ICIPE–2024–0095
Bayesian Variational Inference in the Inverse Biflux Anomalous Diffusion Problem

Douglas F. Corrêa
Josiele S. Teixeira
Instituto Politécnico - Universidade do Estado do Rio de Janeiro
douglas.correa@iprj.uerj.br

Carlos P. T. Zanini
Helio S. Migon
Universidade Federal do Rio de Janeiro
carloszanini@dme.ufrj.br, hsmigon@gmail.com

David A. Pelta
Universidad de Granada
dpelta@ugr.es

Claudio F. M. Toledo
Universidade de São Paulo
claudio@icmc.usp.br

Antônio J. Silva Neto
Universidade do Estado do Rio de Janeiro
ajsneto@iprj.uerj.br

Abstract. *This study presents an Inverse Problem in the Biflux Anomalous Diffusion context employing the Bevilacqua-Galeão Model. The numerical solution of the associated direct problem was obtained with the Finite Difference Method and the discretization scheme was detailed. Subsequently, the inverse problem was solved with the aid of the Bayesian Variational Inference Method. Furthermore, a brief comparative analysis of the results with those derived from alternative methodologies documented in the literature on the Bevilacqua-Galeão Model is presented. Additionally, a statistical evaluation of the results accompanied by insightful discussions about the practical implementation and the evaluation of the employed model is stated concisely.*

Keywords: *Inverse problem. Bayesian Variational Inference. Biflux Anomalous Diffusion. BG Model. Parameter estimation.*

1. INTRODUCTION

Diffusion processes are usually divided into three categories: subdiffusive, normal/linear, and superdiffusive (Evangelista and Lenzi, 2023). Normal phenomena are commonly modeled using Fick's diffusion equation (Fick, 1855) and are usually found in near-equilibrium situations, while phenomena in the other two categories are often reported in non-equilibrium situations, which are called anomalous diffusion.

Recently, a new model has been proposed to deal with anomalous diffusion problems. This new model proposed by Bevilacqua, Galeão, and collaborators (Bevilacqua *et al.*, 2011), denominated here the BG Model, was studied by Lugon Junior *et al.* (2020) who, using the second-moment method, showed that the BG Model can represent both subdiffusive and superdiffusive phenomena.

The present work deals with an inverse problem of estimating parameters in the BG model with synthetic experimental data, where the redistribution coefficient is estimated via Bayesian Variational Inference with the aid of the Differential Evolution algorithm.

2. DIRECT PROBLEM

The unidimensional anomalous bi-flux diffusion equation is written here as

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(\beta \Psi_D - (1-\beta)\Psi_R\right) \tag{1}$$

where Ψ_D is the well-known Fickian flux and Ψ_R is the retention/subsidiary flux from the BG Model.

$$\Psi_D = K_2 \frac{\partial c}{\partial x} \tag{2}$$

and

$$\Psi_R = \beta K_4 \frac{\partial^3 c}{\partial x^3} \tag{3}$$

replacing Eqs.(2) and (3) in Eq. (1) gives

$$\frac{\partial c}{\partial t} = \beta K_2 \frac{\partial^2 c}{\partial x^2} - (1-\beta)\beta K_4 \frac{\partial^4 c}{\partial x^4} \tag{4}$$

A more detailed explanation of the nature of the bi-flux process can be found in Jiang *et al.* (2018).

For the generation of synthetic data in an unidimensional domain of length L, the initial condition was

$$c(x, t=0) = sin(\pi \frac{x}{L}) \tag{5}$$

and the boundary condition was

$$c(x=0, t) = c(x=L, t) = 0 \tag{6}$$

and

$$\left.\frac{\partial^3 c}{\partial x^3}\right|_{x=0} = \left.\frac{\partial^3 c}{\partial x^3}\right|_{x=L} = 0 \tag{7}$$

2.1 Numerical Solution

With the conditions presented, the Direct Problem was numerically solved with the Finite Difference Method using the implicit scheme. The discretization choice was studied by Vasconcellos *et al.* (2017), who suggested the use of a discretization of order greater than $\mathcal{O}((\Delta x)^4)$ for the first and second order derivatives.

For the synthetic data used, random noise from a Gaussian distribution with a mean of 0 and standard deviation of 0.01 was added to the generated solution. Figure 1 shows the generated synthetic data used as the observed data in the inverse problem.

3. INVERSE PROBLEM

Bayes theorem is widely used in the literature on Inverse Problems to, based on observed data, update a *priori* knowledge over latent variables. This updated knowledge is referred to here as the *posteriori*. The Bayes theorem can be written as

$$p(\theta \,|\, c) = \frac{p(c \,|\, \theta)p(\theta)}{p(c)} \tag{8}$$

where c is the observed data, θ is the vector of latent variable, $p(c|\theta)$ is the likelihood, $p(\theta)$ the *priori* distribution of θ and $p(c)$ is the analytically unavailable marginal likelihood of c. Since $p(\theta \,|\, c)$ is not easily obtained, the Variational Inference approach presents as an alternative: finding a distribution $q_{\boldsymbol{\Psi}}(\theta)$ that best approximates $p(\theta \,|\, c)$ and as a consequence also approximates the solution of the inverse problem.

3.1 Bayesian Variational Inference

The measure of the information loss when $q_{\boldsymbol{\Psi}}(\theta)$ is used to approximate $p(\theta \,|\, c)$ is evaluated with the Kullback-Leibler Divergence, written here as

$$D_{KL}\left(q_{\boldsymbol{\Psi}}(\theta) \,||\, p(\theta|c)\right) = \int_{-\infty}^{\infty} q_{\boldsymbol{\Psi}}(\theta) \log \frac{q_{\boldsymbol{\Psi}}(\theta)}{p(\theta \,|\, c)} d\theta \tag{9}$$

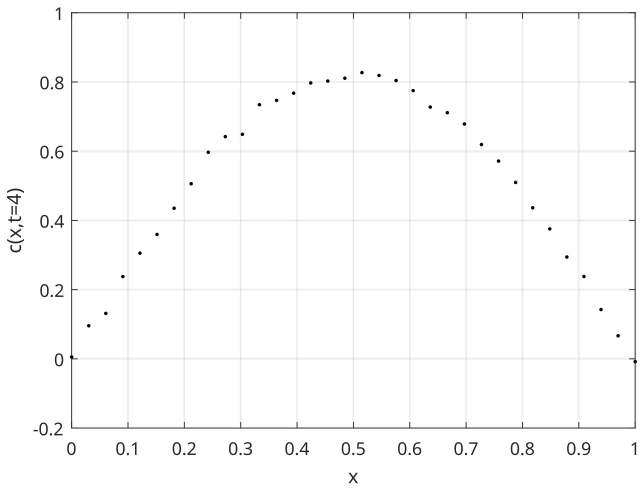


Figure 1: Experimental data generated with the solution of the direct problem with $\beta = 0.5$ and noise added from a Gaussian distribution with mean 0 and standard deviation 0.01.

with the Bayes theorem and some algebraic manipulation, we can get

$$D_{KL}\left(q_{\mathbf{\Psi}}(\theta)\,||\,p(\theta\,|\,c)\right) = -\int_{-\infty}^{\infty} q_{\mathbf{\Psi}}(\theta)\log\frac{p(c\,|\,\theta)p(\theta)}{q_{\mathbf{\Psi}}(\theta)}d\theta + \int_{-\infty}^{\infty} q_{\mathbf{\Psi}}(\theta)\log p(c)d\theta \tag{10}$$

since $p(c)$ does not depend on θ and $q_{\mathbf{\Psi}}(\theta)$ is a probability distribution we can write the above equation as

$$D_{KL}\left(q_{\mathbf{\Psi}}(\theta)\,||\,p(\theta\,|\,c)\right) = -\int_{-\infty}^{\infty} q_{\mathbf{\Psi}}(\theta)\log\frac{p(c\,|\,\theta)p(\theta)}{q_{\mathbf{\Psi}}(\theta)}d\theta + \log p(c) \tag{11}$$

the focus of the present work is now to find the variational parameters that characterize $q_{\mathbf{\Psi}}(\theta)$ and minimizes Eq. (11), more specifically since $p(c)$ is constant we approach it as finding optimum values of $\mathbf{\Psi}$ that maximizes the integral term on the right-hand side of Eq. (11). This Integral term is known in the literature as the Evidence Lower Bound (ELBO) and numerically solved via Monte Carlo and Gauss-Hermite for validation purposes. In the present work, we assume that the *posteriori* can be approximated by a Gaussian distribution which allows us to define the vector of variational parameters as

$$\mathbf{\Psi} = (m, L) \tag{12}$$

where m is the mean and LL^T is the standard deviation of the *posteriori* approximation.

3.2 Transformed Parameter Space

Since we are trying to characterize the physical parameters by a distribution, a transformation is needed for the parameters to lie on $\mathbb{R}^d$. As the latent variable of interest from the BG Model is β it is clear that $\theta \equiv \beta$ and due to the bi-flux nature of the process $\beta \in (0, 1)$ then one possible transformation is

$$\xi = T(\theta) = \log\frac{\theta}{1-\theta} \tag{13}$$

where T is a differentiable and invertible transformation operator that puts the parameter θ in the $\mathbb{R}^d$ space.

As a consequence, the resulting probability needs to be scaled by the derivative of the inverse transformation, and then Eq. (11) becomes

$$D_{KL}\left(q_{\mathbf{\Psi}}(\xi)\,||\,p(\xi\,|\,c)\right) = -\int_{-\infty}^{\infty} q_{\mathbf{\Psi}}(\xi)\log\frac{p(c\,|\,\theta)p(\theta)}{q_{\mathbf{\Psi}}(\theta)}\bigg|_{\theta=T^{-1}(\xi)} \times |J_{T^{-1}(\xi)}|d\xi + \log p(c) \tag{14}$$

and by reparameterizing $\xi = m + L\epsilon$, the same way as done by Zanini *et al.* (2022), we can approximate the ELBO with Monte Carlo as

$$ELBO \approx \frac{1}{M}\sum_{l=1}^{M}\left[\log p(c, T^{-1}(\xi)) + \log|J_{T^{-1}(\xi)}| - logN(\xi; m, LL^T)\big|_{\xi=m+L\epsilon}\right] \tag{15}$$

where $\epsilon^{(l)} \overset{iid}{\sim} \mathcal{N}(0, I), l = 1, ..., M.$

3.3 Numerical Integration

When writing the integral in Eq. (14) as an expectation over $\theta \sim q_\Psi(\theta)$, the result can be obtained via Monte Carlo, here we make a brief comparison of this approach with the use of Gauss Hermite Method. Using the Monte Carlo Method

Table 1: Relative error obtained with the Numerical Methods tested in comparison with Monte Carlo with 100.000 samples, considered here as the exact answer.

Method	Relative Error (%)
Monte Carlo (M = 5 samples)	1.233471
Monte Carlo (M = 50 samples)	1.141598
Monte Carlo (M = 500 samples)	0.900000
Gauss-Hermite (7 points)	1.034771
Gauss-Hermite (12 points)	0.713791

with 5 samples the ELBO calculation was pretty accurately calculated and due to its simplicity it was our choice of solution.

3.4 Differential Evolution

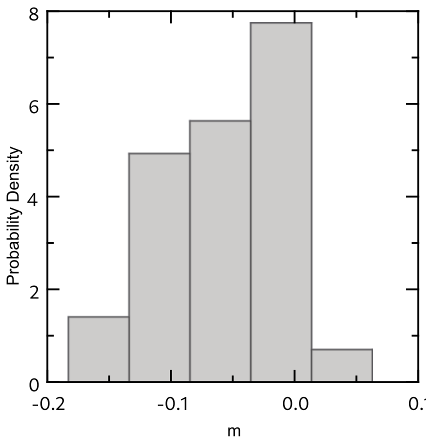
To maximize the ELBO the Differential Evolution(DE) Algorithm developed by Storn and Price (1997) was used since, as a metaheuristics, it allows to obtain a set of potentially near-optimal solutions when used as a solution generator Raoui *et al.* (2021). With that in mind the objective function is the ELBO itself and the vector of parameters to be optimized by the DE Algorithm are the variational parameters, that is, the vector $\boldsymbol{\Psi}$. Our implemented DE Algorithm can be described as

1. Generation of an initial random population of size N_{pop};
2. Mutation operation between random members of the population for generating test candidates, with a mutation factor of 0.35;
3. Crossover operation with p_{cross} probability of occurrence defined, for the present study, as 0.7.
4. If the new member provides a better value for the objective function than the latter it is replaced in the next generation, otherwise it remains in the population for one more generation, this is known as elitism.
5. Repeat steps 2-4 until the difference in the objective function is less than 10^{-4}

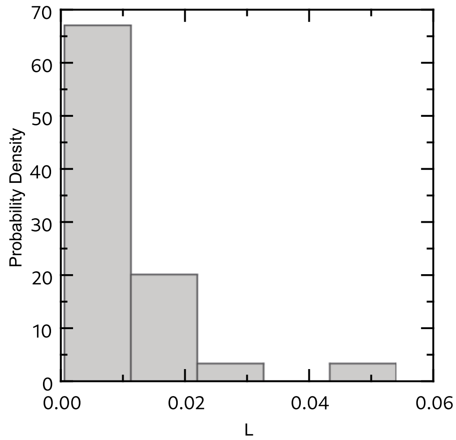
4. RESULTS

Figure 2 shows the results obtained after 30 executions of the DE Algorithm, with the same parameters and conditions and with M=5 samples from the Monte Carlo Method while Figure 4 shows the results obtained with M=20 and Figure 3 with M=10. Only 20 of the 30 executions were used to generate the histogram. The main difference between executions was the initial population from the DE Algorithm that was randomized each time.

Figure 5 presents the updated knowledge (*posteriori*) over β when the synthetic data is observed, that is $p(\theta|c)$. If a point estimation were the main goal then the result absolute error between the exact value and the estimation would have been 12×10^{-3}.

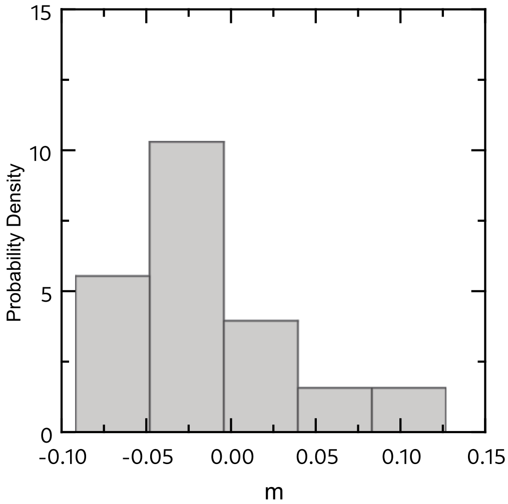


(a) Histogram of m from the 20 best candidates from the Differential Evolution Algorithm.

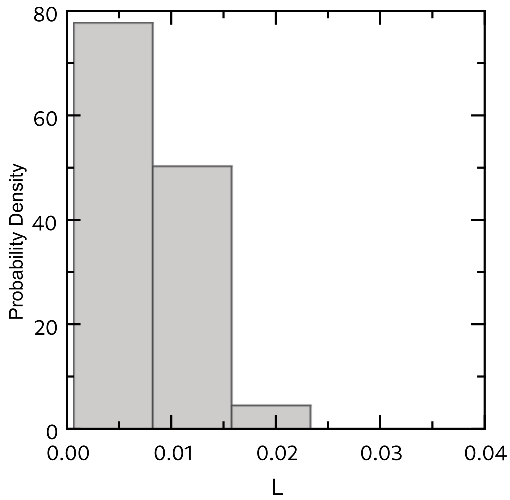


(b) Histogram of L from the 20 best candidates from the Differential Evolution Algorithm.

Figure 2: Histograms from results obtained with $M = 5$ after 30 executions of the DE Algorithm.

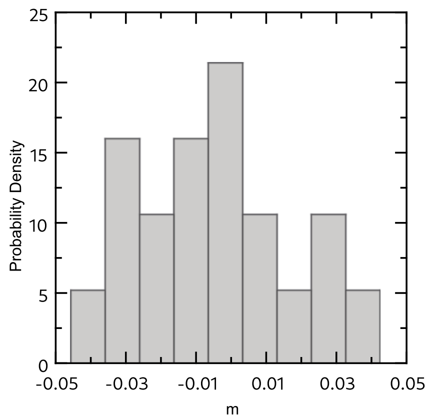


(a) Histogram of m from the 20 best candidates from the Differential Evolution Algorithm.

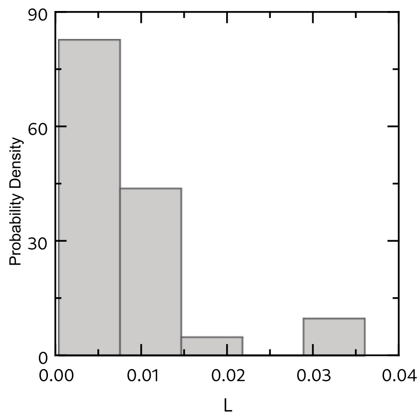


(b) Histogram of L from the 20 best candidates from the Differential Evolution Algorithm.

Figure 3: Histograms from results obtained with $M = 10$ after 30 executions of the DE Algorithm.



(a) Histogram of m from the 20 best candidates from the Differential Evolution Algorithm.



(b) Histogram of L from the 20 best candidates from the Differential Evolution Algorithm.

Figure 4: Histograms from results obtained with $M = 20$ after 30 executions of the DE Algorithm.

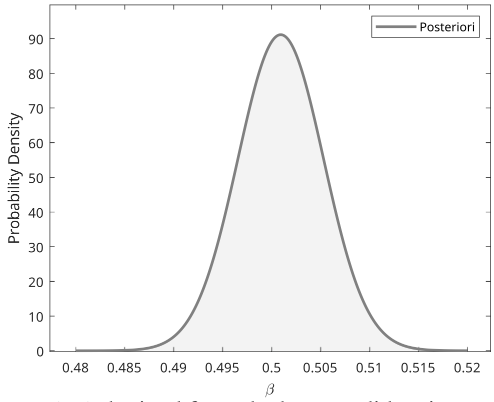


Figure 5: The approximate *posteriori* obtained from the best candidate in terms of objective function value

5. CONCLUDING REMARKS

The results obtained showed that the method was able to estimate a very good approximation for the *posteriori* of β, the best result according to the objective function value was with a Gaussian distribution with mean -0.001, that when the inverse transformation is applied gives a mean of 0.50012 which means a small absolute error. The Bayesian Variational Inference approach showed good potential as a method to estimate properties and parameters in the BG Model since not only the exact parameter was inside the 95% confidence interval of the *posteriori*.

For future works we aim to do a deep analysis of the method and comparisons, highlighting its strengths and weaknesses against, for example, the Monte Carlo Markov Chain with the Metropolis Hasting algorithm.

6. ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

A. J. Silva Neto also acknowledges the financial support provided by SENAI CIMATEC.

D. Corrêa also acknowledges CAPES for the scholarship for his stay at the University of Granada (Grant CAPES/PrInt No. 88887.717186/2022-00) and the MODO Group from UGR for the collaboration.

D. Pelta acknowledges support from projects PID2020-112754GB-I00, MCIN /AEI/ 10.13039 /501100011033 and FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades / Proyecto (BTIC-640-UGR20)

7. REFERENCES

Bevilacqua, L., Galeão, A.C. and Costa, F.P., 2011. “A new analytical formulation of retention effects on particle diffusion processes”. *Anais da Academia Brasileira de Ciências*, Vol. 83, No. 4, pp. 1443–1464. doi:10.1590/s0001-37652011005000033. URL `https://doi.org/10.1590/s0001-37652011005000033`.

Evangelista, L.R. and Lenzi, E.K., 2023. *An Introduction to Anomalous Diffusion and Relaxation*. Springer International Publishing. ISBN 9783031181504. doi:10.1007/978-3-031-18150-4. URL `http://dx.doi.org/10.1007/978-3-031-18150-4`.

Fick, A., 1855. “Ueber diffusion”. *Annalen der Physik*, Vol. 170, No. 1, p. 5986. ISSN 1521-3889. doi: 10.1002/andp.18551700105. URL `http://dx.doi.org/10.1002/andp.18551700105`.

Jiang, M., Bevilacqua, L., Neto, A.J.S., Galeão, A.C.R. and Zhu, J., 2018. “Bi-flux theory applied to the dispersion of particles in anisotropic substratum”. *Applied Mathematical Modelling*, Vol. 64, pp. 121–134. doi: 10.1016/j.apm.2018.07.022. URL `https://doi.org/10.1016/j.apm.2018.07.022`.

Lugon Junior, J., Vasconcellos, J.F.V., Knupp, D.C., Marinho, G.M., Bevilacqua, L. and da Silva Neto, A.J., 2020. “Solution of fourth order diffusion equations and analysis using the second moment”. *Defect and Diffusion Forum*, Vol. 399, p. 1020. ISSN 1662-9507. doi:10.4028/www.scientific.net/ddf.399.10. URL `http://dx.doi.org/10.4028/www.scientific.net/DDF.399.10`.

Raoui, H.E., Cabrera-Cuevas, M. and Pelta, D.A., 2021. “The role of metaheuristics as solutions generators”. *Symmetry*, Vol. 13, No. 11, p. 2034. doi:10.3390/sym13112034. URL `https://doi.org/10.3390/sym13112034`.

Storn, R. and Price, K., 1997. *Journal of Global Optimization*, Vol. 11, No. 4, pp. 341–359. doi: 10.1023/a:1008202821328. URL `https://doi.org/10.1023/a:1008202821328`.

Vasconcellos, J., Marinho, G. and Zani, J., 2017. “Análise numérica da equação da difusão anômala com fluxo bimodal”. *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería*, Vol. 33, No. 34, p. 242249. ISSN 0213-1315. doi:10.1016/j.rimni.2016.05.001. URL `http://dx.doi.org/10.1016/j.rimni.2016.05.001`.

Zanini, C.T.P., Migon, H.d.S. and Dias, R., 2022. “Variational inference for bayesian bridge regression”. doi: 10.48550/ARXIV.2205.09515. URL `https://arxiv.org/abs/2205.09515`.

8. RESPONSIBILITY NOTICE

The authors are solely responsible for the printed material included in this paper.