



## Statistical Finite Element Methods for Nonlinear PDEs

*Connor Duffin*

BSc (Hons), The University of Western Australia

Department of Mathematics and Statistics

ARC Industrial Transformation Hub for Offshore Floating Facilities

Supervisors:

*A/Prof. Edward Cripps*

*Dr. Thomas Stemler*

*Prof. Mark Girolami*

May 2022

This thesis is presented for the degree of Doctor of Philosophy of The University of Western Australia



---

## Thesis Declaration

I, Connor Duffin, certify that:

This thesis does not contain material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution.

No part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of The University of Western Australia and where applicable, any partner institution responsible for the joint-award of this degree.

This thesis does not contain any material previously published or written by another person, except where due reference has been made in the text.

The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

This thesis contains published work and/or work prepared for publication, some of which has been co-authored.

Signature:

Date:

---

This research was carried out while the author was in receipt of a Bruce and Betty Green Postgraduate Research Scholarship and an Australian Government Research Training Program Scholarship at The University of Western Australia. The author was also supported by the ARC Industrial Transformation Hub for Offshore Floating Facilities, funded by the Australian Research Council, Woodside Energy, Shell Australia, Bureau Veritas, and Lloyds Register. Resources were provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

## Abstract

The central role of physically derived nonlinear, time-dependent partial differential equations (PDEs) in scientific and engineering research is undisputed, as is the need for numerical approaches to understand their behaviour. The finite element method (FEM) has emerged as the foremost approach to numerically solve PDEs, yet when discretised solutions are compared to empirical evidence, elements of model mismatch are revealed that require coherent statistical formalisms which respect the underlying governing equations, whilst also providing probabilistic quantification of uncertainties induced by model mismatch. One such method is the recently developed *statistical finite element method* (statFEM), initially studied in the context of static, linear PDEs. This thesis extends these developments to the wider class of nonlinear, time-dependent PDEs in one and two dimensions and reports evidence of the applicability of statFEM to systems of higher, and more realistic, model complexity.

We call our extension NL-statFEM, and the first contribution of this thesis is the derivation of this methodology, which sequentially updates numerical solutions to nonlinear PDEs with data. Additive Gaussian process forcing is added into the governing equations, taken to represent incomplete knowledge of the physics. We take a Bayesian approach and leverage the finite element method to cast the statistical problem as a nonlinear Gaussian state–space model, updating the solution, in receipt of data, in a filtering framework. To compute the approximate Gaussian posterior, we use the extended and ensemble Kalman filters.

NL-statFEM is applicable to problems across science and engineering for which finite element methods are appropriate, and we demonstrate its application on canonical nonlinear PDEs. Results with the Burgers, Kuramoto Sivashinsky and Korteweg-de Vries equations show that statFEM: (1) corrects for model misspecification; (2) provides good approximations to the data generating process; and (3) can leverage physical information to serve as an inferential regulariser in sparse observational settings.

Although extending FEM to high-dimensional systems is computationally feasible, for NL-statFEM a challenge still remains when dealing with such systems of increasing model complexity. The second contribution of this thesis is, therefore, a scalable filtering procedure that melds a low-rank covariance approximation with NL-statFEM. The posterior covariance matrix is represented through its leading-order modes, giving an efficient filter that captures a majority of the variance. This low-rank filter overcomes the computational burdens associated with filtering in high dimensions and enables the application of NL-statFEM to high-dimensional problems that are typical in physical and industrial contexts. Demonstrated on a series of reaction-diffusion problems of increasing dimension, using experimental and simulated data, the method reconstructs the sparsely observed data-generating processes with minimal loss of information, in both posterior mean and the variance.



---

## Acknowledgments

First, to my supervisors and friends Thomas and Ed: I am forever grateful for the opportunity that you gave me to explore this problem, taking me under your wing and letting me “cut loose” on something that I know had been simmering for a while. Thomas: thank you, above all, for making this whole experience some of the most fun I have ever had. Thanks for pulling my head out of the weeds when I was in desperate need of it, and for probing my often incoherent thoughts into something vaguely coherent. Thanks also for the geeky computer/bike chats. Ed: thanks for being a mentor in so much more than the technical stuff. I think it’s fair to say that choosing to work with you, way back in the Honours days, was one of the best decisions I have ever made. It’s hard to imagine what my life would be like without your support and guidance over the years. Thanks a million.

To my external supervisor Mark: thanks for taking a chance on a rogue West Australian, and for your unbridled enthusiasm and energy. Haven’t things turned out well — here’s to many more years of fun!

To my academic “big brothers” Mike and Lachy: thanks for showing me the ropes, helping me adjust into PhD life, and welcoming me into the burger fraternity. I couldn’t have asked for better officemates, nor for a more fun way to start up this chapter of my life. Thanks also to Domenic and Jan, whose proofreading of this thesis was exceedingly helpful.

To my best friends Jayden and Ben, thanks for all the adventures, climbing, and conversations. Thanks must also go to my friends Mihali, Matt, Paul, Jack, Conor, and my Honours comrades. Thanks also to my newer friends in the UK: Arnaud, Alex, Ben, Lawrence, Thomas, Deniz, and Andrew, for making this crazy adjustment feel like it was no big deal at all.

This work would not have happened without the help of various climbing gyms around the world, whose chalky environments helped me to escape my head every now and again. Also, big love to all the artists who made the music that this body of work was done under the influence of. The (perhaps subconscious) affect of all this should not be understated!

To my family: thank you for looking out for me, and for providing the best support network I could have asked for in Perth. Mum and Dad, thanks for your never-ending belief, support, and love through the years. To Sinead and Teagan, thank you for being my closest friends, and for teaching me more than I could have ever imagined. I am so grateful to have you in my life.

Lastly, to Molly. Thanks for being such an amazing addition at the tail-end of this whole ordeal. Thanks for helping me to see the grand scheme, and for pulling me back when I have drifted too far inward. You’re the best :).

---

*To my family.*

---

## Authorship Declaration

This thesis contains work that has been published.

---

**Work 1** Duffin, C., Cripps, E., Stemler, T., and Girolami, M. (2021). Statistical finite elements for misspecified models. *Proceedings of the National Academy of Sciences*, 118(2).  
<https://doi.org/10.1073/pnas.2015006118>

*Location in thesis:*

Chapters 3 and 4.

*Student contribution to work:*

Primary author.

*Co-author signatures and dates:*

| | |

---

**Work 2** Duffin, C., Cripps, E., Stemler, T., and Girolami, M. (2022). Low-rank statistical finite elements for scalable model-data synthesis. *Journal of Computational Physics*.  
<https://doi.org/10.1016/j.jcp.2022.111261>

*Location in thesis:*

Chapter 5.

*Student contribution to work:*

Primary author.

*Co-author signatures and dates:*

| | |

---

---

Student signature:

Date:

---

I, Edward Cripps, certify that the student statements regarding their contribution to each of the works listed above are correct.

Coordinating supervisor signature:

Date:

---

# Contents

<b>Contents</b>	<b>vii</b>
<b>List of Figures</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Inversion . . . . .	2
1.2 Data Assimilation . . . . .	4
1.3 Related work, contributions, and outline . . . . .	4
<b>2 Background and fundamentals</b>	<b>9</b>
2.1 PDE Modelling and Finite Element Methods . . . . .	9
2.1.1 Partial Differential Equations . . . . .	9
2.1.2 Finite Element Methods . . . . .	11
2.2 Probability theory and stochastic differential equations . . . . .	18
2.2.1 Probability theory . . . . .	18
2.2.2 Stochastic Differential Equations . . . . .	20
2.3 Bayesian Inference . . . . .	24
2.3.1 Time-evolving inference . . . . .	26
2.3.2 Gaussian Processes . . . . .	30
2.3.3 Inference with Gaussian Processes . . . . .	32
<b>3 Statistical finite element methods</b>	<b>37</b>
3.1 Statistical finite element methods for static problems . . . . .	37
3.1.1 The linear statFEM prior . . . . .	39
3.1.2 The linear statFEM posterior . . . . .	41
3.2 Linear, time-dependent statFEM . . . . .	46
3.2.1 The linear, time-dependent posterior . . . . .	48
3.3 Nonlinear, time-dependent statFEM . . . . .	50
3.3.1 The nonlinear, time-dependent posterior . . . . .	54
<b>4 Applying NL-statFEM to address misspecification</b>	<b>57</b>
4.1 Burgers equation . . . . .	57
4.1.1 Discretisation . . . . .	60
4.1.2 Example I: estimating hyperparameters . . . . .	60

4.1.3	Example II: mismatched viscosity . . . . .	62
4.2	The Kuramoto Sivashinsky equation . . . . .	64
4.2.1	Discretisation . . . . .	65
4.2.2	Example: mismatched viscosity . . . . .	66
4.3	The Korteweg-de Vries equation . . . . .	67
4.3.1	Discretisation . . . . .	70
4.3.2	Mismatched nonlinearity . . . . .	71
4.3.3	Experimental data . . . . .	72
<b>5</b>	<b>Low-rank filtering for scalability</b>	<b>79</b>
5.1	Reaction-diffusion systems . . . . .	79
5.2	Low-rank NL-statFEM . . . . .	81
5.3	Case studies . . . . .	86
5.3.1	Experimental data: verification . . . . .	86
5.3.2	Mismatch via initial conditions: spiral regime . . . . .	89
5.3.3	Mismatch via initial conditions: oscillatory regime . . . . .	92
<b>6</b>	<b>Discussion</b>	<b>95</b>
6.1	Conclusions . . . . .	95
6.2	Future work . . . . .	97
<b>A</b>	<b>Appendices</b>	<b>99</b>
A.1	Linearisation results . . . . .	99
A.2	Ensemble Kalman filter: additional examples . . . . .	100
A.2.1	Burgers equation . . . . .	100
A.2.2	Kuramoto-Sivashinsky equation . . . . .	100
A.2.3	Korteweg-de Vries equation . . . . .	101
A.3	Low-rank filtering: additional examples . . . . .	103
A.3.1	Influence of prior modes in the cell example . . . . .	103
A.3.2	Verification of parameter estimation . . . . .	104
A.3.3	Catastrophic filter divergence in the spiral wave regime . . . . .	107
	<b>Bibliography</b>	<b>111</b>

# List of Figures

2.1	Finite elements: the piecewise linear $P_1$ interval element, the piecewise linear $P_1$ triangular element, and the piecewise quadratic $P_2$ triangular element. . . . .	15
2.2	Poisson equation example: finite element discretisation results. . . . .	17
2.3	Time integrator stability regions for those used in this thesis. . . . .	23
3.1	Generated data for the Poisson example. Left: single dataset of the $n_{\text{obs}} = 100$ datasets, with the FEM mesh. Points marked with crosshairs are shown as a 1D slice on the figure on the left, which are ordered according to their $x_1$ value. The left figure shows all $n_{\text{obs}} = 100$ at each location. . . . .	46
3.2	StatFEM posterior results for the Poisson example. Left: statFEM prior mean. Centre: statFEM posterior mean. Note that both posterior mean plots are overlaid on the FEM mesh. Right: posterior pointwise variance field. . . . .	46
4.1	Burgers prior results. Left: ExKF estimated prior vs. MC estimated prior, shown with estimated means and 95% credible intervals. Right: MC estimate of the prior under uncertainty refinement, shown with the deterministic FEM solution and the respective 95% credible intervals. . . . .	59
4.2	Burgers example I results. . . . .	61
4.3	Burgers example I results: hyperparameter estimation. . . . .	61
4.4	Burgers example II results: verification quantities. . . . .	62
4.5	Burgers example II: estimated hyperparameters. . . . .	63
4.6	Burgers example II results. Top row: posterior means, 95% credible intervals, data. Bottom row: covariance matrices $\mathbf{C}_{n n}$ . . . . .	63
4.7	KS example: verification quantities. . . . .	67
4.8	KS example: posterior mean results. . . . .	68
4.9	KdV cubic example: verification quantities. . . . .	72
4.10	KdV cubic example: visualising the posterior mean. . . . .	73
4.11	Schematic diagram of the experimental apparatus. Wave-gauges are labelled WG1, WG2, and WG3, and the initial conditions are shown as a grey line, labelled with initial angle $\vartheta^\circ$ . . . . .	73
4.12	Observed data at wave-gauges and NL-statFEM prior mean (given by the deterministic FEM solution to assumed eKdV equation). . . . .	74



4.13	Extended domain extended KdV: mapping from the periodic domain $2L$ into the reflective subdomain $L$ allows for reflective boundary conditions to be incorporated. . . . .	75
4.14	KdV experimental data: verification diagnostics. . . . .	76
4.15	KdV experimental data. Left: data, prior mean, and posterior means and 95% credible intervals across time, at the three observation locations. Right: Posterior means and 95% credible intervals, across the domain, at three timepoints. The posterior, in this case, uses the estimated hyperparameters. . . . .	77
4.16	KdV experimental data: heatmap of the posterior mean $\mathbf{m}_{n n}$ for estimated hyperparameters filter. . . . .	77
5.1	Numerical solutions of the $u_2$ component of the Gray-Scott equation over the domain $\mathcal{D} = [0, 300] \times [0, 300]$ , with $\kappa_1 = 0.1$ , $\kappa_2 = 0.05$ . Different choices of the parameters $(F, k)$ result in variations in the dynamics. . . . .	81
5.2	Leading eigenvalues of $\mathbf{K}_\theta$ over the $128 \times 128$ mesh inside the domain $\mathcal{D} = [0, 50] \times [0, 50]$ . Values are truncated to those which are positive, where the negative values arise due to single-precision floating points being used. . . . .	83
5.3	Cells example: observed data, posterior means, and 95% posterior credible intervals for the four observed data times (top), and posterior means and 95% posterior credible intervals for five hours after the observation times (bottom). . . . .	87
5.4	Cells example: space-time plot of the posterior means for each component. Times at which data are observed are shown with a dashed black line. On the $u$ component there is an immediate correction from data conditioning, not present in the $v$ component: the prior model for $v$ is more accurate. . . . .	88
5.5	Cells example: relative error comparisons between the low-rank and full-rank EnKF variants. . . . .	89
5.6	Spiral regime example: initial conditions on the $u$ component. StatFEM posterior initial condition shown on the left, and the true initial condition shown on the right. . . . .	89
5.7	Spiral regime example: posterior summary plots for time $t = 5$ . . . . .	91
5.8	Spiral regime example: diagnostic plots for the spiral wave example. . . . .	92
5.9	Oscillatory regime example: initial conditions for $u$ component. StatFEM posterior initial condition shown on the left (same for each filter configuration) and the true initial conditions shown on the right. . . . .	93
5.10	Oscillatory regime example: relative $l^2$ errors (top left), effective rank $D_{\text{eff}}$ (top right), and leading posterior covariance modes (bottom) at time $t = 10$ , oscillatory mismatched initial condition. . . . .	94
A.1	Burgers example I results (EnKF). . . . .	101
A.2	Burgers example I results (EnKF): hyperparameter estimation. . . . .	101
A.3	Burgers example II results (EnKF): verification quantities. . . . .	102
A.4	Burgers example II results (EnKF): estimated hyperparameters. . . . .	102
A.5	Burgers example II results (EnKF). Top row: posterior means, 95% credible intervals, data. Bottom row: covariance matrices $\mathbf{C}_{n n}$ . . . . .	103

---

A.6	KS example (EnKF): verification quantities. . . . .	103
A.7	KS example (EnKF): posterior mean results. . . . .	104
A.8	KdV cubic example: results computed with the EnKF. . . . .	105
A.9	KdV experimental data results: computed with the EnKF. . . . .	106
A.10	NL-statFEM low-rank filter, experimental data: influence of prior modes on filter accuracy. Shown are the relative errors on the posterior mean (left) and variance (right) as both the number of prior modes ( $k'$ ) and observation noise ( $\sigma$ ) is increased. . . . .	106
A.11	FEM mesh (purple) and observation locations (blue), for A.3.2. . . . .	107
A.12	Leading 128 eigenvalues of $\mathbf{G}_\theta$ , for A.3.2. . . . .	107
A.13	Diagnostic plots for the stochastic forcing example. . . . .	108
A.14	Norm of the NL-statFEM mean $\mathbf{m}_n^u$ ( $u$ -component) (top), and effective rank $D_{\text{eff}}$ (bottom). . . . .	109



# Chapter 1

## Introduction

Scientific theories are described through mathematical models, which encode knowledge of the system under study through a set of relations between state variables. These relations synthesise prior knowledge into some mathematical form and detail expected behaviours of the system — for example, conservation, dissipation, and/or nonlinearity. However, demonstrating such scientific theories to be true is impossible (Popper, 1959): at best a mathematical model can be verified against data in certain configurations, demonstrating correctness on finitely many specific cases. There exists an asymmetry in the weight of the evidence: if a single observation, or set of observations, is able to show that the theory does not hold, then it has been *falsified*. Due to this asymmetry it is difficult to discuss models as perfectly capturing the true data generating process. This is captured in the well-known statistical aphorism “all models are wrong, but some are useful”, attributed to Box (1979).

Since mathematical models are imperfect, discrepancies are observed in the comparison of the model outputs to measurements. Elements of model mismatch may result from a variety of sources: there may be a missing model subspace (Judd and Smith, 2004), or a mismatch in model state (e.g. discrete vs continuum states) (Berger and Smith, 2019). The model may also be never truly closed, requiring the specification of incompletely known parameters, or certain physics may be deliberately ignored for computational feasibility (Oreskes et al., 1994).

To deal with model discrepancies, statistical protocols are required and in this thesis we take a Bayesian approach to the inference procedure. We do so because (1) Bayesian approaches provide probabilistically coherent uncertainty quantification (UQ); and (2) they allow for the online updating of parameter estimates as more data becomes available. As the main setting of this thesis is within nonlinear, time-dependent PDEs, where data arrives sequentially, the capability of online updating is particularly useful. We develop a Bayesian statistical methodology that combines ideas from the inversion and data assimilation literatures and embeds itself within finite element modelling. More specifically, we introduce function-valued uncertainty with the forcing inside the governing PDE, as in inversion. This uncertainty is taken to represent epistemic model imperfection. However, instead of estimating this forcing, we leverage it to describe a state-space model via finite element discretisation. Numerical solutions are then updated in a filtering context, as in the field of data assimilation. Therefore, to set the scene for this work, we provide an overview of the inverse and data assimilation problems.

## 1.1 Inversion

The goal of the *inverse problem* is to estimate, or infer, model parameters from data, after they have been pushed through some forward model (Tarantola, 2005). The forward model is a possibly nonlinear mapping of the model parameters, commonly given by the solution operator of an ordinary or partial differential equation (ODE or PDE, respectively). Parameters are estimated through the minimisation of some cost function: a classical approach uses the squared Euclidean distance,  $\|\cdot\|_2^2$ , between the model predictions and the data. Denote by  $\mathcal{M}(\mathbf{x}; \Lambda)$  the model with finite-dimensional parameters  $\Lambda$  and fixed inputs  $\mathbf{x}$ . For measurements  $\mathbf{y} = (y_1, \dots, y_n)$  the inverse problem is to estimate the  $\Lambda^*$  such that

$$\Lambda^* = \operatorname{argmin}_{\Lambda} \left\{ \frac{1}{2} \|\mathbf{y} - \mathcal{M}(\mathbf{x}; \Lambda)\|_2^2 + \frac{1}{2\gamma^2} \|\Lambda\|_2^2 \right\}. \quad (1.1)$$

The addition of a regularisation term avoids ill-conditioning, as typically  $\Lambda$  is of higher dimensionality than  $\mathbf{y}$  (for example,  $\Lambda$  is commonly a function, an infinite-dimensional object).

Due to model nonlinearity, the optimisation problem of (1.1) may be nonconvex and possibly poorly conditioned. Therefore scalable gradient-based optimisation schemes are required to estimate  $\Lambda^*$ , such as gradient descent or quasi-Newton schemes (Boyd and Vandenberghe, 2004; Nocedal and Wright, 2006). Gradients can be computed using either the *adjoint sensitivity method* (Plessix, 2006) and/or *automatic differentiation* (Baydin et al., 2018; Farrell et al., 2013).

The problem as formulated above is deterministic and does not introduce any notion of uncertainty. Uncertainties may arise due to observing noisy measurements of the system state, with a data generating process such as

$$\mathbf{y} = \mathcal{M}(\mathbf{x}; \Lambda) + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}). \quad (1.2)$$

Thus the inverse problem can be cast as one of statistical inference, formulated in terms of the *likelihood function*  $p(\mathbf{y} | \Lambda)$ , instead of the cost function (Abramovich and Ritov, 2013). The likelihood is the corresponding joint probability density function of the data,  $\mathbf{y}$ , given  $\Lambda$ ; the inference problem is to determine  $\Lambda$  from the data. In this case,  $p(\mathbf{y} | \Lambda) = \mathcal{N}(\mathcal{M}(\mathbf{x}; \Lambda), \sigma^2 \mathbf{I})$ . If we assume that, *a priori*, our knowledge about  $\Lambda$  is represented by a prior distribution  $p(\Lambda)$ , we can combine this with the likelihood, using Bayes theorem, to give the posterior distribution

$$p(\Lambda | \mathbf{y}) \propto p(\mathbf{y} | \Lambda)p(\Lambda).$$

The choice of the prior may be weakly or strongly informative, depending on the modelling scenario, and may also be *expert elicited* (see, e.g., Oakley and O'Hagan, 2007). Often a Gaussian prior is used, as, in the infinite dimensional context, only Gaussian measure provides an adequate reference measure in the function space (Da Prato and Zabczyk, 2014).

For the finite-dimensional case, with densities taken with respect to Lebesgue measure, combining a mean-zero Gaussian prior with the Gaussian likelihood gives the same *negative log-posterior density* as the cost function of (1.1)<sup>1</sup>. Solving the Bayesian inverse problem involves estimating the posterior distribution  $p(\Lambda | \mathbf{y})$ , providing a complete characterisation of the uncertainty associated with  $\Lambda$

<sup>1</sup>In (1.1) this requires setting  $\sigma = 1$  and, for the prior, specifying  $\Lambda \sim \mathcal{N}(\mathbf{0}, \gamma^2 \mathbf{I})$ . The negative log-posterior density is also called the *potential function* in physics, giving rise to a Gibbs measure. These measures also arise in the energy based models community, in machine learning (LeCun et al., 2006).

conditioned on model assumptions.

However, estimating the posterior distribution is challenging as it is commonly known only up to a normalising constant. Variables of interest are also likely to be high-dimensional, and model nonlinearities will also complicate estimation: the resulting posterior may be possibly multimodal or poorly-conditioned. A standard estimation methodology is Markov Chain Monte Carlo (MCMC) sampling (Casella and George, 1992; Chib and Greenberg, 1995; Tierney, 1994). MCMC constructs a Markov Chain whose stationary distribution is the posterior distribution of interest. The *Metropolis-Hastings condition* (Hastings, 1970; Metropolis et al., 1953) is used to construct the transition kernel, in combination with a requisite proposal distribution. Popular MCMC algorithms for inverse problems include the preconditioned Crank-Nicolson (Cotter et al., 2013), the stochastic Newton algorithm (Martin et al., 2012), and multilevel MCMC (Dodwell et al., 2015).

However, even under optimal model parameters there may still be structural problems that result in mismatch. To deal with this problem, the framework of Bayesian inversion was generalised in the seminal paper of Kennedy and O’Hagan (2001). In this structural model imperfections are dealt with via an additive Gaussian Process ( $\mathcal{GP}$ ). The  $\mathcal{GP}$  is a continuous random process which places a prior distribution over an appropriate space of functions. Upon receipt of data, this prior is updated to give a posterior distribution (Williams and Rasmussen, 2006). The  $\mathcal{GP}$ , in this case, is assumed to capture additive smooth model error that arises from model inadequacy that is statistically independent from the model itself. This functional model error is typically assumed to be smooth. Such a framework posits the data generating process of

$$\mathbf{y} = \rho\mathcal{M}(\mathbf{x}; \mathbf{\Lambda}) + \xi(\mathbf{x}) + \boldsymbol{\eta}, \quad (1.3)$$

where, borrowing the notation of the model introduced above,  $\rho \in \mathbb{R}$ ,  $\xi(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ ,  $\boldsymbol{\eta} \sim \mathcal{N}(0, \mathbf{R})$ . The  $\mathcal{GP}$ ,  $\xi$ , is completely specified by its *mean function*  $m(\mathbf{x})$  and *covariance function*  $k(\mathbf{x}, \mathbf{x}')$ . The mean function defines how we expect  $\mathcal{GP}$  sample paths to behave on average and the covariance function describes the properties of these sample paths (e.g, smoothness). For more details see Chapter 2, Sections 2.3.2 and 2.3.3, of this thesis. The data are thus generated according to a scaled version of the model, plus a functional error term, and observational noise.

The goal of the Kennedy-O’Hagan framework is to estimate the joint posterior distribution over all parameters in the above (model parameters,  $\mathcal{GP}$  mean parameters,  $\mathcal{GP}$  covariance hyperparameters, noise parameters). This jointly quantifies the uncertainty associated with all unknown parameters and provides an approximation to the true data generating process, given by  $\rho\mathcal{M}(\mathbf{x}; \mathbf{\Lambda}) + \xi(\mathbf{x})$ . However, the model parameters  $\mathbf{\Lambda}$  and model error  $\xi(\mathbf{x})$ , as formulated in (1.3), are not identifiable (Bayarri et al., 2007a,b). Estimating the posterior also inherits the same difficulties as the Bayesian inverse problem as it will typically be known only up to a normalising constant, and may be multimodal and poorly conditioned. Nevertheless, the framework has been widely adopted within the Bayesian statistics literature (Goldstein and Rougier, 2006; Higdon et al., 2004, 2008; Plumlee et al., 2016), and there has been some work done on rectifying the identifiability problem (Plumlee, 2017; Tuo and Wu, 2015).

## 1.2 Data Assimilation

Data Assimilation (DA) can be considered a special case of the inverse problem, in which the goal is to infer the underlying data generating process given noisy, subsampled measurements of a temporally-evolving system (Law et al., 2015). For model parameters  $\Lambda$  (now taken as fixed), an example DA model is, for the time index  $n = 1, \dots$

$$\begin{aligned}\mathbf{u}_n &= \mathcal{F}(\mathbf{u}_{n-1}; \Lambda) + \mathbf{q}_n, & \mathbf{q}_n &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \\ \mathbf{y}_n &= \mathbf{H}\mathbf{u}_n + \mathbf{r}_n, & \mathbf{r}_n &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}).\end{aligned}\tag{1.4}$$

The object of interest is the posterior distribution  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \Lambda)$ , where  $\mathbf{y}_{1:n} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ . For linear  $\mathcal{F}(\cdot; \Lambda)$  this distribution is Gaussian,  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \Lambda) = \mathcal{N}(\mathbf{m}_n, \mathbf{C}_n)$ , and can be estimated by the Kalman filter (Kalman, 1960).

The nonlinear Bayesian filtering problem requires two steps: *marginalisation* and *updating*. Assuming that  $p(\mathbf{u}_{n-1} | \mathbf{y}_{1:n-1}, \Lambda)$  is known, the marginalisation step first integrates over the uncertainty in the previous timestep, computing  $p(\mathbf{u}_n | \mathbf{y}_{1:n-1}, \Lambda) = \int p(\mathbf{u}_n | \mathbf{u}_{n-1}, \Lambda) p(d\mathbf{u}_{n-1} | \mathbf{y}_{1:n-1}, \Lambda)$ . This distribution forms the prior for the data observed at the current timestep, so the posterior, up to a normalising constant, is  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \Lambda) \propto p(\mathbf{y}_n | \mathbf{u}_n) p(\mathbf{u}_n | \mathbf{y}_{1:n-1}, \Lambda)$ . Exact computation of  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \Lambda)$  is difficult due to the high dimensionality of  $\mathbf{u}_n$  and the nonlinearity of  $\mathcal{F}(\cdot; \Lambda)$ . The efficient approximation of this posterior distribution is the focus of the DA literature.

The *approximate Gaussian filters* are methods to estimate this posterior in an online fashion: two common examples are the extended and ensemble Kalman filters (ExKF and EnKF, respectively), which each recursively compute the approximation  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \Lambda) \approx \mathcal{N}(\mathbf{m}_n, \mathbf{C}_n)$ . The ExKF linearises about the current timestep to compute a first-order Gaussian approximation  $p(\mathbf{u}_n | \mathbf{y}_{1:n-1}, \Lambda)$ , and uses the Gaussian update to give  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \Lambda)$ . The EnKF constructs a Monte Carlo approximation to  $p(\mathbf{u}_n | \mathbf{y}_{1:n-1}, \Lambda)$  through a set of *particles* or *ensemble members*, and assumes that this *ensemble* has a Gaussian distribution. Each ensemble member is updated with the standard Gaussian update. A similar non-Gaussian approach is the particle filter (Doucet et al., 2000), which may struggle in high dimensions due to weight degeneracy (Bengtsson et al., 2008). Therefore, we focus on the approximate Gaussian filters only. These approaches can be added on top of existing finite element libraries whilst also providing reasonable estimates of the posterior distribution of interest (Law and Stuart, 2012). Computational scalability is also ensured through making a low-rank approximation in the ExKF case, and, through utilising small numbers of particles (relative to the state dimension) in the EnKF case.

## 1.3 Related work, contributions, and outline

Recently, there has been a trend toward combining statistical learning and physics, with these methods coined *physics-informed learning* (Karniadakis et al., 2021). Physical knowledge is incorporated into the inference procedure, enabling data-efficient learning in possibly sparse observational settings, where classical deep/machine learning approaches may fail. The methodology we develop in this thesis can be thought of as a physics-informed online regression methodology, and adds to this literature. Thus we include a brief review of these works to establish our contribution.



A popular example is the SINDy algorithm (Brunton et al., 2016), which uses sparse regression to identify ODE systems from noisy time series measurements. This builds upon placing a sparsity-enforcing regularisation term when determining the ODE from a library of candidate functions. This is extended to the PDE setting in Rudy et al. (2017), which addresses system identification from spatio-temporal measurements inside the problem domain.

Neural network (NN) approaches are common. A popular method is to model the PDE variable with an NN, setting  $u(\mathbf{x}, t) = \text{NN}_\theta(\mathbf{x}, t)$ , for  $\mathbf{x} \in \mathbb{R}^d$ ,  $t \in \mathbb{R}_+$ . Termed *physics-informed neural networks* (Raissi et al., 2019), the cost function includes terms for the PDE dynamics, the initial conditions, and the boundary conditions; physics regularisation ensures the NN is data-efficient. NNs can also be used for system identification (Long et al., 2018), and also for estimating optimal discretisations from data (Bar-Sinai et al., 2019). A latent dynamical model can alternatively be used to facilitate continuous-depth NNs (Chen et al., 2018), allowing for temporal dependencies in NN models. Universal differential equations (UDEs) (Rackauckas et al., 2021) provide a generic formalism that incorporates universal approximators (of which NNs are an example of) into differential equations. These allow for the combination of NN methodology whilst also being able to leverage classical numerical schemes. Adjacent to the direct inclusion of NNs in physical models are neural operators, which construct solution operators to differential equation from NNs (Kovachki et al., 2021; Li et al., 2020).

Gaussian processes can be used in a similar fashion. In Raissi and Karniadakis (2018), a  $\mathcal{GP}$  prior is placed over the PDE variable so that the covariance function encodes the *a priori* specified physical laws. This allows for system identification through transforming PDE coefficients into  $\mathcal{GP}$  hyperparameters, whose estimation is well-studied (see, e.g., Williams and Rasmussen, 2006).

*Probabilistic numerical methods* (Hennig et al., 2015) casts standard tasks in computational mathematics as inference problems, for example, evaluating integrals (Briol et al., 2019), solving linear systems (Cockayne et al., 2019), and solving differential equations (Cockayne et al., 2017). The associated uncertainty quantifies discretisation error from the numerical method and can be included in the computational pipeline when estimating quantities of interest. This provides complete UQ which considers known parametric and numerical sources of uncertainty.

The main contribution of this thesis is the development of a novel statistical methodology to adjust for model misspecification in nonlinear, time-dependent PDEs; the methodology uses a combination of ideas from both Bayesian inversion and data assimilation, and lies within the physics-informed learning literature. A  $\mathcal{GP}$  is stipulated that encodes our *a priori* belief in the model parameters  $\Lambda$ . We do not try and solve the system identification problem (as in, e.g., Rudy et al., 2017), nor, as in Bayesian inversion, do we try and estimate these parameters. Instead, we marginalise to derive a prior distribution over the model solutions  $\mathbf{u}_n$ . This provides a principled way to derive a physics-informed prior distribution from the governing equations, through marginalising over uncertainty inside of them.

Another interpretation of this approach is that we shift the additive  $\mathcal{GP}$  model error of Kennedy and O’Hagan (2001) into the governing equations. However instead of placing the  $\mathcal{GP}$  prior over the PDE variable, as in Raissi and Karniadakis (2018), we leverage the induced uncertainty over the assumed unknown forcing to give a prior. Similar to a UDE, having expressed this uncertainty inside of the PDE, we use classical numerical methods to discretise the problem, in our case the



finite element method, the most widely-used PDE discretisation technique in science and engineering. Hence unlike probabilistic numerics, discretisation uncertainty is not explicitly included in the derived prior measure.

Subsequently, the prior is updated in a sequential fashion, as in data assimilation, through standard nonlinear filtering algorithms (the EnKF and ExKF as above). This gives a posterior distribution  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \mathbf{\Lambda})$ , which describes the belief in the model solutions given the data up to the current time. In this thesis we study the derivation and computation of this posterior distribution, deriving the algorithms in the finite-dimensional setting.

We title our approach *NL-statFEM*, as we build upon the *statistical finite element method (statFEM)*, as given in [Girolami et al. \(2021\)](#). StatFEM introduces similar methodology for linear, static PDEs and studied its application to various scenarios with the Poisson equation. The methodological contribution of this thesis is the computationally scalable extension of statFEM into nonlinear, time-dependent problems, demonstrating efficacy on nonlinear PDEs in one and two dimensions, with experimental and synthetic data. Note that to distinguish our work we refer to our approach as the NL-statFEM exclusively; we reserve the statFEM label for the work of [Girolami et al. \(2021\)](#) only. We apply our NL-statFEM to a number of classical nonlinear PDEs, and these results show that:

1. NL-statFEM is able to sequentially combine measurements of phenomena modelled using PDEs with discretised solutions to their governing equations, giving an approximation to the data generating process with a statistically coherent uncertainty quantification.
2. Both physics and data can be leveraged to give an interpretable posterior distribution, which utilises data to reconstruct incompletely observed phenomena.
3. NL-statFEM performs effectively in sparse observation settings.
4. NL-statFEM enables the application of simpler physical models, correcting for model inadequacy through data, in both weakly and highly nonlinear/chaotic modelling regimes.
5. Taking a low-rank approximation to the covariance matrix scales NL-statFEM to high degrees-of-freedom models, enabling its application to large-scale systems.

The presentation proceeds as follows. In Chapter 2, the background material of this thesis is covered, introducing partial differential equations and the finite element method, and probabilistic inference. Chapter 3 goes through the full construction of the methodology, beginning with the previous work of [Girolami et al. \(2021\)](#) for the linear, static case, before we detail our extension — NL-statFEM — as the expansion of this methodology into both time-dependent and nonlinear problems. Chapter 4 then demonstrates NL-statFEM on three nonlinear, time-dependent PDEs — the Burgers, Kuramoto-Sivashinsky, and Korteweg-de Vries equations — using synthetic and experimental data. Chapter 5 further extends the NL-statFEM as presented in Chapter 3, scaling the method to higher-dimensional systems. In this chapter, we illustrate NL-statFEM on a series of coupled reaction-diffusion problems in one and two dimensions — a cell reaction-diffusion model and the Oregonator system — as these provide the requisite increase in dimensionality whilst also containing sufficient model complexity. The thesis is concluded in Chapter 6 with suggestions for future research.

All code to replicate these results is contained in two public repositories, corresponding to the two publications on which this thesis is based. For results pertaining to Chapter 4 this code is hosted at <https://github.com/connor-duffin/statkdv-paper>, and for Chapter 5 this is hosted at <https://github.com/connor-duffin/low-rank-statfem>.



# Chapter 2

## Background and fundamentals

*In order to present our NL-statFEM methodology, various background concepts must be introduced: this is the focus of this chapter. The physics-informed component of NL-statFEM is assumed to be given by a PDE. Therefore, we begin with a discussion on partial differential equations, with an emphasis on the classical linear PDE systems. We give a brief overview of nonlinear PDEs, deferring their exposition to the examples given in Chapter 4 and 5. Discretisation with FEM is introduced in the context of the Poisson equation, and an example is given to demonstrate the modelling process. Next, we introduce probability theory and stochastic differential equations (SDEs) as the framework we use to model uncertainty. After FEM discretisation, the latent dynamical model of NL-statFEM is an SDE. We introduce the SDE notation which is used throughout the remainder of this thesis, and cover the classical Euler-Maruyama discretisation methods. The Bayesian statistical framework is then introduced, as this is the framework that we use to sequentially update discretised SDE trajectories. Both static and dynamic Bayesian inference is introduced, for finite-dimensional unknown quantities. We conclude this chapter with a discussion on infinite-dimensional Bayesian inference as motivated by Gaussian Processes. As discussed in the introduction, in NL-statFEM we induce uncertainty in the governing PDE through Gaussian Processes as these describe uncertainties associated with functions, enabling the inference procedure.*

### 2.1 PDE Modelling and Finite Element Methods

#### 2.1.1 Partial Differential Equations

In this thesis we develop Bayesian statistical methodology to infer over numerical solutions of physical models. The physical models we work with are partial differential equations (PDEs), which provide a mathematical description of the rates of change of some quantity of interest with respect to, generally, some spatial and/or temporal variables. These rates of change are expressed by partial derivatives and are combined to express some physical law or approximation thereof. A PDE can be defined as follows (Evans, 2010; Strauss, 2007).

**Definition 1.** *Take some domain  $\mathcal{D} \subset \mathbb{R}^d$ , with a boundary  $\partial\mathcal{D}$ , and let  $k \geq 1$  be some integer. Then the expression*

$$F(\nabla^k u(\mathbf{x}), \nabla^{k-1} u(\mathbf{x}), \dots, u(\mathbf{x}), \mathbf{x}) = 0, \quad \mathbf{x} \in \mathcal{D}, \quad (2.1)$$

where  $F : \mathbb{R}^{d^k} \times \mathbb{R}^{d^{k-1}} \times \dots \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$ , defines a  $k$ th order PDE.

In this section a primer on PDE models is provided, focussing on the canonical linear PDE systems

most often seen in introductory contexts. PDEs arise in all fields of science and engineering and in this thesis we are particularly interested in those which include nonlinearity. However, it is difficult to discuss nonlinear PDEs in the same fashion as the classical linear PDEs, so in this chapter only the definition of a nonlinear PDE is given. To gain some intuition with their general form, four linear PDEs are introduced, whose qualitative behaviour is reflected in various ways in the chapters to come. Historical details can be found in [Brezis and Browder \(1998\)](#).

An early example, first introduced by d'Alembert in 1752, for the one-dimensional case, is the *wave equation*

$$u_{tt} - \Delta u = 0, \quad (2.2)$$

where  $\Delta$  denotes the Laplacian operator  $\Delta u = \sum_{i=1}^d u_{x_i x_i}$ . Extensions into two- and three-dimensions were covered by Euler (in 1759) and Bernoulli (in 1762). The wave equation describes the evolution of the displacement  $u$  on some  $d$ -dimensional manifold that arises from the balance of elastic forces, in the absence of damping. For example, in  $d = 1$  with  $x \in \mathbb{R}$ , Equation (2.2) describes the oscillation of an infinitely long string. The *d'Alembert solution* can be derived in this setting and states that for  $u(x, 0) = u_0(x)$ ,  $u_t(x, 0) = v_0(x)$ ,

$$u(x, t) = \frac{1}{2} (u_0(x-t) + u_0(x+t)) + \frac{1}{2} \int_{x-t}^{x+t} v_0(s) \, ds.$$

The wave equation propagates the initial conditions through the domain. There is no damping, with the initial profile being horizontally displaced along the left- and right-moving components plus a contribution from the initial velocity. There is some inherent *conservation* as the energy integral  $\frac{1}{2} \int_{\mathbb{R}} u_t^2(x, t) + u_x^2(x, t) \, dx$  is conserved for all  $t$ .

The *Laplace equation*

$$-\Delta u = 0, \quad (2.3)$$

was first introduced by Laplace in 1780, being applied to the study of gravitational potential fields. The inclusion of a forcing term in (2.3), first seen in the context of electromagnetism, gives the *Poisson equation*

$$-\Delta u = f. \quad (2.4)$$

The Poisson and Laplace equations describe the steady-state behaviour of a material property, and can be derived through equating the flux of the material property out of  $\mathcal{D}$ . In the case of the Poisson equation, the additional forcing  $f$  represents the generation and removal of the material property due to sources and sinks.

Adding in a time derivative to (2.3) gives the *heat equation*

$$u_t - \Delta u = 0, \quad (2.5)$$

first derived by [Fourier \(1822\)](#). The heat equation describes the diffusion of some material property throughout the medium and is associated with *dissipation*. Similar to the Poisson equation, it can be derived through equating the rate of change of this property to the flux of the quantity out of the domain  $\mathcal{D}$ . The steady-state distribution of this quantity satisfies the Laplace equation (in steady

state, there is no temporal evolution so  $u_t = 0$ ). The heat equation also arises in stochastic differential equations (SDEs), where, in one dimension, it describes the evolution of the probability density function of the standard Brownian motion process. In this context it is known as the *Fokker-Planck equation* (Øksendal, 2003).

These second-order PDEs are the prototypical examples of *hyperbolic*, *elliptic*, and *parabolic* systems, respectively. For a general second order equation with  $\mathbf{A}_{ij} = \mathbf{A}_{ji}$

$$\sum_{i,j=1}^d \mathbf{A}_{ij} \frac{\partial^2 u(\mathbf{x})}{\partial x_i \partial x_j} + \sum_{i=1}^d \mathbf{a}_i \frac{\partial u(\mathbf{x})}{\partial x_i} + u(\mathbf{x}) = 0$$

these PDE classifications can be defined from the eigenvalues of  $\mathbf{A}$ . The elliptic equations are those whose eigenvalues are all the same sign. If one of these eigenvalues is zero and the rest have the same sign, then the equation is parabolic. The equation is called hyperbolic if one of the eigenvalues has the opposite sign to the  $d - 1$  others.

Boundary conditions may also need to be specified. The *Dirichlet* boundary condition of  $u = g$  on  $\partial\Omega$  means that  $u$  is fixed to  $g$  on the boundary. Alternatively we may specify the *Neumann* condition that  $\nabla u \cdot \mathbf{n} = g$ , for  $\mathbf{n}$  normal to the boundary  $\partial\mathcal{D}$ , or the *Robin* condition  $\alpha u + \beta \nabla u \cdot \mathbf{n} = g$ , for nonzero scalar constants  $\alpha, \beta$ . For time-dependent PDEs such as (2.2) and (2.5), initial conditions should also be specified:  $u(\mathbf{x}, 0) = u_0(\mathbf{x})$ .

Despite the rich phenomena that (2.2)-(2.4) can exhibit, incorporating nonlinear effects may also be required. For completeness, nonlinearity is now defined here.

**Definition 2.** A nonlinear PDE is one in which  $F(\cdot)$  violates the condition of linearity in any one of its arguments. That is, letting  $\mathbf{u} = (\nabla^k u(\mathbf{x}), \nabla^{k-1} u(\mathbf{x}), \dots, u(\mathbf{x}), \mathbf{x})$  if we cannot write

$$\alpha F(\mathbf{u}) + \beta F(\mathbf{u}) = F((\alpha + \beta)\mathbf{u}),$$

then the PDE  $F(\nabla^k u(\mathbf{x}), \nabla^{k-1} u(\mathbf{x}), \dots, u(\mathbf{x}), \mathbf{x}) = 0$  is nonlinear.

Nonlinearity is commonly encountered in practice and complicates theoretical treatments. General systems of nonlinear PDEs can usually only be studied analytically on a case-by-case basis, where the analyst is able to make some assumptions on the parametric form and/or regularity of the nonlinearity. Therefore, the full discussion of the example nonlinear PDEs considered in this thesis is adjourned until the examples are presented, due to the different physics that arises in each system.

### 2.1.2 Finite Element Methods

Ideally, solutions to (2.1) can be obtained analytically, using a separation of variables or a transform-type method. However, often numerical approaches have to be taken due to nonlinearities or complex geometries. Additionally — and most pertinent to the contribution of this thesis — is that in order to update the physical model solutions with observed data, a finite-dimensional numerical solution is required.

In order to construct these numerical solutions we use the finite element method (FEM), taking some time here to describe the process of finite element discretisation for the motivating example of

the Poisson equation. Numerical approaches replace the infinite-dimensional problem of finding the appropriate function  $u$  that satisfies (2.1) with a finite-dimensional substitute that is computationally tractable.

FEM combines a discrete approximation to the domain  $\mathcal{D}$  (the *mesh*) with a specified PDE to be solved on this domain. The PDE is solved in terms of the *weak form*, which lowers the order of differentiability of the solutions whilst still obtaining valid solutions to the PDE of interest (see, e.g., Larson and Bengzon, 2013). Instead of providing a pointwise approximation to the PDE as in finite differences, the FEM solves for a function in a *trial space* of possible functions. This trial space is commonly a space spanned by compactly supported polynomial basis functions. For computational tractability, this function space has a finite number of *degrees-of-freedom* (DOFs) which parameterise the functional approximation. The number of DOFs is typically a function of the mesh refinement and the degree of the polynomial approximation.

In this section the finite element modelling procedure is described, using the two-dimensional Poisson equation as an example. We focus on spatial discretisation, as temporal discretisation will be taken care of via Euler-Maruyama methods, discussed in the proceeding Section 2.2. This is because all time-evolving systems that are studied in the later chapters are in fact SDEs which arise from FEM discretisation of stochastic PDEs (developed fully in Chapter 3). For some references to FEM modelling, see Brenner and Scott (2008); Ciarlet (2002); Langtangen and Mardal (2019); Suli (2020); Thomée (2006), from which the following presentation draws from.

As a motivating example throughout this section, consider the Poisson equation

$$\begin{cases} -\Delta u = f, \\ u = 0, & \text{for } \mathbf{x} \in \partial\mathcal{D}, \\ u := u(\mathbf{x}), \mathbf{x} \in \mathcal{D} \subseteq \mathbb{R}^2. \end{cases} \quad (2.6)$$

Various function spaces are required to discuss the FEM solution to (2.6). To begin, we introduce the spaces of continuous functions  $C^p(\mathcal{D}; \mathbb{R})$ .

**Definition 3.** The  $C^p(\mathcal{D}; \mathbb{R})$  spaces are the spaces of functions  $u : \mathcal{D} \rightarrow \mathbb{R}$  such that  $u$  is  $p$ -times continuously differentiable.  $C_0^p(\mathcal{D}; \mathbb{R})$  is the space of functions which are  $p$ -times continuously differentiable and vanish on the boundary  $\partial\mathcal{D}$ . Let

$$C^\infty(\mathcal{D}; \mathbb{R}) = \bigcap_{p \geq 0} C^p(\mathcal{D}; \mathbb{R}), \quad C_0^\infty(\mathcal{D}; \mathbb{R}) = \bigcap_{p \geq 0} C_0^p(\mathcal{D}; \mathbb{R}).$$

The  $L^p(\mathcal{D}; \mathbb{R})$  spaces,  $1 \leq p \leq \infty$  are defined next.

**Definition 4.** The  $L^p(\mathcal{D}; \mathbb{R})$  spaces, for  $1 \leq p < \infty$ , are the spaces of functions  $f : \mathcal{D} \rightarrow \mathbb{R}$  such that

$$L^p(\mathcal{D}; \mathbb{R}) = \left\{ f : \int_{\mathcal{D}} |f(\mathbf{x})|^p \, d\mathbf{x} < \infty, \text{ almost everywhere} \right\}$$

with norm

$$\|f\|_p = \left( \int_{\mathcal{D}} |f(\mathbf{x})|^p \, d\mathbf{x} \right)^{1/p}.$$

The integral is to be understood as a Lebesgue integral. The elements of  $L^p(\mathcal{D}; \mathbb{R})$  are the sets of equivalence

classes of square-integrable functions, which are the same up to sets of measure zero. The space  $L^2(\mathcal{D}; \mathbb{R})$  is a Hilbert space with inner product  $\langle \cdot, \cdot \rangle$ , defined by

$$\langle f, g \rangle = \int_{\mathcal{D}} f(\mathbf{x})g(\mathbf{x}) \, d\mathbf{x}.$$

Note that  $\langle f, f \rangle = \|f\|_2^2$ . The space  $L^\infty(\mathcal{D}; \mathbb{R})$  denotes the set of functions which are bounded almost everywhere. The norm is defined by

$$\|u\|_\infty = \inf \{C : |u(\mathbf{x})| \leq C, \text{ for almost every } \mathbf{x}\}.$$

For clarity, when referring to multiple integrals over  $\mathcal{D} \subset \mathbb{R}^d$  we use the notation

$$\int_{\mathcal{D}} f(\mathbf{x}) \, d\mathbf{x} := \int \dots \int_{\mathcal{D}} f(x_1, \dots, x_d) \, dx_1 \dots dx_d,$$

where the individual integral bounds are defined from the region  $\mathcal{D}$ .

Next, the Sobolev spaces are defined, which are the spaces in which we “look” for weak solutions to (2.6). They rely on the notion of a weak derivative, which generalises the derivative through integration by parts. Consider the multiindex  $\alpha = (\alpha_1, \dots, \alpha_d)$  such that  $|\alpha| = \alpha_1 + \dots + \alpha_d = k$ . A partial derivative for some test function  $v \in C_0^\infty(\mathcal{D}; \mathbb{R})$  is

$$D^\alpha v = \frac{\partial^k}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} v.$$

The weak derivative of some function  $u \in L^1(\mathcal{D}; \mathbb{R})$  is the function  $w \in L^1(\mathcal{D}; \mathbb{R})$  such that

$$\int_{\mathcal{D}} u D^\alpha v \, d\mathbf{x} = (-1)^{|\alpha|} \int_{\mathcal{D}} w v \, d\mathbf{x},$$

for all  $v \in C_0^\infty(\mathcal{D}; \mathbb{R})$ . Denoted by  $w = D^\alpha u$ , the weak derivative coincides with the partial derivative for  $u$  sufficiently smooth.

The Sobolev spaces are those which have  $L^p(\mathcal{D}; \mathbb{R})$  weak derivatives up to order  $k$ , and are defined here.

**Definition 5.** The Sobolev space  $W^{k,p}(\mathcal{D}; \mathbb{R})$  is the space of functions  $f : \mathcal{D} \rightarrow \mathbb{R}$ , defined by

$$W^{k,p}(\mathcal{D}; \mathbb{R}) = \{f : \|D^\alpha f\|_p < \infty, |\alpha| \leq k\}.$$

They have the associated norm

$$\|f\|_{k,p} = \left( \sum_{|\alpha| \leq k} \|D^\alpha f\|_p^p \right)^{1/p}.$$

The special case of  $W^{k,2}(\mathcal{D}; \mathbb{R})$  is a Hilbert space and is denoted by  $H^k(\mathcal{D}; \mathbb{R})$ . For functions which vanish on  $\partial\mathcal{D}$ , these are denoted by  $H_0^k(\mathcal{D}; \mathbb{R})$ .

The weak form of (2.6) is now derived. From here on in it is assumed  $f \in L^2(\mathcal{D}; \mathbb{R})$ . First, take



(2.6), multiply with some test function  $v \in C_0^\infty(\mathcal{D})$ , and integrate by parts

$$\int_{\mathcal{D}} \nabla u \cdot \nabla v \, d\mathbf{x} = \int_{\mathcal{D}} f v \, d\mathbf{x}. \quad (2.7)$$

The Dirichlet conditions result in terms due to the boundaries vanishing. The right hand side is the  $L^2(\mathcal{D}; \mathbb{R})$  inner product  $\langle f, v \rangle$ . Next, it is assumed that  $u \in H_0^1(\mathcal{D}; \mathbb{R})$  and  $v \in H_0^1(\mathcal{D}; \mathbb{R})$ <sup>1</sup>. From (2.6) a bilinear form  $\mathcal{A}_\Lambda : H_0^1 \times H_0^1 \rightarrow \mathbb{R}$  is then defined,

$$\mathcal{A}_\Lambda(u, v) := \int_{\mathcal{D}} \nabla u \cdot \nabla v \, d\mathbf{x}, \quad (2.8)$$

which is generated from  $L_\Lambda$ . Pulling all this together we define the weak problem for the Poisson equation: find  $u \in H_0^1(\mathcal{D}; \mathbb{R})$  such that for all  $v \in H_0^1(\mathcal{D}; \mathbb{R})$ ,  $\mathcal{A}_\Lambda(u, v) = \langle f, v \rangle$ . Through the Lax-Milgram theorem unique solutions are guaranteed: the problem is well-posed on function space (Evans, 2010).

To discretise the problem, the *finite element* is first introduced. A finite element is a triple  $(D, \mathcal{P}, \mathcal{N})$ , where  $D \subset \mathcal{D}$  is a bounded closed set with a smooth boundary,  $\mathcal{P}$  is a finite-dimensional space of functions, defined on  $D$ , and  $\mathcal{N} = (N_1, \dots, N_m)$  is a set of *nodal values* that are defined on the dual space  $\mathcal{P}^*$ . A useful basis is the *nodal basis* of  $\mathcal{P}$ ,  $\{\phi_1, \dots, \phi_m\}$ , that is dual to  $\mathcal{N}$ . This nodal basis is defined by  $N_i(\phi_j) = \delta_{ij}$  and using the nodal basis we define the *local interpolant* inside of the finite element  $(D, \mathcal{P}, \mathcal{N})$  as

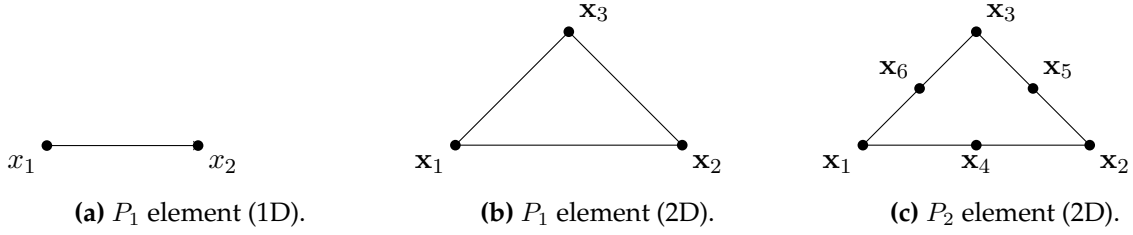
$$I_D v = \sum_{i=1}^m N_i(v) \phi_i.$$

Some example finite elements are as follows (see also Chapter 3 from Brenner and Scott (2008) for more examples). Their domains and nodal locations are shown in Figure 2.1.

1. Consider  $D = [x_1, x_2]$  and suppose that  $\mathcal{N} = (N_1, N_2)$ ,  $N_1(v) = v(x_1)$ ,  $N_2(v) = v(x_2)$ , for some function  $v$ . It can be shown that the basis functions  $\{\phi_1, \phi_2\}$  defined by  $\mathcal{N}$  are the linear polynomials  $\phi_1(x) = (x - x_2)/(x_1 - x_2)$ ,  $\phi_2(x) = (x_2 - x)/(x_2 - x_1)$ . As  $\mathcal{P} = \text{span}\{\phi_1, \phi_2\}$ , the function space  $\mathcal{P}$  is the space of piecewise linear polynomial functions on  $[x_1, x_2]$ . This is known as the  $P_1$  element in 1D.
2. Consider some triangle  $D \subset \mathbb{R}^2$  with nodal points  $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ . Setting  $\mathcal{N} = (N_1, N_2, N_3)$ , where  $N_i(v) = v(\mathbf{x}_i)$  for some function  $v$ , gives  $\mathcal{P}$  as the space of piecewise continuous linear polynomials over  $D$ . This is known as a the  $P_1$  element in 2D.
3. Over the same triangle  $D$  with the nodal points  $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$  and midpoints  $\{\mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6\}$  then  $\mathcal{N} = (N_1, \dots, N_6)$ , where  $N_i(v) = v(\mathbf{x}_i)$ , gives  $\mathcal{P}$  as the space of piecewise quadratic functions over  $D$ . This is the  $P_2$  element in 2D.

As the dimensionality of  $\mathcal{N}$  is increased the order of the space of polynomials given in  $\mathcal{P}$  also increases. For degree  $k$  polynomials in  $\mathcal{P}$ , it is required that  $\dim \mathcal{P} = (k + 1)(k + 2)/2$ , to give the  $P_k$  elements.

<sup>1</sup>These spaces are known as the *test* and *trial* spaces, respectively.



**Figure 2.1:** Finite elements: the piecewise linear  $P_1$  interval element, the piecewise linear  $P_1$  triangular element, and the piecewise quadratic  $P_2$  triangular element.

Finite elements are the objects that the finite element discretisation is constructed from. However, some additional structure is needed to combine these over the PDE domain  $\mathcal{D}$ . First, a *subdivision* is required, which is a finite collection of finite element domains  $D_i$  such that  $\text{int}D_i \cap \text{int}D_j = \emptyset$  if  $i \neq j$  and  $\bigcup_i D_i = \mathcal{D}$ . The subdivision is known as the *mesh* in the finite element literature. A particular case of a subdivision, and the one that we use in this thesis, is a *triangulation*. The triangulation  $\mathcal{D}_h$  is a subdivision of  $\mathcal{D}$  that consists of triangles  $D_i$  in which no vertex of any triangle lies in the interior of an edge of another triangle. The parameter  $h$  is the maximal length of any side of the triangulation  $\mathcal{D}_h$  and parameterises the degree of mesh-refinement.

On each triangle  $D_i$  there is a finite element,  $(D_i, \mathcal{P}_i, \mathcal{N}_i)$ , and from this set of finite elements the *global interpolator* can be constructed. The global interpolator  $I_h$  is defined from the combination of all local interpolators and, for an arbitrary function  $f$ , is given by

$$I_h f|_{D_i} = I_{D_i} f,$$

where  $f|_{D_i}$  denotes the restriction of the function  $f$  to the subdomain  $D_i$ . Across the domain  $\mathcal{D}_h$  it constructs an approximation to  $v$  from the nodal values  $\mathcal{N}_i$ . To simplify the derivations in the following chapters we write the global interpolant as

$$I_h f(\mathbf{x}) = \sum_{j=1}^n f(\mathbf{x}_j) \phi_j(\mathbf{x}), \quad (2.9)$$

where  $\{\mathbf{x}_j\}_{j=1}^n$  is the set of all unique nodes defined from the set of element nodal values  $\mathcal{N}_i$  and their respective evaluation functionals.

The basis functions  $\{\phi_j\}_{j=1}^n$  are the piecewise combinations of the nodal basis functions which satisfy  $\phi_j(\mathbf{x}_i) = \delta_{ij}$ ; these are typically the piecewise continuous polynomials of degree  $r$ . They are related to the local basis functions on an arbitrary finite element  $(D_i, \mathcal{P}_i, \mathcal{N}_i)$  through the *local-to-global* mapping (Langtangen and Mardal, 2019). The local-to-global mapping  $q(i, j)$  takes an element number  $i$  and local basis function index  $j$  and maps into the set of  $\{1, \dots, n\}$ , to give the basis function's index in the global index set. This allows for the global interpolant on a subdomain  $D_i$  to be written as

$$I_h f|_{D_i} = \sum_{j=1}^m f(\mathbf{x}_{q(i,j)}) \phi_{q(i,j)}(\mathbf{x})|_{D_i}.$$

We now return to the weak form of (2.7). Taking the  $P_1$  finite elements over the triangulation

$\mathcal{D}_h$  as the discretisation of the domain gives the trial function space  $V_h = \text{span}\{\phi_j\}_{j=1}^n$ . This defines the finite-dimensional test and trial spaces. An additional requirement is added, so that functions in  $V_h$  are those which vanish on the boundary  $\partial\mathcal{D}_h$  and so  $V_h \subset H_0^1(\mathcal{D}; \mathbb{R})$ . From (2.9) we write the global finite element approximations to  $u(\mathbf{x})$  and  $v(\mathbf{x})$  as  $u(\mathbf{x}) \approx u_h(\mathbf{x}) = \sum_{j=1}^n u_j \phi_j(\mathbf{x})$  and  $v(\mathbf{x}) \approx v_h(\mathbf{x}) = \sum_{j=1}^n v_j \phi_j(\mathbf{x})$  respectively. Substituting these functions into the weak form gives  $\mathcal{A}_\Lambda(u_h, v_h) = \langle f, v_h \rangle$  which expands to

$$\sum_{i=1}^n u_i \int_{\mathcal{D}_h} \nabla \phi_i \cdot \nabla v_h \, d\mathbf{x} = \int_{\mathcal{D}_h} f v_h \, d\mathbf{x}, \quad \forall v_h \in V_h.$$

As this must hold for all  $v_h \in V_h$ , this condition is equivalent to holding for all basis functions  $\phi_j$ , for  $j = 1, \dots, n$ , so

$$\sum_{i=1}^n u_i \int_{\mathcal{D}_h} \nabla \phi_i \cdot \nabla \phi_j \, d\mathbf{x} = \int_{\mathcal{D}_h} f \phi_j \, d\mathbf{x}, \quad \text{for } j = 1, \dots, n.$$

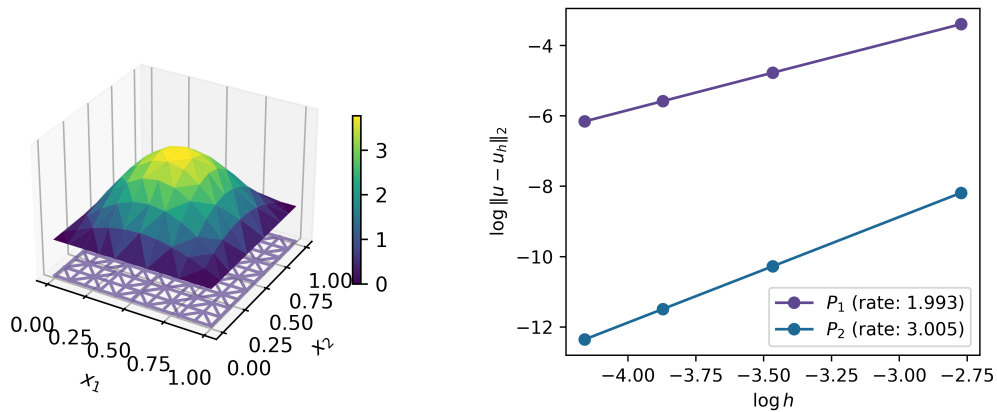
This is a finite-dimensional linear system  $\mathbf{A}\mathbf{u} = \mathbf{b}$ , with  $\mathbf{A}_{ji} = \int_{\mathcal{D}_h} \nabla \phi_i \cdot \nabla \phi_j \, d\mathbf{x}$  and  $\mathbf{b}_j = \int_{\mathcal{D}_h} f \phi_j \, d\mathbf{x}$ . The vector of FEM coefficients  $\mathbf{u} = (u_1, \dots, u_n)$  is to be solved for.

As an aside, as  $V_h \subset H_0^1(\mathcal{D}; \mathbb{R})$  we also have  $\mathcal{A}_\Lambda(u, v_h) = \langle f, v_h \rangle$ . Thus  $\int_{\mathcal{D}} \nabla(u - u_h) \cdot \nabla v_h \, d\mathbf{x} = 0$ , for all  $v_h \in V_h$ . The solution  $u_h$  is the orthogonal projection of  $u$  into  $V_h$  with respect to the Dirichlet inner product.

The construction of the *stiffness matrix*  $\mathbf{A}$  and *load vector*  $\mathbf{b}$  is done elementwise as the integral can be decomposed via  $\mathbf{A}_{ij} = \sum_{k=1}^n \mathbf{A}_{ij}^k$ , with  $\mathbf{A}_{ij}^k = \int_{D_k} \nabla \phi_i \cdot \nabla \phi_j \, d\mathbf{x}$ . The individual elements  $D_i$  are iterated over to assemble the matrix  $\mathbf{A}$ . Due to the delta property of the basis functions,  $\mathbf{A}_{ij}^k$  is nonzero if and only if the  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are nodes in element  $D_k$  (Langtangen and Mardal, 2019). This results in the matrix  $\mathbf{A}$  being sparse. The load vector is assembled in a similar elementwise fashion. Typically some sort of numerical quadrature routine (e.g. Gauss quadrature) is used to evaluate the integrals.

Two main approaches are available to solve the linear system  $\mathbf{A}\mathbf{u} = \mathbf{b}$ : direct and iterative methods. Direct methods compute a factorisation of the form  $\mathbf{A} = \mathbf{L}\mathbf{U}$ , where  $\mathbf{L}$  and  $\mathbf{U}$  are lower and upper triangular matrices, respectively, and then solve  $(\mathbf{L}\mathbf{U})^{-1} \mathbf{b}$  using forward and backward substitution, respectively (Davis, 2006). If  $\mathbf{A}$  is symmetric and positive definite, as is the case here, then this factorisation is the Cholesky decomposition  $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$ . Up to numerical precision this computes the exact solution to the linear system. Direct approaches tend to use more memory than iterative approaches, but are more reliable and give similar performance for low to medium size problems. They are also preferred when solving for multiple right hand sides (i.e. a system of the form  $\mathbf{A}\mathbf{U} = \mathbf{B}$ ) as the factorisation only needs to be computed once; the factors are reused to run forward and back substitution.

Iterative methods take a different approach and construct an approximate solution — up to some specified tolerance — through an iterative procedure (Saad, 2003). Two popular approaches are the conjugate gradient (CG) method for symmetric, positive definite systems, and the generalised minimum residual method (GMRES) for nonsymmetric systems. Iterative solutions are constructed through taking matrix-vector products (“matvecs”) of the form  $\mathbf{A}\mathbf{x}$ , which in this case are efficient due to the sparsity of  $\mathbf{A}$ . Note also that  $\mathbf{A}$  need not be assembled; it only needs to be available as a function that computes the mapping  $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$ . This allows, in implementation, for the abstract



(a) FEM solution (shown as surface) and regular mesh, over the solution domain  $\mathcal{D} = [0, 1] \times [0, 1]$ . (b) Verification of  $L_2(\mathcal{D}; \mathbb{R})$  error under mesh refinement, as a function of the maximal edge-length  $h$ . Shown also are the estimated rates of convergence for the  $P_1$  and  $P_2$  elements.

**Figure 2.2:** Poisson equation example: finite element discretisation results.

representation of these operators without explicit storage. Iterative approaches are preferred for large systems as factorisations may not be able to be stored in memory due to fill-in<sup>2</sup>. However, convergence is typically determined by the condition number of  $\mathbf{A}$  and if the matrix is poorly conditioned (as is the case for FEM) then a *preconditioner* may be required. A preconditioner  $\mathbf{M}$  is applied to both sides, giving the equivalent problem  $\mathbf{MAu} = \mathbf{Mb}$ <sup>3</sup>. The idea is to choose  $\mathbf{M}$  to be an approximation to the inverse of  $\mathbf{A}$  that is able to be easily applied to a vector. The composition of the operators  $\mathbf{MA}$  is thus better-conditioned than the original  $\mathbf{A}$  and so the iterative approaches converge faster. Popular preconditioning approaches include, among others, incomplete factorisations, sparse approximate inverses, and multigrid methods (Saad, 2003).

Given that the FEM constructs a numerical approximation to the solution of (2.6), there will be some error between the approximate and exact solutions. The error will be a function of how refined the mesh is, parameterised by  $h$ , and the order of the polynomial approximation, parameterised by  $r$ . For an exact solution  $u(\mathbf{x})$  and approximate solution  $u_h(\mathbf{x}) \in V_h$ , with  $V_h$  being the finite element solution space with degree  $r$  polynomials, then, for the Poisson equation, the error rate can be derived to be (see, e.g., Brenner and Scott (2008); Thomée (2006))

$$\|u - u_h\|_2 \leq Ch^{r+1} \|u\|_{r+1,2}. \quad (2.10)$$

We now show an example of finite element modelling, in which this rate of convergence is illustrated.

**Example 1.** Consider the specific example of a Poisson equation over the unit square with Dirichlet

<sup>2</sup>Fill-in is where zero entries in the original sparse matrix end up being set to some non-zero value upon factorisation. This increases the memory requirements.

<sup>3</sup>Known as *left preconditioning*.

boundaries and inhomogeneous forcing:

$$\begin{cases} -\Delta u = 8\pi^2 \sin(\pi x_1) \sin(\pi x_2), \\ u = 0, \quad \text{for } \mathbf{x} \in \partial\mathcal{D}, \\ u := u(\mathbf{x}), \quad \mathbf{x} \in \mathcal{D} = [0, 1] \times [0, 1]. \end{cases} \quad (2.11)$$

This equation has the exact solution  $u(x_1, x_2) = 4 \sin(\pi x_1) \sin(\pi x_2)$ .

In this example, we use the  $P_1$  elements over a uniform mesh with  $10 \times 10$  cells. Cells in this case are the individual triangular elements that make up the triangulation; this mesh is shown on the  $(x_1, x_2)$  plane in Figure 2.2a. The estimated solution  $u_h(\mathbf{x})$  is shown in Figure 2.2a. At the furthest distance away from the (fixed to zero) boundary  $\partial\mathcal{D}$ , the approximate solution reaches its maximum due to the positive forcing term. The solution demonstrates the balance between the zero BC's, which behave like a sink, and the source that results from the forcing function. The error rates are shown in Figure 2.2b. The  $P_1$  elements give  $r + 1 \approx 2$  and the  $P_2$  elements give  $r + 1 \approx 3$ . These results accord with the convergence result of (2.10) with the substitution of  $r = 1$  and  $r = 2$  for the  $P_1$  and  $P_2$  elements, respectively.

## 2.2 Probability theory and stochastic differential equations

NL-statFEM updates the FEM discretised PDE solutions with observed data via a probabilistic approach. Section 2.2.1 therefore introduces the basic machinery of probability theory which will be used throughout the rest of this thesis. Following this, we introduce SDEs in Section 2.2. In NL-statFEM, the FEM discretisation of the assumed stochastic physical model results in an SDE over the finite element coefficients. Thus, we introduce the necessary material in order to write out these induced SDEs, and we cover their discretisation with Euler-Maruyama methods. These temporal discretisation methods are used in the remainder of this thesis, and we therefore include some discussion on their convergence properties as well as their stability.

### 2.2.1 Probability theory

Probability theory is the appropriate calculus to manipulate the state of knowledge of a system as it is refined or updated through conditional relationships between variables. This state of knowledge is quantified via a probability measure (Jaynes, 2003), a component of the *probability space*<sup>4</sup>. A probability space is a triple  $(\Omega, \mathcal{F}, P)$ , consisting of the sample space  $\Omega$ , a  $\sigma$ -field  $\mathcal{F}$ , and a probability measure  $P$ . The sample space consists of the space of possible outcomes of some random experiment, with atoms  $\omega$ . The  $\sigma$ -field is the set of all subsets of  $\Omega$  such that (i)  $\Omega \in \mathcal{F}$ ; (ii)  $A \in \mathcal{F} \implies A^c \in \mathcal{F}$ ; and (iii)  $A_1, A_2, \dots \in \mathcal{F} \implies \bigcup_i A_i \in \mathcal{F}$ . Note that  $A^c$  denotes the *complement* of the set  $A$ . The elements of  $\mathcal{F}$  are known as *measurable sets*. The probability measure  $P : \mathcal{F} \rightarrow [0, 1]$  assigns a probability to the events in  $\mathcal{F}$ , satisfying (i)  $P(\Omega) = 1$ ; and (ii)  $P(\bigcup_i A_i) = \sum_i P(A_i)$  for disjoint  $A_i$ .

A *random variable*  $X : \Omega \rightarrow E$  is a *measurable function* that maps samples  $\omega$  to the output space  $E$ . This output space is most often the real line  $\mathbb{R}$  or the Euclidean space  $\mathbb{R}^n$ . To define a measurable

<sup>4</sup>For some general references to this material see Billingsley (2012); Durrett (2019); Williams (1991).

function we require first the idea of a general  $\sigma$ -field over sets with some topology. On  $\mathbb{R}$  we introduce the Borel  $\sigma$ -field  $\mathcal{B}$ , the  $\sigma$ -field generated by the open sets of  $\mathbb{R}$ . For a general space  $E$ , denote by  $\mathcal{B}(E)$  the Borel  $\sigma$ -field, which is the  $\sigma$ -field generated by the open sets of  $E$ . The random variable  $X$  being measurable thus means that, for all  $A \in \mathcal{B}(E)$ ,  $X^{-1}(A) = \{\omega : X(\omega) \in A\} \in \mathcal{F}$ .

A random variable  $X$  is distributed according to a law  $\mathcal{L} := P \circ X^{-1}$  which is a probability measure on  $(E, \mathcal{B}(E))$  (and thus  $\mathcal{L} : \mathcal{B}(E) \rightarrow [0, 1]$ ). The law is completely defined by the *distribution function*  $F : E \rightarrow [0, 1]$ , where  $P(\omega : X(\omega) \leq c) = F(c)$ . We use the notation that  $X \sim \mathcal{L}$  to state that  $X$  is distributed according to the law  $\mathcal{L}$ . As an example of this notation, two common distributions that are used throughout this thesis are the univariate Gaussian distribution  $X \sim \mathcal{N}(m, \sigma^2)$  with mean  $m$  and variance  $\sigma^2$ , and the multivariate Gaussian distribution  $\mathbf{X} \sim \mathcal{N}(\mathbf{m}, \mathbf{K})$  with mean  $\mathbf{m}$  and covariance matrix  $\mathbf{K}$ .

The *expected value* of a random variable  $X$  is the Lebesgue integral<sup>5</sup>

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) \, dP(\omega),$$

and the  $q^{\text{th}}$  *moment* is

$$\mathbb{E}[X^q] = \int_{\Omega} X^q(\omega) \, dP(\omega).$$

The spaces of random variables that have finite  $q^{\text{th}}$  moments are known as the  $L^q(P)$  random variables. These spaces are the same as those defined for functions in the previous section, except that in this case integration is done with respect to a probability measure and not the Lebesgue measure

$$L^q(P) = \left\{ X : \left( \int_{\Omega} |X(\omega)|^q \, dP(\omega) \right)^{1/q} < \infty \right\}.$$

We say that  $\{X_n(\omega)\}_n$  converges to  $X(\omega)$  in the  $L^q(P)$  sense if  $\lim_{n \rightarrow \infty} \int (X_n(\omega) - X(\omega))^q \, dP(\omega) = \lim_{n \rightarrow \infty} \mathbb{E}[(X_n(\omega) - X(\omega))^q] = 0$ .

The *Radon-Nikodym derivative* of two measures  $Q$  and  $P$ , defined on the same measurable space  $(\Omega, \mathcal{F})$ , is the function  $f$  such that

$$Q(A) = \int_A f(\omega) \, dP(\omega).$$

If  $P(A) = 0 \implies Q(A) = 0$ , for  $A \in \mathcal{F}$  then the positive measurable function  $f$  exists and is known as the Radon-Nikodym derivative. For shorthand, where needed we will write  $\frac{dQ}{dP} = f$ . For random variables  $X : \Omega \rightarrow \mathbb{R}^n$ , which are measurable on the space  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ , then the *probability density function*  $p : \mathbb{R}^n \rightarrow \mathbb{R}$  is the Radon-Nikodym derivative of the law with respect to Lebesgue measure over  $\mathbb{R}^n$

$$P(\omega : X(\omega) \in A) = \mathcal{L}(A) = \int_A p(\mathbf{x}) \, d\mathbf{x}.$$

For shorthand the terms density and measure are used interchangeably. In this vein, the notations  $X \sim \mathcal{L}$  and  $p(X) = \mathcal{L}$  are also used interchangeably throughout, to denote that the random variable  $X$  is distributed according to  $\mathcal{L}$  with density  $p(\cdot)$ .

<sup>5</sup>In this we use the Lebesgue integral to refer to integration with respect to any measure; it should be clear by context the measure to which we are integrating against.



An important class of measures are the *Gaussian measures*, which are probability measures that can be used to conduct inference in the infinite-dimensional (function space) setting. When formulating our NL-statFEM inference problem, we will derive a Gaussian measure to represent the state of knowledge of the system in the absence of data. Thus, the definition of Gaussian measure is included.

**Definition 6.** Consider the Hilbert space  $H$  and suppose that  $X \sim \mu$ . Then the measure  $\mu$  is Gaussian if for any element  $b \in H^*$ ,<sup>6</sup>  $b(X) \sim \mathcal{N}(m_b, \sigma_b^2)$ , for  $m_b \in \mathbb{R}$ ,  $\sigma_b \in \mathbb{R}$ .

This definition does not require  $X$  to be a Hilbert space, but it is included as, in this thesis, Hilbert spaces are the only function spaces that we work with. This definition follows through when considering the finite dimensional case, in which Gaussian measure is given by a multivariate Gaussian.

**Example 2.** Take  $H := \mathbb{R}^n$  and let  $\mathbf{x} \sim \mu = \mathcal{N}(\mathbf{m}, \mathbf{K})$  be a multivariate Gaussian. We show that this does indeed define an Gaussian measure. If we consider that  $H$  is the space of column vectors, the dual space,  $H^*$ , consists of the space of row vectors. Thus, the functional  $b(\mathbf{x})$  can be written as  $b(\mathbf{x}) = \mathbf{a}^\top \mathbf{x}$ , for some  $\mathbf{a} \in \mathbb{R}^n$ . Therefore,  $\mathbf{a}^\top \mathbf{x} \sim \mathcal{N}(\mathbf{a}^\top \mathbf{m}, \mathbf{a}^\top \mathbf{K} \mathbf{a})$ , due to being an affine mapping of a Gaussian, so that  $m_b = \mathbf{a}^\top \mathbf{m}$  and  $\sigma_b^2 = \mathbf{a}^\top \mathbf{K} \mathbf{a}$ . Positive definiteness of  $\mathbf{K}$  ensures that  $\sigma_b^2$  is also positive, so  $\mu$  is a Gaussian measure.

Finally, we mention the Dirac measure. This is given by the function  $\delta_x(\cdot) : \Omega \rightarrow \{0, 1\}$ , such that  $\delta_x(A) = 0$  if  $x \notin A$  and  $\delta_x(A) = 1$  if  $x \in A$ .

## 2.2.2 Stochastic Differential Equations

In this thesis, inference over the discretised PDE solutions mostly takes place in the time-evolving setting. For our NL-statFEM the model formulation results in systems of stochastic differential equations (SDEs) which describe the dynamics over the finite element coefficients. In this section the background SDE material is developed, building upon the basic foundations of probability theory. For further details see Øksendal (2003); Särkkä and Solin (2019), from which this material is based upon.

SDEs are ODEs that include stochastic terms<sup>7</sup>. Stochasticity is most commonly found through additive or multiplicative noise, in which an ordinary differential equation is augmented to include some noisy additive or multiplicative term. Informally, a simple scalar SDE may be written as

$$\frac{du}{dt} = f(u, t) + b(u, t)w(t),$$

where  $\{w(t)\}_{t \geq 0}$  is a mean-zero stochastic process that has delta correlations:  $\mathbb{E}[w(t)w(t')] = \delta(t - t')$ <sup>8</sup>. The function  $f(\cdot, \cdot)$  is known as the *drift coefficient* and the function  $b(\cdot, \cdot)$  is known as the *diffusion coefficient*. The solution  $u(t)$ , for  $t > 0$ , can be obtained through integrating (2.2.2) with respect to time

$$u(t) = u(0) + \int_0^t f(u(s), s) ds + \int_0^t b(u(s), s)w(s) ds,$$

<sup>6</sup>The notation  $H^*$  denotes the dual space of  $H$ . For  $H$  defined over the scalar field  $\mathbb{R}$ , the dual space is the space of functionals  $\phi \in H^*$  such that  $\phi(x) \in \mathbb{R}$  for  $x \in H$

<sup>7</sup>In full generality this can be extended to PDEs, however this theory is not required for NL-statFEM and is beyond the scope of this thesis.

<sup>8</sup>This is somewhat pathological: this process is not even measurable (Øksendal, 2003) however it does provide the basic motivation for the desired properties of the stochastic process under consideration

yet the second integral is problematic. Approaching it as a Riemann integral gives the approximation

$$\int_0^t w(s) ds = \lim_{N \rightarrow \infty} \sum_{n=0}^{N-1} w(t_n^*)(t_{n+1} - t_n),$$

where  $0 = t_0 < t_1 < \dots < t_N = t$ , and  $t_n^* \in [t_n, t_{n+1}]$ . Recall that the “upper sums” consist of those sums where the term  $w(t_n^*)$  is its maximum point inside  $[t_n, t_{n+1}]$ , and the “lower sums” are those in which we take  $w(t_n^*)$  to be the minimum point inside  $[t_n, t_{n+1}]$ . The Riemann integral converges if these two sums converge to the same value. In the case of the white noise process this process is unbounded (having infinite variance) so these two sums do not converge.

To resolve this a *stochastic integral* is needed, which integrates with respect to a stochastic process in a similar fashion to the Riemann-Stieltjes integral. To do so, however, the Brownian motion process needs to be defined, which is the stochastic process against which we will integrate.

**Definition 7.** *The Brownian motion process<sup>9</sup>  $\beta : \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}^n$  is a stochastic process which is continuously indexed by time  $t \in \mathbb{R}_+$ . We write the Brownian motion as  $\beta(t, \omega)$ , however where dependency on  $\omega$  is superfluous we write  $\beta(t, \omega) \equiv \beta(t)$ . Brownian motion has the following properties:*

1.  $P(\beta(\mathbf{0}) = \mathbf{0}) = 1$  (starts at  $\mathbf{0}$ ).
2.  $\beta(t + \Delta_t) - \beta(t) \perp \beta(t)$  (independent increments).
3.  $\beta(t + \Delta_t) - \beta(t) \sim \mathcal{N}(\mathbf{0}, \Delta_t \mathbf{G})$  (Gaussian increments).

Given a single element  $\omega$ ,  $\beta(t, \omega)$  defines a sample path indexed with  $t$ . For a given  $t$ ,  $\beta(t, \omega)$  is a random variable. The matrix  $\mathbf{G}$  is sometimes referred to as the diffusion matrix (Särkkä and Solin, 2019).

Some motivation for the Riemann-Stieltjes approach can be obtained through discretising (2.2.2) using a standard Euler discretisation (Hairer et al., 1993):

$$u_{n+1} - u_n = \Delta_t f(u_n, t_n) + \Delta_t b(u_n, t_n) w_n,$$

where  $u_n = u(t_n)$ ,  $w_n = w(t_n)$ ,  $t_{n+1} - t_n = \Delta_t$ . Next, we want to replace  $\Delta_t w_n$  with some stochastic process that has a mean of zero and independent draws. One such approach is to therefore replace  $\Delta_t w_n = B_{n+1} - B_n$ , so that the white noise is the increments of a Brownian motion process  $B(t, \omega)$ . Summing over all times gives the approximation

$$u(t) = u(0) + \sum_{n=1}^N \Delta_t f(u(t_{n-1}), t_{n-1}) + \sum_{n=1}^N b(u(t_{n-1}), t_{n-1})(B_n - B_{n-1})$$

which motivates that as  $N \rightarrow \infty$ ,

$$u(t) = u(0) + \int_0^t f(u(s), s) ds + \int_0^t b(u(s), s) dB(s),$$

<sup>9</sup>In the case in which  $\beta \in \mathbb{R}$  with variance  $\mathbf{G} \equiv 1$  is known as the *standard Brownian motion*.



through replacing the sums with integrals. However this is still not able to be realised, as the Riemann-Stieltjes integral with respect to Brownian motion also does not exist. This is due to the limit

$$\int_0^t b(u(s), s) dB(s) = \lim_{N \rightarrow \infty} \sum_{n=1}^N b(u(t_n^*), t_n^*)(B_n - B_{n-1})$$

being dependent on the value of  $t_n^*$  on which is taken — this cannot be the case for the Riemann-Stieltjes integral. To make this limit unique fix  $t_n^* = t_{n-1}$  and define the Itô stochastic integral as

$$\int_0^t b(u(s), s) dB(s) = \lim_{N \rightarrow \infty} \sum_{n=1}^N b(u(t_{n-1}), t_{n-1})(B_n - B_{n-1}),$$

where the limit is taken in the  $L^2(P)$  sense. The Itô integral provides the motivation for expressing the SDE (2.2.2) in the shorthand

$$du(t) = f(u, t) dt + b(u, t) dB(t), \quad (2.12)$$

which implies the integral equation

$$u(t) = u(0) + \int_0^t f(u, t) dt + \int_0^t b(u, t) dB(t).$$

For the vector-valued case the construction is the same, replacing the standard Brownian motion with its vector-valued counterpart.

To discretise (2.12) we use three standard schemes throughout the remainder of this thesis. The first has already been introduced, being the *explicit Euler* or *Euler-Maruyama* method (Higham, 2001; Kloeden and Platen, 1992), which is given by

$$u_{n+1} - u_n = \Delta_t f(u_n, t_n) + \Delta_t b(u_n, t_n) \Delta B_n,$$

where  $\Delta B_n = B(t_{n+1}) - B(t_n) \sim \mathcal{N}(0, \Delta_t)$ . The *implicit Euler* can also be used

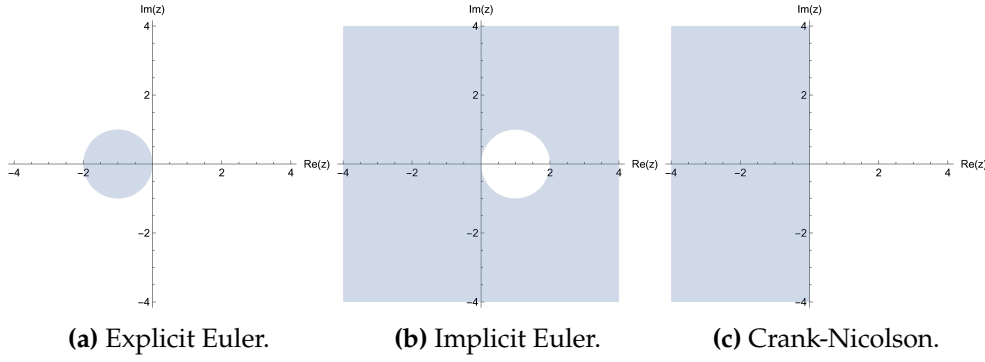
$$u_{n+1} - u_n = \Delta_t f(u_{n+1}, t_{n+1}) + \Delta_t b(u_n, t_n) \Delta B_n,$$

so too the *Crank-Nicolson*

$$u_{n+1} - u_n = \Delta_t f(u_{n+1/2}, t_{n+1/2}) + \Delta_t b(u_n, t_n) \Delta B_n,$$

where  $u_{n+1/2} = (u_{n+1} + u_n)/2$ . These schemes are at most *semi-implicit* as we do not treat the noise terms implicitly. As noted in Kloeden and Platen (1992), due to fully-implicit SDE schemes often requiring reciprocals of Gaussians (which do not have finite second moments) we restrict implicit schemes to being implicit in the drift component only (i.e. the component arising from the PDE discretisation).

We also require notions of how the discretisation converges to the true process as the discretisation interval  $\Delta_t$  is refined. Consider the *time-discrete approximation*  $u_\Delta(t)$ , the linear interpolation of the simulation timepoints  $\{u_n\}_{n=1}^N$ . This is itself a stochastic process. A time-discrete approximation  $u_\Delta(t)$



**Figure 2.3:** Time integrator stability regions for those used in this thesis.

is said to converge *weakly* with order  $\beta$  if for all polynomials  $g$  there is a  $C > 0$  so that

$$|\mathbb{E}[g(u(t)) - g(u_\Delta(t))]| \leq C\Delta t^\beta.$$

The time-discrete approximation is said to converge *strongly* with order  $\beta$  if

$$\mathbb{E}[|u(t) - u_\Delta(t)|] \leq C\Delta t^\beta.$$

Weak convergence only ensures that expectations of functionals of the time-discrete approximation converge at the rate  $\beta$ ; for error rates on the time-discrete approximation itself, the notion of strong convergence is required. Weak convergence relaxes the requirements that the time-discrete approximations are pathwise accurate, and can allow for non-Gaussian noise simulations which preserve the Gaussian moments. All schemes that we consider in this thesis have a strong order of convergence of 0.5 and a weak order of convergence of 1.

These schemes can also be studied through their *stability*, which describes how we expect the numerical scheme to behave for large time. This proceeds similarly to that of  $A$ -stability in the deterministic setting (Hairer et al., 1993), and begins with the complex-valued scalar test SDE

$$dx(t) = \lambda x dt + dB(t), \quad \lambda \in \mathbb{C}. \quad (2.13)$$

Assuming that (2.13) is discretised with a uniform timestep  $\Delta_t$  gives a recursive updating scheme of the form

$$x_n = G(\lambda\Delta_t)x_{n-1} + \Delta B_n, \quad \Delta B_n \sim \mathcal{N}(0, \Delta_t),$$

for some  $G : \mathbb{R} \rightarrow \mathbb{R}$ . The *region of absolute stability* is  $\{z : |G(z)| \leq 1, \Re(z) < 0\}$ , building from the intuition that  $|x_n - \hat{x}_n| \leq |G(\lambda\Delta_t)|^n |x_0 - \hat{x}_0|$ , for two discretisations, starting at  $x_0$  and  $\hat{x}_0$ . Note that in this thesis we use  $\Re(\cdot)$  and  $\Im(\cdot)$  to denote the real and imaginary components of some quantity.

For the explicit Euler this region is  $\{z : |1 + z| \leq 1\}$ , for the implicit Euler this is  $\{z : 1/|1 - z| \leq 1\}$ , and for the Crank-Nicolson this is  $\{z : |1 + z/2|/|1 - z/2| \leq 1\}$ . These regions are plotted in Figure 2.3. The region of stability for both the explicit Euler is small (the unit disk centred at  $-1$ ). The stepsize is naturally constrained by  $\lambda$ . For  $\lambda < 0$  (i.e. for decaying systems) both the implicit methods are stable for all  $\Delta_t$ ; in practice we tend to prefer these methods.

## 2.3 Bayesian Inference

After deriving an SDE over the finite element coefficients we update these coefficients given a set of observed data. This can be framed as a Bayesian inference problem where the SDE implicitly defines a *prior distribution*. The update refers to the computation of the *posterior distribution*: the probability distribution of these coefficients given the dataset. For our NL-statFEM, this is done in a sequential, online fashion through leveraging the filtering algorithms introduced in Chapter 1. However in order to motivate our algorithmic choices we first discuss Bayesian inference in the static parameter setting, establishing the general Bayesian workflow and the difficulties associated with moving this workflow into the time-evolving case. We then discuss the previously introduced filtering algorithms, expanding upon the presentation given in Chapter 1. This section concludes with the required background information on Gaussian Processes, which we leverage in NL-statFEM to induce uncertainty inside of the assumed governing PDE.

We encode our assumptions about the data generating process in the *likelihood* function, which we denote by  $p(\mathbf{y} | \boldsymbol{\theta})$ . Taking a Bayesian approach means that this likelihood is combined with some *a priori* specification of knowledge about  $\boldsymbol{\theta}$ . This knowledge is encoded in a prior distribution with density  $p(\boldsymbol{\theta})$ , and is updated with data via Bayes rule to give the posterior

$$p(\boldsymbol{\theta} | \mathbf{y}) = \frac{p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathbf{y} | \boldsymbol{\theta})p(\boldsymbol{\theta}) d\boldsymbol{\theta}}. \quad (2.14)$$

The prior forms the basis for inference and the posterior describes the state of knowledge of  $\boldsymbol{\theta}$  conditioned on the assumed likelihood function and prior distribution. Modelling assumptions present in the likelihood and the prior therefore manifest in the posterior.

This general update rule was derived by Thomas Bayes (Bayes, 1763) with parallel developments by Laplace (Laplace, 1812). Each of these foundational works focused on the estimation of the unknown parameter in a Binomial model and derived expressions that resemble the current presentation of Bayes theorem. Modern Bayesian methods have grown from the foundational texts of Lindley (1965), Savage (1972), and Jaynes (2003), among others. These texts motivate the Bayesian interpretation of probability as being one of a *degree-of-belief* in the state of some system, thus permitting the combination of multiple sources of information. As data is observed, *a priori* beliefs are updated to give an *a posteriori* characterisation of the state of knowledge. As such a wider variety of probabilistic scenarios can be described within a Bayesian framework (Ramsey, 1926).

Important practical contributions also came from de Finetti and Jeffreys. De Finetti provided motivation for statistical inference as proceeding from the principle of *exchangeability* (de Finetti, 1974). Exchangeability is the property that the joint distribution of a set of random objects  $\{y_j\}_{j=1}^n$  is invariant to permutations of the orderings of this set. De Finetti's theorem states that in the instance that these random variables are Bernoulli random variables, then they are exchangeable if and only if they are conditionally independent conditioned on another random variable. Exchangeability thus ensures that  $p(y_1, \dots, y_n | \theta) = \prod_{j=1}^n p(y_j | \theta)$ , guaranteeing the existence of some parameter  $\theta$  which can thus be inferred.

The *Jeffreys' prior* (Jeffreys, 1961) is a parameterisation-independent way of describing the prior state of knowledge of the system, motivating an "objective" approach to Bayesian statistics that was

free of subjectivity in the choice of the prior. The subjectivity inherent in the choice of prior remained a point of contention between Bayesian and non-Bayesian statisticians. Yet it should be noted that no method is truly objective; all statistical models require specification at some level (Gelman and Hennig, 2017).

Until the mid-1990s Bayesian models were not widely used due to difficulties in computing the posterior distribution, being relegated to cases in which the marginal likelihood,  $p(\mathbf{y}) = \int p(\mathbf{y} | \boldsymbol{\theta}) p(\mathbf{d}\boldsymbol{\theta})$ , was able to be analytically computed. This changed with the advent of Markov Chain Monte Carlo (MCMC) algorithms, which construct a Markov Chain whose stationary distribution coincides with the posterior distribution of interest (Casella and George, 1992; Chib and Greenberg, 1995; Tierney, 1994). The development and uptake of MCMC has led to the proliferation of Bayesian statistical methodology throughout academic disciplines and has led to widespread application (Faber, 2012; Gelman et al., 2013; Gill, 2002; Ivezic et al., 2014; Lesaffre and Lawson, 2012; Sharma, 2017).

MCMC forms a Markov Chain  $\{\boldsymbol{\theta}_n\}_{n=1}^{\infty}$  which as  $n \rightarrow \infty$  ensures  $\boldsymbol{\theta}_n \sim p(\boldsymbol{\theta} | \mathbf{y})$ . For a finite set of MCMC samples  $\{\boldsymbol{\theta}_n\}_{n=1}^N$  this allows for the Monte Carlo approximation of the integral via

$$\mathbb{E}[f(\boldsymbol{\theta})] = \int f(\boldsymbol{\theta}) p(\mathbf{d}\boldsymbol{\theta} | \mathbf{y}) \approx \frac{1}{N} \sum_{j=1}^N f(\boldsymbol{\theta}_j).$$

MCMC relies on the derivation of an appropriate Markov transition kernel  $\Pi(\cdot, \cdot)$ , which is defined to be a measurable function, with respect to the parameter space, in the first argument, and a probability measure in the second argument. Thus supposing we have a Markov Chain  $\{\boldsymbol{\theta}_n\}_n$  which is currently at  $\boldsymbol{\theta}$  the kernel states  $P(\boldsymbol{\theta}_{n+1} \in A | \boldsymbol{\theta}_n = \boldsymbol{\theta}) = \Pi(\boldsymbol{\theta}, A)$ .

To construct such a kernel the Metropolis-Hastings criterion can be used. This combines a proposal density  $q(\boldsymbol{\theta}_n, \boldsymbol{\theta}^*)$ , which takes the current state of the Markov Chain  $\boldsymbol{\theta}_n$  and proposes a new value  $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}_n, \cdot)$ , with a bias-adjusted correction given by

$$\alpha(\boldsymbol{\theta}_n, \boldsymbol{\theta}^*) = \min \left\{ 1, \frac{p(\boldsymbol{\theta}^* | \mathbf{y}) q(\boldsymbol{\theta}_n, \boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}_n | \mathbf{y}) q(\boldsymbol{\theta}_n, \boldsymbol{\theta}^*)} \right\}. \quad (2.15)$$

Note that the marginal likelihood cancels in (2.15) and thus the posterior distribution only needs to be known up to normalising constant. The proposed  $\boldsymbol{\theta}^*$  is accepted with probability  $\alpha(\boldsymbol{\theta}_n, \boldsymbol{\theta}^*)$  and thus  $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}^*$ , otherwise it is rejected and  $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n$ .

Under suitable conditions on  $q(\cdot, \cdot)$  and  $\alpha(\cdot, \cdot)$ <sup>10</sup> the density of the transition kernel  $\Pi(\cdot, \cdot)$  can then be defined as

$$\pi(\boldsymbol{\theta}, \mathbf{d}\boldsymbol{\theta}^*) = \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) q(\boldsymbol{\theta}, \mathbf{d}\boldsymbol{\theta}^*) + r(\boldsymbol{\theta}) \delta_{\boldsymbol{\theta}}(\mathbf{d}\boldsymbol{\theta}^*)$$

where  $r(\boldsymbol{\theta}) = 1 - \int \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) q(\boldsymbol{\theta}, \mathbf{d}\boldsymbol{\theta}^*)$ . We use the notation  $\pi(\boldsymbol{\theta}, \mathbf{d}\boldsymbol{\theta}^*)$  indicate that, in the second argument,  $\pi(\cdot, \cdot)$  is a probability density function. This methodology yields a Markov Chain  $\{\boldsymbol{\theta}_n\}_n$  which as  $n \rightarrow \infty$  converges to the true posterior. MCMC is thus an *approximately exact* method; samples drawn are guaranteed to be drawn from the true posterior in the limit as  $n \rightarrow \infty$ . Approximations arise from truncating the chain to finite time.

For a symmetric  $q(\cdot, \cdot)$  this methodology was first introduced in Metropolis et al. (1953). The

<sup>10</sup>We use a slight abuse of notation as in the case in which  $\boldsymbol{\theta} = \boldsymbol{\theta}^*$  then we set  $q(\boldsymbol{\theta}, \boldsymbol{\theta}^*) \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = 0$ . We define  $\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = 1$  if  $p(\boldsymbol{\theta} | \mathbf{y}) q(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = 0$ .

generalisation to non-symmetric  $q(\cdot, \cdot)$  was covered in [Hastings \(1970\)](#) and allowed for bespoke algorithms that have favourable scaling properties and more rapid mixing. These algorithms work to combat the drawbacks associated with MCMC sampling. These drawbacks include samples being possibly highly correlated due to the underlying Markovian transition kernel, and samplers not converging within finite-time to the measure of interest. Algorithms that may ameliorate these drawbacks include Hamiltonian Monte Carlo (HMC) ([Neal, 2012](#)), Langevin Monte Carlo ([Roberts and Tweedie, 1996](#)), and Riemann manifold HMC ([Girolami and Calderhead, 2011](#)). Each of these make use of geometric information to construct a proposal distribution that makes use of the curvature of the posterior. An alternative MCMC approach, which requires iterative sampling from the “full conditional” distributions in turn was first described in [Geman and Geman \(1984\)](#) and has since been dubbed *Gibbs sampling*. A statistical review of this approach is covered in [Casella and George \(1992\)](#).

### 2.3.1 Time-evolving inference

The Bayesian framework can also be extended into the time-evolving setting, taking  $\boldsymbol{\theta} \equiv \boldsymbol{\theta}_n$ , for the time index  $n$ . Inference in this setting focusses on estimating a sequence of posterior distributions, typically either the *filtering posteriors*  $p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n})$ , the posterior of each  $\boldsymbol{\theta}_n$  given the data up to time  $n$ , or the *smoothing posteriors*  $p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n_t})$ , the posterior of each  $\boldsymbol{\theta}_n$  given all the data up to the final time  $n_t$ . MCMC approaches can target the smoothing distribution through estimating the joint posterior  $p(\boldsymbol{\theta}_{1:n_t} | \mathbf{y}_{1:n_t})$ , however this will be computationally demanding due to high dimensionality. Also, the choice of a proposal distribution  $q(\cdot, \cdot)$  which balances both computational implementability and desirable convergence properties is unclear, and will only be able to be considered on a case-by-case basis.

To avoid these challenges, recursive approaches can be used to estimate the filtering and smoothing distributions without MCMC sampling. These are applicable for common model structures with conditionally independent observations and latent Markovian dynamics, which is the assumed structure in NL-statFEM. In brief, discretisation of the governing SDE, in NL-statFEM, results in Markovian dynamics on the finite element coefficients. Observed data, which is assumed to be observed at the discretisation timepoints, is assumed to be conditionally independent given the model values at this observation time. In this section, the inference problem for our model structure is introduced, focussing on the *approximate Gaussian filters* and their application in estimating the filtering posterior  $p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n})$ .

Data  $\mathbf{y}_n$  is arriving via some assumed data generating process and is typically corrupted with observational noise. Temporal evolution for the unknown parameters  $\boldsymbol{\theta}_n$  proceeds from a known stochastic model, represented by the first-order Markovian transition density  $\boldsymbol{\theta}_n \sim p(\boldsymbol{\theta}_n | \boldsymbol{\theta}_{n-1})$ <sup>11</sup>. Data is assumed to be arriving at time  $n$  via the observation process  $p(\mathbf{y}_n | \boldsymbol{\theta}_n)$  and the distribution over the initial condition is known,  $\boldsymbol{\theta}_0 \sim p(\boldsymbol{\theta}_0)$ . The full model is thus

$$\boldsymbol{\theta}_0 \sim p(\boldsymbol{\theta}_0), \quad \boldsymbol{\theta}_n \sim p(\boldsymbol{\theta}_n | \boldsymbol{\theta}_{n-1}), \quad \mathbf{y}_n \sim p(\mathbf{y}_n | \boldsymbol{\theta}_n), \quad (2.16)$$

for the time index  $n = 1, \dots, n_t$ . Due to the sequential nature of the problem the filtering posterior can

<sup>11</sup>Following the notation of previous section gives  $\pi(\boldsymbol{\theta}_{n-1}, d\boldsymbol{\theta}_n) \equiv p(\boldsymbol{\theta}_n | \boldsymbol{\theta}_{n-1})$ . Higher-order Markovian models can also be used; we stick with first-order models for exposition.

be computed in a recursive manner. This requires two steps: *marginalisation* and *updating*. Assuming that  $p(\boldsymbol{\theta}_{n-1} | \mathbf{y}_{1:n-1})$  is known, then the marginalisation step first integrates over the uncertainty in the previous timestep, computing

$$p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n-1}) = \int p(\boldsymbol{\theta}_n | \boldsymbol{\theta}_{n-1}) p(d\boldsymbol{\theta}_{n-1} | \mathbf{y}_{1:n-1}).$$

This distribution forms the prior for the data observed at the current timestep. This step is also known as the *prediction step* as we predict the next timestep  $n$  given the data up to time  $n - 1$ . This is then updated to give the filtering posterior, up to a normalising constant, as

$$p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n}) \propto p(\mathbf{y}_n | \boldsymbol{\theta}_n) p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n-1}).$$

Computing this posterior updates the prediction distribution  $p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n-1})$  with the data observed at the current time to give the filtering distribution  $p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n})$ .

The joint completion of both the marginalisation and update steps can be complicated as each involve a potentially intractable integral. In the marginalisation step this arises due to the potentially nonstandard distributions being used to model the stochastic dynamics  $p(\boldsymbol{\theta}_n | \boldsymbol{\theta}_{n-1})$ . These may arise from the nonlinearity in an underlying physical model, or from nonstandard noise processes. The update step encounters the standard problem of not knowing the normalising constant of the posterior density, as in (2.14).

Various methods have been proposed to overcome these difficulties (see, e.g., [Law et al., 2015](#)). We describe two particular algorithms in detail — the extended and ensemble Kalman filters — which are applicable to the model structure used in the remainder of this thesis. Discussions on other algorithms are also given to provide context and motivate our choice in using these. Our model structure is the nonlinear Gaussian model, restated here for convenience:

$$\begin{aligned} \boldsymbol{\theta}_n &= \mathcal{F}(\boldsymbol{\theta}_{n-1}) + \mathbf{q}_n, & \mathbf{q}_n &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \\ \mathbf{y}_n &= \mathbf{H}\boldsymbol{\theta}_n + \mathbf{r}_n, & \mathbf{r}_n &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \end{aligned} \tag{2.17}$$

for  $n = 1, \dots, n_t$ . We assume that  $\mathbf{q}_n$  are independent and identically distributed (i.i.d.) and that  $\mathbf{q}_n \perp \boldsymbol{\theta}_n$ , for all  $n$ . It is also assumed that  $\mathbf{r}_n$  are i.i.d. and that  $\mathbf{r}_n \perp \boldsymbol{\theta}_n$ .

The stochastic dynamics is based upon additive Gaussian error over the nonlinear model  $\mathcal{F}(\cdot)$  and gives the transition density as  $p(\boldsymbol{\theta}_n | \boldsymbol{\theta}_{n-1}) = \mathcal{N}(\mathcal{F}(\boldsymbol{\theta}_{n-1}), \mathbf{Q})$ . Observations are taken as a linear mapping of the underlying model state, which uses the *observation operator*  $\mathbf{H}$ , with additive Gaussian observational error.

Setting linear dynamics  $\mathcal{F}(\boldsymbol{\theta}_n) := \mathbf{F}\boldsymbol{\theta}_n$  means that the posterior can be computed with the Kalman filter ([Kalman, 1960](#)). The Kalman filter was the first methodology that addressed the problem of physics-data synthesis and extended the previous filtering approaches of Wiener (see, e.g., [Wiener, 1964](#)) from the frequency domain into the state space. This extension allowed for much wider application as the filter now operated in the same parameterisation as the physical model.

The Kalman filter gives a sequential updating procedure for the posterior mean  $\mathbf{m}_n$  and covariance matrix  $\mathbf{C}_n$ . Assuming  $p(\boldsymbol{\theta}_0) = \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0)$  and that the distribution on the previous state is known,



$p(\boldsymbol{\theta}_{n-1} | \mathbf{y}_{1:n-1}) = \mathcal{N}(\mathbf{m}_{n-1}, \mathbf{C}_{n-1})$ , the Kalman filter proceeds by computing the predicted mean and covariance and time  $n$

$$\hat{\mathbf{m}}_n = \mathbf{F}\mathbf{m}_{n-1}, \quad \hat{\mathbf{C}}_n = \mathbf{F}\mathbf{C}_{n-1}\mathbf{F}^\top + \mathbf{Q},$$

and then updating these predictions with data

$$\begin{aligned} \mathbf{m}_n &= \hat{\mathbf{m}}_n + \hat{\mathbf{C}}_n \mathbf{H}^\top (\mathbf{H} \hat{\mathbf{C}}_n \mathbf{H}^\top + \mathbf{R})^{-1} (\mathbf{y}_n - \mathbf{H} \hat{\mathbf{m}}_n), \\ \mathbf{C}_n &= \hat{\mathbf{C}}_n - \hat{\mathbf{C}}_n \mathbf{H}^\top (\mathbf{H} \hat{\mathbf{C}}_n \mathbf{H}^\top + \mathbf{R})^{-1} \mathbf{H} \hat{\mathbf{C}}_n. \end{aligned}$$

This gives the posterior  $p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n}) = \mathcal{N}(\mathbf{m}_n, \mathbf{C}_n)$ . The prediction-updating sequence can be derived through the affine transformation property of Gaussians and from the multivariate Gaussian conditioning formula (for further detail see [Meinhold and Singpurwalla, 1983](#)). The mean  $\mathbf{m}_n$  is the optimal estimator of the state which maximises the log-posterior density defined by (2.17),

$$\log p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n}) \propto -\frac{1}{2} (\mathbf{y}_n - \mathbf{H}\boldsymbol{\theta}_n)^\top \mathbf{R}^{-1} (\mathbf{y}_n - \mathbf{H}\boldsymbol{\theta}_n) - \frac{1}{2} (\boldsymbol{\theta}_n - \hat{\mathbf{m}}_n)^\top \hat{\mathbf{C}}_n^{-1} (\boldsymbol{\theta}_n - \hat{\mathbf{m}}_n).$$

This is a general framework which arises in many different settings. For example the matrix  $\mathbf{F}$  may be a discretised evolution operator from some high-dimensional PDE, or, a scalar coefficient<sup>12</sup>. The variety of scenarios in which this framework is applicable has led to widespread application across engineering and the physical sciences. The framework has also been extended to deal with continuous-time processes where the dynamics and observation processes are given by stochastic differential equations ([Øksendal, 2003](#)). This is the *Kalman-Bucy* filter ([Kalman and Bucy, 1961](#)), which is again optimal for linear drift and diffusion terms.

To handle the nonlinearity in (2.17) many approaches can be used, though, for those which employ Gaussian approximations, optimality of the resultant estimator can no longer be guaranteed. The extended Kalman filter (ExKF) ([Jazwinski, 2007](#)) linearises the nonlinear  $\mathcal{F}(\cdot)$  at each timestep to give Gaussian approximation to the posterior distribution. For (2.17), with  $p(\boldsymbol{\theta}_{n-1} | \mathbf{y}_{1:n-1}) = \mathcal{N}(\mathbf{m}_{n-1}, \mathbf{C}_{n-1})$ , the update step is therefore given by

$$\hat{\mathbf{m}}_n = \mathcal{F}(\mathbf{m}_{n-1}), \quad \hat{\mathbf{C}}_n = \mathbf{J}_n \mathbf{C}_{n-1} \mathbf{J}_n^\top + \mathbf{Q},$$

where  $\mathbf{J}_n = \partial \mathcal{F}(\boldsymbol{\theta}_n) / \partial \boldsymbol{\theta}_n$ , and  $p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n-1}) = \mathcal{N}(\hat{\mathbf{m}}_n, \hat{\mathbf{C}}_n)$ . The update step proceeds as for the Kalman filter, via the standard Bayes update for a Gaussian:

$$\begin{aligned} \mathbf{m}_n &= \hat{\mathbf{m}}_n + \hat{\mathbf{C}}_n \mathbf{H}^\top (\mathbf{H} \hat{\mathbf{C}}_n \mathbf{H}^\top + \mathbf{R})^{-1} (\mathbf{y}_n - \mathbf{H} \hat{\mathbf{m}}_n), \\ \mathbf{C}_n &= \hat{\mathbf{C}}_n - \hat{\mathbf{C}}_n \mathbf{H}^\top (\mathbf{H} \hat{\mathbf{C}}_n \mathbf{H}^\top + \mathbf{R})^{-1} \mathbf{H} \hat{\mathbf{C}}_n. \end{aligned} \tag{2.18}$$

This gives  $p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n}) = \mathcal{N}(\mathbf{m}_n, \mathbf{C}_n)$ . Accuracy of the approximation depends on how well the action of the nonlinear  $\mathcal{F}(\cdot)$  is approximated by the Jacobian. Scalability can be ensured through making a low-rank approximation to the covariance  $\mathbf{C}_n$  (see Chapter 5, as well as [Law and Stuart, 2012](#); [Rozier et al., 2007](#)). The mean of the ExKF gives a reasonable approximation to the true posterior

<sup>12</sup>In this case, the dynamics are given from a standard AR(1) process, e.g.  $\theta_n = \varphi \theta_{n-1} + q_n$ ,  $q_n \sim \mathcal{N}(0, Q^2)$  ([Shumway and Stoffer, 2017](#)).

mean, and for regularly observed systems the variance also provides a sensible estimate to the true variance (Law and Stuart, 2012). However computing a low-rank approximation of the covariance leads to underestimation (Gillijns et al., 2006) which can result in placing too much belief in the model predictions. To ameliorate this, ad-hoc covariance inflation methods can be used in the prediction step, multiplying the approximated covariance matrix with a scalar  $\rho > 1$  or adding a scaled identity matrix. This has the effect of increasing uncertainty in the predictions.

Sample-based approaches, such as the unscented Kalman filter (UKF) (Wan and Van Der Merwe, 2000) and the particle filter (PF) (Doucet et al., 2000), can be used if linearisation is not possible. The UKF represents  $p(\boldsymbol{\theta}_{n-1} | \mathbf{y}_{1:n-1})$  through a fixed number of particles and corresponding weights (for  $\boldsymbol{\theta}_n \in \mathbb{R}^m$ , this requires  $2m + 1$  particle/weight pairs) in the prediction step. These particles are chosen to give a close approximation to nonlinear mappings of this measure and are pushed through  $\mathcal{F}(\cdot)$  in the prediction step. The predicted mean  $\hat{\mathbf{m}}_n$  and covariance  $\hat{\mathbf{C}}_n$  are approximated through weighted sums of the resultant mapped particles; this procedure is at least second-order accurate. The Gaussian assumption is made so that as with the ExKF,  $p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n-1}) = \mathcal{N}(\hat{\mathbf{m}}_n, \hat{\mathbf{C}}_n)$ , and the update step proceeds as in (2.18).

The PF also provides an approximation to  $p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n})$ , through a set of  $N$  particles  $\{\boldsymbol{\theta}_n^{(j)}\}_{j=1}^N$  and their corresponding weights  $\{w_n^{(j)}\}_{j=1}^N$ . The PF constructs an approximate measure  $p_N(\boldsymbol{\theta}_n | \mathbf{y}_{1:n}) = \sum_{j=1}^N w_n^{(j)} \delta_{\boldsymbol{\theta}_n^{(j)}}(d\boldsymbol{\theta}_n)$ , using an importance sampling approach (Doucet et al., 2000). Particle positions  $\{\boldsymbol{\theta}_n^{(j)}\}_{j=1}^N$  are first shifted in the prediction step and the weights  $\{w_n^{(j)}\}_{j=1}^N$  are updated in the update step. Different to the UKF, the PF makes no distributional assumption and can take arbitrarily many particles: as  $N \rightarrow \infty$ ,  $p_N$  converges to the exact  $p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n})$ .

Both the UKF and PF struggle in high dimensions. The UKF requires taking and storing a matrix square-root of  $\mathbf{C}_n$  which, for high-dimensional systems, is infeasible. The PF results in weight degeneracy when  $N$  is low in comparison to the state dimension (Bengtsson et al., 2008), leading to all but one weight approaching zero.

To circumvent these degenerations in high-dimensional systems, approaches for the nonlinear filtering problem have been proposed in the data assimilation literature (Law et al., 2015). As covered previously, a standard approach is the ensemble Kalman filter (EnKF) which, similar to the UKF and PF, constructs a sample-based characterisation of the filtering posterior. Different to these approaches, however, all particles or *ensemble members* have equal weights, and the ensemble is assumed to be Gaussian distributed.

We detail the EnKF for (2.17) now. Assuming there is an ensemble with  $N_{\text{ens}}$  members, such that  $p(\boldsymbol{\theta}_{n-1} | \mathbf{y}_{1:n-1}) = \frac{1}{N_{\text{ens}}} \sum_{j=1}^{N_{\text{ens}}} \delta_{\boldsymbol{\theta}_{n-1}^{(j)}}(d\boldsymbol{\theta}_{n-1})$ , then the ensemble prediction step is given by

$$\begin{aligned} \hat{\boldsymbol{\theta}}_n^{(i)} &= \mathcal{F}(\boldsymbol{\theta}_{n-1}^{(i)}) + \boldsymbol{\xi}_n^{(i)}, \quad \boldsymbol{\xi}_n^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \quad i = 1, \dots, N_{\text{ens}}, \\ \mathbb{E}(\hat{\boldsymbol{\theta}}_n) &= \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} \hat{\boldsymbol{\theta}}_n^{(i)}, \quad \hat{\mathbf{C}}_n = \frac{1}{N_{\text{ens}} - 1} \sum_{i=1}^{N_{\text{ens}}} \left( \hat{\boldsymbol{\theta}}_n^{(i)} - \mathbb{E}(\hat{\boldsymbol{\theta}}_n) \right) \left( \hat{\boldsymbol{\theta}}_n^{(i)} - \mathbb{E}(\hat{\boldsymbol{\theta}}_n) \right)^\top, \end{aligned}$$

and the update step follows

$$\boldsymbol{\theta}_n^{(i)} = \hat{\boldsymbol{\theta}}_n^{(i)} + \hat{\mathbf{C}}_n \mathbf{H}^\top (\mathbf{H} \hat{\mathbf{C}}_n \mathbf{H}^\top + \mathbf{R})^{-1} (\mathbf{y}_n + \boldsymbol{\eta}_n - \mathbf{H} \hat{\boldsymbol{\theta}}_n^{(i)}), \quad \boldsymbol{\eta}_n^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \quad i = 1, \dots, N_{\text{ens}}.$$



This approach has been popular due to performing well in small-ensemble configurations and being amenable to parallelisation (Evensen, 2003; Mandel, 2006). For small  $N_{\text{ens}}$  relative to the state dimension the posterior mean estimates are accurate to the true mean, however the variance estimates may not reflect the variance of the true posterior in temporally sparse observation regimes (Law and Stuart, 2012). Due to employing small  $N_{\text{ens}}$ , also, spurious correlations may arise in the prediction covariance matrix  $\hat{\mathbf{C}}_n$  leading to the requirement of *covariance localisation*. This applies an elementwise product  $\mathbf{U} \odot \hat{\mathbf{C}}_n$  (Houtekamer and Mitchell, 2001), with  $\mathbf{U}$  chosen to encode physical knowledge of the system at hand — for example, points which are at a large distance from one another should have near zero covariance. Furthermore, as the ensemble-approximated covariance is a low-rank approximation, there may be some under-estimation of the covariance matrix. Again, a covariance inflation can be used to provide an ad-hoc amelioration of this, as with the ExKF. The EnKF can also be formulated in terms of the square root of the covariance matrix (Tippett et al., 2003)<sup>13</sup>, which gives the square-root ensemble Kalman filters. This leads to non-stochastic versions of the EnKF that do not perturb the observations. Instead the ensemble is updated in a deterministic fashion that ensures that Kalman update for the covariance is computed exactly (i.e., it is free of Monte Carlo error).

Additional approaches include the variational algorithms of 3DVAR (Lorenc, 1986; Lorenc et al., 2000) and 4DVAR (Zupanski, 1997). These algorithms are so named because their respective formulations lead to problems in the calculus of variations. The 3D and 4D prefixes result from these methods being applied to problems with three spatial dimensions, as well as the incorporation of temporal dependencies in the 4D case. 3DVAR constructs an approximation to the true filtering posterior through treating the model prediction covariance as fixed, with  $\hat{\mathbf{C}}_n = \hat{\mathbf{C}}$  for all  $n$ . For the nonlinear filter this sets  $p(\boldsymbol{\theta}_n | \mathbf{y}_{1:n-1}) \sim \mathcal{N}(\hat{\mathbf{m}}_n, \hat{\mathbf{C}})$ , where  $\hat{\mathbf{m}}_n = \mathcal{F}(\mathbf{m}_{n-1})$ , for the prediction step. In this update step only the mean of the Gaussian distribution is updated, so that  $\mathbf{m}_n = \hat{\mathbf{m}}_n + \hat{\mathbf{C}}\mathbf{H}^\top (\mathbf{H}\hat{\mathbf{C}}\mathbf{H}^\top + \mathbf{R})^{-1}(\mathbf{y}_n - \mathbf{H}\hat{\mathbf{m}}_n)$ . 4DVAR, which extends into the time domain, estimates the *maximum-a-posteriori* (MAP) point for the smoothing posterior (for a discussion on MAP estimation in the context of inverse problems, see Kaipio and Somersalo (2006)). The MAP estimate is the most probable point in the support of the posterior distribution. In this case we compute

$$\hat{\boldsymbol{\theta}}_0 = \underset{\boldsymbol{\theta}_0}{\operatorname{argmax}} p(\boldsymbol{\theta}_0 | \mathbf{y}_{1:n_t}).$$

This can be computed through standard optimisation methods, which may be accelerated with gradient information (see, e.g., Nocedal and Wright (2006)). If stochastic dynamics is assumed, as in (1.4), then this is known as *weak constraint* 4DVAR. Deterministic dynamics gives the *strong constraint* 4DVAR.

### 2.3.2 Gaussian Processes

In order to build the statistical methodology on top of the finite element method, we require the introduction of some notion of uncertainty for *functions*. In NL-statFEM, the addition of function-valued uncertainty inside of the governing PDE enables, upon finite element discretisation, the derivation of an SDE which can then be numerically integrated (c.f. Section 2.2) and updated with data (c.f.

<sup>13</sup>This is the scaled prediction deviations from the mean:  $\hat{\mathbf{C}}_n = \mathbf{X}_n \mathbf{X}_n^\top$ ,  $\mathbf{X}_n = \sqrt{\frac{1}{N_{\text{ens}} - 1}} \left[ \hat{\boldsymbol{\theta}}_n^{(1)} - \mathbb{E}(\hat{\boldsymbol{\theta}}_n), \dots, \hat{\boldsymbol{\theta}}_n^{(N_{\text{ens}})} - \mathbb{E}(\hat{\boldsymbol{\theta}}_n) \right]$ .

Section 2.3.1). To introduce this uncertainty, we use *Gaussian Processes (GPs)*. A  $\mathcal{GP}$  is a continuously indexed stochastic process whose joint distributions are multivariate Gaussians. They allow for functions to be expressed as being uncertain, instead of fixed, deterministic objects.

$\mathcal{GP}$ s are used for statistical regression and classification (Williams and Rasmussen, 2006), and are common in uncertainty quantification workflows (see, e.g., Ghanem et al., 2017). They can be used to emulate expensive physical models (Sacks et al., 1989), which then enables sensitivity analysis to be conducted (Oakley and O'Hagan, 2004), or as a way of parameterising function-valued model mismatch (Kennedy and O'Hagan, 2001). Due to the connection to Gaussian measure — which is a way to define measure in infinite dimensional spaces where the classical Lebesgue measure cannot be constructed —  $\mathcal{GP}$ s also allow for infinite-dimensional inference to proceed via the Radon-Nikodym derivative (Stuart, 2010, Section 6). This allows for the development of algorithms on the function space, guaranteeing scalability.

We now introduce  $\mathcal{GP}$ s below and state their connection to Gaussian measure. We then discuss the inference problem for  $\mathcal{GP}$ s, and derive the posterior  $\mathcal{GP}$  distribution, and conclude with a discussion on computational techniques. For a general reference to this material see, for example, Williams and Rasmussen (2006).

**Definition 8.** A function  $\xi : \mathcal{D} \rightarrow \mathbb{R}$ , where  $\mathcal{D} \subset \mathbb{R}^d$ , is a *Gaussian Process (GP)* if any finite number of input locations  $\mathcal{X} = (\mathbf{x}_1, \dots, \mathbf{x}_m)$  gives a finite-dimensional multivariate Gaussian density over the associated function values, i.e.

$$p(\xi(\mathbf{x}_1), \dots, \xi(\mathbf{x}_m)) = \mathcal{N}(\mathbf{m}, \mathbf{K}),$$

where  $\mathbf{m} \in \mathbb{R}^m$ ,  $\mathbf{K} \in \mathbb{R}^{m \times m}$ , and  $\mathbf{K}$  is positive semidefinite.

A  $\mathcal{GP}$  is completely specified by its *mean function*  $m(\mathbf{x})$  and *covariance function* or *covariance kernel*  $k(\mathbf{x}, \mathbf{x}')$ , and is commonly written as  $\xi(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ . The mean function and covariance function are related to  $\xi$  via

$$\begin{aligned} \mathbb{E}[\xi(\mathbf{x})] &= m(\mathbf{x}), \\ \mathbb{E}[(\xi(\mathbf{x}') - m(\mathbf{x}))(\xi(\mathbf{x}') - m(\mathbf{x}'))] &= k(\mathbf{x}, \mathbf{x}'). \end{aligned} \tag{2.19}$$

Therefore on the input set  $\mathcal{X}$  we have  $\mathbf{m}_i = m(\mathbf{x}_i)$ ,  $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ .

The covariance function is required to be positive definite, in order to define a valid covariance matrix. Positive definiteness for the covariance function is given by the property that for any function  $h(\mathbf{x}) \in L^1(\mathbb{R}; \mathbb{R})$ , we have

$$\int \int h(\mathbf{x})k(\mathbf{x}, \mathbf{x}')h(\mathbf{x}') \, d\mathbf{x} \, d\mathbf{x}' \geq 0. \tag{2.20}$$

Some common choices of the covariance function are listed below, along with their *hyperparameters*.

1. Squared-exponential:

$$k(\mathbf{x}, \mathbf{x}') = \rho^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{2\ell^2}\right),$$

with hyperparameters  $\boldsymbol{\theta} = (\rho, \ell)$ .

2. Matern:

$$k(\mathbf{x}, \mathbf{x}') = \rho^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{\sqrt{2\nu} \|\mathbf{x} - \mathbf{x}'\|_2}{\ell} \right) K_\nu \left( \frac{\sqrt{2\nu} \|\mathbf{x} - \mathbf{x}'\|_2}{\ell} \right),$$

with hyperparameters  $\boldsymbol{\theta} = (\nu, \rho, \ell)$ ;  $\nu$  is typically referred to as the *smoothness* parameter.  $K_\nu$  in this case is a modified Bessel function of the second kind (Abramowitz and Stegun, 1964).

3. Periodic (defined for 1D):

$$k(x, x') = \rho^2 \exp \left( -\frac{2 \sin^2(\pi(x - x')/p)}{\ell^2} \right),$$

with hyperparameters  $\boldsymbol{\theta} = (\rho, \ell, p)$ .

The choice of the covariance kernel decides the properties of the  $\mathcal{GP}$  sample paths. Choosing the squared-exponential kernel, for example, means that a  $\mathcal{GP}$  with this covariance structure will have sample paths that are  $C^\infty(\mathcal{D})$ . For the Matérn kernel paths are  $p$ -times mean-square differentiable if and only if  $\nu > p$ , and choosing the periodic kernel means that paths are infinitely differentiable and have period  $p$ . The hyperparameters  $\boldsymbol{\theta}$  allow for variation in the expected behaviour of sample paths and must be chosen at the same time as the covariance function. For example, the length-scale parameter  $\ell$  encodes the correlation lengths. Large values of  $\ell$  mean that points further away from one another have higher correlations, and vice-versa. Kernels can also be combined in various ways (for example, they may be added or multiplied) and can still define valid covariance kernels.

Given that  $\mathcal{GP}$ s provide a way to express uncertainty about functions, it is perhaps unsurprising that  $\mathcal{GP}$ s implicitly define a probability measure on an appropriate function space. The function space chosen depends on the covariance function, though is not unique to the specific covariance function, and the measure derived is a Gaussian measure. We now show that for a  $\mathcal{GP}$  that has a covariance function that is  $L^2(\mathcal{D} \times \mathcal{D}; \mathbb{R})$ , a resultant probability measure can be derived on  $L^2(\mathcal{D}; \mathbb{R})$ . For more detail, see Section 6 of Stuart (2010).

**Example 3.** Consider, without loss of generality, a mean-zero  $\mathcal{GP}$   $\xi(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$  with  $k(\cdot, \cdot) \in L^2(\mathcal{D} \times \mathcal{D}; \mathbb{R})$ . For any Hilbert space with inner product  $\langle \cdot, \cdot \rangle$ , the covariance operator  $K$  has the identity  $\langle \psi, K\phi \rangle = \mathbb{E} \langle \psi, \xi \rangle \langle \phi, \xi \rangle$ . For  $L^2(\mathcal{D}; \mathbb{R})$  this inner product is

$$\begin{aligned} \langle \psi, K\phi \rangle &= \mathbb{E} \int \psi(\mathbf{x}) \xi(\mathbf{x}) \, d\mathbf{x} \int \phi(\mathbf{x}') \xi(\mathbf{x}') \, d\mathbf{x}' \\ &= \mathbb{E} \int \int \psi(\mathbf{x}) \xi(\mathbf{x}) \xi(\mathbf{x}') \phi(\mathbf{x}') \, d\mathbf{x}' \, d\mathbf{x} \\ &= \int \psi(\mathbf{x}) \int k(\mathbf{x}, \mathbf{x}') \phi(\mathbf{x}') \, d\mathbf{x}' \, d\mathbf{x}, \end{aligned}$$

and thus we see that the operator  $K$  is an integral operator with kernel  $k(\cdot, \cdot)$ , i.e.  $K\phi = \int k(\mathbf{x}, \mathbf{x}') \phi(\mathbf{x}') \, d\mathbf{x}'$ . Therefore on  $L^2(\mathcal{D}; \mathbb{R})$ , the  $\mathcal{GP}$   $\xi(\mathbf{x})$  defines the Gaussian measure  $\mu = \mathcal{N}(0, K)$ .

### 2.3.3 Inference with Gaussian Processes

Given that the  $\mathcal{GP}$   $\xi \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$  defines a prior over an appropriate function space, the next thing of interest is the derivation of the posterior. One way is to first define a finite dimensional approximation

to this posterior through evaluating the posterior on a set of testing locations  $\mathcal{X}_* = (\mathbf{x}_1, \dots, \mathbf{x}_{n_\xi})$ . Consider the case in which data points  $\mathbf{y} = (y_1, \dots, y_{n_y})$  are observed at locations  $\mathcal{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{n_y})$ . Then the data generating process may be written as

$$y_i = \xi(\mathbf{x}_i) + \eta_i, \quad \eta_i \sim \mathcal{N}(0, \sigma_y^2), \quad \boldsymbol{\eta} \perp \xi. \quad (2.21)$$

The posterior at the testing locations,  $p(\boldsymbol{\xi} \mid \mathbf{y}, \boldsymbol{\theta}, \sigma_y)$ , is thus given through conditioning in the joint Gaussian:

$$\begin{pmatrix} \boldsymbol{\xi} \\ \mathbf{y} \end{pmatrix} \sim \mathcal{N} \left( \mathbf{0}, \begin{bmatrix} \mathbf{K}_{**} & \mathbf{K}_* \\ \mathbf{K}_*^\top & \mathbf{K} + \sigma_y^2 \mathbf{I} \end{bmatrix} \right),$$

where  $\mathbf{K}_{**}$  is the  $n_\xi \times n_\xi$  covariance matrix defined on the testing locations  $\mathcal{X}_*$ ; the covariance matrices  $\mathbf{K}_*$  and  $\mathbf{K}$  are defined similarly using the training locations. The posterior is

$$\begin{aligned} p(\boldsymbol{\xi} \mid \mathbf{y}, \boldsymbol{\theta}, \sigma_y) &= \mathcal{N}(\mathbf{m}_y, \mathbf{K}_y), \\ \mathbf{m}_y &= \mathbf{K}_*(\mathbf{K} + \sigma_y^2 \mathbf{I})^{-1} \mathbf{y}, \\ \mathbf{K}_y &= \mathbf{K}_{**} - \mathbf{K}_*(\mathbf{K} + \sigma_y^2 \mathbf{I})^{-1} \mathbf{K}_*^\top. \end{aligned}$$

In this case the posterior is a multivariate normal distribution, having “discretised” the  $\mathcal{GP}$  onto the grid  $\mathcal{X}_*$ . The extension to the  $\mathcal{GP}$  posterior follows from considering that the above construction holds for any set of testing points  $\mathcal{X}_*$ . [Kanagawa et al. \(2018\)](#) note that from the Kolmogorov extension theorem ([Øksendal, 2003](#), Theorem 2.1.5) and the definition of the  $\mathcal{GP}$  (Definition 8), the posterior is also a  $\mathcal{GP}$

$$p(\boldsymbol{\xi} \mid \mathbf{y}, \boldsymbol{\theta}, \sigma_y) = \mathcal{GP}(m_y(\mathbf{x}), k_y(\mathbf{x}, \mathbf{x}')), \quad (2.22)$$

with mean and covariance functions

$$\begin{aligned} m_y(\mathbf{x}) &= k(\mathbf{x}, \mathcal{X})(\mathbf{K} + \sigma_y^2 \mathbf{I})^{-1} \mathbf{y}, \\ k_y(\mathbf{x}, \mathbf{x}') &= k(\mathbf{x}, \mathbf{x}') - k(\mathbf{x}, \mathcal{X})(\mathbf{K} + \sigma_y^2 \mathbf{I})^{-1} k(\mathcal{X}, \mathbf{x}'), \end{aligned}$$

where the row vector  $k(\mathbf{x}, \mathcal{X}) = k(\mathcal{X}, \mathbf{x})^\top = (k(\mathbf{x}, \mathbf{x}_1), \dots, k(\mathbf{x}, \mathbf{x}_{n_y}))$ .

This can be applied in scenarios in which the likelihood is degenerate (i.e., there is no noise) and circumvents having to derive the posterior on the function space using the Gaussian reference measure. Because  $\mathcal{GP}$ s involve the evaluation of the function at specific points in the domain  $\mathcal{D}$ , the appropriate function space setting for the posterior is the reproducing kernel Hilbert space (RKHS) ([Berlinet and Thomas-Agnan, 2003](#)) and not the  $L^p(\mathcal{D}; \mathbb{R})$  spaces. The defining property of the RKHS is that the evaluation functionals are continuous, which allow for the infinite-dimensional Bayes theorem to be applied ([Stuart, 2010](#), Theorem 6.31). This is not the case for the  $L^p(\mathcal{D}; \mathbb{R})$  spaces.

The infinite dimensional Bayes theorem proceeds from the prior measure  $\mu_0$  defined on a Banach space  $X$  with norm  $\|\cdot\|$ . Data  $\mathbf{y} \in \mathbb{R}^{n_y}$  are assumed to be arriving via the *observation operator*  $H : X \rightarrow \mathbb{R}^{n_y}$  a continuous operator that maps from the function space  $X$  to the data space  $\mathbb{R}^{n_y}$ . The data is assumed to be observed similarly to (2.21),  $\mathbf{y} = H(u) + \boldsymbol{\eta}$  with  $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ . The posterior

measure  $\mu^y$ , is defined from the Radon-Nikodym derivative of the prior with respect to the posterior,

$$\frac{d\mu^y}{\mu_0} = \frac{1}{Z} \exp \left( \frac{1}{2\sigma^2} (\mathbf{y} - H(\mathbf{u}))^\top (\mathbf{y} - H(\mathbf{u})) \right),$$

and is therefore absolutely continuous with respect to the prior.

In practice, to derive the posterior the hyperparameters need to be either set, estimated, or integrated over (MacKay, 1996). A popular approach is to estimate the hyperparameters through the log marginal likelihood (Williams and Rasmussen, 2006), which is analytically tractable and given by

$$\log p(\mathbf{y} \mid \boldsymbol{\theta}, \sigma_y) = -\frac{1}{2} \mathbf{y}^\top (\mathbf{K} + \sigma_y^2 \mathbf{I})^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K} + \sigma_y^2 \mathbf{I}| - \frac{n_y}{2} \log 2\pi.$$

Parameters are estimated through maximising this log marginal likelihood, with which we can include a prior term for regularisation

$$\boldsymbol{\theta}^* = \operatorname{argmax}_{\boldsymbol{\theta}} \{ \log p(\mathbf{y} \mid \boldsymbol{\theta}, \sigma_y) + \log p(\boldsymbol{\theta}) \}.$$

Computing  $\log p(\mathbf{y} \mid \boldsymbol{\theta}, \sigma_y)$  is the main computational bottleneck for Gaussian Processes, due to the inversion of the covariance matrix  $\mathbf{K} + \sigma_y^2 \mathbf{I}$ . This requires  $O(n_y^3)$  operations and  $O(n_y^2)$  memory using the Cholesky decomposition (Golub and Van Loan, 2013). There exist various approaches to scale the  $\mathcal{GP}$  regression problem up to high dimensions and we cover a selection of techniques here; see Liu et al. (2020) for a review of available methods.

Structured grid methods make use of the observation locations  $\mathcal{X}$  being regularly spaced across  $\mathcal{D}$  and are only applicable in this setting. A common example is circulant embedding (Dietrich and Newsam, 1997) which — for stationary covariance functions<sup>14</sup> — embeds the Toeplitz covariance matrix inside of a larger circulant matrix, whose eigendecomposition can be computed using the fast Fourier transform (FFT) (Cooley and Tukey, 1965). Another structured grid approach is when the kernel is separable across dimensions (for a review of these methods, see Chapter 5 from Saatci, 2011), and the grid  $\mathcal{X}$  is expressed as a Cartesian product across these dimensions. The covariance matrix can be written as a Kronecker product of covariance matrices on each subgrid. The full covariance matrix thus never needs to be formed, and properties of the Kronecker product are used to speed up computation.

Low-rank or *sparse* methods are also popular. These use various approximations to bring the scaling from  $O(n_y^3)$  down to  $O(m^2 n_y)$ , where  $m \ll n_y$  (Snelson and Ghahramani, 2006). These include the Nystrom approximation (Williams and Seeger, 2001), which uses a subset of the training data to approximate the eigenvalues and eigenfunctions of the covariance kernel, and the inducing point methods, which construct an approximation through a set of points that lie inside  $\mathcal{D}$ , whose location can be learnt through variational methods (Hensman et al., 2013; Titsias, 2009). Additional techniques are covered in Chapter 8 of Williams and Rasmussen (2006).

Finally the recent combination of iterative solvers and GPU acceleration is showing promise (Wang et al., 2019). Conjugate gradients can be used to compute  $(\mathbf{K} + \sigma_y \mathbf{I})^{-1} \mathbf{y}$ , with matrix-vector products of the form  $(\mathbf{K} + \sigma_y \mathbf{I}) \mathbf{x}$  being done on the GPU. This operation is highly parallelisable as entries are

<sup>14</sup>A stationary covariance function is one that can be written as  $k(\mathbf{x}, \mathbf{x}') = \kappa(\mathbf{x} - \mathbf{x}')$ .

row-independent, and memory limitations can be circumvented through batched ([Gardner et al., 2021](#)) or on-the-fly computation ([Charlier et al., 2021](#)).



## Chapter 3

# Statistical finite element methods

*In this chapter we develop our methodology, coined NL-statFEM. We start by introducing the linear, static statistical finite element method (statFEM), given in [Girolami et al. \(2021\)](#), proceeding from the governing equations to the derived prior and posterior distributions. This is illustrated using an example elliptic PDE. Next, our first extension of statFEM — into linear, time-evolving PDEs — is studied. In this case we develop our approach as motivated for a generic parabolic PDE. Again, the prior distribution is derived and we discuss the data-conditioning procedure; in this case the standard Kalman filter. We then extend this approach further, presenting our NL-statFEM for nonlinear, time-evolving PDEs, as in [Duffin et al. \(2021\)](#). We arrive at the extended and ensemble Kalman filters as the data-conditioning algorithms. We derive our NL-statFEM for a general nonlinear, time-dependent PDE, so as to enable its application to the diverse systems seen in physical and engineering contexts. We defer examples to Chapter 4, in which we demonstrate the method on three classical nonlinear PDEs.*

### 3.1 Statistical finite element methods for static problems

The statistical finite element method (statFEM) of [Girolami et al. \(2021\)](#) provides methodology to deal with model misspecification, through embedding *a priori* uncertainty inside of a static mathematical model. We build upon this methodology in the remainder of this thesis. As such, this section discusses this initial contribution, with additional detail taken from the Supplement of [Duffin et al. \(2021\)](#). In statFEM, FEM solutions are updated upon receipt of data within a fully Bayesian framework. As introduced previously, FEM solves mathematical models given by PDEs, which describe the relationships between partial derivatives of some function, under appropriate boundary conditions, initial conditions, and forcing terms ([Evans, 2010](#)). Uncertainty may thus enter the PDE through any or all of these terms, and may also arise due to model imperfection ([Judd and Smith, 2004](#); [Raissi et al., 2017](#)). In the statFEM framework, this uncertainty will be represented through modelling the PDE coefficients and forcing as stochastic processes ([Ghanem and Spanos, 2003](#); [Gunzburger et al., 2014](#)).

For exposition, it is assumed that the uncertainty arises through the forcing only. A Bayesian interpretation of probability is taken (see, e.g., [Jaynes, 2003](#)), so that the uncertainty represents the degree to which the model is thought to be correctly specified. For an example elliptic linear PDE,



with diffusivity coefficient  $\Lambda(\cdot)$  and Dirichlet boundary conditions, this can be expressed as

$$\begin{aligned} -\nabla \cdot (\Lambda \nabla u) &= \xi, \\ u &= 0, \quad \mathbf{x} \in \partial \mathcal{D}, \end{aligned} \tag{3.1}$$

$u := u(\mathbf{x})$ ,  $\Lambda := \Lambda(\mathbf{x})$ ,  $\xi := \xi(\mathbf{x})$ ,  $\mathbf{x} \in \mathcal{D} \subseteq \mathbb{R}^d$ . Equation (3.1) is taken as the example PDE in this section, assuming that  $\partial \mathcal{D}$  is smooth and that  $\Lambda \in L^\infty(\mathcal{D}; \mathbb{R})$ . It is assumed that  $\Lambda$  is known, however the framework may be extended to deal with  $\Lambda$  also being modelled by a stochastic process.

The stochastic forcing  $\xi(\mathbf{x}) \sim \mathcal{GP}(f(\mathbf{x}), k_\theta(\mathbf{x}, \mathbf{x}'))$ , with  $\xi \in L^2(\mathcal{D}; \mathbb{R})$ , absorbs any deterministic forcing into  $f$ <sup>1</sup>. Note, also, that this implies that  $f \in L^2(\mathcal{D}; \mathbb{R})$ . The push-forward of the  $\mathcal{GP}$   $\xi$  induces a probability measure over the space of solutions to (3.1) and characterises prior belief in the model. The covariance kernel  $k_\theta(\cdot, \cdot)$  can be decided upon by domain experts so that the uncertainty induced is physically sensible. For example,  $k_\theta(\cdot, \cdot)$  can be chosen to be a Matérn covariance function to reflect the unknown forcing having derivatives up to a known order. Throughout this thesis, the forcing is taken to be smooth<sup>2</sup> with a squared-exponential covariance kernel (Williams and Rasmussen, 2006, Chapter 4)

$$\mathbb{E}[\xi(\mathbf{x})\xi(\mathbf{x}')] = k_\theta(\mathbf{x}, \mathbf{x}') = \rho^2 \exp\left(-\frac{1}{2\ell^2}\|\mathbf{x} - \mathbf{x}'\|^2\right), \tag{3.2}$$

governed by hyperparameters  $\boldsymbol{\theta} = (\rho, \ell)$ .

The prior induced by (3.1), on the function space  $L^2(\mathcal{D}; \mathbb{R})$ , is now derived. Given the regularity conditions on the coefficients and forcing, Equation (3.1) has a weak solution  $u \in H_0^1(\mathcal{D}; \mathbb{R})$  which is given by  $u = L_\Lambda^{-1}\xi$ . The operator  $L_\Lambda^{-1}$  is a bounded linear operator that maps  $L^2(\mathcal{D}; \mathbb{R})$  into itself (Evans, 2010) and thus  $u \sim \mu_0(u) = \mathcal{N}(m_u, C_u)$ . The mean  $m_u = L_\Lambda^{-1}f$  is the weak solution to the deterministic elliptic problem, and the covariance operator  $C_u$  is derived from

$$\langle \psi, C_u \phi \rangle = \int_{\mathcal{D}} \psi(\mathbf{x}) L_\Lambda^{-1} K_\theta L_\Lambda^{-1} \phi(\mathbf{x}) \, d\mathbf{x},$$

where  $[K_\theta \phi](\mathbf{x}) = \int k_\theta(\mathbf{x}, \mathbf{x}') \phi(\mathbf{x}') \, d\mathbf{x}'$ .

**Proposition 1.** *The measure  $\mu_0$  defines a Gaussian measure on  $L^2(\mathcal{D}; \mathbb{R})$ .*

*Proof.* We require first that  $m_u \in L^2(\mathcal{D}; \mathbb{R})$ . This follows from  $f \in L^2(\mathcal{D}; \mathbb{R})$  and  $L_\Lambda^{-1} : L^2(\mathcal{D}; \mathbb{R}) \rightarrow L^2(\mathcal{D}; \mathbb{R})$ .

The covariance operator is required to be self-adjoint, positive semidefinite, and trace-class on  $L^2(\mathcal{D}; \mathbb{R})$ . The operators  $L_\Lambda^{-1}$  and  $K_\theta$  are both self-adjoint and positive. Thus  $\langle L_\Lambda^{-1} K_\theta L_\Lambda^{-1} \psi, \phi \rangle = \langle \psi, L_\Lambda^{-1} K_\theta L_\Lambda^{-1} \phi \rangle$  through repeated application of the self-adjoint property, so  $C_u$  is itself self-adjoint. Positivity of  $C_u$  is ensured through again applying the self-adjoint property of  $L_\Lambda^{-1}$  to give  $\langle K_\theta L_\Lambda^{-1} \psi, L_\Lambda^{-1} \psi \rangle$ . Writing  $\hat{\psi} = L_\Lambda^{-1} \psi$  gives  $\langle K_\theta \hat{\psi}, \hat{\psi} \rangle \geq 0$ . Therefore,  $\langle K_\theta L_\Lambda^{-1} \psi, L_\Lambda^{-1} \psi \rangle \geq 0$ . To prove that the operator is trace-class we begin with the diagonalisation of  $L_\Lambda$ ,  $L_\Lambda e_i = \lambda_i e_i$ , which

<sup>1</sup>For example, the RHS of (3.1) may be able to decomposed as  $f_1(\mathbf{x}) + \xi(\mathbf{x})$ ,  $\xi(\mathbf{x}) \sim \mathcal{GP}(f_2(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ , which for  $f_1 \in L^2(\mathcal{D}; \mathbb{R})$  is equivalent to  $\xi(\mathbf{x}) \sim \mathcal{GP}(f_1(\mathbf{x}) + f_2(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$

<sup>2</sup>For the squared-exponential kernel,  $\mathcal{GP}$  sample paths are  $C^\infty(\mathcal{D}; \mathbb{R})$ .

forms a countable orthonormal basis of  $L^2(\mathcal{D}; \mathbb{R})$ . The eigenvalues satisfy (Evans, 2010, Theorem 6.5.1)

$$0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots, \quad \lambda_k \rightarrow \infty, \quad k \rightarrow \infty.$$

The trace of  $C_u$  can be written in terms of this basis, so that

$$\text{Tr} (L_\Lambda^{-1} K_\theta L_\Lambda^{-1}) = \sum_{i=1}^{\infty} \langle L_\Lambda^{-1} K_\theta L_\Lambda^{-1} e_i, e_i \rangle.$$

Again, making use of  $L_\Lambda$  being self-adjoint we write  $\langle L_\Lambda^{-1} K_\theta L_\Lambda^{-1} e_i, e_i \rangle = \langle K_\theta L_\Lambda^{-1} e_i, L_\Lambda^{-1} e_i \rangle = \frac{1}{\lambda_i^2} \langle K_\theta e_i, e_i \rangle$ . Since  $\lambda_i$  are positive and nondecreasing

$$\text{Tr} (L_\Lambda^{-1} K_\theta L_\Lambda^{-1}) = \sum_{i=1}^{\infty} \frac{1}{\lambda_i^2} \langle K_\theta e_i, e_i \rangle \leq \frac{1}{\lambda_1^2} \sum_{i=1}^{\infty} \langle K_\theta e_i, e_i \rangle = \frac{1}{\lambda_1^2} \text{Tr}(K_\theta) < \infty.$$

Thus  $\mu_0$  defines a Gaussian measure on  $L^2(\mathcal{D}; \mathbb{R})$ . □

The posterior measure,  $\mu^y$ , on  $L^2(\mathcal{D}; \mathbb{R})$  (Stuart, 2010), is defined using the Radon-Nikodym derivative under the condition that the posterior is absolutely continuous with respect to the prior. This posterior measure conditions on the PDE parameters  $\Lambda$  and covariance hyperparameters  $\theta$ . Denote by  $H : L^2(\mathcal{D}; \mathbb{R}) \rightarrow \mathbb{R}^{n_y}$  the *observation operator*, which is assumed to be continuous and maps the PDE variable of interest from the solution space to the data space. Assuming the data-generating process  $\mathbf{y} = H\mathbf{u} + \boldsymbol{\eta}$ , for data  $\mathbf{y} \in \mathbb{R}^{n_y}$ , and observational noise  $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{n_y})$ , gives the infinite-dimensional posterior

$$\frac{d\mu^y}{d\mu_0} = \frac{1}{Z} \exp \left( -\frac{1}{2\sigma^2} (\mathbf{y} - H\mathbf{u})^\top (\mathbf{y} - H\mathbf{u}) \right), \quad (3.3)$$

for some normalising constant  $Z$ . The posterior measure enables model deficiencies to be corrected for by observations, synthesising *a priori* knowledge of the physics with observed data. However this is an abstract mathematical object which is not able to be realised without some form of discretisation; one such discretisation technique is now discussed.

### 3.1.1 The linear statFEM prior

In practice, computing the posterior requires discretisation at some stage. In statFEM this is done before observing data, to derive a finite-dimensional prior distribution through FEM. FEM first requires deriving the weak form of (3.1) which proceeds through multiplying by test functions  $\psi \in H_0^1(\mathcal{D}; \mathbb{R})$  and integrating over  $\mathcal{D}$

$$-\int_{\mathcal{D}} \nabla \cdot (\Lambda(\mathbf{x}) \nabla u(\mathbf{x})) \psi(\mathbf{x}) \, d\mathbf{x} = \int_{\mathcal{D}} \xi(\mathbf{x}) \psi(\mathbf{x}) \, d\mathbf{x},$$

from which the first part can be simplified through integrating by parts and recognising that the boundary terms are zero due to Dirichlet conditions

$$\int_{\mathcal{D}} \Lambda(\mathbf{x}) \nabla u(\mathbf{x}) \cdot \nabla \psi(\mathbf{x}) \, d\mathbf{x} = \int_{\mathcal{D}} \xi(\mathbf{x}) \psi(\mathbf{x}) \, d\mathbf{x}.$$

Hence, for this example,

$$\mathcal{A}_\Lambda(u, \psi) = \int_{\mathcal{D}} \Lambda(\mathbf{x}) \nabla u(\mathbf{x}) \cdot \nabla \psi(\mathbf{x}) \, d\mathbf{x}, \quad \langle \xi, \psi \rangle = \int_{\mathcal{D}} \xi(\mathbf{x}) \psi(\mathbf{x}) \, d\mathbf{x},$$

where  $u \in H_0^1(\mathcal{D}; \mathbb{R})$  (c.f. Section 2.1.2). The weak form relaxes derivative requirements on solutions; first order weak derivatives and not second order strong derivatives are required for the trial functions.

The finite element mesh is given by subdividing the domain  $\mathcal{D}$  into the triangulation  $\mathcal{D}_h \subseteq \mathcal{D}$  with vertices  $\{\mathbf{x}_j\}_{j=1}^{n_u}$ , where the maximal length of the sides of the triangulation is  $h$ . In this thesis, it is assumed that the model domains do not require bespoke methods to generate the mesh  $\mathcal{D}_h$ . Mesh discretisations for models with complex domain geometry may lead to an additional source of model misspecification<sup>3</sup>, which is left untreated in this work.

The first-order polynomial basis functions  $\{\phi_j\}_{j=1}^{n_u}$  are defined on the mesh, having the property that  $\phi_j(\mathbf{x}_i) = \delta_{ij}$  and  $\phi_j(\mathbf{x}) = 0$  if  $\mathbf{x} \in \partial\mathcal{D}$ . Denote by  $V_h \subset H_0^1(\mathcal{D}; \mathbb{R})$  the span of these basis functions, and write the approximations  $u_h(\mathbf{x}) = \sum_{i=1}^{n_u} u_i \phi_i(\mathbf{x})$ ,  $\psi_h(\mathbf{x}) = \sum_{i=1}^{n_u} \psi_i \phi_i(\mathbf{x})$ . In the finite-dimensional space  $V_h$  the weak form gives the linear system

$$\sum_{i=1}^{n_u} u_i \mathcal{A}_\Lambda(\phi_i, \phi_j) = \langle \xi, \phi_j \rangle, \quad j = 1, \dots, n_u.$$

Defining a Gaussian over the FEM coefficients  $\mathbf{u} = (u_1, \dots, u_{n_u})^\top \in \mathbb{R}^{n_u}$ ,

$$p(\mathbf{u} \mid \boldsymbol{\theta}, \Lambda) = \mathcal{N}(\mathbf{m}_u, \mathbf{C}_u),$$

$$\mathbf{m}_u = \mathbf{A}^{-1} \mathbf{b}, \quad \mathbf{C}_u = \mathbf{A}^{-1} \mathbf{G}_\theta \mathbf{A}^{-\top},$$

where  $\mathbf{A}_{ji} = \mathcal{A}_\Lambda(\phi_i, \phi_j)$ ,  $\mathbf{b}_j = \langle f, \phi_j \rangle$ . The covariance  $\mathbf{G}_\theta$  is

$$\mathbf{G}_{\theta,ij} = \mathbb{E}[\langle \xi, \phi_i \rangle \langle \xi, \phi_j \rangle] = \int_{\mathcal{D}} \phi_i(\mathbf{x}) \int_{\mathcal{D}} k_\theta(\mathbf{x}, \mathbf{x}') \phi_j(\mathbf{x}') \, d\mathbf{x} \, d\mathbf{x}',$$

to yield the statFEM prior distribution. The mean is the deterministic FEM solution,  $\mathbf{A}^{-1} \mathbf{b}$ , and the covariance is the action of the discretised PDE operator on the prior  $\mathcal{GP}$  covariance, giving a “physics-informed” uncertainty quantification (UQ). This unbiased measure summarises the *a-priori* uncertainty associated with the FEM coefficients, as stipulated by prior modelling assumptions.

**Remark.** The choice of basis functions defines the discretisation. For example, if  $\mathcal{D} = [0, 1]$  then choosing the Fourier basis of  $L^2([0, 1]; \mathbb{R})$ ,  $\{e^{2\pi i k x}\}_{k=1}^\infty$ , gives the spectral Galerkin method (Boyd, 2001).

The covariance can be approximated by  $\mathbf{G}_\theta = \mathbf{M} \mathbf{K}_\theta \mathbf{M}^\top$ , where  $\mathbf{M}_{ij} = \langle \phi_i, \phi_j \rangle$ . The matrix  $\mathbf{M}$  is known as the mass matrix and often appears in finite element computations. This approximation can

<sup>3</sup>Misspecification, in this sense, may come from the mesh being inappropriately refined in parts of the domain where phenomena of interest are occurring.

be derived from

$$\begin{aligned}
 \mathbf{G}_{\theta,ij} &= \int_{\mathcal{D}} \phi_i(\mathbf{x}) \int_{\mathcal{D}} k_{\theta}(\mathbf{x}, \mathbf{x}') \phi_j(\mathbf{x}') \, d\mathbf{x}' \, d\mathbf{x} \\
 &\approx \sum_{k=1}^{n_u} \int_{\mathcal{D}} \phi_i(\mathbf{x}) \sum_{l=1}^{n_u} \int_{\mathcal{D}} \phi_k(\mathbf{x}) k_{\theta}(\mathbf{x}_k, \mathbf{x}_l) \phi_l(\mathbf{x}') \phi_j(\mathbf{x}') \, d\mathbf{x}' \, d\mathbf{x} \\
 &\approx \sum_{k=1}^{n_u} \int_{\mathcal{D}} \phi_i(\mathbf{x}) \phi_k(\mathbf{x}) \sum_{l=1}^{n_u} \int_{\mathcal{D}} k_{\theta}(\mathbf{x}_k, \mathbf{x}_l) \phi_l(\mathbf{x}') \phi_j(\mathbf{x}') \, d\mathbf{x}' \, d\mathbf{x} \\
 &= \sum_{k=1}^{n_u} \mathbf{M}_{ik} \sum_{l=1}^{n_u} \mathbf{K}_{\theta,kl} \mathbf{M}_{lj}.
 \end{aligned}$$

Discretisation of the covariance  $\mathbf{G}_{\theta}$  requires some care to implement as this is a  $2d$  dimensional integral. In general this does not give a sparse matrix; for common choices of  $k_{\theta}(\cdot, \cdot)$  — such as the Matérn and squared-exponential kernels — the resultant covariance matrix is dense. A sparsity constraint can be imposed by assuming that the forcing is a white noise process,  $k_{\theta}(\mathbf{x}, \mathbf{x}') = \theta^2 \delta(\mathbf{x} - \mathbf{x}')$ ,

$$\begin{aligned}
 \mathbf{G}_{\theta,ij} &= \int_{\mathcal{D}} \phi_i(\mathbf{x}) \int_{\mathcal{D}} k_{\theta}(\mathbf{x}, \mathbf{x}') \phi_j(\mathbf{x}') \, d\mathbf{x}' \, d\mathbf{x} \\
 &= \theta^2 \int_{\mathcal{D}} \phi_i(\mathbf{x}) \int_{\mathcal{D}} \delta(\mathbf{x}, \mathbf{x}') \phi_j(\mathbf{x}') \, d\mathbf{x}' \, d\mathbf{x} \\
 &= \theta^2 \int_{\mathcal{D}} \phi_i(\mathbf{x}) \phi_j(\mathbf{x}) \, d\mathbf{x} = \theta^2 \mathbf{M}_{ij}.
 \end{aligned}$$

Using the white noise kernel gives the mass matrix  $\mathbf{M}$  as the covariance matrix. Covariance localisation (see, e.g., [Houtekamer and Mitchell \(2001\)](#)) may also be used to enforce a sparsity constraint, which would set entries below some threshold to zero. An alternative is our scalable approach of Chapter 5, which makes a low-rank approximation such that  $\mathbf{G}_{\theta} \approx \mathbf{L}_{\theta} \mathbf{L}_{\theta}^{\top}$ ,  $\mathbf{L}_{\theta} \in \mathbb{R}^{n_u \times k}$ , with  $k \ll n_u$ . If  $\mathbf{G}_{\theta}$  has a rapidly decaying spectrum this can be very efficient.

### 3.1.2 The linear statFEM posterior

Observations are also assumed to be taken, following the data generating process  $\mathbf{y} = \mathbf{H}\mathbf{u} + \boldsymbol{\eta}$ . Instead of solving the classical forward problem of UQ (which is given by the statFEM prior), statFEM synthesises measurements and the prior model to give a posterior distribution over the finite element coefficients. This posterior distribution provides a probabilistic description of the belief in the system state, conditioned on measurements and modelling assumptions.

The data  $\mathbf{y} \in \mathbb{R}^{n_y}$  are assumed to be corrupted by an independent measurement noise process,  $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{n_y})$ , and are observed via the linear observation operator  $\mathbf{H} : \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_y}$ . This defines the likelihood  $p(\mathbf{y} | \mathbf{u}, \sigma) = \mathcal{N}(\mathbf{H}\mathbf{u}, \sigma^2 \mathbf{I}_{n_y})$ , and the statFEM posterior distribution is the combination of this likelihood with the statFEM prior

$$\begin{aligned}
 p(\mathbf{u} | \mathbf{y}, \boldsymbol{\theta}, \sigma, \Lambda) &\propto p(\mathbf{y} | \mathbf{u}, \sigma) p(\mathbf{u} | \boldsymbol{\theta}, \Lambda) \\
 &= \mathcal{N}(\mathbf{m}_{u|y}, \mathbf{C}_{u|y}),
 \end{aligned}$$

where

$$\begin{aligned}\mathbf{m}_{u|y} &= \mathbf{m}_u + \mathbf{C}_u^\top \mathbf{H}^\top (\mathbf{H} \mathbf{C}_u \mathbf{H}^\top + \sigma^2 \mathbf{I}_{n_y})^{-1} (\mathbf{y} - \mathbf{H} \mathbf{m}_u), \\ \mathbf{C}_{u|y} &= \mathbf{C}_u - \mathbf{C}_u^\top \mathbf{H}^\top (\mathbf{H} \mathbf{C}_u \mathbf{H}^\top + \sigma^2 \mathbf{I}_{n_y})^{-1} \mathbf{H} \mathbf{C}_u.\end{aligned}$$

This update is a weighted linear combination of information available from prior physical assumptions and observed data. If multiple datasets are observed,  $\mathbf{y}_i = \mathbf{H} \mathbf{u} + \boldsymbol{\eta}_i$ , for  $i = 1, \dots, n_{\text{obs}}$ , and the complete dataset is  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_{n_{\text{obs}}}) \in \mathbb{R}^{n_y \times n_{\text{obs}}}$ . Assuming that these  $\mathbf{y}_i$  are independent gives the statFEM posterior as

$$\begin{aligned}p(\mathbf{u} | \mathbf{y}, \boldsymbol{\theta}, \sigma, \Lambda) &= \mathcal{N}(\mathbf{m}_{u|y}, \mathbf{C}_{u|y}), \\ \mathbf{m}_{u|y} &= \mathbf{C}_u \left( \mathbf{A}^\top \mathbf{G}_\theta^{-1} \mathbf{A} \mathbf{m}_u + \frac{1}{\sigma^2} \sum_{i=1}^{n_{\text{obs}}} \mathbf{H}^\top \mathbf{y}_i \right), \\ \mathbf{C}_{u|y}^{-1} &= \mathbf{A}^\top \mathbf{G}_\theta^{-1} \mathbf{A} + \frac{n_{\text{obs}}}{\sigma^2} \mathbf{H}^\top \mathbf{H}.\end{aligned}$$

In practice this may arise when a series of measurements are taken at the same sensor locations across time, whilst retaining a static physical model.

### Discussion

Similar to the classical convergence analysis of FEM solutions (c.f. Section 2.1.2), convergence rates of the statFEM prior and posterior measures are contained in [Papandreou et al. \(2021\)](#). However instead of distances between the true and approximate PDE solutions, rates of convergence in this case are bounds on the distances between the true and discretised measures. The distance between measures is the Wasserstein-2 distance. For two probability measures,  $(\mu, \nu)$ , defined on the same underlying Banach space<sup>4</sup>  $X$  with norm  $\|\cdot\|$ , the Wasserstein-2 distance is

$$W_2(\mu, \nu) = \left( \inf_{\gamma \in \Gamma(\mu, \nu)} \int \|x - x'\|^2 \gamma(dx, dx') \right)^{1/2},$$

where  $\Gamma(\cdot, \cdot)$  is the space of probability measures that have  $\mu$  and  $\nu$  as the first and second marginals, respectively ([Villani, 2009](#)). To compute these distances the statFEM prior and posterior distributions are first derived as Gaussian processes, as opposed to the multivariate Gaussians introduced above.

Representing (3.1) in the operator form gives

$$L_\Lambda u(\mathbf{x}) = \xi(\mathbf{x}), \quad \xi(\mathbf{x}) \sim \mathcal{GP}(f(\mathbf{x}), k_\theta(\mathbf{x}, \mathbf{x})),$$

with the assumed Dirichlet boundary conditions  $u = 0$  for  $\mathbf{x} \in \partial\mathcal{D}$ . The true prior Gaussian process is

$$\begin{aligned}p(u(\mathbf{x}) | \boldsymbol{\theta}, \Lambda) &\sim \mathcal{GP}(m_u(\mathbf{x}), k_u(\mathbf{x}, \mathbf{x}')), \\ m_u(\mathbf{x}) &= L_\Lambda^{-1} f(\mathbf{x}), \\ k_u(\mathbf{x}, \mathbf{x}') &= L_{\Lambda, \mathbf{x}}^{-1} k_\theta(\mathbf{x}, \mathbf{x}') L_{\Lambda, \mathbf{x}'}^{-1}.\end{aligned} \tag{3.4}$$

<sup>4</sup>The Banach structure is not necessary; one can use a metric space and the definition still holds.

Depending on the context it is assumed that  $L_\Lambda^{-1}$  is defined from the weak form (as in Section 3.2) or from the strong form (being given from the Greens function of (3.1)). Making a finite element discretisation gives the statFEM prior

$$\begin{aligned} p(u_h(\mathbf{x}) \mid \boldsymbol{\theta}, \Lambda) &= \mathcal{GP}(m_{u,h}(\mathbf{x}), k_{u,h}(\mathbf{x}, \mathbf{x}')), \\ m_{u,h}(\mathbf{x}) &= \boldsymbol{\Phi}(\mathbf{x})^\top \mathbf{A}^{-1} \mathbf{b}, \\ k_{u,h}(\mathbf{x}, \mathbf{x}') &= \boldsymbol{\Phi}(\mathbf{x})^\top \mathbf{A}^{-1} \mathbf{G}_\theta \mathbf{A}^{-\top} \boldsymbol{\Phi}(\mathbf{x}'), \\ \boldsymbol{\Phi}(\mathbf{x}) &= [\phi_1(\mathbf{x}), \dots, \phi_{n_u}(\mathbf{x})] \in \mathbb{R}^{n_u}, \end{aligned} \tag{3.5}$$

which induces a Gaussian process from the Gaussian distribution over the FEM coefficients.

Writing out the data generating process as  $y_i = u(\mathbf{x}_i) + \eta_i$ ,  $\eta_i \sim \mathcal{N}(0, \sigma^2)$ , for observation locations  $\mathcal{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{n_y})$ , gives the posterior  $\mathcal{GP}$

$$\begin{aligned} p(u(\mathbf{x}) \mid \mathbf{y}, \boldsymbol{\theta}, \Lambda, \sigma) &\sim \mathcal{GP}(m_{u|y}(\mathbf{x}), k_{u|y}(\mathbf{x}, \mathbf{x}')), \\ m_{u|y}(\mathbf{x}) &= m_u(\mathbf{x}) + \mathbf{k}_u(\mathbf{x}, \mathcal{X})(\mathbf{K}_u + \sigma^2 \mathbf{I})^{-1} (\mathbf{y} - \mathbf{m}_u(\mathcal{X})), \\ k_{u|y}(\mathbf{x}, \mathbf{x}') &= k_u(\mathbf{x}, \mathbf{x}') - \mathbf{k}_u(\mathbf{x}, \mathcal{X})(\mathbf{K}_u + \sigma^2 \mathbf{I})^{-1} \mathbf{k}_u(\mathcal{X}, \mathbf{x}'), \end{aligned} \tag{3.6}$$

for  $\mathbf{k}_u(\mathbf{x}, \mathcal{X}) = \mathbf{k}_u(\mathcal{X}, \mathbf{x})^\top = (k_u(\mathbf{x}, \mathbf{x}_1), \dots, k_u(\mathbf{x}, \mathbf{x}_{n_y}))$ ,  $\mathbf{K}_{u,ij} = k_u(\mathbf{x}_i, \mathbf{x}_j)$ . The finite element discretised posterior has the similar form

$$\begin{aligned} p(u_h(\mathbf{x}) \mid \mathbf{y}, \boldsymbol{\theta}, \Lambda, \sigma) &\sim \mathcal{GP}(m_{u|y,h}(\mathbf{x}), k_{u|y,h}(\mathbf{x}, \mathbf{x}')), \\ m_{u|y,h}(\mathbf{x}) &= m_{u,h}(\mathbf{x}) + \mathbf{k}_{u,h}(\mathbf{x}, \mathcal{X})(\mathbf{K}_{u,h} + \sigma^2 \mathbf{I})^{-1} (\mathbf{y} - \mathbf{m}_{u,h}(\mathcal{X})), \\ k_{u|y,h}(\mathbf{x}, \mathbf{x}') &= k_{u,h}(\mathbf{x}, \mathbf{x}') - \mathbf{k}_{u,h}(\mathbf{x}, \mathcal{X})(\mathbf{K}_{u,h} + \sigma^2 \mathbf{I})^{-1} \mathbf{k}_{u,h}(\mathcal{X}, \mathbf{x}'), \end{aligned} \tag{3.7}$$

for  $\mathbf{k}_{u,h}(\mathbf{x}, \mathcal{X}) = \mathbf{k}_{u,h}(\mathcal{X}, \mathbf{x})^\top = (k_{u,h}(\mathbf{x}, \mathbf{x}_1), \dots, k_{u,h}(\mathbf{x}, \mathbf{x}_{n_y}))$ ,  $\mathbf{K}_{u,h,ij} = k_{u,h}(\mathbf{x}_i, \mathbf{x}_j)$ .

Denote by  $\mu_0$  and  $\mu_{0,h}$  the Gaussian measures on  $L^2(\mathcal{D}; \mathbb{R})$  which are induced by (3.4) and (3.5), respectively. Note that in this analysis it is assumed that  $L_\Lambda^{-1} : L^2(\mathcal{D}; \mathbb{R}) \rightarrow L^2(\mathcal{D}; \mathbb{R})$ . Papandreou et al. (2021) show that on  $L^2(\mathcal{D}; \mathbb{R})$  the finite-dimensional statFEM prior  $\mu_{0,h}$  and the true prior  $\mu_0$  have a bounded Wasserstein-2 distance  $W_2(\mu_{0,h}, \mu_0) \leq Ch^2$ , for some  $C > 0$ . Similarly, denote by  $\mu^y$  and  $\mu_h^y$  as the true and discretised statFEM posterior Gaussian measures induced by (3.6) and (3.7), respectively, on  $L^2(\mathcal{D}; \mathbb{R})$ . The authors show that the Wasserstein-2 distance of bounded linear functionals  $b : L^2(\mathcal{D}; \mathbb{R}) \rightarrow \mathbb{R}$  can again be bounded  $W_2(b(\mu_h^y), b(\mu^y)) \leq C(b)h^2 + \mathcal{O}(h^4)$  with the scale factor  $C(\cdot)$  independent of  $h$ . The rate of convergence in the prior measure is the same as that of the true and discrete PDE solutions; the same rate of convergence also holds for functionals of the posterior measure.

Error rates on the statFEM posterior distribution with and without discretisation are given in Karvonen et al. (2022). In this case it is assumed that the observations are drawn according to a deterministic elliptic PDE  $L_\Lambda \tilde{u} = \tilde{f}$  where there may be some mismatch between  $\tilde{f}$  and  $f$ . Different to Papandreou et al. (2021) the function-space setting is taken to be the Sobolev space  $H^m(\mathcal{D}; \mathbb{R})$ , which is assumed to be *norm-equivalent*<sup>5</sup> to the RKHS induced via  $k_\theta(\cdot, \cdot)$ . This holds for the Matérn covariance

<sup>5</sup>On a space  $X$ , the norms  $\|\cdot\|_a$  and  $\|\cdot\|_b$  are defined to be equivalent if there exist positive constants  $C_1, C_2$  such that  $C_1 \|x\|_a \leq \|x\|_b \leq C_2 \|x\|_a$  for all  $x \in X$ .

kernel with smoothness parameter  $\nu$ , being norm-equivalent to  $H^{\nu+d/2}(\mathcal{D}; \mathbb{R})$ . It is assumed that  $L_{\Lambda}^{-1} : H^m(\mathcal{D}; \mathbb{R}) \rightarrow C^2(\mathcal{D}; \mathbb{R})$ , and that  $f \in H^m(\mathcal{D}; \mathbb{R})$ ,  $\tilde{f} \in H^m(\mathcal{D}; \mathbb{R}) \cap C^k(\mathcal{D}; \mathbb{R})$ , where  $m > d/2$ . In the absence of finite element discretisation the  $L^2(\mathcal{D}; \mathbb{R})$  error between the posterior mean and true solution  $\tilde{u}$  can be bounded

$$\mathbb{E} [\|\tilde{u} - m_{u|y}\|] \leq C n_y^{-1/2+d/(4m)},$$

under the additional assumption of quasi-uniform data (Wendland, 2004).

Next, the posterior is discretised using finite elements as in (3.7). As previous  $n_u$  basis functions are used with degree  $p$  polynomials, to give the  $L^2(\mathcal{D}; \mathbb{R})$  discretised error

$$\mathbb{E} [\|\tilde{u} - m_{u|y,h}\|] \leq C(n_y^{-1/2+d/(4m)} + \sigma^{-1} n_u^{-(p+1)} n_y^{3/2}).$$

Therefore to have the optimal rate of convergence it is desirable to choose  $n_u$  such that  $n_u = n_y^{(2-d/(4k))/(p+1)}$ .

$$\mathbb{E} [\|\tilde{u} - m_{u|y,h}\|] \leq C(1 + \sigma^{-1}) n_y^{-1/2+d/(4m)}.$$

Karvonen et al. (2022) note that despite this choice of  $n_u$  giving the smallest possible *rate* of convergence, this may not give the smallest RHS due to the constants involved.

Hyperparameters,  $\theta$ , and noise standard deviation,  $\sigma$ , can be fixed *a priori* or estimated from the log-marginal likelihood. For the single-dataset case, the marginal likelihood is given by  $p(\mathbf{y} | \theta, \sigma, \Lambda) = \mathcal{N}(\mathbf{H}\mathbf{m}_u, \mathbf{H}\mathbf{C}_u\mathbf{H}^\top + \sigma^2\mathbf{I}_{n_y})$ , and for multiple datasets this is  $p(\mathbf{y} | \theta, \sigma, \Lambda) = \prod_{i=1}^{n_{\text{obs}}} p(\mathbf{y}_i | \theta, \sigma, \Lambda)$ . Depending on the computational budget and/or the requirement for UQ,  $\theta$  may be estimated using either optimisation (e.g. maximum marginal likelihood (Williams and Rasmussen, 2006), maximum-a-posteriori (MAP) (Murphy, 2012)) or via sampling (e.g. MCMC (Casella and George, 1992; Chib and Greenberg, 1995; Tierney, 1994)). In the original statFEM paper, the sampling approach is taken.

The statFEM prior provides a physically-motivated regularisation, through interpreting the prior as an  $l^2$  regulariser with mean  $\mathbf{m}_u$ , weighted by  $\mathbf{C}_u$  (Tarantola, 2005):

$$\log p(\mathbf{u} | \mathbf{y}, \theta, \sigma, \Lambda) \propto -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{H}\mathbf{u})^\top (\mathbf{y} - \mathbf{H}\mathbf{u}) - \frac{1}{2} (\mathbf{u} - \mathbf{m}_u)^\top \mathbf{C}_u^{-1} (\mathbf{u} - \mathbf{m}_u).$$

Without this regularisation, estimating  $n_u$  FEM coefficients from  $n_y$  observations may be ill-posed if  $n_u \gg n_y$ ; the statFEM prior ensures well-posedness.

The posterior is a multivariate Gaussian due to the Gaussian prior and the Gaussian likelihood. The Gaussian prior results from the use of a linear PDE with stochasticity in the forcing function only. Discretisation ensures that this stochastic forcing induces a multivariate Gaussian<sup>6</sup>. Nonlinearity would result in a non-Gaussian prior, whose distributional form depends on the underlying PDE nonlinearity (see, e.g., Snelson et al., 2004). A non-Gaussian prior could also result from stochasticity entering from the diffusivity function  $\Lambda$  or from the boundary conditions. For example if the diffusivity parameter is modelled by a log-Gaussian process, with  $\log \Lambda \sim \mathcal{GP}(m_\Lambda, k_\Lambda)$  (Gunzburger et al., 2014), the prior  $p(\mathbf{u} | \theta) = \int p(\mathbf{u} | \Lambda) p(\Lambda) d\Lambda$  and posterior  $p(\mathbf{u} | \mathbf{y}, \theta, \sigma) = \int p(\mathbf{u} | \Lambda, \mathbf{y}) p(\Lambda) d\Lambda$  are not available in closed form. Marginalising over  $\Lambda$  is required in order to fully quantify the induced uncertainty in the model specification. Monte Carlo methods are required in order to characterise these measures,

<sup>6</sup>As linear mappings of Gaussians and  $\mathcal{GP}$ s are also Gaussian (subject to regularity conditions on the kernel).



using, for example, multilevel Monte Carlo (Cliffe et al., 2011; Teckentrup et al., 2013), quasi Monte Carlo (Graham et al., 2011), or the more recent unadjusted Langevin algorithm (Akyildiz et al., 2022).

StatFEM shares the same foundation as the Bayesian inverse problem of determining the forcing  $\xi$  from data  $\mathbf{y}$ , however there is a difference in the measure of interest. In the inverse problem, this is the posterior  $p(\xi | \mathbf{y}, \boldsymbol{\theta}, \sigma, \Lambda)$ , and it is assumed that  $u := u(\xi)$  where the parameter-to-data mapping is given through pushing  $\xi$  through the PDE in Equation (3.1). The measure of interest in the statFEM approach, however, is  $p(\mathbf{u} | \mathbf{y}, \boldsymbol{\theta}, \sigma, \Lambda)$  which integrates over the unknown forcing  $\xi$  to give the posterior over the FEM coefficients. Depending on the inferential situation one approach may be preferred over the other. Given the statFEM discretisation and DGP, if the estimation of  $\xi$  is of interest, then  $\boldsymbol{\xi} = (\langle \xi, \phi_1 \rangle, \dots, \langle \xi, \phi_{n_\xi} \rangle)$ ,  $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{b}, \mathbf{G}_\theta)$ . As  $\mathbf{u} = \mathbf{A}^{-1}\boldsymbol{\xi}$ , the posterior is

$$p(\boldsymbol{\xi} | \mathbf{y}, \boldsymbol{\theta}, \sigma, \Lambda) = \mathcal{N}(\mathbf{m}_{\xi|y}, \mathbf{C}_{\xi|y}), \quad (3.8)$$

where

$$\begin{aligned} \mathbf{m}_{\xi|y} &= \mathbf{b} + \mathbf{G}_\theta^\top \mathbf{H}^\top \mathbf{A}^{-\top} (\mathbf{H} \mathbf{A}^{-1} \mathbf{G}_\theta \mathbf{A}^{-\top} \mathbf{H}^\top + \sigma^2 \mathbf{I}_{n_y})^{-1} (\mathbf{y} - \mathbf{H} \mathbf{A}^{-1} \mathbf{b}), \\ \mathbf{C}_{u|y} &= \mathbf{G}_\theta - \mathbf{G}_\theta^\top \mathbf{H}^\top \mathbf{A}^{-\top} (\mathbf{H} \mathbf{A}^{-1} \mathbf{G}_\theta \mathbf{A}^{-\top} \mathbf{H}^\top + \sigma^2 \mathbf{I}_{n_y})^{-1} \mathbf{H} \mathbf{A}^{-1} \mathbf{G}_\theta. \end{aligned}$$

As in the statFEM problem *a-priori* assumptions on  $\xi$  act to regularise the otherwise ill-posed estimation problem.

**Example 4.** To illustrate the conditioning procedure, the previously considered elliptic partial differential equation with random forcing is updated with data. This equation is

$$\begin{aligned} -\nabla \cdot (\Lambda(\mathbf{x}) \nabla u(\mathbf{x})) &= \xi(\mathbf{x}), \\ u(\mathbf{x}) &= 0, \quad \mathbf{x} \in \partial \mathcal{D}, \end{aligned} \quad (3.9)$$

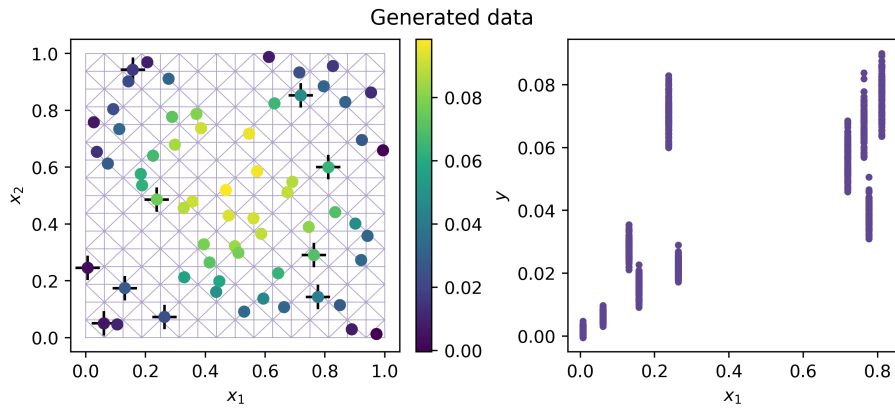
where the domain is the unit square,  $\mathcal{D} = [0, 1] \times [0, 1]$ , so  $\mathbf{x} = (x_1, x_2)$  where  $\Lambda(\mathbf{x}) = 1 + 0.3 \sin(\pi(x_1 + x_2))$ ,  $\xi(\mathbf{x}) \sim \mathcal{GP}(1, \theta^2 \delta(\mathbf{x} - \mathbf{x}'))$ , with  $\theta = 0.05$ .

Discretisation with finite elements gives the prior measure  $p(\mathbf{u} | \theta, \Lambda) = \mathcal{N}(\mathbf{A}^{-1}\mathbf{b}, \mathbf{A}^{-1}\mathbf{G}_\theta\mathbf{A}^{-\top})$ , which uses a  $16 \times 16$  cell mesh with piecewise linear polynomial basis functions, so that the state dimension is  $n_u = 289$ . A diagonal approximation to  $\mathbf{G}_\theta$  is made, setting  $\mathbf{G}_{\theta,ii} = \theta^2 \sum_j \mathbf{M}_{ij}$ . This is known as *lumping* in the FEM literature (Zienkiewicz et al., 2013).

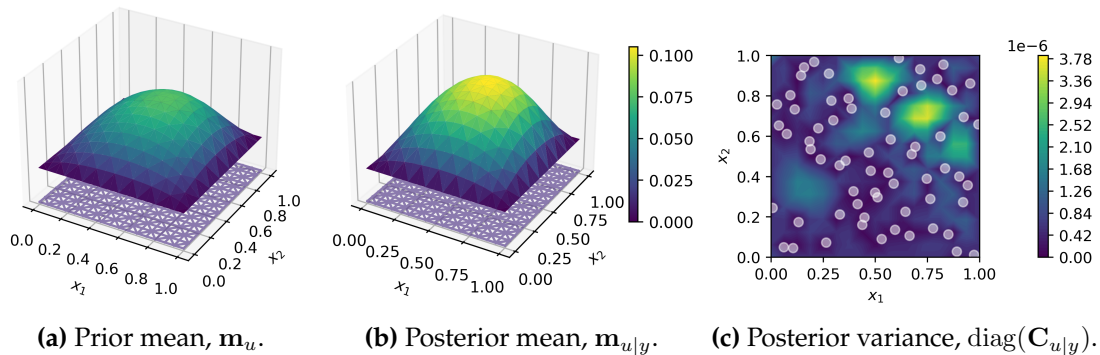
In this example, synthetic data is used. These synthetic data are generated from Equation (3.9) with the additional stochasticity,  $\log \Lambda(\mathbf{x}) \sim \mathcal{GP}(\log(1 + 0.3 \sin(\pi(x_1 + x_2))), k_\Lambda(\mathbf{x}, \mathbf{x}'))$ , setting  $k_\Lambda(\mathbf{x}, \mathbf{x}') = 0.2^2 \exp(-\|\mathbf{x} - \mathbf{x}'\|^2 / (2 \cdot 0.2^2))$ . To generate the data,  $n_{\text{obs}} = 100$  stochastic PDE solutions are interpolated onto  $n_y = 64$  observation locations. Model mismatch is given by scaling the solutions by a known constant of 1.4, and adding on observational noise  $\eta \sim \mathcal{N}(0, 0.001^2 \mathbf{I}_{n_y})$ . These data are shown in Figure 3.1. Observations approximately cover the finite element mesh, with a value-dependent variance: the  $\mathcal{GP}$ 's inside of the PDE dominate the observational noise. In this case the data are not necessarily sparsely observed, covering  $\approx 22\%$  of the total state dimension.

In computing the posterior, hyperparameters are fixed to  $\sigma = 0.001$  and  $\theta = 0.05$ . Owing to the conjugate Gaussian model, the posterior is  $p(\mathbf{u} | \mathbf{y}, \theta, \sigma, \Lambda) = \mathcal{N}(\mathbf{m}_{u|y}, \mathbf{C}_{u|y})$ . Results are shown





**Figure 3.1:** Generated data for the Poisson example. Left: single dataset of the  $n_{\text{obs}} = 100$  datasets, with the FEM mesh. Points marked with crosshairs are shown as a 1D slice on the figure on the left, which are ordered according to their  $x_1$  value. The left figure shows all  $n_{\text{obs}} = 100$  at each location.



**Figure 3.2:** StatFEM posterior results for the Poisson example. Left: statFEM prior mean. Centre: statFEM posterior mean. Note that both posterior mean plots are overlaid on the FEM mesh. Right: posterior pointwise variance field.

in Figure 3.2, which shows the statFEM prior and posterior mean (Figures 3.2a and 3.2b), and the posterior variance (Figure 3.2c). The posterior mean more closely aligns with that of the data generating process, appearing sensibly updated with observations. The posterior variance field shrinks to small values near observation locations, growing larger where data is not observed. In comparison to the mean the magnitude of the variance is small; given observations, the statFEM model is highly certain.

### 3.2 Linear, time-dependent statFEM

We now provide our first extension to the statFEM methodology, via extension to time-dependent problems. This was first detailed in the Supplementary Information of [Duffin et al. \(2021\)](#). As a motivating example, we consider the general parabolic PDE with Dirichlet boundaries and space-time stochastic forcing. Denote by  $L_\Lambda$  the general linear elliptic differential operator, with known parameters  $\Lambda$ . Unless otherwise mentioned,  $\Lambda$  is implicitly conditioned on through the remainder of this chapter.

The general parabolic equation is

$$\begin{aligned}\partial_t u + L_\Lambda u &= f + \xi, \\ u &= 0, \quad \mathbf{x} \in \partial\mathcal{D},\end{aligned}\tag{3.10}$$

where  $\mathbf{x} \in \mathcal{D} \subseteq \mathbb{R}^d$ ,  $t \in [0, T]$ ,  $\xi := \xi(\mathbf{x}, t)$ ,  $u := u(\mathbf{x}, t)$ ,  $u : \mathcal{D} \rightarrow [0, T]$ . It is assumed  $u(\mathbf{x}, 0) = u_0(\mathbf{x})$ ,  $f \in L^2(\mathcal{D}; \mathbb{R})$ , and, without loss of generality,  $u \in \mathbb{R}$ . Multidimensional states  $\mathbf{u} \in \mathbb{R}^s$  are most common and are studied in Chapters 4 and 5.

For tractability the covariance function is assumed to have the separable form

$$\xi(\mathbf{x}, t) \sim \mathcal{GP}(0, \delta(t - t') \cdot k_\theta(\mathbf{x}, \mathbf{x}')), \quad \xi(\mathbf{x}, 0) \equiv 0,\tag{3.11}$$

implying that stochastic forcing is white noise in time and spatially regular as per  $k_\theta(\cdot, \cdot)$ ; for fixed  $t$ ,  $\xi(\mathbf{x}, t) \in L^2(\mathcal{D}; \mathbb{R})$ . A white noise process is assumed to write the above as a stochastic differential equation, forced with a spatially regular Brownian motion process.

**Remark.** The stochastic forcing is an  $L^2(\mathcal{D}; \mathbb{R})$ -valued Wiener process with Gaussian process increments, i.e.  $\xi(\mathbf{x}, t + \Delta t) - \xi(\mathbf{x}, t) \sim \mathcal{GP}(0, \Delta t k_\theta(\mathbf{x}, \mathbf{x}'))$ , and thus (3.10) is a function-valued stochastic differential equation (SDE).

Multiplying by test functions  $\psi(\mathbf{x}) \in V = H_0^1(\mathcal{D}; \mathbb{R})$  and integrating over the problem domain  $\mathcal{D}$  gives the time-dependent weak form

$$\langle \partial_t u, \psi \rangle + \mathcal{A}_\Lambda(u, \psi) = \langle f, \psi \rangle + \langle \xi, \psi \rangle,$$

where  $\mathcal{A}_\Lambda(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$  is the bilinear form generated from  $L_\Lambda$ .

Choosing a triangulation  $\mathcal{D}_h$  with vertices  $\{\mathbf{x}_j\}_{j=1}^{n_u}$  discretises the domain, upon which the basis functions  $\{\phi_j\}_{j=1}^{n_u}$  are defined. The basis functions have  $\phi_j(\mathbf{x}_i) = \delta_{ij}$  and  $\phi_j(\mathbf{x}) = 0$  for  $\mathbf{x} \in \partial\mathcal{D}$ . The finite-dimensional test and trial spaces is  $V_h := \text{span}\{\phi_j\}_{j=1}^{n_u}$ , with  $V_h \subset V$ . Writing out the solution as  $u_h(\mathbf{x}, t) = \sum_{i=1}^{n_u} u_i(t) \phi_i(\mathbf{x})$  and replacing the testing function with  $\phi_j$  gives

$$\langle \partial_t u_h, \phi_j \rangle + \mathcal{A}_\Lambda(u_h, \phi_j) = \langle f, \phi_j \rangle + \langle \xi, \phi_j \rangle, \quad j = 1, \dots, n_u,$$

implicitly defining an  $n_u$ -dimensional SDE for the finite element coefficients. The Gaussian process  $\boldsymbol{\xi}(t) = (\langle \xi, \phi_1 \rangle, \dots, \langle \xi, \phi_{n_u} \rangle)^\top \sim \mathcal{GP}(\mathbf{0}, \delta(t - t') \cdot \mathbf{G}_\theta)$  can be informally thought of as the derivative of a Brownian motion process  $\boldsymbol{\beta}(t)$  with diffusion matrix  $\mathbf{G}_\theta$  (Øksendal, 2003; Särkkä and Solin, 2019). The covariance is

$$\mathbb{E}[\langle \xi, \phi_i \rangle \langle \xi, \phi_j \rangle] = \delta(t - t') \cdot \int_{\mathcal{D}} \phi_i(\mathbf{x}) \int_{\mathcal{D}} k(\mathbf{x}, \mathbf{x}') \phi_j(\mathbf{x}') \, d\mathbf{x}' \, d\mathbf{x} = \delta(t - t') \cdot \mathbf{G}_{\theta, ij},$$

inheriting the spatial correlation structure from  $k_\theta(\cdot, \cdot)$ , which imbues some smoothness.

Concatenating the FEM coefficients into the vector  $\mathbf{u}(t) = (u_1(t), \dots, u_{n_u}(t))^\top \in \mathbb{R}^{n_u}$  gives the SDE

$$\mathbf{M} \, d\mathbf{u} + \mathbf{A} \mathbf{u} \, dt = \mathbf{b} \, dt + d\boldsymbol{\beta}(t),$$

and to discretise in time we use the notation  $\mathbf{u}_n = (u_1(n\Delta t), \dots, u_{n_u}(n\Delta t))^\top$  to represent the solution

vector at time  $n\Delta_t$ . The explicit Euler discretisation (Hairer et al., 1993; Kloeden and Platen, 1992) is

$$\mathbf{M}(\mathbf{u}_n - \mathbf{u}_{n-1}) + \Delta_t \mathbf{A} \mathbf{u}_{n-1} = \Delta_t \mathbf{b} + \mathbf{e}_{n-1},$$

where  $\mathbf{e}_{n-1} = \beta_n - \beta_{n-1} \sim \mathcal{N}(\mathbf{0}, \Delta_t \mathbf{G}_\theta)$ . To avoid possible instabilities we also introduce the *implicit Euler*

$$\mathbf{M}(\mathbf{u}_n - \mathbf{u}_{n-1}) + \Delta_t \mathbf{A} \mathbf{u}_n = \Delta_t \mathbf{b} + \mathbf{e}_{n-1},$$

and the *Crank-Nicolson*

$$\mathbf{M}(\mathbf{u}_n - \mathbf{u}_{n-1}) + \Delta_t \mathbf{A} \mathbf{u}_{n-1/2} = \Delta_t \mathbf{b} + \mathbf{e}_{n-1},$$

with  $\mathbf{u}_{n-1/2} := (\mathbf{u}_n + \mathbf{u}_{n-1})/2$ . For exposition in the rest of this section we consider only the explicit Euler method, for which the updating equation can be written as

$$\mathbf{u}_n = (\mathbf{I} - \Delta_t \mathbf{M}^{-1} \mathbf{A}) \mathbf{u}_{n-1} + \Delta_t \mathbf{M}^{-1} \mathbf{b} + \mathbf{M}^{-1} \mathbf{e}_{n-1},$$

which defines the conditional measure at time  $n$  to be

$$p(\mathbf{u}_n | \mathbf{u}_{n-1}, \boldsymbol{\theta}) = \mathcal{N}\left((\mathbf{I} - \Delta_t \mathbf{M}^{-1} \mathbf{A}) \mathbf{u}_{n-1} + \Delta_t \mathbf{M}^{-1} \mathbf{b}, \Delta_t \mathbf{M}^{-1} \mathbf{G}_\theta \mathbf{M}^{-\top}\right).$$

Sequentially marginalising over the uncertainty in the previous timestep defines the prior at timestep  $n$ ,  $p(\mathbf{u}_n | \boldsymbol{\theta}, \Lambda) = \int p(\mathbf{u}_n | \mathbf{u}_{n-1}, \boldsymbol{\theta}, \Lambda) p(d\mathbf{u}_{n-1} | \boldsymbol{\theta}, \Lambda)$ . From here on in PDE coefficients  $\Lambda$  are implicitly conditioned on, writing  $p(\mathbf{u}_n | \boldsymbol{\theta}, \Lambda) \equiv p(\mathbf{u}_n | \boldsymbol{\theta})$ .

### 3.2.1 The linear, time-dependent posterior

Data  $\mathbf{y}_n \in \mathbb{R}^{n_y}$  are observed at time  $n\Delta_t$  on the grid  $\mathbf{x}_{\text{obs}}$ . To synthesise the FEM model with data the prior distribution is updated to give a sequence of posterior distributions  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \boldsymbol{\theta}, \sigma)$ , where  $\mathbf{y}_{1:n} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ . The posterior distribution describes the belief in the model given observations up to and including the current time point and, as in the linear case, quantifies all assumed sources of uncertainty in the model and the data generating process. This solves the *filtering* problem and the posterior distribution of interest is referred as the *filtering distribution* or just the *posterior*, when the context is clear.

These data are corrupted with independent and identically distributed (i.i.d.) noise  $\boldsymbol{\eta}_n \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_y})$  independent to the model  $\mathbf{u}_n$ , to give the data generating process  $\mathbf{y}_n = \mathbf{H}_n \mathbf{u}_n + \boldsymbol{\eta}_n$ . As in the static case, the linear observation operator  $\mathbf{H}_n : \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_y}$  maps from the computed solution grid to the observation grid using the FEM interpolant. Statistically, this is a linear Gaussian state-space model<sup>7</sup> (Shumway and Stoffer, 2017), so the standard Kalman filter can be applied to obtain the posterior distribution (Kalman, 1960).

The initial conditions are known and are described by a Dirac measure,  $p(\mathbf{u}_0) = \delta_{\mathbf{m}_0}(d\mathbf{u}_0)$ . For time  $n$ , the prediction step is first made, propagating uncertainty in the previous timestep to give the prediction distribution  $p(\mathbf{u}_n | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}, \sigma)$ . Data observed at time  $n\Delta_t$  is conditioned on to give the updated filtering distribution  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \boldsymbol{\theta}, \sigma)$ . Unless stated otherwise,  $\mathcal{GP}$  hyperparameters  $\boldsymbol{\theta}$  and

<sup>7</sup>With the underlying evolution being defined from a PDE.

observational noise variance  $\sigma$  are known.

If the previous filtering distribution is  $p(\mathbf{u}_{n-1} \mid \mathbf{y}_{1:n-1}, \boldsymbol{\theta}, \sigma) = \mathcal{N}(\mathbf{m}_{n-1|n-1}, \mathbf{C}_{n-1|n-1})$ , then for  $n = 1, \dots, n_t$  the filter proceeds as:

1. (Prediction step) Compute the predicted mean and covariance:

$$\begin{aligned} p(\mathbf{u}_n \mid \mathbf{y}_{1:n-1}, \boldsymbol{\theta}, \sigma) &= \mathcal{N}(\mathbf{m}_{n|n-1}, \mathbf{C}_{n|n-1}), \\ \mathbf{m}_{n|n-1} &= (\mathbf{I} - \Delta_t \mathbf{M}^{-1} \mathbf{A}) \mathbf{m}_{n-1|n-1} + \Delta_t \mathbf{M}^{-1} \mathbf{b}, \\ \mathbf{C}_{n|n-1} &= (\mathbf{I} - \Delta_t \mathbf{M}^{-1} \mathbf{A}) \mathbf{C}_{n-1|n-1} (\mathbf{I} - \Delta_t \mathbf{M}^{-1} \mathbf{A})^\top + \Delta_t \mathbf{M}^{-1} \mathbf{G}_\theta \mathbf{M}^{-\top}. \end{aligned}$$

2. (Update step) Update the mean and covariance with data  $\mathbf{y}_n$ :

$$\begin{aligned} p(\mathbf{u}_n \mid \mathbf{y}_{1:n}, \boldsymbol{\theta}, \sigma) &= \mathcal{N}(\mathbf{m}_{n|n}, \mathbf{C}_{n|n}), \\ \mathbf{m}_{n|n} &= \mathbf{m}_{n|n-1} + \mathbf{C}_{n|n-1} \mathbf{H}_n^\top \left( \mathbf{H}_n \mathbf{C}_{n|n-1} \mathbf{H}_n^\top + \sigma^2 \mathbf{I} \right)^{-1} (\mathbf{y}_n - \mathbf{H}_n \mathbf{m}_{n|n-1}), \\ \mathbf{C}_{n|n} &= \mathbf{C}_{n|n-1} - \mathbf{C}_{n|n-1} \mathbf{H}_n^\top \left( \mathbf{H}_n \mathbf{C}_{n|n-1} \mathbf{H}_n^\top + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{H}_n \mathbf{C}_{n|n-1}. \end{aligned}$$

If no data is observed at time point  $n$  then only the prediction step needs to be computed as by definition  $p(\mathbf{u}_n \mid \mathbf{y}_{1:n}, \boldsymbol{\theta}, \sigma) = p(\mathbf{u}_n \mid \mathbf{y}_{1:n-1}, \boldsymbol{\theta}, \sigma)$ .

## Discussion

First, we note the following operation counts. The above filter stores the full covariance matrix  $\mathbf{C}_{n|n}$  at each iteration, so the memory cost  $\mathcal{O}(n_u^2)$ . Each iteration requires the solution of  $\mathbf{M}$ , for which an LU decomposition of  $\mathbf{M}$  can be taken and stored before running the filter. Thus, solving for  $\mathbf{M}^{-1} \mathbf{z}$  can be computed in the filter with forward- and back-substitution. This procedure is  $\mathcal{O}(b^2 n_u)$  to factorise, and  $\mathcal{O}(b n_u)$  to solve, for a mass matrix with bandwidth  $b$ . An alternative method *lumps* the mass matrix, replacing  $\mathbf{M}$  with the diagonal  $\tilde{\mathbf{M}}$  with elements  $\tilde{M}_{ii} = \sum_j M_{ij}$ . However, for the remainder of this discussion we will assume that the LU approach is taken, as this is able to be generalised to implicit discretisation schemes, which require inverting  $\mathbf{A}$ .

In the prediction step, computing the prediction mean has the cost  $\mathcal{O}(b n_u)$ , and computing the prediction covariance, which requires matrices of the form  $\mathbf{M}^{-1} \mathbf{B} \mathbf{M}^{-\top}$ , requires  $\mathcal{O}(b n_u^2)$  operations per timestep. Next, in the update step,  $\mathbf{S}_n = \mathbf{H}_n \mathbf{C}_{n|n-1} \mathbf{H}_n^\top + \sigma^2 \mathbf{I}$  is computed. Recalling that  $\mathbf{H}_n$  is  $(n_y \times n_u)$ , if we assume that  $\mathbf{H}_n$  has  $\mathcal{O}(1)$  entries per row then this has the cost of  $\mathcal{O}(n_y n_u)$ . To solve  $\mathbf{S}_n$ , the Cholesky decomposition is used, which has cost  $\mathcal{O}(n_y^3/3)$  to form  $\mathbf{S}_n = \mathbf{L}_n \mathbf{L}_n^\top$ , and cost  $\mathcal{O}(n_y^2)$  to solve  $\mathbf{S}_n^{-1} \mathbf{z}$ . Computing the mean update thus has the cost  $\mathcal{O}(n_y^2) + \mathcal{O}(n_u n_y)$ , and the covariance update has the cost  $\mathcal{O}(n_u n_y^2) + \mathcal{O}(n_u^2 n_y)$ . In our experience, the main cost has been associated with computing the products of the form  $\mathbf{M}^{-1} \mathbf{B} \mathbf{M}^{-\top}$ . In scaling up the method, the  $\mathcal{O}(b n_u^2)$  complexity of computing this product is prohibitively expensive, as is the  $\mathcal{O}(n_u^2)$  memory requirement.

For static  $\mathcal{GP}$  hyperparameters,  $\boldsymbol{\theta}$ , and noise standard deviation,  $\sigma$ , estimation can proceed via MAP methods (Murphy, 2012). In this case for a single timestep  $n$  the marginal likelihood is

$$p(\mathbf{y}_n \mid \mathbf{y}_{1:n-1}, \boldsymbol{\theta}, \sigma) = \mathcal{N} \left( \mathbf{H} \mathbf{m}_{n|n-1}, \mathbf{H} \mathbf{C}_{n|n-1} (\boldsymbol{\theta}) \mathbf{H}^\top + \sigma^2 \mathbf{I}_{n_y} \right),$$

as  $\mathbf{C}_{n|n-1} \equiv \mathbf{C}_{n|n-1}(\boldsymbol{\theta}_n)$ . Estimation uses the complete-data marginal likelihood  $p(\mathbf{y}_{1:n_t} | \boldsymbol{\theta}, \sigma) = \prod_{k=0}^{n_t-1} p(\mathbf{y}_{n_t-k} | \mathbf{y}_{1:(n_t-k-1)})$ , where  $p(\mathbf{y}_1 | \mathbf{y}_0) := p(\mathbf{y}_1)$ , from which taking logarithms gives (up to a proportionality constant)

$$\log p(\mathbf{y}_{1:n_t} | \boldsymbol{\theta}, \sigma) = -\frac{1}{2} \sum_{n=1}^{n_t} \log |\boldsymbol{\Sigma}_n(\boldsymbol{\theta}, \sigma)| - \frac{1}{2} \sum_{n=1}^{n_t} (\mathbf{y}_n - \mathbf{H}_n \mathbf{m}_{n|n-1})^\top \boldsymbol{\Sigma}_n(\boldsymbol{\theta}, \sigma)^{-1} (\mathbf{y}_n - \mathbf{H}_n \mathbf{m}_{n|n-1}),$$

where  $\boldsymbol{\Sigma}_n(\boldsymbol{\theta}, \sigma) = \mathbf{H}_n \mathbf{C}_{n|n-1}(\boldsymbol{\theta}) \mathbf{H}_n^\top + \sigma^2 \mathbf{I}_{n_y}$ . The MAP estimates are

$$(\hat{\boldsymbol{\theta}}, \hat{\sigma}) = \underset{(\boldsymbol{\theta}, \sigma)}{\operatorname{argmax}} \{ \log p(\mathbf{y}_{1:n_t} | \boldsymbol{\theta}, \sigma) + \log p(\boldsymbol{\theta}) + \log p(\sigma) \}.$$

Note that each evaluation of the log-marginal likelihood requires running the filter with fixed  $(\boldsymbol{\theta}, \sigma)$  to evaluate  $\log p(\mathbf{y}_{1:n_t} | \boldsymbol{\theta}, \sigma)$ .

Alternatively we can assume the time evolving structure  $\boldsymbol{\theta} := \boldsymbol{\theta}_n$ ,  $\sigma := \sigma_n$  for all  $n$ , and that  $\boldsymbol{\theta}_n$  and  $\sigma_n$  are independent across time, i.e.  $p(\boldsymbol{\theta}_n | \boldsymbol{\theta}_{n-1}) \equiv p(\boldsymbol{\theta}_n)$ . To estimate these hyperparameters the prediction step can be split into two, with an additional optimisation step added in. Denote by  $(\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n)$  the estimated hyperparameters at time  $n$ . To estimate these, for each  $n$  we first compute the prediction mean  $\mathbf{m}_{n|n-1}$ . Then,  $(\boldsymbol{\theta}_n, \sigma_n)$  are estimated via

$$(\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n) = \underset{(\boldsymbol{\theta}_n, \sigma_n)}{\operatorname{argmax}} \{ \log p(\mathbf{y}_n | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}_n, \hat{\boldsymbol{\theta}}_{1:n-1}, \sigma_n, \hat{\sigma}_{1:n-1}) + \log p(\boldsymbol{\theta}_n) + \log p(\sigma_n) \},$$

at each iteration  $n$ . After obtaining estimates  $(\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n)$  the prediction covariance is computed,  $\mathbf{C}_{n|n-1} = \mathbf{C}_{n|n-1}(\hat{\boldsymbol{\theta}}_n)$ , and we complete the update step to estimate  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \hat{\boldsymbol{\theta}}_{1:n}, \hat{\sigma}_{1:n})$ .

Knowledge of the physical problem can inform the hyperparameters  $\boldsymbol{\theta}$ , which can be encoded in a prior distribution  $p(\boldsymbol{\theta})$ . Throughout this thesis, we have used the weakly informative priors  $\rho_n \sim \mathcal{N}_+(1, 1^2)$ ,  $\ell_n \sim \mathcal{N}_+(1, 1^2)$ , and  $\sigma_n \sim \mathcal{N}_+(0, 1^2)$ , to reflect our *a priori* uncertainty. For the optimisation routine, L-BFGS-B (Nocedal and Wright, 2006) with positivity constraints, as implemented in SciPy (Virtanen et al., 2020), has worked well in our experience.

Further discussion of this filtering approach is contained in the following section, of which the above presented algorithm arises as a special case of the extended Kalman Filter.

### 3.3 Nonlinear, time-dependent statFEM

We now detail the NL-statFEM, our second extension to the original statFEM, first presented in Duffin et al. (2021). In this case we consider a general nonlinear, time-dependent PDE with stochastic forcing and Dirichlet boundary conditions. As previous the parameters  $\Lambda$  are assumed known and are implicitly conditioned upon. Denote by  $L_\Lambda$  and  $N_\Lambda(\cdot)$  the linear and nonlinear differential operators, respectively, which comprise the dynamics; the PDE is then

$$\begin{aligned} \partial_t u + L_\Lambda u + F_\Lambda(u) + f + \xi &= 0, \\ u &= g(\mathbf{x}), \quad \mathbf{x} \in \partial\mathcal{D}, \end{aligned} \tag{3.12}$$

as previous  $\mathbf{x} \in \mathcal{D} \subseteq \mathbb{R}^d$ ,  $t \in [0, T]$ ,  $f := f(\mathbf{x}) \in L^2(\mathcal{D}; \mathbb{R})$ ,  $u := u(\mathbf{x}, t)$  and  $u : \mathcal{D} \times [0, T] \rightarrow \mathbb{R}$ . We fix the initial conditions  $u(\mathbf{x}, 0) = u_0(\mathbf{x})$ , and set  $\xi(\mathbf{x}, t) \sim \mathcal{GP}(0, \delta(t - t') \cdot k_\theta(\mathbf{x}, \mathbf{x}'))$ .

As in the previous linear cases, the NL-statFEM starts with the finite element spatial discretisation to give a semidiscrete problem (a vector SDE), which is then discretised in time. Utilising the same basis functions as previous, we first multiply by test functions  $\phi_j \in V_h$ , recalling  $V_h = \text{span}\{\phi_i\}_{i=1}^{n_u}$ . In this case,  $\phi_i(\mathbf{x}) = g(\mathbf{x})$  for  $\mathbf{x} \in \partial\mathcal{D}$ . Integrating over the discrete problem domain  $\mathcal{D}_h$  and making the finite-dimensional approximation  $u_h(\mathbf{x}, t) = \sum_{i=1}^{n_u} u_{h,i}(t)\phi_i(\mathbf{x}) \in V_h \subset V$  yields the unsimplified weak form

$$\langle \partial_t u_h, \phi_j \rangle + \langle \mathcal{L}_\Lambda u_h, \phi_j \rangle + \langle F_\Lambda(u_h), \phi_j \rangle + \langle f, \phi_j \rangle + \langle \xi, \phi_j \rangle = 0, \quad j = 1, \dots, n_u.$$

This implicitly defines a nonlinear coupled system of stochastic differential equations over the finite element coefficients  $\mathbf{u}(t) = (u_{h,1}(t), \dots, u_{h,n_u}(t))^\top$ . Due to the nonlinear  $F_\Lambda(\cdot)$ , for fixed  $t$ , the resultant probability measure on function space is not of any known distributional form.

As in the linear case we can write this system as a nonlinear vector SDE

$$\mathbf{M} d\mathbf{u} + \mathbf{A}\mathbf{u} dt + \mathcal{F}(\mathbf{u}) dt + \mathbf{b} dt + d\boldsymbol{\beta}_t = 0, \quad (3.13)$$

where  $\mathbf{A} \in \mathbb{R}^{n_u \times n_u}$  encodes the operator  $L_\Lambda$  through  $\mathbf{A}_{ji} = \langle L_\Lambda \phi_i, \phi_j \rangle$ ,  $\mathcal{F} : \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_u}$  encodes the action of the nonlinear  $F_\Lambda$  through  $\mathcal{F}_j(\mathbf{u}) = \langle F_\Lambda(u_h), \phi_j \rangle$ , and  $\mathbf{b}_j = \langle f, \phi_j \rangle$ <sup>8</sup>.

Discretisation in time uses the same schemes introduced in the linear case. That is, the explicit Euler for  $\mathbf{u}_n := \mathbf{u}(n\Delta_t)$  is

$$\mathbf{M}(\mathbf{u}_n - \mathbf{u}_{n-1}) + \Delta_t \mathbf{A}\mathbf{u}_{n-1} + \Delta_t \mathcal{F}(\mathbf{u}_{n-1}) + \Delta_t \mathbf{b} + \mathbf{e}_{n-1} = 0, \quad (3.14)$$

in which  $\mathbf{e}_{n-1} = \boldsymbol{\beta}_n - \boldsymbol{\beta}_{n-1} \sim \mathcal{N}(0, \Delta_t \mathbf{G}_\theta)$  are i.i.d. Gaussian (timestep size  $\Delta_t$  is the same through the simulation). The implicit Euler is

$$\mathbf{M}(\mathbf{u}_n - \mathbf{u}_{n-1}) + \Delta_t \mathbf{A}\mathbf{u}_n + \Delta_t \mathcal{F}(\mathbf{u}_n) + \Delta_t \mathbf{b} + \mathbf{e}_{n-1} = 0, \quad (3.15)$$

and the Crank-Nicolson is

$$\mathbf{M}(\mathbf{u}_n - \mathbf{u}_{n-1}) + \Delta_t \mathbf{A}\mathbf{u}_{n-1/2} + \Delta_t \mathcal{F}(\mathbf{u}_{n-1/2}) + \Delta_t \mathbf{b} + \mathbf{e}_{n-1} = 0. \quad (3.16)$$

The implicit Euler and Crank-Nicolson methods both require the solution of a nonlinear system of equations at each timestep. This can be done through using Newton iterations, and, depending on the problem at hand, may be too computationally expensive to consider. As an alternative, Runge-Kutta methods can also be used, although these require some care to implement in order to ensure the covariance is properly propagated. General combinations of implicit and explicit discretisations may also be useful; these are known as the general IMEX family of discretisation (Ascher et al., 1995). For example, to avoid solving a nonlinear system of equations one possibility is

$$\mathbf{M}(\mathbf{u}_n - \mathbf{u}_{n-1}) + \Delta_t \mathbf{A}\mathbf{u}_n + \Delta_t \mathcal{F}(\mathbf{u}_{n-1}) + \Delta_t \mathbf{b} + \mathbf{e}_{n-1} = 0.$$

<sup>8</sup>Depending on the forms of  $L_\Lambda$  and  $F_\Lambda(\cdot)$  additional simplifications on these functionals may be possible; their derivation is thus problem-specific.

In this case nonlinear terms are handled with explicit Euler, and linear terms with implicit Euler. This can be useful in regimes where there is weak nonlinearity, and dominant linear behaviour (e.g. dominant diffusive behaviour) so long as the nonlinear component is not too numerically stiff (see, e.g., [Shampine and Gear, 1979](#); [Söderlind et al., 2015](#)). We have typically ensured stability through employing Crank-Nicolson; for the models considered in this thesis, the overhead of having to solve nonlinear systems of equations has been considered a reasonable tradeoff for unconditional stability.

Denote by  $\mathcal{M} : \mathbb{R}^{n_u} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_u}$  the deterministic PDE component of the stochastic dynamics. For example, for Crank-Nicolson this gives  $\mathcal{M}(\mathbf{u}_n, \mathbf{u}_{n-1}) = \mathbf{M}(\mathbf{u}_n - \mathbf{u}_{n-1}) + \Delta_t \mathbf{A} \mathbf{u}_{n-1/2} + \Delta_t \mathcal{F}(\mathbf{u}_{n-1/2}) + \Delta_t \mathbf{b}$ . The evolution can thus be written as an nonlinear additive Gaussian model

$$\mathcal{M}(\mathbf{u}_n, \mathbf{u}_{n-1}) + \mathbf{e}_{n-1} = 0, \quad (3.17)$$

implicitly defining a prior distribution over the FEM coefficients through marginalising over the previous timestep distribution  $p(\mathbf{u}_n | \boldsymbol{\theta}) = \int p(\mathbf{u}_n | \mathbf{u}_{n-1}, \boldsymbol{\theta}) p(d\mathbf{u}_{n-1} | \boldsymbol{\theta})$ . This prior is not available in closed form and two approximations are used, based on the extended Kalman filter (ExKF) and the ensemble Kalman filter (EnKF). These are named after their filtering extensions as they form the basis for the subsequent filtering methods (c.f. Section 3.3.1). The ExKF approach linearises about the current solution with the Jacobian of  $\mathcal{M}(\cdot, \cdot)$  to give a Gaussian approximation  $p(\mathbf{u}_n | \boldsymbol{\theta}) \approx \mathcal{N}(\mathbf{m}_n, \mathbf{C}_n)$ . The EnKF approach is standard Monte Carlo; we simulate  $\mathbf{e}_n$  at each timestep and build an ensemble of these simulated trajectories. Summary statistics can be computed from this ensemble.

The ExKF algorithm is now detailed. At the previous timestep  $t = (n-1)\Delta_t$  the associated measure on the FEM coefficients is  $p(\mathbf{u}_{n-1} | \boldsymbol{\theta}) = \mathcal{N}(\mathbf{m}_{n-1}, \mathbf{C}_{n-1})$ . Taking a Taylor series about the solution to  $\mathcal{M}(\mathbf{m}_n, \mathbf{m}_{n-1}) = 0$ ,  $\mathbf{m}_n$ , and writing the Jacobian as  $\mathbf{J}_n = \frac{\partial}{\partial \mathbf{u}_n} \mathcal{M}(\mathbf{u}_n, \mathbf{u}_{n-1})^9$  gives

$$\mathcal{M}(\mathbf{u}_n, \mathbf{u}_{n-1}) = \mathcal{M}(\mathbf{m}_n, \mathbf{m}_{n-1}) + \mathbf{J}_n(\mathbf{u}_n - \mathbf{m}_n) + \mathbf{J}_{n-1}(\mathbf{u}_{n-1} - \mathbf{m}_{n-1}) + \mathbf{r}(\mathbf{u}_n) + \mathbf{e}_{n-1} = 0,$$

where  $\mathbf{r}(\mathbf{u}_n)$  denotes the remainder terms of second order and above. Ignoring these remainder terms gives the prior at time  $n$

$$\begin{aligned} \mathbf{u}_n &\sim \mathcal{N}(\mathbf{m}_n, \mathbf{C}_n), \\ \mathbf{C}_n &= \mathbf{J}_n^{-1} \left( \mathbf{J}_{n-1} \mathbf{C}_{n-1} \mathbf{J}_{n-1}^\top + \mathbf{G}_\theta \right) \mathbf{J}_n^{-\top}, \end{aligned}$$

via the standard affine transformation of a Gaussian random variable (see, e.g., [Petersen and Pedersen, 2012](#)). This linear approximation extends the FEM via the propagation of induced uncertainty serially through the simulation, with the mean coinciding with the FEM solution. The covariance is given by the action of the Jacobians on the previous timestep covariance, which will induce some error between the true covariance and this approximation; systems with stronger nonlinearity will induce more error.

Alternatively, the EnKF (Monte Carlo) approach takes an ensemble  $\{\mathbf{u}_{n-1}^{[i]}\}_{i=1}^{N_{\text{ens}}}$  and at each timestep draws independent samples  $\mathbf{e}_{n-1}^{[i]} \sim \mathcal{N}(\mathbf{0}, \Delta_t \mathbf{G}_\theta)$  for  $i = 1, \dots, N_{\text{ens}}$ . For each sampled  $\mathbf{e}_{n-1}^{[i]}$ , the model

<sup>9</sup>This Jacobian can be computed via taking a Gateaux derivative of the weak form; see Section A.1 for a full discussion of this. In practice this can be done by hand or through automatic differentiation ([Baydin et al., 2018](#); [Farrell et al., 2013](#)).



**Algorithm 1** ExKF algorithm (with parameter estimation).

---

```

for  $n \leq n_t$  do
  (Prediction step)
  Solve  $\mathcal{M}(\mathbf{m}_{n|n-1}, \mathbf{m}_{n-1|n-1}) = \mathbf{0}$ .
   $\hat{\mathbf{C}}_{n|n-1} = \mathbf{J}_n^{-1} (\mathbf{J}_{n-1} \mathbf{C}_{n-1|n-1} \mathbf{J}_{n-1}^\top) \mathbf{J}_n^{-\top}$ .
  Estimate:  $(\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n) = \operatorname{argmax}_{\boldsymbol{\theta}_n, \sigma_n} \{\log p(\mathbf{y}_n | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}_n, \hat{\boldsymbol{\theta}}_{1:n-1}, \sigma_n, \hat{\sigma}_{1:n-1}) + \log p(\boldsymbol{\theta}_n) + \log p(\sigma_n)\}$ .
  Update  $\mathbf{C}_{n|n-1} = \hat{\mathbf{C}}_{n|n-1} + \Delta_t \mathbf{J}_n^{-1} \mathbf{G}_{\hat{\boldsymbol{\theta}}_n} \mathbf{J}_n^{-\top}$ .
  (Analysis step)
   $\mathbf{S}_n = \mathbf{H}_n \mathbf{C}_{n|n-1} \mathbf{H}_n^\top + \hat{\sigma}_n^2 \mathbf{I}$ .
   $\mathbf{m}_{n|n} = \mathbf{m}_{n|n-1} + \mathbf{C}_{n|n-1} \mathbf{H}_n^\top \mathbf{S}_n^{-1} (\mathbf{y}_n - \mathbf{H}_n \mathbf{m}_{n|n-1})$ .
   $\mathbf{C}_{n|n} = \mathbf{C}_{n|n-1} - \mathbf{C}_{n|n-1} \mathbf{H}_n^\top \mathbf{S}_n^{-1} \mathbf{H}_n \mathbf{C}_{n|n-1}$ .
end for

```

---

**Algorithm 2** EnKF algorithm (with parameter estimation).

---

```

for  $n \leq n_t$  do
  (Prediction step)
  for  $i \leq N_{\text{ens}}$  do
    Solve  $\mathcal{M}(\hat{\mathbf{u}}_{n,\text{pred}}^{[i]}, \mathbf{u}_{n-1}^{[i]}) = 0$ .
  end for
  Compute  $\hat{\mathbf{m}}_{n|n-1}$  and  $\hat{\mathbf{C}}_{n|n-1}$  from  $\{\hat{\mathbf{u}}_{n,\text{pred}}^{[i]}\}_{i=1}^{N_{\text{ens}}}$ .
  Estimate:  $(\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n) = \operatorname{argmax}_{\boldsymbol{\theta}_n, \sigma_n} \{\log p(\mathbf{y}_n | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}_n, \hat{\boldsymbol{\theta}}_{1:n-1}, \sigma_n, \hat{\sigma}_{1:n-1}) + \log p(\boldsymbol{\theta}_n) + \log p(\sigma_n)\}$ .
  for  $i \leq N_{\text{ens}}$  do
    Solve  $\mathcal{M}(\mathbf{u}_{n,\text{pred}}^{[i]}, \mathbf{u}_{n-1}^{[i]}) + \mathbf{e}_{n-1}^{[i]} = 0$ .
  end for
  Compute  $\mathbf{m}_{n|n-1}$  and  $\mathbf{C}_{n|n-1}$  from  $\{\mathbf{u}_{n,\text{pred}}^{[i]}\}_{i=1}^{N_{\text{ens}}}$ .
  Let  $\mathbf{S}_n = \mathbf{H}_n \mathbf{C}_{n|n-1} \mathbf{H}_n^\top + \hat{\sigma}_n^2 \mathbf{I}$ .
  (Analysis step)
  for  $i \leq N_{\text{ens}}$  do  $\mathbf{u}_n^{[i]} = \mathbf{u}_{n,\text{pred}}^{[i]} + \mathbf{C}_{n|n-1} \mathbf{H}_n^\top \mathbf{S}_n^{-1} (\mathbf{y}_n + \boldsymbol{\eta}_n^{[i]} - \mathbf{H}_n \mathbf{u}_{n,\text{pred}}^{[i]})$ 
  end for
end for

```

---

is solved for  $\mathbf{u}_n^{[i]}$

$$\mathcal{M}(\mathbf{u}_n^{[i]}, \mathbf{u}_{n-1}^{[i]}) + \mathbf{e}_{n-1}^{[i]} = 0, \quad i = 1, \dots, N_{\text{ens}}.$$

The associated approximate measure  $p(\mathbf{u}_n | \boldsymbol{\theta}) = \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} \delta_{\mathbf{u}_n^{[i]}}(d\mathbf{u}_n)$  can be empirically summarised by its first and second centred moments

$$\mathbb{E}[\mathbf{u}_n] = \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} \mathbf{u}_n^{[i]},$$

$$\operatorname{cov}(\mathbf{u}_n) = \frac{1}{N_{\text{ens}} - 1} \sum_{i=1}^{N_{\text{ens}}} (\mathbf{u}_n^{[i]} - \mathbb{E}(\mathbf{u}_n)) (\mathbf{u}_n^{[i]} - \mathbb{E}(\mathbf{u}_n))^\top.$$

To ensure that  $\operatorname{cov}(\mathbf{u}_n)$  is not rank-deficient we require  $N_{\text{ens}} \geq n_u + 1$ . The EnKF trajectories are those given by the Euler-Maruyama simulation of the SDE (Kloeden and Platen, 1992) using the three discretisations given above, in Equations (3.14), (3.15), and (3.16).



### 3.3.1 The nonlinear, time-dependent posterior

These schemes form the basis for computing the filtering posterior  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \boldsymbol{\theta}, \sigma)$ , describing the belief in the FEM solution conditioned on data observed up to and including time  $n\Delta_t$ . The data generating process is the same as for the linear case,  $\mathbf{y}_n = \mathbf{H}_n \mathbf{u}_n + \boldsymbol{\eta}_n$ . Data  $\mathbf{y}_n \in \mathbb{R}^{n_y}$  is observed via  $\mathbf{H}_n : \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_y}$ , the observation operator that maps from the FEM solution mesh to the observed points  $\mathbf{x}_{\text{obs}}$ . Observation noise is Gaussian,  $\boldsymbol{\eta}_n \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ , and thus so too is the likelihood  $p(\mathbf{y}_n | \mathbf{u}_n, \sigma) = \mathcal{N}(\mathbf{H}_n \mathbf{u}_n, \sigma^2 \mathbf{I})$ .

The following procedure we describe is a generic nonlinear filtering algorithm which makes the Gaussian assumption after the prediction step; both the ExKF and EnKF approaches are described fully in Algorithms 1 and 2. For the initial presentation, we assume that hyperparameters,  $\boldsymbol{\theta}$ , and noise standard deviation,  $\sigma$ , are known. As in the linear case these parameters may also be estimated, which is discussed after the algorithm presentation.

At time  $n$  assume that the measure on the previous time is

$$p(\mathbf{u}_{n-1} | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}, \sigma) = \mathcal{N}(\mathbf{m}_{n-1|n-1}, \mathbf{C}_{n-1|n-1}).$$

Proceed as follows for  $n = 1, \dots, n_t$ :

1. Compute the prediction distribution:

$$\begin{aligned} p(\mathbf{u}_n | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}, \sigma) &= \int p(\mathbf{u}_n | \mathbf{u}_{n-1}, \boldsymbol{\theta}) p(d\mathbf{u}_{n-1} | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}, \sigma) \\ &\approx \mathcal{N}(\mathbf{m}_{n|n-1}, \mathbf{C}_{n|n-1}). \end{aligned}$$

2. Complete the analysis step:

$$\begin{aligned} p(\mathbf{u}_n | \mathbf{y}_{1:n}, \boldsymbol{\theta}, \sigma) &\propto p(\mathbf{y}_n | \mathbf{u}_n, \sigma) p(\mathbf{u}_n | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}, \sigma) \\ &= \mathcal{N}(\mathbf{m}_{n|n}, \mathbf{C}_{n|n}). \end{aligned}$$

The prior  $p(\mathbf{u}_n | \boldsymbol{\theta})$  is recovered if only the prediction step (step 1), is completed at each iteration. In this case the algorithm reduces to those given in the subsection above.

As in the linear case the hyperparameters may be set to the MAP estimates (Murphy, 2012). In this case the exact marginal likelihood has no tractable form and the Gaussian approximation is constructed using the ExKF linearisation:

$$p(\mathbf{y}_n | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}, \sigma) = \mathcal{N}(\mathbf{H}_n \hat{\mathbf{m}}_{n|n-1}, \mathbf{H}_n \hat{\mathbf{C}}_{n|n-1} \mathbf{H}_n^\top + \Delta_t \mathbf{H}_n \hat{\mathbf{G}}_n(\boldsymbol{\theta}) \mathbf{H}_n^\top + \sigma_n^2 \mathbf{I}_{n_y}),$$

in which  $\hat{\mathbf{G}}_n(\boldsymbol{\theta}) = \mathbf{J}_n^{-1} \mathbf{G}_\theta \mathbf{J}_n^{-\top}$ ; for definitions of  $\hat{\mathbf{m}}_{n|n-1}$  and  $\hat{\mathbf{C}}_{n|n-1}$  we refer to Algorithms 1 and 2. Using this the complete-data likelihood can be computed (up to a normalising constant)

$$p(\mathbf{y}_{1:n_t} | \boldsymbol{\theta}, \sigma) = -\frac{1}{2} \sum_{n=1}^{n_t} \log |\boldsymbol{\Sigma}_n(\boldsymbol{\theta}, \sigma)| - \frac{1}{2} \sum_{n=1}^{n_t} (\mathbf{y}_n - \mathbf{H}_n \hat{\mathbf{m}}_{n|n-1})^\top \boldsymbol{\Sigma}_n(\boldsymbol{\theta}, \sigma)^{-1} (\mathbf{y}_n - \mathbf{H}_n \hat{\mathbf{m}}_{n|n-1}),$$

where  $\boldsymbol{\Sigma}_n(\boldsymbol{\theta}, \sigma) = \mathbf{H}_n \hat{\mathbf{C}}_{n|n-1} \mathbf{H}_n^\top + \Delta_t \mathbf{H}_n \hat{\mathbf{G}}_n(\boldsymbol{\theta}) \mathbf{H}_n^\top + \sigma^2 \mathbf{I}_{n_y}$ . This, again, requires running the filter up

to timestep  $n_t$  to compute it. If no temporal evolution is thought to occur on the hyperparameters, then they can be estimated from this complete-data likelihood.

If the parameters are assumed to be time-evolving, then we write  $(\boldsymbol{\theta}, \sigma) := (\boldsymbol{\theta}_n, \sigma_n)$ , with  $(\boldsymbol{\theta}_n, \sigma_n)$  independent across time. As in the linear case, to estimate  $(\boldsymbol{\theta}_n, \sigma_n)$  we modify the algorithms by splitting the prediction step in two, adding an optimisation step in between. If the measure on the previous timestep is given by  $p(\mathbf{u}_{n-1} | \mathbf{y}_{1:n-1}, \hat{\boldsymbol{\theta}}_{1:n-1}, \hat{\sigma}_{1:n-1}) = \mathcal{N}(\mathbf{m}_{n-1}, \mathbf{C}_{n-1})$ , then the tentative prediction step, which predicts the mean and covariance using  $\hat{\boldsymbol{\theta}}_{n-1}$ , is given by

$$\begin{aligned} p(\mathbf{u}_n | \mathbf{y}_{1:n-1}, \hat{\boldsymbol{\theta}}_{1:n-1}, \hat{\sigma}_{1:n-1}) &= \int p(\mathbf{u}_n | \mathbf{u}_{n-1}) p(d\mathbf{u}_{n-1} | \mathbf{y}_{1:n-1}, \hat{\boldsymbol{\theta}}_{1:n-1}, \hat{\sigma}_{1:n-1}) \\ &\approx \mathcal{N}(\hat{\mathbf{m}}_{n|n-1}, \hat{\mathbf{C}}_{n|n-1}), \end{aligned}$$

where  $\hat{\mathbf{m}}_{n|n-1}$  and  $\hat{\mathbf{C}}_{n|n-1}$  are defined in Algorithms 1 and 2 for the ExFF and EnKF methods, respectively. The approximate marginal likelihood is

$$p(\mathbf{y}_n | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}_n, \hat{\boldsymbol{\theta}}_{1:n-1}, \sigma_n, \hat{\sigma}_{1:n-1}) = \mathcal{N}(\mathbf{H}_n \hat{\mathbf{m}}_{n|n-1}, \mathbf{H}_n \hat{\mathbf{C}}_{n|n-1} \mathbf{H}_n^\top + \Delta_t \mathbf{H}_n \hat{\mathbf{G}}(\boldsymbol{\theta}_n) \mathbf{H}_n^\top + \sigma_n^2 \mathbf{I}_{n_y})$$

and the hyperparameters  $(\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n)$  are then estimated, using the log-posterior:

$$(\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n) = \underset{(\boldsymbol{\theta}_n, \sigma_n)}{\operatorname{argmax}} \{ \log p(\mathbf{y}_n | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}_n, \hat{\boldsymbol{\theta}}_{1:n-1}, \sigma_n, \hat{\sigma}_{1:n-1}) + \log p(\boldsymbol{\theta}_n) + \log p(\sigma_n) \}.$$

After obtaining the MAP estimates, the full prediction step can be done,

$$\begin{aligned} p(\mathbf{u}_n | \mathbf{y}_{1:n-1}, \hat{\boldsymbol{\theta}}_{1:n}, \hat{\sigma}_{1:n-1}) &= \int p(\mathbf{u}_n | \mathbf{u}_{n-1}, \hat{\boldsymbol{\theta}}_n) p(d\mathbf{u}_{n-1} | \mathbf{y}_{1:n-1}, \hat{\boldsymbol{\theta}}_{1:n-1}, \hat{\sigma}_{1:n-1}) \\ &\approx \mathcal{N}(\mathbf{m}_{n|n-1}, \mathbf{C}_{n|n-1}), \end{aligned}$$

and the update step remains the same, to give  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \hat{\boldsymbol{\theta}}_{1:n}, \hat{\sigma}_{1:n})$ . As in the linear case optimisation uses L-BFGS-B (Nocedal and Wright, 2006), with starting points set to the previous estimates  $(\hat{\boldsymbol{\theta}}_{n-1}, \hat{\sigma}_{n-1})$ . Unless otherwise specified, weakly informative truncated Gaussian priors are used:  $\rho_n \sim \mathcal{N}_+(1, 1^2)$ ,  $\ell_n \sim \mathcal{N}_+(1, 1^2)$ , and  $\sigma_n \sim \mathcal{N}_+(0, 1^2)$ .

## Discussion

In what follows, operation counts for the presented algorithms are discussed and some notes on possible points of interest on NL-statFEM are included. These pertain to potential issues that may arise for practitioners. When giving the operation counts, the cost associated with completing a single model solve assumed to be  $\mathcal{O}(M)$  operations, and the Jacobian is assumed to be a sparse matrix with bandwidth  $b$ . Note also that the hyperparameters are fixed and we assume that  $n_y < n_u$ , as is the case when the number of sensor locations, for example, is small in comparison to the finite element mesh.

**ExKF:** The total number of operations for the ExKF is similar to the number of operations required for the standard Kalman filter, with the mass matrix solves of the linear case being replaced with Jacobian solves. The prediction step for the mean has the cost of  $\mathcal{O}(M)$ , and for the covariance this is

$\mathcal{O}(b^2 n_u) + \mathcal{O}(bn_u^2)$  (for factorisation and forward- and back-substitution, respectively). The updating step requires  $\mathcal{O}(n_y n_u)$  operations to form  $\mathbf{S}_n = \mathbf{H}_n \mathbf{C}_{n|n-1} \mathbf{H}_n^\top + \sigma^2 \mathbf{I}_{n_y}$ , and  $\mathcal{O}(n_y^3/3)$  operations to compute  $\mathbf{S}_n = \mathbf{L}_n \mathbf{L}_n^\top$ . Using this factorisation, computing the mean update is  $\mathcal{O}(n_y^2) + \mathcal{O}(n_u n_y)$ , and the covariance update is  $\mathcal{O}(n_u n_y^2) + \mathcal{O}(n_u^2 n_y)$ .

**EnKF:** The covariance matrix is represented through the mean perturbations (Mandel, 2006)

$$\mathbf{C}_n = \mathbf{E}_n \mathbf{E}_n^\top, \quad \mathbf{E}_n = \frac{1}{\sqrt{N_{\text{ens}} - 1}} \left[ \mathbf{u}_n^{[1]} - \mathbb{E}[\mathbf{u}_n], \dots, \mathbf{u}_n^{[N_{\text{ens}}]} - \mathbb{E}[\mathbf{u}_n] \right] \in \mathbb{R}^{n_u \times N_{\text{ens}}},$$

which avoids having to store and manipulate the dense  $\mathbf{C}_n$ . The prediction step is  $\mathcal{O}(N_{\text{ens}} M)$ , however this can be parallelised in order to save compute time, as each prediction step is conditionally independent. The computation of  $\mathbf{H} \mathbf{E}_n$  and  $\mathbf{H} \mathbf{U}_n$ , where  $\mathbf{U}_n = [\mathbf{u}_n^{[1]}, \dots, \mathbf{u}_n^{[N_{\text{ens}}]}]$ , are each  $\mathcal{O}(N_{\text{ens}} n_y)$ . Then  $\mathbf{S}_n = \mathbf{H} \mathbf{E}_n (\mathbf{H} \mathbf{E}_n)^\top + \sigma_n^2 \mathbf{I}_{n_y}$  is computed, which is  $\mathcal{O}(n_y^2 N_{\text{ens}})$ . Again this requires  $\mathcal{O}(n_y^3/3)$  operations to factorise  $\mathbf{S}_n = \mathbf{L}_n \mathbf{L}_n^\top$ . Having the Cholesky factor, solving  $\tilde{\mathbf{y}}_n^{[i]} = \mathbf{S}_n^{-1} (\mathbf{y}_n - \mathbf{H} \mathbf{u}_n + \boldsymbol{\eta}_n^{[i]})$  is  $\mathcal{O}(n_y^2)$  and computing  $\mathbf{u}_n^{[i]} + \mathbf{E}_n (\mathbf{H} \mathbf{E}_n)^\top \tilde{\mathbf{y}}_n^{[i]}$  is  $\mathcal{O}(N_{\text{ens}} n_u) + \mathcal{O}(N_{\text{ens}} n_y)$ . Completing this for the full ensemble gives the total complexity  $\mathcal{O}(N_{\text{ens}} n_y^2) + \mathcal{O}(N_{\text{ens}}^2 n_u) + \mathcal{O}(N_{\text{ens}}^2 n_y)$ .

Note that  $\mathbf{R}_n$  forms a low-rank square-root of the covariance matrix, estimated from the ensemble. As we may have  $N_{\text{ens}} < n_u$ , spurious correlations can arise between model variables, which may be physically insensible (for example, correlations between variables which are substantially distanced from one another). This can be offset using covariance localisation (Houtekamer and Mitchell, 2001).

For a linear model, the ExKF approach of Algorithm 1 gives the standard Kalman filter of the previous section (as  $\partial_{\mathbf{u}} \mathbf{A} \mathbf{u} = \mathbf{A}$ ). The discretised linear PDE operator is replaced with the tangent linear PDE operator, to give similar updating equations. The ExKF approach may be preferred when working with code that can automatically compute the requisite Jacobian matrix without manual user intervention. When working with models for which the Jacobian matrix cannot be computed, then the EnKF may be preferred. Also, in scenarios in which the posterior mean  $\mathbf{m}_{n|n}$  is the main quantity of interest, then the EnKF may be more efficient due to being able to use small ensemble sizes  $N_{\text{ens}} \ll n_u$  (Houtekamer and Mitchell, 1998). This trades off efficient estimation of  $\mathbf{m}_{n|n}$  for possibly inaccurate UQ (Law and Stuart, 2012).

Finally, we note the following bibliographic details. As it stands, our work is the only methodological extension of the initial statFEM of Girolami et al. (2021). However, there has been additional theoretical work for the static case, giving various guarantees on the derived measures (see Papandreou et al. (2021) and Karvonen et al. (2022), which was briefly mentioned in Section 3.1). Additionally, an application of statFEM to digital railway monitoring is given in Febrianto et al. (2021), and the combination of statFEM with Langevin samplers is given in Akyildiz et al. (2022). As noted throughout the previous chapters, our scalable extension to NL-statFEM is given in Duffin et al. (2022), where we detail applications to reaction-diffusion systems. This is the focus of Chapter 5.

## Chapter 4

# Applying NL-statFEM to address misspecification

*In this chapter we investigate the application of our NL-statFEM to a variety of 1D nonlinear, time-dependent systems, studying the data assimilation procedure for Burgers equation, the Kuramoto-Sivashinsky (KS) equation, and the Korteweg-de Vries (KdV) equation. Synthetic and experimental data is used. We gradually increase the severity of model misspecification, proceeding through:*

1. *Unknown stochastic forcing (Burgers);*
2. *Misspecified linear viscosity in a weakly nonlinear system (Burgers);*
3. *Misspecified linear viscosity in a highly nonlinear system (KS);*
4. *Misspecified nonlinear advection in a weakly nonlinear system (KdV);*
5. *Experimental data assimilation in a weakly nonlinear system (KdV).*

*In this chapter simulations are run with the ExKF only; results using the EnKF are shown in Appendix A.2.*

*Results, first presented in [Duffin et al. \(2021\)](#), show that conditioning on data improves the model accuracy. This is verified through computing the relative error on the NL-statFEM posterior mean, the root mean square error, the Kalman gain norm, and a signal-to-noise ratio. The posterior mean is able to accurately recover the data generating process, and, where appropriate, hyperparameter estimates appear to coincide with the true values. Code to replicate all results is available at <https://github.com/connor-duffin/statkdv-paper>.*

### 4.1 Burgers equation

Burgers equation is a nonlinear PDE that describes the balance of nonlinear steepening and viscous damping. It is named after [Burgers \(1948\)](#), who initially studied the equation in the context of turbulence modelling. The Burgers equation can be derived through making simplifications inside of the incompressible Navier-Stokes equations, which, ignoring boundary and initial conditions, are

given by

$$\mathbf{u}_t + \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla \mathbf{p} + \nu \nabla^2 \mathbf{u} + \mathbf{f}, \quad (4.1)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (4.2)$$

$$\mathbf{u} := \mathbf{u}(\mathbf{x}, t), \quad \mathbf{x} \in \mathcal{D} \subset \mathbb{R}^3, \quad t \in \mathbb{R}_+,$$

for  $\mathbf{u}(x_1, x_2, x_3, t) := (u(x_1, x_2, x_3, t), v(x_1, x_2, x_3, t), w(x_1, x_2, x_3, t))$ ,  $p := p(x_1, x_2, x_3, t)$ . The scalar coefficient  $\nu$  is the *kinematic viscosity* which parameterises the strength of viscous effects. The velocity field inside of  $\mathcal{D}$  is given by  $\mathbf{u}$ , and the pressure field is  $p$ . These equations can be derived from the conservation of momentum (giving (4.1)) and the conservation of mass (giving (4.2)), in combination with the assumption of incompressibility and the fluid being Newtonian.

Writing out the equations in terms of the scalar velocity components gives

$$\begin{aligned} u_t + uu_x + vv_y + ww_z &= -p_x + \nu(u_{xx} + u_{yy} + u_{zz}), \\ v_t + uv_x + vv_y + ww_z &= -p_y + \nu(v_{xx} + v_{yy} + v_{zz}), \\ w_t + uw_x + vw_y + ww_z &= -p_z + \nu(w_{xx} + w_{yy} + w_{zz}) + g, \\ u_x + u_y + u_z &= 0, \end{aligned}$$

where  $g = 9.81\text{ms}^{-2}$  is the acceleration due to gravity. Ignoring the pressure terms, gravitational acceleration, and the incompressibility condition gives the 1D system

$$\begin{cases} u_t + uu_x = \nu u_{xx}, \\ u := u(x, t), \quad x \in \mathbb{R}, \quad t \in \mathbb{R}_+, \end{cases} \quad (4.3)$$

describing the balance of nonlinear advection  $uu_x$  and damping  $\nu u_{xx}$ . This is the *Burgers equation*.

Through this simplification, the chaotic effects seen in highly turbulent regimes of the Navier-Stokes equations (i.e. high Reynolds number flow) are not present. This enables the study of nonlinearity and viscosity in a simplified context, where analytical solutions are available through the *Hopf-Cole transformation* (Cole, 1951; Hopf, 1950). The Hopf-Cole transformation makes the substitution  $u = -2\nu \hat{u}_x / \hat{u}$ , which then gives the heat equation for  $\hat{u}$  over  $(x, t) \in \mathbb{R} \times \mathbb{R}_+$

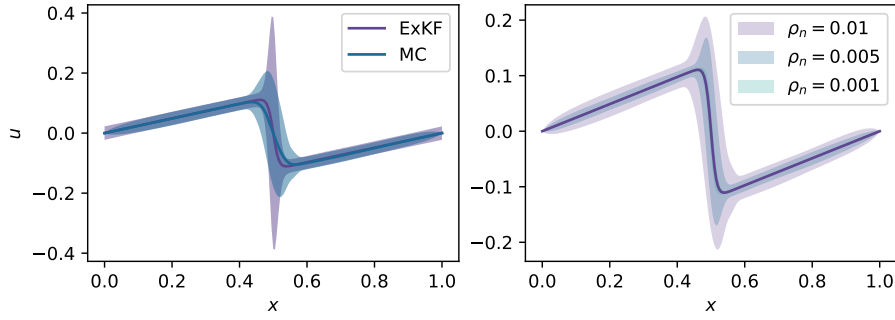
$$\hat{u}_t = \nu \hat{u}_{xx}, \quad \hat{u} := \hat{u}(x, t), \quad u(x, 0) = u_0(x).$$

Solving via the Fourier transform gives (Debnath, 2012)

$$u(x, t) := -2\nu \partial_x \log \left( (4\pi\nu t)^{-1/2} \int \exp \left[ -\frac{(x-x')^2}{4\nu t} - \frac{1}{2\nu} \int_0^{x'} u_0(x'') dx'' \right] dx' \right). \quad (4.4)$$

Strong solutions to the linear heat equation can be mapped via (4.4) to nonlinear Burgers equation solutions. This motivates that there is some reduction in the degree of nonlinearity present, through using a Burgers model in lieu of Navier-Stokes.

For  $\nu > 0$  this is a dissipative system due to the viscosity term. The *inviscid* Burgers equation



**Figure 4.1:** Burgers prior results. Left: ExKF estimated prior vs. MC estimated prior, shown with estimated means and 95% credible intervals. Right: MC estimate of the prior under uncertainty refinement, shown with the deterministic FEM solution and the respective 95% credible intervals.

is obtained if  $\nu = 0$ . This is a scalar conservation law, having the conservative form  $u_t + (u^2)_x/2 = 0$ . Solutions to the inviscid equation develop shocks, and can be studied using the method of characteristics (Strauss, 2007). It is a common model for traffic flow; in this scenario,  $u$  is the *traffic density*, being, for example, cars per unit length.

In this subsection we study the viscous Burgers' equation on a finite spatiotemporal domain, given by

$$\begin{cases} u_t + uu_x - \nu u_{xx} = 0, \\ u := u(x, t), \quad x \in [0, 1], \quad t \in [0, T], \\ u(0, t) = u(1, t) = 0, \\ u(x, 0) = \sin(2\pi x). \end{cases} \quad (4.5)$$

In this we take the interpretation that  $u(x, t)$  is the velocity of some fluid flow, so that  $u$  has the dimensions (length)/(time). The kinematic viscosity coefficient  $\nu$  has the dimension (length)<sup>2</sup>/(time), and parameterises the strength of viscous effects.

A statistical augmentation of (4.5) is

$$\begin{aligned} u_t + uu_x - \nu u_{xx} &= \xi, \\ \xi &:= \xi(x, t) \sim \mathcal{GP}(0, \delta(t - t') \cdot k_\theta(x, x')), \end{aligned} \quad (4.6)$$

and retaining the boundary and initial conditions of (4.5), this is the system we consider in the rest of this section. We use the squared-exponential structure of Equation (3.2), with  $k_\theta(x, x') = \rho^2 \exp(-\|x - x'\|^2/(2\ell^2))$ , so that  $\theta = (\rho, \ell)$ .

The remainder of this section investigates the application of NL-statFEM to compute the posterior distribution  $p(\mathbf{u}_n \mid \mathbf{y}_{1:n}, \hat{\boldsymbol{\theta}}_{1:n}, \hat{\sigma}_{1:n})$ , for two cases with synthetic data. The first is a sanity check on the parameter estimation process, with data simulated from (4.6). Hyperparameters are recovered and the NL-statFEM posterior mean provides a sensible estimate of the data generating process (DGP). The second considers misspecification coming from the kinematic viscosity  $\nu$ , with the NL-statFEM dynamic model being underdamped in comparison to the DGP. In this case data is simulated from the deterministic Equation (4.5), with synthetic observation noise added. In this example we introduce various verification quantities that are used to check filter performance.

### 4.1.1 Discretisation

First the domain  $\mathcal{D} = [0, 1]$  is discretised into the mesh with nodes  $x_i = ih$ , for  $i = 0, \dots, n_c$  such  $n_c h = 1$ . We refer to  $n_c$  as the number of finite element cells in the mesh, with the finite dimensional set of basis functions  $V_h^1(\mathcal{D}) = \text{span}\{\phi_i(x)\}_{i=1}^{n_u}$  being defined such that  $\phi_i(x_j) = \delta_{ij}$ . These basis functions are piecewise linear polynomials. Next, writing  $u_h(x, t) = \sum_{i=1}^{n_u} u_i(t)\phi_i(x)$ , and substitute into the weak form

$$\langle \partial_t u_h, \phi_j \rangle + \langle u_h \partial_x u_h, \phi_j \rangle + \nu \langle \partial_x u_h, \partial_x \phi_j \rangle = \langle \xi, \phi_j \rangle, \quad j = 1, \dots, n_u.$$

For time discretisation, we use implicit Euler to give the updating equation, where  $u_h^n = u_h(x, n\Delta_t)$ ,

$$\langle u_h^n - u_h^{n-1}, \phi_j \rangle + \Delta_t \langle u_h^n \partial_x u_h^n, \phi_j \rangle + \nu \Delta_t \langle \partial_x u_h^n, \partial_x \phi_j \rangle = \langle \Delta \xi, \phi_j \rangle, \quad j = 1, \dots, n_u,$$

where  $\Delta \xi \sim \mathcal{GP}(0, \Delta_t k_\theta(x, x'))$ . This defines the updating rule for the finite element coefficients  $\mathbf{u}_n = (u_1(n\Delta_t), \dots, u_{n_u}(n\Delta_t))^\top$ . This discretisation is implemented in Fenics (Logg et al., 2012), and for the numerics we use the default Fenics Newton solver, which solves the linearised system using the LU decomposition. Convergence is typically reached in  $< 5$  Newton iterations.

### 4.1.2 Example I: estimating hyperparameters

In this example, we set  $\nu = 0.01$ , and we numerically integrate the Burgers equation up to  $t = 4$ , using, for the finite element discretisation,  $n_c = 200$  cells, and a timestep of  $\Delta_t = 0.02$ .

First, before showing the posterior, in Figure 4.1 we plot the estimated prior  $p(\mathbf{u}_n | \boldsymbol{\theta})$  at the final simulation time  $t = 4$ . Prior estimates obtained using the ExKF method and the EnKF<sup>1</sup> method are compared in Figure 4.1 (left). In this case we set  $(\rho_n, \ell_n) = (0.01, 0.1)$ . The ExKF prior mean coincides with the deterministic FEM solution, and the EnKF mean appears slightly dampened in comparison. The ExKF 95% credible intervals are sharply peaked about the steepened central feature, due to the large gradients seen at this point. The EnKF 95% credible intervals also inflate about this region of high gradient but are more spatially diffuse. In this instance, repeated application of the tangent linear covariance approximation results in inaccurate UQ. The EnKF approximation is again shown in Figure 4.1 (right), for three levels of prior uncertainty,  $\rho_n \in \{10^{-2}, 5 \times 10^{-3}, 10^{-3}\}$ . As  $\rho_n$  decreases, the uncertainty shrinks about the known FEM solution. This hyperparameter acts as a measure of the *a priori* accuracy of the model predictions.

For the posterior, data are simulated using (4.6) as the DGP, with  $(\rho_n, \ell_n) = (0.05, 0.1)$ , for all  $n$ . Synthetic data  $\mathbf{y}_n$  are given by subsampling the resultant finite element coefficients on the subgrid  $\mathbf{x}_{\text{obs}}$ , with  $n_y = 101$ . Synthetic noise is also added, which has  $\sigma_n = 0.01$ . As in the previous section, the DGP is

$$\mathbf{y}_n = \mathbf{H}\mathbf{u}_n + \boldsymbol{\eta}_n, \quad \boldsymbol{\eta}_n \sim \mathcal{N}(0, \sigma_n \mathbf{I}_{n_y}).$$

We compute the posterior  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \hat{\boldsymbol{\theta}}_{1:n}, \hat{\sigma}_{1:n})$ , using the ExKF, fixing  $\ell_n = 0.1$  and estimating  $(\rho_n, \sigma_n)$ , using the MAP estimation method, assuming that  $\rho_n$  and  $\sigma_n$  are respectively i.i.d. for all times. We use the weakly informative truncated Gaussian prior distributions  $p(\rho_n) \sim \mathcal{N}_+(1, 1^2)$ ,  $p(\sigma_n) \sim \mathcal{N}_+(0, 1^2)$ ,

<sup>1</sup>Recall that the EnKF for the prior is the Euler-Maruyama discretisation (Kloeden and Platen, 1992), which gives a Monte Carlo approximation.



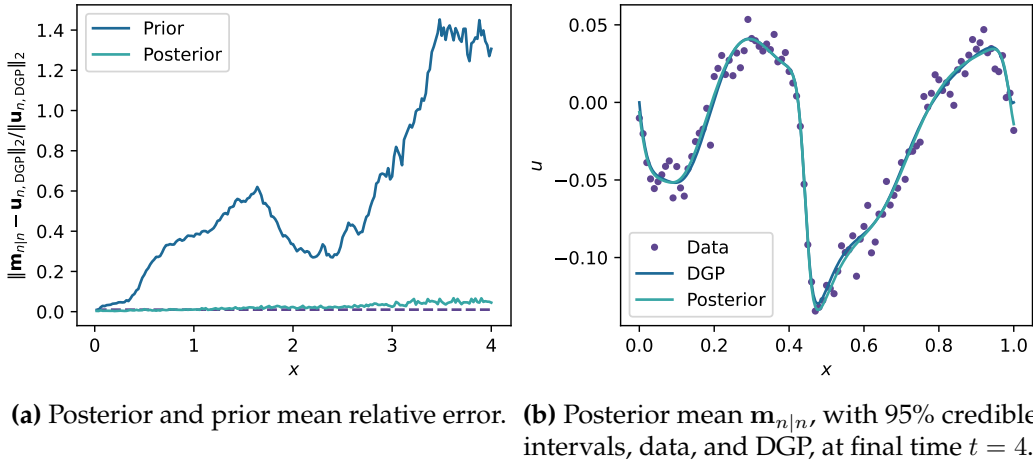


Figure 4.2: Burgers example I results.

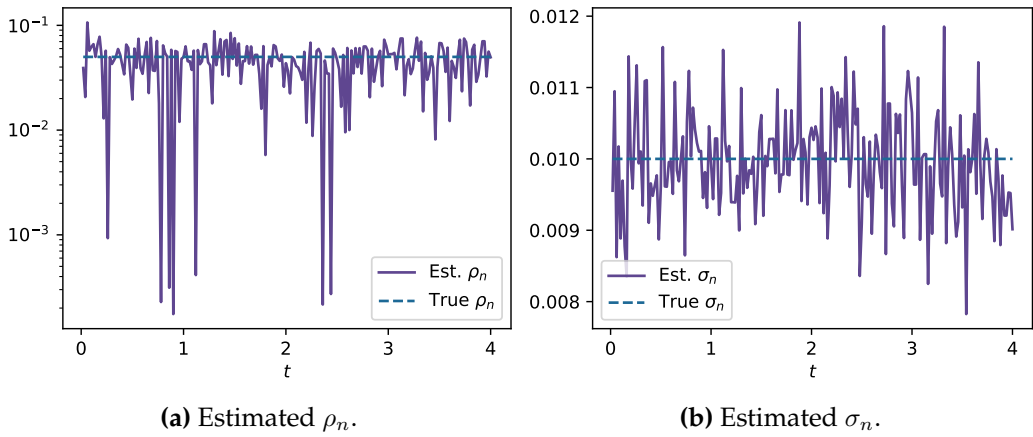


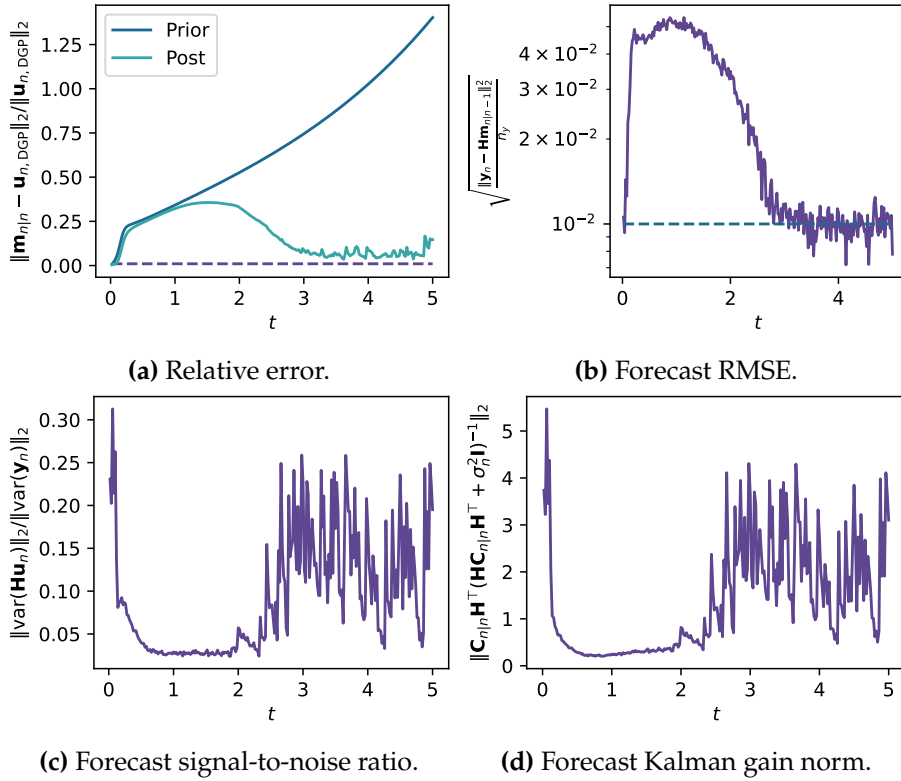
Figure 4.3: Burgers example I results: hyperparameter estimation.

reflecting that *a priori* we do not have information available on these parameters.

Figure 4.2 shows the ExKF results. As a first measure of filter performance we introduce the relative error in the  $l^2$  norm,  $\|\cdot\|_2$ . For a NL-statFEM model with mean  $\mathbf{m}_n$  this is  $\|\mathbf{m}_n - \mathbf{u}_{n,\text{DGP}}\|_2 / \|\mathbf{u}_{n,\text{DGP}}\|_2$ , providing a verification of the accuracy of the posterior mean. Figure 4.2a shows that conditioning on data improves the relative error. Without conditioning on data errors accumulate over time, which is seen in the prior. This is perhaps unsurprising given that the accumulation of stochastic perturbations is what results in model mismatch. For the posterior mean, the relative error seems to increase slightly as the time increases, hovering at  $\approx 1\%$ . A visual check of this accuracy is provided in Figure 4.2b, which plots the posterior mean  $\mathbf{m}_{n|n}$  and 95% credible intervals with the data and DGP. The mean visually provides a reasonable approximation to the underlying DGP and appears to effectively “denoise” the data.

Parameter estimates  $(\hat{\rho}_n, \hat{\sigma}_n)$  are shown in Figure 4.3 and appear accurate. The  $\mathcal{GP}$  variance hyperparameter is occasionally underestimated (Figure 4.3a), implying that at these particular time points the observed prediction mean  $\mathbf{H}\mathbf{m}_{n|n-1}$  is close to the data  $\mathbf{y}_n$ . Noise estimates (Figure 4.3b) are stable and show that in this case NL-statFEM is able to separate signal from noise.





**Figure 4.4:** Burgers example II results: verification quantities.

### 4.1.3 Example II: mismatched viscosity

In this example data  $y_n$  are generated by taking the deterministic (4.5) as the DGP, with damping set to  $\nu = 0.01$ . All discretisations are the same as in the previous, and we integrate up to  $t = 5$ . Data is subsampled on a grid with  $n_y = 52$ , to give the observed values  $y_n$ . Simulated observational error is also added on, with  $\sigma_n = 0.01$ . The underlying NL-statFEM model induces model mismatch through setting  $\nu = 0.001$  in (4.6), so the model is misspecified due to incorrectly specified damping. We use the ExKF to compute the posterior  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \hat{\boldsymbol{\theta}}_{1:n}, \hat{\sigma}_{1:n})$ , estimating  $(\rho_n, \sigma_n)$  at each time using MAP estimation. Again we fix  $\ell = 0.1$  and retain the priors of the previous subsection.

Verification quantities are shown in Figure 4.4. In terms of the relative error (Figure 4.4a), the posterior error increases initially but is always enveloped by the prior error. After  $t \approx 1.5$ , the error begins to decrease and by  $t \approx 3$  the error reaches a stable value of  $\approx 0.1$ ; incorporating data corrects for coefficient misspecification. In terms of the forecast root mean square error (RMSE, Figure 4.4b)

$$\text{RMSE}_n = \sqrt{\frac{\|\mathbf{y}_n - \mathbf{H}\mathbf{m}_{n-1}\|_2^2}{n_y}},$$

the filter reaches the known value of the standard deviation ( $\sigma_n = 0.01$ ) as the filter enters the stable filtering configuration at  $t \approx 3.0$ . These results concur with parameter estimates (Figure 4.5), which also stabilise at  $t \approx 3.0$ .

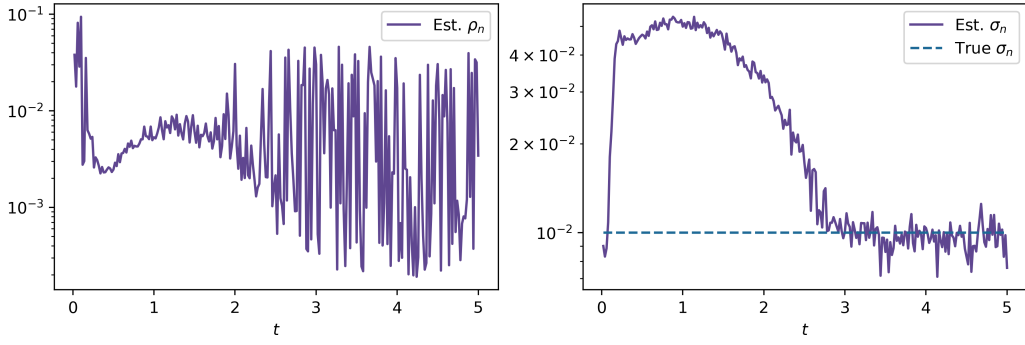


Figure 4.5: Burgers example II: estimated hyperparameters.

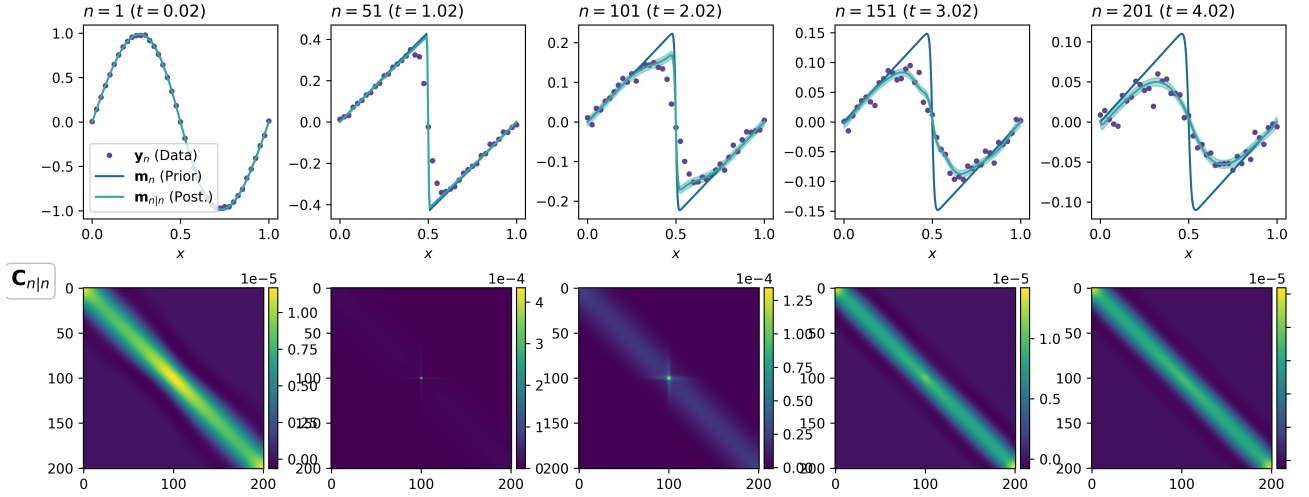


Figure 4.6: Burgers example II results. Top row: posterior means, 95% credible intervals, data. Bottom row: covariance matrices  $\mathbf{C}_{n|n}$ .

Also plotted is an estimated signal-to-noise ratio (Figure 4.4c)

$$\text{SNR}_n = \frac{\|\text{var}(\mathbf{H}\mathbf{u}_n)\|_2}{\|\text{var}(\mathbf{y}_n)\|_2},$$

as well as the norm of the Kalman gain matrix (Figure 4.4d),

$$\|\mathbf{C}_{n|n-1}\mathbf{H}^\top (\mathbf{H}\mathbf{C}_{n|n-1}\mathbf{H}^\top + \sigma_n^2\mathbf{I})^{-1}\|_2.$$

These quantities qualitatively accord with one another and suggest that where the forecast RMSE is higher, the correction (as measured by the magnitude of the Kalman gain) is smaller. Again, after  $t \approx 3$ , these quantities reach stable configurations.

With the parameter estimates in Figure 4.5, these results indicate that where the noise standard deviation  $\sigma_n$  is overestimated, the increased measurement uncertainty results in smaller corrections as measured by the Kalman gain norm. In these instances, NL-statFEM is not able to recognise the model error and instead treats this as noise. Unsurprisingly, the larger noise standard deviation estimates at these points results in smaller signal-to-noise ratios. Upon the noise being accurately estimated the relative error and RMSE both shrink to what appear to be asymptotic values.

Shown in Figure 4.6 (top), the estimated posterior means provide a reasonable approximation to the underlying DGP. Up to  $t \approx 3.0$ , the means retain the highly steepened feature of the assumed model, which arises due to the incorrect model specification of more dominant nonlinear effects. After  $t \approx 3.0$ , the data is seen to have corrected for this misspecification, resulting in posterior means that accord with the DGP. This then results in the covariance structure (Figure 4.6, bottom) changing from a highly localised region about the steepened front, to a more spatially uniform pattern by the end time of the simulation. Note also that we do not observe the same increases in the 95% credible intervals as we do in the prior (c.f. Figure 4.1); in this case conditioning on data refines the uncertainty.

## 4.2 The Kuramoto Sivashinsky equation

The Kuramoto-Sivashinsky (KS) equation is a chaotic, biharmonic PDE that is used to model pattern formation in a variety of physical contexts. Kuramoto (Kuramoto, 1978; Kuramoto and Tsuzuki, 1975, 1976) derived the equation in the context of chemical reaction kinetics, and, in parallel, Sivashinsky (Michelson and Sivashinsky, 1977; Sivashinsky, 1977, 1980) derived the equation in the study of diffusive instabilities in laminar flame fronts. The equation was also derived in Benney (1966a), in the context of liquid films flowing down an inclined plane. The KS equation is given by

$$\begin{cases} u_t + u_{xx} + \mu u_{xxxx} + \frac{1}{2}u_x^2 = 0, \\ u := u(x, t), \quad x \in [0, L], \quad t \in [0, T], \\ u(x, 0) = u_0(x), \quad u(x, t) = u(x + L, t). \end{cases} \quad (4.7)$$

Alternatively, the KS equation can be specified in terms of  $v = u_x$ , which gives a PDE with similar nonlinearity to the Burgers and Navier-Stokes equations

$$v_t + v_{xx} + \mu v_{xxxx} + vv_x = 0. \quad (4.8)$$

This is sometimes referred to as KS being in “derivative form” (Collet et al., 1993). The KS equation defines a dissipative dynamical system whose solutions are contained in an absorbing ball as  $t \rightarrow \infty$ . Nicolaenko et al. (1985) show that as  $\lim_{t \rightarrow \infty} \|v\| \leq CL^{5/2}$  for some  $C > 0$ . The KS equation also displays finite-dimensional dynamics, with the resultant inertial manifold dimension  $d_i$  being bounded by  $d_i \leq CL^{3.75} + 1$  (Foias et al., 1988; Hyman and Nicolaenko, 1986).

The coefficient  $\mu$  parameterises the strength of the elasticity term  $u_{xxxx}$ . The dimensionless parameter  $\tilde{L} = L/(2\pi\sqrt{\mu})$  is taken as the bifurcation parameter (Hyman and Nicolaenko, 1986). As  $\tilde{L}$  increases the KS equation solutions transition to spatiotemporal chaos. The transition to chaos is marked by alternating windows of laminar behaviour — where one solution is globally attracting — and oscillatory or chaotic behaviour. Globally attracting solutions at these regions of laminar behaviour appear to increase linearly in the number of dominant modes. After  $\tilde{L} \gtrsim 5.415$  spatiotemporal chaos is seen, which is the situation that will be studied when considering the NL-statFEM augmentation of (4.7).

This transition to chaos can be analysed through studying the modes of (4.7). This proceeds through expanding  $u(x, t) = \sum_{k=1}^{\infty} \hat{u}_k(t) \exp(2\pi i k x / L) / \sqrt{L}$ . The Fourier basis  $\{\exp(2\pi i k x / L) / \sqrt{L}\}_{k=0}^{\infty}$  forms

an orthonormal basis of the  $L$ -periodic functions in  $L^2([0, L]; \mathbb{R})$ . Making the substitution into (4.7) and setting  $\mu = 1$  admits the coupled system of ODEs over the Fourier coefficients

$$\partial_t \hat{u}_j = \left(\frac{2\pi j}{L}\right)^2 \hat{u}_j - \left(\frac{2\pi j}{L}\right)^4 \hat{u}_j + \frac{1}{2\sqrt{L}} \left(\frac{2\pi}{L}\right)^2 \sum_k \hat{u}_k \hat{u}_{j-k} k(j-k). \quad (4.9)$$

This equation provides some insight into how each term contributes to the dynamics. The second-order term  $u_{xx}$  increases the energy on the lower order modes whilst on the higher order modes, the  $u_{xxxx}$  term dampens. The nonlinear term  $u_x^2$  acts to transfer energy between modes. The balance between this damping and energy transfer, as parameterised by the domain length  $L$  and damping  $\mu$ , thus controls the dynamical regime of the model. As  $L$  increases the damping shifts to higher order modes, allowing the lower order modes to have a greater effect.

We choose to study this PDE due to chaotic behaviour being present in a scalar 1D system. To study chaos in other PDEs, dimensionality and complexity need to be increased — for example, Navier-Stokes requires at least 2D model domains, and requires solving over both velocity and pressure fields. This requires computational development which is beyond the scope of this chapter. Also, different to the Burgers examples of the previous section, KS solutions do not exhibit drastic dissipation in the timescales that we study in this chapter and are qualitatively more oscillatory than the previous examples.

In this section we consider the 1D stochastic KS equation in derivative form, given by

$$\begin{cases} u_t + uu_x + \nu u_{xx} + u_{xxxx} = \xi, \\ u := u(x, t), \quad x \in [0, 32\pi], \quad t \in [0, 100], \\ u(x, 0) = u_0(x), \quad u(0, t) = u(32\pi, t), \end{cases} \quad (4.10)$$

where  $\xi(x, t) \sim \mathcal{GP}(0, \delta(t - t') \cdot k_\theta(x, x'))$ , with  $k_\theta(\cdot, \cdot)$  set as previous. We study the efficacy of NL-statFEM in assimilating data in a highly nonlinear dynamical system with model misspecification. As in the previous subsection, misspecification arises from an incorrect viscosity coefficient. We show that NL-statFEM is able to appropriately assimilate data, providing small ( $\approx 1\%$ ) relative errors and stable hyperparameter estimates.

### 4.2.1 Discretisation

To get the weak form we multiply by a test function  $v \in V$ , and integrate by parts. We discretise in time, writing  $u^n(x, t) = u(x, n\Delta_t)$ , to give the implicit Euler semi-discrete (time-discretised) weak form

$$\begin{aligned} \langle u^n - u^{n-1}, v \rangle + \Delta_t \langle u^n \partial_x u^n, v \rangle - \Delta_t \nu \langle \partial_x u^n, \partial_x v \rangle - \Delta_t \langle \partial_x w^n, \partial_x v \rangle &= \langle \Delta \xi^n, v \rangle, \\ -\langle \partial_x u^n, \partial_x v \rangle - \langle w^n, v \rangle &= \langle \Delta \xi^n, v \rangle, \quad \forall v \in V. \end{aligned} \quad (4.11)$$

where  $\Delta \xi^n \sim \mathcal{GP}(0, \Delta_t k_\theta(x, x'))$ . Note that all boundary terms disappear due to periodic boundaries. As in Burgers we discretise the domain  $\mathcal{D} = [0, 32\pi]$  into the mesh with nodes  $x_i = ih$ , for  $i = 0, \dots, n_c$  such that  $n_c h = 32\pi$ . In this case we set  $u_h(x, t) = \sum_{i=1}^{n_u} u_i(t) \phi_i(x) \in V_h$ , where  $V_h = \text{span}\{\phi_i(x)\}_{i=1}^{n_u}$ , for the piecewise linear polynomial basis functions. This expansion is defined similarly for the  $v_h(x, t)$

and  $w_h(x, t)$  and gives

$$\begin{aligned} \langle u_h^n - u_h^{n-1}, \phi_j \rangle + \Delta_t \langle u_h^n \partial_x u_h^n, \phi_j \rangle - \Delta_t \nu \langle \partial_x u_h^n, \partial_x \phi_j \rangle - \Delta_t \langle \partial_x w_h^n, \partial_x \phi_j \rangle &= \langle \Delta \xi^n, \phi_j \rangle, \\ -\langle \partial_x u_h^n, \partial_x \phi_j \rangle - \langle w_h^n, \phi_j \rangle &= \langle \Delta \xi^n, \phi_j \rangle, \quad j = 1, \dots, n_u, \end{aligned} \quad (4.12)$$

implicitly defining the updating method for  $\mathbf{u}_n = (u_1(n\Delta_t), \dots, u_{n_u}(n\Delta_t))^\top$ . Instead of using higher-order basis functions we double the state dimension to account for the biharmonic term. Fenics is again used to implement the discretisation, with Newtons method to solve the nonlinear system and the LU decomposition to solve the inner linear system. In practice, we also set  $n_c = 512$  and  $\Delta_t = 0.02$ .

### 4.2.2 Example: mismatched viscosity

Data is generated according to the deterministic (4.8). In this case we set  $\nu = 0.95$ . We observe data  $\mathbf{y}_n \in \mathbb{R}^{n_y}$ , with  $n_y = 52$ , taken at uniformly spaced values on the mesh. Simulated i.i.d. observational noise is added on, with  $\boldsymbol{\eta}_n \sim \mathcal{N}(\mathbf{0}, 0.05^2 \mathbf{I}_{n_y})$ , to give the DGP  $\mathbf{y}_n = \mathbf{H}\mathbf{u}_n + \boldsymbol{\eta}_n$ .

For the NL-statFEM model, we induce mismatch by setting  $\nu = 1$ . As before, we set the covariance of  $\xi$  to that of (3.2), with parameters  $\boldsymbol{\theta}_n = (\rho_n, \ell_n)$ . We estimate  $\boldsymbol{\theta}_n$  and  $\sigma_n$  at each timestep  $n$ , using the weakly informative truncated Gaussian priors

$$p(\rho_n) \sim \mathcal{N}_+(1, 1^2), \quad p(\ell_n) \sim \mathcal{N}_+(1, 1^2), \quad p(\sigma_n) \sim \mathcal{N}_+(0, 1^2).$$

The posterior distribution  $p(\mathbf{u}_n \mid \mathbf{y}_{1:n}, \hat{\boldsymbol{\theta}}_{1:n}, \hat{\sigma}_{1:n})$  is computed with the ExKF. The initial conditions for both the DGP and the posterior are the same and are given the Dirac measure,  $p(\mathbf{u}_0) = \delta_{\mathbf{m}_0}(\mathbf{d}\mathbf{u}_0)$ . These initial conditions are set from simulating (4.8), with  $\nu = 0.95$  and  $u_0(x) = \sin(x/16)$ , for 2000 timesteps to skip over transient behaviour<sup>2</sup>.

The verification quantities introduced in Section 4.1.3 are shown in Figure 4.7. For the posterior the relative errors (Figure 4.7a) hover at  $\approx 0.015$ ; for the prior these are more variable but appear to have a mean value of  $\approx 1$ . Note also the rapid increase in the prior relative error that results from the incorrectly specified viscosity. Incorporating data through the NL-statFEM formulation corrects for this. Due to the oscillatory nature of KS solutions, the relative error is stable over the solution runtime as no dissipative effects are seen on this timescale.

The forecast RMSE (Figure 4.7b) appears close to the known value of  $\sigma_n = 0.05$  (shown as a dashed line). In this case, the prediction step is accurate and gives reasonable predictions to the underlying data generating process. This also suggests that the model with misspecified viscosity is locally accurate but leads to divergences as  $t$  increases. This is also supported by Figure 4.7a: the relative errors appear visually indistinguishable for small  $t$ , yet diverge as  $t$  increases. The norm of the Kalman gain (Figure 4.7c)<sup>3</sup> does not see the same initial drop as in Burgers equation. In this case the noise hyperparameter is accurately estimated at all times, so the Kalman updates are stable and noise is not mistaken for dynamics. Hyperparameter estimates (Figure 4.7d) show that the noise standard deviation is accurately estimated (c.f. Figure 4.7b). In this case the  $\mathcal{GP}$  hyperparameter estimates ( $\hat{\rho}_n, \hat{\ell}_n$ ) are not identified and are estimated near their prior mean values.

<sup>2</sup>In some fields this is called the “spin-up” period (see, e.g., Kalnay and Yang, 2010; Yang et al., 1995).

<sup>3</sup>In this case, given the similarity of the Kalman gain norm to the signal-to-noise ratio, we plot only the Kalman gain norm.

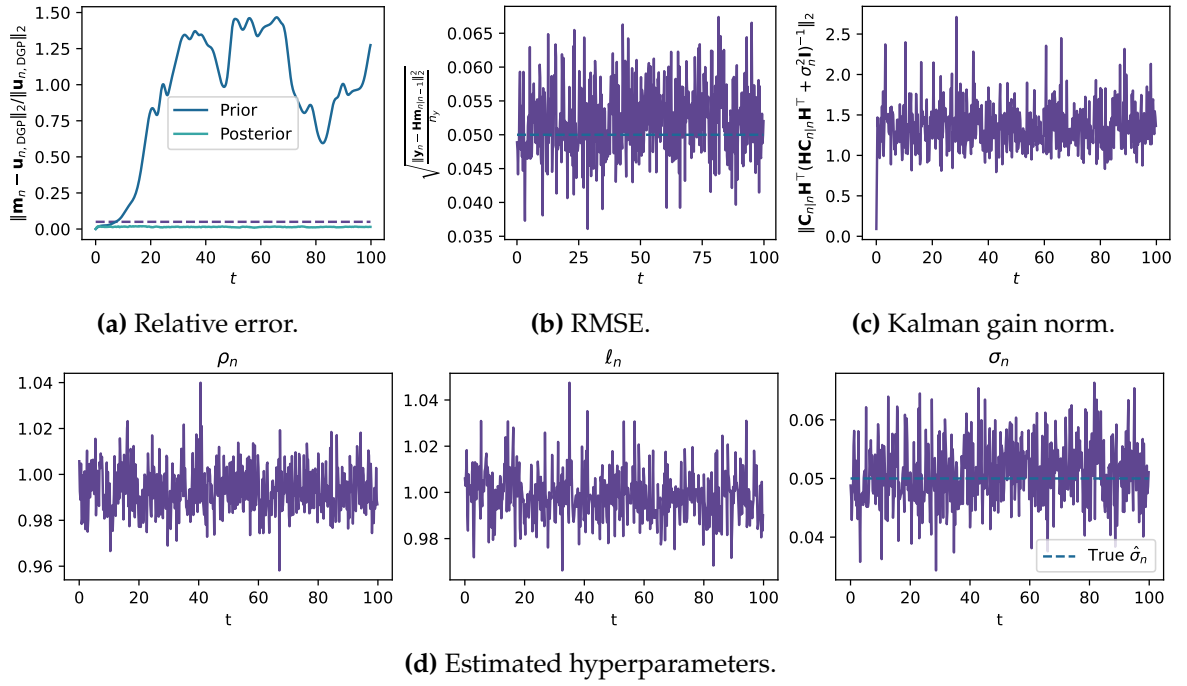


Figure 4.7: KS example: verification quantities.

The prior and posterior means are compared in Figure 4.8. On the space-time plot of Figure 4.8a, the means are similar up to  $t \approx 20$ , at which point they diverge due to the different values of  $\nu$ . The dramatic divergence is due to chaotic dynamics. For a collection of temporal snapshots, shown in Figure 4.8b, the NL-statFEM posterior mean acts as a physics-informed interpolator for the data that corrects for the mismatch present in the prior. In this case the uncertainty bounds are not visible and the posterior is highly certain about the mean  $\mathbf{m}_{n|n}$ .

### 4.3 The Korteweg-de Vries equation

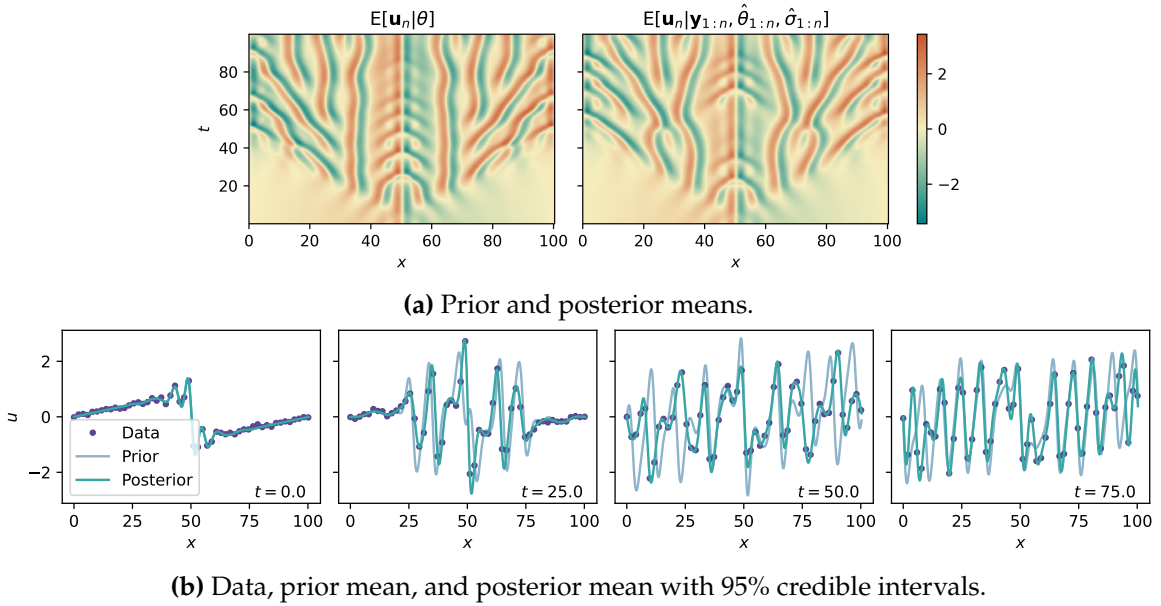
The Korteweg-de Vries (KdV) equation is one of the most widely-studied nonlinear PDEs. For classical parameter values it is given by (Drazin and Johnson, 1989)

$$\begin{cases} u_t + 6uu_x + u_{xxx} = 0, \\ u := u(x, t), \quad x \in \mathbb{R}, \quad t \in \mathbb{R}_+. \end{cases} \quad (4.13)$$

Solutions are characterised by their *nonlinear* and *dispersive* behaviour due to the  $uu_x$  and  $u_{xxx}$  terms, respectively.

The KdV equation has a long history, intertwined with the study of solitary waves (Allen, 1998). Observations of these were made by John Scott Russell in 1834, and theoretical foundations were subsequently laid by Boussinesq (1871) and Lord Rayleigh (1876). Building on these works, Korteweg and de Vries (1895) derived their namesake equation to model the solitary wave propagating in a rectangular canal. The KdV equation was then studied by a group of post-WWII physicists (see, e.g., Gardner et al., 1967; Lax, 1968; Miura et al., 1968; Zabusky and Kruskal, 1965), where analytical properties of the KdV equation were of interest. This line of research culminated in a proof that KdV





**Figure 4.8:** KS example: posterior mean results.

was integrable and resulted in significant developments in the field of integrable systems.

In this section, we study the KdV equation in the oceanographic context, so we provide a brief introduction to its use in this setting. For more detail see, for example, [Gerkema and Zimmerman \(2008\)](#). In oceanography, KdV arises as a model of nonlinear internal waves ([Rayson et al., 2011, 2019](#)), which are observed as waves of depression or elevation along a pycnocline. Nonlinear internal waves are of broad interest in the physical sciences, due to the resulting sediment resuspension and mixing ([Jones et al., 2020](#)), which has both biological and geological effects ([Boegman and Stastna, 2019; Cacchione et al., 2002](#)). They also result in currents and bores, which draw attention from the engineering community due to the resultant interactions with offshore facilities ([Huang et al., 2016; Osborne and Burch, 1980](#)).

KdV can be derived from considering a two-dimensional case of the Navier-Stokes equations for an incompressible Newtonian fluid ([Benney, 1966b; Lamb and Yan, 1996; Lee and Beardsley, 1974](#)), under the additional assumption that the flow is inviscid

$$\begin{cases} \partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho} \nabla \mathbf{p} + \mathbf{g}, \\ \nabla \cdot \mathbf{u} = 0, \\ \mathbf{u} := \mathbf{u}(\mathbf{x}, t), \quad \mathbf{u} \in \mathbb{R}^2, \quad \mathbf{x} = (x, z) \in [0, L] \times [0, H], \quad t \in \mathbb{R}_+. \end{cases} \quad (4.14)$$

In this case  $\mathbf{u}$  is a two-dimensional velocity field which describes the flow in the horizontal and vertical directions, and  $\mathbf{g}$  is a gravitational acceleration vector. To derive the KdV equation we assume that (i)  $\rho \approx \bar{\rho} + \Delta_\rho \rho_0(z)$ ; and (ii) the flow is both shallow in the vertical height-scale and long in the horizontal length-scale.

Under these assumptions (4.14) is solved using the *streamfunction*, which is assumed to have the separable form  $\psi(x, z, t) = u(x, t)\phi(z)$ . The *vertical mode* is  $\phi(\cdot)$  and the *wave profile* is given by  $u(\cdot)$ . The

wave profile  $u$  gives the displacement from the pycnocline<sup>4</sup>, which for a two-layer system is the internal wave displacement from the equilibrium density interface. Asymptotics give the KdV equation

$$u_t + \alpha uu_x + \beta u_{xxx} + cu_x = 0, \quad (4.15)$$

which is readily interpretable: waves propagate at wave speed  $c$ , nonlinear steepening results from  $uu_x$ , and dispersion is due to  $u_{xxx}$ . Relative coefficient values determine the dominating regime and waves can vary from quasi-linear to highly nonlinear. Coefficients are determined from the solution of the eigenvalue problem

$$\phi''(z) + \frac{N^2(z)}{c^2} \phi(z) = 0, \quad N^2(z) = -\frac{g}{\bar{\rho}} \frac{d\rho_0}{dz},$$

where  $N^2(z)$  is the Brunt-Väisälä frequency (Cushman-Roisin and Beckers, 2011). The coefficients are then given by

$$\alpha = \frac{3c}{2} \frac{\int (\phi'(z))^3 dz}{\int (\phi'(z))^2 dz}, \quad \beta = \frac{c}{2} \frac{\int (\phi(z))^2 dz}{\int (\phi'(z))^2 dz}. \quad (4.16)$$

These require knowledge of the underlying density profile  $\rho(z)$ , which can be either fixed *a priori*, or estimated using in-situ measurements (Manderson et al., 2019).

For boundary conditions, one possibility is to specify an inflow boundary condition at  $x = 0$ , so that  $u(0, t) = I(t)$ , and a normal outflow boundary condition of  $\partial_x u(x, t)|_{x=L} = 0$ , or, an absorbing boundary layer condition  $\partial_x u(x, t)|_{x=L'} = 0$ , where  $L' > L$ . Periodic boundary conditions can also be used, which is our choice for the remainder of this section.

For NL-statFEM, we consider the stochastic version of (4.15)

$$\begin{cases} u_t + \alpha uu_x + \beta u_{xxx} + cu_x + \nu u = \xi, \\ u := u(x, t), \quad x \in [0, L], \quad t \in [0, T], \\ u(x, t) = u(x + L, t), \\ u(x, 0) := u_0(x). \end{cases} \quad (4.17)$$

Again with  $\xi(x, t) \sim \mathcal{GP}(0, \delta(t - t') \cdot k_\theta(x, x'))$ , with  $k_\theta(x, x') = \rho^2 \exp(-\|x - x'\|_2^2 / (2\ell^2))$ . We include an additional linear dissipation term,  $\nu$ , in (4.17), as for the second example, the model becomes impractically mismatched by the end time of the simulation if this term is ignored. We acknowledge that for laminar boundaries other methods are preferred (Grimshaw et al., 2003; Horn et al., 2002). The dissipation coefficient  $\nu$  is an inverse time-scale, having units 1/(time).

From a modelling perspective, KdV-type models offer a less accurate but computationally less burdensome mathematical description of reality. In this section we study two examples of model misspecification. The first is brought about through incorrect assumptions on model nonlinearity, where a cubic nonlinear term is added into (4.15) to define the DGP. The second misspecification occurs with experimental data, which uses measurements obtained in a two-layer stratified system within a tank. In this scenario there is no “true” model and all models are misspecified to some extent (Judd and Smith, 2004). These examples show that: (a) NL-statFEM is effective at correcting for

<sup>4</sup>The pycnocline is the point at which the vertical density profile gradient is largest; for a two-layer system the pycnocline is the interface between the two layers.



misspecification which is brought about through nonlinearity; and (b) NL-statFEM is able to deal with real-world misspecification of physics in sparse observation regimes.

### 4.3.1 Discretisation

We discretise using the scheme outlined in [Debussche and Printems \(1999\)](#). We proceed first by time discretisation, then spatial discretisation. With  $u^n(x) = u(x, n\Delta_t)$ , we use Crank-Nicolson for time integration and continue as usual by multiplying with test functions  $\psi \in V^0$  and integrating with respect to  $x$  over  $[0, L]$  to give the weak form

$$\begin{aligned} \langle u^{n+1} - u^n, \psi \rangle + \Delta_t \alpha \langle u^{n+1/2} u_x^{n+1/2}, \psi \rangle + \Delta_t \beta \langle u_{xxx}^{n+1/2}, \psi \rangle \\ + \Delta_t c \langle u_x^{n+1/2}, \psi \rangle + \Delta_t \nu \langle u^{n+1/2}, \psi \rangle = \langle \Delta \xi^n, \psi \rangle, \quad \forall \psi \in V^0, \end{aligned}$$

with  $\Delta \xi^n \sim \mathcal{GP}(0, \Delta_t k_\theta(x, x'))$ . As with KS, we split the above into a coupled system of three first order equations

$$\begin{aligned} \langle u^{n+1} - u^n, \psi \rangle + \Delta_t \alpha \langle u^{n+1/2} u_x^{n+1/2}, \psi \rangle + \Delta_t \beta \langle u_x^{n+1/2}, \psi \rangle + \Delta_t c \langle u_x^{n+1/2}, \psi \rangle + \Delta_t \nu \langle u^{n+1/2}, \psi \rangle &= \langle \Delta \xi^n, \psi \rangle, \\ \langle u_x^{n+1/2}, \psi \rangle &= \langle v^{n+1/2}, \psi \rangle, \\ \langle v_x^{n+1/2}, \psi \rangle &= \langle w^{n+1/2}, \psi \rangle, \quad \forall \psi \in V^0. \end{aligned}$$

Discretise the domain  $\mathcal{D} = [0, L]$  into the mesh with nodes  $x_i = ih$ , for  $i = 0, \dots, n_c$ , so that  $n_c h = L$ . In this case we specify separate function spaces for  $u_h(x, t)$  and  $\psi_h(x, t)$ . For the trial functions,  $u_h(x, t) = \sum_{i=1}^{n_u} u_i(t) \phi_i(x) \subset C^0([0, L]; \mathbb{R})$ , so that  $V_h^1 = \text{span}\{\phi_i\}_{i=1}^{n_u}$ , for the piecewise linear basis functions. This expansion is defined similarly for  $v_h(x, t)$  and  $w_h(x, t)$ , and we use the same number of basis functions for each, so the total state dimension is  $3n_u$ . For the test function space we set  $V_h^0 = \text{span}\{\varphi_i(x)\}_{i=1}^{n_u}$ , where  $\varphi_i(x)$  are the piecewise constant functions which are defined as (for  $0 \leq i < n_c$ )

$$\varphi_i(x) = \begin{cases} \varphi_i(x) = 1, & x \in [x_i, x_{i+1}), \\ \varphi_i(x) = 0, & \text{otherwise.} \end{cases}$$

Thus,  $\psi_h(x) = \sum_{i=1}^{n_u} \psi_i \varphi_i(x)$ .

In this case, we do not use Fenics, but have instead developed our own in-house Python package to solve the system, built on top of Scipy. Solving the nonlinear system at each time is implemented using Newton's method, and convergence is typically achieved in  $< 5$  Newton iterations. The linear system to be solved at each timestep is solved directly, with the sparse LU decomposition. [Debussche and Printems \(1999\)](#) note that using the Crank-Nicolson scheme avoids numerical dissipation present in explicit schemes.

### 4.3.2 Mismatched nonlinearity

In this section, data is generated from an extended KdV equation with a cubic nonlinear term:

$$\begin{cases} u_t + \alpha uu_x + \beta u_{xxx} + \varepsilon u^3 u_x = 0, \\ u := u(x, t), \quad x \in [0, 20], \quad t \in [0, 50], \\ u(x, t) = u(x + 20, t), \\ u(x, 0) := u_0(x). \end{cases} \quad (4.18)$$

We set  $\alpha = 1$ ,  $\beta = 0.01$ , and  $\varepsilon = 20$ . The NL-statFEM model solves (4.17) with the same  $\alpha$  and  $\beta$ , and  $c = \nu \equiv 0$ , on the same grid as (4.18). Initial conditions, for both the DGP and NL-statFEM model, are set to a wave of depression  $u(x, 0) = -0.3 \operatorname{sech}^2(x - 15)$ . For the discretisation settings we use  $n_c = 400$ ,  $\Delta_t = 0.025$ .

The data generating process is simulated using Dedalus (Burns et al., 2020) with 1024 gridpoints in space. This is then downsampled to 20 gridpoints and jittered with synthetic Gaussian observational noise,  $\boldsymbol{\eta}_n \sim \mathcal{N}(0, 0.001^2 \mathbf{I}_{n_y})$ , to give the simulated dataset.  $\mathcal{GP}$  hyperparameters  $\boldsymbol{\theta}_n = (\rho_n, \ell_n)$  and noise standard deviation  $\sigma_n$  are estimated at each step by maximising the log-marginal posterior with the weakly informative truncated Gaussian priors  $\rho_n \sim \mathcal{N}_+(1, 1^2)$ ,  $\ell_n \sim \mathcal{N}_+(1, 1^2)$ , and  $\sigma_n \sim \mathcal{N}_+(0, 1^2)$ . The ExKF is used to compute the posterior  $p(\mathbf{u}_n \mid \mathbf{y}_n, \hat{\boldsymbol{\theta}}_{1:n}, \hat{\sigma}_n)$ .

Relative errors (Figure 4.9a) show that after  $t \approx 15$  the filter reaches a stable configuration with errors at around 10%. RMSE values (Figure 4.9b) stabilise after an initial jump, yet remain above the known noise value at each iteration. The Kalman gain (Figure 4.9c) suggests that in this case model corrections are large in comparison to those seen previously. As a result, this example contains a larger degree of model misspecification. Parameter estimates (Figure 4.9d) indicate that the length and noise parameters are both stable, with the noise being slightly overestimated. Times at which the noise is not identified result in it being set to the lower bound (shown as the vertical bars on the figure). In this case model predictions, as quantified by  $\rho_n$ , vary in their accuracy and appear approximately bounded to within  $(10^{-5}, 10^{-1})$ . The length hyperparameter is poorly identified and is often set to near the prior mean value of  $\ell_n = 1$ .

Next, we compare the posterior results in terms of the estimated wave profiles. For comparison, we begin with the estimated prior, which is shown Figure 4.10a (left). For the fixed set of hyperparameters  $\rho_n = 0.0025$ ,  $\ell_n = 1$  for all  $n$ , the prior is visually mismatched in the mean via a horizontal displacement, increased oscillations, and increased wave interactions. Wave interactions in this case arise due to the periodic boundary conditions, which result in higher amplitude oscillations (and hence faster travelling waves due to nonlinear advection) from high- and low-amplitude wavefronts interacting. Note that the stochastic forcing induces an uncertainty about the PDE solution, represented by the 95% credible intervals shown. Note also that the data generating process is approximately contained within the credible intervals. Figure 4.10a (right) shows that the posterior mean approximates the data generating process, and the posterior uncertainty bounds have shrunk as a result of conditioning, indicating high certainty about the posterior mean values. Model discrepancy between the data and the NL-statFEM model has been corrected for. The space-time view of the posterior mean, shown in Figure 4.10b, shows that the NL-statFEM posterior has incorporated the complex soliton interactions

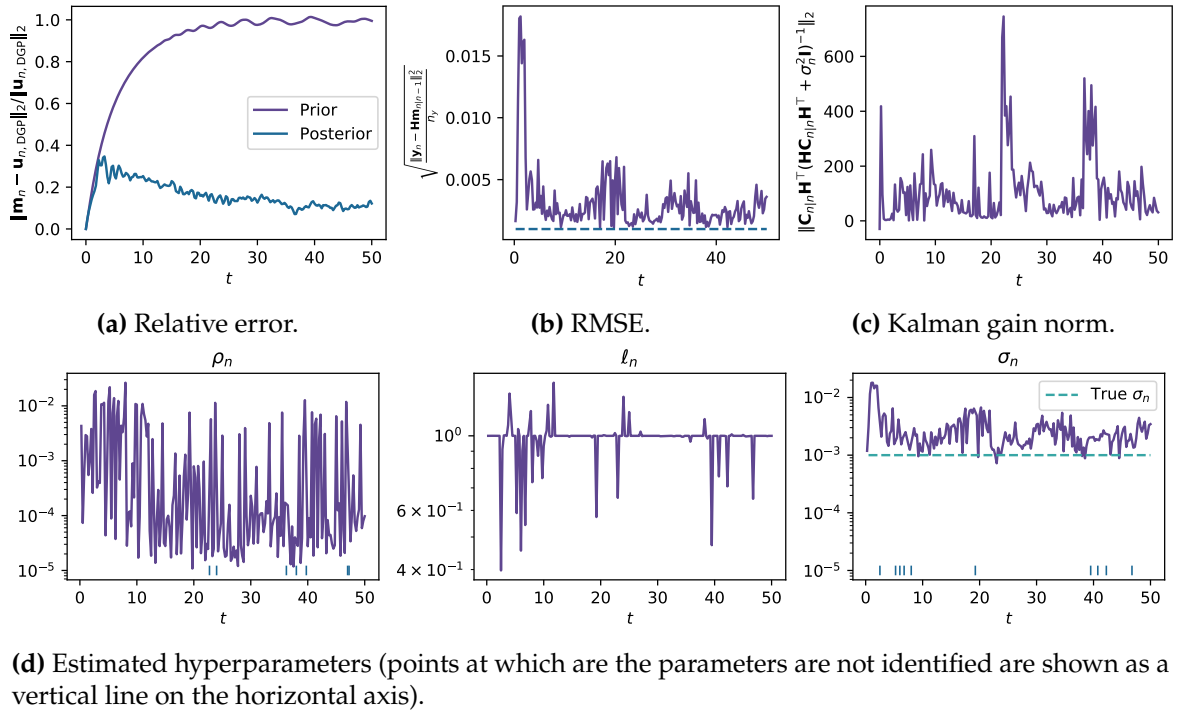


Figure 4.9: KdV cubic example: verification quantities.

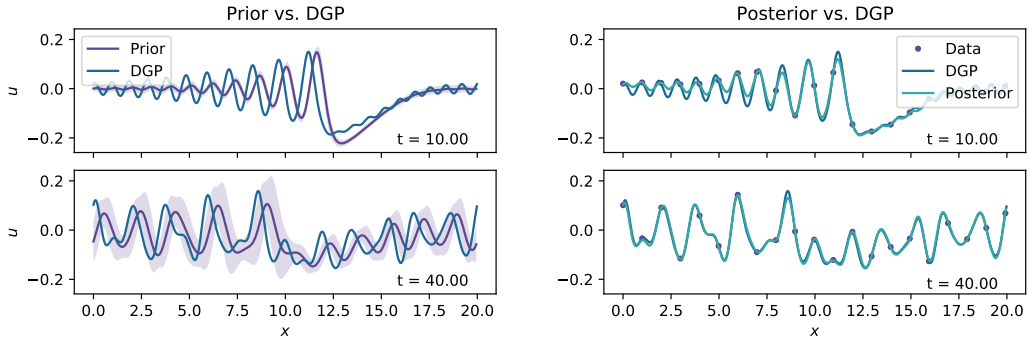
in the data, not present in the prior.

### 4.3.3 Experimental data

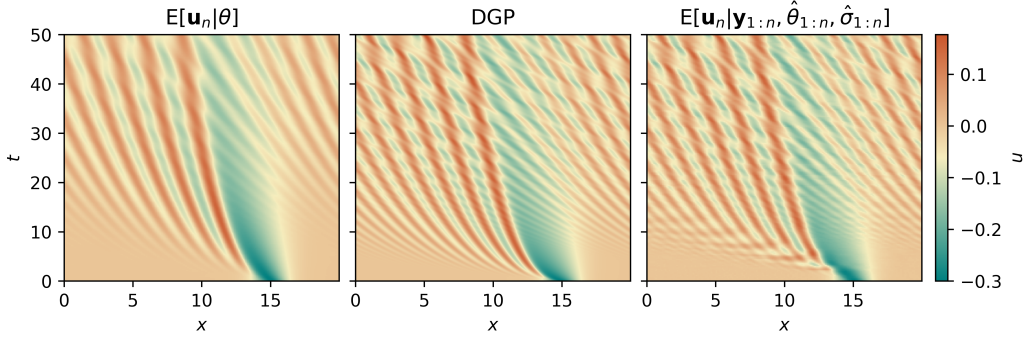
We now apply the method to the experimental data collected in [Horn et al. \(2002\)](#). Experiments were conducted to study weakly nonlinear models for internal waves in lakes and consisted of generating internal waves in a two-layer stratified system, inside of a clear acrylic tank of dimensions  $6 \times 0.3 \times 0.29$  m. The tank contained an upper layer of fresh water and a lower layer of saline water, with a density gradient of  $\Delta\rho = 20 \text{ kg m}^{-3}$ . The tank was able to rotate in order to establish the initial conditions, which were an inclined plane of angle  $\vartheta = 0.5^\circ$ . This initial condition mimics the shear induced by strong winds in lakes. At time  $t = 0$  the tank is rotated to restore it to the horizontal.

Data were recorded at three spatially equidistant locations in the tank using ultrasonic wave gauges, taking measurements approximately every  $0.01$  s, up to  $T = 1000$  s. We use data up to  $T = 300$  s, as beyond this point the leading wave train begins to dissipate and slowly approaches equilibrium; we are mainly interested in the behaviour up to and after the initial wave train passes through. Data are measured in voltages and are post-processed to give pycnocline displacements in metres. A schematic of the experimental setup is shown in [Figure 4.11](#). In this case we have  $H = 0.29$  m and  $L = 6$  m. Initial conditions result in soliton trains forming after an initial transient period. This is seen when plotting the data, shown in [Figure 4.12](#), where the small measurement error is visually apparent. Transient behaviour is observed before steepening and a soliton wave train forms; three such steepening events are observed in the data we analyse. As  $T \rightarrow 1000$  s dissipation results in the wave profile approaching a flat steady-state profile.

Our physical model is the extended KdV (eKdV) equation of [\(4.17\)](#), taking  $x \in [0, 2L]$ ,  $t \in [0, T]$ ,



(a) Left: Data generating process, and prior mean with 95% credible intervals. Right: data, data generating process, posterior means, and 95% credible intervals.



(b) Means across space-time grid.

Figure 4.10: KdV cubic example: visualising the posterior mean.

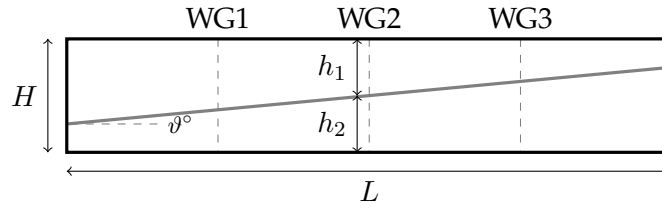
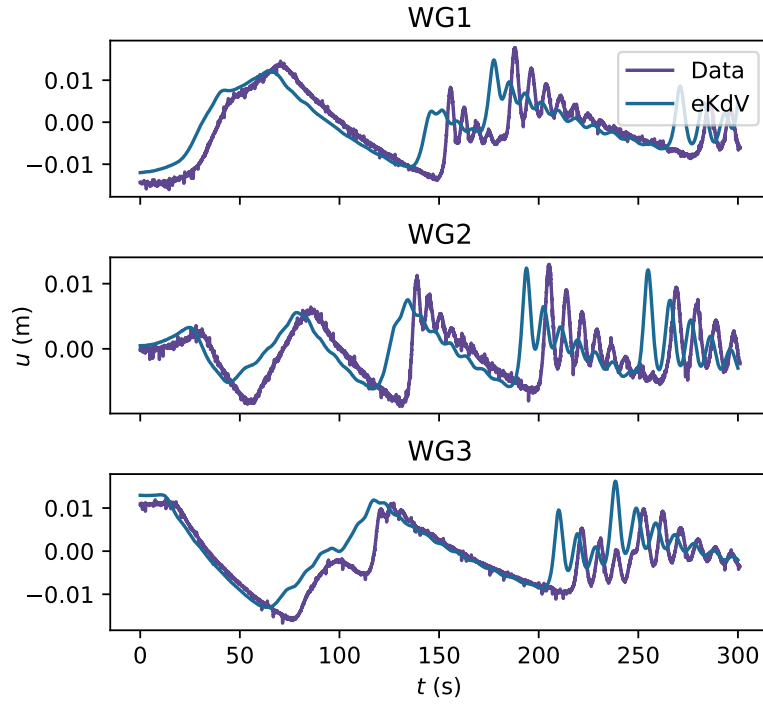


Figure 4.11: Schematic diagram of the experimental apparatus. Wave-gauges are labelled WG1, WG2, and WG3, and the initial conditions are shown as a grey line, labelled with initial angle  $\vartheta^\circ$ .

and with coefficients

$$\alpha = \frac{3c(h_1 - h_2)}{2h_1h_2}, \quad \beta = \frac{ch_1h_2}{6}, \quad c = \sqrt{\frac{g'h_1h_2}{H}}, \quad g' = \frac{\Delta\rho g}{\rho_0}.$$

For the experiment under consideration, we have  $h_1 = 0.232$  m,  $h_2 = 0.058$  m,  $H = h_1 + h_2 = 0.29$  m,  $\rho_0 = 1000$  kgm<sup>-3</sup>. The dissipation coefficient  $\nu$  is an inverse time-scale which is set to  $3 \times 10^{-3}$  s<sup>-1</sup>. The NL-statFEM prior mean, at the locations of the wave-gauges, is shown in Figure 4.12. This model does not capture the observed behaviour exactly, becoming apparent after the initial transient dynamical period. In this instance, model mismatch gradually appears as a process that acts over long time-scales. Model waves have higher velocity than the observations and model amplitudes are slightly larger than observed amplitudes. It is conjectured that this is due to misparameterisation of dissipation, but in any case the model is misspecified.



**Figure 4.12:** Observed data at wave-gauges and NL-statFEM prior mean (given by the deterministic FEM solution to assumed eKdV equation).

Incorporation of reflective boundary conditions is done by solving the eKdV equation across the extended domain  $[0, 2L]$ , with periodic boundary conditions, and summing solutions in the (reflected) subdomains  $[0, L]$ ,  $[L, 2L]$ :

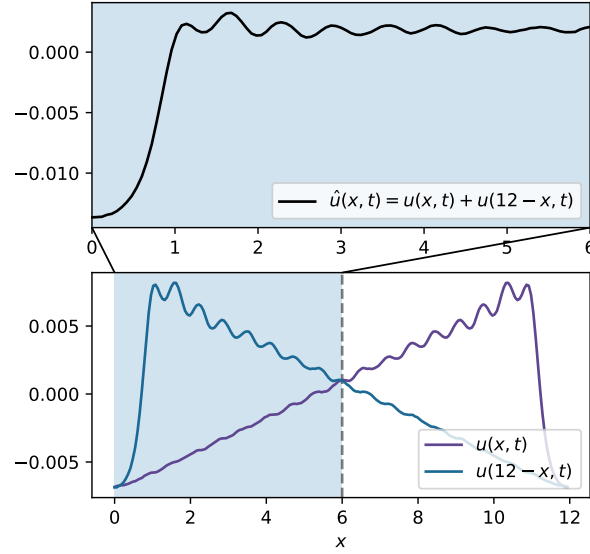
$$\hat{u}(x, t) := u(x, t) + u(2L - x, t), \quad x \in [0, L].$$

For details on the derivation see [Horn et al. \(2002\)](#). An illustration of this procedure is shown in [Figure 4.13](#). Solving on the extended periodic domain allows for reflective boundary conditions to be considered as the sum of left-moving and right-moving wave components.

Rather than estimating the eKdV parameters using inversion techniques we sequentially update the model with observations to give the posterior  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \hat{\boldsymbol{\theta}}_{1:n}, \sigma)$ . Note that this posterior is over the full extended domain  $[0, 2L]$ , and we can access the posterior over the physical model domain through summing the appropriate components of the estimated  $\mathbf{u}_n$ , to give the estimated  $p(\hat{\mathbf{u}}_n | \mathbf{y}_{1:n}, \hat{\boldsymbol{\theta}}_{1:n}, \sigma)$ . As before we assume  $\mathbf{y}_n = \mathbf{H}\mathbf{u}_n + \boldsymbol{\eta}_n$  with known noise  $\boldsymbol{\eta}_n \sim \mathcal{N}(0, 1.3588 \times 10^{-8} \mathbf{I}_{n_y})$ .

Due to small data in space (three observations at each timestep) we use a projection method to estimate hyperparameters. This linearly projects the predicted mean  $\mathbf{m}_{n|n-1}$  forward, estimated from the data points:  $y_{n,i} = a_n + b_n \mathbf{m}_{n|n-1}(x_i)$ . Parameters  $a_n, b_n$  are estimated to give the best least-squares linear projection from the prediction to the data. This gives a projected dataset,  $\tilde{\mathbf{y}}_n$ , using the linear shift:  $\tilde{\mathbf{y}}_n = a_n + b_n \mathbf{m}_{n|n-1}$ . The estimated hyperparameters are then given by

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}_n}{\operatorname{argmax}} \{ \log p(\tilde{\mathbf{y}}_n | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}_n, \hat{\boldsymbol{\theta}}_{1:n-1}, \sigma) + \log p(\boldsymbol{\theta}_n) \},$$



**Figure 4.13:** Extended domain extended KdV: mapping from the periodic domain  $2L$  into the reflective subdomain  $L$  allows for reflective boundary conditions to be incorporated.

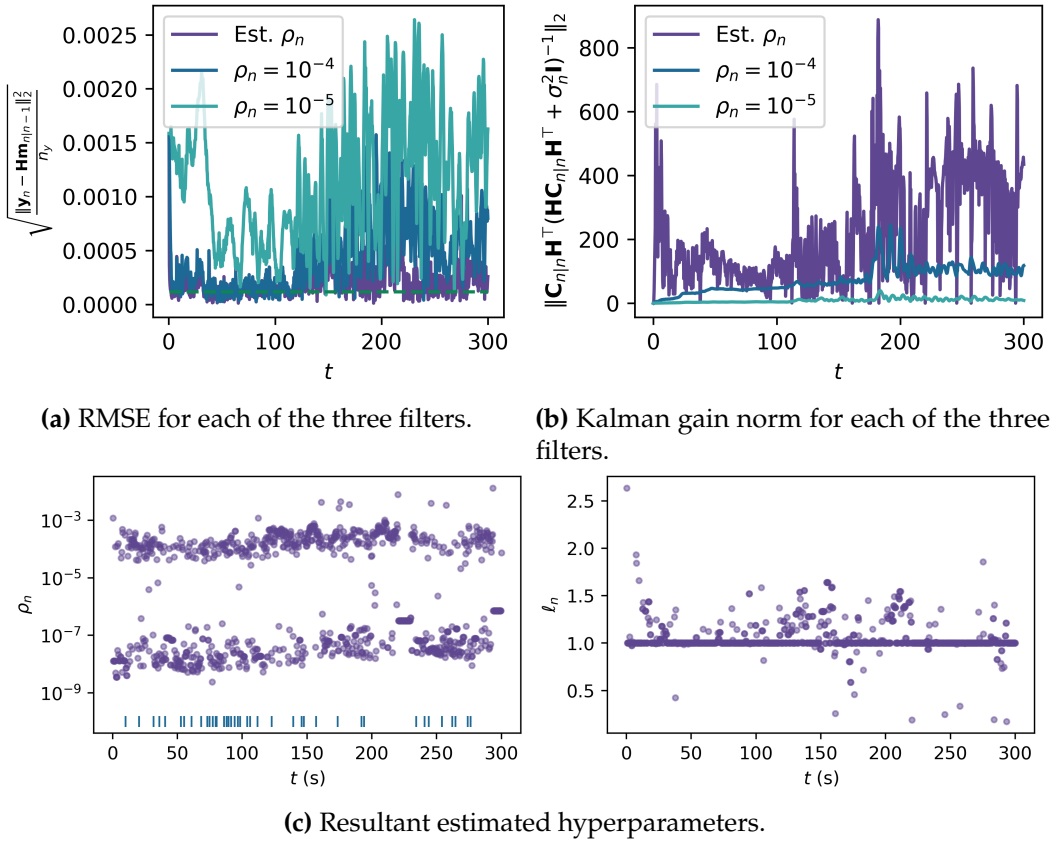
in which the observed data  $\mathbf{y}_n$  is replaced with the projected data  $\tilde{\mathbf{y}}_n$ . We project to a grid of 100 points uniformly spaced across the solution grid. Note that this is only for the parameter estimation step and we do not use this  $\tilde{\mathbf{y}}_n$  as the data in the update step. We set weakly informative priors:  $\rho_n \sim \mathcal{N}_+(1, 1^2)$  and  $\ell_n \sim \mathcal{N}_+(1, 1^2)$ .

In this case, we consider three separate filters, each with different hyperparameters. These are: (i)  $(\rho_n, \ell_n) = (10^{-5}, 1)$ , (ii)  $(\rho_n, \ell_n) = (10^{-4}, 1)$ , and (iii)  $(\rho_n, \ell_n)$  estimated for all times. For ease of reference, we refer to these as the small- $\rho_n$ , the medium- $\rho_n$  and the estimated- $\rho_n$  filters, respectively. We run three separate filters to ensure that filter performance is reasonable, given the sparsity of observations and the parameter estimation procedure. In each case, the NL-statFEM posterior is computed using the ExKF method with  $n_c = 200$  FEM cells and  $n_t = 1001$  timesteps.

Diagnostics are shown in Figure 4.14a, and it is seen that the filters stratify in terms of the RMSE. The small- $\rho_n$  filter has the largest RMSE values, consistently larger than the medium- $\rho_n$  and the estimated- $\rho_n$ . At the data observation locations, the small- $\rho_n$  filter is the least accurate. The medium- $\rho_n$  and the estimated- $\rho_n$  appear similar at the start of the simulation, but separate by the end time  $T = 300$  s, with the medium- $\rho_n$  larger than the estimated- $\rho_n$ . In Figure 4.14b we see this stratification again, with the small- $\rho_n$  filter having the smallest corrections, followed by the medium- $\rho_n$ , and the estimated- $\rho_n$ .

Estimated hyperparameters are shown in Figure 4.14c. The  $\rho_n$  parameter is seen to vary between two distinct levels, hinting that model predictions vary in their accuracy. In this case  $\ell_n$  again struggles to be identified from the prior mean. Taken with the parameter estimates of Figure 4.14c we see that with decreasing values of  $\rho_n$ , NL-statFEM trusts the underlying model more and is less uncertain of the predictions. In terms of the estimated log marginal likelihood, the results are:

- Small- $\rho_n$ :  $\log p(\mathbf{y}_{1:n_t} | \boldsymbol{\theta}_{1:n_t}, \sigma) = -138,584.5440$ .



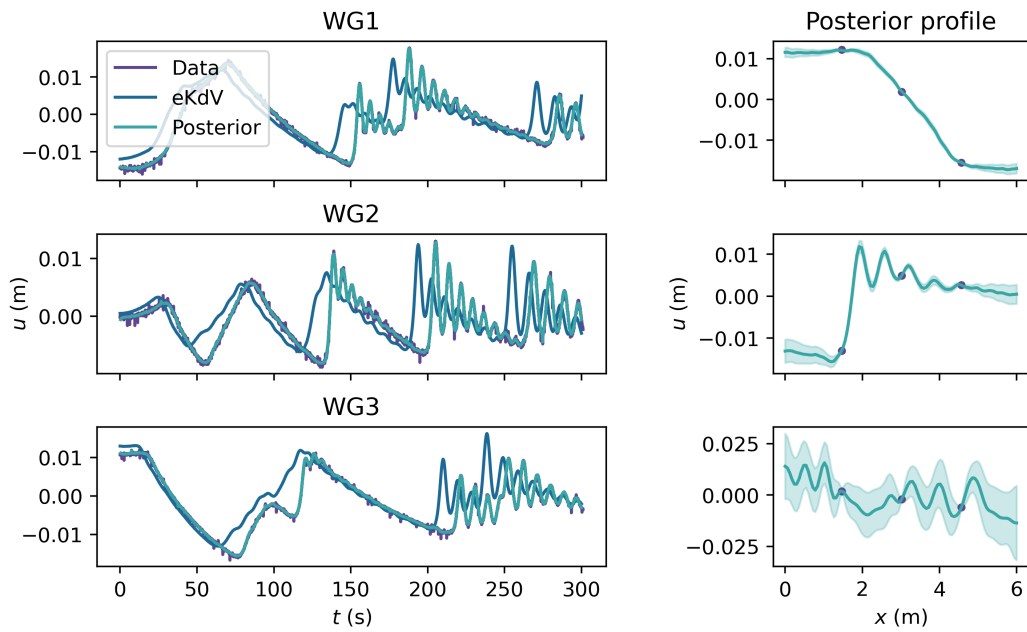
**Figure 4.14:** KdV experimental data: verification diagnostics.

- Medium- $\rho_n$ :  $\log p(\mathbf{y}_{1:n_t} | \boldsymbol{\theta}_{1:n_t}, \sigma) = 1,758.9706$ .
- Estimated- $\rho_n$ :  $\log p(\mathbf{y}_{1:n_t} | \hat{\boldsymbol{\theta}}_{1:n_t}, \sigma) = 20,914.4421$ .

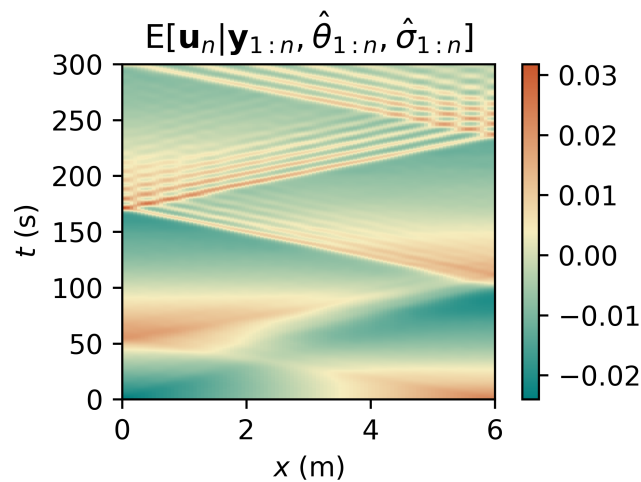
We see that the estimated- $\rho_n$  filter gives the best values of the log marginal likelihood; optimising the log marginal posterior gives the most performant model. For the remainder of this section, we therefore discuss the results of the estimated- $\rho_n$  filter only.

At the wave-gauge locations, the posterior mean  $\mathbf{m}_{n|n}$  is shown in Figure 4.15 (left). In comparison to the prior mean, the posterior mean offers a close fit to the data. The credible intervals shrink about the data and are not seen on the figure. Posterior wave profiles, shown in Figure 4.15 (right), demonstrate that given the data, the method is able to yield a sensible estimate for the underlying wave profile and is hence able to reconstruct the wave profile given sparse observations in space. The uncertainty bounds are also sensible, in this case, shrinking about data observation locations, and growing where there are no observations. Through incorporating *a priori* physical information, NL-statFEM provides a physics-informed online regression methodology, that works in scenarios in which observations are spatiotemporally sparse. A space-time view of the posterior mean wave profile is shown in Figure 4.16, and demonstrates that the general behaviours of the flow (e.g. reflective boundary conditions, dissipation, wave train formation) are all indeed captured.





**Figure 4.15:** KdV experimental data. Left: data, prior mean, and posterior means and 95% credible intervals across time, at the three observation locations. Right: Posterior means and 95% credible intervals, across the domain, at three timepoints. The posterior, in this case, uses the estimated hyperparameters.



**Figure 4.16:** KdV experimental data: heatmap of the posterior mean  $\mathbf{m}_{n|n}$  for estimated hyperparameters filter.





# Chapter 5

## Low-rank filtering for scalability

*Despite the appeal of NL-statFEM, computational scalability is a challenge to its application to high-dimensional problems typically experienced in physical and industrial contexts. This chapter overcomes this hurdle by embedding a low-rank approximation of the underlying dense covariance matrix, obtained from the leading order modes of the full-rank alternative (Duffin et al., 2022). We focus on the ExKF, deriving a low-rank variant abbreviated by LR-ExKF. The scalable NL-statFEM, which uses the LR-ExKF, is presented in the context of reaction-diffusion systems: canonical examples of nonlinear, time-dependent phenomena in 2D. This setting provides the necessary increase in state dimension from the 1D examples considered in Chapter 4, being also amenable to finite element discretisation due to parabolicity.*

*This chapter proceeds by first introducing reaction-diffusion equations. Next, we derive the LR-ExKF using the NL-statFEM construction and discuss the algorithm. The chapter is concluded with three case studies using experimental and synthetic data. The first is a 1D example with experimental data, and we show that the LR-ExKF is able to accurately reproduce the full-rank ExKF, with small relative errors on the mean and the variance. The next two examples deal with synthetic data, both using the Oregonator, a 2D system of two coupled PDEs, and demonstrate scalability, taking the state dimension up to 132,098 degrees-of-freedom. Examples show that the NL-statFEM accurately reproduces the data generating process under two scenarios of misspecified initial conditions. We also use the effective rank to verify the effective dimensionality of the low-rank covariance matrix, ensuring that the number of modes chosen is adequate. Code to reproduce these results is publicly available at <https://github.com/connor-duffin/low-rank-statfem>.*

### 5.1 Reaction-diffusion systems

Reaction-diffusion (RD) systems are semilinear parabolic PDEs that describe the evolution of the system state as it diffuses throughout the problem domain, whilst undergoing nonlinear interactions with itself. In the simplest form, a general RD equation is

$$\mathbf{u}_t = \mathbf{r}(\mathbf{u}; \Lambda) + \kappa \Delta \mathbf{u}, \quad \mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}). \quad (5.1)$$

with  $\mathbf{u} := \mathbf{u}(\mathbf{x}, t) \in \mathbb{R}^s$ ,  $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^d$ ,  $t \in [0, T]$ . As in the previous chapters, parameters  $\Lambda$  are assumed known.

[Turing \(1952\)](#) discusses how reaction and diffusion processes, inside some organic medium, form the basis for chemical morphogenesis<sup>1</sup>. This was experimentally verified in [Castets et al. \(1990\)](#). The setting in which these ideas developed was that of molecular biology. In this setting differential equations can be derived from the chemical principle of mass action, where the rate of a chemical reaction to the concentration of the reactants raised to the power of their stoichiometric coefficients. These principles can be used to derive ODE systems which describe the evolution of chemical concentrations, as specified by some governing chemical process which is spatially homogenous. Adding in spatial inhomogeneity, whose dynamical representation arises through Fick's law of diffusion ([Fick, 1855](#)), allows for spatial variation to be taken into account and gives, instead of an ODE system, a PDE system. This is the case in the original biological context of Turing.

A common example of a RD system — and the one that Turing studied initially — is an *activator-inhibitor* system. An activator-inhibitor system comprises of two states — so that  $\mathbf{u} \in \mathbb{R}^2$  — where one component is autocatalytic (i.e., it creates more of itself) and the other slows down this autocatalysis. These two components form the *activator* and *inhibitor*, respectively.

To illustrate the ideas of the *Turing instability*, we follow the approach of [De Wit \(2007\)](#). A steady-state solution to (5.1),  $\bar{\mathbf{u}}$ , is assumed, which is temporally and spatially homogeneous. This solution also satisfies  $\mathbf{r}(\bar{\mathbf{u}}) = 0$ . This state is perturbed by some function  $\mathbf{v}(\mathbf{x}, t) = \sum_{n=1}^{\infty} \mathbf{c}_n \exp(\omega_n t) \phi_n(\mathbf{x})$ , where  $\Delta \phi_n = \lambda_n \phi_n$  for each  $n$ . Linearising (5.1) about  $\bar{\mathbf{u}}$  gives a PDE over the perturbed state  $\bar{\mathbf{u}} + \mathbf{v}$

$$\mathbf{v}_t = \mathbf{J}(\bar{\mathbf{u}}; \Lambda) \mathbf{v} + \kappa \Delta \mathbf{v} \quad (5.2)$$

where  $\mathbf{J}(\bar{\mathbf{u}}; \Lambda) = \frac{\partial \mathbf{r}}{\partial \mathbf{u}}(\bar{\mathbf{u}}; \Lambda)$ . Making the appropriate substitutions gives the equation in series form, which gives the condition  $\|\omega_n \mathbf{I} - \kappa \lambda_n \mathbf{I} - \mathbf{J}(\bar{\mathbf{u}}; \Lambda)\|_2 = 0$  in order for the PDE to hold.

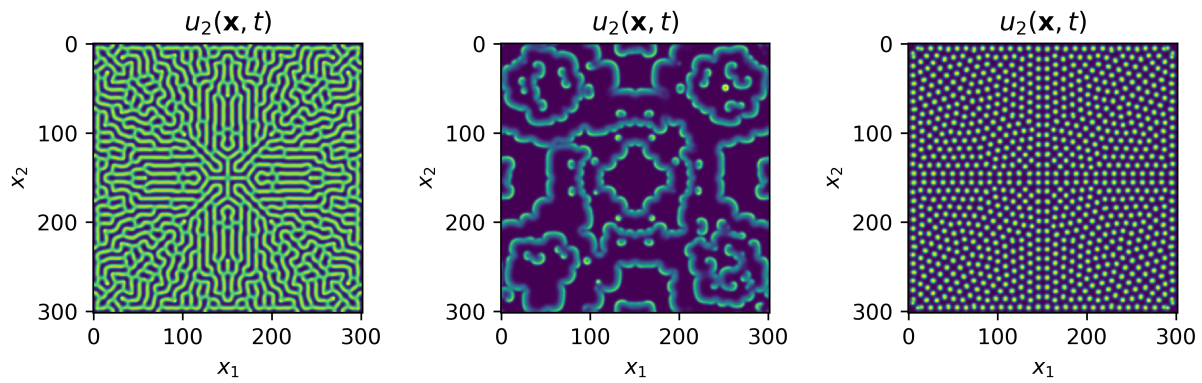
Observe that in order for solutions to decay to some steady state we require that  $\Re(\omega_n) < 0$ . If this is the case the system is in the stable regime. The *Turing instability* occurs when the transition to  $\Re(\omega_n) > 0$  occurs for some  $n$ , meaning that solutions are now unstable and at least one mode drifts away at some exponential rate. This results from changing parameters  $\Lambda$  or  $\kappa$  and gives rise to a bifurcation. The *Hopf instability* occurs when  $\Re(\omega_n) = 0$ , resulting in temporal oscillations which do not decay.

*Turing patterns* are the spatiotemporal patterns that emerge as a result of the balance of reaction and diffusion processes. To illustrate the formation of these patterns, we consider the *Gray-Scott (GS) model* ([Pearson, 1993](#))

$$\begin{aligned} \frac{\partial u_1}{\partial t} &= -u_1 u_2^2 + F(1 - u_1) + \kappa_1 \Delta u_1, \\ \frac{\partial u_2}{\partial t} &= u_1 u_2^2 - (F + k)u_2 + \kappa_2 \Delta u_2. \end{aligned} \quad (5.3)$$

The GS model is an activator-inhibitor system with  $u_2$  the activator and  $u_1$  the inhibitor. The behaviour of the model solutions depends on the parameters  $\Lambda = (F, k, \kappa_1, \kappa_2)$ ; the parameter  $F$  is known as the feed rate and is the rate that  $u_1$  is pumped into the domain. The parameter  $k$  is the kill rate and is the rate that  $u_2$  is removed from the domain. Depending on the choices of these parameters, the dynamics vary between various qualitative spatiotemporal patterns; three archetypal examples of the dynamics

<sup>1</sup>Morphogenesis is the process by which cells change and form into the specific structure that the resultant organism becomes.



(a) “Soliton”-type pattern with  $(F, k) = (0.055, 0.062)$ . (b) “Spiral”-type pattern with  $(F, k) = (0.014, 0.047)$ . (c) “Spot”-type pattern with  $(F, k) = (0.028, 0.062)$ .

**Figure 5.1:** Numerical solutions of the  $u_2$  component of the Gray-Scott equation over the domain  $\mathcal{D} = [0, 300] \times [0, 300]$ , with  $\kappa_1 = 0.1$ ,  $\kappa_2 = 0.05$ . Different choices of the parameters  $(F, k)$  result in variations in the dynamics.

are shown in Figure 5.1 and reflect various patterns seen in biology.

RD systems are semilinear parabolic equations that in general do not have closed form solutions. They are an appropriate candidate for finite element methods, due to being parabolic, and have potentially complex geometries. They can also be misspecified, due to approximations being employed in their derivation which neglect certain chemical reaction kinetics and variables. For example, in the Oregonator, a model for the dynamics of the Belousov-Zhabotinskii (BZ) reaction (Field and Noyes, 1974; Field et al., 1972; Jahnke et al., 1989; Tyson and Fife, 1980), a state dimension can be set to its steady state value, thereby reducing  $s$  from three to two. RD systems are appropriate for studying the effects of model misspecification in systems which, despite nonlinearity, are not highly nonlinear. This also makes RD systems appropriate for testing the filtering methodology, as the LR-ExKF can be applied without modifications required to account for strongly nonlinear, chaotic dynamics.

## 5.2 Low-rank NL-statFEM

We start with the FEM discretisation of (5.1). Without loss of generality, we deal with single-state equations with  $u \in \mathbb{R}$ , discretising the system to yield a coupled system of ODEs over the FEM coefficients. As for the previous examples, we triangulate the domain  $\mathcal{D}$  into  $\mathcal{D}_h$  and look for solutions  $u_h(\mathbf{x}, t) = \sum_{i=1}^{n_u} u_i(t) \phi_i(\mathbf{x})$ , in  $V_h = \text{span}\{\phi_i\}_{i=1}^{n_u}$ , the solution space. In this section we again take  $\phi_i$  to be the linear polynomial  $C^1(\mathcal{D}; \mathbb{R})$  hat-functions. Multiplying by  $\phi_j$  and integrating Equation (5.1) over space admits the weak form

$$\langle \partial_t u_h, \phi_j \rangle + \kappa \mathcal{A}(u_h, \phi_j) = \langle r(u_h), \phi_j \rangle, \quad j = 1, \dots, n_u,$$

which corresponds to the  $n_u$ -dimensional ODE for the FEM coefficients  $\mathbf{u}(t) = (u_1(t), \dots, u_{n_u}(t))^T$

$$\mathbf{M} \frac{d\mathbf{u}}{dt} + \kappa \mathbf{A} \mathbf{u} = \tilde{\mathbf{r}}(\mathbf{u}),$$

where  $\mathbf{M}_{ij} = \langle \phi_i, \phi_j \rangle$ ,  $\mathbf{A}_{ij} = \mathcal{A}(\phi_i, \phi_j)$ , and  $\tilde{\mathbf{r}}(\mathbf{u})_j = \langle r(u_h), \phi_j \rangle$ . Under bounded derivative conditions on  $r(\cdot)$ , this ODE will have unique solutions by the Picard-Lindelöf theorem (Thomée, 2006, Chapter 13)

For stability, Crank-Nicolson is used for the time discretisation. As in the previous chapters, we set the timestep size  $\Delta_t > 0$ ,  $\mathbf{u}_n = (u_1(n\Delta_t), \dots, u_{n_u}(n\Delta_t))^\top$ , and  $\mathbf{u}_{n-1/2} = (\mathbf{u}_n + \mathbf{u}_{n-1})/2$ . The fully discrete system is therefore

$$\mathbf{M}(\mathbf{u}_n - \mathbf{u}_{n-1}) + \Delta_t \kappa \mathbf{A} \mathbf{u}_{n-1/2} = \Delta_t \tilde{\mathbf{r}}(\mathbf{u}_{n-1/2}),$$

and is accurate to order  $\mathcal{O}(h^2 + \Delta_t^2)$  (Thomée, 2006, Chapter 14).

Assuming the same initial and boundary conditions, NL-statFEM adds stochastic forcing to Equation (5.1):

$$\begin{aligned} u_t &= \kappa \nabla^2 u + r(u) + \xi, \\ \xi &:= \xi(\mathbf{x}, t) \sim \mathcal{GP}(0, \delta(t - t') \cdot k_\theta(\mathbf{x}, \mathbf{x}')). \end{aligned} \quad (5.4)$$

Spatial discretisation with finite elements gives the stochastic differential equation for the FEM coefficients

$$\mathbf{M} d\mathbf{u} + \kappa \mathbf{A} \mathbf{u} dt = \tilde{\mathbf{r}}(\mathbf{u}) dt + d\xi,$$

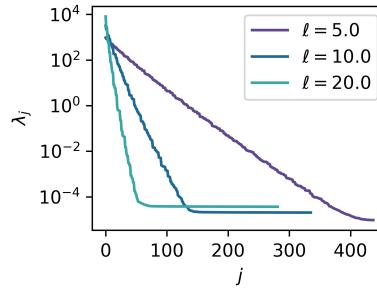
and time discretisation, with the Crank-Nicolson scheme, gives the fully-discrete evolution equation on the FEM coefficients

$$\mathbf{M}(\mathbf{u}_n - \mathbf{u}_{n-1}) + \Delta_t \kappa \mathbf{A} \mathbf{u}_{n-1/2} = \Delta_t \tilde{\mathbf{r}}(\mathbf{u}_{n-1/2}) + \mathbf{e}_n, \quad (5.5)$$

where  $\mathbf{e}_{n,j} = \langle \xi_n - \xi_{n-1}, \phi_j \rangle$ . The  $\mathbf{e}_n$  are i.i.d. with  $\mathbf{e}_n \sim \mathcal{N}(0, \Delta_t \mathbf{G}_\theta)$ , for all  $n$ . Equation (5.5) implicitly defines the NL-statFEM prior  $p(\mathbf{u}_n | \boldsymbol{\theta}, \Lambda)$ , conditioned on the PDE coefficients  $\Lambda$ . An approximation to the prior can be computed using the methods presented in Chapter 3. This would require either estimating the covariance through a tangent linear approximation to the dynamics (the ExKF method), or through Monte Carlo simulation (the EnKF method). We note also that Monte Carlo simulation of (5.5), recovers the standard Euler-Maruyama discretisation (Kloeden and Platen, 1992).

As previous (c.f. Chapter 3) we employ the covariance approximation  $\mathbf{G}_\theta \approx \mathbf{M} \mathbf{K}_\theta \mathbf{M}^\top$ . The covariance matrix  $\mathbf{K}_\theta$  can be represented by a possibly low-rank square root  $\mathbf{K}_\theta = \mathbf{K}_\theta^{1/2} \mathbf{K}_\theta^{\top/2}$ , with  $\mathbf{K}_\theta^{1/2} \in \mathbb{R}^{n_u \times k'}$  for some  $k' \leq n_u$ . If the spectrum of  $\mathbf{K}_\theta$  is rapidly decaying then a majority of the variance can be explained by  $k' \ll n_u$  dominant modes, providing a scalable approach to storing the posterior covariance. In this work we use the squared-exponential covariance function, which is known to have rapidly decaying eigenvalues (Zimmermann, 2015); for efficient methods to decompose  $\mathcal{GP}$  covariance matrices into their leading eigenvalues we refer to Charlier et al. (2021); Dietrich and Newsam (1997); Saatci (2011); Solin and Särkkä (2020).

As an example, for a  $128 \times 128$  finite element mesh over the square domain  $\mathcal{D} = [0, 50] \times [0, 50]$ , the leading eigenvalues of the squared-exponential covariance matrix  $\mathbf{K}_\theta$  are plotted in Figure 5.2, for various values of the length-scale parameter  $\ell \in \{5, 10, 20\}$ . After a maximum of  $\approx 400$  eigenvalues (for  $\ell = 5$ ) the single-precision limit is reached, suggesting that in this case a low-rank approximation will accurately capture the covariance matrix, whilst having much lower storage requirements —  $\mathcal{O}(k'n_u)$



**Figure 5.2:** Leading eigenvalues of  $\mathbf{K}_\theta$  over the  $128 \times 128$  mesh inside the domain  $\mathcal{D} = [0, 50] \times [0, 50]$ . Values are truncated to those which are positive, where the negative values arise due to single-precision floating points being used.

floating points as opposed to  $\mathcal{O}(n_u^2)$ . This motivates the proceeding filtering algorithm.

Now suppose that data is arriving, corrupted with Gaussian noise and observed through the linear observation operator  $\mathbf{H} : \mathbb{R}^{n_y} \rightarrow \mathbb{R}^{n_u}$ . The data generating process is

$$\mathbf{y}_n = \mathbf{H}\mathbf{u}_n + \boldsymbol{\eta}_n, \quad \boldsymbol{\eta}_n \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_y}),$$

and we compute the posterior distribution  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \boldsymbol{\theta}, \sigma, \Lambda)$ , with  $\mathbf{y}_{1:n} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ . To do so we use a low-rank variant of the ExKF, which computes the Gaussian approximation  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \boldsymbol{\theta}, \sigma, \Lambda) \approx \mathcal{N}(\mathbf{m}_n, \mathbf{C}_n)$ , from a low-rank approximation of the state covariance matrix  $\mathbf{C}_n = \mathbf{L}_n \mathbf{L}_n^\top$ . This is constructed with the leading eigenvalues and eigenvectors of the covariance matrix  $\mathbf{K}_\theta$  and the previous timestep covariance  $\mathbf{C}_{n-1}$ . For similar approaches see [Gillijns et al. \(2006\)](#); [Law and Stuart \(2012\)](#); [Rozier et al. \(2007\)](#); [Verlaan and Heemink \(1997\)](#).

Alternative filtering methods may be used in this context: common examples include the particle filter and the EnKF. However for high-dimensional systems the particle filter is known to suffer from particle collapse ([Bengtsson et al., 2008](#)) and may require a computationally infeasible number of particles to give accurate UQ. In terms of computational performance, the EnKF is known to scale favourably as the state dimension is increased, providing estimates of the posterior mean which accurately estimate the underlying data generating process. It is also able to make effective use of distributed computing, due to conditional independence of the ensemble in the prediction step. However due to the Monte Carlo structure of the EnKF the estimates of the posterior variance may be inaccurate due to small ensemble sizes ([Law and Stuart, 2012](#)). Our interest in the LR-ExKF is thus motivated by additional results in [Law and Stuart \(2012\)](#), which show that using a low-rank covariance matrix in the ExKF can provide accurate UQ for regularly observed nonlinear dissipative systems, similar to what is considered here.

Assume that the distribution of the previous state is given by  $p(\mathbf{u}_{n-1} | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}, \sigma, \Lambda) = \mathcal{N}(\mathbf{m}_{n-1}, \mathbf{C}_{n-1})$ , with  $\mathbf{C}_{n-1} = \mathbf{L}_{n-1} \mathbf{L}_{n-1}^\top$ ,  $\mathbf{L}_{n-1} \in \mathbb{R}^{n_u \times k}$ . Furthermore, assume that a low-rank square root of  $\mathbf{G}_\theta = \mathbf{G}_\theta^{1/2} \mathbf{G}_\theta^{\top/2}$  is also available, where  $\mathbf{G}_\theta^{1/2} \in \mathbb{R}^{n_u \times k'}$ ; this is constructed via  $\mathbf{G}_\theta^{1/2} = \mathbf{M} \mathbf{K}_\theta^{1/2}$ ,  $\mathbf{K}_\theta^{1/2} \in \mathbb{R}^{n_u \times k'}$ . Also note that  $D_n \tilde{\mathbf{r}} := \frac{\partial \tilde{\mathbf{r}}(\mathbf{u}_{n+1/2})}{\partial \mathbf{u}_n}$ .

For all timesteps  $n = 1, \dots, n_t$ , the LR-ExKF proceeds as:

1. (Prediction step) Solve

$$\mathbf{M}(\hat{\mathbf{m}}_n - \mathbf{m}_{n-1}) + \Delta_t \kappa \mathbf{A}(\hat{\mathbf{m}}_{n-1/2}) = \Delta_t \tilde{\mathbf{r}}(\hat{\mathbf{m}}_{n-1/2}),$$

for the prediction mean  $\hat{\mathbf{m}}_n$  and compute the prediction covariance square root:

$$\tilde{\mathbf{L}}_n = \left[ (\mathbf{M} + \Delta_t(\kappa \mathbf{A} + D_n \tilde{\mathbf{r}}))^{-1} (\mathbf{M} + \Delta_t D_{n-1} \tilde{\mathbf{r}}) \mathbf{L}_{n-1}, \right. \\ \left. \Delta_t (\mathbf{M} + \Delta_t \kappa \mathbf{A} + \Delta_t D_n \tilde{\mathbf{r}})^{-1} \mathbf{G}_\theta^{1/2} \right]$$

noting that  $\tilde{\mathbf{L}}_n \tilde{\mathbf{L}}_n^\top = \hat{\mathbf{C}}_n$  (the prediction covariance), and that  $\tilde{\mathbf{L}}_n \in \mathbb{R}^{n_u \times (k+k')}$ . Each column of  $\tilde{\mathbf{L}}_n$  can be formed in parallel; see the following discussion for further details.

2. (Truncation step) Take the eigendecomposition  $\tilde{\mathbf{L}}_n^\top \tilde{\mathbf{L}}_n = \mathbf{V}_n \boldsymbol{\Sigma}_n \mathbf{V}_n^\top$ , where  $\boldsymbol{\Sigma}_n = \text{diag}(\varsigma_1, \dots, \varsigma_{k+k'})$ . Approximate  $\hat{\mathbf{L}}_n = \tilde{\mathbf{L}}_n [\mathbf{V}_n]_{:,1:k}$  for the highest magnitude  $k$  modes, so the prediction covariance is  $\hat{\mathbf{C}}_n = \hat{\mathbf{L}}_n \hat{\mathbf{L}}_n^\top$ .
3. (Update step) Update the mean:

$$\mathbf{m}_n = \hat{\mathbf{m}}_n + (\mathbf{H} \hat{\mathbf{C}}_n)^\top \left( \mathbf{H} \hat{\mathbf{C}}_n \mathbf{H}^\top + \sigma^2 \mathbf{I}_{n_y} \right)^{-1} (\mathbf{y}_n - \mathbf{H} \hat{\mathbf{m}}_n).$$

And the covariance:

$$\mathbf{L}_n = \hat{\mathbf{L}}_n \mathbf{R}_n, \\ \mathbf{R}_n \mathbf{R}_n^\top = \mathbf{I}_k - \hat{\mathbf{L}}_n^\top \mathbf{H}^\top (\mathbf{H} \hat{\mathbf{C}}_n \mathbf{H}^\top + \sigma^2 \mathbf{I}_{n_y})^{-1} \mathbf{H} \hat{\mathbf{L}}_n,$$

using a Cholesky decomposition or otherwise.

If  $k = k' = n_u$  the LR-ExKF recovers the full ExKF exactly. If the datum  $\mathbf{y}_n$  is missing only the prediction and truncation steps are completed, to produce the posterior  $p(\mathbf{u}_n | \mathbf{y}_{1:n}, \boldsymbol{\theta}, \sigma, \Lambda) \equiv p(\mathbf{u}_n | \mathbf{y}_{1:n-1}, \boldsymbol{\theta}, \sigma, \Lambda)$ .

## Discussion

First we compare the ExKF and LR-ExKF in terms of memory and operation counts. For a general reference to these matrix computations we refer to [Golub and Van Loan \(2013\)](#). The standard ExKF requires that  $\mathbf{C}_n$  and  $\mathbf{G}_\theta$  are stored in memory, which is  $\mathcal{O}(n_u^2)$  in space. For large DOF problems this is infeasible, and provides the main motivation for the low-rank approach. If one employs the standard ExKF, the prediction step for the covariance matrix requires the solution of the sparse  $n_u \times n_u$  matrix  $\mathbf{M} + \Delta_t(\kappa \mathbf{A} + D_n \tilde{\mathbf{r}})$ ,  $2n_u$  times for each timestep. For large  $n_u$  this becomes prohibitively expensive. If using a direct solver is feasible, this can be slightly mitigated by computing the LU factorisation of  $\mathbf{M} + \Delta_t(\kappa \mathbf{A} + D_n \tilde{\mathbf{r}})$ , and reusing the factors when solving for each column of  $\tilde{\mathbf{L}}_{n-1}$ . However this still requires computing the factorisation, and running forward- and back-substitution  $2n_u$  times per timestep.

Furthermore, the update step requires the solution of the system  $\mathbf{H} \hat{\mathbf{C}}_n \mathbf{H}^\top + \sigma^2 \mathbf{I}_{n_y}$ ,  $n_u$  times, for each timestep, which requires  $\mathcal{O}(n_y^3/3)$  operations to take the Cholesky decomposition and  $\mathcal{O}(n_y^2 n_u)$  to solve.



In comparison, LR-ExKF requires storing only the low-rank square root  $\mathbf{L}_n$ , having the memory requirement of  $\mathcal{O}((k+k')n_u)$  floating points. Also,  $\mathbf{M} + \Delta_t(\kappa\mathbf{A} + D_n\tilde{\mathbf{r}})$  is solved  $k+k'$  times for each timestep, and the truncation step incurs a cost of  $\mathcal{O}((k+k')^3)$  to compute the eigendecomposition. The update step has a cost of  $\mathcal{O}(n_y^3/3)$  operations to take the Cholesky factorisation of  $\mathbf{H}\hat{\mathbf{C}}_n\mathbf{H}^\top + \sigma^2\mathbf{I}_{n_y}$ , but requires only  $k$  solves, to give the cost  $\mathcal{O}(n_y^2k)$ . Finally, the cost of computing the Cholesky factor  $\mathbf{R}_n$  is  $\mathcal{O}(k^3/3)$ , however, as  $k \ll n_u$ , computing the decomposition is dwarfed, in practice, by the prediction step.

Note also that each column of the prediction covariance square root  $\tilde{\mathbf{L}}_n$  is able to be computed in parallel. As each column requires solving  $\mathbf{M} + \Delta_t(\kappa\mathbf{A} + D_n\tilde{\mathbf{r}})$  this can result in lower runtimes, especially so when combined with, for example, algebraic multigrid solvers (Saad, 2003). This parallelisation is likely to be necessary when scaling up the LR-ExKF to larger systems than those considered here, and shares similar parallelisation potential to the EnKF prediction step (Evensen, 2009).

As the filter is initialised with the  $l^2$ -optimal low-rank square-root  $\mathbf{K}_\theta^{1/2}$ , and the truncation step preserves the dominant modes of variation, it is thought that this scheme is able to provide accurate uncertainty quantification. An avenue of future work is verifying this conjecture under regimes of stronger nonlinearity.

The low-rank approximation of the covariance matrix will lead to underestimation. If we write the prediction covariance as  $\hat{\mathbf{C}}_n + \hat{\mathbf{C}}_{n,\text{err}}$  the norm of the discarded component is given by  $\|\hat{\mathbf{C}}_{n,\text{err}}\|_2 = \varsigma_{k+1}$ . This is noted in Gillijns et al. (2006), where it is also commented that this underestimation could lead to similar problems as encountered in ensemble Kalman filtering, such as catastrophic filter divergence (Gottwald and Majda, 2013). We have observed this when using unstable time-integration schemes; we refer to Appendix A.3.3 for full details. There is additional symmetry to the EnKF: both algorithms propagate a low-rank approximation to the covariance square root ( $\mathbf{L}_n$  in the LR-ExKF; the ensemble in the EnKF), and make the Gaussian assumption in the update step. One can think of the LR-ExKF as an EnKF where the ensemble members are chosen to optimally represent the variance (in the  $l^2$  sense) and the covariance propagation is done by the tangent linear model.

To avoid inverting  $\mathbf{H}\hat{\mathbf{C}}_n\mathbf{H}^\top + \sigma^2\mathbf{I}_{n_y}$ , when  $n_y \gg k$ , the Woodbury matrix identity (Mandel, 2006; Rozier et al., 2007) can be used

$$\begin{aligned} \left(\mathbf{H}\hat{\mathbf{C}}_n\mathbf{H}^\top + \sigma^2\mathbf{I}_{n_y}\right)^{-1} &= \left(\mathbf{H}\hat{\mathbf{L}}_n(\mathbf{H}\hat{\mathbf{L}}_n)^\top + \sigma^2\mathbf{I}_{n_y}\right)^{-1} \\ &= \frac{1}{\sigma^2} \left(\mathbf{I}_{n_y} + \mathbf{H}\hat{\mathbf{L}}_n \left(\sigma^2\mathbf{I}_k + (\mathbf{H}\hat{\mathbf{L}}_n)^\top(\mathbf{H}\hat{\mathbf{L}}_n)\right)^{-1} (\mathbf{H}\hat{\mathbf{L}}_n)^\top\right). \end{aligned}$$

The dense  $k \times k$  symmetric positive definite matrix  $\sigma^2\mathbf{I}_k + (\mathbf{H}\hat{\mathbf{L}}_n)^\top(\mathbf{H}\hat{\mathbf{L}}_n)$  can be solved using the Cholesky decomposition (Golub and Van Loan, 2013).

$\mathcal{GP}$  hyperparameters  $\theta$  and noise standard deviation  $\sigma$  can be estimated through the methods discussed in Chapter 3. In this chapter, we do not estimate hyperparameters, choosing to study the performance of the filter for the fixed- $\theta$  case. Hyperparameter estimation, in the LR-ExKF context, is deferred to Appendix A.3.2.



### 5.3 Case studies

The methodology is now demonstrated on three examples. The first is a verification of the method on a 1D example of cell reaction-diffusion, using the experimental data of [Simpson et al. \(2020\)](#), and shows a coherent synthesis of data with a prior physical model. It also confirms that the LR-ExKF accurately reproduces the full-rank ExKF, with relative errors of  $\mathcal{O}(10^{-6})$  for the mean, and  $\mathcal{O}(10^{-4})$  for the variance.

The remaining two examples are in 2D, and use the Oregonator model ([Field et al., 1972](#); [Tyson and Fife, 1980](#)). In each case, we induce model misspecification through the initial conditions. The first demonstrates that observations of a single component of the system can correct for misspecification on the unobserved component. The second example studies the affect of increasing the mismatch variance  $\rho$  on filter performance and also shows scalability, using a state dimension of  $n_u = 132,098$ . Conditioning on data corrects for misspecification, and in both cases NL-statFEM recovers the underlying data generating process to relative errors of  $\mathcal{O}(10^{-2})$ . In these examples, running the full-rank filter is prohibitively expensive so the *effective rank* of  $\hat{\mathbf{L}}_n$  is used as a measure for filter performance.

The appendix contains two additional case studies (Appendices [A.3.2](#) and [A.3.3](#)), which discuss the parameter estimation methodology, and a case of catastrophic filter divergence, respectively.

#### 5.3.1 Experimental data: verification

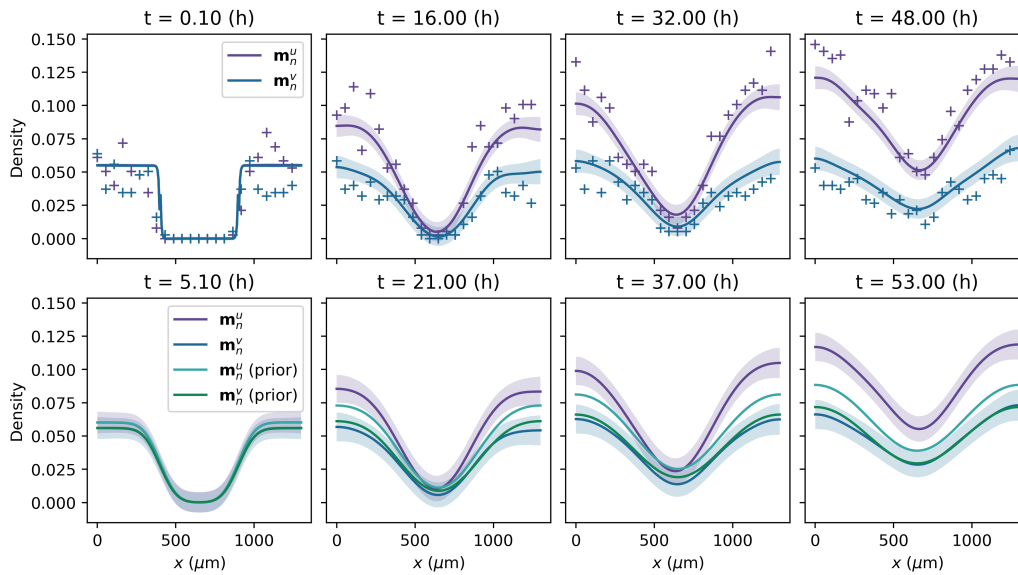
A system of two coupled nonlinear reaction-diffusion equations is considered, which model the densities of two different cell populations, as cells react with one another and diffuse throughout the domain ([Simpson et al., 2020](#)). The combined system state is given by  $\mathbf{w} = (u, v) \in \mathbb{R}^2$  and, with stochastic forcing, the governing equations are

$$\begin{cases} u_t = Du_{xx} - k_u u + 2k_v v(1 - u - v) + \xi_u, \\ v_t = Dv_{xx} + k_u u - k_v v(1 - u - v) + \xi_v, \\ u_x(0, t) = 0, \quad u_x(1300, t) = 0, \\ u := u(x, t), \quad x \in [0, 1300] \mu m, \quad t \in [0, 60] h. \end{cases} \quad (5.6)$$

Coefficients are set to  $D = 700 \mu m^2/h$ ,  $k_u = 0.025$ ,  $k_v = 0.0725$ , and the initial conditions are

$$\begin{cases} u(x, 0) = v(x, 0) = 0, & x \in [400, 900], \\ u(x, 0) = v(x, 0) = 0.055, & \text{otherwise.} \end{cases}$$

In contrast to [Simpson et al. \(2020\)](#), who linearly interpolate the data to give their initial conditions, we assume the fixed piecewise initial conditions as above. Instead of interpolating the data at time  $t = 0$ , we condition on it. This approximately represents the unknown initial conditions. Equation (5.6) is discretised using the  $C^0(\mathcal{D}; \mathbb{R})$  piecewise linear basis functions, for each component, on a regular mesh with 200 cells on the interval  $[0, 1300]$ . Crank-Nicolson is used for the time discretisation, with timestep size  $\Delta_t = 0.1$ .



**Figure 5.3:** Cells example: observed data, posterior means, and 95% posterior credible intervals for the four observed data times (top), and posterior means and 95% posterior credible intervals for five hours after the observation times (bottom).

Hyperparameters of  $\xi_u$  and  $\xi_v$  are set to the same values, which are constant across time. This is to avoid propagating poor estimates through the simulation as, in this case, observations are temporally sparse. The covariance structure of Equation (5.4), is used, and cross-correlations are assumed to be zero,  $\mathbb{E}(\xi_u \xi_v) = 0$ . Hyperparameters are set to  $\rho = 2 \times 10^{-3}$  and  $\ell = 100$ .

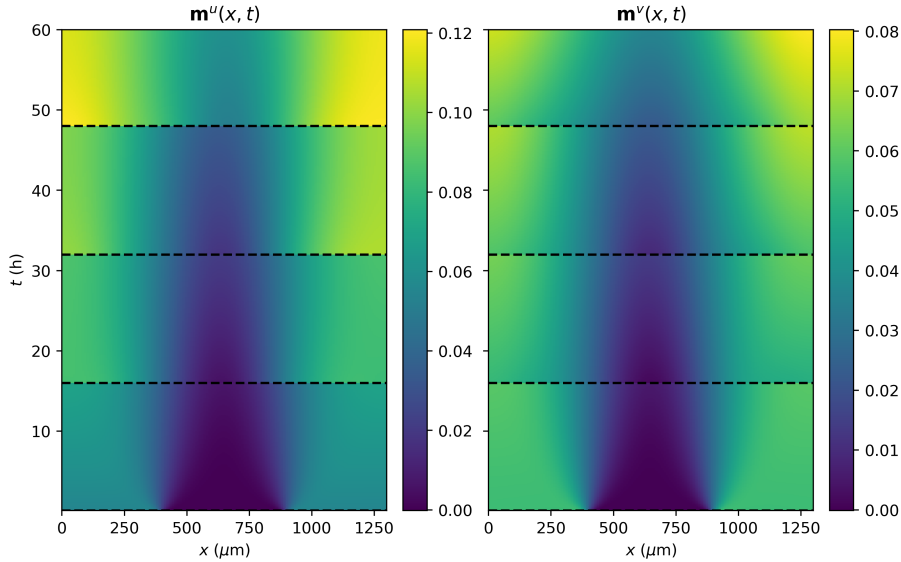
The data consist of observations of the two species at 4 times,  $t_{\text{obs}} = 0, 16, 32, 48$  hours. The concatenated state,  $\mathbf{w}_n = (\mathbf{u}_n^\top, \mathbf{v}_n^\top)^\top$ , gives the data generating process  $\mathbf{y}_n = \mathbf{H}\mathbf{w}_n + \boldsymbol{\eta}_n$  at the observation times, with noise  $\boldsymbol{\eta}_n \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{n_y})$ ,  $\sigma = 0.01$ .

The NL-statFEM posterior  $p(\mathbf{w}_n | \mathbf{y}_{1:n}, \boldsymbol{\theta}, \sigma, \Lambda) = \mathcal{N}(\mathbf{m}_n, \mathbf{L}_n \mathbf{L}_n^\top)$  is computed with the LR-ExKF, using  $k = k' = 32$  modes for both the state covariance, and for the covariances of the  $\mathcal{G}\mathcal{P}$ s  $\xi_u$  and  $\xi_v$ . More than 99% of the variance is retained in the variance truncation at each timestep. The resulting posterior means  $\mathbf{m}_n^u$ ,  $\mathbf{m}_n^v$ , and 95% posterior credible intervals are shown for both the observation times and times five hours after the observation times, in Figure 5.3. For the  $v$  component, there is little discrepancy between the data and the prior assumed model, however for the  $u$  component there is some degree of model mismatch, which conditioning on data can partially account for. The posterior means for each component across the entire space-time grid are also plotted in Figure 5.4, and demonstrate the immediate effects of conditioning on data at the times at which these data are observed.

The low- and full-rank ExKFs are now compared through the posterior means and variances (diagonal of the covariance matrix), in terms of the relative error

$$\frac{\|\mathbf{m}_n^u - \mathbf{m}_{n,\text{LR}}^u\|_2}{\|\mathbf{m}_n^u\|_2}, \quad \frac{\|\text{var}(\mathbf{u}_n) - \text{var}(\mathbf{u}_{n,\text{LR}})\|_2}{\|\text{var}(\mathbf{u}_n)\|_2}.$$

Shown in Figure 5.5a, for a fixed number of modes ( $k = k' = 32$ ), the relative errors are small, at approximately  $10^{-7}$  for the posterior mean, and  $10^{-5}$  for the posterior variance. Sharp increases are



**Figure 5.4:** Cells example: space-time plot of the posterior means for each component. Times at which data are observed are shown with a dashed black line. On the  $u$  component there is an immediate correction from data conditioning, not present in the  $v$  component: the prior model for  $v$  is more accurate.

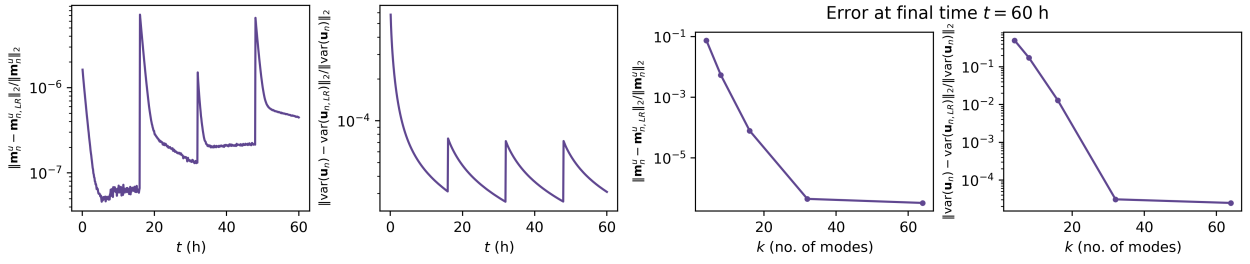
observed at the times at which data is observed, in both the posterior mean and the posterior variance. The relative error at the end time of the simulation is also shown in Figure 5.5 as  $k$  is increased, with  $k' = 32$ . As  $k$  increases the error decreases. Past  $k = k' = 32$  there is a small increase in the accuracy of the filter, with a minor increase from  $k = 32$  to  $k = 64$ . This minor increase is thought to be due to the inclusion of information available from the data, not present in the prior alone. Beyond some effective number of modes, taking larger  $k$  does not yield significant gains in accuracy.

These results also suggest that the low-rank approximation of the prior covariance matrix has a large affect on the accuracy of the estimated posterior covariance matrix, as once  $k = k'$  gains in filter accuracy are small. Recall that in the problem specification the uncertainty is induced via the additive  $\mathcal{GP}$ ,  $\xi$ , and no other sources of uncertainty are considered; if the covariance of  $\xi$  is approximated to a sufficient degree then the low-rank approximation of the filter is accurate.

This can be seen from the ExKF approximation of the prior measure, which for exposition is considered in the case in which  $\mathbf{G}_\theta$  is rank-deficient but is stored fully in memory. This gives an iterative approximation for the covariance matrix (c.f. Chapter 3)

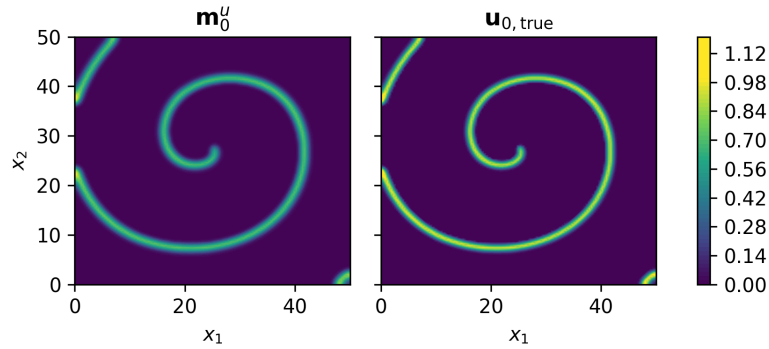
$$\hat{\mathbf{C}}_n = \mathbf{J}_n^{-1}(\mathbf{J}_{n-1}\mathbf{C}_{n-1}\mathbf{J}_{n-1}^\top + \mathbf{G}_\theta)\mathbf{J}_n^{-\top},$$

where  $\mathbf{J}_n$  is the Jacobian matrix of the FEM model with respect to  $\mathbf{u}_n$ , evaluated at  $(\hat{\mathbf{m}}_n, \mathbf{m}_{n-1})$ . Given the recursive nature of computation and the fact that  $\mathbf{C}_0 = \mathbf{0}$  (due to the assumed initial conditions being exact) the prior covariance is a sum of matrix products with  $\mathbf{G}_\theta$  and the Jacobian matrices, at each timestep. The prior covariance will be accurate if the low-rank approximation of  $\mathbf{G}_\theta$  is accurate. Empirically, this is also seen for the posterior covariance. This phenomenon is analysed further in Appendix A.3.1, in which we investigate the errors on the mean and the variance as both  $k'$  and  $\sigma$  are varied.



(a) Relative error between the low-rank and full-rank ExKF, across all times, using  $k = k' = 32$  modes, on ExKF, at the final time  $t = 60$  h, as  $k$  is increased the  $u$  component. (b) Relative error between the low-rank and full-rank ExKF, across all times, using  $k = k' = 32$  modes, on ExKF, at the final time  $t = 60$  h, as  $k$  is increased the  $u$  component. (c) Error at final time  $t = 60$  h, as  $k$  is increased the  $u$  component. (d) Error at final time  $t = 60$  h, as  $k$  is increased the  $u$  component. ( $k' = 32$  for all  $k$ ).

**Figure 5.5:** Cells example: relative error comparisons between the low-rank and full-rank ExKF variants.



**Figure 5.6:** Spiral regime example: initial conditions on the  $u$  component. StatFEM posterior initial condition shown on the left, and the true initial condition shown on the right.

### 5.3.2 Mismatch via initial conditions: spiral regime

We now consider a two-dimensional example of misspecified initial conditions with the Oregonator (Field et al., 1972; Tyson and Fife, 1980), a coupled PDE with state  $\mathbf{w} = (u, v) \in \mathbb{R}^2$ . Adding stochastic forcing on the observed  $v$ -component gives the two-dimensional system

$$\begin{cases} u_t = \frac{1}{\varepsilon} \left( u(1-u) - fv \frac{u-q}{u+q} \right) + D_u \nabla^2 u, \\ v_t = u - v + D_v \nabla^2 v + \xi_v, \\ \nabla u \cdot \mathbf{n} = 0, \quad \nabla v \cdot \mathbf{n} = 0, \quad x \in \partial \mathcal{D}, \\ u := u(\mathbf{x}, t), \quad v := v(\mathbf{x}, t), \quad \mathbf{x} = (x_1, x_2) \in \mathcal{D}, \\ \mathcal{D} = [0, 50] \times [0, 50], \quad t \in [0, 10]. \end{cases}$$

The Oregonator has been well-studied, after being derived as a simplified model for the chemical reaction kinetics of the BZ reaction (Field and Noyes, 1974; Field et al., 1972; Jahnke et al., 1989; Tyson and Fife, 1980). It is a classical example of an activator-inhibitor system, sharing similar behaviour to the Fitzhugh-Nagumo model in certain parameter regimes (Gong and Christini, 2003). We first study the Oregonator in the excitable regime, setting  $f = 2$ ,  $q = 0.002$ , and  $\varepsilon = 0.02$ . Diffusion constants are set to  $D_u = 1$ ,  $D_v = 0.6$ . The  $\mathcal{GP}$   $\xi_v$  has the covariance kernel of Equation (5.4), with  $\theta = (\rho, \ell) = (0.001, 5)$ .

To set the initial conditions, we use the procedure of [Jahnke et al. \(1989\)](#). To induce some degree of mismatch, initial conditions of the NL-statFEM filtering posterior are set by pushing the initial condition of the truth through a heat equation with the diffusion coefficient  $D_u = 1$  (i.e. the Oregonator sans reaction terms), for  $t = 0.1$ . This results in the amplitude of the initial conditions being dampened and blurred (see [Figure 5.6](#)).

The spatial discretisation uses the standard  $C^0(\mathcal{D}; \mathbb{R})$  hat functions, over a regular mesh with  $128 \times 128$  cells (16,641 nodes) so that the total state dimension is 33,282. As previous Crank-Nicolson is used for time discretisation, with  $\Delta_t = 0.001$ .

Data is observed on the slow  $v$  component at 1041 locations inside  $\mathcal{D}_h$ , and is observed every 5 timesteps (the time between observations is 0.005). Data is sparse in space, having approximately 3% of the state dimension observed at observation times. The excitable  $u$  is not observed at all, and is only updated through the observed  $v$  values.

As previous we concatenate the discretised state vector into  $\mathbf{w}_n = (\mathbf{u}_n^\top, \mathbf{v}_n^\top)^\top$  so that the data generating process is  $\mathbf{y}_n = \mathbf{H}\mathbf{w}_n + \boldsymbol{\eta}_n$ ,  $\boldsymbol{\eta}_n \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{n_y})$ . The measurement fidelity is  $\sigma = 0.01$ . We run the LR-ExKF to obtain the posterior  $p(\mathbf{w}_n | \mathbf{y}_{1:n}, \boldsymbol{\theta}, \sigma, \Lambda) = \mathcal{N}(\mathbf{m}_n, \mathbf{L}_n \mathbf{L}_n^\top)$  for all  $n$ , using  $k = 250$  ( $\approx 0.75\%$  of the state dimension) modes to represent the state covariance  $\mathbf{L}_n \mathbf{L}_n^\top$ , and using  $k' = 150$  modes to approximate the covariance matrix of  $\xi_v$ ,  $\mathbf{G}_\theta$ . When running the filter, at least 99% of the variance is retained at each truncation step.

To give a snapshot of results, the data, posterior mean, posterior variance, and leading order modes are each shown in [Figure 5.7](#), at time  $t = 5$ . The dominant region of variance appears to be at the spiral tip, with additional variation observed about the boundary of the spiral on the  $u$  component. This hierarchy is seen in both the colour intensities of the variance plots in [Figure 5.7a](#) and in the columns of  $\mathbf{L}_n$ , in [Figure 5.7b](#); for UQ, there is a hierarchy of variance regions of decreasing importance.

Performance of the filter is verified through computing the relative errors on the means  $\mathbf{m}_n^u$ ,  $\mathbf{m}_n^v$  against the data generating process (DGP):

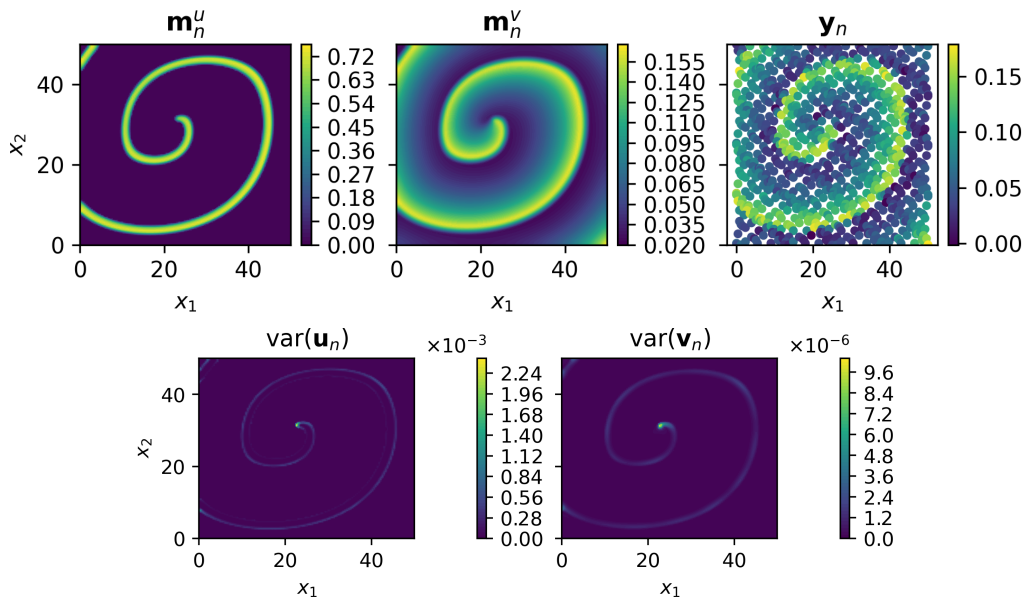
$$\frac{\|\mathbf{m}_n^u - \mathbf{u}_{n,\text{DGP}}\|_2}{\|\mathbf{u}_{n,\text{DGP}}\|_2}, \quad \frac{\|\mathbf{m}_n^v - \mathbf{v}_{n,\text{DGP}}\|_2}{\|\mathbf{v}_{n,\text{DGP}}\|_2}.$$

These are shown in [Figure 5.8a](#). After an initial period of disparity, the NL-statFEM mean  $\mathbf{m}_n$  closely tracks the truth, reaching a stable configuration after this initial warm-up period. Observations can thus correct for misspecification on the unobserved component, with small ( $O(10^{-2})$ ) relative errors in the mean of the NL-statFEM posterior.

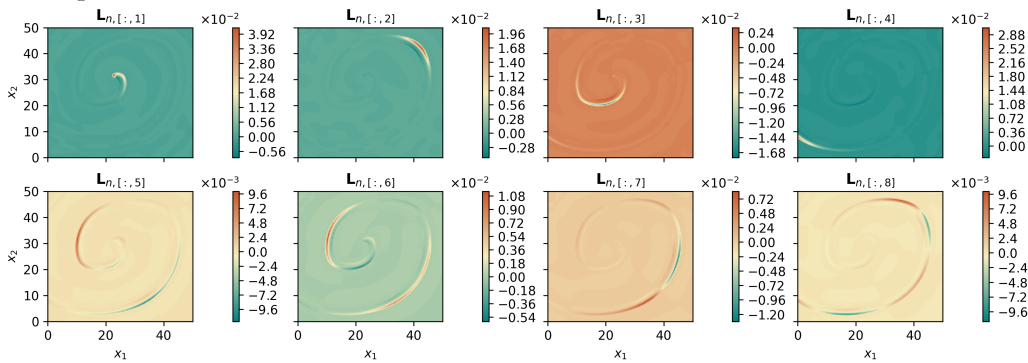
We also compute the effective rank  $D_{\text{eff}}$  ([Gottwald and Majda, 2013](#); [Patil et al., 2001](#)) of the prediction covariance matrix square-root  $\hat{\mathbf{L}}_n$ , using the eigenvalues from the truncation step. We define the effective rank from the eigenvalues (diagonal of  $\boldsymbol{\Sigma}_n$ )  $\varsigma_1 \geq \varsigma_2 \geq \dots \geq \varsigma_k$  of the  $k \times k$  matrix  $\hat{\mathbf{L}}_n^\top \hat{\mathbf{L}}_n$

$$D_{\text{eff}} = \frac{\left(\sum_{i=1}^k \sqrt{\varsigma_i}\right)^2}{\sum_{i=1}^k \varsigma_i},$$

which takes values  $D_{\text{eff}} \in [1, \min\{k, n_u\}]$ . This measures the alignment of the columns of  $\hat{\mathbf{L}}_n$  and can be used to diagnose problems (for example filter collapse), or verify performance (for example, checking that  $k$  and  $k'$  are not over- or under-specified).



(a) Top subplot: Posterior means  $\mathbf{m}_n^u$  (top left),  $\mathbf{m}_n^v$  (top centre), and observed data  $\mathbf{y}_n$  (top right). Bottom subplot: posterior variances  $\text{var}(\mathbf{u}_n)$  (bottom left) and  $\text{var}(\mathbf{v}_n)$  (bottom right) (diagonal of the covariance matrix  $\mathbf{L}_n \mathbf{L}_n^\top$ ). All are reported for time  $t = 5$ .



(b) First eight columns of the posterior covariance square-root  $\mathbf{L}_n$  for the  $u$ -component. These are the leading modes of variation at time  $t = 5$ .

**Figure 5.7:** Spiral regime example: posterior summary plots for time  $t = 5$ .

For this example, this is plotted in Figure 5.8b, and appears to be stable after an initial drop, further suggesting that the filter has reached a stable configuration. The initial drop in the effective rank appears almost immediately after the filter is started, whereas for the relative errors this is at time  $t = 2$ . This suggests that reaching a stable configuration in the covariance is perhaps necessary before the same occurs in the mean. The estimated  $D_{\text{eff}}$  has a mean value of approximately  $D_{\text{eff}} \approx 70$ , indicating that the choice of  $k = 250$  is perhaps excessive for this example.



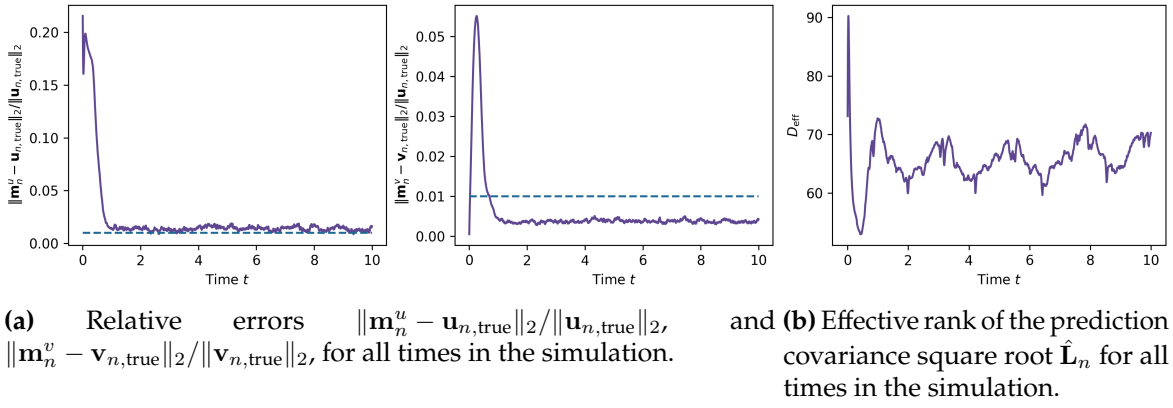


Figure 5.8: Spiral regime example: diagnostic plots for the spiral wave example.

### 5.3.3 Mismatch via initial conditions: oscillatory regime

In the oscillatory regime, the Oregonator has  $f = 0.95$ ,  $\varepsilon = 0.75$ ,  $q = 0.002$ , with diffusion coefficients  $D_u = D_v = 0.001$  (Gong and Christini, 2003). Stochastic forcing is now included on the  $u$  component

$$u_t = \frac{1}{\varepsilon} \left( u(1-u) - fv \frac{u-q}{u+q} \right) + D_u \nabla^2 u + \xi_u,$$

$$v_t = u - v + D_v \nabla^2 v,$$

and the same spatio-temporal domain  $\mathbf{x} \in \mathcal{D} = [0, 50] \times [0, 50]$ ,  $t \in [0, 10]$ , is used. In this instance the initial conditions  $(u_{0,\text{true}}, v_{0,\text{true}})$  are perturbed via setting

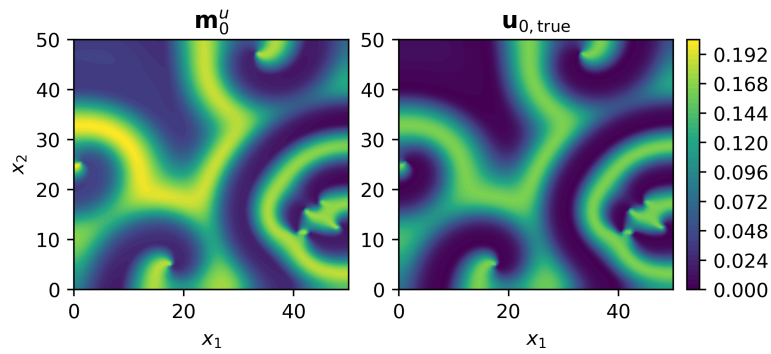
$$u_0 = u_{0,\text{true}} + \varkappa \left( 1 + \sin \left( \frac{\pi x_1}{50} + \zeta_1 \right) \sin \left( \frac{\pi x_2}{50} + \zeta_2 \right) \right), \quad v_0 = v_{0,\text{true}},$$

where  $(\zeta_1, \zeta_2) \sim \mathcal{N}(0, 1^2)$ . The amplitude  $\varkappa$  is set to 0.02 in our case studies, and for the truth  $(u_{0,\text{true}}, v_{0,\text{true}})$ , these are set from running a pilot simulation for 100,000 timesteps, where the pilot simulation has randomly generated initial conditions  $u, v \sim \text{Unif}(0, 0.15)$ . The upper and lower bounds on the uniform distribution are determined from the attractor of the corresponding Oregonator ODE. Both the perturbed and true initial conditions are shown in Figure 5.9. This initial condition is included in the accompanying GitHub repository, available from <https://github.com/connor-duffin/low-rank-statfem>.

As previous the spatial discretisation uses  $C^0(\mathcal{D}; \mathbb{R})$  basis functions, on a mesh with  $256 \times 256$  cells, with Crank-Nicolson for the time discretisation. The timestep size is  $\Delta_t = 0.01$ .

The  $u$ -component is observed at each timestep up to  $t = 10$ , at 512 observation locations. As usual,  $\mathbf{y}_n = \mathbf{H}\mathbf{w}_n + \boldsymbol{\eta}_n$ ,  $\boldsymbol{\eta}_n \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{n_y})$ , with  $\sigma = 0.01$ . The posterior  $p(\mathbf{w}_n | \mathbf{y}_{1:n}, \boldsymbol{\theta}, \sigma, \Lambda) = \mathcal{N}(\mathbf{m}_n, \mathbf{L}_n \mathbf{L}_n^\top)$  is computed using the LR-ExKF with  $k = 128$  and  $k' = 64$ . The initial leading-order eigendecomposition of the  $\mathcal{GP}$  covariance matrix  $\mathbf{G}_\theta$  is done using Lanczos iterations (Saad, 2003) as implemented in Scipy (Virtanen et al., 2020), with the matrix-vector products done on the GPU via KeOps (Charlier et al., 2021).

We compare three filters, each with the  $\mathcal{GP}$  covariance kernel of Equation (5.4), which set  $\rho \in \{10^{-2}, 10^{-3}, 2 \times 10^{-4}\}$ , with  $\ell = 10$  and  $\sigma = 10^{-2}$  for each. In each filter more than 99% of the



**Figure 5.9:** Oscillatory regime example: initial conditions for  $u$  component. StatFEM posterior initial condition shown on the left (same for each filter configuration) and the true initial conditions shown on the right.

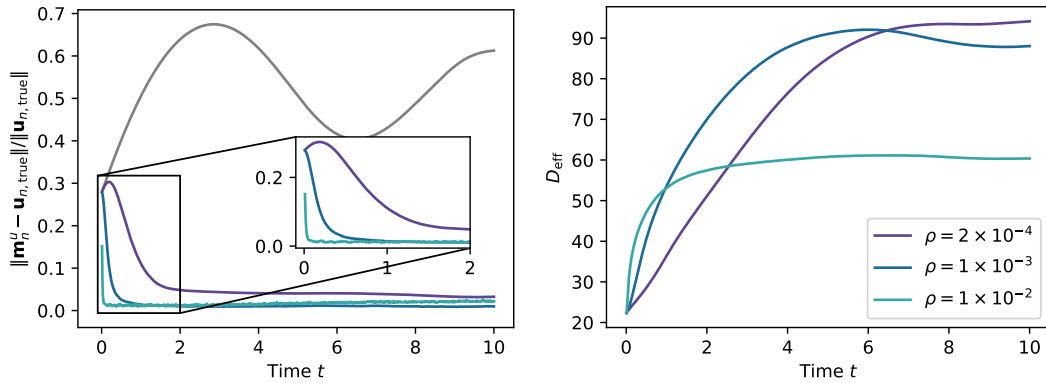
variance is retained at each truncation step. For the remainder of this discussion these are referred to as the large- $\rho$  filter, the moderate- $\rho$  filter, and the small- $\rho$  filter, respectively.

Relative errors for the NL-statFEM posterior means are shown alongside the NL-statFEM prior mean in Figure 5.10a. The small- $\rho$  filter shows the slowest misspecification correction compared to the others; the filter is more certain of the model predictions and takes longer than the others to reach a stable filtering configuration. It also results in the largest relative error by the end of the simulation. The moderate  $\rho$ -filter, however, despite slower initial misspecification correction, results in the lowest relative error by  $t = 10$ . The large- $\rho$  filter displays the most rapid convergence, yet results in relative errors increasing after the initial correction. This is thought to be due to  $\rho$  being equal to the noise  $\rho = \sigma = 0.01$ , which results in assimilation of spurious noise perturbations. These results confirm the role of the  $\rho$  hyperparameter in controlling the belief in (or uncertainty of) the underlying dynamical model. Higher variance implies less certainty and more rapid corrections for model misspecification.

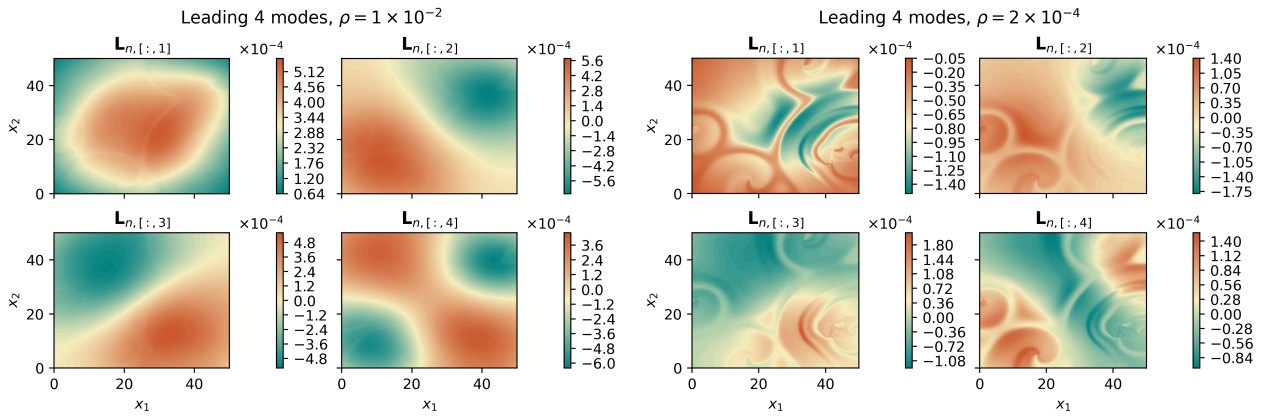
This trend of slower corrections, for more certain models, is also seen in the effective rank  $D_{\text{eff}}$  of the prediction covariance matrix  $\hat{\mathbf{L}}_n \hat{\mathbf{L}}_n^\top$ , plotted for each of the filters in Figure 5.10b. In terms of  $D_{\text{eff}}$ , the large- $\rho$  filter has the most rapid convergence to a stable filtering configuration, with the small- $\rho$  filter the slowest. Once in this stable filtering configuration the small- $\rho$  filter has a larger effective rank compared to the large- $\rho$  filter, and it is posited that a higher effective rank implies a more complex covariance structure. To check this we plot the leading modes of the covariance matrix at the end time  $t = 10$ , in Figure 5.10c. This is indeed the case, and we see that the modes for the small- $\rho$  filter display more localised structures pertaining to the dynamics of the model, when compared with those of the large- $\rho$  filter, which appear more similar to the eigenfunctions of the  $\mathcal{GP}$  covariance kernel  $k_\theta(\cdot, \cdot)$ .

This also merits another interpretation of the variance hyperparameter  $\rho$ , that of controlling the weight of the *a priori* misspecification covariance matrix  $\mathbf{G}_\theta$  in comparison to the (tangent linear) dynamical evolution of the previous timestep covariance  $\mathbf{C}_{n-1} = \mathbf{L}_{n-1} \mathbf{L}_{n-1}^\top$ . In this case, due to the complex spatial oscillations seen in the dynamical model, by decreasing the weight of the simpler  $\mathcal{GP}$  covariance, this results in the more complex dynamical interactions being present in the posterior covariance, resulting in an increased  $D_{\text{eff}}$ .





(a) Relative  $l^2$  errors for the large- $\rho$ , moderate- $\rho$ , small- $\rho$  filters, and the prior model. (b) Effective rank of the prediction covariance square root  $\hat{\mathbf{L}}_n$  for the large- $\rho$ , moderate- $\rho$ , small- $\rho$  filters, and the prior model.



(c) Leading four covariance modes (first four columns of the posterior covariance matrix square root  $\mathbf{L}_n$ ) for the large- $\rho$  filter (left) and the small- $\rho$  filter (right). Shown are the modes for the  $u$ -component.

**Figure 5.10:** Oscillatory regime example: relative  $l^2$  errors (top left), effective rank  $D_{\text{eff}}$  (top right), and leading posterior covariance modes (bottom) at time  $t = 10$ , oscillatory mismatched initial condition.

# Chapter 6

## Discussion

*This thesis was about a procedure to synthesise physical models with data, focussing on the specific case of nonlinear PDEs as discretised with finite element methods. Significant research has been dedicated to inversion problems, which aim to recover model parameters from data, and to data assimilation problems, which aim to sequentially update a time-evolving model state as data becomes available. We built upon the previous work of Girolami et al. (2021) and introduced the NL-statFEM (Duffin et al., 2021): a coherent Bayesian statistical approach to deal with model misspecification through the combination of techniques from both inversion and data assimilation. Through the admission of uncertainty inside of the governing equation — as in inversion — model misspecification is recognised and is sequentially corrected for through updating the resultant stochastic physical model with data using classical data assimilation algorithms (the extended and ensemble Kalman filters). Numerical discretisation of the governing equations with finite element methods enabled the application of the method to a variety of systems in one and two dimensions. Scalability, ensured through making a low-rank approximation to the posterior covariance matrix (Duffin et al., 2022), allows for widespread application of NL-statFEM across engineering and the physical sciences. In this chapter, the main contributions of this thesis are summarised in Section 6.1, and extensions to future work are covered in Section 6.2.*

### 6.1 Conclusions

Partial differential equations are one of the most common ways to describe physical laws or approximations thereof. Solutions to these equations are often realised, in practice, via finite element discretisation, due to their nonlinearity and/or the presence of complex geometries of the solution domains. However, due to model imperfections, in the comparison of model outputs to observed data, mismatch is often seen between the posited physical model and observations. This suggests that the model is misspecified.

A common approach to reconcile model misspecification is to estimate the PDE parameters such that they minimise the error between observed data and the model output. This solves the *inverse problem*. A Bayesian approach to the inverse problem may also be taken. In this, a prior distribution is placed over the model parameters, which is subsequently updated with data based upon a particular data generating process. This gives the posterior distribution. Solving the inverse problem deals with model misspecification due to possibly incorrectly specified model parameters, but does not allow for the correction of model misspecification from non-parameter sources. An alternative approach

— used in the field of data assimilation — is to instead treat the model as a black box object with some extrusive additive noise process, and sequentially update the model solutions as data becomes available. This corrects for general model error (i.e., not just from parameters) in a time-evolving fashion, but sacrifices interpretability due to uncertainties being extraneous to the model.

In [Girolami et al. \(2021\)](#) the *statistical finite element method (statFEM)* is presented, which corrects for model error via observations in an interpretable, statistically coherent fashion. As in Bayesian inversion, all model parameters are modelled stochastically via  $\mathcal{GP}$ s. However, instead of being estimated, these parameters are marginalised over to give an induced prior distribution over the FEM discretised PDE solution. This prior is then updated with data to give the posterior distribution over the FEM coefficients. Due to linearity of the underlying assumed elliptic PDE, the posterior distribution derived is also Gaussian and can be arrived at via the standard Gaussian update.

FEM solutions are directly updated with data to correct for arbitrarily-sourced model mismatch, similar to how this is done in the field of data assimilation. The posterior quantifies all assumed uncertainties associated with both the model and the data generating process. However, in [Girolami et al. \(2021\)](#), statFEM did not make use of the temporal structure as in common data assimilation algorithms, being limited to linear, static PDEs in low-dimensional settings. In this thesis we address these shortcomings, providing a computationally scalable approach to statFEM in nonlinear, time-dependent problems.

Our first contribution was the extension of statFEM to nonlinear, time-evolving systems. Denoted *NL-statFEM*, this methodology broadens the applicability of the original work to a much wider amount of problems. We leveraged the same model uncertainty structure as in the original statFEM, but, using methods from the data assimilation literature, derived algorithms to sequentially compute the posterior distribution of interest in an online manner. Model misspecification uncertainty was given by a stochastic forcing function, modelled with a  $\mathcal{GP}$ . However to sequentially update models with data, our NL-statFEM made use of the extended and ensemble Kalman filters, which compute Gaussian approximations to the filtering posterior. These approximate Gaussian distributions describe the probability of the finite element discretised PDE solution, conditioned on the data up to and including the current point in time. Hyperparameter estimation also proceeds for the  $\mathcal{GP}$ , via the ExKF log-marginal likelihood. This log-marginal likelihood also quantifies uncertainty associated with the model forecasts. The derivation of this methodology was provided in Chapter 3, being based upon the previously published work in [Duffin et al. \(2021\)](#).

In Chapter 4, our NL-statFEM methodology was demonstrated on three canonical nonlinear PDEs: Burgers Equation, the Kuramoto-Sivashinsky equation, and the Korteweg-de Vries equation. These results showed that the method can approximate the data generating process to give an interpretable posterior distribution, which can correct for model mismatch in regimes of gradually increasing model misspecification. The method was also applied to the experimental data collected in [Horn et al. \(2002\)](#), and provided a sensible physics-informed interpolator, with uncertainty quantification, in a spatially sparse observational regime. Hyperparameter estimation was checked, empirically verifying unbiasedness the ExKF log-marginal-likelihood estimates. Code to replicate these results is contained in an accompanying repository, available at <https://github.com/connor-duffin/statkdv-paper>.

Our second contribution was the derivation of a computationally scalable NL-statFEM, by making

a low-rank approximation to the posterior covariance matrix through representing this covariance by its leading-order modes. Combining this with the previously developed ExKF methodology gives a low-rank extended Kalman filter (LR-ExKF) algorithm that is highly scalable, whilst also providing a sensible and interpretable uncertainty quantification. This was presented in Chapter 5, where we again provide extensive numerical evidence of the efficacy of this approach. Results in 1D, with experimental data collected in [Simpson et al. \(2020\)](#), showed that the LR-ExKF can approximate the full-rank alternative with small relative errors in both the mean and the variance. These results also confirm the efficacy of NL-statFEM in temporally sparse observation regimes. Additional results with the 2D Oregonator system demonstrate scalability, with the NL-statFEM correcting for model misspecification in the initial conditions through observing a single component of the coupled system. These results are robust under different model parameter regimes and under increasing state dimension. We also showed that the effective rank of the covariance matrix can verify the number of modes taken in the low-rank approximation, providing a visual check of filter convergence. Code to reproduce these results is available at <https://github.com/connor-duffin/low-rank-statfem>.

## 6.2 Future work

This thesis sets the foundation for future studies of embedding data within nonlinear, time-evolving FEM models. However we deliberately took the approach of discretising the problem first, then conducting inference. This trades mathematical complexity in the problem specification — avoiding having to derive posterior measures over the appropriate function space of PDE solutions — for potential scalability. Scalability, in this sense, means that the problems are well-posed on the infinite-dimensional space and are thus robust to mesh refinement (see, e.g., [Cotter et al., 2013](#)). An alternate approach would be to derive filtering methods on the appropriate function space, which sequentially update the filtering measure as data becomes available. Analogues of the approximate Gaussian filters could be derived and theoretical guarantees may be given, showing, for example, stability and convergence. The RKHS setting of [Papandreou et al. \(2021\)](#) and [Karvonen et al. \(2022\)](#) could be leveraged in this work. This infinite-dimensional formulation would also have to acknowledge the connection to stochastic PDEs (see, e.g., [Hairer, 2009](#)), which was outside the scope of this thesis.

Various theoretical guarantees could also be proven for NL-statFEM. The first is the convergence and stability of the ExKF/EnKF for generic nonlinear PDEs, under appropriate assumptions on model nonlinearity (for example, assuming that the nonlinearity arises as a polynomial of the PDE state variable). Further analysis should also be conducted on the LR-ExKF in both the linear and nonlinear cases, analytically studying convergence to both the full-rank approximate posterior (if applicable) and the true filtering posterior, as the number of modes is increased.

Despite labelling our NL-statFEM as a statistical finite element method, the underlying framework is discretisation-agnostic and is amenable to any appropriate PDE discretisation method. Extensions, therefore, into spectral methods (e.g., spectral Galerkin methods and tau methods) ([Boyd, 2001](#)) and finite volume methods ([LeVeque, 2002](#)) would be of interest. When extending into these domains, studying alternate time-integrators would also be of use, going beyond the first-order schemes considered in this thesis. Discretisation, in space and/or time, may also be handled via probabilistic

numerical methods (see, e.g., [Hennig et al., 2015](#)), giving a fully probabilistic, statistically coherent approach to assimilating data into PDE models. The associated posterior measure, in this case, would quantify uncertainty associated the PDE solution given the data and a chosen level of mesh-refinement.

Estimation of the PDE parameters from the marginal likelihood is not considered in this work and would be a useful extension that allows for a completely data-driven methodology. MAP estimation, the same as for the  $\mathcal{GP}$  hyperparameters, may be used. Alternatively, a fully Bayesian approach may be taken for both the  $\mathcal{GP}$  hyperparameters ([Filippone and Girolami, 2014](#)) and the PDE parameters, with the posterior distribution of these parameters estimated.

The further application of NL-statFEM to high-dimensional, highly nonlinear systems is, finally, of course of interest. Augmentations of the presented filtering algorithms may be required to deal with significantly nonlinear systems with sparse observation regimes ([Verlaan and Heemink, 2001](#)). Specifically for the methods considered in this thesis, the application of a NL-statFEM-type method to higher-dimensional fluids problems governed by conservation laws for experimental and in-situ measurements is of particular import, being an important setting in which temporally evolving nonlinear PDEs are applied.

# Appendix A

## Appendices

This appendix contains various additional results that supplement the main text. First, we present a proof that the linearisation of the weak form returns a Gateaux derivative which coincides with the Jacobian matrix. The next section presents the same simulation study results as in Chapter 4 but with the NL-statFEM posterior distributions computed with the EnKF instead of the ExKF. The final section presents three additional simulation studies using the LR-ExKF of Chapter 5. The first demonstrates the effect of the number of prior covariance modes on the covariance approximation, using the cells example studied previously. The next two use the Oregonator. These examples verify the parameter estimation methodology and give an example of filter divergence, respectively.

### A.1 Linearisation results

When solving nonlinear FEM systems, at each timestep the system of equations requires solving, for which we use Newton iterations.<sup>1</sup> This derivative is also required for NL-statFEM, when we compute the ExKF posterior covariance. In the notation of Chapter 3 this is the derivative of the system with respect to  $\mathbf{u}_n$ ,  $\frac{\partial \mathcal{M}(\mathbf{u}_n, \mathbf{u}_{n-1})}{\partial \mathbf{u}_n}$ , which is equivalent to taking a Gateaux derivative in the direction of the test function. We demonstrate this equivalency in the following proposition.

**Proposition 2.** Consider a general nonlinear weak form,  $\mathcal{W} : V_h \times V_h \rightarrow \mathbb{R}$ , which we will write as  $\mathcal{W}(u_h; \phi_j)$ ,  $j = 1, \dots, n_u$ , assumed to be of the form  $\mathcal{W}(u_h; \phi_j) = \langle L_\Lambda u_h, \phi_j \rangle + \langle F_\Lambda(u_h), \phi_j \rangle$ . Then for  $W : \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_u}$ , in which  $W(\mathbf{u})_j = \mathcal{W}(u_h; \phi_j)$ , the Gateaux derivative  $DW(u_h; \phi_i, \phi_j)$ , in direction  $\phi_i$ , defines the  $(j, i)$  entry of the Jacobian matrix  $\mathbf{J}(\mathbf{u}) = \frac{\partial W(\mathbf{u})}{\partial \mathbf{u}}$ .

*Proof.* For a single  $j$ , in some direction of  $v_h$ , the Gateaux derivative of the nonlinear weak form is given by:

$$\begin{aligned} DW(u_h; v_h, \phi_j) &= \frac{d}{d\epsilon} [\langle L_\Lambda(u_h + \epsilon v_h), \phi_j \rangle + \langle F_\Lambda(u_h + \epsilon v_h), \phi_j \rangle + \langle \xi, \phi_j \rangle]_{\epsilon=0} \\ &= [\langle L_\Lambda v_h, \phi_j \rangle + \langle \partial F_\Lambda(u_h + \epsilon v_h)v_h, \phi_j \rangle]_{\epsilon=0} \\ &= \langle L_\Lambda v_h, \phi_j \rangle + \langle \partial F_\Lambda(u_h)v_h, \phi_j \rangle \end{aligned}$$

---

<sup>1</sup>The nonlinear solve can include a line search component, which chooses a locally optimal stepsize according to some criteria, such as the Armijo condition (Nocedal and Wright, 2006). We have found that in practice adding a line search increases stability.

Therefore  $DW(u_h; \phi_i, \phi_j) = \langle L_\Lambda \phi_i, \phi_j \rangle + \langle \partial F_\Lambda(u_h) \phi_i, \phi_j \rangle$ . Hence making the substitution  $v_h = u_h = \sum_{i=1}^{n_u} u_i \phi_i(x)$  we get

$$DW(u_h; u_h, \phi_j) = \sum_{i=1}^{n_u} u_i \langle L_\Lambda \phi_i, \phi_j \rangle + \sum_{i=1}^{n_u} u_i \langle \partial F_\Lambda(u_h) \phi_i, \phi_j \rangle, \quad j = 1, \dots, n_u.$$

On the other hand, the  $(j, i)$  entry of the Jacobian matrix is given by

$$\begin{aligned} \frac{\partial W(\mathbf{u})_j}{\partial u_i} &= \frac{\partial}{\partial u_i} [\langle L_\Lambda u_h, \phi_j \rangle + \langle F_\Lambda(u_h), \phi_j \rangle] \\ &= \langle L_\Lambda \phi_i, \phi_j \rangle + \langle \partial F_\Lambda(u_h) \phi_i, \phi_j \rangle. \end{aligned}$$

Implying that  $(\mathbf{J}(\mathbf{u})\mathbf{u})_j = \sum_{i=1}^{n_u} u_i \langle L_\Lambda \phi_i, \phi_j \rangle + \sum_{i=1}^{n_u} u_i \langle \partial F_\Lambda(u_h) \phi_i, \phi_j \rangle$ , which is equal to  $DW(u_h; u_h, \phi_j)$ .  $\square$

## A.2 Ensemble Kalman filter: additional examples

In this section we reproduce the results given in Chapter 4, computing the NL-statFEM posterior distribution using the EnKF instead of the ExKF. In these examples, unless otherwise mentioned, all discretisation parameters (e.g., number of FEM cells  $n_c$ , timestep size  $\Delta_t$ , and so on) and PDE parameters  $\Lambda$  are set to the same values as given in the respective sections in Chapter 4. In general, results are similar to those given by the ExKF, with some notable differences arising due to the Monte Carlo approximations given in the EnKF, and small ensemble sizes  $N_{\text{ens}}$ .

### A.2.1 Burgers equation

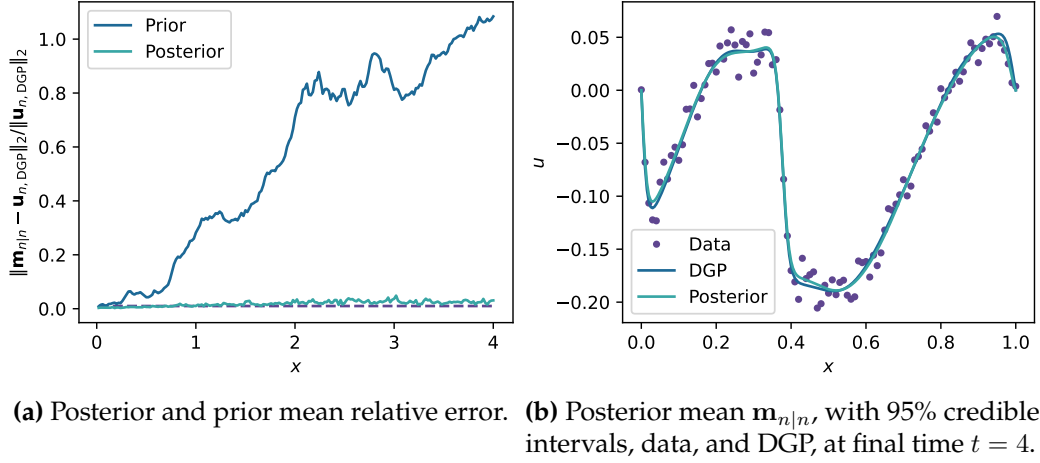
First, we replicate the simulation study of Sections 4.1.2 and 4.1.3, replacing the NL-statFEM posterior ExKF computations with the EnKF. In each case we use  $N_{\text{ens}} = 32$  ensemble members for a state dimension of  $n_u = 200$ . We first redo the hyperparameter identification experiment, for which the relative errors and the estimated posterior mean profile are shown in Figure A.1. The relative errors appear less smooth than in the ExKF case (c.f. Figure 4.2a). Hyperparameter estimates are shown in Figure A.2 and appear accurate to a similar degree as in the ExKF case.

The second simulation study, which investigates the specific case of model misspecification being introduced via a misspecified viscosity coefficient, replicates the results obtained in Section 4.1.3 with the EnKF. The various verification quantities are shown in Figure A.3, and the estimated hyperparameters are shown in Figure A.4. Results are comparable for both the ExKF and the EnKF. Posterior means and covariances for select timepoints are shown in Figure A.5. Due to the ensemble approximation of the covariance matrix, spurious correlations arise in the posterior covariance due to rank-deficiency, which are not present in the ExKF covariance matrix.

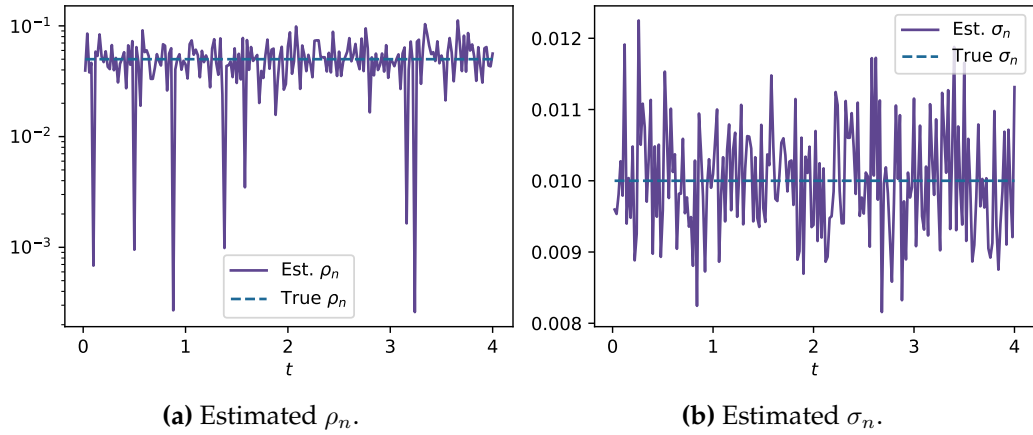
### A.2.2 Kuramoto-Sivashinsky equation

In this section the results of Section 4.2.2 are replicated. The NL-statFEM posterior is now computed with the EnKF, with  $N_{\text{ens}} = 128$  for a state dimension of  $n_u = 400$ . Recall that in this case, we deal with





**Figure A.1:** Burgers example I results (EnKF).



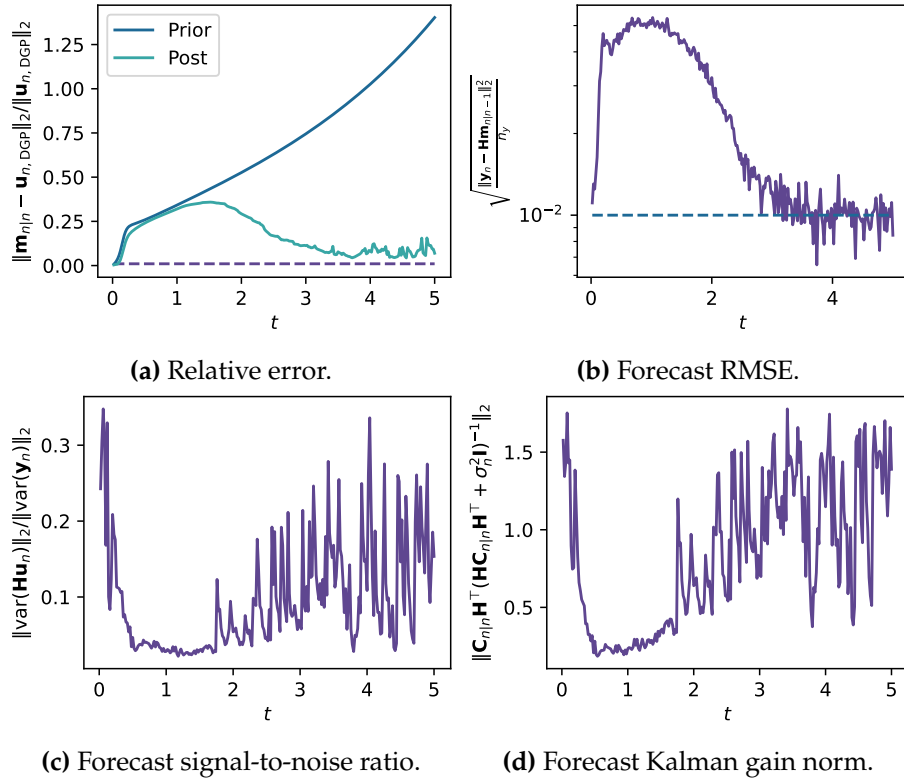
**Figure A.2:** Burgers example I results (EnKF): hyperparameter estimation.

another example of a misspecified viscosity, as in the previous Burgers equation example. The various verification quantities are shown in Figure A.6. Relative errors are slightly larger than when using the ExKF and the estimated  $\sigma_n$  increases over time. This increase is thought to be due to the rank-deficient nature of the ensemble; there is less uncertainty inside of the ensemble, so this residual uncertainty ends up being modelled through the additive noise. Due to the noise overestimation, there is less information being included in the filter as some dynamics are mistaken for noise, resulting in higher relative errors. The Kalman gain appears slightly lower, suggesting that less information is gleaned from observations than with the ExKF. The prior and posterior means are shown in Figure A.7 and appear visually akin to those computed via the ExKF.

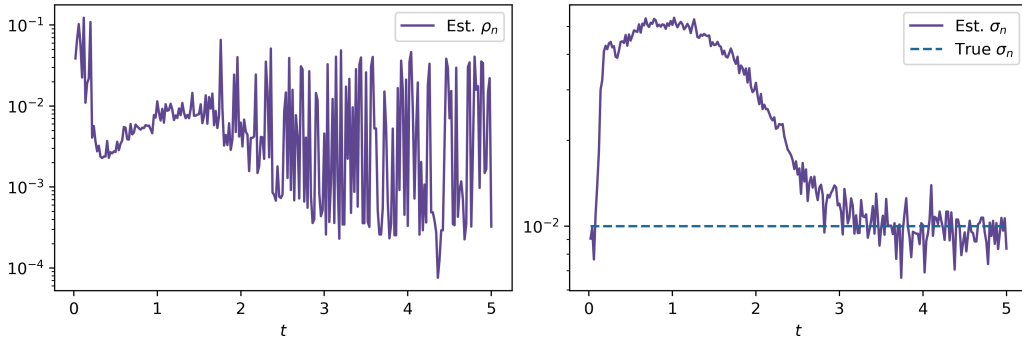
### A.2.3 Korteweg-de Vries equation

The results of Sections 4.3 are now recomputed with the EnKF. First the cubic example of Section 4.3.2 is repeated using the EnKF, with  $N_{\text{ens}} = 400$ . Results are shown in Figure A.8 and are similar to those obtained with the ExKF. Posterior mean profiles are visually equivalent to those computed with the ExKF, and parameter estimates are also comparable. Relative errors appear less smooth for the EnKF results in comparison to the ExKF results (c.f. Figure 4.9a), thought to be due to Monte Carlo





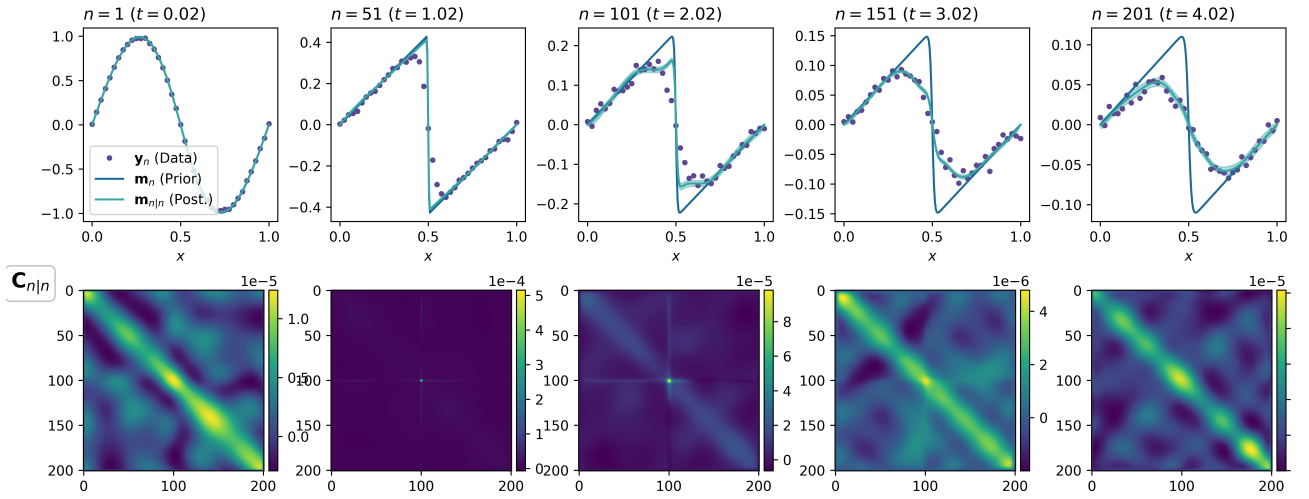
**Figure A.3:** Burgers example II results (EnKF): verification quantities.



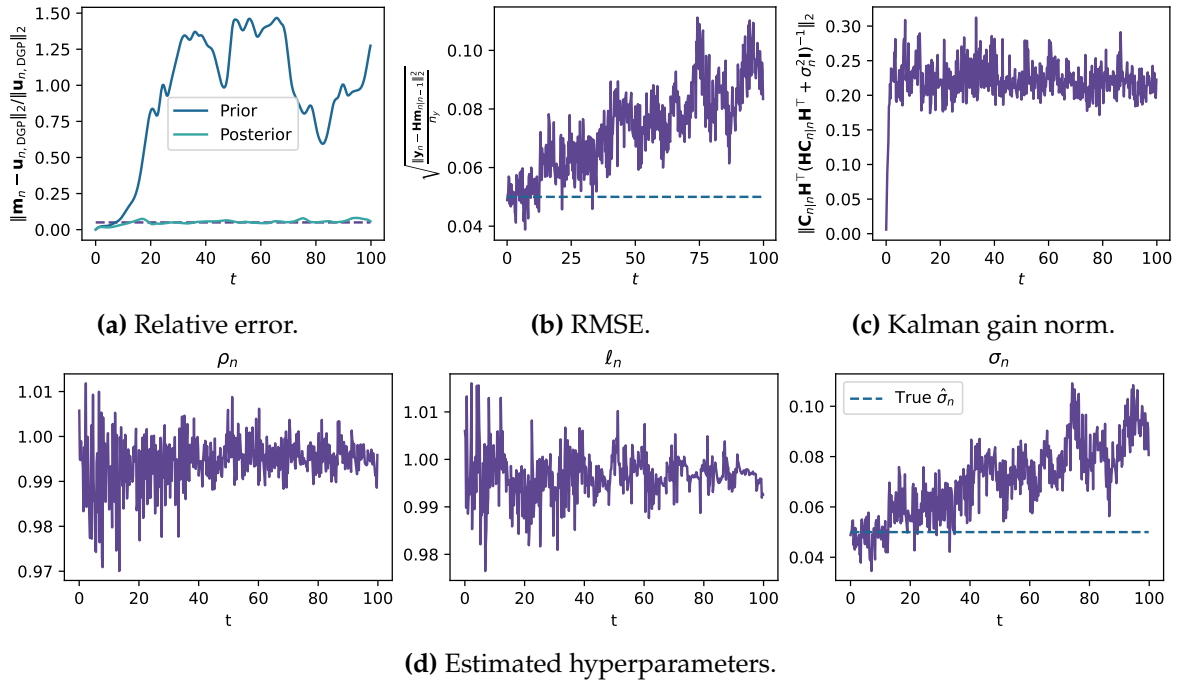
**Figure A.4:** Burgers example II results (EnKF): estimated hyperparameters.

approximation.

Using the EnKF, the results of Section 4.3.3 are shown in Figure A.9. Posterior wave profiles (Figure A.9a) appear less smooth than for the ExKF (c.f. Figure 4.15) with larger uncertainty bounds. This decreased smoothness is also seen in the space-time view of the posterior mean, shown in Figure A.9c. The ExKF hyperparameter estimates are more varied, with  $\sigma_n = 10^{-3}$  estimated more often in comparison to the EnKF (c.f. Figure A.9b).



**Figure A.5:** Burgers example II results (EnKF). Top row: posterior means, 95% credible intervals, data. Bottom row: covariance matrices  $\mathbf{C}_{n|n}$ .

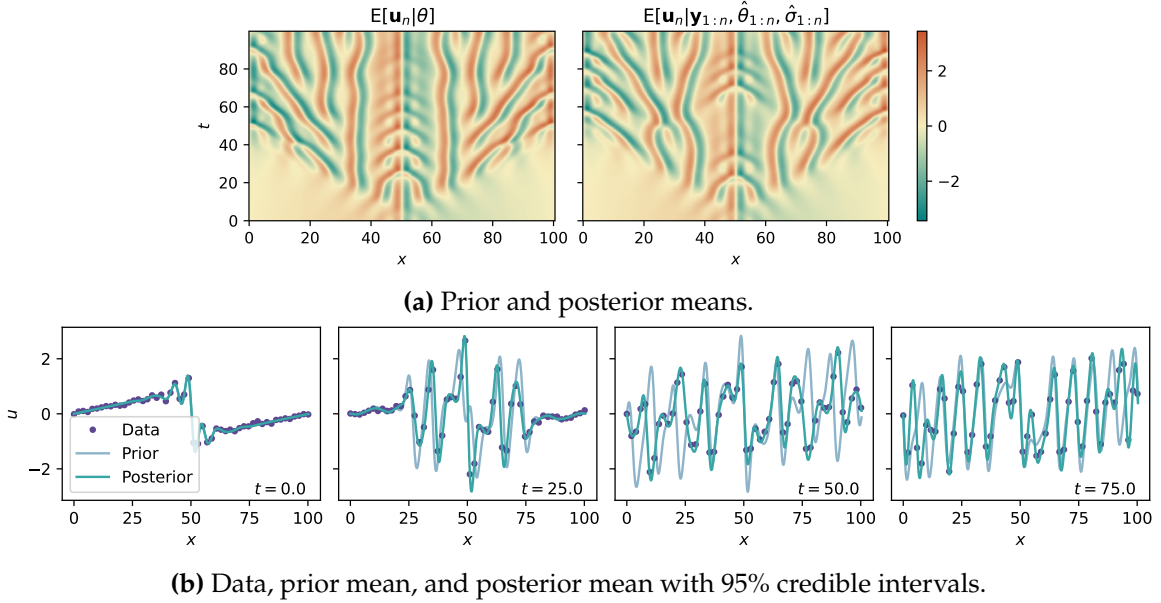


**Figure A.6:** KS example (EnKF): verification quantities.

## A.3 Low-rank filtering: additional examples

### A.3.1 Influence of prior modes in the cell example

As discussed in the main text (Section 5.3.1), in the cell example, the prior modes control the accuracy of the UQ given by the LR-ExKF. Here we investigate this further, through checking the errors in the posterior mean and variance as we vary the prior modes  $k'$  and the observational noise  $\sigma$ . For each  $k'$  we fix the filter modes  $k$  to  $k = k' + 16$ , to allow for sufficient extra modes to capture information present in the data, not included in the prior modes. Note that in this instance the variation of  $\sigma$  is only to investigate the computational performance of the filtering method; the known noise value



**Figure A.7:** KS example (EnKF): posterior mean results.

of  $\sigma = 10^{-2}$  should be used (as it is in the main text) when attempting to infer the system state from measurements.

The results of running the filters for  $k' \in \{16, 32, 48, 64\}$ ,  $\sigma \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$  are shown in Figure A.10. These suggest that as the noise decreases (i.e., the data becomes more informative) then more modes are required to accurately capture the variance. Despite the additional “overhead modes” on  $k = k' + 16$ , gains are seen when increasing  $k'$ , confirming that the number of prior modes have a large affect on the performance on the filter. Errors in the posterior mean appear to increase more rapidly with the decreasing  $\sigma$ , in comparison to the variance. It is posited that this discrepancy is seen due to the posterior mean update using the full nonlinear dynamics when completing the prediction step — the covariance propagation, in comparison, uses the tangent linear dynamics.

### A.3.2 Verification of parameter estimation

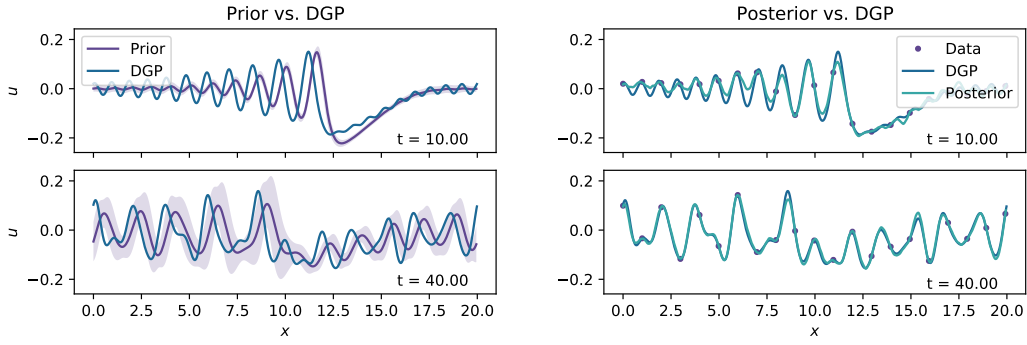
For this example, the Oregonator equations in the oscillatory regime are used (recall  $f = 0.95$ ,  $\varepsilon = 0.75$ ,  $q = 0.002$ , and the diffusion coefficients  $D_u = D_v = 0.001$  (Gong and Christini, 2003)). Data is generated according to a stochastic Oregonator with stochastic forcing on the  $u$  component

$$u_t = \frac{1}{\varepsilon} \left( u(1-u) - fv \frac{u-q}{u+q} \right) + D_u \nabla^2 u + \xi_u,$$

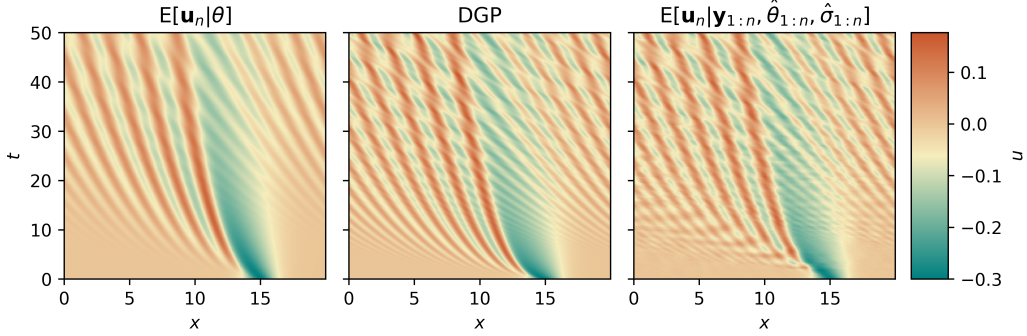
$$v_t = u - v + D_v \nabla^2 v,$$

and we verify the hyperparameter estimation routine of Chapter 3.

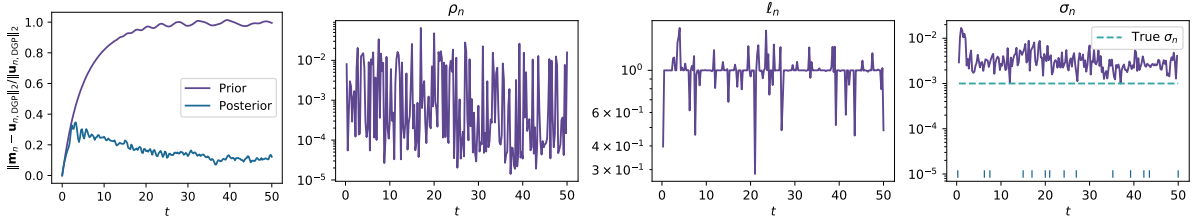
Initial conditions are the same as for the previous oscillatory example, and the same space-time domain is used as previous:  $\mathbf{x} \in \Omega = [0, 50] \times [0, 50]$ ,  $t \in [0, 10]$ . Timesteps use an implicit-explicit



(a) Left: Data generating process, and prior mean with 95% credible intervals. Right: data, data generating process, posterior means, and 95% credible intervals.



(b) Means across space-time grid.



(c) Relative error. (d) Estimated hyperparameters (points at which the noise is not identified are shown as a vertical line on the horizontal axis).

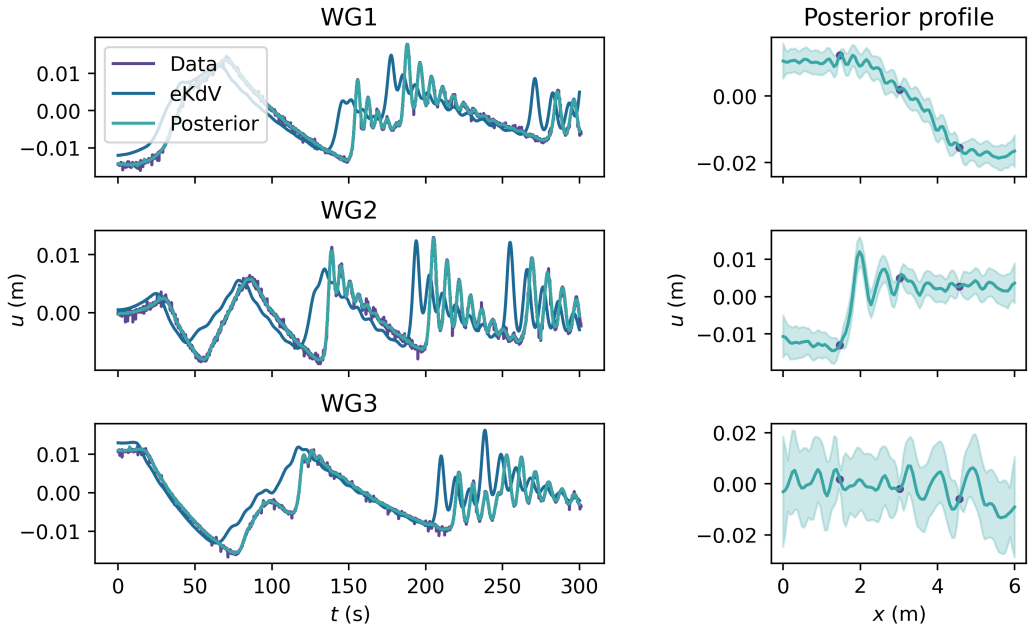
**Figure A.8:** KdV cubic example: results computed with the EnKF.

(IMEX) scheme (Ascher et al., 1995) with forward and backward Euler:

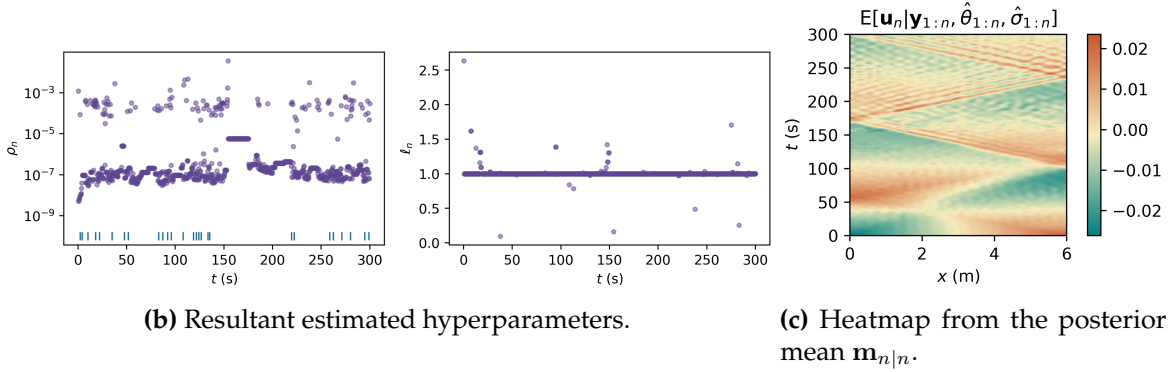
$$\mathbf{M}(\mathbf{w}_{n+1} - \mathbf{w}_n) + \Delta_t \kappa \mathbf{A} \mathbf{w}_{n+1} = \Delta_t \tilde{\mathbf{r}}(\mathbf{w}_n) + \mathbf{e}_n,$$

Recalling the concatenated state  $\mathbf{w}_n = (\mathbf{u}_n^\top, \mathbf{v}_n^\top)^\top$ . Timestep size is set to  $\Delta_t = 10^{-2}$  or  $\Delta_t = 10^{-4}$ , and we investigate the accuracy of the estimated hyperparameters for each (each filter is run for 1000 timesteps). Note that this does not yield the same integration window, one being two orders of magnitude smaller than the other.

Data, of the  $u$  component, is observed at 512 locations at each timestep (locations shown in Figure A.11). The assumed data generating process is  $\mathbf{y}_n = \mathbf{H} \mathbf{w}_n + \boldsymbol{\eta}_n$ ,  $\boldsymbol{\eta}_n \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_y})$ . Data is generated with  $\sigma = 0.01$ , and  $\boldsymbol{\theta} = (\rho, \ell) = (10^{-3}, 10)$ . Filtering is done using LR-ExKF, with  $k = k' = 128$  modes, to compute the NL-statFEM posterior  $p(\mathbf{w}_n | \mathbf{y}_{1:n}, \boldsymbol{\theta}_{1:n}, \sigma_{1:n}, \Lambda) \sim \mathcal{N}(\mathbf{m}_n, \mathbf{L}_n \mathbf{L}_n^\top)$ . For the various filters in this section, each retain at least 99% of the variance at each timestep. The leading eigenvalues of  $\mathbf{K}_\theta$ , which determine the accuracy of the low-rank approximation  $\mathbf{G}_\theta^{1/2}$ , are



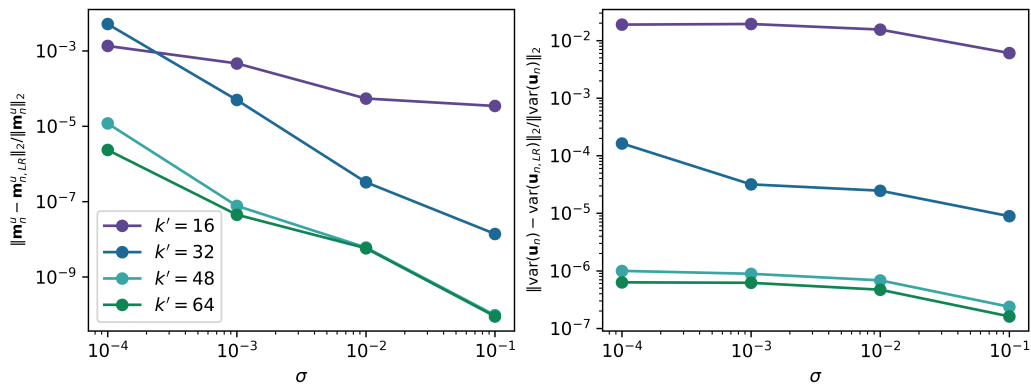
(a) Left: data, prior mean, and posterior means and 95% credible intervals across time, at the three observation locations. Right: Posterior means and 95% credible intervals, across the domain, at three timepoints. The posterior, in this case, uses the estimated hyperparameters.



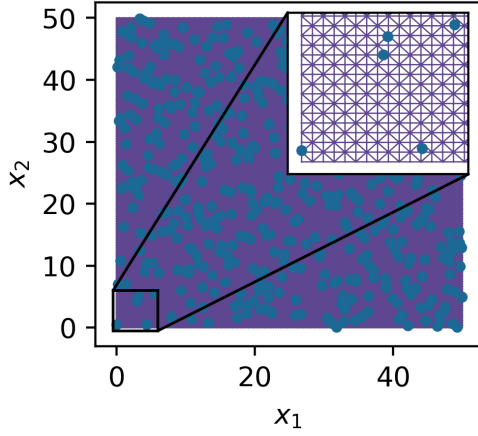
(b) Resultant estimated hyperparameters.

(c) Heatmap from the posterior mean  $\mathbf{m}_{n|n}$ .

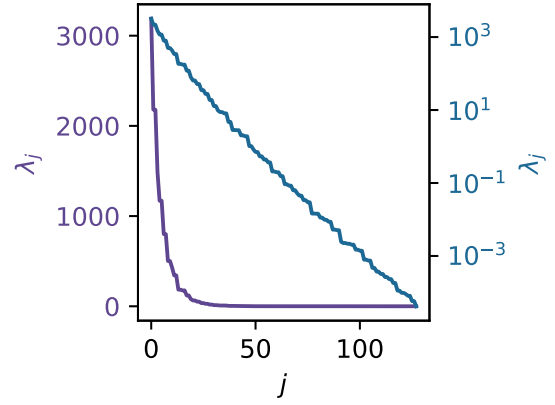
**Figure A.9:** KdV experimental data results: computed with the EnKF.



**Figure A.10:** NL-statFEM low-rank filter, experimental data: influence of prior modes on filter accuracy. Shown are the relative errors on the posterior mean (left) and variance (right) as both the number of prior modes ( $k'$ ) and observation noise ( $\sigma$ ) is increased.



**Figure A.11:** FEM mesh (purple) and observation locations (blue), for A.3.2.



**Figure A.12:** Leading 128 eigenvalues of  $\mathbf{G}_\theta$ , for A.3.2.

shown in Figure A.12, and have a range of approximately  $(10^{-4}, 10^3)$ .

To avoid recomputing  $\mathbf{G}_\theta$  at each iteration,  $\ell$  is fixed at  $\ell = 10$  for all  $n$ , and we estimate  $\rho_n$  and  $\sigma_n$  at each  $n$ , assuming that  $\rho_n$  and  $\sigma_n$  are independent for each  $n$ . Incorporating more complex temporal structure on these hyperparameters is of interest and is a possible avenue for future research. Priors are set to the weakly informative Gaussian priors  $\rho_n \sim \mathcal{N}_+(1, 1^2)$  and  $\sigma_n \sim \mathcal{N}_+(0, 1^2)$ .

The hyperparameter estimates are shown in Figures A.13a and A.13b, and demonstrate that the noise is identified at each timestep. For the larger timestep  $\Delta_t = 10^{-2}$  the hyperparameter  $\rho_n$  is poorly identified. Estimates appear to be contained within two point clouds, the topmost cloud being identified with the noise  $\sigma_n$ . This inaccuracy is thought to be due to the combination of linearised dynamics for the covariance update combined with the low-rank approximation for the square root  $\mathbf{G}_\theta$ . Note also that the inclusion of the truncated Gaussian prior  $\rho_n \sim \mathcal{N}_+(1, 1)$  will also have an effect.

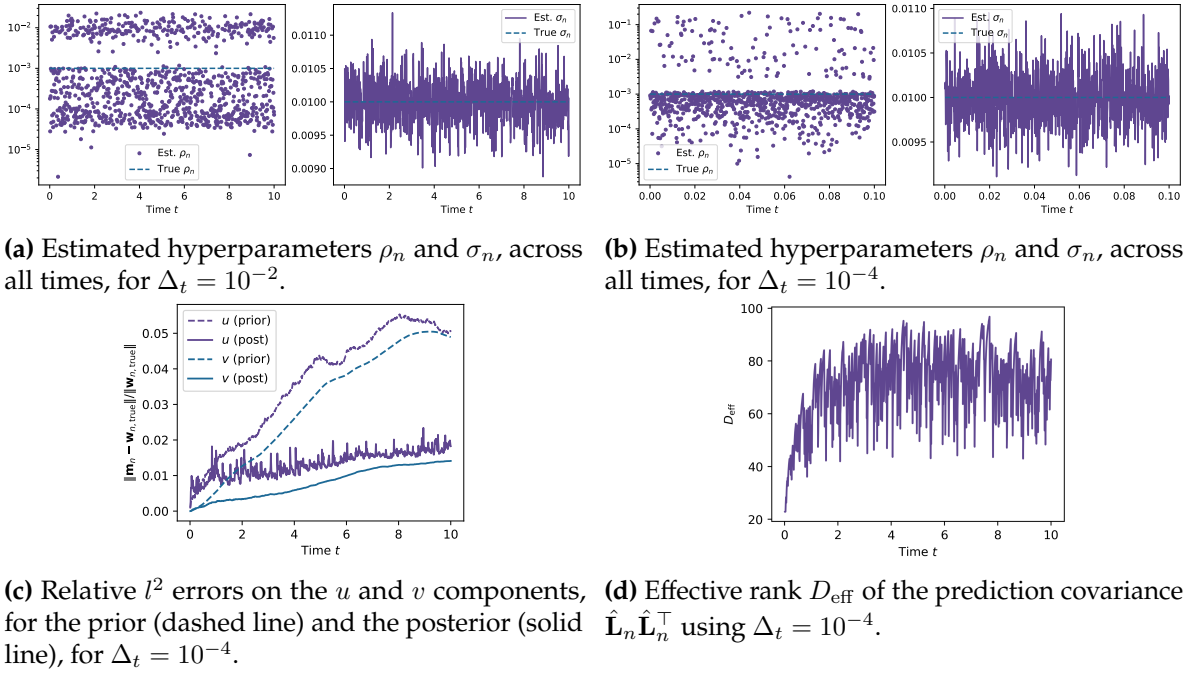
This is confirmed by running the same filter with the smaller timestep  $\Delta_t = 10^{-4}$ , thought to increase the accuracy of the linearised covariance prediction step. The filter appears to better identify the scale hyperparameter  $\rho_n$  (see Figure A.13b). Some variation remains, however, which is thought to be due to the low-rank approximation to  $\mathbf{G}_\theta$  and the truncated Gaussian prior. For completeness we also plot the relative  $l^2$  errors in Figure A.13c and the effective rank of the covariance matrix in Figure A.13d. We see that the variation of the effective rank appears to be due to the variation in the estimates of the hyperparameters (c.f. the smooth effective rank results for fixed hyperparameter values as in, e.g., Figure 5.10b).

### A.3.3 Catastrophic filter divergence in the spiral wave regime

In this case study we show that the LR-ExKF can have catastrophic filter divergence occur, which is also observed with the EnKF (Gottwald and Majda, 2013; Harlim and Majda, 2010). Catastrophic filter divergence is where the posterior mean estimate  $\mathbf{m}_n$  diverges to machine infinity in finite time. Previous studies are contextually similar, with EnKF divergence occurring in sparsely observed dissipative nonlinear systems with small noise. The mechanism of the divergence is the use of an unstable time-integration scheme (Gottwald and Majda, 2013), which we verify.

We simulate data according to a deterministic Oregonator in the spiral wave regime (recall  $f = 2$ ,





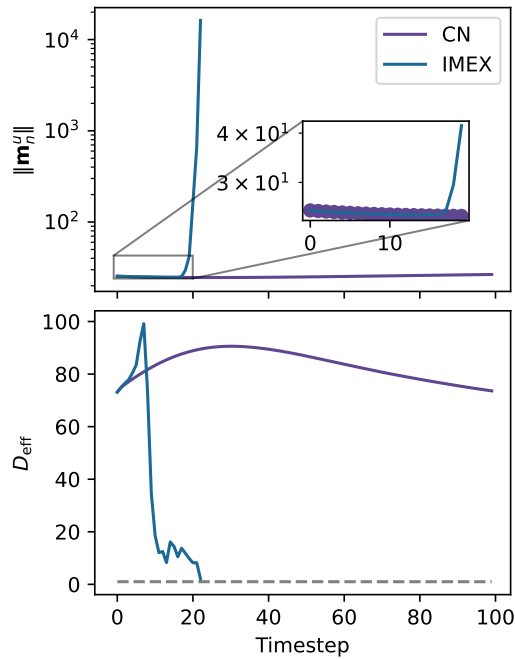
**Figure A.13:** Diagnostic plots for the stochastic forcing example.

$q = 0.002$ , and  $\varepsilon = 0.02$ , and the diffusion coefficients  $D_u = 1$ ,  $D_v = 0.6$ ). These data are observed at every timestep, on the slow  $v$  component, across 512 observation locations. The initial conditions are the same as those in the second case study in the main text. Additive Gaussian noise is added, so that  $\mathbf{y}_n = \mathbf{H}\mathbf{w}_n + \boldsymbol{\eta}_n$ ,  $\boldsymbol{\eta}_n \sim \mathcal{N}(\mathbf{0}, 10^{-4}\mathbf{I}_{n_y})$ . There is no model mismatch between the NL-statFEM model and the underlying data generating process for  $\mathbf{y}_n$ ; the only difference is that the NL-statFEM model includes the stochastic process  $\xi_v$  on the component  $v$ , which has the covariance kernel of Equation (3.2).  $\mathcal{GP}$  hyperparameters are fixed, with  $\boldsymbol{\theta} = (\rho, \ell) = (10^{-3}, 10)$ . For the numerics, both models use an FEM mesh with  $128 \times 128$  cells, and the timestep size is  $\Delta_t = 10^{-3}$ .

Filtering is done using the LR-ExKF, with  $k = 512$  and  $k' = 128$  modes, to compute the posterior  $p(\mathbf{w}_n | \mathbf{y}_{1:n}, \boldsymbol{\theta}, \sigma, \Lambda) = \mathcal{N}(\mathbf{m}_n, \mathbf{L}_n \mathbf{L}_n^\top)$ . For each filter more than 99% of the variance is retained at each timestep.. We say that the filter has diverged if any of elements  $\mathbf{u}_{n,i} \geq 10^4$  for  $i = 1, \dots, n_u$ , and two separate filters are run: one with the IMEX scheme of A.3.2 for timesteps, and the other with the Crank-Nicolson (CN) scheme for timesteps.

The IMEX filter diverges to machine infinity after 23 timesteps (see Figure A.14, top). The proposed mechanism of this divergence is that the posterior estimates of the mean do not accord with the underlying attractor, resulting in stiffness in the underlying dynamical model (Gottwald and Majda, 2013). Hence the time integration becomes stiff, which the IMEX scheme is not able to resolve, and the filter diverges. In Gottwald and Majda (2013) this is accompanied by the effective rank of the covariance matrix reducing to one, which gives spurious correlations and thus poor posterior estimates in the update step. This is observed in this scenario, too, with the effective rank dropping to near unity in finite time by the divergent timestep  $n = 23$  (see Figure A.14, bottom). Note also the resemblance to particle filter degeneracy, which results in the collapse of the particle weights to a Dirac measure.

As in examples in the main text, the behaviour of the relative norm also ‘‘lags’’ the behaviour of the effective rank; the effective rank seems to drop sharply whilst the relative norms of the IMEX and



**Figure A.14:** Norm of the NL-statFEM mean  $\mathbf{m}_n^u$  ( $u$ -component) (top), and effective rank  $D_{\text{eff}}$  (bottom).

CN are both visually indistinguishable (Figure A.14, top). Only after the effective rank drops to near unity is the divergence seen, giving a reminder of the influence of the posterior covariance matrix  $\mathbf{C}_n$  on posterior estimates of the mean  $\mathbf{m}_n$ . Changing the time integrator to CN results in catastrophic filter divergence being avoided, and the relative  $l^2$  norm and  $D_{\text{eff}}$  appear to asymptotically approach some limiting value after an initial increase (Figure A.14, bottom). Note also the smoothness of the effective rank  $D_{\text{eff}}$  resulting from the fixed choice of hyperparameter  $\rho$ .





# Bibliography

- F. Abramovich and Y. Ritov. *Statistical Theory: A Concise Introduction*. CRC Press, Apr. 2013. ISBN 978-1-4822-1184-9.
- M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. U.S. Government Printing Office, 1964.
- Ö. D. Akyildiz, C. Duffin, S. Sabanis, and M. Girolami. Statistical Finite Elements via Langevin Dynamics. *SIAM/ASA Journal on Uncertainty Quantification*, 2022.
- J. Allen. The early history of solitons (solitary waves). *Physica Scripta*, 57(3):436, 1998.
- U. M. Ascher, S. J. Ruuth, and B. T. R. Wetton. Implicit-Explicit Methods for Time-Dependent Partial Differential Equations. *SIAM Journal on Numerical Analysis*, 32(3):797–823, 1995.
- Y. Bar-Sinai, S. Hoyer, J. Hickey, and M. P. Brenner. Learning data-driven discretizations for partial differential equations. *Proceedings of the National Academy of Sciences*, 116(31):15344–15349, July 2019. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1814058116.
- M. J. Bayarri, J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C.-H. Lin, and J. Tu. A Framework for Validation of Computer Models. *Technometrics*, 49(2):138–154, May 2007a. ISSN 0040-1706. doi: 10.1198/004017007000000092.
- M. J. Bayarri, D. Walsh, J. O. Berger, J. Cafeo, G. Garcia-Donato, F. Liu, J. Palomo, R. J. Parthasarathy, R. Paulo, and J. Sacks. Computer model validation with functional output. *The Annals of Statistics*, 35(5):1874–1906, Oct. 2007b. ISSN 0090-5364, 2168-8966. doi: 10.1214/009053607000000163.
- A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind. Automatic Differentiation in Machine Learning: A Survey. *Journal of Machine Learning Research*, 18(153):1–43, 2018. ISSN 1533-7928.
- T. Bayes. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, 53:370–418, Jan. 1763. doi: 10.1098/rstl.1763.0053.
- T. Bengtsson, P. Bickel, and B. Li. Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *Probability and Statistics: Essays in Honor of David A. Freedman*, pages 316–334, Jan. 2008. doi: 10.1214/1939403070000000518.
- D. J. Benney. Long Waves on Liquid Films. *Journal of Mathematics and Physics*, 45(1-4):150–155, 1966a. ISSN 1467-9590. doi: 10.1002/sapm1966451150.

- D. J. Benney. Long Non-Linear Waves in Fluid Flows. *Journal of Mathematics and Physics*, 45(1-4):52–63, Apr. 1966b. ISSN 00971421. doi: 10.1002/sapm196645152.
- J. O. Berger and L. A. Smith. On the Statistical Formalism of Uncertainty Quantification. *Annual Review of Statistics and Its Application*, 6(1):433–460, 2019. doi: 10.1146/annurev-statistics-030718-105232.
- A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer, Boston, 2004th edition edition, Dec. 2003. ISBN 978-1-4020-7679-4.
- P. Billingsley. *Probability and Measure: 894*. Wiley, Hoboken, N.J, 4th edition edition, Mar. 2012. ISBN 978-1-118-12237-2.
- L. Boegman and M. Stastna. Sediment Resuspension and Transport by Internal Solitary Waves. *Annual Review of Fluid Mechanics*, 51(1):129–154, 2019. doi: 10.1146/annurev-fluid-122316-045049.
- J. Boussinesq. Théorie de l’intumescence liquide appelée onde solitaire ou de translation se propageant dans un canal rectangulaire. *CR Acad. Sci. Paris*, 72(755–759), 1871.
- G. E. P. Box. Robustness in the Strategy of Scientific Model Building. In R. L. Launer and G. N. Wilkinson, editors, *Robustness in Statistics*, pages 201–236. Academic Press, Jan. 1979. ISBN 978-0-12-438150-6. doi: 10.1016/B978-0-12-438150-6.50018-2.
- J. P. Boyd. *Chebyshev and Fourier Spectral Methods*. Dover Publications Inc., Mineola, N.Y, 2nd edition edition, Feb. 2001. ISBN 978-0-486-41183-5.
- S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK ; New York, 2004. ISBN 978-0-521-83378-3.
- S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer New York, New York, NY, 2008. ISBN 978-0-387-75933-3 978-0-387-75934-0. doi: 10.1007/978-0-387-75934-0.
- H. Brezis and F. Browder. Partial Differential Equations in the 20th Century. *Advances in Mathematics*, 135(1):76–144, Apr. 1998. ISSN 00018708. doi: 10.1006/aima.1997.1713.
- F.-X. Briol, C. J. Oates, M. Girolami, M. A. Osborne, and D. Sejdinovic. Probabilistic Integration: A Role in Statistical Computation? *Statistical Science*, 34(1):1–22, Feb. 2019. ISSN 0883-4237, 2168-8745. doi: 10.1214/18-STS660.
- S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, Apr. 2016. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1517384113.
- J. M. Burgers. A Mathematical Model Illustrating the Theory of Turbulence. In R. Von Mises and T. Von Kármán, editors, *Advances in Applied Mechanics*, volume 1, pages 171–199. Elsevier, Jan. 1948. doi: 10.1016/S0065-2156(08)70100-5.

- K. J. Burns, G. M. Vasil, J. S. Oishi, D. Lecoanet, and B. P. Brown. Dedalus: A flexible framework for numerical simulations with spectral methods. *Physical Review Research*, 2(2):023068, Apr. 2020. doi: 10.1103/PhysRevResearch.2.023068.
- D. Cacchione, L. F. Pratson, and A. Ogston. The shaping of continental slopes by internal tides. *Science*, 296(5568):724–727, 2002.
- G. Casella and E. I. George. Explaining the Gibbs Sampler. *The American Statistician*, 46(3):167, Aug. 1992. ISSN 00031305. doi: 10.2307/2685208.
- V. Castets, E. Dulos, J. Boissonade, and P. De Kepper. Experimental evidence of a sustained standing Turing-type nonequilibrium chemical pattern. *Physical Review Letters*, 64(24):2953–2956, June 1990. doi: 10.1103/PhysRevLett.64.2953.
- B. Charlier, J. Feydy, J. A. Glaunès, F.-D. Collin, and G. Durif. Kernel Operations on the GPU, with Autodiff, without Memory Overflows. *Journal of Machine Learning Research*, 22(74):1–6, 2021. ISSN 1533-7928.
- R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural Ordinary Differential Equations. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.
- S. Chib and E. Greenberg. Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49(4):327, Nov. 1995. ISSN 00031305. doi: 10.2307/2684568.
- P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. SIAM, 2002.
- K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 14(1):3–15, Jan. 2011. ISSN 1432-9360, 1433-0369. doi: 10.1007/s00791-011-0160-x.
- J. Cockayne, C. Oates, T. Sullivan, and M. Girolami. Probabilistic Numerical Methods for Partial Differential Equations and Bayesian Inverse Problems. *arXiv:1605.07811 [cs, math, stat]*, July 2017.
- J. Cockayne, C. J. Oates, I. C. Ipsen, M. Girolami, et al. A Bayesian conjugate gradient method (with discussion). *Bayesian Analysis*, 14(3):937–1012, 2019.
- J. D. Cole. On a Quasi-Linear Parabolic Equation Occurring in Aerodynamics. *Quarterly of Applied Mathematics*, 9(3):225–236, 1951. ISSN 0033-569X.
- P. Collet, J.-P. Eckmann, H. Epstein, and J. Stubbe. A global attracting set for the Kuramoto-Sivashinsky equation. *Communications in Mathematical Physics*, 152(1):203–214, Jan. 1993. ISSN 0010-3616, 1432-0916.
- J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301, 1965. ISSN 0025-5718, 1088-6842. doi: 10.1090/S0025-5718-1965-0178586-1.

- S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. *Statistical Science*, 28(3):424–446, Aug. 2013. ISSN 0883-4237, 2168-8745. doi: 10.1214/13-STS421.
- B. Cushman-Roisin and J.-M. Beckers. *Introduction to Geophysical Fluid Dynamics: Physical and Numerical Aspects*. Academic Press, Aug. 2011. ISBN 978-0-08-091678-1.
- G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, Cambridge, second edition, 2014. ISBN 978-1-107-05584-1. doi: 10.1017/CBO9781107295513.
- T. A. Davis. *Direct Methods for Sparse Linear Systems*. Fundamentals of Algorithms. Society for Industrial and Applied Mathematics, Jan. 2006. ISBN 978-0-89871-613-9. doi: 10.1137/1.9780898718881.
- B. de Finetti. *Theory of Probability*. Wiley, New York, 1974.
- A. De Wit. Spatial Patterns and Spatiotemporal Dynamics in Chemical Systems. In I. Prigogine and S. A. Rice, editors, *Advances in Chemical Physics*, pages 435–513. John Wiley & Sons, Inc., Hoboken, NJ, USA, Mar. 2007. ISBN 978-0-470-14168-7 978-0-471-32920-6. doi: 10.1002/9780470141687.ch5.
- L. Debnath. *Nonlinear Partial Differential Equations for Scientists and Engineers*. Birkhäuser Boston, Boston, 2012. ISBN 978-0-8176-8264-4 978-0-8176-8265-1. doi: 10.1007/978-0-8176-8265-1.
- A. Debussche and J. Printems. Numerical simulation of the stochastic Korteweg–de Vries equation. *Physica D: Nonlinear Phenomena*, 134(2):200–226, Oct. 1999. ISSN 0167-2789. doi: 10.1016/S0167-2789(99)00072-X.
- C. R. Dietrich and G. N. Newsam. Fast and Exact Simulation of Stationary Gaussian Processes through Circulant Embedding of the Covariance Matrix. *SIAM Journal on Scientific Computing*, 18(4): 1088–1107, July 1997. ISSN 1064-8275. doi: 10.1137/S1064827592240555.
- T. J. Dodwell, C. Ketelsen, R. Scheichl, and A. L. Teckentrup. A Hierarchical Multilevel Markov Chain Monte Carlo Algorithm with Applications to Uncertainty Quantification in Subsurface Flow. *arXiv:1303.7343 [math]*, Aug. 2015.
- A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, July 2000. ISSN 1573-1375. doi: 10.1023/A:1008935410038.
- P. G. Drazin and R. S. Johnson. *Solitons: An Introduction*, volume 2. Cambridge university press, 1989.
- C. Duffin, E. Cripps, T. Stemler, and M. Girolami. Statistical finite elements for misspecified models. *Proceedings of the National Academy of Sciences*, 118(2), Jan. 2021. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2015006118.
- C. Duffin, E. Cripps, T. Stemler, and M. Girolami. Low-rank statistical finite elements for scalable model-data synthesis. *Journal of Computational Physics*, 463, Aug. 2022. ISSN 0021-9991. doi: 10.1016/j.jcp.2022.111261.

- R. Durrett. *Probability: Theory and Examples*: 49. Cambridge University Press, Cambridge ; New York, NY, 5th edition edition, Apr. 2019. ISBN 978-1-108-47368-2.
- L. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, second edition, Mar. 2010. ISBN 978-0-8218-4974-3 978-1-4704-1144-2. doi: 10.1090/gsm/019.
- G. Evensen. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, Nov. 2003. ISSN 1616-7228. doi: 10.1007/s10236-003-0036-9.
- G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag, Berlin Heidelberg, second edition, 2009. ISBN 978-3-642-03710-8. doi: 10.1007/978-3-642-03711-5.
- M. H. Faber. *Statistics and Probability Theory*, volume 18 of *Topics in Safety, Risk, Reliability and Quality*. Springer Netherlands, Dordrecht, 2012. ISBN 978-94-007-4055-6 978-94-007-4056-3. doi: 10.1007/978-94-007-4056-3.
- P. E. Farrell, D. A. Ham, S. W. Funke, and M. E. Rognes. Automated Derivation of the Adjoint of High-Level Transient Finite Element Programs. *SIAM Journal on Scientific Computing*, 35(4): C369–C393, Jan. 2013. ISSN 1064-8275. doi: 10.1137/120873558.
- E. Febrianto, L. Butler, M. Girolami, and F. Cirak. A Self-Sensing Digital Twin of a Railway Bridge using the Statistical Finite Element Method. *arXiv:2103.13729 [cs, math]*, Mar. 2021.
- A. Fick. V. On liquid diffusion. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10(63):30–39, July 1855. ISSN 1941-5982. doi: 10.1080/14786445508641925.
- R. J. Field and R. M. Noyes. Oscillations in chemical systems. IV. Limit cycle behavior in a model of a real chemical reaction. *The Journal of Chemical Physics*, 60(5):1877–1884, Mar. 1974. ISSN 0021-9606. doi: 10.1063/1.1681288.
- R. J. Field, E. Koros, and R. M. Noyes. Oscillations in chemical systems. II. Thorough analysis of temporal oscillation in the bromate-cerium-malonic acid system. *Journal of the American Chemical Society*, 94(25):8649–8664, Dec. 1972. ISSN 0002-7863. doi: 10.1021/ja00780a001.
- M. Filippone and M. Girolami. Pseudo-Marginal Bayesian Inference for Gaussian Processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2214–2226, Nov. 2014. ISSN 1939-3539. doi: 10.1109/TPAMI.2014.2316530.
- C. Foias, G. R. Sell, and R. Temam. Inertial manifolds for nonlinear evolutionary equations. *Journal of Differential Equations*, 73(2):309–353, June 1988. ISSN 0022-0396. doi: 10.1016/0022-0396(88)90110-6.
- J. B. J. Fourier. *Théorie Analytique de La Chaleur*, volume 504. Didot Paris, 1822.
- C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura. Method for Solving the Korteweg-deVries Equation. *Physical Review Letters*, 19(19):1095–1097, Nov. 1967. doi: 10.1103/PhysRevLett.19.1095.

- J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson. GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration. *arXiv:1809.11165 [cs, stat]*, June 2021.
- A. Gelman and C. Hennig. Beyond subjective and objective in statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(4):967–1033, Oct. 2017. ISSN 09641998. doi: 10.1111/rssa.12276.
- A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, 3rd edition edition, Nov. 2013. ISBN 978-1-4398-4095-5.
- S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, Nov. 1984. ISSN 1939-3539. doi: 10.1109/TPAMI.1984.4767596.
- T. Gerkema and J. T. F. Zimmerman. An introduction to internal waves. *Lecture Notes, Royal NIOZ, Texel*, 207:207, 2008.
- R. Ghanem, D. Higdon, and H. Owhadi, editors. *Handbook of Uncertainty Quantification*. Springer International Publishing, Cham, 2017. ISBN 978-3-319-12384-4 978-3-319-12385-1. doi: 10.1007/978-3-319-12385-1.
- R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Courier Corporation, 2003.
- J. Gill. *Bayesian Methods: A Social and Behavioral Sciences Approach*. Chapman and Hall/CRC, New York, May 2002. ISBN 978-0-429-12347-4. doi: 10.1201/9781420057478.
- S. Gillijns, D. Bernstein, and B. De Moor. The Reduced Rank Transform Square Root Filter for Data Assimilation. *IFAC Proceedings Volumes*, 39(1):1252–1257, 2006. ISSN 14746670. doi: 10.3182/20060329-3-AU-2901.00202.
- M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods: Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, Mar. 2011. ISSN 13697412. doi: 10.1111/j.1467-9868.2010.00765.x.
- M. Girolami, E. Febrianto, G. Yin, and F. Cirak. The statistical finite element method (statFEM) for coherent synthesis of observation data and model predictions. *Computer Methods in Applied Mechanics and Engineering*, 375:113533, Mar. 2021. ISSN 0045-7825. doi: 10.1016/j.cma.2020.113533.
- M. Goldstein and J. Rougier. Bayes Linear Calibrated Prediction for Complex Systems. *Journal of the American Statistical Association*, 101(475):1132–1143, Sept. 2006. ISSN 0162-1459. doi: 10.1198/016214506000000203.
- G. H. Golub and C. F. Van Loan. *Matrix Computations*. JHU Press, Feb. 2013. ISBN 978-1-4214-0859-0.



- Y. Gong and D. J. Christini. Antispiral Waves in Reaction-Diffusion Systems. *Physical Review Letters*, 90(8):088302, Feb. 2003. doi: 10.1103/PhysRevLett.90.088302.
- G. A. Gottwald and A. J. Majda. A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks. *Nonlinear Processes in Geophysics*, 20(5):705–712, Sept. 2013. ISSN 1607-7946. doi: 10.5194/npg-20-705-2013.
- I. G. Graham, F. Y. Kuo, D. Nuyens, R. Scheichl, and I. H. Sloan. Quasi-Monte Carlo methods for elliptic PDEs with random coefficients and applications. *Journal of Computational Physics*, 230(10):3668–3694, May 2011. ISSN 0021-9991. doi: 10.1016/j.jcp.2011.01.023.
- R. Grimshaw, E. Pelinovsky, and T. Talipova. Damping of large-amplitude solitary waves. *Wave Motion*, 37(4):351–364, Apr. 2003. ISSN 01652125. doi: 10.1016/S0165-2125(02)00093-8.
- M. D. Gunzburger, C. G. Webster, and G. Zhang. Stochastic finite element methods for partial differential equations with random input data\*. *Acta Numerica*, 23:521–650, May 2014. ISSN 0962-4929, 1474-0508. doi: 10.1017/S0962492914000075.
- E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*. Springer Series in Computational Mathematics, Springer Ser.Comp.Mathem. Hairer,E.:Solving Ordinary Diff. Springer-Verlag, Berlin Heidelberg, second edition, 1993. ISBN 978-3-540-56670-0. doi: 10.1007/978-3-540-78862-1.
- M. Hairer. An Introduction to Stochastic PDEs. *arXiv:0907.4178 [math]*, July 2009.
- J. Harlim and A. J. Majda. Catastrophic filter divergence in filtering nonlinear dissipative systems. *Communications in Mathematical Sciences*, 8(1):27–43, Mar. 2010. ISSN 1539-6746, 1945-0796. doi: cms/1266935012.
- W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, Apr. 1970. ISSN 0006-3444. doi: 10.1093/biomet/57.1.97.
- P. Hennig, M. A. Osborne, and M. Girolami. Probabilistic numerics and uncertainty in computations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2179):20150142, 2015.
- J. Hensman, N. Fusi, and N. D. Lawrence. Gaussian Processes for Big Data. *arXiv:1309.6835 [cs, stat]*, Sept. 2013.
- D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafoe, and R. D. Ryne. Combining Field Data and Computer Simulations for Calibration and Prediction. *SIAM Journal on Scientific Computing*, 26(2):448–466, Jan. 2004. ISSN 1064-8275. doi: 10.1137/S1064827503426693.
- D. Higdon, J. Gattiker, B. Williams, and M. Rightley. Computer Model Calibration Using High-Dimensional Output. *Journal of the American Statistical Association*, 103(482):570–583, 2008. ISSN 0162-1459.

- D. J. Higham. An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations. *SIAM Review*, 43(3):525–546, Jan. 2001. ISSN 0036-1445. doi: 10.1137/S0036144500378302.
- E. Hopf. The partial differential equation  $u_t + uu_x = \mu u_{xx}$ . *Communications on Pure and Applied Mathematics*, 3(3):201–230, 1950. ISSN 1097-0312. doi: 10.1002/cpa.3160030302.
- D. A. Horn, J. Imberger, G. N. Ivey, and L. G. Redekopp. A weakly nonlinear model of long internal waves in closed basins. *Journal of Fluid Mechanics*, 467:269–287, Sept. 2002. ISSN 1469-7645, 0022-1120. doi: 10.1017/S0022112002001362.
- P. L. Houtekamer and H. L. Mitchell. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3):796–811, 1998.
- P. L. Houtekamer and H. L. Mitchell. A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, 129(1):123–137, Jan. 2001. ISSN 0027-0644. doi: 10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.
- X. Huang, Z. Chen, W. Zhao, Z. Zhang, C. Zhou, Q. Yang, and J. Tian. An extreme internal solitary wave event observed in the northern South China Sea. *Scientific Reports*, 6(1):30041, July 2016. ISSN 2045-2322. doi: 10.1038/srep30041.
- J. M. Hyman and B. Nicolaenko. The Kuramoto-Sivashinsky equation: A bridge between PDE'S and dynamical systems. *Physica D: Nonlinear Phenomena*, 18(1):113–126, Jan. 1986. ISSN 0167-2789. doi: 10.1016/0167-2789(86)90166-1.
- Ž. Ivezić, A. J. Connolly, J. T. VanderPlas, and A. Gray. *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data*. Princeton University Press, Jan. 2014. ISBN 978-1-4008-4891-1. doi: 10.1515/9781400848911.
- W. Jahnke, W. E. Skaggs, and A. T. Winfree. Chemical vortex dynamics in the Belousov-Zhabotinskii reaction and in the two-variable oregonator model. *The Journal of Physical Chemistry*, 93(2):740–749, Jan. 1989. ISSN 0022-3654, 1541-5740. doi: 10.1021/j100339a047.
- E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, UK; New York, NY, 2003. ISBN 978-0-511-06589-7 978-0-521-59271-0 978-0-511-06802-7 978-0-511-79042-3 978-1-139-63632-2 978-1-280-41722-1.
- A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Courier Corporation, 2007.
- H. Jeffreys. *Theory of Probability*. Oxford University Press, Oxford, third edition, 1961.
- N. L. Jones, G. N. Ivey, M. D. Rayson, and S. M. Kelly. Mixing Driven by Breaking Nonlinear Internal Waves. *Geophysical Research Letters*, 47(19):e2020GL089591, 2020. ISSN 1944-8007. doi: 10.1029/2020GL089591.
- K. Judd and L. Smith. Indistinguishable states II. The imperfect model scenario. *Physica D: Nonlinear Phenomena*, 196(3-4):224–242, Sept. 2004. ISSN 01672789. doi: 10.1016/S0167-2789(04)00182-4.

- J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*, volume 160. Springer Science & Business Media, 2006.
- R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, Mar. 1960. ISSN 0021-9223. doi: 10.1115/1.3662552.
- R. E. Kalman and R. S. Bucy. New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering*, 83(1):95–108, Mar. 1961. ISSN 0021-9223. doi: 10.1115/1.3658902.
- E. Kalnay and S.-C. Yang. Accelerating the spin-up of ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 136(651):1644–1651, 2010.
- M. Kanagawa, P. Hennig, D. Sejdinovic, and B. K. Sriperumbudur. Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences. *arXiv:1807.02582 [cs, stat]*, July 2018.
- G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, June 2021. ISSN 2522-5820. doi: 10.1038/s42254-021-00314-5.
- T. Karvonen, F. Cirak, and M. Girolami. Error analysis for a statistical finite element method. *arXiv:2201.07543 [cs, math, stat]*, Jan. 2022.
- M. C. Kennedy and A. O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, Aug. 2001. ISSN 1369-7412, 1467-9868. doi: 10.1111/1467-9868.00294.
- P. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations. Applications of Mathematics*. Springer Berlin Heidelberg, 1992. ISBN 978-3-540-54062-5.
- D. J. Korteweg and G. de Vries. XLI. On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 39(240):422–443, 1895.
- N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar. Neural Operator: Learning Maps Between Function Spaces. *arXiv:2108.08481 [cs, math]*, Dec. 2021.
- Y. Kuramoto. Diffusion-Induced Chaos in Reaction Systems. *Progress of Theoretical Physics Supplement*, 64:346–367, Feb. 1978. ISSN 0375-9687. doi: 10.1143/PTPS.64.346.
- Y. Kuramoto and T. Tsuzuki. On the Formation of Dissipative Structures in Reaction-Diffusion Systems: Reductive Perturbation Approach. *Progress of Theoretical Physics*, 54(3):687–699, Sept. 1975. ISSN 0033-068X. doi: 10.1143/PTP.54.687.
- Y. Kuramoto and T. Tsuzuki. Persistent Propagation of Concentration Waves in Dissipative Media Far from Thermal Equilibrium. *Progress of Theoretical Physics*, 55(2):356–369, Feb. 1976. ISSN 0033-068X. doi: 10.1143/PTP.55.356.

- K. G. Lamb and L. Yan. The evolution of internal wave undular bores: Comparisons of a fully nonlinear numerical model with weakly nonlinear theory. *Journal of physical oceanography*, 26(12):2712–2734, 1996.
- H. P. Langtangen and K.-A. Mardal. *Introduction to Numerical Methods for Variational Problems*, volume 21 of *Texts in Computational Science and Engineering*. Springer International Publishing, Cham, 2019. ISBN 978-3-030-23787-5 978-3-030-23788-2. doi: 10.1007/978-3-030-23788-2.
- P. S. Laplace. *Théorie analytique des probabilités*. V. Courcier, 1812.
- M. G. Larson and F. Bengzon. *The Finite Element Method: Theory, Implementation, and Applications*, volume 10 of *Texts in Computational Science and Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-33286-9 978-3-642-33287-6. doi: 10.1007/978-3-642-33287-6.
- K. Law, A. Stuart, and K. Zygalakis. *Data Assimilation: A Mathematical Introduction*, volume 62. Springer, Cham, Switzerland, 2015. ISBN 978-3-319-20325-6.
- K. J. H. Law and A. M. Stuart. Evaluating Data Assimilation Algorithms. *Monthly Weather Review*, 140(11):3757–3782, Nov. 2012. ISSN 0027-0644. doi: 10.1175/MWR-D-11-00257.1.
- P. D. Lax. Integrals of nonlinear equations of evolution and solitary waves. *Communications on Pure and Applied Mathematics*, 21(5):467–490, 1968. ISSN 1097-0312. doi: 10.1002/cpa.3160210503.
- Y. LeCun, S. Chopra, R. Hadsell, M. A. Ranzato, and F. J. Huang. A tutorial on energy-based learning. In G. Bakir, T. Hofman, B. Scholkopt, A. Smola, and B. Taskar, editors, *Predicting structured data*. MIT Press, 2006.
- C.-Y. Lee and R. C. Beardsley. The generation of long nonlinear internal waves in a weakly stratified shear flow. *Journal of Geophysical Research*, 79(3):453–462, Jan. 1974. ISSN 01480227. doi: 10.1029/JC079i003p00453.
- E. Lesaffre and A. B. Lawson. *Bayesian Biostatistics*. John Wiley & Sons, Aug. 2012. ISBN 978-0-470-01823-1.
- R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*, volume 31. Cambridge University Press, 2002.
- Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Neural Operator: Graph Kernel Network for Partial Differential Equations. *arXiv:2003.03485 [cs, math, stat]*, Mar. 2020.
- D. V. Lindley. *Introduction to Probability and Statistics from a Bayesian Viewpoint*. Cambridge University Press, 1965.
- H. Liu, Y.-S. Ong, X. Shen, and J. Cai. When Gaussian Process Meets Big Data: A Review of Scalable GPs. *IEEE Transactions on Neural Networks and Learning Systems*, 31(11):4405–4423, Nov. 2020. ISSN 2162-2388. doi: 10.1109/TNNLS.2019.2957109.

- A. Logg, K.-A. Mardal, and G. Wells. *Automated Solution of Differential Equations by the Finite Element Method: The FEniCS Book*, volume 84. Springer Science & Business Media, 2012.
- Z. Long, Y. Lu, X. Ma, and B. Dong. PDE-Net: Learning PDEs from Data. In *Proceedings of the 35th International Conference on Machine Learning*, pages 3208–3216. PMLR, July 2018.
- A. C. Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1177–1194, 1986. ISSN 1477-870X. doi: 10.1002/qj.49711247414.
- A. C. Lorenc, S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, and F. W. Saunders. The Met. Office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126(570): 2991–3012, 2000. ISSN 1477-870X. doi: 10.1002/qj.49712657002.
- D. J. C. MacKay. Hyperparameters: Optimize, or Integrate Out? In G. R. Heidbreder, editor, *Maximum Entropy and Bayesian Methods*, pages 43–59. Springer Netherlands, Dordrecht, 1996. ISBN 978-90-481-4407-5 978-94-015-8729-7. doi: 10.1007/978-94-015-8729-7\_2.
- J. Mandel. Efficient Implementation of the Ensemble Kalman Filter. *University of Colorado at Denver and Health Sciences Center, Center for Computational Mathematics*, page 9, 2006.
- A. Manderson, M. D. Rayson, E. Cripps, M. Girolami, J. P. Gosling, M. Hodkiewicz, G. N. Ivey, and N. L. Jones. Uncertainty Quantification of Density and Stratification Estimates with Implications for Predicting Ocean Dynamics. *Journal of Atmospheric and Oceanic Technology*, 36(7):1313–1330, July 2019. ISSN 0739-0572, 1520-0426. doi: 10.1175/JTECH-D-18-0200.1.
- J. Martin, L. C. Wilcox, C. Burstedde, and O. Ghattas. A Stochastic Newton MCMC Method for Large-Scale Statistical Inverse Problems with Application to Seismic Inversion. *SIAM Journal on Scientific Computing*, 34(3):A1460–A1487, Jan. 2012. ISSN 1064-8275. doi: 10.1137/110845598.
- R. J. Meinhold and N. D. Singpurwalla. Understanding the Kalman Filter. *The American Statistician*, 37(2):123–127, 1983. ISSN 0003-1305. doi: 10.2307/2685871.
- N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, June 1953. ISSN 0021-9606. doi: 10.1063/1.1699114.
- D. M. Michelson and G. I. Sivashinsky. Nonlinear analysis of hydrodynamic instability in laminar flames—II. Numerical experiments. *Acta Astronautica*, 4(11):1207–1221, Nov. 1977. ISSN 0094-5765. doi: 10.1016/0094-5765(77)90097-2.
- R. M. Miura, C. S. Gardner, and M. D. Kruskal. Korteweg-de Vries Equation and Generalizations. II. Existence of Conservation Laws and Constants of Motion. *Journal of Mathematical Physics*, 9(8): 1204–1209, Aug. 1968. ISSN 0022-2488. doi: 10.1063/1.1664701.
- K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning Series. MIT Press, Cambridge, MA, 2012. ISBN 978-0-262-01802-9.

- R. M. Neal. MCMC using Hamiltonian dynamics. *arXiv:1206.1901 [physics, stat]*, June 2012.
- B. Nicolaenko, B. Scheurer, and R. Temam. Some global dynamical properties of the Kuramoto-Sivashinsky equations: Nonlinear stability and attractors. *Physica D: Nonlinear Phenomena*, 16(2): 155–183, June 1985. ISSN 0167-2789. doi: 10.1016/0167-2789(85)90056-9.
- J. Nocedal and S. Wright. *Numerical Optimization*. Springer Science & Business Media, 2006.
- J. E. Oakley and A. O’Hagan. Probabilistic sensitivity analysis of complex models: A Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):751–769, Aug. 2004. ISSN 1369-7412, 1467-9868. doi: 10.1111/j.1467-9868.2004.05304.x.
- J. E. Oakley and A. O’Hagan. Uncertainty in prior elicitation: A nonparametric approach. *Biometrika*, 94(2):427–441, June 2007. ISSN 0006-3444. doi: 10.1093/biomet/asm031.
- B. Øksendal. Stochastic differential equations. In *Stochastic Differential Equations*, pages 65–84. Springer, 2003.
- N. Oreskes, K. Shrader-Frechette, and K. Belitz. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, 263(5147):641–646, Feb. 1994. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.263.5147.641.
- A. R. Osborne and T. L. Burch. Internal Solitons in the Andaman Sea. *Science*, 208(4443):451–460, May 1980. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.208.4443.451.
- Y. Papandreou, J. Cockayne, M. Girolami, and A. B. Duncan. Theoretical Guarantees for the Statistical Finite Element Method. *arXiv:2111.07691 [cs, math]*, Nov. 2021.
- D. J. Patil, B. R. Hunt, E. Kalnay, J. A. Yorke, and E. Ott. Local Low Dimensionality of Atmospheric Dynamics. *Physical Review Letters*, 86(26):5878–5881, June 2001. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.86.5878.
- J. E. Pearson. Complex Patterns in a Simple System. *Science*, 261(5118):189–192, July 1993. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.261.5118.189.
- K. B. Petersen and M. S. Pedersen. *The Matrix Cookbook*. Nov. 2012.
- R.-E. Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503, Nov. 2006. ISSN 0956-540X. doi: 10.1111/j.1365-246X.2006.02978.x.
- M. Plumlee. Bayesian Calibration of Inexact Computer Models. *Journal of the American Statistical Association*, 112(519):1274–1285, July 2017. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2016.1211016.
- M. Plumlee, V. R. Joseph, and H. Yang. Calibrating Functional Parameters in the Ion Channel Models of Cardiac Cells. *Journal of the American Statistical Association*, 111(514):500–509, Apr. 2016. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2015.1119695.

- K. R. Popper. *The Logic of Scientific Discovery*. The Logic of Scientific Discovery. Basic Books, Oxford, England, 1959.
- C. Rackauckas, Y. Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, D. Skinner, A. Ramadhan, and A. Edelman. Universal Differential Equations for Scientific Machine Learning. *arXiv:2001.04385 [cs, math, q-bio, stat]*, Nov. 2021.
- M. Raissi and G. E. Karniadakis. Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357:125–141, Mar. 2018. ISSN 0021-9991. doi: 10.1016/j.jcp.2017.11.039.
- M. Raissi, P. Perdikaris, and G. E. Karniadakis. Numerical Gaussian Processes for Time-dependent and Non-linear Partial Differential Equations. *arXiv:1703.10230 [cs, math, stat]*, Mar. 2017.
- M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, Feb. 2019. ISSN 0021-9991. doi: 10.1016/j.jcp.2018.10.045.
- F. P. Ramsey. Truth and Probability. In R. Braithwaite, editor, *The Foundations of Mathematics and Other Logical Essays*, pages 156–198. London: Kegan Paul, Trench, Trubner, & Co, 1926.
- L. Rayleigh. XXXII. On waves. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1(4):257–279, Apr. 1876. ISSN 1941-5982. doi: 10.1080/14786447608639037.
- M. D. Rayson, G. N. Ivey, N. L. Jones, M. J. Meuleners, and G. W. Wake. Internal tide dynamics in a topographically complex region: Browse Basin, Australian North West Shelf. *Journal of Geophysical Research: Oceans*, 116(C1), 2011. ISSN 2156-2202. doi: 10.1029/2009JC005881.
- M. D. Rayson, N. L. Jones, and G. N. Ivey. Observations of Large-Amplitude Mode-2 Nonlinear Internal Waves on the Australian North West Shelf. *Journal of Physical Oceanography*, 49(1):309–328, Jan. 2019. ISSN 0022-3670, 1520-0485. doi: 10.1175/JPO-D-18-0097.1.
- G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, Dec. 1996. ISSN 1350-7265.
- D. Rozier, F. Birol, E. Cosme, P. Brasseur, J. M. Brankart, and J. Verron. A Reduced-Order Kalman Filter for Data Assimilation in Physical Oceanography. *SIAM Review*, 49(3):449–465, Jan. 2007. ISSN 0036-1445. doi: 10.1137/050635717.
- S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, Apr. 2017. ISSN 2375-2548. doi: 10.1126/sciadv.1602614.
- Y. Saad. *Iterative Methods for Sparse Linear Systems*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics, Jan. 2003. ISBN 978-0-89871-534-7. doi: 10.1137/1.9780898718003.

- Y. Saatici. *Scalable Inference for Structured Gaussian Process Models*. PhD thesis, University of Cambridge, 2011.
- J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and Analysis of Computer Experiments. *Statistical Science*, 4(4):409–423, Nov. 1989. ISSN 0883-4237, 2168-8745. doi: 10.1214/ss/1177012413.
- S. Särkkä and A. Solin. *Applied Stochastic Differential Equations*, volume 10. Cambridge University Press, 2019.
- L. J. Savage. *The Foundations of Statistics*. Courier Corporation, 1972.
- L. F. Shampine and C. W. Gear. A user’s view of solving stiff ordinary differential equations. *SIAM review*, 21(1):1–17, 1979.
- S. Sharma. Markov Chain Monte Carlo Methods for Bayesian Data Analysis in Astronomy. *Annual Review of Astronomy and Astrophysics*, 55(1):213–259, 2017. doi: 10.1146/annurev-astro-082214-122339.
- R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer Texts in Statistics. Springer International Publishing, fourth edition, 2017. ISBN 978-3-319-52451-1. doi: 10.1007/978-3-319-52452-8.
- M. J. Simpson, R. E. Baker, S. T. Vittadello, and O. J. Maclaren. Practical parameter identifiability for spatio-temporal models of cell invasion. *Journal of The Royal Society Interface*, 17(164):20200055, Mar. 2020. doi: 10.1098/rsif.2020.0055.
- G. I. Sivashinsky. Nonlinear analysis of hydrodynamic instability in laminar flames—I. Derivation of basic equations. *Acta Astronautica*, 4(11):1177–1206, Nov. 1977. ISSN 0094-5765. doi: 10.1016/0094-5765(77)90096-0.
- G. I. Sivashinsky. On Flame Propagation Under Conditions of Stoichiometry. *SIAM Journal on Applied Mathematics*, 39(1):67–82, 1980. ISSN 0036-1399.
- E. Snelson and Z. Ghahramani. Sparse Gaussian Processes using pseudo-inputs. In *In NIPS*, 2006.
- E. Snelson, Z. Ghahramani, and C. Rasmussen. Warped Gaussian Processes. In *Advances in Neural Information Processing Systems*, volume 16. MIT Press, 2004.
- G. Söderlind, L. Jay, and M. Calvo. Stiffness 1952–2012: Sixty years in search of a definition. *BIT Numerical Mathematics*, 55(2):531–558, June 2015. ISSN 0006-3835, 1572-9125. doi: 10.1007/s10543-014-0503-3.
- A. Solin and S. Särkkä. Hilbert space methods for reduced-rank Gaussian process regression. *Statistics and Computing*, 30(2):419–446, Mar. 2020. ISSN 0960-3174, 1573-1375. doi: 10.1007/s11222-019-09886-w.
- W. A. Strauss. *Partial Differential Equations: An Introduction*. John Wiley & Sons, 2007.
- A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, May 2010. ISSN 0962-4929, 1474-0508. doi: 10.1017/S0962492910000061.



- E. Suli. Lecture Notes on Finite Element Methods for Partial Differential Equations. *University of Oxford*, page 106, Aug. 2020.
- A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Jan. 2005. ISBN 978-0-89871-572-9 978-0-89871-792-1. doi: 10.1137/1.9780898717921.
- A. L. Teckentrup, R. Scheichl, M. B. Giles, and E. Ullmann. Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients. *Numerische Mathematik*, 125(3):569–600, Nov. 2013. ISSN 0029-599X, 0945-3245. doi: 10.1007/s00211-013-0546-4.
- V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*. Number v. 25 in Springer Series in Computational Mathematics. Springer, Berlin ; New York, 2nd ed edition, 2006. ISBN 978-3-540-33121-6.
- L. Tierney. Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*, 22(4):1701–1728, Dec. 1994. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1176325750.
- M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker. Ensemble Square Root Filters\*. *Monthly Weather Review*, 131(7):1485–1490, July 2003. ISSN 0027-0644, 1520-0493. doi: 10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.
- M. Titsias. Variational Learning of Inducing Variables in Sparse Gaussian Processes. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, pages 567–574. PMLR, Apr. 2009.
- R. Tuo and C. F. J. Wu. A theoretical framework for calibration in computer models: Parametrization, estimation and convergence properties. *arXiv:1508.07155 [stat]*, Aug. 2015.
- A. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72, Aug. 1952. doi: 10.1098/rstb.1952.0012.
- J. J. Tyson and P. C. Fife. Target patterns in a realistic model of the Belousov–Zhabotinskii reaction. *The Journal of Chemical Physics*, 73(5):2224–2237, Sept. 1980. ISSN 0021-9606. doi: 10.1063/1.440418.
- M. Verlaan and A. W. Heemink. Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrology and Hydraulics*, 11(5):349–368, Oct. 1997. ISSN 0931-1955, 1436-3259. doi: 10.1007/BF02427924.
- M. Verlaan and A. W. Heemink. Nonlinearity in Data Assimilation Applications: A Practical Method for Analysis. *Monthly Weather Review*, 129(6):1578–1589, June 2001. ISSN 1520-0493, 0027-0644. doi: 10.1175/1520-0493(2001)129<1578:NIDAAA>2.0.CO;2.
- C. Villani. *Optimal Transport: Old and New*, volume 338. Springer, 2009.
- P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, Í. Polat, Y. Feng, E. W. Moore,

- J. Van der Plas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and S. . . Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3):261–272, 2020. doi: 10.1038/s41592-019-0686-2.
- E. A. Wan and R. Van Der Merwe. The unscented Kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*, pages 153–158. Ieee, 2000.
- K. Wang, G. Pleiss, J. Gardner, S. Tyree, K. Q. Weinberger, and A. G. Wilson. Exact Gaussian Processes on a Million Data Points. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
- H. Wendland. *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, 2004. ISBN 978-0-521-84335-5. doi: 10.1017/CBO9780511617539.
- N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. MIT Press, Cambridge/Mass, Mar. 1964. ISBN 978-0-262-73005-1.
- C. Williams and M. Seeger. Using the Nyström Method to Speed Up Kernel Machines. In *Advances in Neural Information Processing Systems*, volume 13. MIT Press, 2001.
- C. K. Williams and C. E. Rasmussen. *Gaussian Processes for Machine Learning*, volume 2. MIT press Cambridge, MA, 2006.
- D. Williams. Probability with Martingales. <https://www.cambridge.org/highereducation/books/probability-with-martingales/B4CFCE0D08930FB46C6E93E775503926>, Feb. 1991.
- Z.-L. Yang, R. E. Dickinson, A. Henderson-Sellers, and A. J. Pitman. Preliminary study of spin-up processes in land surface models with the first stage data of Project for Intercomparison of Land Surface Parameterization Schemes Phase 1(a). *Journal of Geophysical Research: Atmospheres*, 100 (D8):16553–16578, 1995. ISSN 2156-2202. doi: 10.1029/95JD01076.
- N. J. Zabusky and M. D. Kruskal. Interaction of “solitons” in a collisionless plasma and the recurrence of initial states. *Physical review letters*, 15(6):240, 1965.
- O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu, editors. *The Finite Element Method: Its Basis and Fundamentals*. Butterworth-Heinemann, Oxford, Jan. 2013. ISBN 978-1-85617-633-0. doi: 10.1016/B978-1-85617-633-0.00019-8.
- R. Zimmermann. On the condition number anomaly of Gaussian correlation matrices. *Linear Algebra and its Applications*, 466:512–526, Feb. 2015. ISSN 0024-3795. doi: 10.1016/j.laa.2014.10.038.
- D. Zupanski. A General Weak Constraint Applicable to Operational 4DVAR Data Assimilation Systems. *Monthly Weather Review*, 125(9):2274–2292, Sept. 1997. ISSN 0027-0644, 1520-0493. doi: 10.1175/1520-0493(1997)125<2274:AGWCAT>2.0.CO;2.