

A BAYESIAN GENERATIVE MODEL WITH GAUSSIAN PROCESS PRIORS FOR THERMOMECHANICAL ANALYSIS OF MICRO-RESONATORS

Maximillian F. Vording^{†*}✉, Peter O. Okeyo^{‡*}, Juan J. R. Guillamón^{†*},
Peter E. Larsen^{*}, Mikkel N. Schmidt[†], and Tommy S. Alstrøm^{†*}

[†] DTU Compute, Technical University of Denmark

[‡] Dep. of Pharmacy, University of Copenhagen

ABSTRACT

Thermal analysis using resonating micro-electromechanical systems shows great promise in characterizing materials in the early stages of research. Through thermal cycles and actuation using a piezoelectric speaker, the resonant behaviour of a model drug, theophylline monohydrate, is measured across the surface whilst using a laser-Doppler vibrometer for readout. Acquired is a sequence of spectra that are strongly correlated in time, temperature and spatial location of the readout. Traditionally, each spectrum is analyzed individually to locate the resonance peak. We propose a Bayesian model using a warped Gaussian process prior taking the correlations into account and demonstrate on both synthetic and experimental data, that it yields better estimates of both location and amplitude of the resonance peak. Thus, the proposed model can give a more precise characterization of drugs, which is important in drug discovery and development.

Index Terms— Bayesian learning and modeling, Gaussian processes, drug characterisation, thermomechanical analysis

1. INTRODUCTION

In the pharmaceutical industry, the majority of drugs are solid dosage forms and when they are exposed to varying environmental conditions such as temperature and humidity this influences their physiochemical and mechanical properties. In particular, drugs that form hydrates (water-solid interactions) can have very different properties in comparison to their anhydrous counterparts [1]. The current standard thermal methods that are used for drug characterisation include differential

scanning calorimetry (DSC) and thermogravimetric analysis (TGA), which are limited in sensitivity and require a minimum of a few milligrams for analysis [2]. In drug discovery, where the amount of drug can be limited due to impurities after synthesis, methods that are highly sensitive are needed. Due to these limitations, micro-resonators are being investigated [3, 4] for conducting thermal analysis of drugs due to their high sensitivity and low sample mass for analysis [5].

Micro-electromechanical systems (MEMS) have been the subject of intense research for decades. Due to their high responsiveness, high quality factors, high self-referencing abilities and fast response times, resonant MEMS have shown promise as sensors in many different fields of research [5, 6].

Standard techniques using phase-locked loop (PLL) is not suitable for tracking resonance peaks with low signal-to-noise-ratio (SNR) and fluctuations in resonance frequency (Ω) and quality factor (Q) during the phase transformation of theophylline monohydrate (TP MH) [7]. As an alternative, this paper introduces a Bayesian generative model with Gaussian process priors for estimating the physical parameters in the function governing the system behaviour. We test this model on thermomechanical experiments performed on micro-resonators, with the system characterized by the magnitude of the frequency response.

Fig. 1) shows the experimental setup, which uses a piezoelectric speaker as an actuator for the characterisation of TP MH during a thermal cycle, and a laser Doppler vibrometer (LDV) is used as an optical readout. A scanning grid is defined on TP MH, then its vibrational mode of vibration is identified. After this, TP MH is heated (25-90 °C, 5 °C/min) using a Linkam hotstage, while continuously obtaining spectra. The data analysis is performed by manually tracking Ω and Q . The obtained changes in the frequency response of the system relates to the thermal changes by an unknown function characterising the properties of the material itself.

The proposed model relates the underlying structure in spatial, temporal, and thermal covariates to the governing physical parameters through warped Gaussian process (GP) priors [8–10]. The warped GP priors are used to express information about the smoothness of the physical parameters with respect to the covariates, as well as prior constraints on

[†] Department of Applied Mathematics and Computer Science, Technical University of Denmark, DTU, Richard Petersens Plads 321, 2800 Kgs. Lyngby, Denmark

^{*} The Danish national research foundation and Villum foundations center for intelligent drug delivery and sensing using microcontainers and nanomechanics (IDUN), Department of Health technology, Technical University of Denmark, Ørsted Plads, 2800 Kgs Lyngby, Denmark

[‡] Department of Pharmacy, University of Copenhagen, Universitetsparken 2, 2100, Copenhagen, Denmark.

✉ Corresponding author e-mail: maxvo@dtu.dk

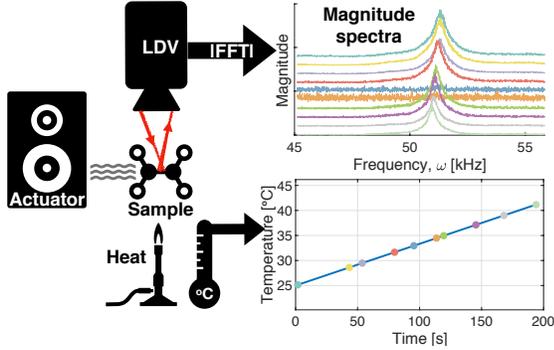


Fig. 1. Schematic illustration of a typical experimental setup measuring a resonators (samples) velocity response to the actuators (left) vibration frequency with a laser Doppler vibrometer (LDV upper/central) transformed by the fast Fourier transform (FFT) into a magnitude spectrum (upper right plot) at every new temperature (lower right plot) reached by the heat source (lower).

their possible values, allowing consistent tracking of peaks like [11] attempts for NMR spectra varying over time. For an ideal and clean driven damped resonator, the governing function can be described by second order differential equation [12]. As this is a simplification of the actual system, including informative prior constraints on the parameters will help solving the problem and allow leaving unrelated noise signals and other modes of variation in the residuals.

2. METHODS

2.1. Generative model

The observed data can be recorded in the form of a matrix $\mathbf{Z} \in \mathbb{R}^{N \times M}$, where the i 'th row corresponds to a measured spectrum and the j 'th column to a frequency. The frequencies are equally spaced within some predefined range $\omega_1, \dots, \omega_M$. In addition to this, for each measurement the temperature t_i , and spatial location of the measurement position x_i, y_i is also recorded.

We model the observed spectra using a resonance peak model $f(\omega, \theta)$ corrupted by additive noise ϵ ,

$$z_{i,j} = f(\omega_j, \theta_i) + \epsilon_{i,j}. \quad (1)$$

The peak model depends on the frequency ω as well as a set of parameters θ which govern the location, amplitude, and shape of the resonance peak. The additive noise is assumed independent and identically distributed over observations and frequencies. With a Gaussian noise model, the likelihood can be written as

$$p(\mathbf{Z}|\Theta) = \prod_{i=1}^N \prod_{j=1}^M \mathcal{N}(z_{i,j}; f(\omega_j, \theta_i), \sigma_\epsilon^2), \quad (2)$$

where the parameters of the peak model for each observation are collected in $\Theta = \{\theta_1, \dots, \theta_N\}$, and σ_ϵ^2 denotes the variance of the noise.

We assume that the resonance peak parameters are correlated, such that the location, amplitude and shape of resonance peaks are similar for observations with similar temperature and spatial location. We formulate this correlation through a prior on the parameters $p(\Theta)$. Combining this prior with the likelihood, we estimate the model parameters by maximizing the posterior distribution $p(\Theta|\mathbf{Z})$, computing the maximum a posteriori (MAP) estimate,

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}} p(\mathbf{Z}|\Theta)p(\Theta). \quad (3)$$

2.2. Resonance peak model

In the microresonator experiments, the governing function is derived from the solution, $z(t) = z_0 \exp\{i\omega t\}$, to the ordinary differential equation that governs a driven damped vibration on a linear resonator [12]. We use the magnitude as our frequency response function,

$$f(\omega_j, \theta_i) = \frac{F_i \Omega_i^2}{\sqrt{Q_i^2 (\Omega_i^2 - \omega_j^2)^2 + \Omega_i^2 \omega_j^2}}. \quad (4)$$

The parameters of this model are $\theta_i = \{F_i, \Omega_i, Q_i\}$, where F_i is the amplitude, Ω_i is the resonance frequency, and Q_i is the quality factor which describes the shape of the peak.

2.3. Prior distributions

For the amplitudes $\mathbf{F} = \{F_1, \dots, F_N\}$, resonance frequencies $\mathbf{\Omega} = \{\Omega_1, \dots, \Omega_N\}$, and quality factors $\mathbf{Q} = \{Q_1, \dots, Q_N\}$ we use independent prior distributions of the same form, namely warped Gaussian process (GP) priors. For each of the three parameter sequences, the GP allows us to specify a covariance function that defines the degree of correlation between the observations dependent on the observation index, temperature, and spatial coordinates. Specifically we use a radial basis function (RBF) for covariance on the following form

$$c(\phi, \phi') = \exp\left(-\frac{\Delta_i^2}{2\lambda_i^2} - \frac{\Delta_t^2}{2\lambda_t^2} - \frac{\Delta_x^2}{2\lambda_x^2} - \frac{\Delta_y^2}{2\lambda_y^2}\right) \quad (5)$$

where $\phi = \{i, t, x, y\}$ are the covariates, and $\Delta_i = (i - i')$ is shorthand for the difference between i and i' (and similarly for t, x , and y). The signal variance is here $\sigma_f^2 = 1$ and the noise variance term is omitted, since variance is introduced, when mapping the GP to the prior distribution. The length scale parameters $\lambda_i, \lambda_t, \lambda_x$, and λ_y control the degree of correlation with respect to the observation index or time, temperature, and spatial covariates respectively. The covariance function thus describes the covariance between parameters at

two different observation indices i and i' , and the covariance between each pair of observations can be collected in a matrix $\Sigma \in \mathbb{R}^{N \times N}$.

Since the resonance peak parameters F , Ω , and Q are non-negative by definition, an unconstrained Gaussian process is not directly suited as a prior distribution. To handle this constraint and provide a large degree of flexibility in the prior specification, we use a warped Gaussian process, in which the parameter we model is related to the GP prior through a parameterised link function. As the treatment is identical for the three model parameters, in the following we will use \mathbf{H} to denote one of the three parameters. We thus have the relation that $\ell(\mathbf{H}) = \mathbf{h}$ is distributed according to a GP with zero mean and covariance as described above. Here $\ell(\cdot)$ is a strictly increasing link function which is chosen such that its inverse $\ell^{-1}(\cdot)$ maps the Normal distribution onto a suitable truncated Normal distribution on the non-negative real numbers.

In particular we use the following inverse link function [9]

$$\ell^{-1}(\mathbf{h}) = P_H^{-1}(P_h(\mathbf{h})), \quad (6)$$

where $P_h(\cdot)$ is the cumulative distribution function (CDF) of the standard Gaussian distribution [13]

$$P_h(h_i) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{h_i}{\sqrt{2}\sigma_i} \right) \right), \quad (7)$$

where σ_i is the i th diagonal element of Σ_h . $P_H^{-1}(\cdot)$ is the inverse CDF of the truncated Normal distribution [13]

$$P_H^{-1}(\xi) = \mu + \sigma \sqrt{2} \operatorname{erf}^{-1} \left(\xi \operatorname{erf}(\beta) + (1 - \xi) \operatorname{erf}(\alpha) \right), \quad (8)$$

where $\beta = \frac{b-\mu}{\sigma\sqrt{2}}$ and $\alpha = \frac{a-\mu}{\sigma\sqrt{2}}$. Since the marginal distributions of \mathbf{h} are standard Normal, the standard Normal CDF maps \mathbf{h} onto a distribution with uniform marginals, and from here the inverse CDF of the truncated Normal maps this onto a distribution with truncated Normal marginals. The truncated Normal distribution has parameters a and b which control the upper and lower truncation points as well as μ and σ which control the location and scale of the distribution. Thus, this allows us to define the covariance of \mathbf{H} through the GP model while enforcing desired marginal distributions through the link function. We use the change of variables trick, $\mathbf{h} = \mathbf{C}_h^T \boldsymbol{\eta}$, from [9], where \mathbf{C}_h is the matrix square root (Cholesky decomposition) of the GP covariance matrix, $\Sigma_h = \mathbf{C}_h \mathbf{C}_h^T$. This trick ensures a simplification of the optimisation in a space of standard i.i.d. $\boldsymbol{\eta}$.

2.4. Optimisation

The MAP estimates are computed through standard numerical gradient constrained optimisation using `fmincon` in MATLAB. To get a warm start we set our initial \mathbf{F} to the maximum magnitudes in each spectra, Ω to the corresponding frequencies and all Q to 300 based on the usual Q -values of

our resonator. The optimisation is run as a block coordinate ascent, where we maximize the log-posterior objective with respect to Ω and $\mathbf{F} \cup \mathbf{Q}$ separately in alternating order until convergence. Convergence thresholds are set with an objective function tolerance of 0.001 and a parameter step tolerance of 0.001, with a max of $5 \cdot 10^7$ function evaluations and 20 iterations per block.

3. EXPERIMENTS

All experiments are carried out using the model described above with the governing resonance peak function in Eq. 4. To assess the importance of modeling the correlation between observations, we test the model with either the radial basis function (RBF) for the kernel (referred to as *GPP w. RBF kernel*) or a diagonal kernel for the covariance in the GP (referred to as *GPP w. diag. kernel*). With a diagonal kernel, the model of covariance is effectively disabled, corresponding to a standard Bayesian peak fitting model run on each observation independently.

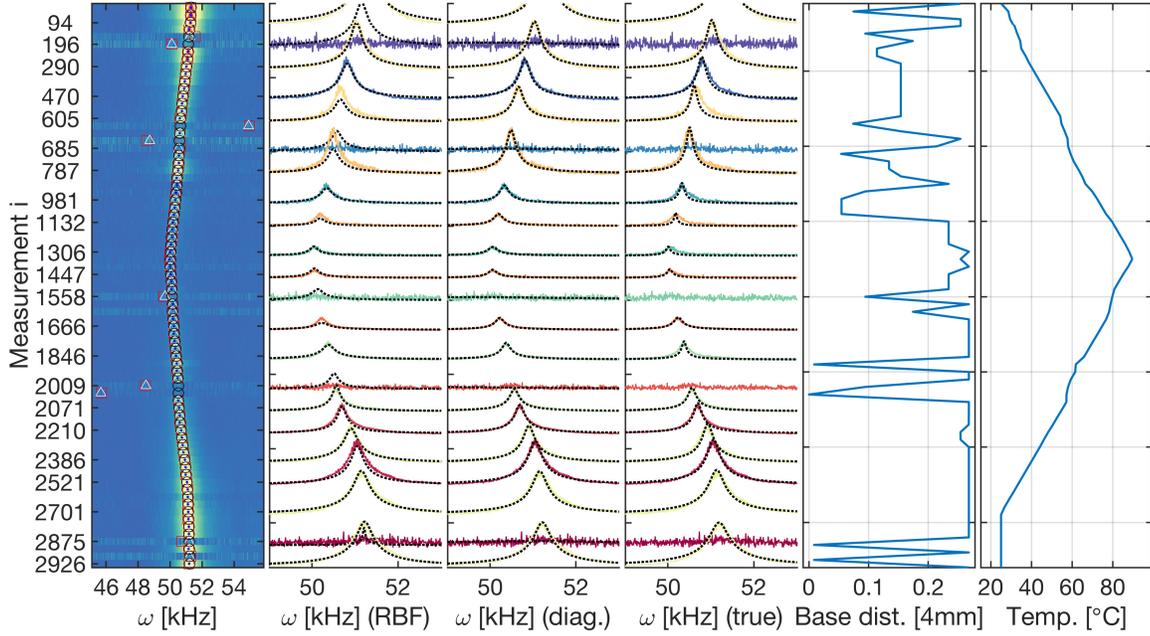
3.1. Micro-resonator

To demonstrate the model on realistic data under non-ideal conditions, we apply it to a data set from an experiment with a micro-resonator, that has been manually annotated by the experimenter, so we can validate using the annotations as ground truth.

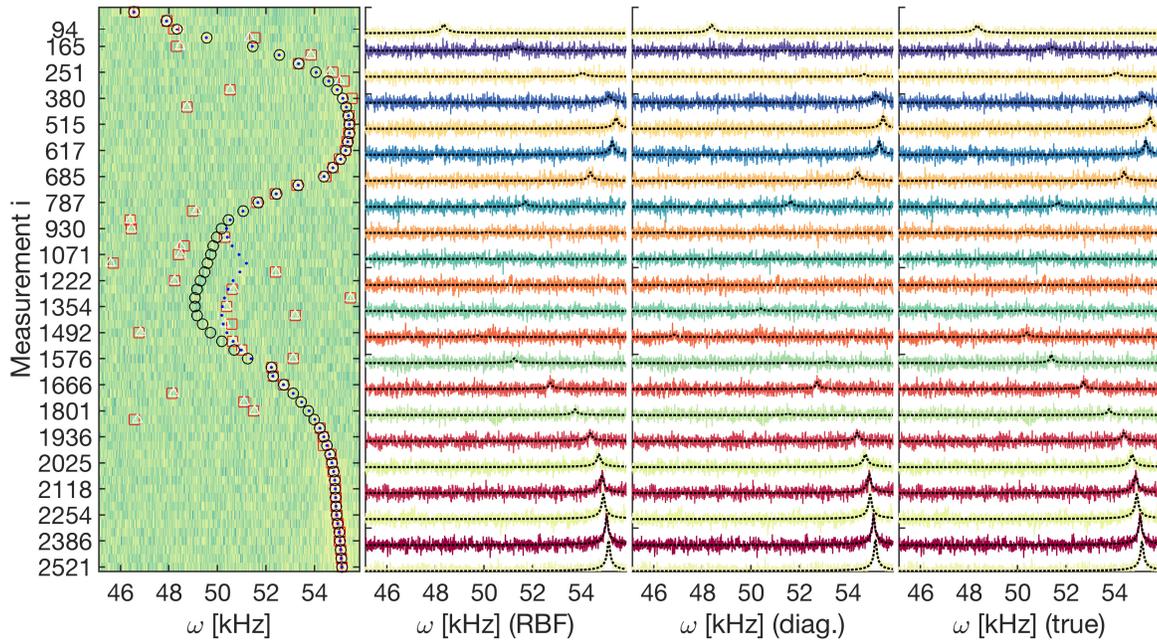
The data stems from an experiment using a laser-Doppler vibrometer (LDV) to measure the frequency response of a micro-resonator, which is shown in Fig. 2. Here the covariates are seen as the temperature profile and spatial coordinates of the measurement points on the surface of the resonator. While the temperature is turned up and down, the LDV scans across the surface of the resonator, here 67 times on a grid of 45 points leading to $N = 67 \times 45 = 3015$ measurements. After filtering out the measurements with low signal-to-noise-ratio (SNR) and poor conditions for the annotator, we end up with 65 spectra and corresponding parameters manually annotated and fitted by a domain expert among the authors. To test our model's ability to interpolate parameters in regions with low SNR, we reintroduce 11 of the left out spectra with lowest SNR, which has indices $i = [165, 196, 617, 680, 685, 1558, 1591, 2009, 2025, 2875, 2920]$. None of these has annotated parameters provided and are not used for comparison in Tab. 1.

3.2. Choice of prior parameters

The model specification and parameter values are decided upon given a mix of prior knowledge and statistical measures easily obtained from the data. The noise is modelled through the normal likelihood, with $\sigma_\epsilon = 2.969 \cdot 10^{-5}$ being the average standard deviation of spectral regions away



(a) Micro-resonator data



(b) Synthetic data

Fig. 2. (a) Data from real experiment with micro-resonator. (b) Data generated from model with RBF kernel. Both data sets use the same covariates, y , where the spatial (x and y) is represented by the euclidean distance to the attachment point of the resonator in (a) as the 2nd to rightmost plots. In the rightmost plot we show the temperature (t) growing linearly with the temporal (i) changes as seen in the leftmost labels. In the leftmost columns, the full spectrograms are presented with Ω estimates on top, using black circles for RBF kernel model, white triangles for diag. kernel model, red squares for initial Ω and blue dots for true parameters. A subset of spectra are stacked and aligned in the 3 plots to the right of these with resonance peak shapes (in black dotted lines) generated from Eq. 4 with the parameter estimates.

from peaks. For the length scales we chose values in the range of the covariates and for all 3 parameters (F, Ω, Q) they are the same $[\lambda_x, \lambda_y, \lambda_t, \lambda_i] = [1, 1, 15, 1000]$, except for Ω , which has $[\lambda_x, \lambda_y] = 10^{16}$ to avoid effects from the measurement position on the resonance frequency, which should not be correlated for this kind of resonator. For the truncated normal priors over Ω we use $\mu_\Omega = 50.647$ kHz based on the mean of our initial resonance frequencies, $\sigma_\Omega = 5.3672$ kHz, which is half width of our frequency range and $[a_\Omega, b_\Omega] = [45.217, 55.736]$ kHz, which is the frequency range with 1% margin. Similar for F we use the mean, minimum and maximum of the magnitude spectra to set $[\mu_F, \sigma_F, a_F, b_F] = [4.3951 \cdot 10^{-4}, 10^{-2}, 10^{-10}, 10^{-3}]$. Based on our prior knowledge for ranges in Q given this kind of resonator, we set the distribution parameters to $[\mu_Q, \sigma_Q, a_Q, b_Q] = [500, 100, 10, 1000]$.

3.3. Validation on synthetic data

To validate the inference procedure and demonstrate the benefit of the warped Gaussian process for modeling correlations in the resonance peak parameters in spectra with low SNR, we generate 10 synthetic data sets from the model using the same prior parameters and RBF kernel, but double the noise variance, $\sigma_\epsilon = 4.198 \cdot 10^{-5} = \sqrt{2} \cdot 2.969 \cdot 10^{-5}$. Results for one of these are shown in Fig. 2.b and the summarised error statistics for all in Tab. 1.

4. RESULTS AND DISCUSSION

In Fig. 2 we see how for both real and generated data sets, the GPP w. RBF kernel leads to MAP estimates of all parameters following the global shape of the spectra interpolating coherent peak shapes through all measurements, even when SNR is too low to spot the peak out of context. We also see how the model with a diagonal kernel tends to get stuck at local minima, near the initial starting points, which is not the case when using the RBF kernel. This suggest, that utilizing the underlying structure of the experiment given by the relation between the covariates in the experiment and parameters in the governing function, can improve the model performance. This is especially the case, when the SNR is low and or some spectra are missing, where the GPP exhibits interpolating behaviour.

In Fig. 3, we see how the model with RBF kernel gives estimates of Ω close to ground truth for both the synthetic and real data. Both models estimate Q parameters lower than the ground truth, but inspection of how well the peak shapes fit the spectra in Fig. 2 reveals how the annotated Q 's seems overestimated in comparison to our to model estimates.

For synthetic data we see in Table 1 how the model achieves MAP estimates closest to the ground truth Ω when

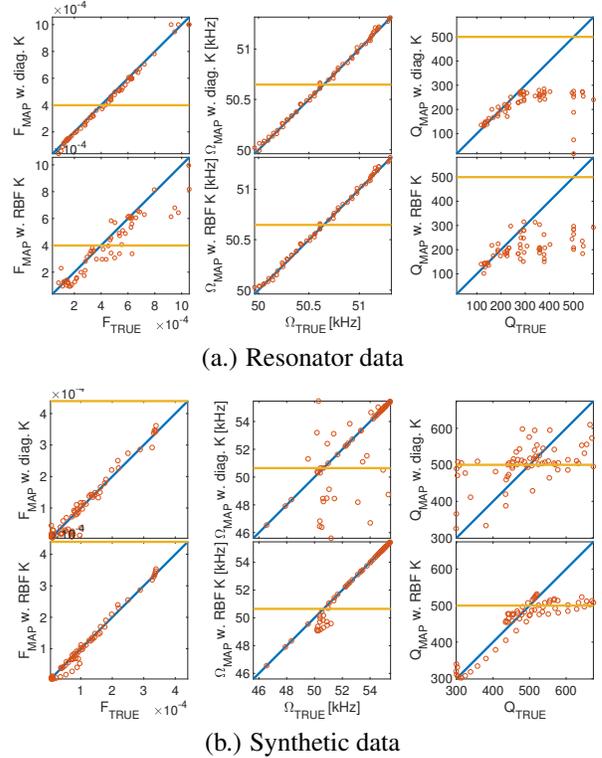


Fig. 3. (a.) Expert annotations (x-axis) plotted against MAP parameter estimates for the real data. (b.) The true generated parameters (x-axis) plotted against MAP parameter estimates for the generated data set seen in Fig. 2.b. A good estimator has points close to and along the blue diagonal. The yellow lines mark μ for each corresponding prior.

using the RBF kernel, but is better at estimating Q , when using the diagonal kernel.

As seen in Fig. 3, both models tend to over- or underestimate the Q parameters. This could be due to using a wrong likelihood for modeling real data, which is not necessarily shaped according to the ideal driven damped resonator used in these models. The background noise level seems to have a trend shifting the spectra up, which is not modelled by our likelihood and can for our constant noise model only be alleviated by lowering the Q . We wish to model the background noise as well with warped GP priors.

The quality factor, Q , is often hard to estimate using only the magnitude spectra. Other literature [5, 12] suggest fitting the slope of the phase at the peak location, to estimate the Q . Given that our model can estimate Ω very well, our model could work as a precursor for this in future work. Furthermore we will like to investigate the use of a Metropolis-Hastings algorithm and Gibbs sampling for uncertainty estimates on the parameters allowing for the design of an efficient user-annotation scheme and choice of prior parameters by Bayesian optimisation.

	Real data				Synthetic data			
	GPP w. RBF Σ		GPP w. diag. Σ		GPP w. RBF Σ		GPP w. diag. Σ	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
\hat{F} [10^{-5}]	9.701	6.080	2.339	1.736	1.002(331)	0.673(228)	1.123(394)	0.704(240)
$\hat{\Omega}$ [10^2]	0.213	0.154	0.209	0.146	9.743(3.596)	5.085(2.076)	23.54(7.88)	15.95(6.39)
\hat{Q} [10^2]	1.297	0.847	1.281	0.722	2.066(483)	1.451(395)	1.011(210)	0.686(136)

Table 1. For 10 synthetic data sets (right) and one real (left) the performance is reported in root mean squared error (RMSE) and mean absolute error (MAE between the ground truth and MAP estimates for the 65 spectra with high SNR. Standard errors are in parentheses and lowest non-overlapping errors in bold.

5. CONCLUSION

From experiments with our proposed Bayesian model, we can conclude, that the model is appropriate and efficient for generative modelling of the magnitude spectra of the frequency response of micro-resonators. Using a Bayesian formulation of the model, we can introduce regularisation towards expected parameters and enforce non-negativity through the truncated priors. Introducing covariance between function parameters through the warped GP prior, helps interpolating model parameters in areas with low SNR. Furthermore some experiments showed improvement in model performance, when using the warped GP priors and a proper kernel function relating covariates to the governing function parameters, thereby utilizing the underlying structure of the experiment. The proposed model shows the potential of being applied in various different fields of research that require tracking of peaks in low SNR and characterization of materials.

6. REFERENCES

- [1] R. Hilfiker *et al.*, *Polymorphism in the pharmaceutical industry*. Wiley Online Library, 2006, vol. 2.
- [2] J. Scheirs, *Compositional and failure analysis of polymers: a practical approach*. John Wiley & Sons, 2000.
- [3] S. Bose, S. Schmid, T. Larsen, S. Sylvest Keller, A. Boisen, and K. Almdal, “Micromechanical fast quasi-static detection of and relaxations with nanograms of polymer,” *Journal of Polymer Science Part B: Polymer Physics*, vol. 53, no. 15, pp. 1035–1039, 2015.
- [4] S. A. Manzoor Bukhari, M. F. Khan, A. Goswami, R. McGee, and T. Thundat, “Thermomechanical analysis of picograms of polymers using a suspended microchannel cantilever,” *RSC Adv.*, vol. 7, pp. 8415–8420, 2017.
- [5] M. Karl, P. E. Larsen, V. P. Rangacharya, E. T. Hwu, J. Rantanen, A. Boisen, and T. Rades, “Ultrasensitive microstring resonators for solid state thermomechanical analysis of small and large molecules,” *Journal of the American Chemical Society*, vol. 140, no. 50, pp. 17 522–17 531, 2018.
- [6] K. Ekinici and M. Roukes, “Nanoelectromechanical systems,” *Review of scientific instruments*, vol. 76, no. 061101, 2005.
- [7] P. O. Okeyo, O. Ilchenko, R. Slipets, P. E. Larsen, A. Boisen, T. Rades, and J. Rantanen, “Imaging of dehydration in particulate matter using raman line-focus microscopy,” *Scientific reports*, vol. 9, no. 1, p. 7525, 2019.
- [8] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT Press Cambridge, MA, 2006, vol. 2, no. 3.
- [9] M. N. Schmidt and H. Laurberg, “Non-negative matrix factorization with gaussian process priors,” *Computational Intelligence and Neuroscience*, 2008.
- [10] M. N. Schmidt, “Function factorization using warped gaussian processes,” in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 921–928.
- [11] M. Zieba, P. Klukowski, A. Gonczarek, Y. Nikolaev, and M. J. Walczak, “Gaussian process regression for automated signal tracking in step-wise perturbed nuclear magnetic resonance spectra,” *Applied Soft Computing*.
- [12] S. Schmid, L. G. Villanueva, and M. L. Roukes, *Fundamentals of nanomechanical resonators*. Springer, 2016, vol. 49.
- [13] J. Burkardt, “The truncated normal distribution,” *Department of Scientific Computing Website, Florida State University*, 2014.