Bayesian Inference of Plasma Diffusion Parameters: An LSSVM based approach

Mandar H. Chandorkar*
Multiscale Dynamics
Centrum Wiskunde Informatica
Amsterdam 1098XG, the Netherlands
mandar.chandorkar@cwi.nl

Enrico Camporeale
Multiscale Dynamics
Centrum Wiskunde Informatica
e.camporeale@cwi.nl

Cyril Furthlener
INRIA, Paris-Saclay
furthlen@inria.fr

Michele Sebag
INRIA, Paris-Saclay
sebag@inria.fr

Abstract

We present a method which combines sparse irregular observations with physics based models, for performing Bayesian inference over plasma diffusion parameters. Our method uses a basis function approach coupled with a least squares support vector machine objective function which weighs differently errors arising due to data fitting and satisfaction of physical constraints. The method incorporates physical models into classical least squares techniques for the purpose of data assimilation and uncertainty quantification of latent parameters.

1 Introduction

The Earth's radiation belts are the regions of space near the Earth that extend between 2 and 8 Earth's radii, where the terrestrial magnetic field traps electrons and ions in complex electromagnetic orbits [1]. Since their discovery, the belts have been the subject of intensive research due to their complex behavior and damaging effects on spacecraft [2–4].

Radiation belt particles generally execute three types of periodic motion, each with its own corresponding adiabatic invariant: gyration about magnetic field lines, bounce along field lines, and drift around the Earth. During active times, when conditions change on time scales shorter than the periods of motion, adiabaticity can be broken and particle motion can not be simply decomposed into the aforementioned components. Particle motion can then be represented diffusively along each component via the *Fokker-Planck* equation yielding a powerful picture of radiation belt dynamics [5].

The third invariant represents the total magnetic flux enclosed within a full particle orbit. It is common to use a normalized form of this, the so called Roederer L^* [6]. It is analogous to radial distance from the center of the Earth (in Earth radii) to the equatorial crossing point of the bouncing particle. Diffusion in L^* alone (the other invariants shall be considered conserved) accounts for the capture and inward radial transport of radiation belt particles ([6], [7]).

One of the main difficulties of using a physics-based model for studying and forecasting energetic electrons in the radiation belt is that the parameters that characterize the Fokker-Planck equation, namely diffusion tensor and loss term, are not directly observable. Hence, their determination represents an inverse problem, which is generally difficult to solve and can often become ill-posed.

*mandar2812.github.io

in preparation for 32st Conference on Neural Information Processing Systems (NIPS 2018), Montreal, Canada.

In this paper we propose an inference model which can learn from sparse data while taking into account prior knowledge of the system dynamics in the form of a linear partial differential equation. The method replaces the finite difference solver with a surrogate model which tries to fit the observations and the system dynamics. The surrogate is expressed as a basis function expansion whose coefficients are computed by formulating a *Least Squares Support Vector Machine*(LSSVM) like optimization objective.

In the proceeding sections, we give a short introduction to the radial diffusion equation used in magnetospheric physics. After an overview of the parameterizations of the radial diffusion unknowns used by the research community, we give a detailed formulation of our proposed method and demonstrate how one may use it for performing inference over said diffusion parameters.

1.1 Plasma Diffusion

The radial diffusion system is a simplified one-dimensional version of the *Fokker-Planck* equation. It tracks the time evolution of the *phase space density* of particles, f which is governed by the differential equation (1) known as *radial diffusion* in the radiation belt community[8].

$$\frac{\partial f}{\partial t} = l^2 \frac{\partial}{\partial l}\left(\frac{\kappa(l,t)}{l^2}\frac{\partial f}{\partial l}\right) - \lambda(l,t)f + Q(l,t) \tag{1}$$

The *phase space density* f is a function of the spatial coordinate l which denotes the *Roederer* L^* or *L-shell*, and time t.

The key quantities in the system above are.

1. f: The density of particles as a function of space l and time t.
2. $\kappa(l,t)$: Diffusion field.
3. $\lambda(l,t)$: Loss rate, this is a non-negative quantity which indicates how quickly particles are lost from the radiation belts.
4. $Q(l,t)$: Particle injection rate.

Diffusion Parameters

To solve the radial diffusion system 1, the quantities $\kappa(l,t)$, $\lambda(l,t)$ and $Q(l,t)$ need to be specified. It is a common practice (see Selesnick et al. [9], Brautigam and Albert [10], Fei et al. [11] and Shprits et al. [12]) to parameterise the diffusion field κ and loss rate λ in the following manner.

$$\kappa(l,t), \lambda(l,t) \sim \alpha l^{\beta} 10^{bKp(t)} \tag{2}$$
$$\tag{3}$$

The quantities α, β and b are parameters which define the diffusion field and loss rate while the quantity $Kp(t)$ is known as the Kp index, a measured quantity which stands as a proxy for the global geomagnetic activity [13].

2 Inverse Problems

The *inverse problem* can be stated as follows: given a set of noisy observations y scattered in the space time domain, of a physical quantity f governed by the dynamical system $\mathcal{L}_\theta f = Q_\theta$, estimate the parameters θ of the dynamical system $(\mathcal{L}_\theta, Q_\theta)$.

In this formalism $\mathcal{L}_\theta$ is a differential operator, Q_θ is a *source term* and θ is a collection of parameters which specify the operator and the source term.

In the radial diffusion system (1), $\mathcal{L}_\theta = \frac{\partial}{\partial t} - l^2 \frac{\partial}{\partial l}\left(\frac{\kappa(l,t)}{l^2}\frac{\partial}{\partial l}\right) + \lambda(l,t)$. In this case θ would be a collection of parameters which would specify the analytic expressions for $\kappa(l,t)$, $\lambda(l,t)$ and $Q(l,t)$.

2.1 Related Work

Meshfree PDE solutions

Least Squares Support Vector Machines (LSSVM) have been applied to calculating approximate solutions to PDEs [14], [15] as well as parameter estimation of delay differential equations [16], the approach taken in the aforementioned research (Mehrkanoon et al. [16]) was expressing the parameter estimation of time delay as an algebraic optimization problem resulting in closed-form approximation for the time varying parameters while avoiding iterative simulation of the dynamical system (governed by the delay differential equations) in the parameter estimation process.

Radial Basis Functions (RBF) were first applied for solution of PDE problems in [17], the authors used colocation with *multiquadric* basis functions for approximating solutions of boundary value problems.

Radial basis functions have been applied for the mesh-free solutions of Poisson PDE systems ([18], [19], [20] & [21]), as well as the Poisson control problem Pearson [22]. Further applications of RBFs include atmospheric flow [23], convection-diffusion [24], and Schrödinger's equation ([25, 26]). See ref. [27] for a recent textbook with geoscience applications.

Gaussian Processes Gaussian Process (GP) models [28] have a rich theory which has much overlapping with linear systems and deterministic and stochastic differential equations.

Skilling [29] presented one of the earliest works which focused on calculating solutions of ordinary differential equations (ODE) systems with Gaussian Process methodology, Graepel [30] applied it for solving linear partial differential equations with Dirichlet and Von Neumann boundary conditions.

Interplay between linear operators and GP models applied to Bayesian filtering was investigated by Särkkä [31]. Dondelinger et al. [32] proposed an adaptive gradient matching technique to used Gaussian Process models for infering parameters of coupled ODE systems.

Neural Networks were also employed for solutions of boundary value problems in the works such as Lagaris et al. [33], Aarts and van der Veer [34] & Tsoulos et al. [35]. Baymani et al. [36] used Feedforward networks for calculating solutions to the Stokes problem. These approaches generally revolved around decomposing the solution into two components, i.e. the first one satisfying the boundary conditions and the second one represented by the feedforward network.

3 Methodology

Performing Bayesian inference over parameters of physical systems, involves synthesizing pre-exsiting knowledge of the physical system in question i.e. the *partial differential equation* (PDE), with statistical techniques. The aim of such an exercise is often the quantification of uncertainty over system parameters from an often sparse set of observations which are quantities of interest in the physical system.

3.1 Model Formulation

We approach the radial diffusion inference problem by formulating a modified version of the *least squares support vector machine* predictor for obtaining a closed form approximation to the phase space density f which tries to satisfy the radial diffusion PDE 1 on a fixed set of *colocation* points while minimizing error on a set of sparse noisy observations. Using the surrogate phase space density estimator as a baseline, we specify the likelihood of the observations.

Surrogate Phase Space Density Model

Let $\mathcal{D} = (x_i^o, y_i) : i = 1 \cdots n_o$ be a set of noisy observations of the phase space density f, where $x_i = (l_i, t_i)$ are points in the space time domain. We seek a linear estimator for f of the form $\hat{f}(x) = w^T \varphi(x) + b$, where $\varphi(.) : \mathbb{R}^2 \rightarrow \mathbb{R}^d$ is a d dimensional feature map and b is a scalar intercept.

Further let $\mathcal{C} = (x_i^c, q_i) : i = 1 \cdots n_c$ be a set of colocation points on which we aim to enforce radial diffusion dynamics. The values q_i represent the particle injection rate Q at x^c and are calculated each time the parameters of Q are sampled, by evaluating the expression $Q(l, t) = (\alpha_Q l^{\beta_Q} + \gamma_Q) 10^{b_Q Kp(t)}$.

We exploit the linearity of the differential operator $\mathcal{L}_\theta$ and note that $\mathcal{L}_\theta[\hat{f}(x)] = w^T \mathcal{L}_\theta[\varphi(x)]$, yielding an estimator $\hat{Q}(x) = w^T \psi(x)$ where $\psi_\theta(x) = \mathcal{L}_\theta[\varphi(x)]$. Calculating $w \in \mathbb{R}^d$ can now be cast as the following constrained L_2 regularized least squares problem.

$$min_{w,e,\epsilon}\, \mathcal{J}(w, e, \epsilon; \theta) = \frac{1}{2} w^T w + \frac{1}{2\gamma_o} \sum_{k=1}^{n_o} e_k^2 + \frac{1}{2\gamma_c} \sum_{k=1}^{n_c} \epsilon_k^2 \tag{4}$$

$$s.t$$

$$y_i = w^T \varphi(x_i^o) + b + e_i, \;\; i = 1 \cdots n_o \tag{5}$$

$$q_i = w^T \psi_\theta(x_i^c) + \epsilon_i, \;\; i = 1 \cdots n_c \tag{6}$$

It can be seen that system 4 is similar to the formulation of the LSSVM model, while incorporating the dynamics of linear PDE systems into its loss function.

The quantities γ_o and γ_c are weights attached to the errors on observations and colocation points respectively, thus by smoothly varying them one may assign higher or lower importance for the surrogate model to fit the observational data and the dynamics of the physical system.

In order to solve this system one must construct its *Lagrangian*.

$$\begin{aligned}\mathfrak{L}(w, e, \epsilon, \alpha_{1\cdots k}, \beta_{1\cdots k}; \theta; \gamma_o; \gamma_c) = {} & \frac{1}{2} w^T w + \frac{1}{2\gamma_o} \sum_{k=1}^{n_o} e_k^2 + \frac{1}{2\gamma_c} \sum_{k=1}^{n_c} \epsilon_k^2 \\ & + \sum_{k=1}^{n_o} \alpha_k (y_k - w^T \varphi(x_k^o) - b - e_k) \\ & + \sum_{k=1}^{n_c} \beta_k (q_j - w^T \psi_\theta(x_j^c) - \epsilon_j)\end{aligned}$$

Given fixed values for γ_o, γ_c and PDE parameters θ, the equation above expresses the Lagrangian of system 4. The quantities $\alpha_1, \cdots, \alpha_{n_o}$ and $\beta_1, \cdots, \beta_{n_c}$ are the *Lagrange multipliers* introduced for equality constraints of the system. Applying the *Karush-Kuhn-Tucker*(KKT) conditions the solution of the optimization problem 4 can be expressed in terms of the Lagrange multipliers $\alpha = (\alpha_1, \cdots, \alpha_{n_o})$ $\beta = (\beta_1, \cdots, \beta_{n_c})$.

$$\begin{bmatrix} 0 & \mathbf{1}^T & \mathbf{0} \\ \mathbf{1} & \Omega + \gamma_o I & \Omega_* \\ \mathbf{0} & \Omega_*^T & \Omega_{**} + \gamma_c I \end{bmatrix} \begin{bmatrix} b \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ y \\ q \end{bmatrix} \tag{7}$$

The components of the symmetric block matrix system on the left hand side of 7 are

1. $\Omega \in \mathbb{R}^{n_o \times n_o} : \omega_{ij} = \varphi(x_i^o)^T \varphi(x_j^o)$
2. $\Omega_{**} \in \mathbb{R}^{n_c \times n_c} : \omega_{ij}^{**} = \psi(x_i^c)^T \psi(x_j^c)$
3. $\Omega_* \in \mathbb{R}^{n_o \times n_c} : \omega_{ij}^{*} = \varphi(x_i^o)^T \psi(x_j^c)$

The surrogate model can now be used to estimate the phase space density at a point $x = (l, t)$.

$$\hat{f}(x; \theta) = \sum_{k=1}^{n_o} \alpha_k \varphi(x)^T \varphi(x_k^o) + \sum_{k=1}^{n_c} \beta_k \varphi(x)^T \psi_\theta(x_k^c) + b \tag{8}$$

Choice of Basis

There exist several choices regarding the basis $\varphi(.)$, they are but not limited to orthogonal polynomials, Fourier series, radial basis functions etc.

For our problem we choose a basis which is a product of a multiquadric radial basis in space and an inverse multiquadric basis in time.

$$\varphi_i(l,t) = \frac{\sqrt{1 + (|l - l_i|/\rho_l)^2}}{(1 + (|t - t_i|/\rho_t)^2)^{s/2}} \tag{9}$$

The nodes of the basis are placed on a regular rectangular grid and their length scales ρ_l and ρ_t are set to the distance between adjacent node points. The value of the exponent s in equation 9 should be a positive value between 0 and 1, we have seen that setting $s = 0.75$ works well in practice.

Role of γ_o and γ_c

The quantities γ_o and γ_c serve to control the importance assigned to of errors made on the observations and colocation points respectively. Varying them gives the modeler the ability to vary the behavior of the surrogate model. In the limiting case of γ_c tending to zero, the model behaves as if the PDE dynamics is enforced as a hard constraint. This case is equivalent to the following formulation.

$$min_{w,e}\ \mathcal{J}(w, e, \epsilon; \theta) = \frac{1}{2}w^T w + \frac{1}{2\gamma_o}\sum_{k=1}^{n_o} e_k^2 \tag{10}$$

$$s.t$$

$$y_i = w^T \varphi(x_i^o) + b + e_i,\ \ i = 1 \cdots n_o \tag{11}$$

$$q_i = w^T \psi_\theta(x_i^c),\ \ i = 1 \cdots n_c \tag{12}$$

Although choosing $\gamma_c = 0$ is an appropriate choice if one wishes to enforce the physical dynamics as a constraint, it can possibly lead to numerical instabilities in inverting system 7 and hence choosing a small value like $\gamma_c = 10^{-8}$ works better in practice.

3.2 Quantifying Observation Likelihood

We assume a multivariate Gaussian distribution for calculating the likelihood of the observations conditioned on the system parameters θ.

The surrogate model 8 gives a baseline or mean value for the phase space density, we use a hybrid RBF covariance function $C(x_i, x_j) = \sigma^2 exp(-\frac{1}{2}(\frac{|l_i - l_j|^2}{s} + \frac{|t_i - t_j|}{r}))$ to quantify the covariance of the phase space density f over two points $x_i = (l_i, t_i)$ and $x_j = (l_j, t_j)$ in the domain.

$$\begin{bmatrix} y_1 \\ \vdots \\ y_{n_o} \end{bmatrix} | x_1, \cdots, x_{n_o}, \theta \sim \mathcal{N}\left(\begin{bmatrix} \hat{f}(x_1) \\ \vdots \\ \hat{f}(x_{n_o}) \end{bmatrix}, \begin{bmatrix} C(x_1, x_1) & \cdots & C(x_1, x_{n_o}) \\ \vdots & \ddots & \vdots \\ C(x_{n_o}, x_{n_1}) & \cdots & C(x_{n_o}, x_{n_o}) \end{bmatrix}\right) \tag{13}$$

The values of s and r, the length scales of the covariance function, can be fixed to the size of the space-time grid of the radial basis functions, alternatively they can also be treated as system parameters which can be sampled by the inference procedure. Since the core aims of this research was the quantification of the uncertainty over the parameters of the radial diffusion system, we treat the covariance function parameters as fixed.

3.3 Inference

We employ the adaptive Metropolis algorithm as proposed by Haario et al. [37], for sampling system parameters. The adaptive Metropolis algorithm adapts the exploration variance according to the running sample statistics of the Markov Chain procedure.

Table 1: Parameters: Prior

Quantity	α	β	γ	b
λ	$Lognormal(0,1)$	$Gamma(1,1)$	N.A	$\mathcal{N}(0,2)$
Q	$Lognormal(0,2)$	$Gamma(1,1)$	$Lognormal(0,2)$	$\mathcal{N}(0,2)$

4 Experiment

For the purposes of the experiment, the parameters of κ fixed while MCMC inference is performed on the parameters of λ and Q. The prior distributions chosen for the parameters are shown in table 1. The posterior distribution over the parameters of λ is sampled via adaptive Metropolis, the first 1500 samples are kept aside as the *burn in* period of the Markov Chain.

Data Generation

The technique presented is applied on synthetic data generated using a radial diffusion solver, the ground truth values of the radial diffusion parameters are listed in table 2.

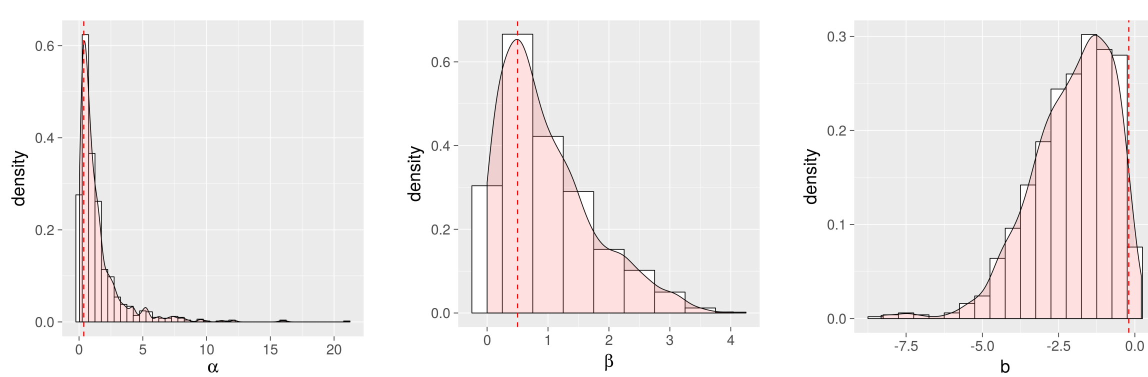
A slightly modified parameterization is adopted for the particle injection Q as shown in equation 14 below. The particle injection has a strictly time varying component due to presence of a constant γ.

$$Q(l,t) \sim (\alpha l^{\beta} + \gamma) 10^{bKp(t)} \tag{14}$$

The initial phase space density $f(t=0)$ and the time trajectory of the Kp index are assumed to be set as follows.

$$f(t=0) = 100(1 + I_1(l)) \tag{15}$$

$$Kp(t) = \begin{cases} 2.5 + 4t & 0 \leq t < 1.5 \\ 8.5 & 1.5 < t \leq 3 \\ 17.5 - 3t & 3 < t \leq 5 \\ 2.5 & 5 < t \end{cases} \tag{16}$$



(a) Posterior Samples α (b) Posterior Samples β (c) Posterior Samples b

Figure 1: Posterior distributions for parameters of λ, the red dotted line indicates ground truth for each parameter

Where I_1 is the modified Bessel function of the first kind with parameter 1. The evolution of the Kp index [13] is assumed such that it mimics a geomagnetic storm event. The radial diffusion solver is run for domain limits $l \in [1,7], t \in [0,5]$ with 200 bins in the spatial and 50 bins in the temporal domains respectively.

Table 2: Parameters: Ground Truth

Quantity	α	β	γ	b
Q	1	0.5	0.05	0.45
κ	4.731×10^{-10}	10	0	0.506
λ	0.3678	0.5	0	-0.2

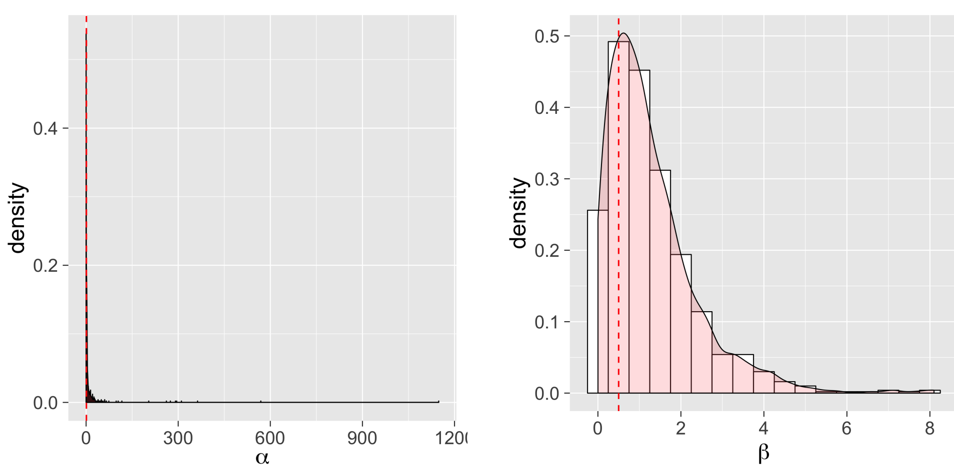
After the approximate solution profiles of f are generated, the points are sub-sampled uniformly such that 50 points lie in the interior of the domain and 20 points at the initial time step (t = 0) are selected. These observations are then perturbed by Gaussian noise to yield the final observation set $\mathcal{D}$ which is fed to the surrogate model $\hat{f}(x)$.

Results

The posterior inferred for parameters of the particle loss rate λ are shown in 1(a), 1(b) and 1(c) respectively. Figures 2(a), 2(b), 3(a) and 3(b) show the same for parameters of the particle injection Q.

We see that the marginal posterior distributions of each parameter have high probability density near the ground truth, and that they are heavy tailed because there are often large regions of the parameter space which produce similar solutions for the phase space density f.

The strength of the proposed method is the ability to quantify the uncertainty in diffusion parameters from a sparse set of observations. Due to the formulation of the surrogate as a dual optimization problem, it allows the inference to scale well with respect to high dimensional basis function expansions. The method shows promise for modeling and inference of physical systems and warrants further research in its improvement.



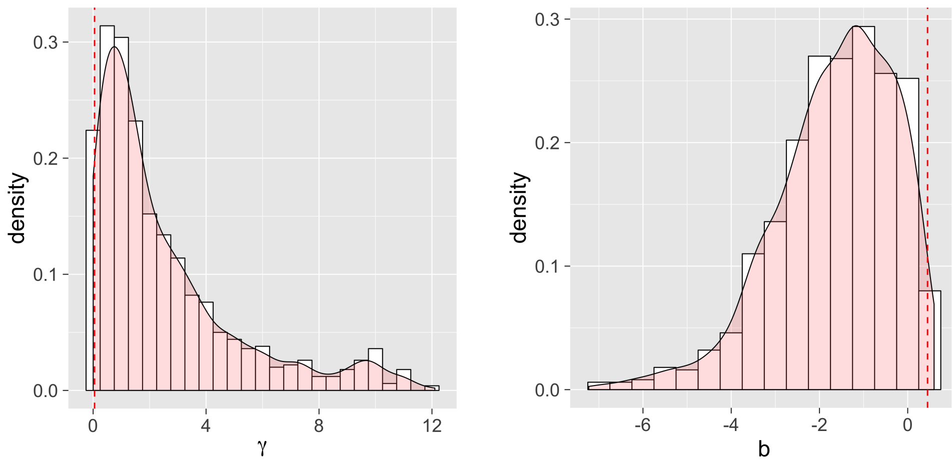
(a) Posterior Samples α

(b) Posterior Samples β

Figure 2: Posterior distributions for parameters of Q, the red dotted line indicates ground truth for each parameter

Acknowledgments

The authors would like to acknowledge INRIA associate tram grant EA-544 for enabling scientific collaboration.



(a) Posterior Samples γ

(b) Posterior Samples b

Figure 3: Posterior distributions for parameters of Q, the red dotted line indicates ground truth for each parameter

References

[1] James A. Van Allen and Louis A. Frank. Radiation around the earth to a radial distance of 107,400 km. *Nature*, 183(4659):430–434, 02 1959. URL `http://dx.doi.org/10.1038/183430a0`.

[2] Robin Gubby and John Evans. Space environment effects and satellite design. *Journal of Atmospheric and Solar-Terrestrial Physics*, 64(16):1723 – 1733, 2002. ISSN 1364-6826. doi: https://doi.org/10.1016/S1364-6826(02)00122-0. URL `http://www.sciencedirect.com/science/article/pii/S1364682602001220`. Space Weather Effects on Technological Systems.

[3] D. T. Welling. The long-term effects of space weather on satellite operations. *Annales Geophysicae*, 28: 1361–1367, June 2010. doi: 10.5194/angeo-28-1361-2010.

[4] Daniel N Baker. How to cope with space weather. *Science*, 297(5586):1486–1487, 2002.

[5] Michael Schulz and Louis J Lanzerotti. *Particle diffusion in the radiation belts*, volume 7. Springer Science & Business Media, 2012.

[6] J. G. Roederer. *Periodic Drift Motion and Conservation of the Third Adiabatic Invariant*, pages 72–83. Springer Berlin Heidelberg, Berlin, Heidelberg, 1970. ISBN 978-3-642-49300-3. doi: 10.1007/978-3-642-49300-3_3. URL `https://doi.org/10.1007/978-3-642-49300-3_3`.

[7] Carl-Gunne Fälthammar. Effects of time-dependent electric fields on geomagnetically trapped radiation. *Journal of Geophysical Research*, 70(11):2503–2516, 1965. ISSN 2156-2202. doi: 10.1029/JZ070i011p02503. URL `http://dx.doi.org/10.1029/JZ070i011p02503`.

[8] L. R. Lyons and M. Schulz. Access of energetic particles to storm time ring current through enhanced radial “diffusion”. *Journal of Geophysical Research: Space Physics*, 94(A5):5491–5496, 1989. ISSN 2156-2202. doi: 10.1029/JA094iA05p05491. URL `http://dx.doi.org/10.1029/JA094iA05p05491`.

[9] R. S. Selesnick, J. B. Blake, W. A. Kolasinski, and T. A. Fritz. A quiescent state of 3 to 8 mev radiation belt electrons. *Geophysical Research Letters*, 24(11):1343–1346, 1997. ISSN 1944-8007. doi: 10.1029/97GL51407. URL `http://dx.doi.org/10.1029/97GL51407`.

[10] D. H. Brautigam and J. M. Albert. Radial diffusion analysis of outer radiation belt electrons during the october 9, 1990, magnetic storm. *Journal of Geophysical Research: Space Physics*, 105(A1):291–309, 2000. ISSN 2156-2202. doi: 10.1029/1999JA900344. URL `http://dx.doi.org/10.1029/1999JA900344`.

[11] Yue Fei, Anthony A. Chan, Scot R. Elkington, and Michael J. Wiltberger. Radial diffusion and mhd particle simulations of relativistic electron transport by ulf waves in the september 1998 storm. *Journal of Geophysical Research: Space Physics*, 111(A12):n/a–n/a, 2006. ISSN 2156-2202. doi: 10.1029/2005JA011211. URL `http://dx.doi.org/10.1029/2005JA011211`. A12209.

[12] Yuri Y. Shprits, Nigel P. Meredith, and Richard M. Thorne. Parameterization of radiation belt electron loss timescales due to interactions with chorus waves. *Geophysical Research Letters*, 34(11):n/a–n/a, 2007. ISSN 1944-8007. doi: 10.1029/2006GL029050. URL `http://dx.doi.org/10.1029/2006GL029050`. L11110.

[13] J. Bartels, N. H. Heck, and H. F. Johnston. The three-hour-range index measuring geomagnetic activity. *Terrestrial Magnetism and Atmospheric Electricity (Journal of Geophysical Research)*, 44:411, 1939. doi: 10.1029/TE044i004p00411.

[14] Siamak Mehrkanoon and Johan A.K. Suykens. Learning solutions to partial differential equations using ls-svm. *Neurocomputing*, 159(Supplement C):105 – 116, 2015. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2015.02.013. URL `http://www.sciencedirect.com/science/article/pii/S0925231215001629`.

[15] Siamak Mehrkanoon and Johan A.K. Suykens. Ls-svm approximate solution to linear time varying descriptor systems. *Automatica*, 48(10):2502 – 2511, 2012. ISSN 0005-1098. doi: https://doi.org/10.1016/j.automatica.2012.06.095. URL `http://www.sciencedirect.com/science/article/pii/S0005109812003652`.

[16] Siamak Mehrkanoon, Saeid Mehrkanoon, and Johan A.K. Suykens. Parameter estimation of delay differential equations: An integration-free ls-svm approach. *Communications in Nonlinear Science and Numerical Simulation*, 19(4):830 – 841, 2014. ISSN 1007-5704. doi: https://doi.org/10.1016/j.cnsns.2013.07.024. URL `http://www.sciencedirect.com/science/article/pii/S1007570413003444`.

[17] E.J. Kansa. Multiquadrics—a scattered data approximation scheme with applications to computational fluid-dynamics—ii solutions to parabolic, hyperbolic and elliptic partial differential equations. *Computers & Mathematics with Applications*, 19(8):147 – 161, 1990. ISSN 0898-1221. doi: https://doi.org/10.1016/0898-1221(90)90271-K. URL `http://www.sciencedirect.com/science/article/pii/089812219090271K`.

[18] A. Aminataei and M.M. Mazarei. Numerical solution of poisson's equation using radial basis function networks on the polar coordinate. *Computers & Mathematics with Applications*, 56(11):2887 – 2895, 2008. ISSN 0898-1221. doi: https://doi.org/10.1016/j.camwa.2008.07.026. URL `http://www.sciencedirect.com/science/article/pii/S0898122108004616`.

[19] Yong Duan. A note on the meshless method using radial basis functions. *Computers & Mathematics with Applications*, 55(1):66 – 75, 2008. ISSN 0898-1221. doi: https://doi.org/10.1016/j.camwa.2007.03.011. URL `http://www.sciencedirect.com/science/article/pii/S089812210700332X`.

[20] Yong Duan and Yong-Ji Tan. A meshless galerkin method for dirichlet problems using radial basis functions. *Journal of Computational and Applied Mathematics*, 196(2):394 – 401, 2006. ISSN 0377-0427. doi: https://doi.org/10.1016/j.cam.2005.09.018. URL `http://www.sciencedirect.com/science/article/pii/S0377042705005832`.

[21] M. Elansari, D. Ouazar, and A. H.-D. Cheng. Boundary solution of poisson's equation using radial basis function collocated on gaussian quadrature nodes. *Communications in Numerical Methods in Engineering*, 17(7):455–464, 2001. ISSN 1099-0887. doi: 10.1002/cnm.419. URL `http://dx.doi.org/10.1002/cnm.419`.

[22] John W. Pearson. A radial basis function method for solving pde-constrained optimization problems. *Numerical Algorithms*, 64(3):481–506, Nov 2013. ISSN 1572-9265. doi: 10.1007/s11075-012-9675-6. URL `https://doi.org/10.1007/s11075-012-9675-6`.

[23] Martin Tillenius, Elisabeth Larsson, Erik Lehto, and Natasha Flyer. A scalable rbf–fd method for atmospheric flow. *Journal of Computational Physics*, 298:406 – 422, 2015. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2015.06.003. URL `https://www.sciencedirect.com/science/article/pii/S0021999115003824`.

[24] Ali Safdari-Vaighani, Alfa Heryudono, and Elisabeth Larsson. A radial basis function partition of unity collocation method for convection–diffusion equations arising in financial applications. *Journal of Scientific Computing*, 64(2):341–367, Aug 2015. ISSN 1573-7691. doi: 10.1007/s10915-014-9935-9. URL `https://doi.org/10.1007/s10915-014-9935-9`.

[25] Katharina Kormann and Elisabeth Larsson. A galerkin radial basis function method for the schrödinger equation. *SIAM Journal on Scientific Computing*, 35(6):A2832–A2855, 2013. doi: 10.1137/120893975. URL `https://doi.org/10.1137/120893975`.

[26] Katharina Kormann and Elisabeth Larsson. Radial basis functions for the timedependent schrödinger equation. *AIP Conference Proceedings*, 1389(1):1323–1326, 2011. doi: 10.1063/1.3637863. URL `http://aip.scitation.org/doi/abs/10.1063/1.3637863`.

[27] Bengt Fornberg and Natasha Flyer. *A primer on radial basis functions with applications to the geosciences*. SIAM, 2015.

[28] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN 026218253X.

[29] John Skilling. *Bayesian Solution of Ordinary Differential Equations*, pages 23–37. Springer Netherlands, Dordrecht, 1992. ISBN 978-94-017-2219-3. doi: 10.1007/978-94-017-2219-3_2. URL `https://doi.org/10.1007/978-94-017-2219-3_2`.

[30] Thore Graepel. Solving noisy linear operator equations by gaussian processes: Application to ordinary and partial differential equations. January 2003. URL `https://www.microsoft.com/en-us/research/publication/solving-noisy-linear-operator-equations-by-gaussian-processes-application-to-ordinary-and-partial-dif`

[31] Simo Särkkä. *Linear Operators and Stochastic Partial Differential Equations in Gaussian Process Regression*, pages 151–158. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-21738-8. doi: 10.1007/978-3-642-21738-8_20. URL `https://doi.org/10.1007/978-3-642-21738-8_20`.

[32] Frank Dondelinger, Dirk Husmeier, Simon Rogers, and Maurizio Filippone. Ode parameter inference using adaptive gradient matching with gaussian processes. In Carlos M. Carvalho and Pradeep Ravikumar, editors, *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, volume 31 of *Proceedings of Machine Learning Research*, pages 216–228, Scottsdale, Arizona, USA, 29 Apr–01 May 2013. PMLR. URL `http://proceedings.mlr.press/v31/dondelinger13a.html`.

[33] I. E. Lagaris, A. Likas, and D. I. Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. *IEEE Transactions on Neural Networks*, 9(5):987–1000, Sep 1998. ISSN 1045-9227. doi: 10.1109/72.712178.

[34] Lucie P. Aarts and Peter van der Veer. Neural network method for solving partial differential equations. *Neural Processing Letters*, 14(3):261–271, Dec 2001. ISSN 1573-773X. doi: 10.1023/A:1012784129883. URL `https://doi.org/10.1023/A:1012784129883`.

[35] Ioannis G. Tsoulos, Dimitris Gavrilis, and Euripidis Glavas. Solving differential equations with constructed neural networks. *Neurocomputing*, 72(10):2385 – 2391, 2009. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2008.12.004. URL `http://www.sciencedirect.com/science/article/pii/S0925231208005560`. Lattice Computing and Natural Computing (JCIS 2007) / Neural Networks in Intelligent Systems Designn (ISDA 2007).

[36] M. Baymani, S. Effati, and A. Kerayechian. A feed-forward neural network for solving stokes problem. *Acta Applicandae Mathematicae*, 116(1):55, Jul 2011. ISSN 1572-9036. doi: 10.1007/s10440-011-9627-5. URL `https://doi.org/10.1007/s10440-011-9627-5`.

[37] Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive metropolis algorithm. *Bernoulli*, 7(2): 223–242, 04 2001. URL `https://projecteuclid.org:443/euclid.bj/1080222083`.