Life-Cycle of Structures and Infrastructure Systems – Biondini & Frangopol (Eds)

Efficient posterior estimation for stochastic SHM using transport maps

J. Grashorn
Institute for Risk and Reliability, Leibniz University Hannover, Germany; International Research Training Group (IRTG) 2657, Leibniz University Hannover, Germany

M. Broggi
Institute for Risk and Reliability, Leibniz University Hannover, Germany

L. Chamoin
Laboratoire de Mécanique Paris-Saclay (LMPS), Université Paris-Saclay, CentraleSupélec, ENS Paris-Saclay, CNRS, France

M. Beer
Institute for Risk and Reliability, Leibniz University Hannover, Germany; Institute for Risk and Uncertainty, University of Liverpool; Shanghai Institute for Disaster Prevention and Relief, Tongji University

ABSTRACT: Accurate parameter estimation is a challenging task that demands realistic models and algorithms to obtain the parameter's probability distributions. The Bayesian theorem in conjunction with sampling methods proved to be invaluable here since it allows for the formulation of the problem in a probabilistic framework. This opens up the possibilities of using prior information and knowledge about parameter distributions as well as the natural incorporation of aleatory and epistemic uncertainties. Traditionally, Markov Chain Monte Carlo (MCMC) methods are used to approximate the posterior distribution of samples given some data. However, these methods usually need a large amount of samples and therefore a large amount of model evaluations. Recent advances in transport theory and its application in the context of Bayesian model updating (BMU) make it possible to approximate the posterior distribution analytically and hence eliminate the need for sampling methods. This paves the way for the usage in real-time applications and for fast parameter estimation. We investigate here the application of transport maps to a simple analytical model as well as a structural dynamics model. The performance is compared to an MCMC approach to assess the accuracy and efficiency of transport maps. A discussion about requirements for the implementation of transport maps as well as details on the implementation are also given.

1 INTRODUCTION

Model parameter estimation is a daily task in many engineering disciplines. In SHM it proved to be very useful to track the changes in a system during its lifetime, thus making it possible to evaluate occurring damage or degradation of components in the system. In applications with uncertainties and stochastic parameters the model updating is usually done by using the Bayesian theorem; however depending on the system at hand the posterior density can not be calculated analytically. For the exploration of the posterior distribution then algorithms based on Markov Chain Monte Carlo (MCMC) are used, since these approaches do not require knowledge about the posterior topology. However, a downside of MCMC is that the convergence can not easily be assessed and sometimes many samples are needed in order to fully reach an adequate result. Some MCMC

DOI: 10.1201/9781003323020-82

algorithms also suffer from burn-in. Recently there have been advances in optimal transport theory (Villani 2009) which were applied in the Bayesian updating context (Parno and Marzouk 2018; Spantini et al. 2018). This opens up the possibility of circumventing some of the issues of MCMC methods, since transport maps provide a means of formulating an analytical relationship between some chosen, easy to evaluate reference distribution and the posterior distribution. Integrals can thus be evaluated on the reference distribution and then be transported to the posterior. In addition, sampling from the posterior becomes a simple evaluation of the map. The problem of finding this map is solved by optimization. Previous works have implemented transport map (TM) approximation in various use-cases, using synergies of this approach with model order reduction techniques to speed up the process (Rubio et al. 2019a,b).

In the following chapters the advantages and disadvantages of the TM approach when compared to an MCMC estimation are investigated. For this purpose we use two academic examples, one exponential function with two parameters and one structural model with three degrees of freedom.

2 PARAMETER ESTIMATION

2.1 *Bayesian model updating*

Let $\theta \in \mathbb{R}^d$ be a d-dimensional random variable with probability $p(\boldsymbol{\theta})$ describing uncertain parameters of a model $\mathcal{M}(\theta)$. Given measured data $\mathbb{D}$ the probability of observing θ in $\mathcal{M}(\theta)$ under the condition of can $\mathbb{D}$ be calculated using Bayes' theorem (Beck et al. 1998)

$$p(\boldsymbol{\theta}|\mathbb{D}) = \frac{p(\mathbb{D}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\mathbb{D})} \tag{1}$$

where the likelihood $p(\mathbb{D}|\boldsymbol{\theta})$ describes the probability of observing the data under the assumption of $\boldsymbol{\theta}$ and is usually modeled as a stochastic distance between $\mathcal{M}(\boldsymbol{\theta})$ and $\mathbb{D}$. One key difficulty in Bayesian model updating (BMU) is the irregular and unknown shape of $p(\boldsymbol{\theta}|\mathbb{D})$ and the fact that it can only be evaluated point-wise. Therefore, MCMC methods are employed to explore the probability space (Beck and Au 2002). The obtained posterior distribution $p(\boldsymbol{\theta}|\mathbb{D})$ is an expression for the updated probability for $\boldsymbol{\theta}$ constrained by the observation of $\mathbb{D}$. $p(\mathbb{D})$ is constant for any given set of model and data so Eq. (1) is also used in the non-normalized form

$$p(\boldsymbol{\theta}|\mathbb{D}) \propto p(\mathbb{D}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}). \tag{2}$$

This poses no issue for MCMC methods since the posterior's shape is not affected.

The MCMC algorithm used in this paper is the Transitional MCMC (TMCMC) method (Ching and Chen 2007). The main idea is to introduce an exponent $\alpha_j \in [0, 1]$ to the likelihood

$$p(\boldsymbol{\theta}|\mathbb{D}) \propto p(\mathbb{D}|\boldsymbol{\theta})^{\alpha_j}\, p(\boldsymbol{\theta}) \tag{3}$$

and increasing α_j with each level j starting from $\alpha_1 = 0$, which is equal to sampling from the prior density. For $\alpha_j = 1$ Eq. (3) becomes Eq. (2). Values for α_j for the intermediate levels are chosen based on the variance of the drawn samples. After drawing samples from the prior density, the Adaptive Metropolis-Hastings algorithm is used to draw samples for the next level until $\alpha_j = 1$ is reached. The main motivation behind TMCMC is to avoid the problem of sampling from difficult target PDFs but sampling from a series of PDFs that converge to the target PDF and that are easier to sample (Ching and Chen 2007).

2.2 *Transport maps*

A transport map M is a deterministic coupling between a reference density ρ and the target density π

$$\int f(y)\pi(y)\mathrm{d}y = \int (M(x))\rho(x)\mathrm{d}x \;\text{ with } Y = M(X) \tag{4}$$

where the target density in the case of BMU is the posterior distribution. The reference density can be chosen freely by the analyst. Common choices are standard normal or standard uniform distributions (Spantini et al. 2018). Any integrals on the target density can thus be calculated on the reference density by use of the map M. Moreover, samples from the target density Y can be drawn by drawing samples X from the reference density and then evaluating the map M. This makes it possible to find an analytical formulation for the posterior density in BMU, which is usually difficult or impossible. The task now becomes to find the map M. A map can be any invertible function $M: \mathbb{R}^d \to \mathbb{R}^d$, e.g. polynomials or even neural networks (Parno and Marzouk 2018). Using the notation $M_\#$ for the push-forward operation the mismatch of the approximation $\pi \approx M_\# \rho$ can be expressed with the Kullback-Leibler (KL) divergence

$$\mathcal{D}_{\mathrm{KL}}(M_\#\rho||\pi) = \mathcal{D}_{\mathrm{KL}}(\rho||M_\#^{-1}\pi) \tag{5}$$

$$= \mathbb{E}_\rho\left[\log\frac{\rho}{\mathrm{M}_\#^{-1}\pi}\right] \tag{6}$$

where the invertibility of the map is used in Eq. (5). With $\mathbf{a}$ as map parameters Eq. (6) becomes

$$\mathcal{D}_{\mathrm{KL}}(M_\#\rho||\pi) = \int_X [\log\rho(x) - \log\pi(M(\mathbf{a},x)) - \log[|\det\nabla M(\mathbf{a},x)|]]\rho(x)\mathrm{d}x \tag{7}$$

Due to optimality and uniqueness properties, maps M were proposed to be monotonic, lower-triangular and constructed from components

$$M^k(\boldsymbol{\theta}_{1,\ldots,k},\mathbf{a}) = f(\theta_1,\ldots,\theta_{k-1},0,\mathbf{a}) + \int_0^{\theta_k} g\left(\partial_k f\left(\theta_1,\ldots,\theta_{k-1},\bar{\theta},\mathbf{a}\right)\right)d\bar{\theta} \tag{8}$$

where f: $\mathbb{R}^d \to \mathbb{R}$ and g: $\mathbb{R}^d \to \mathbb{R}_+$ so that the resulting map has the structure

$$M(\boldsymbol{\theta}) = \begin{bmatrix} M^1(\theta_1) \\ \cdots \\ M^d(\theta_1,\ldots,\theta_d) \end{bmatrix} \tag{9}$$

A good approximation of the posterior density results in a small KL divergence, so that Eq. (7) can be transformed into a minimization problem. Note that $\rho(x)$ does not depend on the map parameters and instead of the full posterior $\pi = p(\boldsymbol{\theta}|\mathbb{D})$ the non-normalized form

$$\tilde{\pi} = p(\mathbb{D}|\boldsymbol{\theta})\rho(\boldsymbol{\theta}) \tag{10}$$

can be used. Furthermore, since M consists of analytical functions, the involved integrals can easily be computed by suitable quadrature rules (i.e. Gauss, Monte Carlo etc.). The final minimization problem to obtain the needed map parameters $\mathbf{a}$ is then

$$\min_{\mathbf{a}} \sum_i w_{q,i}\left[-\log\left(\tilde{\pi}\left(M\left(\boldsymbol{a},\boldsymbol{\theta}_{q,i}\right)\right)\right) - \log\left(\left|\det\nabla M\left(\mathbf{a},\boldsymbol{\theta}_{q,i}\right)\right|\right)\right] \tag{11}$$

where $w_{q,i}$ and $\boldsymbol{\theta}_{q,i}$ are weights and integration points for the quadrature rule.

For computations, the framework MParT was used (MParT Development Team 2022). As a basis for the functions f in (8) Hermite polynomials with a selectable degree were chosen. Higher order maps usually are better able to approximate the target function, however since more parameters need to be optimized their computation is also more costly. The maximum map order n can be chosen adaptively, since maps can be combined in the form

$$\mathcal{M}(\boldsymbol{\theta}) = M_n \circ M_{n-1} \circ \ldots \circ M_1(\boldsymbol{\theta}) \tag{12}$$

where the subscript denotes the map order and each M_i is of the form (9). If the accuracy of $\mathcal{M}$ after calculation of the i-th component is deemed too low, another component of order $i+1$ can be computed.

2.2.1 *Laplace approximation*

As a first approximation a map to a Gaussian with mean $\boldsymbol{\theta}_0$ and covariance Σ_0 can be calculated by use of Laplace approximation. $\boldsymbol{\theta}_0$ and Σ_0 are found by solving the optimization problem

$$\boldsymbol{\theta}_0 = \arg\min_{\theta} -\log\tilde{\pi}(\boldsymbol{\theta}) \tag{13}$$

which corresponds to finding the mode of the posterior. Σ_0 corresponds to finding the Hessian $\mathcal{H}$ at $\boldsymbol{\theta}_0$, the final Laplace map thus becomes (Rubio et al. 2019a)

$$L(\boldsymbol{\theta}) = \boldsymbol{\theta}_0 - \mathcal{H}^{-\frac{1}{2}}\boldsymbol{\theta} \tag{14}$$

The Laplace approximation can be used to regularize the problem when combined with (12) to give the final form of $\mathcal{M}$

$$\mathcal{M}(\boldsymbol{\theta}) = M_n \circ M_{n-1} \circ \ldots \circ M_1 \circ L(\boldsymbol{\theta}) \tag{15}$$

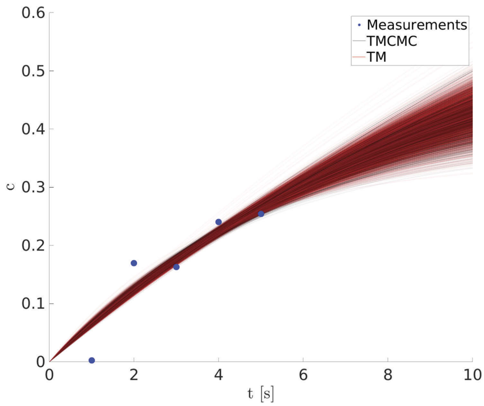

Figure 1. Plot of c over t for the analytical exponential model. Blue dots are noisy measurements, red and black lines are outputs of samples taken from TM and TMCMC approaches respectively.

Table 1. Number of model, gradient and hessian evaluations for TMCMC and both cases of TM approximations.

		Model evaluations
TMCMC		45000
TM	1st order	12636
	2nd order	20900
	3rd order	12500
	Total	46036

3 EXAMPLES

3.1 *Analytical exponential model*

As a simple example we show the application of transport maps to an analytical problem of the form

$$c = A\left(1 - e^{Bt}\right) + \zeta \tag{16}$$

with prior distributions $A, B \sim N(0, 1)$, $\zeta \sim \mathcal{N}\left(0, \sigma_M^2\right)$, where ζ is a zero-mean Gaussian noise with variance σ_M^2. The model was taken from (Parno and Marzouk 2018). The parameters to estimate are thus A and B. Data is taken from evaluating Eq. (16) at times $t = \{1,2,3,4,5\}$ with parameters $A = 0.4$ and $B = 0.2$. Letting $\boldsymbol{c}_M$ be the vector of measurements, $\boldsymbol{\theta} = [A,\ B]^T$ and $\sigma_M = 0.1$ the problem to solve for obtaining the log-posterior becomes

$$\begin{aligned} \log p(\boldsymbol{\theta}|c_M) &= log[p(c_M|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})] \\ &= -\frac{1}{\sigma_M^2}\sum_{i=1}^{5}\left(c_{M,j} - \left[\boldsymbol{\theta}_1\left(1 - e^{\theta_2 t_i}\right)\right]\right)^2 - \frac{1}{2}\boldsymbol{\theta}^2 \end{aligned} \tag{17}$$

Note that for the optimization procedure in Eq. (11) the gradient $\nabla_{\boldsymbol{a}}\mathcal{D}_{\text{KL}}$ is needed, which by applying the chain rule ultimately also needs $\nabla_{\boldsymbol{\theta}}\mathcal{M}(\boldsymbol{\theta})$. Since this problem is analytical the derivatives are readily available, however special care needs to be taken in situations where this is not the case.

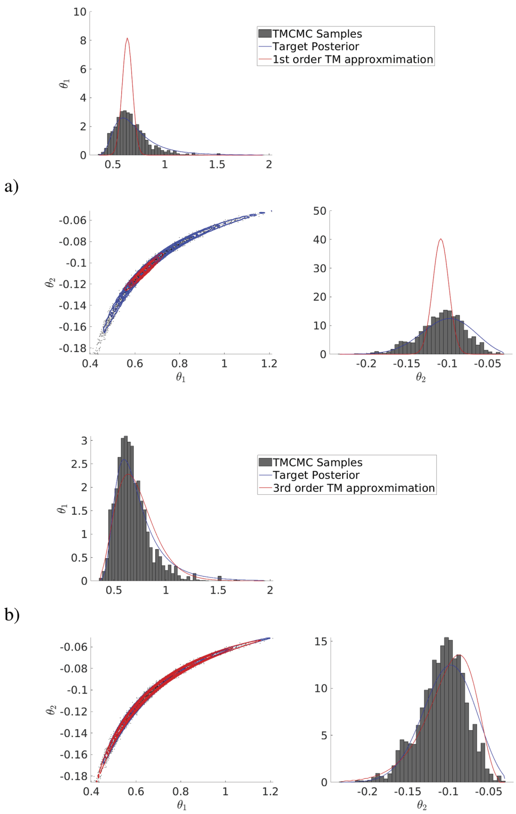

Figure 2. Marginal posteriors for updating results of TMCMC and TM approximations for the analytical exponential system. Figure a) shows a first order map approximation, b) shows a subsequent approximation with a third order map.

Figure 1 shows a result plot with indicated measurements and results taken from samples from transport maps and TMCMC. For TMCMC 5000 samples were taken per level, for the TM approach a MC-integration with 100 samples was used to calculate parameters of subsequent maps up to order 3. The results of the posterior approximation with first, second and third order TM are shown in Figure 2, together with samples taken from TMCMC and the true posterior function. In addition, Table 1 shows the number of model calls for TMCMC and for the calculation of each order of TM, together with the total amount of model calls for all three calculated maps. These results show that the first order map is only able to capture the mean of the posterior and some of its correlation structure, whereas second and third order map are able to make a better approximation. Since a first order map defines a linear relationship between the reference Gaussian and the target density the result can also only be a Gaussian. The computational cost of TM and TMCMC is comparable, however no care was taken to optimize the numerical integration scheme in the TM approach. The results show however some bias in the TMCMC samples which is a common occurrence since the convergence of MCMC methods can not be assessed easily.

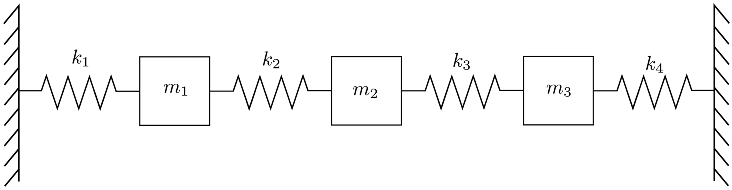

Figure 3. 3-DOF system with three masses m_1, m_2 and m_3 connected by four springs.

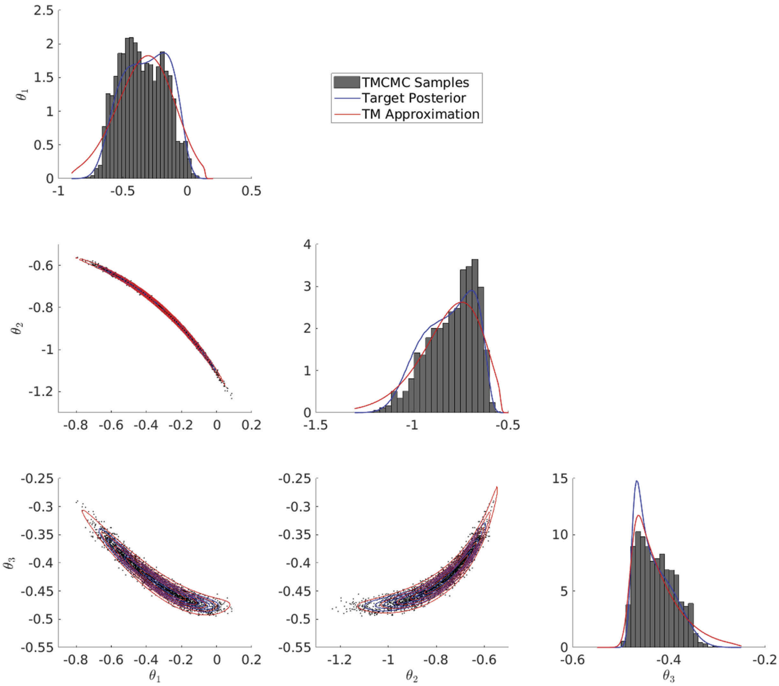

Figure 4. Marginal posterior results for posterior estimation of 3-DOF-system with TMCMC and 5th order TM approach.

Note that after calculation of the map coefficients, a fully analytical expression is obtained which allows for cheap drawing of new samples if needed. Doing the same with MCMC methods would require further model evaluations.

3.2 *3-DOF-SYSTEM*

For the second example we chose a three degree of freedom system with four springs (see Figure 3) with m_1 = 100 kg, m_2 = 200 kg and m_3 = 300 kg. k_1, k_2, and k_3 are uncertain with a prior distribution of $k_{1,2,3} \sim U(100, 3500)\,\frac{\mathrm{N}}{\mathrm{m}}$. As data we chose the three natural frequencies that can handily be calculated from the system's equations with parameters $k_1 = 1500\,\frac{\mathrm{N}}{\mathrm{m}}$, $k_2 = 750\,\frac{\mathrm{N}}{\mathrm{m}}$, and $k_4 = 1000\,\frac{\mathrm{N}}{\mathrm{m}}$. For regularization purposes $k_1 - k_3$ are transformed to standard normal space using the inverse CDF method (Devroye 1986)

$$k_i(\theta_i) = 100 + (3500 - 100)\,\Phi(\theta_i), \quad i \in \{1, 2, 3\} \tag{18}$$

where θ_i are the random variables to be updated and Φ is the CDF of the standard normal distribution. With $\boldsymbol{\omega}_m$ as the vector of measured natural frequencies the posterior becomes

$$\log p(\boldsymbol{\theta}|\omega_m) = \frac{1}{\sigma_M^2}(\omega_m - \mathcal{M}(\boldsymbol{\theta}))^2 - \frac{1}{2}\boldsymbol{\theta}^2 \tag{19}$$

where $\mathcal{M}(\boldsymbol{\theta})$ is the model to calculate the natural frequencies ω from the system's matrices.

Note that for the optimization the expression $\partial\mathcal{M}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is needed and the system matrices depend on $\mathbf{k}$, such that

$$\frac{\partial\mathcal{M}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i} = \frac{\partial\mathcal{M}(\boldsymbol{\theta})}{\partial k_i}\frac{\partial k_i}{\partial\boldsymbol{\theta}_i} = (3500 - 100)\,\phi(\theta_i)\frac{\partial\mathcal{M}(\boldsymbol{\theta})}{\partial k_i}. \tag{20}$$

where $\phi(\theta)$ is the PDF of the standard normal distribution.

Results of the updating procedures can be seen in Figure 4. For the TM approach a Gauss-Hermite integration of order five was chosen this time and the maximum order of the map was also five. For TMCMC again 5000 samples were used per level.

The target posterior in this example is more complex and shows almost a bi-modal distribution which is more difficult for the TM approach to approximate since higher order maps are needed. The mean and general covariance structure however are nicely approximated. Because of the more complicated shape of the posterior it was found that the TM approximation needed more model evaluations than the TMCMC approach. However the TMCMC samples also again show some bias.

4 CONCLUSION

In this contribution the transport map approach for estimation of the posterior in Bayesian parameter estimation was compared to standard MCMC. Both methods were used on noisy data from an analytical exponential model and a model of natural frequencies of a 3-DOF dynamical system. The TM estimation shows great promise in circumventing some of the problems arising in MCMC sampling, however the implementation and tuning of the parameters in the TM approach require more care. Especially the need to use optimization methods for finding the maps is prohibitive of their use if the problem at hand is too complex and not analytical. While the latter is not generally the case in engineering applications there exist a wide variety of model order reduction methods to replace the original system with a surrogate model that uses analytical functions. In these cases transport maps show great promise since the model becomes differentiable. The result of the transport map approximation is a fully analytic expression of the posterior distribution, allowing for integration and resampling in an efficient way, which is one of the main advantages over MCMC-based methods. Furthermore, since the KL-divergence is used to optimize the map parameters the convergence between approximation and true posterior is directly evaluated. In MCMC methods the mismatch between approximation and posterior is not quantified which can lead to bias in the drawn samples. Moreover, sequential updating which was not covered here, is naturally possible in the transport map approach by combining multiple maps. Issues with optimization convergence that arise due to the complicated shape of the posterior distribution could be reduced this way, since the change in the approximated posteriors is smaller in the sequential setting when compared to using all data points at once. The usage of transport maps for sequential updating is also interesting for on-line parameter estimation when combined with model-order reduction methods.

ACKNOWLEDGMENTS

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the GRK2657 (grant reference number 433082294) is greatly appreciated.

REFERENCES

Beck, J. L. & S.-K. Au (2002). Bayesian updating of structural models and reliability using markov chain monte carlo simulation. *Journal of engineering mechanics 128*(4), 380–391.

Beck, J. L., L. S. Katafygiotis, et al. (1998). Updating models and their uncertainties. i: Bayesian statistical framework. *Journal of Engineering Mechanics-Proceedings of the ASCE 124*(4), 455–462.

Ching, J. & Y.-C. Chen (2007). Transitional markov chain monte carlo method for bayesian model updating, model class selection, and model averaging. *Journal of engineering mechanics* (7), 816–832.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer New York, NY.

MParT Development Team (2022). Monotone Parameterization Toolkit (MParT). Version 1.2.0.

Parno, M. D. & Y. M. Marzouk (2018). Transport map accelerated markov chain monte carlo. *SIAM/ASA Journal on Uncertainty Quantification 6*(2), 645–682.

Rubio, P.-B., L. Chamoin, & F. Louf (2019b). Real-time bayesian data assimilation with data selection, correction of model bias, and on-the-fly uncertainty propagation. *Comptes Rendus Mécanique 347*(11), 762–779.

Rubio, P.-B., F. Louf, & L. Chamoin (2019a). Transport map sampling with pgd model reduction for fast dynamical bayesian data assimilation. *International Journal for Numerical Methods in Engineering 120*(4), 447–472.

Spantini, A., D. Bigoni, & Y. Marzouk (2018). Inference via low-dimensional couplings. *The Journal of Machine Learning Research 19*(1), 2639–2709.

Villani, C. (2009). *Optimal transport: old and new*, Volume 338. Springer.