Radial Basis Function Approximations of Bayesian Parameter Posterior Densities for Uncertainty Analysis

Fabian Fröhlich[1,2], Sabrina Hross[1,2], Fabian J. Theis[1,2], and Jan Hasenauer[1,2]

[1] Institute of Computational Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany
{fabian.froehlich,sabrina.hock,fabian.theis, jan.hasenauer}@helmholtz-muenchen.de
[2] Department of Mathematics, Technische Universität München, 85748 Garching, Germany

Abstract. Dynamical models are widely used in systems biology to describe biological processes ranging from single cell transcription of genes to the tissue scale formation of gradients for cell guidance. One of the key issues for this class of models is the estimation of kinetic parameters from given measurement data, the so called parameter estimation. Measurement noise and the limited amount of data, give rise to uncertainty in estimates which can be captured in a probability density over the parameter space. Unfortunately, studying this probability density, using e.g. Markov chain Monte-Carlo, is often computationally demanding as it requires the repeated simulation of the underlying model. In the case of highly complex models, such as PDE models, this can render the study intractable. In this paper, we will present novel methods for analysis of such probability densities using networks of radial basis functions. We employed lattice generation algorithms, adaptive interacting particle sampling schemes as well as classical sampling schemes for the generation of approximation nodes coupled to the respective weighting scheme and compared their efficiency on different application examples. Our analysis showed that the novel method can yield an expected L_2 approximation error in marginals that is several orders of magnitude lower compared to classical approximations. This allows for a drastic reduction of the number of model evaluations. This facilitates the analysis of uncertainty for problems with high computational complexity. Finally, we successfully applied our method to a complex partial differential equation model for guided cell migration of dendritic cells.

1 Introduction

In computational biology, parameter estimation is of crucial importance to build predictive models. In Bayesian parameter estimation, parameters are considered to be distributed according to a multi-variate probability density. This interpretation emanates from the application of Bayes' Rule:

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(D)}\,. \tag{1}$$

P. Mendes et al. (Eds.): CMSB 2014, LNBI 8859, pp. 73–85, 2014.

Here, $p(\theta|\mathcal{D})$ is the **posterior** density, a multivariate probability density of the parameter vector θ, given some dataset $\mathcal{D}$. $p(\mathcal{D}|\theta)$ is the **likelihood**, which describes the probability of observing the dataset $\mathcal{D}$ given the parameter vector θ. $p(\theta)$ is the **prior**, which encapsulates information about the parameter which is available before experiments are carried out. $p(\mathcal{D})$ is the **evidence** which normalises the posterior.

To study the uncertainty of individual parameters, the one dimensional marginal densities

$$p(\theta_i|\mathcal{D}) = \int \dots \int p(\theta|\mathcal{D}) d\theta_1 \dots d\theta_{i-1} d\theta_{i+1} \dots d\theta_{n_\theta} \tag{2}$$

are analysed. In many applications, however, no closed form of the likelihood or the posterior can be deduced. In those applications, the computation of the evidence $p(\mathcal{D}) = \int p(\theta|\mathcal{D}) d\theta$ requires the numerical multivariate integration over all parameters. However, it is usually sufficient to only consider the non-normalised posterior $q(\theta|\mathcal{D}) = p(\mathcal{D}|\theta)p(\theta)$, which does not require computation of integrals.

Still, for the computation of marginal densities, a multivariate integration is necessary. For high-dimensional integration, Markov chain Monte Carlo (MCMC) generally is the method of choice [1]. There have been many developments for efficient computation of marginals using MCMC [2,3]. For example, the MCMC method can be combined with the Variational Bayes approach, which assumes that the posterior factorises over a partition of latent variables [4].

For posterior densities with heavy tails or non-linear correlation structure, MCMC methods can have slow convergence rates [5]. Moreover, the derivation of the factorisation of the posterior can be intractable for many densities. Both of these issues frequently arise when estimating parameters in dynamical systems [3]. The evaluation of the likelihood, and thus also of the posterior, requires the solution of the dynamical system which often is not available in analytical form. Hence, one has to rely on numerical solutions to the dynamical systems which can take several seconds up to minutes when considering PDEs. A good approximation to the marginal density might require millions of samples and thus also millions of simulations of the dynamical system. Therefore, the computation of marginal densities can easily be intractable, even for dynamical systems with a small number of parameters.

In this paper we present a novel method for the approximation of posterior densities using radial basis functions. Radial basis functions are commonly employed in scattered data approximation of multivariate functions, solvers for partial differential equations [6,7] and global optimisation algorithms [8], as they yield good approximations even for small numbers of approximation nodes. In the following, we show how this method can be exploited for uncertainty analysis for parameter estimation. In particular we consider the problem of approximation node generation and compared sampling, lattice and interacting particle based methods. Furthermore, we discuss the advantages and disadvantages of the different methods. Using an example from image-based systems biology, we show that for computationally demanding problems, which for instance require

the simulation of PDEs, the proposed scheme is significantly more efficient than classical Monte-Carlo methods.

2 Methods

In the following section, we will first describe the deduction of the multivariate probability densities in Bayesian parameter estimation. Subsequently, we will describe three different methods to approximate the multivariate densities. Eventually we will discuss the approximation of marginal densities, based on marginals of the approximation.

2.1 Bayesian Inference

Recalling Equation (1), the parameter density is the product of likelihood, prior and a normalisation factor. The likelihood $p(\mathcal{D}|\theta)$ describes the probability of observing measurement data $\mathcal{D} = \{(t_k, \bar{y}(t_k))\}_{k=1}^{n_t}$, consisting of n_t time-points t_k and respective n_y-dimensional vector of observations $\bar{y}(t_k)$, given the parameter vector θ:

$$p(\mathcal{D}|\theta) = \prod_{i=1}^{n_y} \prod_{k=1}^{n_t} \sqrt{\frac{1}{2\pi\sigma_{ik}^2}} \exp\left(-\frac{1}{2}\left(\frac{\bar{y}_i(t_k) - y_i(t_k;\theta)}{\sigma_{ik}}\right)^2\right) . \tag{3}$$

Here we assumed that $y_i(t_k;\theta)$ is the simulation of the underlying model for species i, which describes the measurement data up to some additive normally distributed measurement noise with variance σ_{ik}^2 at time-point t_k :

$$\bar{y}_i(t_k) \sim \mathcal{N}(y_i(t_k;\theta), \sigma_{ik}^2) . \tag{4}$$

Bayesian inference is especially challenging for dynamical systems, as the observables are given by the map of the solution to differential equations

$$y(t;\theta) = h(u(t;\theta),\theta) , \tag{5}$$

where u is the solution to some dynamical system with parameters θ, which could be a ordinary or partial differential equation and h is the function that maps states u of the system to the observable components.

2.2 Radial Basis Function Approximation

One of the most common tools for the approximation of probability densities is kernel density estimation. For kernel density estimation, the probability density $p(\theta|\mathcal{D})$ is approximated by a equally weighted convex combination of kernel functions Φ_j. By introducing weighting factors w_j, the approximand φ can be written in a more general form:

$$\varphi^{w,\Phi,N}(\theta) = \frac{1}{\sum_j w_j} \sum_{j=1}^{N} w_j \Phi_j(\theta) . \tag{6}$$

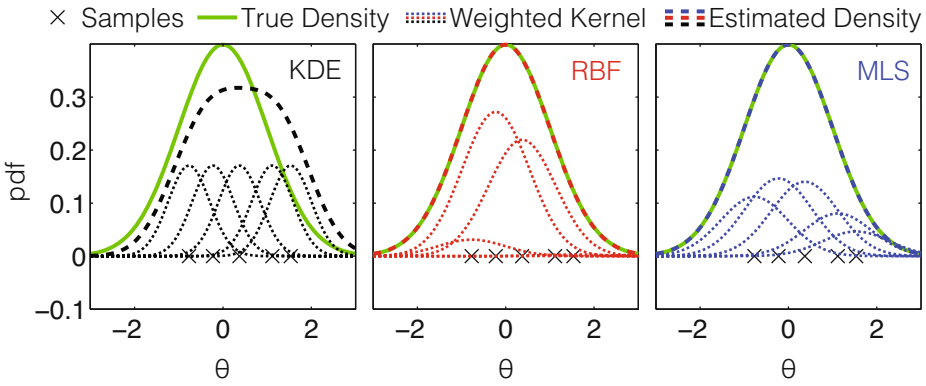

Fig. 1. Illustration of estimated densities from Kernel Density Estimation (KDE), Radial Basis Function interpolation (RBF), and Moving Least Squares Approximation. The approximated density is a univariate normal density with mean 0 and variance 1. The estimation was carried out using 5 random samples from the normal density.

These weighting factors introduce new degrees of freedom which can be used to incorporate additional information on the density such as values of point-wise evaluation.

In the following, we will discuss three different choices for the weighting factors w_j:

- **Kernel density estimation** (KDE)**:** $w_j = 1$
- **Radial Basis Function interpolation** (RBF)**:** $w_j = w_j(\theta^{(j)}, \Phi)$
- **Moving least squares approximation** (MLS)**:** $w_j = \varrho q(\theta^{(j)}|\mathcal{D})$

The approximation of a standard normal density using these kernels is schematically illustrated in Fig. 1.

In this paper, we will only consider Gaussian kernels with mean $\theta^{(j)}$ and covariance matrix $2\varepsilon M$:

$$\Phi_j(\theta) = \sqrt{\frac{1}{\pi^{n_\theta} \det(\varepsilon M)}} \exp\left(-\frac{1}{\varepsilon}(\theta - \theta^{(j)})^T M^{-1}(\theta - \theta^{(j)})\right), \tag{7}$$

in which $\theta^{(j)}$, $j = 1, \dots N$, are the approximation nodes, $M \in \mathbb{R}^{N\times N}_{sym} \succeq 0$ describes the correlation structure, $\varepsilon \in \mathbb{R}_+$ is the kernel bandwidth parameter and n_θ is the number of parameters. Although a multitude of alternatives are available, this choice for the kernel allows for analytical formulas for the marginals of φ and efficient evaluations schemes [9]. Moreover, Gaussian kernels are radial basis functions and hence positive definite kernels, which ensures uniqueness of interpolations and allows for efficient numerical interpolation algorithms [6]. However, all methods presented in the following easily translate to other kernel

functions, such as the non-smooth Wendland Kernels which should be employed when considering a non-smooth posterior [6].

For Gaussian kernels, the one-dimensional marginals are given by

$$\varphi^{w,\Phi,N}(\theta_i) = \frac{\sum_{j=1}^{N} w_j \sqrt{\frac{1}{\pi^{n_\theta} \det(\varepsilon M)}} \exp\left(-\frac{1}{\varepsilon}(\theta - \theta^{(j)})^T M^{-1} (\theta - \theta^{(j)})\right)}{\sum_j w_j} . \quad (8)$$

The convergence rate for such kernel approximations is usually quantified in terms of the asymptotic mean integrated square error (AMISE)

$$\mathrm{AMISE}_i = \int_{-\infty}^{+\infty} \mathbb{E}(\varphi^{w,\Phi,N}(\theta_i) - p(\theta_i|\mathcal{D}))^2 d\theta \,. \quad (9)$$

KDE: Kernel density estimation is the common method for estimating probability densities and has been employed for decades [10,11]. As the weight w_j is always 1, no additional computations are necessary. The free parameter ε is usually determined using Scott's rule [11]. For the convergence of the KDE, it is necessary that the approximation nodes $\Theta = \left\{\theta^{(j)}\right\}_{j=1}^{N}$ are samples of the density $p(\theta|\mathcal{D})$ which is approximated. The theoretical lower convergence bound for all non-negative KDE kernels is

$$\mathrm{AMISE}^* = \mathcal{O}(N^{-\frac{4}{5}}) \,, \quad (10)$$

which is slower than linear convergence. This lower bound is independent of the dimensionality of θ. However, this is only this bound is only asymptotic and for realistic regimes of N, dimensionality dependent effects and will often dominate.

MLS: In contrast to KDE, MLS exploits the available function values $q(\theta^{(j)}|\mathcal{D})$ at the interpolation nodes. Indeed, the MLS approximation is obtained by minimising a locally weighted distance function [6]. Accordingly, the weights are given by the scaled value of the density at that approximation point, $w_j = \varrho q(\theta^{(j)}|\mathcal{D})$. The scaling parameters ε and ϱ are obtained by minimising $\sum_{j=1}^{N} \left(\varphi^{w,\Phi,N}(\theta^{(j)}) - q(\theta|\mathcal{D})\right)^2$, while M is in general provided. The convergence of the MLS approximation has high regularity requirements on the distribution of points in Θ. In the next section we will discuss options for such regular sets of points.

RBF: Radial basis function interpolation is a common tool employed in scattered data approximation [6,7]. Here the weights are computed based on the interpolation conditions:

$$\underbrace{\begin{bmatrix} \Phi_1(\theta^{(1)}) & \cdots & \Phi_N(\theta^{(1)}) \\ \vdots & \ddots & \vdots \\ \Phi_1(\theta^{(N)}) & \cdots & \Phi_N(\theta^{(N)}) \end{bmatrix}}_{A_{\Theta,\Phi}} \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} q(\theta^{(1)}|\mathcal{D}) \\ \vdots \\ q(\theta^{(N)}|\mathcal{D}) \end{bmatrix} . \quad (11)$$

Thus, a linear system of equations must be solved to compute the weights w_j. It can be shown that the condition number of the interpolation matrix $A_{\Theta,\Phi}$

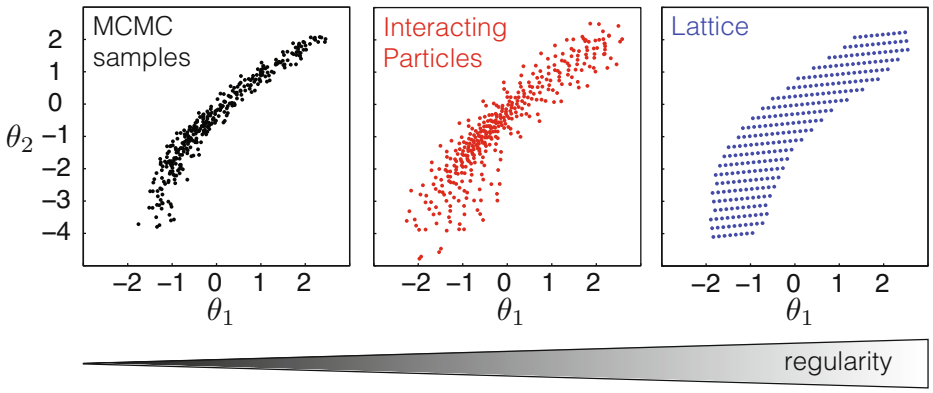

Fig. 2. Illustration of different node construction schemes. The methods are ordered by the regularity in the generated samples. All samples are generated for the density given by (14).

depends on the mesh distance $d_\Theta = \frac{1}{2}\min_{y \neq x} \|x - y\|$ with $x, y \in \Theta$. For the selection of the parameter ε, leave-one-out cross-validation (LOOCV) is the most commonly used approach [6]. For RBF interpolation, the interpolation error depends on the local fill distance $h_\rho(x) := \max_{y \in B(x,\rho)} \min_{\theta \in \Theta} \|y - \theta\| \leq h_0$ of the set Θ [12]. Therefore, the performance of RBF can be tremendously increased by imposing certain regularity based on d_Θ and h_ρ, although RBF interpolation will theoretically work on any set Θ. We will discuss the exact details of these regularity conditions in the next section.

2.3 Construction of Approximation Nodes

As discussed in the previous section, the different approximation schemes for $p(\theta|\mathcal{D})$ have different requirements on the regularity of Θ. In the following we will present methods for the construction of Θ that comply with these requirements. The resulting sets Θ are schematically depicted in Fig. 2.

Sampling: For the convergences of KDEs, the set of approximation nodes Θ has to be a statistically representative sample from the posterior distribution of $p(\theta|\mathcal{D})$. This can be achieved by constructing a Markov chain which has the distribution associated with $p(\theta|\mathcal{D})$ as equilibrium distribution. Markov chain methods enjoy great popularity as method of choice for numerical computation of high-dimensional integrals. Despite numerous improvements [2,3] of the original methods, the number of function evaluations required to obtain a converged Markov chain for multi-modal and heavy-tailed posterior distributions is often large [5].

Lattice: The MLS approach requires regular sets Θ. One possibility to obtain such regular sets are lattices. Lattices are integer linear combinations of basis

vectors and in general infinite and thus must be restricted to finite subsets for all practical purposes. It is reasonable to assume that super-level-sets of the density functions are a reasonable choice, as approximation of the density should be carried out in areas of high mass [13]. For the generation of such lattices we used the algorithms presented in [13] and also used the suggestion to motivate the level of super-level-sets using chi-square based confidence levels [13]. In the case of non-identifiable parameters, further restrictions of the domain of interest are necessary for the validity of the method. Theoretically, there exists an infinite number of different lattices as there is no restriction on the choice of basis vectors. However, [13] describes the optimality of A^* root lattices as minimiser of the mesh distance d_Θ and the global maximiser of the local fill distance h_ρ for dimensions up to three. Hence these lattices also constitute a viable choice for RBF interpolation. However, the number of lattice points grows exponentially with the number of dimensions, which limits their applicability. Sparse grids could also be a valid approach for lattice generation [14], however no efficient methods for the restriction to super-level-sets are available.

Interacting Particles: Also for the RBF method, lattices are a viable choice. RBF interpolation, however, allows for more flexibility in the choice of Θ and should also benefit from local refinements in the point density [7]. In general, the generation of locally finer lattices is difficult and a more elegant solution can be obtained by using interacting particle methods [15].

Such an interacting particle system is given by an energy function

$$E(\Theta) = \sum_{p=1}^{N} \sum_{q=1}^{N} V\left(\left\|\theta^{(q)} - \theta^{(p)}\right\|\right) , \tag{12}$$

where $V(r)$ is the potential which defines the interacting forces between particles. Here, $\operatorname{argmin}_{\Theta} E(\Theta)$ is the ground state of the particle system. By locally rescaling the potential of the function by the inverse of the density function

$$\tilde{D}(\theta) = \frac{D_0}{\sqrt{1 + \|\nabla p(\theta|\mathcal{D})\|}} , \tag{13}$$

a higher density in θ can be achieved in areas where the non-normalised density $p(\theta|\mathcal{D})$ exhibits large fluctuations [16].

In summary, there are three commonly used density approximation methods and three methods for the generation of approximation nodes. In the following, we will study the performance of the four most reasonable combinations: (i) KDE with sampling; (ii) MLS with lattices; and (ii) RBF with lattices and interacting particles.

3 Results

The evaluate the performance of the different methods, we consider two complementary examples. Firstly, we study a 2-dimensional numerical example for which analytical solutions are available. Secondly, the parameter inference for the PDE model from imaging data is considered.

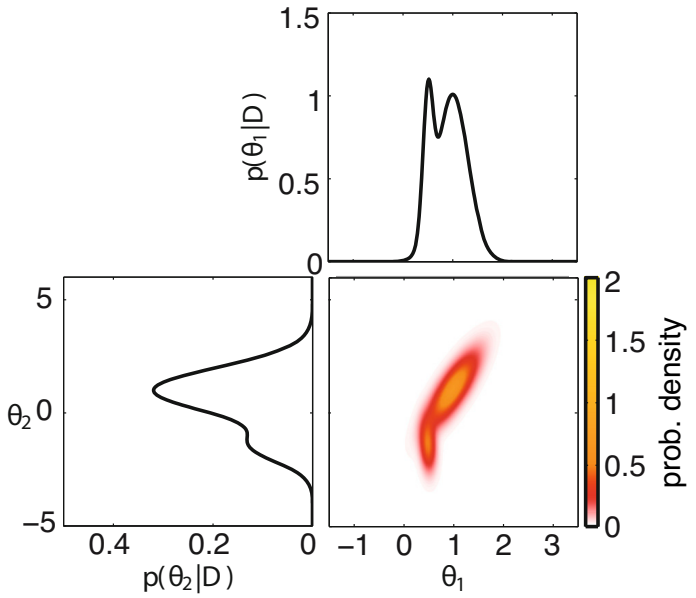

Fig. 3. Posterior density and its marginals for Example 1. The posterior density is plotted in the lower right. The respective marginal densities are plotted above and to the left of the posterior. Both marginal densities and the posterior density are bimodal.

3.1 Example 1: Numerical Example

In the first example, we assume that the posterior distribution $P(\theta|\mathcal{D})$ of the dynamical system is described by the sums of two normal distributions,

$$P(\theta|\mathcal{D}) = (\frac{4}{5}\mathcal{N}(\mu^{(1)}, \Sigma^{(1)}) + \frac{1}{5}\mathcal{N}(\mu^{(2)}, \Sigma^{(2)}) \tag{14}$$

with covariance matrices and means

$$\Sigma^{(1)} = \begin{bmatrix} 0.1 & 0.25 \\ 0.25 & 1 \end{bmatrix}, \Sigma^{(2)} = \begin{bmatrix} 0.01 & -0.01 \\ -0.01 & 0.5 \end{bmatrix}, \mu^{(1)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mu^{(2)} = \begin{bmatrix} 0.5 \\ -1.5 \end{bmatrix}. \tag{15}$$

The posterior density as well as the respective marginals are plotted in Fig. 3.

For this example the marginal densities can be computed explicitly, which allows for exact error analysis. We carried out two different comparisons. First we compared KDE on samples, MLS on lattice and RBF on lattice (Fig. 4 (left)). Secondly, we also considered RBF with interacting particles with different particle numbers (Fig. 4 (right)). The MCMC samples were generated using the DRAM toolbox [2]. For the generation and restriction of the lattice and the particle method the implementation previously described methods [13] have been used.

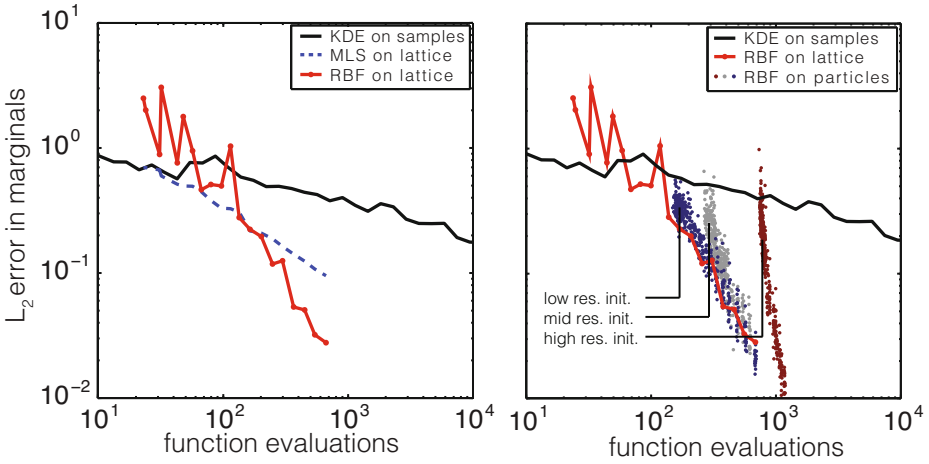

Fig. 4. Comparison of the approximation error in marginals for the KDE, the MLS and the RBF approach. The L_2 error in marginal approximation is plotted for the KDE method using MCMC samples is plotted for several sample numbers of a single Markov chain and compared to the MLS and RBF approach on the lattice. Lattices with different numbers of nodes were obtained by rescaling of the basis vectors. For the RBF method coupled to the particle method, the initialisations were generated using lattices with varying resolution.

The number of function evaluations for the lattice based approaches is determined by the scaling of the basis for the lattice generation. By decreasing the scaling factor of the basis vector, the distance between points in the lattice decreases and the total number of points and thus the total number of function evaluations increases. For low numbers of samples, KDE on MCMC samples and MLS and RBF on lattice yield similar results. However, for higher sample numbers both MLS, as well as RBF yield significantly lower errors. The error which is attained by the MLS approach at about 10^3 samples is obtained by the KDE approach between 10^4 and 10^5 samples. For the same number of samples, the RBF approach already yields an error which is approximately one magnitude lower than both the error from the KDE and the MLS approach.

To avoid additional function evaluations for the interaction particle method, the gradient $\nabla p(\theta|\mathcal{D})$ was approximated using the MLS approach. The number of function evaluation is therefore the number of points in the initial configuration plus the final number of particles. The number of particles is determined by the characteristic length of the inter-particle forces given by D_0 in (13).

Despite the required number of function evaluations for the initialisations, the combination of RBF approximation and interacting particles methods provides slightly better performance, for low and medium resolution, than the lattice-based approach. The fluctuations are a result of stochasticity in the interacting particle scheme [16]. This suggests that, for this example, the resolution of the approximation to $\nabla p(\theta|\mathcal{D})$ has only a negligible influence on the resulting error of the marginal approximation.

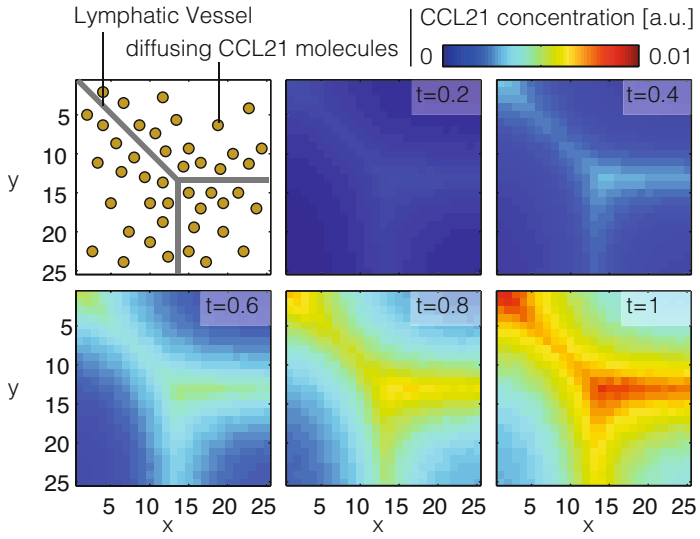

Fig. 5. Schematic of the model and synthetic data for the chemokine gradient model. The upper left plot shows the schematic of the underlying geometry for the generated synthetic data. The synthetic data for the measurement of CCL21 concentration at various time-points is shown in the other subplots.

Summarising, we could show that for this example both the MLS and the RBF approximation schemes coupled to the respective sampling schemes yielded lower errors in marginals compared to the KDE approximation scheme.

3.2 Example 2: PDE Model for Chemokine Gradient Formation

In this section we consider a PDE model describing the formation of gradients of the cytokine CCL21 around lymphatic vessels [17]. Such gradients are, among other processes, responsible for the migration of dendritic cells towards the lymphatic vessels. The subsequent dendritic cell traffic through the lymphoid vessels towards the lymph nodes is a key process in adaptive immune response.

The model considered here utilises a reaction diffusion equation to describe the secretion of CCL21 by the lymphoid vessels and the subsequent diffusion. The secreted CCL21 is then stabilised by complex formation with heparan sulfates in the surrounding tissue, which is modelled by an additional ODE. A schematic of the model as well as the utilised synthetic data is shown in Fig. 5. Of biological interest are the five unknown parameters: diffusion D, secretion α and degradation rate γ of CCL21 as well as the association k_1 and dissociation constant k_{-1} of the complex. For the detailed model, parameter inference and uncertainty analysis for those parameters we refer to the original publication [17]. In the original work the posterior probability was not calculate as it requires the simulation of the discretised PDE with several thousand state variables and is nearly infeasible with traditional methods.

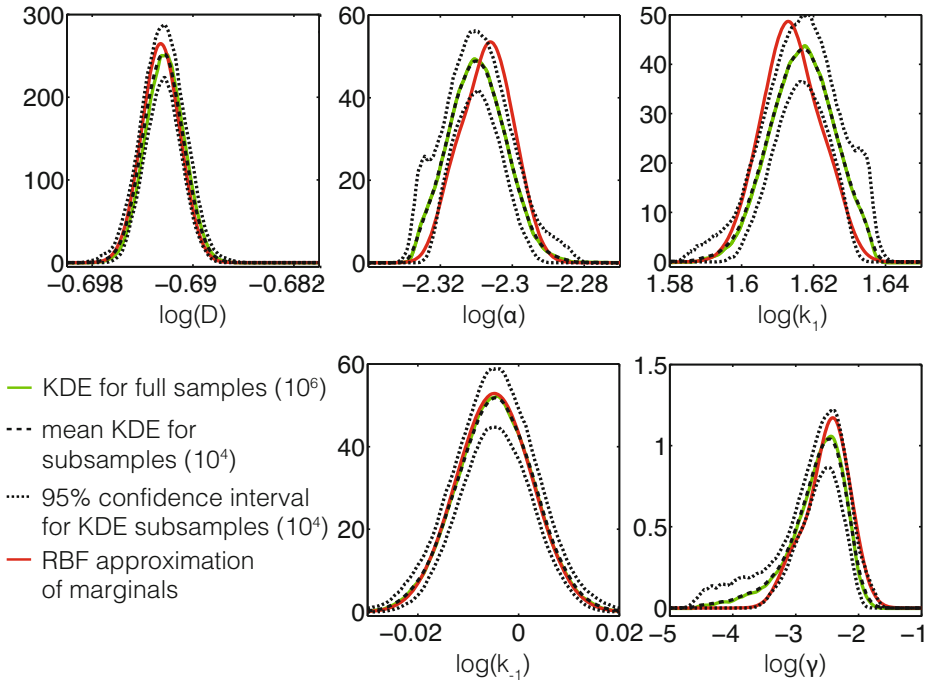

Fig. 6. Comparison of marginal densities computed using KDE and RBF approaches. The RBF approach was employed using a lattice of about 10^4 points. For the KDE approach a total of 10^6 MCMC samples were generated and percentile based confidence intervals calculated using a sliding window bootstrapping scheme with 10^3 bootstraps and a window size of 10^4.

In the following we applied the KDE and RBF approach to this problem to approximate the posterior probability and the resulting estimated marginal densities are shown in Fig 6. First 10^6 MCMC samples were generated using the DRAM toolbox [2], as these samples passed all convergence tests and its approximated marginal was considered as reference solution. Second, we generated approximation nodes for the RBF approach based on 10^4 lattice points. The resulting marginal is a reasonable approximation as it is close to the reference solution (see Fig. 6). Last, we used a sliding window approach to generate subsamples of size 10^4 from the original MCMC. Based on those samples we derived KDE based marginals comparable to the RBF approach.

If we compare both approaches we see that the 95% confidence intervals of the KDE-derived marginal show a rather large uncertainty although a highly efficient adaptive sampling scheme has been employed. In comparison the RBF approximation is often closer to the reference solution. Furthermore, the sample generation took several weeks compared to several days for the RBF lattice points. This suggests that that for this problem, RBF approximation can indeed reduce the number of required function evaluations.

4 Discussion

The problem of posterior densities approximation is omnipresent in Bayesian parameter estimation. One of the common approaches is KDE in combination with MCMC sampling. For some simple examples also analytical approximations are available. In this manuscript we suggest two alternative methods introduced in the context of scattered data approximation, namely MLS and RBF approximation. As these methods require a higher regularity of the approximation nodes, we also introduced lattices and interacting particle methods for node generation.

The resulting methods have been evaluated based on examples. For a 2-dimensional example we showed superior convergence properties for the RBF and MLS approach over the KDE approach. For a computationally demanding PDE model with 5 parameters, for which the MCMC sampling took several weeks, we showed that RBF methods provide a valid alternative which can provide better approximations with smaller samples sizes. This PDE example suggests that the method is suited for problems with a low dimensional parameter space but high computational complexity. In this case, the RBF and MLS approach could significantly reduce the number of required simulations of the systems. As this is one of the first application of MLS and RBF approximations in the context of parameter estimation [18], further studies are necessary to determine the full potential of the methods.

Acknowledgements. We thank Frank Filbir for fruitful discussions on the subject of this paper. This work was supported by a DFG Fellowship through the Graduate School of Quantitative Biosciences Munich (QBM; F.F.), the Federal Ministry of Education and Research (BMBF) within the Virtual Liver project (Grant No. 0315766; J.H.), and the European Union within the ERC grant 'LatentCauses' (F.J.T.).

References

1. Robert, C., Casella, G.: Monte Carlo Statistical Methods. Springer, New York (2004)
2. Haario, H., Laine, M., Mira, A., Saksman, E.: DRAM: Efficient Adaptive MCMC. Statistics and Computing 16(4), 339–354 (2006)
3. Schmidl, D., Czado, C., Hug, S., Theis, F.J.: A Vine-copula Based Adaptive MCMC Sampler for Efficient Inference of Dynamical Systems. Bayesian Analysis 8(1), 1–22 (2013)
4. MacKay, D.: Information Theory, Inference and Learning Algorithms (2003)
5. Jarner, S., Roberts, G.: Convergence of Heavy-tailed Monte Carlo Markov Chain Algorithms. Scandinavian Journal of Statistics 34(1994), 781–815 (2007)
6. Fasshauer, G.: Meshfree Approximation Methods with MATLAB. World Scientific (2007)
7. Sarra, S.A., Kansa, E.J.: Multiquadric Radial Basis Function Approximation Methods for the Numerical Solution of Partial Differential Equations. Advances in Computational Mechanics 2 (2009)

8. Müller, J.: Surrogate Model Algorithms for Computationally Expensive Black-Box Global Optimization Problems. PhD thesis (2012)
9. Potts, D., Steidl, G.: Fast Summation at Nonequispaced Knots by NFFT. SIAM Journal on Scientific Computing 24(6), 2013–2037 (2003)
10. Silverman, B.: Density Estimation for Statistics and Data Analysis. Chapman & Hall/CRC (1986)
11. Ramsay, P.H., Scott, D.W.: Multivariate Density Estimation, Theory, Practice, and Visualization. Wiley (1993)
12. Wu, Z., Schaback, R.: Local Error Estimates for Radial Basis Function Interpolation of Scattered Data. IMA Journal of Numerical Analysis 13(1), 1–15 (1993)
13. Fröhlich, F.: Approximation and Analysis of Probability Densities using Radial Basis Functions. Master's thesis, Technische Universität München, Germany
14. Bungartz, H.J., Griebel, M.: Sparse grids. Acta Numerica 13, 147 (2004)
15. Torquato, S.: Reformulation of the Covering and Quantizer Problems as Ground States of Interacting Particles. Physical Review E 82(5), 1–52 (2010)
16. Reboux, S., Schrader, B., Sbalzarini, I.F.: A Self-Organizing Lagrangian Particle Method for Adaptive Resolution Advection-Diffusion Simulations. Journal of Computational Physics 231(9), 3623–3646 (2012)
17. Hock, S., Hasenauer, J., Theis, F.J.: Modeling of 2D Diffusion Processes based on Microscopy Data: Parameter Estimation and Practical Identifiability Analysis. BMC Bioinformatics 14(suppl.7) (2013)
18. Bliznіouk, N., Ruppert, D., Shoemaker, C., Regis, R.: Bayesian Calibration of Computationally Expensive Models using Optimization and Radial Basis Function Approximation. Journal of Computational and Graphical Statistics 17(2) (2008)