Bayesian calibration of a model of polymer die swell using data from a laser–based measurement system

Julian Hennicker[a], Alex Gansen[b], Clemens Sill[c], Nicolas Verdon[c], Jörg Baller[b], Jack S. Hale[a]

[a]*Institute of Computational Engineering, Department of Engineering, Faculty of Science, Technology and Medicine, University of Luxembourg, Maison du Nombre, 6 avenue de la Fonte, L-4364, Esch-sur-Alzette, Luxembourg,*
[b]*Department of Physics, Faculty of Science, Technology and Medicine, University of Luxembourg, 162a avenue de la Faïencerie, L-1511 Limpertsberg, Luxembourg,*
[c]*Goodyear Innovation Center Luxembourg Goodyear S.A., Avenue Gordon Smith, L-7750 Colmar-Berg, Luxembourg,*

Abstract

Our goal is to produce a predictive model linking apparent shear rate with polymer die swell for a single extrusion system. We propose the use of a Kennedy O'Hagan-type approach for calibrating this model, where the die swell is modelled as a sum of a deterministic simulator, a Gaussian process and an additive white noise term. For the deterministic simulator we use a model that links the shear rate and two parameters related to the shear–thinning and relaxation time of a polymer to the die swell. The role of the Gaussian process is to capture the inherent structural uncertainty induced by the missing physical processes such as wall slip and non-isothermal conditions in the derivation of the simulator. The parameter calibration of the full model is performed using a subjective Bayesian methodology where the solution is characterised by the posterior distribution of the parameters given the observed data. We condition our model on experimental data produced from a capillary rheometer fitted with a laser-based die swell measurement system. We implement the models using a high–level probabilistic programming language and explore the resulting posterior using the No–U–Turn Sampler (NUTS). Our results show that the experimental swell data leads to a contraction in the posterior distribution with respect to the prior on the parameter related to the relaxation time of the polymer. In addition we demonstrate that the Kennedy O'Hagan-type model structure leads to improved fit of the model within the range of experimental data without sacrificing the simulator's extrapolative power outside.

1. Introduction

Extrusion is the process of forcing materials through a die to create products of desired cross sectional shape, e.g. pipes, tyre treads, synthetic fibres, polymer sheets or even pasta [45]. Because of the phenomenon of die swell [50] it is almost always the case that the dimensions of the final extrudate does not correspond with the shape

of the die. In the context of the design of extrusion processes, a model directly linking process parameters (e.g. die geometry, chemical and physical properties of materials, temperature, pressure) with final extrudate shape would be an important tool for engineers involved in optimisation, uncertainty quantification and sensitivity analysis of production processes based on extrusion.

The predictive simulation of complex fluid problems, of which polymer die swell is an important example [50], remains an ongoing challenge both in terms of model specification (e.g. constitutive model development [42, 61] and calibration [43]) and numerical solution techniques (e.g. the development of robust and stable discretisation methods [26, 15, 9]). In particular for viscoelastic systems, the unsolved difficulties in the numerical resolution of viscoelastic flows at high Weissenberg (Wi) numbers (the ratio of intrinsic to experimental time scales) still place limits on their practical range of applicability [30]. In addition to these numerical issues, the inherent complexities of the physical processes present in a typical extrusion system, such as non-isothermal conditions [40, 16], slip boundary conditions [36, 27], multiple time-scales [61] and multi-phase flow [33] make even appropriate model specification challenging. Given these considerable complexities, it is nearly inevitable that even a state-of-the-art calibrated numerical model of die swell will show some level of model discrepancy (structural uncertainty, model inadequacy, model bias), that is, the inability of the model to reproduce reality [38, 31].

The central idea of this paper is to tackle the problem of producing a predictive model of die swell using a data–informed Bayesian calibration approach. Our direction is orthogonal to tackling the important underlying numerical and modelling challenges in viscoelastic flow outlined in the previous paragraph. We simply assume that both a set of experimental data and a reasonable predictive model of die swell exists, and that we would like to incorporate the data into the model such that it provides a reasonable prediction within the range of acquired data, and reasonable extrapolative power outside of that range. It is well-known that purely data-driven approaches often struggle with extrapolation, even with large datasets [1, 6], and that physics-based approaches often do not adequately capture the behaviour of the real system due to incorrect and/or incomplete modelling assumptions.

A concise summary of the key ingredients of our study is as follows:

1. A set of experimental data acquired from a capillary rheometer fitted with a laser-based system for measuring die swell. Our model polymer is an uncured Styrene-Butadiene Rubber (SBR) compound [45]. SBR is the main constituent of tires and is therefore of interest as a model system for our industrial partner, the Goodyear Tire and Rubber Company. The experimental setup is essentially a lab-scale extrusion system and it shares many of the aforementioned complexities of the industrial scale extruders present in Goodyear's production facilities.

2. A simple and cheap semi-empirical model, attributable to Tanner [52], linking shear rate and two (*a priori* unknown) constitutive parameters to die swell in an axisymmetric geometry. Its role in our study is to act as a sufficiently good model capturing the main experimental trend in the die swell data.

3. A Kennedy O'Hagan (KOH) type Bayesian calibration scheme [29], that ac-

counts uncertainty when calibrating Tanner's model (2.) using the experimental data from the lab-scale extrusion system (1.). A key feature of the KOH-type approaches is an additive model structure that combines a physically-motivated simulation with a data-centric Gaussian Process to close the discrepancy between simulation and reality. The full Bayesian posterior is specified using a differentiable probabilistic programming language, TensorFlow Probability [13, 14] (TFP), and explored using the highly efficient No-U-Turn-Sampler (NUTS) [25].

We remark that the semi-empirical Tanner's model could also be replaced with a set of more expensive numerical simulator runs that could more accurately predict die swell with fewer underlying modelling assumptions, e.g. Rheotool [39] or Ansys Polyflow [51]. With this 'expensive simulator' case in mind, we also implement a methodology attributable to Higdon et al. (HIG) [23] for simultaneously calibrating the model and constructing a surrogate model from a limited budget of runs of the Tanner model and compare its performance with the standard KOH approach.

Before outlining the content of the paper, we briefly make a few points about our methodology in the broader context of Bayesian calibration of engineering and industrial systems. Bayesian calibration is a principled way in which measured data and prior information can be combined with model predictions in order to improve model output. The KOH methodology is considered the 'gold standard' [59] for Bayesian calibration methodologies in terms of its simultaneous ability to close model discrepancy, calibrate the parameters of the simulator and provide interpretable output with quantified uncertainty. The KOH methodology has been used to calibrate many types of models e.g. building systems [59, 8, 7], electric motors [28], thermal problems [24], nuclear weapons [22] and spot welding [2]. It is worth noting that model calibration remains an active topic in theoretical research. For example, significant efforts have been made to understand and alleviate identifiability issues with the KOH approach. For example [41] presented an approach to improve identifiability by defining a Gaussian Process that is orthogonal to the gradient of the simulation model, although at considerable extra computational cost. The recent papers [57, 56] develop a frequentist L^2–calibration technique that is consistent and controls issues related to parameter identifiability.

An outline of this paper is as follows. In section 2, we present the experimental setup and procedure to produce the data used in the predictive modelling. We start with introducing the (simple) physics based simulator and its calibration within the Bayesian inference framework in section 3.1. We then present its counterpart, the purely data-driven model based on Gaussian Processes and the calibration of its hyper-parameters in section 3.2. Then, in section 3.3, we present the KOH model, which couples the two previous approaches. Finally, the Higdon approach, as presented in section 3.4, extends the KOH model by introducing a surrogate model for the simulator. The HIG approach enables a numerically cost-efficient implementation of the KOH model even for expensive numerical simulators.

2. Experimental setup

2.1. Samples

A non–oil extended styrene-butadiene rubber (SBR) containing 27% of styrene has been used as a model system for the present study. The SBR has a molecular weight of $310\,000\mathrm{g\,mol}^{-1}$ (measured with gel permeation chromatography). The rubber is unfilled.

2.2. Measurement of die swell

In order to calibrate the proposed models linking applied shear rate with die swell we require a data set from an extrusion process. Figure 1 shows a schematic drawing of the experimental setup used.

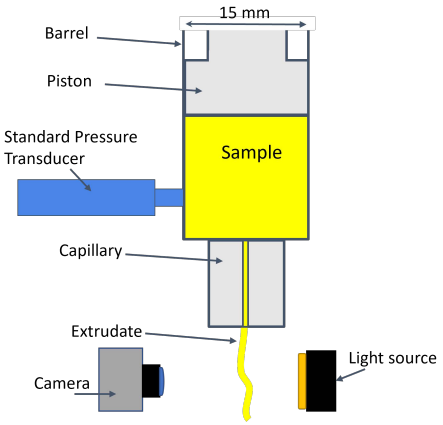

Figure 1: Schematic drawing of experimental setup

We use a commercial capillary rheometer (Rheograph 25 from Göttfert) in combination with a laser based die swell tester (also from Göttfert). Capillary rheometers are based on the same physical principle (see e.g. [32]) as extrusion machines used in production processes: pressure driven flow through dies. In the experimental setup used for this study, the sample is pressed by a piston through a die with the shape of a cylindrical capillary. First, the sample under investigation is placed into the barrel of the capillary rheometer with radius $r_b = 15\mathrm{mm}$ and heated to $T = 120°\mathrm{C}$. The piston, moving at a controlled velocity v^{pist}, pushes the sample through the capillary of radius $r_c = 1\mathrm{mm}$. As the radius of the capillary is significantly smaller than the radius of the barrel the velocity of the extrudate v^{ext} will be higher than the velocity of the piston v^{pist}. The volumetric flow rate inside the barrel is given by

$$Q^{\mathrm{barrel}} = \pi r_b^2 v^{\mathrm{pist}}. \tag{1}$$

Similarly, the volumetric flow rate inside the capillary is given by

$$Q^{\mathrm{cap}} = \pi r_c^2 v^{\mathrm{ext}}. \tag{2}$$

Assuming incompressibility, we can define $Q := Q^{\mathrm{cap}} = Q^{\mathrm{barrel}}$ allowing us to define the apparent shear rate in the capillary as

$$\dot{\gamma}_w = \frac{4Q}{\pi r_c^3}. \tag{3}$$

By controlling the piston's velocity, the apparent shear rate $\dot{\gamma}_w$ is controlled during the experiments. Classical flow curves, i.e. measurement of pressure as a function of increasing shear rates have been recorded. Data is only recorded when steady flow conditions have been reached. For more details about the experimental procedure please refer to [17]. For this study, the pressure data is not used but only data of the extrudate's swell together with the related shear rates. To capture the swell of the extrudate at the die's exit, a commercial system from Göttfert ("Die Swell Tester") is used [3]. This device is capable of directly measuring the die swell at a single given distance ($60\mathrm{mm}$) from the die exit during the extrusion process. The laser emits an almost two-dimensional fan-shaped laser beam that illuminates the extrudate. At the opposite side, there is a line detector which records the extrudate's thickness in a plane perpendicular to the die's exit. Extrudate thicknesses can be measured with this system in a range from $0.15\,\mathrm{mm}$ to $28\,\mathrm{mm}$ with a standard deviation of $\pm 3\mu\mathrm{m}$.

The acquired data used for the modelling in this study is presented in Figure 2.

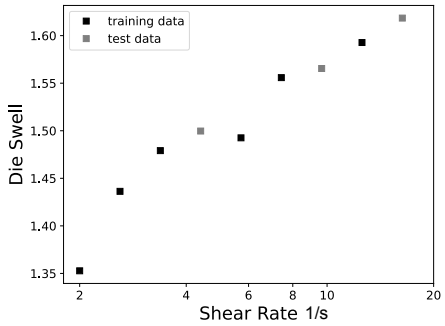

Figure 2: Die swell vs. apparent shear rate data obtained from the capillary extrusion experiment, divided in training and test data, as used for the calibration and validation of the models presented below

3. Mathematical models

For the presentation of the general mathematical models, which are not restricted to die swell, we will use an independent and condensed notation for the experimental data of section 2, consisting of N_{obs} scalar inputs from some input set $\mathcal{U} \subset \mathbb{R}$, and outputs from some output set $\mathcal{V} \subset \mathbb{R}$,

$$\{(x_j^{\mathrm{obs}}, y_j^{\mathrm{obs}}) \mid x_j^{\mathrm{obs}} \in \mathcal{U},\ y_j^{\mathrm{obs}} \in \mathcal{V},\ j = 1, \ldots, N_{\mathrm{obs}}\},$$

and define

$$\mathbf{x}^{\mathrm{obs}} = (x_j^{\mathrm{obs}})_{j=1,\dots,N_{\mathrm{obs}}}, \qquad \mathbf{y}^{\mathrm{obs}} = (y_j^{\mathrm{obs}})_{j=1,\dots,N_{\mathrm{obs}}}.$$

We further assume that the only source of randomness in the data is the measurement noise on the experimental output, and that this noise is Gaussian white noise with a known variance,

$$\epsilon_j^{\mathrm{obs}} \sim_{\mathrm{iid}} \mathcal{N}(0, \sigma_{\mathrm{obs}}^2). \tag{4}$$

We will present in this section four different modelling approaches applied on the experimentally obtained die swell data. Our main focus lies in the comparison of the models' performance in predicting die swell inside as well as outside the training range. We will also regard the calibration of model parameters, where we differentiate between the physics model (material) parameters, and the hyper–parameters of the probabilistic models, which are in our case the parameters defining the Gaussian Processes. We thus conduct a fully Bayesian inference for each of the model problems.

3.1. Deterministic model: Tanner's formula

A simple analytical model for the prediction of die swell for single–mode Oldroyd–B viscoelastic fluids is given by Tanner's formula [52, 53, 58, 46], which relates the recoverable stress at the die wall, σ_{w}, which is the first normal stress difference divided by the wall shear stress, to extrudate swell $S = \frac{r_{\mathrm{extrudate}}}{r_{\mathrm{die}}}$. For axisymmetric dies the model can be written

$$S = \left(1 + \frac{\sigma_{\mathrm{w}}^2}{2}\right)^{\frac{1}{6}} + 0.13, \tag{5}$$

where the last, constant summand is such as to fit measurements for Newtonian fluids, for which the the first normal stress difference is zero. Assuming a fully developed Poiseuille flow in the capillary, σ_{w} can be expressed as a function of relaxation time λ, solvent viscosity ratio $\beta = \frac{\eta_s}{\eta}$, and wall shear rate $x = \dot{\gamma}_{\mathrm{w}}$,

$$\sigma_{\mathrm{w}} = (1 - \beta)\lambda x, \tag{6}$$

by exploiting the analytical solution (cf. e.g. [34]).

For what follows, we introduce the notations for the material parameters, $\boldsymbol{\theta} := (\beta, \lambda)$, and for the simple predictive die swell simulator, obtained by substituting (6) into (5),

$$\eta^{\mathrm{sim}}(x, \boldsymbol{\theta}) := \left(1 + \frac{(1-\beta)^2\lambda^2 x^2}{2}\right)^{\frac{1}{6}} + 0.13. \tag{7}$$

The key advantage of Tanner's formula is that it is an algebraic expression, so no numerical flow simulation has to be carried out. In our experience numerical simulations of die swell typically take on the order of hours to run. On the other hand, Tanner's formula relies on gross oversimplifications, such as neglecting die exit effects and the multiple relaxation times of most complex viscoelastic fluids, which means that it is only strictly valid in suitable regimes [50] (simple fluids e.g. polyvinyl chloride (PVC) and polyethylene (PE) and at low Weissenberg numbers, respectively).

In order to justify this simple model choice as the one proposed by Tanner [52], we make reference to a study [10] that presents in Fig. 7(a) a comparison of numerical planar swell predictions between a state-of-the-art Eulerian free–surface flow solver [10], the Rheotool solver [39], Tanner's model [52, 53], and three other numerical studies [55, 11, 19]. Firstly, starting at even at relatively low Weissenberg numbers ($\mathrm{Wi} > 1.5$) the numerical solvers are prone to small ripples and self-sustained surface oscillations at the free surface of the extrudate, which is attributed by the authors of [10] to numerical, rather than physical instabilities. Secondly, the discrepancy between predicted swell ratios at the largest Weissenberg number $\mathrm{Wi} = 2.5$ is not small (Rheotool [39]: $S \sim 1.63$, Habla et al. [19]: $S \sim 1.52$), suggesting that there is significant differences between numerical simulations of the same phenomenon. Finally, Tanner's model, despite being based on many simplifying assumptions, does seem to provide a reasonable prediction of die swell ($S \sim 1.48$ at $\mathrm{Wi} = 2.5$), suggesting that it could be a parsimonious and computationally inexpensive choice for our Bayesian calibration study. We also remark that after calibration our final Weissenberg numbers are on the order of 10 which constitutes a challenge even for state-of-the-art numerical solvers [10].

3.1.1. Inference on model parameters (calibration)

The parameters $\boldsymbol{\theta} \in \mathbb{R}^{N_{\mathrm{par}}}$, $N_{\mathrm{par}} = 2$, occurring in Tanner's formula are not known, in general, and have to be calibrated by means of the training data.

Let us assume that Tanner's formula can principally reproduce reality, i.e.

$$y_j^{\mathrm{obs}} = \eta^{\mathrm{sim}}(x_j^{\mathrm{obs}}, \boldsymbol{\theta}) + \epsilon_j^{\mathrm{obs}}, \tag{8}$$

where $\eta^{\mathrm{sim}} : \mathcal{U} \times \mathbb{R}^{N_{\mathrm{par}}} \to \mathcal{V}$ is the simulator as defined in (7).

Let us now use the experimental data for inference on $\boldsymbol{\theta}$ in the Bayesian statistics framework, where we want to quantitatively analyse the posterior PDF for this model problem,

$$p(\boldsymbol{\theta} \mid \mathbf{y}^{\mathrm{obs}}) \propto p(\mathbf{y}^{\mathrm{obs}} \mid \boldsymbol{\theta}) p_0(\boldsymbol{\theta}). \tag{9}$$

From equations (8) and (4), we have

$$y_j^{\mathrm{obs}} - \eta^{\mathrm{sim}}(x_j^{\mathrm{obs}}, \boldsymbol{\theta}) \sim_{\mathrm{iid}} \mathcal{N}(0, \sigma_{\mathrm{obs}}^2), \qquad \forall j = 1, \ldots, N_{\mathrm{obs}},$$

and hence,

$$\mathbf{y}^{\mathrm{obs}} \,\big|\, \boldsymbol{\theta} \sim \mathcal{N}(\eta^{\mathrm{sim}}(\mathbf{x}^{\mathrm{obs}}, \boldsymbol{\theta}), \sigma_{\mathrm{obs}}^2 \mathbf{I}),$$

which defines the likelihood function $p(\mathbf{y}^{\mathrm{obs}} \mid \boldsymbol{\theta})$.

The prior PDF $p_0(\boldsymbol{\theta})$ used here is reported in table 5.

Next to a successful regression, we will want to exploit the model for making predictions. Therefore, let us define $\mathbf{x}^{\mathrm{pred}} = (x_j^{\mathrm{pred}})_{j=1,\ldots,N_{\mathrm{pred}}}$, the set of inputs, and $\mathbf{y}^{\mathrm{pred}} = (y_j^{\mathrm{pred}})_{j=1,\ldots,N_{\mathrm{pred}}}$, the set of outputs for the predictions. We assume that the model (8) also holds for the predictions, s.t. we have

$$\mathbf{y}^{\mathrm{pred}} \,\big|\, \boldsymbol{\theta} \sim \mathcal{N}(\eta^{\mathrm{sim}}(\mathbf{x}^{\mathrm{pred}}, \boldsymbol{\theta}), \sigma_{\mathrm{obs}}^2 \mathbf{I}).$$

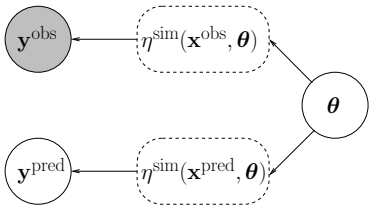

Figure 3: Bayesian graphical model for the calibration of the simulator. The nodes with a full line contour indicate random variables. The filled node represents an observed quantity, and the unfilled nodes represent unobserved quantities. The nodes with a dashed line contour indicate deterministic transformations of the upstream nodes

Concretely, predictions are obtained by sampling from the posterior predictive (PP) distribution,

$$p_{\mathrm{PP}}(\mathbf{y}^{\mathrm{pred}} \mid \mathbf{y}^{\mathrm{obs}}) = \int p(\mathbf{y}^{\mathrm{pred}} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{y}^{\mathrm{obs}}) \mathrm{d}\boldsymbol{\theta}.$$

For the PP expectation, we have

$$\mathbb{E}[\mathbf{y}^{\mathrm{pred}} \mid \mathbf{y}^{\mathrm{obs}}] = \int \mathbb{E}[\mathbf{y}^{\mathrm{pred}} \mid \boldsymbol{\theta}] p(\boldsymbol{\theta} \mid \mathbf{y}^{\mathrm{obs}}) \mathrm{d}\boldsymbol{\theta} = \mathbb{E}[\eta^{\mathrm{sim}}(\mathbf{x}^{\mathrm{pred}}, \boldsymbol{\theta}) \mid \mathbf{y}^{\mathrm{obs}}].$$

Note that the PP distribution can be used for posterior predictive checking (PPC) by setting $\mathbf{x}^{\mathrm{pred}} = \mathbf{x}^{\mathrm{obs}}$. For the numerical tests, we define $\mathbf{x}^{\mathrm{pred}}$ as a homogeneous discretisation of the input space $\mathcal{U}$, and include $\mathbf{y}^{\mathrm{pred}}$ in the MCMC sampling from the (enriched) posterior PDF, which is just the integrand of the PP distribution. This method of making predictions differs from classical statistics, where predictions would be drawn from the calibrated model, as

$$\mathbf{y}^{\mathrm{pred}\star} \mid \mathbf{y}^{\mathrm{obs}} \sim \mathcal{N}(\eta^{\mathrm{sim}}(\mathbf{x}^{\mathrm{pred}}, \boldsymbol{\theta}^{\star}), \sigma^2_{\mathrm{obs}} \mathbf{I}),$$

where $\boldsymbol{\theta}^{\star}$ is the calibrated parameter set, which would usually be the Maximum Likelihood Estimate (MLE), and could be generalised to be, e.g. the posterior mean or the Maximum A Posteriori estimate (MAP), in a BI setting.

3.1.2. Numerical results

We implement the Bayesian models in the TensorFlow Probability (TFP) framework [14], a Python library that allows the straightforward expression of probabilistic models as code. Because TFP builds on top of TensorFlow (TF) [13] the models can automatically be compiled using the XLA (Accelerated Linear Algebra) compiler into optimal machine code targeting a variety of hardware backends. Perhaps most importantly, TF provides forward (tangent-linear) and reverse (adjoint) mode automatic differentiation to efficiently calculate first and higher-order derivatives of the Bayesian posterior with respect to the underlying parameters. These derivatives are necessary in order to use modern high-acceptance rate MCMC algorithms such as Hamiltonian Monte Carlo (HMC) [12, 4] or the No-U-Turn-Sampler (NUTS) [25] that we use to

explore the Bayesian posterior. TFP is not unique; similar software methodologies are can be found in packages such as Stan [54] and PyMC3 [48]. For our purposes we found TFP to be attractive as it is provides high-level building blocks for probabilistic programming while still allowing non-standard operations to be performed using low-level TF functions.

The probabilistic inference problems were executed on a server fitted with two AMD EPYC 7742 CPUs with 64 cores per socket, 2TB of system memory and three NVIDIA Tesla V100 Graphics Processing Units (GPUs) with 32GB of memory each. We used the official TensorFlow 2.6.1 GPU-enabled Docker image with support for the CUDA backend. We used the Podman container engine to run the images [20]. Overall runtime was dominated by the expensive Cholesky decomposition of the covariance matrix in the Gaussian Process model. To ensure numerical stability TensorFlow was set to use double precision for all arithmetic operations.

For each inference problem three MCMC chains were executed in parallel, one per GPU. The chains were started at random points in parameter space and were then advanced using the NUTS sampler [25] in conjunction with a transformed transition kernel [18, 37] which enables the chain to operate in unconstrained space ($\mathbb{R}^n$) rather than the natural but constrained space imposed by the original definition of the probabilistic problem. The step-size was tuned automatically using a dual averaging policy [35] for an acceptance rate of around 0.6. We found this combination of algorithms to be robust, reasonably performant and largely free of problem-specific parameter tuning. Standard MCMC chain diagnostics were performed including auto-correlation plots and trace plots.

Table 1: SIM. Summary statistics on the parameters. Number of training points $N_{\mathrm{obs}} = 6$, Number of predictions $N_{\mathrm{pred}} = 100$, Number of MCMC samples $= 10^4$, MCMC acceptance ratio $= 0.69$.

	Posterior		Prior	
	mean	stddev	mean	stddev
λ	0.58	0.05	1.0	0.45
β	0.1	0.04	0.1	0.04

.

3.2. Purely data–driven ML approach

An alternative to deterministic models like Tanner's formula are purely data driven models, which rely on the experimental data exclusively. In a first step, we postulate

$$y_j^{\mathrm{obs}} = \eta^{\mathrm{GP}}(x_j^{\mathrm{obs}}) + \epsilon_j^{\mathrm{obs}}, \qquad \forall j = 1, \ldots, N_{\mathrm{obs}}, \tag{10}$$

assuming that the experimental observations can be reproduced, up to measurement noise, by evaluating the objective function for our regression, η^{GP}, at the training data abscissas. Common surrogate models for η^{GP} are, e.g. Polynomial Chaos Expansions (Pres) [60] or Gaussian Processes [44].

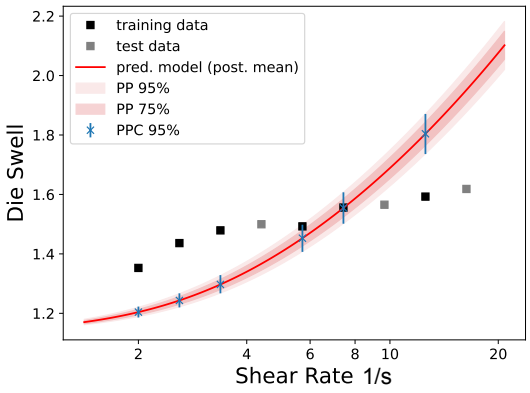

Figure 4: Swell vs. shear rate. Results from the SIM approach. Number of training samples is 6 from physical experiments. Posterior mean for the simulation (dashed line) and the full predictive model (simulation + GP model discrepancy; full line) with its 95% credibility interval. Statistics from the posterior predictive distribution are visualised in blue

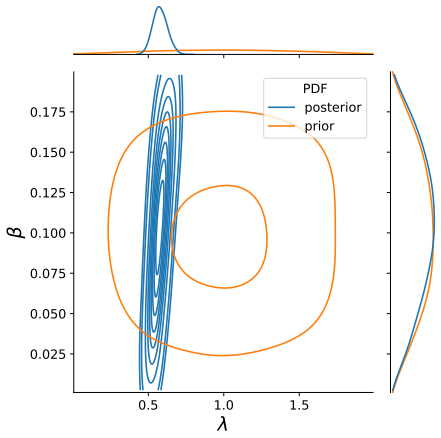

Figure 5: SIM. Approximation of the prior and posterior PDFs for the material parameters. The prior and posterior PDFs are both symmetric in all parameters (horizontal and vertical directions, which means that there is no correlation). The calibration yields a clear shift and tightening of the posterior for the relaxation time λ, whereas the posterior for the solvent viscosity ratio β does barely change. This is due to the exclusive use of die swell data for the calibration, as discussed below in section 4

Gaussian processes (GP).. In the present study, we use a Gaussian Process prior model,

$$\eta^{\mathrm{GP}} \sim \mathcal{GP}(\mu_{\mathrm{GP}}, \Sigma_{\mathrm{GP}}), \tag{11}$$

where $\mu_{\mathrm{GP}} : \mathcal{U} \to \mathbb{R}$ is the GP mean function, and $\Sigma_{\mathrm{GP}} : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ is the symmetric positive-definite GP covariance kernel function. The statement (11) implies that for all finite subsets $U \subset \mathcal{U}$ and for all $x \in U$,

$$\eta^{\mathrm{GP}}(x) \sim \mathcal{N}(\mu_{\mathrm{GP}}(x), (\Sigma_{\mathrm{GP}}(x, x'))_{x' \in U}), \tag{12}$$

which can be used for a discretisation of the GP model.

A Gaussian Process is well defined by its mean and covariance kernel functions. In practice, one often uses parametric functions to represent the GP mean and covariance functions (hierarchical models, see e.g. [29]) and thus reduces the inference on infinite dimensional function spaces to an inference on some finite set of so-called hyper–parameters, which we denote by $\boldsymbol{\nu}_{\mathrm{GP}}$. When needed, we denote this dependence explicitly with square brackets, as $\mu_{\mathrm{GP}}[\boldsymbol{\nu}_{\mathrm{GP}}]$, $\Sigma_{\mathrm{GP}}[\boldsymbol{\nu}_{\mathrm{GP}}]$. An in-depth review on GP can be found, e.g. in the textbook [44].

We now want to conduct a Bayesian regression on the data numerically. To this aim, let us define the discrete set of input abscissas as $\mathbf{x}^{\mathrm{obs}} = (x_j^{\mathrm{obs}})_{j=1,\dots,N_{\mathrm{obs}}}$. Let us furthermore convene an abuse of notation, analogously to operator overloading in programming languages,

$$\eta^{\mathrm{GP}}(\mathbf{x}) = (\eta^{\mathrm{GP}}(x))_{x \in \mathbf{x}}, \quad \mu_{\mathrm{GP}}(\mathbf{x}) = (\mu_{\mathrm{GP}}(x))_{x \in \mathbf{x}}, \quad \Sigma_{\mathrm{GP}}(\mathbf{x}, \mathbf{x}') = (\Sigma_{\mathrm{GP}}(x, x'))_{x \in \mathbf{x}, x' \in \mathbf{x}'},$$

for any $\mathbf{x}, \mathbf{x}' \subset \mathcal{U}$.

The Bayesian model then writes

$$p\Big(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}} \;\Big|\; \mathbf{y}^{\mathrm{obs}}\Big) \propto p\Big(\mathbf{y}^{\mathrm{obs}} \;\Big|\; \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}}\Big) p_0(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}}).$$

From equation (10), the likelihood function has a structure inherited from the noise PDF, and does not depend on the hyper–parameters directly, i.e.

$$\mathbf{y}^{\mathrm{obs}} \mid \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}} = \mathbf{y}^{\mathrm{obs}} \mid \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}) \sim \mathcal{N}\Big(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \sigma^2_{\mathrm{obs}} \mathbf{I}\Big). \tag{13}$$

The prior in the Bayesian model is hierarchically defined. From (12), the prior on the GP is a multivariate Gaussian,

$$p_0(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}) \mid \boldsymbol{\nu}_{\mathrm{GP}}) = \mathcal{N}(\mu_{\mathrm{GP}}[\boldsymbol{\nu}_{\mathrm{GP}}](\mathbf{x}^{\mathrm{obs}}), \Sigma_{\mathrm{GP}}[\boldsymbol{\nu}_{\mathrm{GP}}](\mathbf{x}^{\mathrm{obs}}, \mathbf{x}^{\mathrm{obs}})),$$

The prior on the hyper–parameters, $p_0(\boldsymbol{\nu}_{\mathrm{GP}})$, has to be set as discussed below. By the chain rule of probability, we have for the full joint prior,

$$p_0(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}}) = p_0(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}) \mid \boldsymbol{\nu}_{\mathrm{GP}}) p_0(\boldsymbol{\nu}_{\mathrm{GP}}).$$

The posterior predictive distribution for this model is

$$\begin{aligned} p_{\mathrm{PP}}(\mathbf{y}^{\mathrm{pred}} \mid \mathbf{y}^{\mathrm{obs}}) = \int & p\Big(\mathbf{y}^{\mathrm{pred}} \;\Big|\; \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}})\Big) p\Big(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}) \;\Big|\; \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}}\Big) \\ & \times p\Big(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}} \;\Big|\; \mathbf{y}^{\mathrm{obs}}\Big) \mathrm{d}\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}) \mathrm{d}\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}) \mathrm{d}\boldsymbol{\nu}_{\mathrm{GP}}. \end{aligned}$$

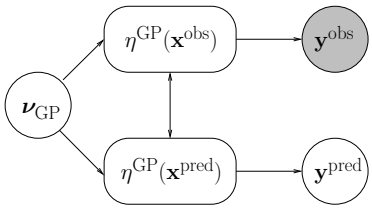

Figure 6: Bayesian graphical model for the GP regression. The nodes with a full line contour indicate random variables. The filled node represents an observed quantity, and the un–filled nodes represent un–observed quantities

We assume the validity of relation (10) for the predictions, s.t. we have

$$\mathbf{y}^{\mathrm{pred}} \mid \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}) \sim \mathcal{N}\Big(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}), \sigma^2_{\mathrm{obs}}\mathbf{I}\Big).$$

Consequently, it holds for the PP expectation,

$$\mathbb{E}[\mathbf{y}^{\mathrm{pred}} \mid \mathbf{y}^{\mathrm{obs}}] = \mathbb{E}[\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}) \mid \mathbf{y}^{\mathrm{obs}}].$$

Furthermore, from [44, Eq. (A.6)] the conditioned GP obeys the relation

$$\begin{aligned}\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}) \,\big|\, \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}} \\ \sim \mathcal{N}\Big(\mu_{\mathrm{GP}}[\boldsymbol{\nu}_{\mathrm{GP}}](\mathbf{x}^{\mathrm{pred}}) - \Sigma^{-1}_{\mathrm{GP}}[\boldsymbol{\nu}_{\mathrm{GP}}](\mathbf{x}^{\mathrm{pred}}, \mathbf{x}^{\mathrm{pred}})\Sigma_{\mathrm{GP}}[\boldsymbol{\nu}_{\mathrm{GP}}](\mathbf{x}^{\mathrm{pred}}, \mathbf{x}^{\mathrm{obs}}) \\ \times \Big(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}) - \mu_{\mathrm{GP}}[\boldsymbol{\nu}_{\mathrm{GP}}](\mathbf{x}^{\mathrm{obs}})\Big), \Sigma^{-1}_{\mathrm{GP}}[\boldsymbol{\nu}_{\mathrm{GP}}](\mathbf{x}^{\mathrm{pred}}, \mathbf{x}^{\mathrm{pred}})\Big).\end{aligned} \tag{14}$$

3.2.1. Numerical Results

The TensorFlow Probability framework (TFP) offers an intuitive way of implementing Bayesian models, in the sense that TFP code typcally resembles the mathematical formlae from above. It contains many PDFs, and as an open source code, it is highly customizable, s.t. we can easily define our priors hierarchically. As covariance kernel function, we choose the so–called "squared exponential", "Gaussian" or "radial basis function"

$$k(\mathbf{x}, \mathbf{y}) = \sigma^2_{\mathrm{GP}} e^{\frac{-\|\mathbf{x}-\mathbf{y}\|^2}{2{L_{\mathrm{GP}}}^2}}, \tag{15}$$

which contains as hyper–parameters the standard deviation at any given input $\mathbf{x}$, $\sigma_{\mathrm{GP}} = \sqrt{k(\mathbf{x}, \mathbf{x})}$, and the correlation length L_{GP}. It is a standard GP kernel and provided by TFP. Following the literature [29, 47, 44], we parametrize the GP mean function to be constant,

$$\mu_{\mathrm{GP}}(x) = c_{\mathrm{GP}}, \tag{16}$$

the posterior mean function being implicitly approximated from discretised Gaussian Process realisations, conditioned on the observations. Hence,

$$\boldsymbol{\nu}_{\mathrm{GP}} = (c_{\mathrm{GP}}, \sigma_{\mathrm{GP}}, L_{\mathrm{GP}}). \tag{17}$$

For the actual choice of the priors (cf. table 5), we rely on the following considerations. The length scale parameter should not be too small, because we believe that extrudate swell is a continuous and rather smooth process, nor should it be too big in order that the model can capture small-scale local effects. We choose a truncated normal prior with a maximum at $1/5$ of the domain of consideration. The amplitude parameter reflects the accuracy of the model output, i.e. the absolute deviation of the model output from its parameterised mean, $|\eta^{\mathrm{GP}} - c_{\mathrm{GP}}|$. Its prior should take this into account, and might not promote values smaller than the experimental noise. We choose an inverse–gamma distribution with a maximum at $1/10$ and a mean at $1/6$. The prior for the constant mean function was set at a value of 1.5 based on the assumption that swell should not drop below 1.0 and will probably not grow too large over the range of applied shear rate.

The number of considered samples and burn–in steps are reported in the corresponding figures.

Table 2: ML. Summary statistics on the parameters. Number of training points $N_{\mathrm{obs}} = 6$, Number of predictions $N_{\mathrm{pred}} = 100$, Number of MCMC samples $= 10^4$, MCMC acceptance ratio $= 0.63$.

	Posterior		Prior	
	mean	stddev	mean	stddev
c_{GP}	1.49	0.04	1.48	0.06
σ_{GP}	0.07	0.03	0.08	0.04
L_{GP}	0.65	0.18	0.61	0.19

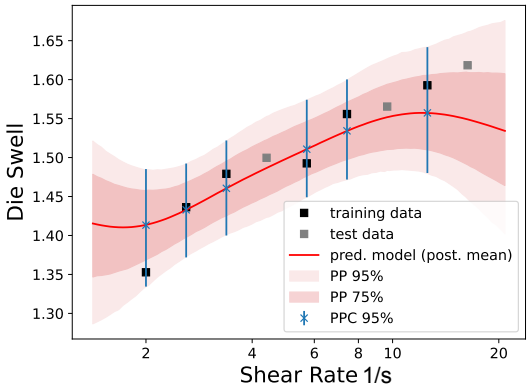

Figure 7: Swell vs. shear rate. Results from the ML approach. Number of training samples is 6 from physical experiments. Posterior mean for the full predictive model with its 95% credibility interval. Statistics from the posterior predictive distribution are visualised in blue

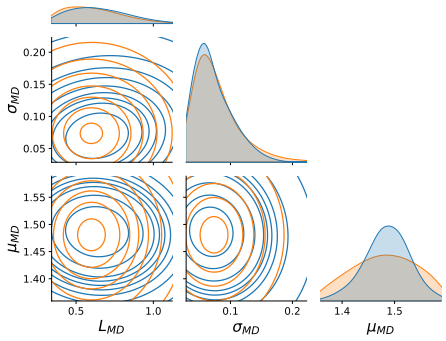

Figure 8: ML. Approximation of the prior (orange) and posterior (blue) PDFs for the GP hyper-parameters. The prior PDF is symmetric in all parameters, meaning that there is no correlation. The posterior PDF does not differ much form the prior specification, which is expected, as the amount of observed data is too low too have a strong impact on a plausible prior

3.3. Kennedy & O'Hagan approach: model discrepancy

In this section, we outline how to take into account the inherent uncertainty of deterministic modelling and simulation in a statistical regression framework. Following the seminal paper [29], we can combine the previous two approaches, in that we conduct a calibration of the physics model parameters θ and introduce a GP model discrepancy term for simultaneous regression.

The model writes

$$y_j^{\text{obs}} = \eta^{\text{sim}}(x_j^{\text{obs}}, \boldsymbol{\theta}) + \eta^{\text{GP}}(x_j^{\text{obs}}) + \epsilon_j^{\text{obs}}, \qquad \forall j = 1, \dots, N_{\text{obs}}, \tag{18}$$

$$\eta^{\text{GP}} \sim \mathcal{GP}(\mu_{\text{GP}}, \Sigma_{\text{GP}}), \tag{19}$$

where $\eta^{\text{sim}} : \mathcal{U} \times \mathbb{R}^{N_{\text{par}}} \to \mathcal{V}$ represents the simulator, and $\eta^{\text{GP}} : \mathcal{U} \to \mathcal{V}$ is the model discrepancy function described as a GP.

The associated Bayesian model then writes

$$p\Big(\eta^{\text{GP}}(\mathbf{x}^{\text{obs}}), \boldsymbol{\nu}_{\text{GP}}, \boldsymbol{\theta} \mid \mathbf{y}^{\text{obs}}\Big) \propto p\Big(\mathbf{y}^{\text{obs}} \mid \eta^{\text{GP}}(\mathbf{x}^{\text{obs}}), \boldsymbol{\nu}_{\text{GP}}, \boldsymbol{\theta}\Big) p_0\Big(\eta^{\text{GP}}(\mathbf{x}^{\text{obs}}), \boldsymbol{\nu}_{\text{GP}}, \boldsymbol{\theta}\Big).$$

From equation (18), we deduce for the likelihood function,

$$\mathbf{y}^{\text{obs}} \mid \eta^{\text{GP}}(\mathbf{x}^{\text{obs}}), \boldsymbol{\theta} \sim \mathcal{N}\Big(\eta^{\text{sim}}(\mathbf{x}^{\text{obs}}, \boldsymbol{\theta}) + \eta^{\text{GP}}(\mathbf{x}^{\text{obs}}), \sigma_{\text{obs}}^2 \mathbf{I}\Big). \tag{20}$$

The prior in the Bayesian model is, again, hierarchically defined, due to the GP. From (12), the prior on $\eta^{\text{GP}}(\mathbf{x}^{\text{obs}})$ is a multivariate Gaussian,

$$p_0(\eta^{\text{GP}}(\mathbf{x}^{\text{obs}}) \mid \boldsymbol{\nu}_{\text{GP}}) = \mathcal{N}(\mu_{\text{GP}}(\mathbf{x}^{\text{obs}}), \Sigma_{\text{GP}}[\boldsymbol{\nu}_{\text{GP}}](\mathbf{x}^{\text{obs}}, \mathbf{x}^{\text{obs}})),$$

The priors for the physical parameters $\boldsymbol{\theta}$ and the hyper–parameters $\boldsymbol{\nu}_{\mathrm{GP}}$ can reasonably be assumed independent, i.e.

$$p_0(\boldsymbol{\nu}_{\mathrm{GP}}, \boldsymbol{\theta}) = p_0(\boldsymbol{\nu}_{\mathrm{GP}}) p_0(\boldsymbol{\theta}),$$

and are set according to table 5. By the chain rule of probability (and the just stated independence), we have for the full joint prior,

$$p_0(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}}, \boldsymbol{\theta}) = p_0(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}) \mid \boldsymbol{\nu}_{\mathrm{GP}}) p_0(\boldsymbol{\nu}_{\mathrm{GP}}) p_0(\boldsymbol{\theta}).$$

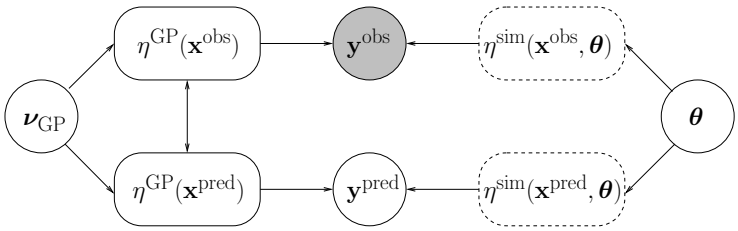

Figure 9: Bayesian graphical model for the KOH approach. The nodes with a full line contour indicate random variables. The filled node represents an observed quantity, and the unfilled nodes represent unobserved quantities. The nodes with a dashed line contour indicate deterministic transformations of the upstream nodes

The posterior predictive distribution is,

$$\begin{aligned} &p_{\mathrm{PP}}\Big(\mathbf{y}^{\mathrm{pred}} \;\Big|\; \mathbf{y}^{\mathrm{obs}}\Big) \\ &\quad = \int p\Big(\mathbf{y}^{\mathrm{pred}} \;\Big|\; \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}), \boldsymbol{\theta}\Big) p\Big(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}) \;\Big|\; \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}}\Big) \\ &\qquad \times p\Big(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}}, \boldsymbol{\theta} \;\Big|\; \mathbf{y}^{\mathrm{obs}}\Big) \mathrm{d}\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}) \mathrm{d}\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}) \mathrm{d}\boldsymbol{\nu}_{\mathrm{GP}} \mathrm{d}\boldsymbol{\theta}. \end{aligned}$$

Assuming the validity of relation (18) for the predictions, we have

$$\mathbf{y}^{\mathrm{pred}} \mid \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}), \boldsymbol{\theta} \sim \mathcal{N}\Big(\eta^{\mathrm{sim}}(\mathbf{x}^{\mathrm{pred}}, \boldsymbol{\theta}) + \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}), \sigma^2_{\mathrm{obs}} \mathbf{I}\Big),$$

and hence

$$\mathbb{E}[\mathbf{y}^{\mathrm{pred}} \mid \mathbf{y}^{\mathrm{obs}}] = \mathbb{E}[\eta^{\mathrm{sim}}(\mathbf{x}^{\mathrm{pred}}, \boldsymbol{\theta}) + \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}) \mid \mathbf{y}^{\mathrm{obs}}].$$

For the conditioned GP, $\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}) \;\big|\; \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}}$, equation (14) holds true.

3.3.1. Numerical results

We again use the squared exponential GP covariance function (15), and a constant mean function, s.t. we have $\boldsymbol{\nu}_{\mathrm{GP}} = (c_{\mathrm{GP}}, \sigma_{\mathrm{GP}}, L_{\mathrm{GP}})$. The prior for c_{GP} is set around zero, as we assume the model discrepancy term to be small, and do not have any prior evidence of its sign.

In TFP, we build up the joint PDF for the full model, step by step. Let us first define the priors for the GP hyper–parameters.

```
def prior_mean():
    return tfd.Normal(loc=0., name="mean")

def prior_amplitude():
    return tfd.InverseGamma(concentration, scale, name="amplitude")

def prior_length_scale():
    return tfd.TruncatedNormal(loc, scale, low, high, name="length_scale" )
```

As already mentioned, we define $\mathbf{x}^{\mathrm{pred}}$ as a homogeneous discretisation of the input space $\mathcal{U}$, and include $\mathbf{y}^{\mathrm{pred}}$ in the MCMC sampling from the (enriched) posterior PDF. We hence define the discretised GP over the input abscissas $\{\mathbf{x}^{\mathrm{obs}}, \mathbf{x}^{\mathrm{pred}}\}$, containing both, inputs for observations and predictions. The GP prior depends now on the hyper–parameters $\boldsymbol{\nu}_{\mathrm{GP}}$ and can be defined as follows.

```
def prior_GP(length_scale, amplitude, mean):
    kernel = tensorflow_probability.math.psd_kernels.ExponentiatedQuadratic(
        amplitude, length_scale)
    kernel_matrix = kernel.matrix(
        tnp.concatenate((x_obs, x_pred)), tnp.concatenate((x_obs, x_pred)))
    kernel_matrix = tf.linalg.set_diag(
        kernel_matrix, tf.linalg.diag_part(kernel_matrix) + 1E-5) # Add jitter
    scale = tf.linalg.LinearOperatorLowerTriangular(
        tf.linalg.cholesky(kernel_matrix))
    return tfd.MultivariateNormalLinearOperator(
        loc=mean, scale=scale, name="GP")
```

The prior for the physical parameters is defined separately,

```
def prior_theta():
    return tfd.Sample(tfd.Beta(concentration1, concentration0),
                      sample_shape=2, name="theta")
```

The likelihood function can be defined as

```
def likelihood(theta, GP):
    loc = GP[0:len(x_obs)] + f_sim(x_obs,theta)
    return tfd.MultivariateNormalDiag(
        loc, scale_identity_multiplier=noise_stddev, name="observations")
```

Now, a full joint PDF for the model problem can be constructed by means of the chain rule of probability,

```
model = tfd.JointDistributionSequentialAutoBatched( [
    prior_mean,
    prior_amplitude,
    prior_length_scale,
    prior_GP,
    prior_theta,
    likelihood])
```

and subsequently be conditioned by the observed data,

```
model_conditioned = model.experimental_pin(observations=y_obs)
```

Table 3: KOH. Summary statistics on the parameters. Number of training points $N_{\text{obs}} = 6$, Number of predictions $N_{\text{pred}} = 100$, Number of MCMC samples $= 10^4$, MCMC acceptance ratio $= 0.64$.

	Posterior		Prior	
	mean	stddev	mean	stddev
λ	0.4	0.15	1.0	0.45
β	0.1	0.04	0.1	0.04
c_{GP}	0.07	0.08	0.0	0.08
σ_{GP}	0.1	0.04	0.08	0.04
L_{GP}	0.66	0.18	0.61	0.18

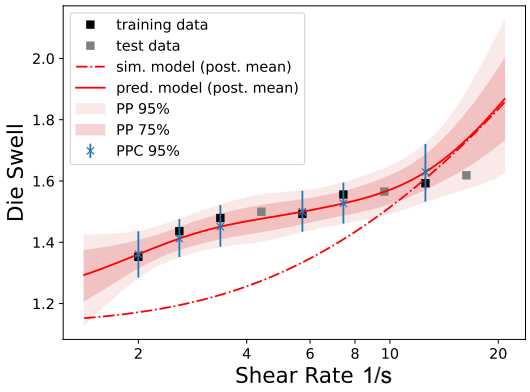

Figure 10: Swell vs. shear rate. Results from the KOH approach. Number of training samples is 6 from physical experiments. Posterior mean for the simulation (dashed line) and the full predictive model (simulation + GP model discrepancy; full line) with its 95% credibility interval. Statistics from the posterior predictive distribution are visualized in blue

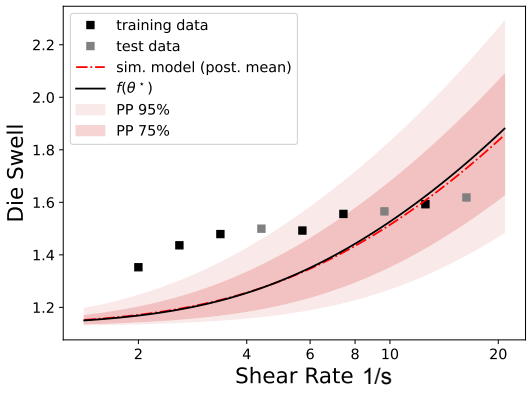

Figure 11: KOH. Posterior mean and uncertainty intervals for the simulation output compared to the exact simulations at the calibrated (posterior mean) model parameters $\theta^\star$. The discrepancy between the predicted vs. calibrated simulation output is due to the non linearity in θ of Tanner's formula. Also note that, consequently, the predicted simulation output need not obey Tanner's formula any longer

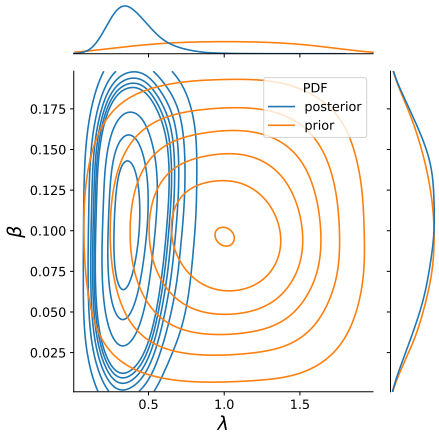

Figure 12: KOH. Approximation of the prior and posterior PDFs for the material parameters. The prior and posterior PDFs are both showing no correlation. The calibration yields a clear shift and tightening of the posterior for the relaxation time λ. The tightening is however less pronounced than in Fig. 5, which is because of the presence of the MD term, which smooths the calibration procedure, as discussed below in section 4

3.4. Higdon approach: model discrepancy and GP surrogate model

In the case of Tanner's formula for die swell "simulation", we can run the "simulator" as often as we want, since it is a simple scalar algebraic formula. For computationally more expensive simulators, e.g. a full fluid dynamics simulation, and for use cases beyond die swell prediction, the number of simulation runs is typically bounded. In this

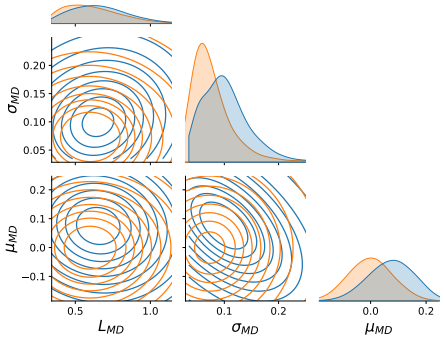

Figure 13: KOH. Approximation of the prior (orange) and posterior (blue) PDFs for the GP hyper-parameters The prior PDF is symmetric in all parameters, meaning that there is no correlation. The posterior PDF differs more form the prior specification than it does for the ML model. This is due to the presence of the physical simulator, which imposes a bias on the ML module, which is the MD term within the KOH model.

case, most of the literature is concerned model order reduction (MOR), i.e. with establishing a cheap surrogate model, which approximates the original simulator (e.g. [47]). In this context, parametric approaches, such as PCEs, and non parametric approaches, such as GPs, are both widely used. Only a few literature exist for combining MOR with the KOH approach. In [23], a method is developed, which represents the model discrepancy term as well as the surrogate model for the simulator as GPs, and implements a simultaneous regression on these stochastic processes, their hyper–parameters, and the physics model parameters. We will also apply and study this method on our example.

In order to define a surrogate $\tilde{\eta}^{\mathrm{sim}}$ for η^{sim} as a GP, the original domain $\mathcal{U} \subset \mathbb{R}$ of a scalar x–observable has to be extended to the whole space $\mathcal{U} \times \mathbb{R}^{N_{\mathrm{par}}}$ of inputs to the simulator. This extension is straightforward, and we refer to the literature for a detailed presentation on multi–dimensional input space GPs [44]. The data from the limited number of simulation runs consists of N_{sim} inputs

$$(\mathbf{x}^{\mathrm{sim}}, \mathbf{t}^{\mathrm{sim}}) = (x_j^{\mathrm{sim}}, \boldsymbol{\theta}_j^{\mathrm{sim}})_{j=1,\ldots,N_{\mathrm{sim}}} \subset \mathcal{U} \times \mathbb{R}^{N_{\mathrm{par}}},$$

and scalar outputs

$$\mathbf{y}^{\mathrm{sim}} = (y_j^{\mathrm{sim}})_{j=1,\ldots,N_{\mathrm{sim}}} \subset \mathcal{V}.$$

We further allow for some Gaussian white noise in the data and assume the variance to be known,

$$\epsilon_j^{\mathrm{sim}} \sim_{\mathrm{iid}} \mathcal{N}(0, \tilde{\sigma}_{\mathrm{sim}}^2). \tag{21}$$

Concerning the model discrepancy term, we will model η^{GP} as a GP, as before in the KOH approach.

The full model then writes,

$$y_j^{\text{obs}} = \tilde{\eta}^{\text{sim}}(x_j^{\text{obs}}, \boldsymbol{\theta}) + \eta^{\text{GP}}(x_j^{\text{obs}}) + \epsilon_j^{\text{obs}} + \epsilon_j^{\text{sim}}, \qquad \forall j = 1, \dots, N_{\text{obs}}, \tag{22}$$

$$\tilde{\eta}^{\text{sim}}(x_j^{\text{sim}}, \boldsymbol{\theta}_j^{\text{sim}}) = \eta^{\text{sim}}(x_j^{\text{sim}}, \boldsymbol{\theta}_j^{\text{sim}}) + \epsilon_j^{\text{sim}}, \qquad \forall j = 1, \dots, N_{\text{sim}}, \tag{23}$$

$$\tilde{\eta}^{\text{sim}} \sim \mathcal{GP}(\tilde{\mu}_{\text{sim}}, \tilde{\Sigma}_{\text{sim}}), \tag{24}$$

$$\eta^{\text{GP}} \sim \mathcal{GP}(\mu_{\text{GP}}, \Sigma_{\text{GP}}). \tag{25}$$

The associated Bayesian model is

$$\begin{aligned} &p\Big(\eta^{\text{GP}}(\mathbf{x}^{\text{obs}}), \boldsymbol{\nu}_{\text{GP}}, \tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{obs}}, \boldsymbol{\theta}), \tilde{\boldsymbol{\nu}}_{\text{sim}}, \boldsymbol{\theta} \mid \mathbf{y}^{\text{obs}}, \mathbf{y}^{\text{sim}}\Big) \\ &\quad \propto p\Big(\mathbf{y}^{\text{obs}}, \mathbf{y}^{\text{sim}} \mid \eta^{\text{GP}}(\mathbf{x}^{\text{obs}}), \boldsymbol{\nu}_{\text{GP}}, \tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{obs}}, \boldsymbol{\theta}), \tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{sim}}, \mathbf{t}^{\text{sim}}), \tilde{\boldsymbol{\nu}}_{\text{sim}}, \boldsymbol{\theta}\Big) \\ &\quad p_0\Big(\eta^{\text{GP}}(\mathbf{x}^{\text{obs}}), \boldsymbol{\nu}_{\text{GP}}, \tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{obs}}, \boldsymbol{\theta}), \tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{sim}}, \mathbf{t}^{\text{sim}}), \tilde{\boldsymbol{\nu}}_{\text{sim}}, \boldsymbol{\theta}\Big). \end{aligned}$$

From equations (22), (23), we deduce for the likelihood function,

$$\begin{bmatrix} \mathbf{y}^{\text{obs}} \\ \mathbf{y}^{\text{sim}} \end{bmatrix} \Big| \ \eta^{\text{GP}}(\mathbf{x}^{\text{obs}}), \tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{obs}}, \boldsymbol{\theta}), \tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{sim}}, \mathbf{t}^{\text{sim}}), \boldsymbol{\theta} \tag{26}$$

$$\sim \mathcal{N}\Big(\begin{bmatrix} \tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{obs}}, \boldsymbol{\theta}) + \eta^{\text{GP}}(\mathbf{x}^{\text{obs}}) \\ \tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{sim}}, \mathbf{t}^{\text{sim}}) \end{bmatrix}, \begin{bmatrix} (\sigma_{\text{obs}}^2 + \tilde{\sigma}_{\text{sim}}^2)\mathbf{I} & 0 \\ 0 & \tilde{\sigma}_{\text{sim}}^2 \mathbf{I} \end{bmatrix} \Big). \tag{27}$$

The priors can be set up independently, for each of the GPs, using again a hierarchical definition. For the model discrepancy prior, we have

$$p_0\Big(\eta^{\text{GP}}(\mathbf{x}^{\text{obs}}), \boldsymbol{\nu}_{\text{GP}}\Big) = p_0\Big(\eta^{\text{GP}}(\mathbf{x}^{\text{obs}}) \mid \boldsymbol{\nu}_{\text{GP}}\Big) p_0(\boldsymbol{\nu}_{\text{GP}}),$$

with (from the GP definition (12)),

$$p_0\Big(\eta^{\text{GP}}(\mathbf{x}^{\text{obs}}) \mid \boldsymbol{\nu}_{\text{GP}}\Big) = \mathcal{N}\Big(\mu_{\text{GP}}(\mathbf{x}^{\text{obs}}), \Sigma_{\text{GP}}[\boldsymbol{\nu}_{\text{GP}}](\mathbf{x}^{\text{obs}}, \mathbf{x}^{\text{obs}})\Big),$$

and a prior on $\boldsymbol{\nu}_{\text{GP}}$ as reported in table 5.

For the surrogate simulator, we have

$$p_0\Big(\tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{obs}}, \boldsymbol{\theta}), \tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{sim}}, \mathbf{t}^{\text{sim}}), \tilde{\boldsymbol{\nu}}_{\text{sim}}, \boldsymbol{\theta}\Big) = p_0\Big(\tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{obs}}, \boldsymbol{\theta}), \tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{sim}}, \mathbf{t}^{\text{sim}}) \mid \tilde{\boldsymbol{\nu}}_{\text{sim}}, \boldsymbol{\theta}\Big) p_0(\tilde{\boldsymbol{\nu}}_{\text{sim}}) p_0(\boldsymbol{\theta}),$$

with (cf. equation (12)),

$$\begin{aligned} &p_0\Big(\tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{obs}}, \boldsymbol{\theta}), \tilde{\eta}^{\text{sim}}(\mathbf{x}^{\text{sim}}, \mathbf{t}^{\text{sim}}) \mid \tilde{\boldsymbol{\nu}}_{\text{sim}}, \boldsymbol{\theta}\Big) \\ &\quad = \mathcal{N}\Big(\begin{bmatrix} \tilde{\mu}_{\text{sim}}(\mathbf{x}^{\text{obs}}, \boldsymbol{\theta}) \\ \tilde{\mu}_{\text{sim}}(\mathbf{x}^{\text{sim}}, \mathbf{t}^{\text{sim}}) \end{bmatrix}, \begin{bmatrix} \tilde{\Sigma}_{\text{sim}}[\tilde{\boldsymbol{\nu}}_{\text{sim}}]\big((\mathbf{x}^{\text{obs}}, \boldsymbol{\theta}), (\mathbf{x}^{\text{obs}}, \boldsymbol{\theta})\big) & \tilde{\Sigma}_{\text{sim}}[\tilde{\boldsymbol{\nu}}_{\text{sim}}]\big((\mathbf{x}^{\text{obs}}, \boldsymbol{\theta}), (\mathbf{x}^{\text{sim}}, \mathbf{t}^{\text{sim}})\big) \\ \tilde{\Sigma}_{\text{sim}}[\tilde{\boldsymbol{\nu}}_{\text{sim}}]\big((\mathbf{x}^{\text{sim}}, \mathbf{t}^{\text{sim}}), (\mathbf{x}^{\text{obs}}, \boldsymbol{\theta})\big) & \tilde{\Sigma}_{\text{sim}}[\tilde{\boldsymbol{\nu}}_{\text{sim}}]\big((\mathbf{x}^{\text{sim}}, \mathbf{t}^{\text{sim}}), (\mathbf{x}^{\text{sim}}, \mathbf{t}^{\text{sim}})\big) \end{bmatrix} \Big). \end{aligned}$$

The priors for the physical parameters $\boldsymbol{\theta}$ and the hyper–parameters $\boldsymbol{\nu}_{\mathrm{GP}}$ can reasonably be assumed independent, i.e.

$$p_0(\boldsymbol{\nu}_{\mathrm{GP}}, \boldsymbol{\theta}) = p_0(\boldsymbol{\nu}_{\mathrm{GP}}) p_0(\boldsymbol{\theta}),$$

and are set according to table 5.

By the chain rule of probability (and the just stated independence), we have for the full joint prior,

$$\begin{aligned} &p_0\Big(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}}, \tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{obs}}, \boldsymbol{\theta}), \tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{sim}}, \mathbf{t}^{\mathrm{sim}}), \tilde{\boldsymbol{\nu}}_{\mathrm{sim}}, \boldsymbol{\theta}\Big) \\ &\quad = p_0\Big(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}) \;\Big|\; \boldsymbol{\nu}_{\mathrm{GP}}\Big) p_0(\boldsymbol{\nu}_{\mathrm{GP}}) p_0\Big(\tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{obs}}, \boldsymbol{\theta}), \tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{sim}}, \mathbf{t}^{\mathrm{sim}}) \mid \tilde{\boldsymbol{\nu}}_{\mathrm{sim}}, \boldsymbol{\theta}\Big) p_0(\tilde{\boldsymbol{\nu}}_{\mathrm{sim}}) p_0(\boldsymbol{\theta}). \end{aligned}$$

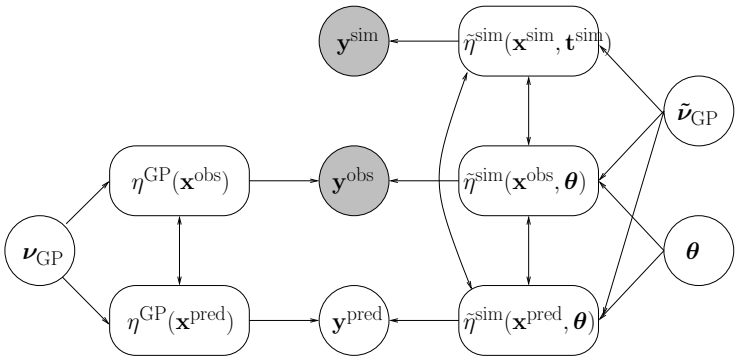

Figure 14: Bayesian graphical model for Higdon's approach. The nodes with a full line contour indicate random variables. The filled nodes represent observed quantities, and the un–filled nodes represent un–observed quantities

The posterior predictive distribution is,

$$\begin{aligned} &p_{\mathrm{PP}}\Big(\mathbf{y}^{\mathrm{pred}} \;\Big|\; \mathbf{y}^{\mathrm{obs}}, \mathbf{y}^{\mathrm{sim}}\Big) \\ &\quad = \int p\Big(\mathbf{y}^{\mathrm{pred}} \;\Big|\; \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}), \tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{pred}}, \boldsymbol{\theta})\Big) p\Big(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}) \;\Big|\; \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}}\Big) \\ &\qquad \times p\Big(\tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{pred}}, \boldsymbol{\theta}) \;\Big|\; \tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{obs}}, \boldsymbol{\theta}), \tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{sim}}, \mathbf{t}^{\mathrm{sim}}), \tilde{\boldsymbol{\nu}}_{\mathrm{sim}}, \boldsymbol{\theta}\Big) \\ &\qquad \times p\Big(\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}}, \tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{obs}}, \boldsymbol{\theta}), \tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{sim}}, \mathbf{t}^{\mathrm{sim}}), \tilde{\boldsymbol{\nu}}_{\mathrm{sim}}, \boldsymbol{\theta} \;\Big|\; \mathbf{y}^{\mathrm{obs}}, \mathbf{y}^{\mathrm{sim}}\Big) \\ &\qquad \times \mathrm{d}\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}) \mathrm{d}\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}) \mathrm{d}\boldsymbol{\nu}_{\mathrm{GP}} \mathrm{d}\tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{pred}}, \boldsymbol{\theta}) \mathrm{d}\tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{obs}}, \boldsymbol{\theta}) \mathrm{d}\tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{sim}}, \mathbf{t}^{\mathrm{sim}}) \mathrm{d}\tilde{\boldsymbol{\nu}}_{\mathrm{sim}} \mathrm{d}\boldsymbol{\theta}. \end{aligned}$$

Applying (23) on the predictive data, we obtain

$$\begin{aligned} \mathbf{y}^{\mathrm{pred}} \;\Big|\; \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}), \tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{pred}}, \boldsymbol{\theta}), \boldsymbol{\theta} \sim \mathcal{N}\Big(\tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{pred}}, \boldsymbol{\theta}) + \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}), (\sigma^2_{\mathrm{obs}} + \tilde{\sigma}^2_{\mathrm{sim}})\mathbf{I}\Big), \\ \mathbb{E}[\mathbf{y}^{\mathrm{pred}} \mid \mathbf{y}^{\mathrm{obs}}] = \mathbb{E}[\tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{pred}}, \boldsymbol{\theta}) + \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}) \mid \mathbf{y}^{\mathrm{obs}}]. \end{aligned}$$

For the conditioned GP, $\eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{pred}}) \mid \eta^{\mathrm{GP}}(\mathbf{x}^{\mathrm{obs}}), \boldsymbol{\nu}_{\mathrm{GP}}$, equation (14) holds true. For $\tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{pred}}, \boldsymbol{\theta}) \mid \tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{obs}}, \boldsymbol{\theta}), \tilde{\eta}^{\mathrm{sim}}(\mathbf{x}^{\mathrm{sim}}, \mathbf{t}^{\mathrm{sim}}), \tilde{\boldsymbol{\nu}}_{\mathrm{sim}}, \boldsymbol{\theta}$, an analogous formula can be derived.

3.4.1. Numerical results

We use the squared exponential GP covariance function (15), for both GPs, s.t. we have $\boldsymbol{\nu}_{\mathrm{GP}} = (\sigma_{\mathrm{GP}}, L_{\mathrm{GP}})$ and $\tilde{\boldsymbol{\nu}}_{\mathrm{sim}} = (\tilde{\sigma}_{\mathrm{sim}}, \tilde{L}_{\mathrm{sim}})$.

The N_{sim} abscissas used for the acquisition of training data from the simulator are obtained taking a tensor product of 10 samples in one-dimensional the x-space (shear rates) with 20 Latin Hypercube samples in the two-dimensional $\boldsymbol{\theta}$-space (material parameters), giving 200 data points in total.

Table 4: Higdon. Summary statistics on the parameters. Number of training points $N_{\mathrm{obs}} = 6$, $N_{\mathrm{sim}} = 200$, Number of predictions $N_{\mathrm{pred}} = 50$, Number of MCMC samples $= 5 \times 10^4$, MCMC acceptance ratio $= 0.6475$.

	Posterior		Prior	
	mean	stddev	mean	stddev
λ	0.4	0.15	1.0	0.45
β	0.1	0.04	0.1	0.04
c_{GP}	0.06	0.07	0.0	0.08
σ_{GP}	0.11	0.04	0.08	0.04
L_{GP}	0.66	0.18	0.61	0.18
$\tilde{c}_{\mathrm{sim}}$	1.77	0.2	1.8	0.38
$\tilde{\sigma}_{\mathrm{sim}}$	0.45	0.07	0.46	0.46
$\tilde{L}_{\mathrm{sim}}$	(1.2, 0.74, 0.93)	(0.1, 0.05, 0.21)	(0.61, 0.37, 0.79)	(0.18, 0.11, 0.23)

4. Discussion

The main two outputs from our modelling and analysis are

- The predictive function linking shear rate (etc.) with extrudate shape (die swell), including uncertainty intervals.
- The calibration of the physics and statistics based parameters, including summary statistics.

We compared different modelling approaches, namely a pure physics–based model (Tanner's formula for die swell; SIM), a purely data–driven model (ML) using Gaussian processes (GPs), and a additive blend of both (KOH and HIG approaches). All of these models yield predictive functions and uncertainty intervals for die swell, but the ML model can not be used for the calibration of material parameters, since of course they do not appear in its mathematical formulation. We used Bayesian statistics to calibrate the model (hyper–)parameters on the data. The priors on the hyper–parameters

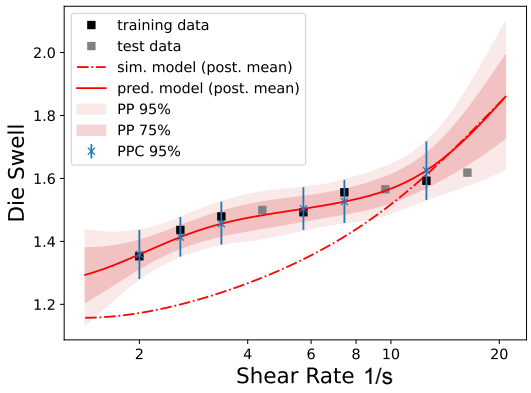

Figure 15: Swell vs. shear rate. Results from the Higdon approach. Number of training samples is 6 from physical experiments and 6^3 from simulation runs. GP posterior mean for the simulation (dashed line) and the full predictive model (simulation + model discrepancy; full line) with its 95% credibility interval. Statistics from the posterior predictive distribution are visualized in blue

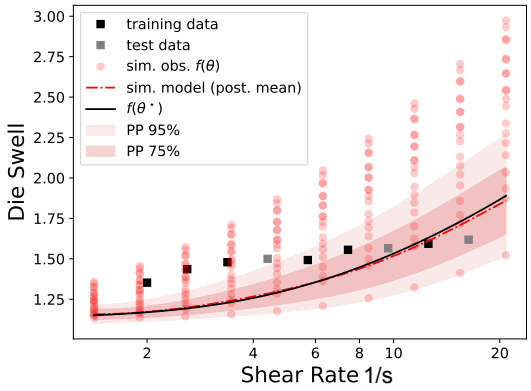

Figure 16: Posterior mean and uncertainty intervals for the simulation GP compared to the exact simulations at the calibrated (posterior mean) model parameters $\theta^\star$. The discrepancy between the predicted vs. calibrated simulation output is due to the non linearity in θ of Tanner's formula

(i.e. the parameters defining the GPs), rely on common–sense arguments, such as assuming a smooth relation between shear rate and swell encoded by the exponential Gaussian function and a relatively long correlation length.

The SIM predictive model exhibits a clear model discrepancy. The best–fit resembles a least–squares fit, meaning that the model error, expressed within the likelihood PDF, overrules the prior specifications on the material parameters. The uncertainty bounds are unrealistically small.

The ML model provides a better fit in the interpolation range of observed data. Out-

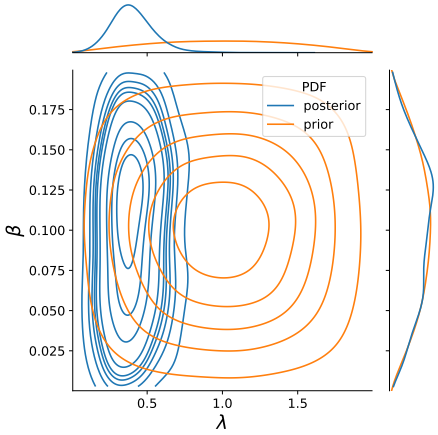

Figure 17: Higdon. Approximation of the prior and posterior PDFs for the material parameters. The setup and results are coinciding with the results from the KOH approach, cf. Fig. 12

side the interpolation range, its predictions converge towards its mean function, which is parameterised as a constant in the present study, being a hyper–parameter in the full BI framework. This is one of the major differences to the coupled physics–based and data–driven KOH approach, which makes use of both the physics model and the Gaussian Process-based MD within and outside the interpolation range. Thus, better results can be expected with the KOH approach, when the physics–based simulator is assumed to represent reality better than the mean of the available data. Tanner's formula being an asymptotic model for die swell at low shear rates, this cannot be affirmed here. Such a study would require numerical simulations at high shear rates, which as we have discussed, suffer from instabilities due to the high Weissenberg number problem, and are a current research challenge.

In the interpolation and close extrapolation range, the KOH predictive model gives better results than the ML model. It is also more robust with respect to the prior specification of the GP hyper–parameters, since the GPs are only used to model the MD, in the KOH approach. Importantly, it provides the calibrated material parameters. The bias observed with the SIM model is due to the prior specification and to the MD term. In our example, the SIM model conducts a calibration very close to least squares fitting restricted to the range of available data. The MD term in the KOH model provides a buffer biasing the fitting of the simulator, which we hope improves the accuracy outside the observation range. It might be argued that this also improves the quality of the calibration, but a general assertion can not be made. An excellent discussion of these issues can be found in [5].

The results from the HIG model were almost identical to those from the KOH model, despite the fact we used a relatively coarse set of points on the parameter space for the evaluation of Tanner's formula. The integration of a surrogate model directly within the Bayesian calibration framework seems to be straightforward.

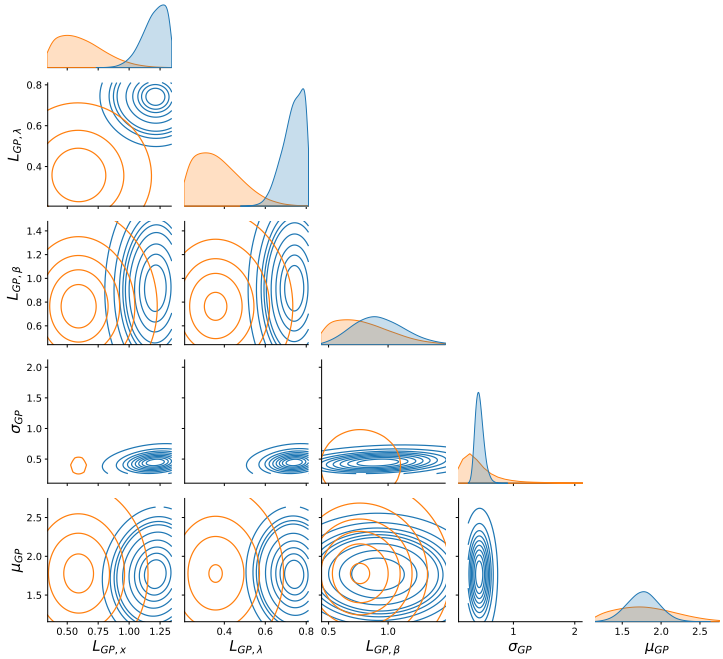

Figure 18: Higdon. Prior and posterior approximations

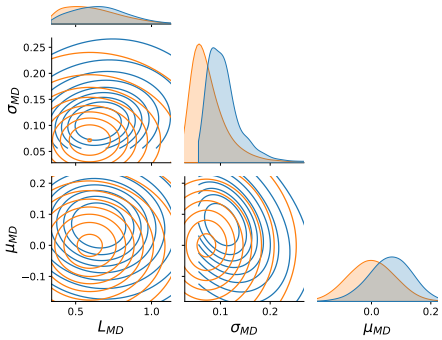

Figure 19: Higdon. Prior and posterior approximations

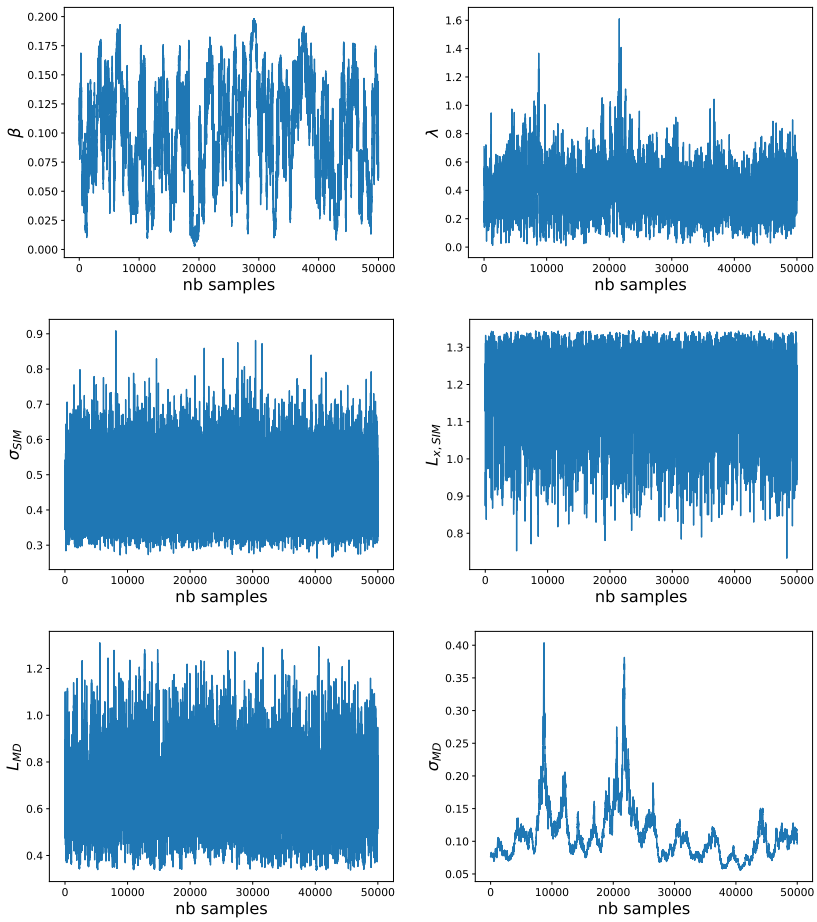

Figure 20: Higdon. MCMC. The sample chains mix well. The reconstructed posterior seems to be converged. The posterior for the amplitude hyper–parameter is clearly not Gaussian, and is skewed towards the low side

We now discuss the parameter calibrations for Tanner's model within the KOH and HIG approaches, which as just discussed are practically identical. The relaxation parameter λ within Tanner's model appears to be reasonably well-informed by the data with a significant shift from prior to posterior. This is not the case for the parameter β related to the solvent viscosity ratio where the data is apparently not informative. This is expected given the relative insensitivity of the swell with respect to the relaxation time in Tanner's model. If it is necessary to better identify β it would be prudent to additionally use the pressure drop data from the capillary rheometer, i.e. to perform a classical capillary rheometry test, in conjunction with a second equation linking β with the pressure drop. We discuss this possibility further in the conclusions.

With regards to the hyper-parameter calibrations in the Gaussian Process models

Table 5: Prior PDF specification and SI units (in square brackets) of the parameters. The SI unit of the shear rate $\dot{\gamma}_{\mathrm{w}}$ is seconds [s]. Ticks / crosses indicate if the parameter appears in the formulation of the model (yes / no)

Parameters	Distribution	Hyper-parameters			SIM	ML	KOH	HIG
λ [s^{-1}]	Beta	scale 2.0	α 2.0	β 2.0	✓	X	✓	✓
β [$-$]	Beta	scale 0.2	α 2.0	β 2.0	✓	X	✓	✓
c_{GP} [$-$]	TruncatedNormal		μ 0.0 (1.5 ML)	σ 0.08 (0.06 ML)	X	✓	✓	✓
σ_{GP} [$-$]	InverseGamma		α 5.5	β 0.35	X	✓	✓	✓
L_{GP} [s]	TruncatedNormal		μ 0.61	σ 0.18	X	✓	✓	✓
$\tilde{c}_{\mathrm{sim}}$ [$-$]	TruncatedNormal		μ 1.8	σ 0.4	X	X	X	✓
$\tilde{\sigma}_{\mathrm{sim}}$ [$-$]	InverseGamma		α 3.0	β 0.9	X	X	X	✓
$\tilde{L}_{\mathrm{sim}}$ [$\mathrm{s}, -, \mathrm{s}^{-1}$]	TruncatedNormal		μ (0.6, 0.4, 0.8)	σ (0.2, 0.1, 0.2)	X	X	X	✓

in the KOH and HIG approaches we note that there is some information content in the data that causes a posterior shift. Nevertheless, as always, in a low-data regime such as this one and using data-centric models such as Gaussian Processes, care should always be taken that the prior assumptions are appropriate, again see [5] for an excellent discussion.

5. Conclusions

We have presented a Bayesian framework following the seminal works of Kennedy and O'Hagan [29] and Higdon et al. [23] for calibrating a simple model of die–swell using calibrated against data acquired from a laser measurement system connected to a capillary rheometer. By comparing the calibrated KOH and HIG models with more classical ML and SIM approaches we have shown that this approach could be a pragmatic tool for creating a predictive model linking process parameters such as shear rate and die swell. We wish to emphasise once more that our approach is not meant to supplant advances in either experimental mechanics, constitutive model development or computational mechanics, but can be seen as being complimentary in the sense that it combines data from simulations and experiments sources in a principled way with quantified uncertainty.

Before finishing, we briefly mention a few directions that we think are promising for future investigations. With respect to obtaining data to identify the parameter β, it

Table 6: Summary statistics from the posterior. Crosses indicate that the parameter does not appear in the formulation of the model

	SIM		ML		KOH		HIG	
	mean	stddev	mean	stddev	mean	stddev	mean	stddev
λ	0.58	0.05	X	X	0.4	0.15	0.4	0.15
β	0.1	0.04	X	X	0.1	0.04	0.1	0.04
c_{GP}	X	X	1.49	0.04	0.07	0.08	0.06	0.07
σ_{GP}	X	X	0.07	0.03	0.1	0.04	0.11	0.04
L_{GP}	X	X	0.65	0.18	0.66	0.18	0.66	0.18
$\tilde{c}_{\mathrm{sim}}$	X	X	X	X	X	X	1.77	0.2
$\tilde{\sigma}_{\mathrm{sim}}$	X	X	X	X	X	X	0.45	0.07
$\tilde{L}_{\mathrm{sim}}$	X	X	X	X	X	X	(1.2, 0.74, 0.93)	(0.1, 0.05, 0.21)

would be interesting to look models using a HIG-type approach with multiple outputs (e.g. pressure drop and swell). This would be a relatively straightforward extension of our work, the only missing piece being a simple model linking apparent shear rate and pressure drop. This is where a numerical simulation of the full flow field would be valuable, as pressure drop can be extracted relatively easily from a numerical simulator.

In this study we used classical Gaussian Process models for the discrepancy term in both the KOH and HIG models. This means that samples of the swell ratio from these models are not monotonically increasing in apparent shear rate, an assumption that computer simulations of die swell, including the Tanner model, do tend to satisfy. With the basic principle of incorporating the best possible prior assumptions into a model (e.g. monotonically increasing die swell), it may be interesting to look at performing KOH-type calibration with GP models that can incorporate such constraints, see e.g. for a recent survey [49].

More broadly in terms of expanding the practical applications of our work, it would be to look at the KOH calibration approach in the context of the 'digital twin' concept, where a model of a physical system is continuously updated as new data is acquired. This would more closely meet the requirements of performing calibration on the factory floor. Aside from the usual practical issues, from a computational point of view he KOH methodology is known to scale at $\mathcal{O}(N^3)$ in the number of data and prediction points N due to the repeated Cholesky decomposition of a large dense covariance matrix. Furthermore, it is well-known that even efficient Markov Chain methods such as NUTS often still require thousands to tens of thousands of posterior evaluations for converged statistics. For example, A recent paper proposes a KOH approach using particle filters [59], instead of Markov Chain methods, that seems to offer model calibration times compatible with the real-time restrictions of digital twins.

6. Backmatter

Abbreviations.

BI - Bayesian Inference.

GP - Gaussian Process.

GPU - Graphics Processing Unit.

HIG - Higdon.

KOH - Kennedy O'Hagan.

NUTS - No-U-Turn sampler.

MD - Model Discrepancy.

ML - Machine Learning.

PCE - Polynomial Chaos Expansion.

PDF - Probability Density Function.

PVC - Polyvinyl Chloride.

SBR - Styrene-Butadiene Rubber.

SIM - Simulation.

TF - TensorFlow.

TFP - TensorFlow Probability.

Wi - Weissenberg number.

XLA - Accelerated Linear Algebra.

Data availability statement. The software and data that support the findings of this study are openly available in Zenodo at `https://doi.org/10.5281/zenodo.10600411` [21].

Author Contributions. Conceptualisation: JB, JSH, AG, JH, NV, CS. Data Curation: JH, AG. Funding acquisition: JB, JSH, NV. Investigation: JH, AG. Software: JH, JSH. Supervision: JB, JSH. Methodology: JH, AG, JB, JSH. Project administration: JB. Writing - original draft: JH, AG, JB, JSH. Writing - review and editing: JH, AG, JSH, JB, NV, CS. All authors approved the final submitted draft.

Funding. This work was funded in whole, or in part, by the Luxembourg National Research Fund (FNR) grant reference EMDD 14263566. For the purposes of open access, and in fulfilment of the obligations arising from the grant agreement, the authors have applied a Creative Commons Attribution 4.0 International (CC BY 4.0) license to any Author Accepted Manuscript version arising from this submission.

Competing interests. Author JSH had a family member working at the Goodyear Innovation Center Luxembourg during the period of the project. The family member was not involved in the implementation of the project.

Authors JH, AG, CS, NV and JB declare no competing interests.

References

[1] Balestriero, R., Pesenti, J., LeCun, Y., 2021. Learning in High Dimension Always Amounts to Extrapolation. arXiv:2110.09485 [cs] URL: http://arxiv.org/abs/2110.09485. arXiv: 2110.09485.

[2] Bayarri, M.J., Berger, J.O., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J., Lin, C.H., Tu, J., 2007. A Framework for Validation of Computer Models. Technometrics 49, 138–154. URL: https://www.jstor.org/stable/25471307. publisher: [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality].

[3] Behzadfar, E., Ansari, M., Konaganti, V.K., Hatzikiriakos, S.G., 2015. Extrudate swell of HDPE melts: I. Experimental. Journal of Non-Newtonian Fluid Mechanics 225, 86–93. URL: https://www.sciencedirect.com/science/article/pii/S0377025715001342, doi:10.1016/j.jnnfm.2015.07.008.

[4] Brooks, S., Gelman, A., Jones, G., Meng, X.L. (Eds.), 2011. Handbook of Markov Chain Monte Carlo. Chapman and Hall/CRC. URL: https://doi.org/10.1201%2Fb10905, doi:10.1201/b10905.

[5] Brynjarsdóttir, J., O'Hagan, A., 2014. Learning about physical parameters: the importance of model discrepancy. Inverse Problems 30, 114007. URL: https://doi.org/10.1088/0266-5611/30/11/114007, doi:10.1088/0266-5611/30/11/114007. publisher: IOP Publishing.

[6] Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L., Zdeborová, L., 2019. Machine learning and the physical sciences. Reviews of Modern Physics 91, 045002. URL: https://link.aps.org/doi/10.1103/RevModPhys.91.045002, doi:10.1103/RevModPhys.91.045002. publisher: American Physical Society.

[7] Chong, A., Menberg, K., 2018. Guidelines for the Bayesian calibration of building energy models. Energy and Buildings 174, 527–547. URL: https://www.sciencedirect.com/science/article/pii/S0378778818307539, doi:10.1016/j.enbuild.2018.06.028.

[8] Chong, A., Xu, W., Chao, S., Ngo, N.T., 2019. Continuous-time Bayesian calibration of energy models using BIM and energy data. Energy and Buildings 194, 177–190. URL: https://www.sciencedirect.com/science/article/pii/S0378778818339094, doi:10.1016/j.enbuild.2019.04.017.

[9] Claus, S., Phillips, T.N., 2013. Viscoelastic flow around a confined cylinder using spectral/hp element methods. Journal of Non-Newtonian Fluid Mechanics 200, 131–146. URL: https://www.sciencedirect.com/science/article/pii/S0377025713000785, doi:10.1016/j.jnnfm.2013.03.004.

[10] Comminal, R., Pimenta, F., Hattel, J.H., Alves, M.A., Spangenberg, J., 2018. Numerical simulation of the planar extrudate swell of pseudoplastic and viscoelastic fluids with the streamfunction and the VOF methods. Journal of Non-Newtonian Fluid Mechanics 252, 1–18. URL: https://www.sciencedirect.com/science/article/pii/S0377025717304883, doi:10.1016/j.jnnfm.2017.12.005.

[11] Crochet, M.J., Keunings, R., 1982. Finite element analysis of die swell of a highly elastic fluid. Journal of Non-Newtonian Fluid Mechanics 10, 339–356. URL: https://www.sciencedirect.com/science/article/pii/0377025782800086, doi:10.1016/0377-0257(82)80008-6.

[12] Delyon, B., Lavielle, M., Moulines, E., 1999. Convergence of a stochastic approximation version of the EM algorithm. The Annals of Statistics 27, 94 – 128. URL: https://doi.org/10.1214/aos/1018031103, doi:10.1214/aos/1018031103.

[13] Developers, T., 2022. TensorFlow. URL: https://zenodo.org/record/6476456, doi:10.5281/zenodo.6476456.

[14] Dillon, J.V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M., Saurous, R.A., 2017. TensorFlow Distributions. arXiv:1711.10604 [cs, stat] URL: http://arxiv.org/abs/1711.10604. arXiv: 1711.10604.

[15] Fattal, R., Kupferman, R., 2004. Constitutive laws for the matrix-logarithm of the conformation tensor. Journal of Non-Newtonian Fluid Mechanics 123, 281–285. URL: http://www.sciencedirect.com/science/article/pii/S0377025704002630, doi:10.1016/j.jnnfm.2004.08.008.

[16] Ferry, J.D., 1980. Viscoelastic Properties of Polymers. John Wiley & Sons. Google-Books-ID: 9dqQY3Ujsx4C.

[17] Gansen, A., Rehor, M., Sill, C., Polinska, P., Westermann, S., Dheur, J., Hale, J.S., Baller, J., 2020. Investigation of the sharkskin melt instability using optical fourier analysis. Journal of Applied Polymer Science 137. URL: <GotoISI>://WOS:000503961000001, doi:10.1002/app.48806.

[18] Girolami, M., Calderhead, B., 2011. Riemann manifold langevin and hamiltonian monte carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73, 123–214. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2010.00765.x, doi:https://doi.org/10.1111/j.1467-9868.2010.00765.x, arXiv:https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2010.00

[19] Habla, F., Marschall, H., Hinrichsen, O., Dietsche, L., Jasak, H., Favero, J.L., 2011. Numerical simulation of viscoelastic two-phase flows using openFOAM®. Chemical Engineering Science 66, 5487–5496. URL: https://www.sciencedirect.com/science/article/pii/S0009250911005422, doi:10.1016/j.ces.2011.06.076.

[20] Hale, J.S., Li, L., Richardson, C.N., Wells, G.N., 2017. Containers for Portable, Productive, and Performant Scientific Computing. Computing in Science & Engineering 19, 40–50. URL: http://ieeexplore.ieee.org/document/7933304/, doi:10.1109/MCSE.2017.2421459.

[21] Hennicker, J., Hale, J.S., 2024. Supplementary Material: Bayesian calibration of a model of polymer die swell using data from a laser–based measurement system. URL: https://doi.org/10.5281/zenodo.10600411, doi:10.5281/zenodo.10600411.

[22] Higdon, D., Gattiker, J., Williams, B., Rightley, M., 2008a. Computer Model Calibration Using High-Dimensional Output. Journal of the American Statistical Association 103, 570–583. URL: https://www.jstor.org/stable/27640080. publisher: [American Statistical Association, Taylor & Francis, Ltd.].

[23] Higdon, D., Kennedy, M., Cavendish, J.C., Cafeo, J.A., Ryne, R.D., 2004. Combining Field Data and Computer Simulations for Calibration and Prediction. SIAM Journal on Scientific Computing 26, 448–466. URL: https://epubs.siam.org/doi/10.1137/S1064827503426693, doi:10.1137/S1064827503426693. publisher: Society for Industrial and Applied Mathematics.

[24] Higdon, D., Nakhleh, C., Gattiker, J., Williams, B., 2008b. A Bayesian calibration approach to the thermal problem. Computer Methods in Applied Mechanics and Engineering 197, 2431–2441. URL: https://www.sciencedirect.com/science/article/pii/S0045782507005087, doi:10.1016/j.cma.2007.05.031.

[25] Hoffman, M.D., Gelman, A., 2014. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. J. Mach. Learn. Res. 15, 1593–1623.

[26] Hwang, W.R., Walkley, M.A., Harlen, O.G., 2011. A fast and efficient iterative scheme for viscoelastic flow simulations with the DEVSS finite element method. Journal of Non-Newtonian Fluid Mechanics 166, 354–362. URL: http://www.sciencedirect.com/science/article/pii/S0377025711000206, doi:10.1016/j.jnnfm.2011.01.003.

[27] Jay, P., Piau, J.M., El Kissi, N., Cizeron, J., 1998. The reduction of viscous extrusion stresses and extrudate swell computation using slippery exit surfaces. Journal of Non-Newtonian Fluid Mechanics 79, 599–617. URL: http://www.sciencedirect.com/science/article/pii/S0377025798001219, doi:10.1016/S0377-0257(98)00121-9.

[28] John, D.N., Schick, M., Heuveline, V., 2019. Learning Model Discrepancy of an Electric Motor with Bayesian Inference. Proceedings 33, 11. URL: https://www.mdpi.com/2504-3900/33/1/11, doi:10.3390/proceedings2019033011. number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[29] Kennedy, M.C., O'Hagan, A., 2001. Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63, 425–464. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00294, doi:https://doi.org/10.1111/1467-9868.00294.

[30] Keunings, R., 1986. On the high Weissenberg number problem. Journal of Non-Newtonian Fluid Mechanics 20, 209–226. URL: https://www.sciencedirect.com/science/article/pii/0377025786800222, doi:10.1016/0377-0257(86)80022-2.

[31] Lei, C.L., Ghosh, S., Whittaker, D.G., Aboelkassem, Y., Beattie, K.A., Cantwell, C.D., Delhaas, T., Houston, C., Novaes, G.M., Panfilov, A.V., Pathmanathan, P., Riabiz, M., dos Santos, R.W., Walmsley, J., Worden, K., Mirams, G.R., Wilkinson, R.D., 2020. Considering discrepancy when calibrating a mechanistic electrophysiology model. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 378, 20190349. URL: https://royalsocietypublishing.org/doi/10.1098/rsta.2019.0349, doi:10.1098/rsta.2019.0349. publisher: Royal Society.

[32] Macosko, C.W., 1994. Rheology. Principles, Measurements and Applications. VCH, New York.

[33] Málek, J., Rajagopal, K., Tůma, K., 2015. A thermodynamically compatible model for describing the response of asphalt binders. International Journal of Pavement Engineering 16, 297–314. URL: https://doi.org/10.1080/10298436.2014.942860, doi:10.1080/10298436.2014.942860. publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10298436.2014.942860.

[34] Na, Y., Yoo, J., 1991. A finite volume technique to simulate the flow of a viscoelastic fluid. Computational Mechanics 8, 43–55. doi:https://doi.org/10.1007/BF00370547.

[35] Nesterov, Y., 2009. Primal-dual subgradient methods for convex problems. Math. Program. 120, 221–259. doi:https://doi.org/10.1007/s10107-007-0149-x.

[36] Nithi-Uthai, N., Manas-Zloczower, I., 2003. Numerical simulation of sharkskin phenomena in polymer melts. Applied Rheology 13, 79–86.

[37] Parno, M.D., Marzouk, Y.M., 2018. Transport map accelerated markov chain monte carlo. SIAM/ASA Journal on Uncertainty Quantification 6, 645–682. URL: https://doi.org/10.1137%2F17m1134640, doi:10.1137/17m1134640.

[38] Peters, E.L., Balin, R., Jansen, K.E., Doostan, A., Evans, J.A., 2022. S-frame discrepancy correction models for data-informed Reynolds stress closure. Journal of Computational Physics 448, 110717. URL: https://www.sciencedirect.com/science/article/pii/S0021999121006124, doi:10.1016/j.jcp.2021.110717.

[39] Pimenta, F., Alves, M.A., 2017. Stabilization of an open-source finite-volume solver for viscoelastic fluid flows. Journal of Non-Newtonian Fluid Mechanics 239, 85–104. URL: https://www.sciencedirect.com/science/article/pii/S0377025716303329, doi:10.1016/j.jnnfm.2016.12.002.

[40] Pimenta, F., Alves, M.A., 2021. Conjugate heat transfer in the unbounded flow of a viscoelastic fluid past a sphere. International Journal of Heat and Fluid Flow 89, 108784. URL: https://www.sciencedirect.com/science/article/pii/S0142727X2100014X, doi:10.1016/j.ijheatfluidflow.2021.108784.

[41] Plumlee, M., Joseph, V.R., 2016. Orthogonal Gaussian process models. arXiv:1611.00203 [stat] URL: http://arxiv.org/abs/1611.00203. arXiv: 1611.00203.

[42] Rajagopal, K.R., Srinivasa, A.R., 2000. A thermodynamic frame work for rate type fluid models. Journal of Non-Newtonian Fluid Mechanics 88, 207–227. URL: http://www.sciencedirect.com/science/article/pii/S0377025799000233, doi:10.1016/S0377-0257(99)00023-3.

[43] Rappel, H., Beex, L.A.A., Bordas, S.P.A., 2018. Bayesian inference to identify parameters in viscoelasticity. Mechanics of Time-Dependent Materials 22, 221–258. URL: https://doi.org/10.1007/s11043-017-9361-0, doi:10.1007/s11043-017-9361-0.

[44] Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. The MIT Press.

[45] Rauwendaal, C., 2014. Polymer Extrusion. Carl Hanser Verlag GmbH Co KG. Google-Books-ID: DKuQAgAAQBAJ.

[46] Russo, G., Phillips, T., 2011. Spectral element predictions of die-swell for oldroyd-b fluids. Computers & Fluids 43, 107–118. URL: https://www.sciencedirect.com/science/article/pii/S0045793010002562, doi:https://doi.org/10.1016/j.compfluid.2010.09.028. symposium on High Accuracy Flow Simulations. Special Issue Dedicated to Prof. Michel Deville.

[47] Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P., 1989. Design and analysis of computer experiments (with discussion). volume 4.

[48] Salvatier, J., Wiecki, T., Fonnesbeck, C., 2016. Probabilistic programming in python using pymc3. Peer J. Computer Science 2:e55. doi:https://doi.org/10.7717/peerj-cs.55.

[49] Swiler, L.P., Gulian, M., Frankel, A.L., Safta, C., Jakeman, J.D., 2020. A survey of constrained gaussian process regression: Approaches and implementation challenges. Journal of Machine Learning for Modeling and Computing 1. URL: https://www.dl.begellhouse.com/journals/558048804a15188a,2cbcbe11139f18e5,0776649265326db4.html, doi:10.1615/JMachLearnModelComput.2020035155. publisher: Begel House Inc.

[50] Tang, D., Marchesini, F.H., Cardon, L., D'hooge, D.R., 2020. State of the-Art for Extrudate Swell of Molten Polymers: From Fundamental Understanding at Molecular Scale toward Optimal Die Design at Final Product Scale. Macromolecular Materials and Engineering 305, 2000340. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/mame.202000340, doi:10.1002/mame.202000340. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mame.202000340.

[51] Tang, D., Marchesini, F.H., Cardon, L., D'hooge, D.R., 2019. Three-dimensional flow simulations for polymer extrudate swell out of slit dies from low to high aspect ratios. Physics of Fluids 31, 093103. URL: https://aip.scitation.org/doi/10.1063/1.5116850, doi:10.1063/1.5116850. publisher: American Institute of Physics.

[52] Tanner, R.I., 1970. A theory of die-swell. Journal of Polymer Science Part A-2: Polymer Physics 8, 2067–2078. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pol.1970.160081203, doi:https://doi.org/10.1002/pol.1970.160081203.

[53] Tanner, R.I., 2005. A theory of die-swell revisited. Journal of Non-Newtonian Fluid Mechanics 129, 85–87. URL: https://www.sciencedirect.com/science/article/pii/S0377025705001618, doi:https://doi.org/10.1016/j.jnnfm.2005.05.010.

[54] Team, S.D., 2019. Stan Modeling Language Users Guide and Reference Manual. URL: https://mc-stan.org.

[55] Tomé, M.F., Mangiavacchi, N., Cuminato, J.A., Castelo, A., McKee, S., 2002. A finite difference technique for simulating unsteady viscoelastic free surface flows. Journal of Non-Newtonian Fluid Mechanics 106, 61–106. URL: https://www.sciencedirect.com/science/article/pii/S0377025702000642, doi:10.1016/S0377-0257(02)00064-2.

[56] Tuo, R., Jeff Wu, C.F., 2016. A Theoretical Framework for Calibration in Computer Models: Parametrization, Estimation and Convergence Properties. SIAM/ASA Journal on Uncertainty Quantification 4, 767–795. URL: https://epubs.siam.org/doi/10.1137/151005841, doi:10.1137/151005841. publisher: Society for Industrial and Applied Mathematics.

[57] Tuo, R., Wu, C.F.J., 2015. Efficient calibration for imperfect computer models. The Annals of Statistics 43, 2331–2352. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-43/issue-6/Efficient-calibration-for-imperfect-computer-models/10.1214/15-AOS1314.full, doi:10.1214/15-AOS1314. publisher: Institute of Mathematical Statistics.

[58] Vlachopoulos, J., Horie, M., Lidorikis, S., 1972. An evaluation of expressions predicting die swell. Transactions of the Society of Rheology 16, 669–685. doi:10.1122/1.549269.

[59] Ward, R., Choudhary, R., Gregory, A., Jans-Singh, M., Girolami, M., 2021. Continuous calibration of a digital twin: Comparison of particle filter and Bayesian calibration approaches. Data-Centric Engineering 2. URL: https://www.cambridge.org/core/journals/data-centric-engineering/article/continuous-calibration-of-a-digital-twin-comparison-of-particle-filter-and-b
BE7304EFCB5006059184852E39EED19A#article, doi:10.1017/dce.2021.12. publisher: Cambridge University Press.

[60] Xiu, D., Karniadakis, G.E., 2002. The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations. SIAM Journal on Scientific Computing 24, 619–644. URL: https://epubs.siam.org/doi/10.1137/S1064827501387826, doi:10.1137/S1064827501387826. publisher: Society for Industrial and Applied Mathematics.

[61] Řehoř, M., Gansen, A., Sill, C., Polińska, P., Westermann, S., Dheur, J., Baller, J., Hale, J.S., 2020. A comparison of constitutive models for describing the flow of uncured styrene-butadiene rubber. Journal of Non-Newtonian Fluid Mechanics 286, 104398. URL: http://www.sciencedirect.com/science/article/pii/S0377025720301555, doi:10.1016/j.jnnfm.2020.104398.