending force coefficient is given by κ_l.