International Journal of Applied Mathematics

Volume 33 No. 5 2020, 783-803
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
doi: http://dx.doi.org/10.12732/ijam.v33i5.4

USING THE BAYESIAN FRAMEWORK FOR INFERENCE IN FRACTIONAL ADVECTION-DIFFUSION TRANSPORT SYSTEM

Edward Boone[1], Ryad Ghanam[2 §],
Nadeem Malik[3], and Joshua Whitlinger[4]

[1] Dept. of Statistical Sci. and Operations Research
Virginia Commonwealth University
Richmond, VA 23284, USA

[2] Dept. of Liberal Arts and Sciences
Virginia Commonwealth University in Qatar
Doha, QATAR

[3] Dept. of Mechanical Engineering
Texas Tech University
Lubbock, TX 74909, USA

[4] Dept. of Statistical Sci. and Operations Research
Virginia Commonwealth University
Richmond, VA 23284, USA

Abstract: This work shows for the first time the viability of using the Bayesian paradigm for both estimation and hypothesis testing when applied to fractional differential equations. Two distinct fractional differential equation models were explored using simulated data sets to determine the performance of the Bayesian inferential methods across values of α (the fractional order) and σ (the experimental error variance). This inferential paradigm shows promise as it has robust estimation, predictions and provides for hypothesis testing to determine whether a fractional process is warranted by the data. A simulation study, applied to a fractional transport system in porous media, demonstrates the robustness of the estimation and the sensitivity of the hypothesis tests to various levels of α and σ^2.

Received: May 18, 2020

§Correspondence author

AMS Subject Classification: 35R11, 35C05, 35C10, 35E15, 35G25, 65M15, 65G99
Key Words: Bayesian estimation; validation; modeling error; analytical approximate solution

1. Introduction

Many physical systems are modeled by partial differential equations, [1] and [2]; applications of such equations arise in many fields, such as, atmospheric chemistry, [3], modeling of fluid flow in homogeneous media, [4], combustion, [5], and porous media [6]. In a single-phase system, let us denote the scalar (typically the concentration) by $p(x,t)$ at the position x and at the time instant t, and let the scalar be transported by the velocity field $u(x,t)$, and by diffusion due to gradients in $c(x,t)$ and the fluxes across the boundaries of the region. Many different types of advection-diffusion arise in different systems. One such system arises in application to flow of gas in porous media, such as shale rocks. Here the system is described by the advection-diffusion equation,

$$\frac{\partial p(x,t)}{\partial t} + U(p,p_x)\frac{\partial}{\partial x}\left(p(x,t)\right)) = \frac{\partial}{\partial x}\left(D(p)\frac{\partial p}{\partial x}\right), \tag{1}$$

where U is the convective velocity, and D is the diffusivity is the non-linear diffusion flux, see [6].

But there are many phenomena in nature which are not described adequately by these models. Among them are crowded systems, such as protein diffusion within cells, [7], and diffusion through porous media, [8]. Here fractional differential equations sometimes provide a better description, see [9], [10]. Though fractional calculus (FC) has a long history, going back 300 years, its application in many fields of science and engineering is relatively new [8]–[18].

A fractional differential equation contains a fractional derivative in at least one of its terms. For example, a fractional diffusion equation reads,

$$\frac{\partial^\alpha}{\partial t^\alpha}\left(\frac{\partial p}{\partial t}\right) = \frac{\partial^{1+\alpha} p}{\partial t^{1+\alpha}} = A\left(\frac{\partial^2 p}{\partial x^2}\right), \tag{2}$$

where $\partial^\alpha/\partial t^\alpha$ represents the Caputo derivative and A is the pseudo-diffusivity defined in [19]. Fractional calculus methods have the ability to represent non-Gaussian continuous time random walk (CTRW) statistical processes [20], [21] which leads to so-called anomalous diffusion, see Metzler and Klafter [22]. It is possible to apply a spatial fractional derivative, discussed in [23].

Our interest in this paper is in estimation and hypothesis testing applied to fractional fluid transport models. We will study a generalization of equation (1) to a factional advection-diffusion equation with the Caputo derivative applied to the time derivative,

$$\frac{\partial^\alpha p(x,t)}{\partial t^\alpha} = \frac{\partial}{\partial x}\left(D(p)\frac{\partial p}{\partial x}\right) - U(p,p_x)\frac{\partial p}{\partial x}, \quad x>0,\ t>0, \tag{3}$$

where $p(x,t)$ is the pressure in tight gas reservoir.

Bayesian statistical methods have been developing rapidly for the last 40 years due to the advent of fast computing. As opposed to Frequentist methods such as Method of Moments [35] or Maximum Likelihood [36] which obtain information only from the data, Bayesian methods consider incorporating all sources of information including the data and any information available prior to experimentation. In the presented context, one has prior information about α, namely $0 < \alpha \leq 1$. All prior information should be incorporated into a *prior* probability distribution so that any uncertainties about α are reflected. Using the prior probability distribution with the Likelihood of the data one can obtain a *posterior* probability distribution via Bayes' theorem. Since posterior probability distributions are typically not analytically tractible, sampling techniques such as Markov chain Monte Carlo (MCMC) are used to obtain samples from the posterior probability distribution. All inferences are made using the samples from the posterior probability distribution. For more on the Bayesian framework and inferences, see [25],[26]

This work is important because it establishes the viability of using the Bayesian paradigm for both estimation and hypothesis testing when applied to fractional differential equations. This work is significant also because we have not found any previous publications directly in this field, although [27] report work in a related field where they focus only on using the Bayesian technique to estimate the fractional order in a growth model and only through time. As such, the work reported here can be considered pioneering as it brings together fractional partial differential equations and statistical techniques for the first time.

The remainder of this work is organized as follows. Section 2 gives the problem statement, model definitions and statistical model. A simulated example is given in Section 3 to illustrate the methodology in a single case example. This is followed in Section 4 by a large scale simulation study on the robustness of the estimation procedure across a variety of α and σ^2 specifications. The details of conducting hypothesis tests are presented in Section 5 with a simulation study illustrating the sensitivity of the hypothesis test to varying values of α and σ^2.

And finally in Section 6 a discussion of the results are given as well as some issues for future work in the area.

2. The problem statement

We will focus on two particular time-fractional advection-diffusion systems given by (3), [24], [6]. Our task is to provide inference on such parameters for the fraction and the experimental error variance using the Bayesian framework. We will explore two systems which are relevant to applications in transport in porous media,

$$\frac{\partial^\alpha p}{\partial t^\alpha} = \frac{\partial}{\partial x}\left(D\frac{\partial p}{\partial x}\right) - U\left(\frac{\partial p}{\partial x}\right), \quad t > 0,\ x > 0. \tag{4}$$

In order to estimate the fractional parameter α and the experimental error variance σ^2 we will use a likelihood approach, namely the Bayesian framework, hence a likelihood needs to be created as:

$$Y(x_i, t_j) = p(x_i, t_j)\epsilon(x_i, t_j), \tag{5}$$

where $Y(x_i, t_j)$ is the stochastic process with mean $p(x_i, t_j)$ at location x_i and time t_j. In this case the experimental error $\epsilon(x_i, t_j)$ will follow some appropriate probability distribution with variance σ^2.

2.1. Models

Model 1.
The first model that we consider comes from equation (3) where we take $D = 1$ and $U = 1$ and the initial condition is $p(x, 0) = e^{-cx}$, $(c > 0)$, and the boundary condition is $p(x, t) \to 0$ as $x \to \infty$. This yields the linear system

$$\frac{\partial^\alpha p}{\partial t^\alpha} = \frac{\partial^2 p}{\partial x^2} - \frac{\partial p}{\partial x}, \qquad t > 0,\ x > 0. \tag{6}$$

For this work we will assume $c = 1$ for simplicity. In this case $p(x, t)$ has a closed form solution under some mild conditions given by a Mittag-Leffler function:

$$p(x, t) = e^{-x} \sum_{k=0}^{\infty} \frac{2t^{\alpha k}}{\Gamma(\alpha k + 1)}. \tag{7}$$

Model 2.
The second model that we consider comes from equation (3) where we take $D(p) = \frac{1}{p}$ and $U(p, p_x) = \frac{1}{p}\frac{\partial p}{\partial x}$, with the initial condition $p(x, 0) = e^{-x}$, and the boundary condition $p(x, t) \to 0$ as $x \to \infty$. This yields the non-linear system

$$\frac{\partial^\alpha p}{\partial t^\alpha} = \frac{\partial}{\partial x}\left(\frac{1}{p}\frac{\partial p}{\partial x}\right) - \frac{1}{p}\left(\frac{\partial p}{\partial x}\right)^2. \tag{8}$$

In this case $p(x, t)$ has a closed form solution under some mild conditions which is given again by a Mittag-Lefler function:

$$p(x, t) = e^{-x}\sum_{k=0}^{\infty}\frac{(-t^\alpha)^k}{\Gamma(\alpha k + 1)}, \qquad t > 0,\ x > 0. \tag{9}$$

The main parameter in this model is α, the fraction of the derivative, and is bound between 0 and 1. If $\alpha = 1$ then no fractionation is needed.

2.2. Statistical model

A Bayesian approach is used to estimate the parameters in the model which requires that both a likelihood be specified as well as prior distributions on the model parameters. For **Model 1** the following likelihood model is specified as:

$$Y(x_i, t_j) = p(x_i, t_j)\epsilon(x_i, t_j) = \left[e^{-x_i}\sum_{k=0}^{\infty}\frac{2t_j^{\alpha k}}{\Gamma(\alpha k + 1)}\right]\epsilon(x_i, t_j). \tag{10}$$

And similarly, **Model 2** has the following likelihood model specification:

$$Y(x_i, t_j) = p(x_i, t_j)\epsilon(x_i, t_j) = \left[e^{-x_i}\sum_{k=0}^{\infty}\frac{\left(-t_j^\alpha\right)^k}{\Gamma(\alpha k + 1)}\right]\epsilon(x_i, t_j), \tag{11}$$

where $\epsilon(x_i, t_j) \overset{iid}{\sim} LogNormal(1, \sigma^2)$. Here the $LogNormal$ likelihood is chosen to ensure that pressure is always a positive value. Let $y(x_i, t_j)$ be the observed value of $Y(x_i, t_j)$ where $i = 1, ..., n_x$ and $j = 1, ..., n_t$. In this specification, the likelihood has two parameters, α and σ^2. The prior distribution for α and σ^2, $\pi(\alpha, \sigma^2)$ are specified as $\alpha \sim Beta\,(\alpha^*, \beta^*)$ to reflect the prior knowledge that α is bound between 0 and 1. For σ^2 the prior distribution is specified as $\sigma^2 \sim \chi^2(df)$ to reflect the prior knowledge that σ^2 must be a positive value. For more on prior distribution and selection, see [25].

For notation, let $\mathbf{x}$ be all x_i values and $\mathbf{t}$ be all the values of t_j and $y(\mathbf{x}, \mathbf{t})$ be all the corresponding values of $y(x_i, t_j)$ observed and $p(\mathbf{x}, \mathbf{t})$ be all the solutions at the corresponding values. The posterior distribution $\pi\left(\alpha, \sigma^2 | p(\mathbf{x}, \mathbf{t}), y(\mathbf{x}, \mathbf{t})\right)$ can be found using Bayes' Theorem [34]:

$$\pi\left(\alpha, \sigma^2 | p(\mathbf{x}, \mathbf{t}), y(\mathbf{x}, \mathbf{t})\right) = \frac{\pi(\alpha, \sigma^2) L(y(\mathbf{x}, \mathbf{t}) | p(\mathbf{x}, \mathbf{t}), \alpha, \sigma^2)}{\int \pi(\alpha, \sigma^2) L(y(\mathbf{x}, \mathbf{t}) | p(\mathbf{x}, \mathbf{t}), \alpha, \sigma^2) d\alpha d\sigma^2}. \tag{12}$$

In the case considered here there is no analytic solution to

$$\pi\left(\alpha, \sigma^2 | p(\mathbf{x}, \mathbf{t}), y(\mathbf{x}, \mathbf{t})\right),$$

and hence sampling method must be employed to draw samples from the posterior distribution, from which all inferences will be made. There are many choices for the algorithm to sample from the posterior distribution such as Acceptance Sampling, Metropolis-Hastings Sampling, Sampling Importance Resampling, etc. For more on sampling algorithms see [28], [26], [29].

3. Simulated examples

Suppose the system of interest is given by (4) where $\alpha = 0.82$. Further suppose that data for $y(\mathbf{x}, \mathbf{t})$ has been observed, with noise, at all combinations of $n_x = 31$ equally spaced levels of x from 0.01 to 10 and $n_t = 11$ equally spaced times t from 0.5 to 1.5 and the noise is multiplicative following a LogNormal distribution with mean 1 and $\sigma = 0.1$. Figure 1 shows the unperturbed data surface in panel (a) and the perturbed data surface in panel (b). Notice that this set parameter specification produces a quite noisy surface. Similarly for the system given by (4) the unperturbed surface is in panel (c) and perturbed surface in panel (d). In real world situations we would expect to observe perturbed data similar to those in panel (b) and (d). The goal of this work is to determine if statistical techniques can be used to adequately estimate the parameters of the underlying surface when presented with noisy data.

The prior distributions for α and σ were specified as follows for both Model 1 and Model 2:

$$\begin{aligned} \alpha &\sim Beta(3,3), \\ \sigma^2 &\sim \chi^2(1). \end{aligned}$$

Using the LogNormal likelihood, the prior above and Bayes Formula the posterior distribution is given as:

$$\begin{aligned} \pi\left(\alpha, \sigma | y(x_i, t_i), p(x_i, t_i)\right) &\propto \alpha^{\alpha^*}(1-\alpha)^{\beta^*}(\sigma^2)^{k/2-1} e^{-\sigma^2/2} \\ &\times e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{nx}\sum_{j=1}^{n_t}(\ln y(x_i,t_j) - p(x_i,t_j))^2} \end{aligned}$$

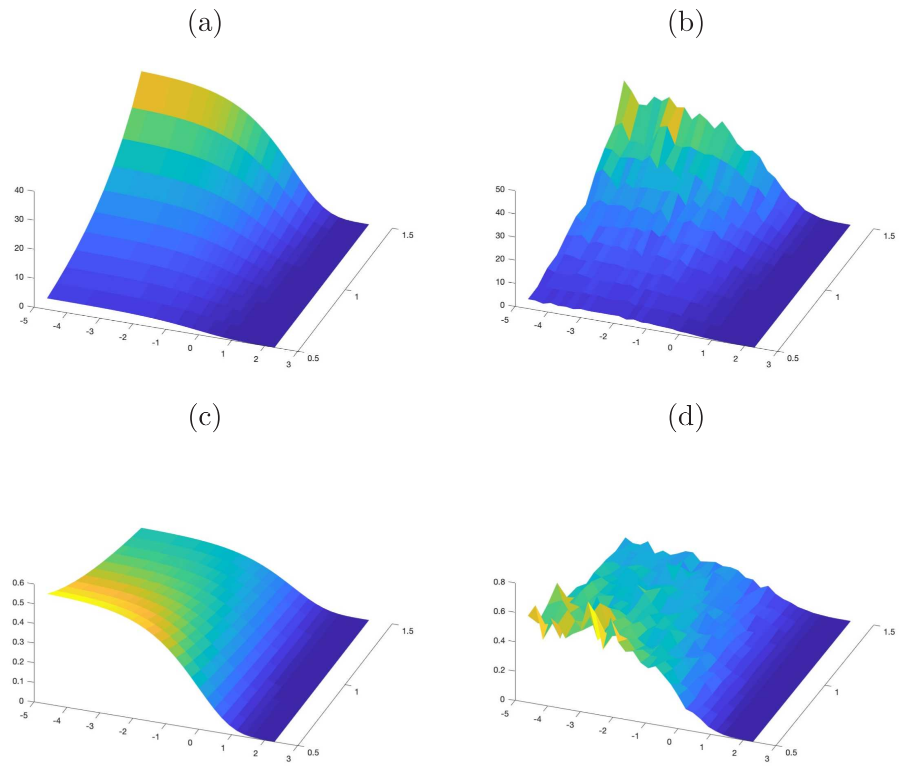

Figure 1: Model 1 ((4)) with no noise (a), $\alpha = 0.82$ and with noise $\sigma = 0.1$ (b) and Model 2 ((4)) with no noise (c), $\alpha = 0.82$ and with noise $\sigma = 0.1$(d).

$$
\begin{aligned}
&\times \quad \sigma^{-n_x-n_t} \prod_{i=1}^{n_x} \prod_{j=1}^{n_t} y(x_i, t_j)^{-1} \qquad (13) \\
&= \quad \alpha^3 (1-\alpha)^3 (\sigma^2)^{-1/2} e^{-\sigma^2/2} \\
&\times \quad e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_t} (\ln y(x_i,t_j) - p(x_i,t_j))^2} \\
&\times \quad \sigma^{-n_x-n_t} \prod_{i=1}^{n_x} \prod_{j=1}^{n_t} y(x_i, t_j)^{-1}.
\end{aligned}
$$

The Sampling Importance Resampling algorithm was employed with 10,000 candidate samples with 1,000 posterior samples drawn. The sampler generated a posterior sample of 89.8% unique samples indicating a high quality sample from the posterior distribution. Histograms of the posterior distributions of α and σ can be found in Figure 2, which shows the true values $\alpha = 0.82$ and $\sigma = 0.1$ in the middle of the distribution. For reference, the 95% posterior credible intervals were generated by taking the 2.5% and 97.5% quantiles from the marginal posterior distributions. This procedure gave parameter estimates for Model 1, α, $(0.8169, 0.8226)$, which contains the true value 0.82, and for σ, $(0.0907, 0.1050)$, which also contains the true value of 0.1. Similarly for Model 2 the credible interval for α is $(0.8020, 0.8389)$ and for σ, $(0.0928, 0.1083)$, both of which contain the true values. This this gives preliminary evidence that the procedure may be able to adequately estimate the model parameters.

Not only can the method proposed be used to estimate the fraction of the differential equation, it can also be used to quantify the prediction uncertainty associated with the model and parameter estimates. To do this the posterior predictive distribution can be employed to generate a distribution for a new observation $p(x_{new}, t_{new})$ at the value of x_{new} and t_{new}. Recall, the posterior predictive distribution is given by:

$$\begin{aligned}\pi\left(p(x^*, t^*)|y(\mathbf{x}, \mathbf{t}), p(\mathbf{x}, \mathbf{t})\right) &= \int \pi(\alpha, \sigma^2|y(\mathbf{x}, \mathbf{t}), p(\mathbf{x}, \mathbf{t})) \\ &\times\ L(p(x^*, t^*)|y(\mathbf{x}, \mathbf{t}), p(\mathbf{x}, \mathbf{t}), \alpha, \sigma^2) \\ &\times\ d\alpha d\sigma^2. \qquad (14)\end{aligned}$$

Across the domain of the inputs profile plots are created of the median surface, the surfaces generated by the 2.5% and 97.5% quantiles for given values of t and given values of x along with the data. Figures 3 and 4 give these profile plots for x and t, respectively for both Model 1 and Model 2. Notice that the posterior predictive intervals capture most of the observed data. This gives evidence that the modeling approach is properly quantifying the uncertainties associated with both estimation as well as inherent noise in the data. Furthermore, since the predictive intervals performance is similar across Model 1 and Model 2, this is another piece of evidence that the approach may be viable across a wide variety of fractional differential equations.

4. Robustness in estimation

In order to determine if the approach proposed for estimating α is robust to the value of α in the underlying process a robustness analysis was conducted for varying values of α and σ. To study this new datasets were simulated using each combination of $\alpha = 0.1, 0.25, 0.5, 0.75$ and 0.9 and $\sigma = 0.01, 0.1$

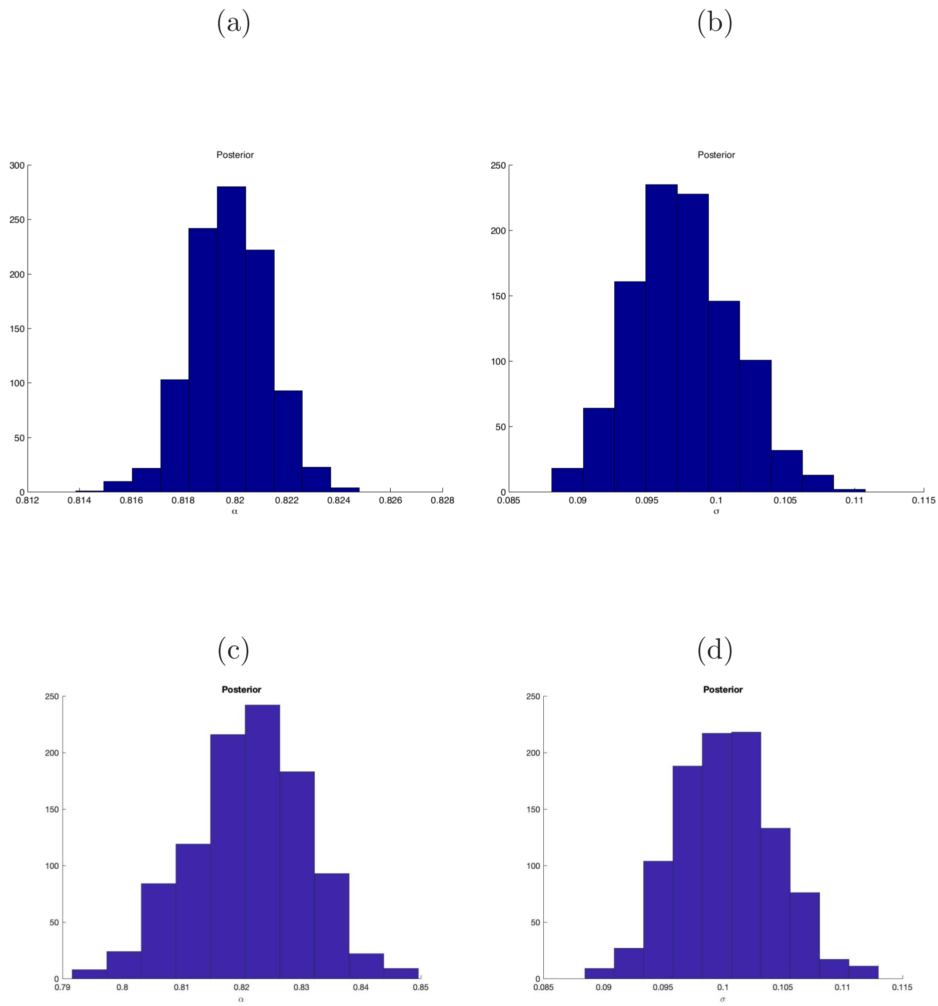

Figure 2: Histograms of the marginal posterior distributions for Model 1 of α (a) and σ (b) and for Model 2 of α (c) and σ (d).

and 0.25. A total of 200 datasets were simulated for each (α, σ) combination and using the Metropolis Sampler 1,500 posterior distribution samples were generated from which ht first 500 sample were discarded as “burn in”. The remaining 1,000 samples were used to make inferences about the parameters by calculating a posterior credible interval using the 2.5% and 97.5% quantiles of the samples for each parameter. The overall coverage probability was calculated

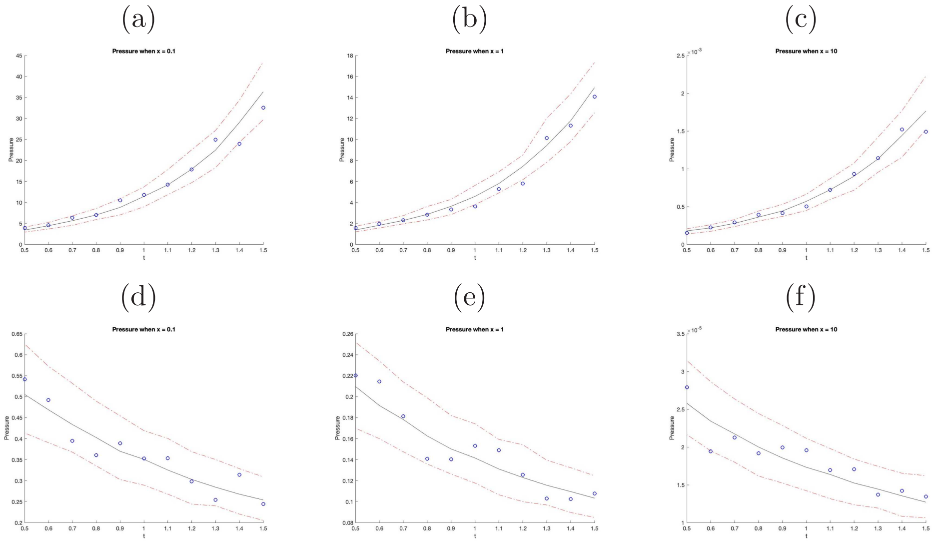

Figure 3: Profiles of predictive distribution for Model 1 with 95% predictive intervals across t. Data values ($\circ$), median ($-$), $Q_{2.5}$ and $Q_{97.5}$ ($- \cdot -$). Panels (a), (b) and (c) correspond to Model 1 profiles when $x = 0.1, 1.0$ and 10, respectively. Panels (d), (e) and (f) correspond to Model 2 profiles when $x = 0.1, 1.0$ and 10, respectively.

for each parameter combination, Coverage(α) and Coverage(β). Similarly, for each (α, σ) combination the mean width of credible interval, Width(α) and Width(σ) was calculated.

Table 4 shows these results for **Model 1** and Table 4 shows the results for **Model 2**. The results for both **Model 1** and **Model 2** show quite narrow interval widths relative to the parameter size. For example, in **Model 1** when $\alpha = 0.25$ and $\sigma = 0.01$ the width of the associated interval for α is 0.0005 and for σ is 0.0014. Also notice that as σ increases the width of the intervals for both α and σ also increase, as expected. In Table 4 one can see that the intervals for α are much wider when σ is large compared for **Model 2** to those from **Model 1**. In terms of coverage probabilities, both **Model 1** and **Model 2** perform reasonably well with proportion of intervals capturing the true value close to the preset 95% level. There are some lower results when σ is large. For example, in **Model 1** when $\alpha = 0.9$ and $\sigma = 0.25$ the coverage probability for α is only 0.865 and for **Model 2** when $\alpha = 0.1$ and $\sigma = 0.25$ the coverage probability for α is only 0.870. Based on these results the estimation procedure

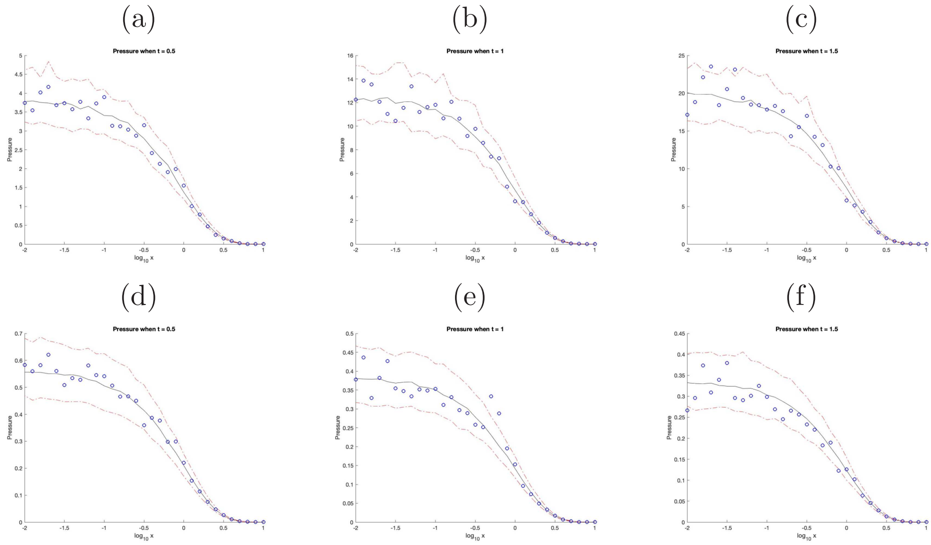

Figure 4: Profiles of predictive distribution for Model 1 with 95% predictive intervals across x. Data values ($\circ$), median ($-$), $Q_{2.5}$ and $Q_{97.5}$ ($- \cdot -$). Panels (a), (b) and (c) correspond to Model 1 profiles when $t = 0.5, 1.0$ and 1.5, respectively. Panels (d), (e) and (f) correspond to Model 2 profiles when $x = 0.5, 1.0$ and 1.5, respectively.

appears to be quite robust across reasonable parameter values.

5. Hypothesis testing

One important question that our approach can answer: "Is it worth using a fractional differential equation model versus a traditional model?" Using the Bayesian paradigm this problem can be framed as a hypothesis test on the fractionation parameter α specifically:

$$\begin{aligned} H_0 : \alpha &= 1, \\ H_A : \alpha &< 1. \end{aligned}$$

If $\alpha = 1$, then there is no fractionation needed. If $\alpha < 1$ then a fractionation is needed and should be included in the model. From a statistics point of view

Table 1: Average credible interval widths and coverage probabilities for α and σ for **Model 1**. All values based on 200 simulations.

α	σ	Width(α)	Width(σ)	Coverage(α)	Coverage(σ)
0.1	0.01	0.0002	0.0014	0.920	0.940
	0.1	0.0027	0.0151	0.975	0.930
	0.25	0.0066	0.0367	0.955	0.960
0.25	0.01	0.0005	0.0014	0.945	0.970
	0.1	0.0053	0.0151	0.955	0.955
	0.25	0.0129	0.0375	0.925	0.935
0.5	0.01	0.0011	0.0014	0.955	0.940
	0.1	0.0107	0.0148	0.975	0.930
	0.25	0.0229	0.0367	0.940	0.930
0.75	0.01	0.0016	0.0014	0.955	0.940
	0.1	0.0156	0.0149	0.925	0.930
	0.25	0.0299	0.0371	0.915	0.925
0.9	0.01	0.0019	0.0014	0.955	0.940
	0.1	0.0180	0.0150	0.915	0.945
	0.25	0.0312	0.0368	0.865	0.935

there are several approaches that could be used, such as likelihood ratio tests. In this work, the posterior hypothesis probability approach will be used as it is consistent with the previous Bayesian estimation approach. For notation simplicity let D denote all the observed data so that $D = \{y(\mathbf{x}, \mathbf{t})\}$. Since the goal is to obtain a posterior probability for the hypothesis, Bayes' Theorem will be used with prior probabilities for each hypothesis denoted as $P(H_0)$ and $P(H_A)$, respectively which gives:

$$P(H_A|D) = \frac{P(D|H_A)P(H_A)}{P(D|H_A)P(H_A) + P(D|H_0)P(H_0)},$$

where $P(H_A|D)$ is the posterior probability of the alternative hypothesis, H_A, given the data D and $P(H_0|D)$ is the posterior probability of the null hypothesis H_0 given the data D. The marginal probability of the data under H_A is denoted $P(D|H_A)$ and is calculated as:

$$P(D|H_A) = \int L(D|\alpha, \sigma, H_A)p(\alpha, \sigma H_A)d\alpha d\sigma,$$

where $L(D|\alpha, \sigma, H_A)$ is the likelihood of the data given α, σ and H_A and $p(\alpha, \sigma|H_A)$ is the prior distribution of α, σ under H_A [37]. Hypothesis

Table 2: Average credible interval widths and coverage probabilities for α and σ for **Model 2**. All values based on 200 simulations.

α	σ	Width(α)	Width(σ)	Coverage(α)	Coverage(σ)
0.1	0.01	0.0010	0.0014	0.955	0.955
	0.1	0.0117	0.0149	0.935	0.955
	0.25	0.0544	0.0372	0.870	0.940
0.25	0.01	0.0059	0.0014	0.945	0.935
	0.1	0.0579	0.0149	0.920	0.945
	0.25	0.1313	0.0375	0.960	0.940
0.5	0.01	0.0049	0.0014	0.930	0.945
	0.1	0.0492	0.0150	0.930	0.950
	0.25	0.1051	0.0374	0.920	0.910
0.75	0.01	0.0040	0.0014	0.935	0.925
	0.1	0.0395	0.0151	0.965	0.955
	0.25	0.0895	0.0372	0.920	0.935
0.9	0.01	0.0034	0.0014	0.960	0.910
	0.1	0.0335	0.0150	0.950	0.920
	0.25	0.0775	0.0371	0.925	0.965

testing from this approach is different from the traditional frequentist approach as there is no sampling distribution nor associated cut off value [38]. Instead, the probability is easily interpreted with values near 1 indicating high probability and values near 0 indicating low probability. If H_A is no better than H_0 then one would expect the $P(H_A|D)$ to be near 1/2. For more on Bayesian Hypothesis testing, see [25], [26]

In testing the hypothesis $\alpha = 1$, the prior distribution considered in Section 2.2 will not work as the Beta distribution assigns $P(\alpha = 1) = 0$. Hence, the hypothesis is not well defined under this prior. Instead of using the traditional Beta distribution, the one inflated Beta distribution will be used and is defined as:

$$OIBeta(\theta|\alpha^*, \beta^*, \phi) = \begin{cases} \phi & \text{if } \theta = 1, \\ (1-\phi)Beta(\alpha^*, \beta^*) & \text{if } \theta \in (0,1), \end{cases} \tag{15}$$

where $Beta(\alpha^*, \beta^*)$ is the traditional Beta distribution with parameters α^* and β^* and ϕ is the weight (mixing) parameter [39]. In this case, the weight parameter in the one inflated Beta distribution can be interpreted as the prior probability of H_0. For this work the weight parameter will be set to $\phi = 0.5$

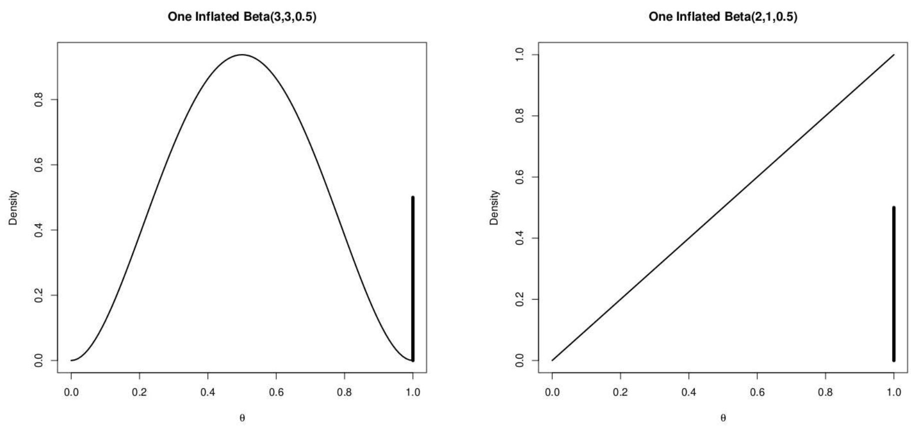

Figure 5: Density plots of the One Inflated Beta(3,3,0.5) and One Inflated Beta(2,1,0.5) distributions.

to reflect no preference for either hypothesis. Figure 5 shows the density plots of $OIBeta(3, 3, 0.5)$ and $OIBeta(2, 1, 0.5)$, which will be used to examine prior sensitivity. Notice that for the $OIBeta(3, 3, 0.5)$ that the density is predominantly in the middle of the domain with a spike at 1 and very low density for values near 1. Hence, this low density may influence the hypothesis test. Whereas, in the $OIBeta(2, 1, 0.5)$ the density is increasing across the interval from 0 to 1 with a spike at 1 as well. Notice that the density near 1 is much higher than the spike at 1. This high density may influence the resulting inferences.

To assess the sensitivity to prior distribution specification both $OIBeta(3, 3, 0.5)$ and $OIBeta(2, 1, 0.5)$ were considered. Figure 5 shows the density plots of these two distributions. Notice that for the $OIBeta(3, 3, 0.5)$ that the density is predominantly in the middle of the domain with a spike at 1 and very low density for values near 1. Hence, this low density may influence the hypothesis test. Whereas, in the $OIBeta(2, 1, 0.5)$ the density is increasing across the interval from 0 to 1 with a spike at 1 as well. Notice that the density near 1 is much higher than the spike. This high density may influence the resulting inferences.

A simulation study considers the distribution of $P(H_A|D)$ under values of $\alpha \in \{0.95, 0.96, 0.97, 0.98, 0.99, 1\}$ and $\sigma \in \{0.01, 0.1, 025\}$. For each combination of α and σ parameter values 200 datasets were generated from both **Model**

1 and **Model 2** from each model and both the null model ($\alpha = 1$) and the alternative models were fit, respectively, using the MCMC algorithm outlined above. For each of the Models and Hypotheses 1,000 posterior samples were obtained after a burn-in of 500 samples. Under H_0 the value of α was fixed at 0 and hence not estimated, however, σ was sampled using the same scheme as above. Using the retained posterior samples the marginal probability of the data under each hypothesis was calculated and $P(H_A|D)$ and $P(H_0|D)$ were obtained and recorded. This resulted in 200 posterior hypothesis probabilities for each parameter combination and the 2.5%, 50% and 97.5% quantiles of the $P(H_A|D)$ were found. This was done for both $OIBeta(3, 3, 0.5)$ and $OIBeta(2, 1, 0.5)$ as prior distributions.

Table 5 shows the results of the sensitivity study for **Model 1** with the quantiles for $P(H_A|D)$ given under each parameter combination and both prior distribution specifications. There are many items to note from this table. First, notice that when there is low noise in the data $\sigma = 0.01$, when $\alpha = 1$ (H_0 true), the quantiles for $OIBeta(3, 3, 0.5)$ are near 1/2, indicating no preference for either model. However, the quantiles for $OIBeta(2, 1, 0.5)$ show a large range of variability with some values around 0.8175 indicating it is likely that $\alpha < 1$. The table also shows fro $\sigma = 0.01$ that for values of $\alpha < 1$, regardless of prior distribution the analysis gives a posterior probability of 1 to H_A. Hence the prior distribution does have an impact on inferences when H_0 is true. For medium noise levels in the data $\sigma = 0.1$ the $OIBeta(2, 1, 0.5)$ prior distribution specification seems to be more sensitive to values of $\alpha < 1$ and is able to clearly detect when $\alpha < 0.97$ as evidenced by the 2.5% quantile being above 0.9. However, the $OIBeta(3, 3, 0.5)$ prior distribution specification doesn't seem to be able to detect $\alpha < 1$ until $\alpha < 0.96$. Also notice for this specification the wide variability in the posterior model probabilities when $\alpha > 0.96$ indicating the method has high uncertainty about the which hypothesis is more likely. For high levels of noise in the data $\sigma = 0.25$, the model with the $OIBeta(3, 3, 0.5)$ prior distribution is unable to consistently determine the more likely hypothesis. Likewise, the model with $OIBeta(2, 1, 0.5)$ as the prior distribution has difficulty as well but performs better than the other specification. Table 5 shows very similar results in that $OIBeta(3, 3, 0.5)$ prior distribution seems to produce considerable variability in the $P(H_A|D)$ for **Model 2** when σ is not a small value. This seems far more pronounced in **Model 2** than in **Model 1**. However, the $OIBeta(2, 1, 0.5)$ prior specification behaves as expected across the spectrum α and σ examined here.

Table 3: Posterior Hypothesis Probability 2.5%, 50% and 97.5% quantiles for $P(H_A|D)$ with varying levels of α and σ for **Model 1** where both $OIBeta(3,3,0.5)$ and $OIBeta(2,1,0.5)$ prior distributions are considered. Quantiles are reported as $(Q_{0.025}, Q_{0.5}, Q_{0.975})$. All values based on 200 simulated data sets.

σ	α	$OIBeta(3,3,0.5)$ $(Q_{0.025}, Q_{0.5}, Q_{0.975})$	$OIBeta(2,1,0.5)$ $(Q_{0.025}, Q_{0.5}, Q_{0.975})$
0.01	1	(0.4579,0.5001, 0.5412)	(0.4655, 0.5965, 0.8175)
	0.99	(1.0000 ,1.0000, 1.0000)	(1.0000 ,1.0000, 1.0000)
	0.98	(1.0000 ,1.0000, 1.0000)	(1.0000 ,1.0000, 1.0000)
	0.97	(1.0000 ,1.0000, 1.0000)	(1.0000 ,1.0000, 1.0000)
	0.96	(1.0000 ,1.0000, 1.0000)	(1.0000 ,1.0000, 1.0000)
	0.95	(1.0000 ,1.0000, 1.0000)	(1.0000 ,1.0000, 1.0000)
0.1	1	(0.0016 , 0.4969 , 0.5159)	(0.5101, 0.5838, 0.9093)
	0.99	(0.0001 , 0.0057 , 0.8776)	(0.5822, 0.8833, 0.9994)
	0.98	(0.0062 , 0.7639 , 0.9999)	(0.8599, 0.9983, 1.0000)
	0.97	(0.5500 , 0.9999 , 1.0000)	(0.9980, 1.0000, 1.0000)
	0.96	(0.9994 , 0.9999 , 1.0000)	(1.0000 ,1.0000, 1.0000)
	0.95	(0.9999 , 1.0000 , 1.0000)	(1.0000 ,1.0000, 1.0000)
0.25	1	(0.0015 , 0.4933 , 0.5292)	(0.5151, 0.5876, 0.9066)
	0.99	(0.0002, 0.0046 , 0.3826)	(0.5242, 0.6800, 0.9909)
	0.98	(0.0004 , 0.0102 , 0.8870)	(0.5311, 0.7616, 0.9958)
	0.97	(0.0017 , 0.1014 , 0.9928)	(0.5992, 0.9310, 0.9997)
	0.96	(0.0079 , 0.6234 , 0.9995)	(0.6853, 0.9910, 1.0000)
	0.95	(0.0364 , 0.8899 , 0.9999)	(0.8621, 0.9990, 1.0000)

6. Discussion

This work has shown the viability of using the Bayesian paradigm for both estimation and hypothesis testing when applied to fractional differential equations. Two distinct fractional differential equation models were explored using simulated data sets to determine the performance of the Bayesian inferential methods across values of α and σ. These models were purposely chosen as the number of parameters is small which allows us to isolate the fractionation parameter and examine the inferences without the complications of other parameters that may be included in a more complex model. In addition, posterior

Table 4: Posterior Hypothesis Probability 2.5%, 50% and 97.5% quantiles for $P(H_A|D)$ with varying levels of α and σ for **Model 2** where both $OIBeta(3,3,0.5)$ and $OIBeta(2,1,0.5)$ prior distributions are considered. Quantiles are reported as $(Q_{0.025}, Q_{0.5}, Q_{0.975})$. All values based on 200 simulated data sets.

σ	α	$OIBeta(3,3,0.5)$ $(Q_{0.025}, Q_{0.5}, Q_{0.975})$	$OIBeta(2,1,0.5)$ $(Q_{0.025}, Q_{0.5}, Q_{0.975})$
0.01	1	(0.4626 0.5023, 0.5413)	(0.4720, 0.5897, 0.9195)
	0.99	(1.0000 1.0000, 1.0000)	(1.0000, 1.0000, 1.0000)
	0.98	(1.0000 1.0000, 1.0000)	(1.0000, 1.0000, 1.0000)
	0.97	(1.0000 1.0000, 1.0000)	(1.0000, 1.0000, 1.0000)
	0.96	(1.0000 1.0000, 1.0000)	(1.0000, 1.0000, 1.0000)
	0.95	(1.0000 1.0000, 1.0000)	(1.0000, 1.0000, 1.0000)
0.1	1	(0.0015 0.4911, 0.5143)	(0.5035, 0.5810, 0.8525)
	0.99	(0.0003 0.0084, 0.5951)	(0.5354, 0.7944, 0.9933)
	0.98	(0.0021 0.1544, 0.9946)	(0.6645, 0.9749, 0.9999)
	0.97	(0.0235 0.9630, 1.0000)	(0.8627, 0.9990, 1.0000)
	0.96	(0.5228 0.9998, 1.0000)	(0.9929, 1.0000, 1.0000)
	0.95	(0.9958 1.0000, 1.0000)	(0.9999, 1.0000, 1.0000)
0.25	1	(0.0021 0.2680, 0.5205)	(0.4898, 0.5853, 0.8878)
	0.99	(0.0017 0.0164, 0.4278)	(0.5090, 0.6259, 0.9700)
	0.98	(0.0014 0.0355, 0.9406)	(0.5336, 0.7282, 0.9863)
	0.97	(0.0047 0.0886, 0.9743)	(0.5533, 0.8195, 0.9985)
	0.96	(0.0064 0.3525, 0.9965)	(0.5857, 0.9194, 0.9995)
	0.95	(0.0132 0.4542, 0.9995)	(0.6443, 0.9642, 1.0000)

predictive distributions were considered in a limited sense to illustrate how they can be utilized in the fractional differential equation paradigm.

An important issue that has arisen from this work is the choice of prior distribution for α when conducting hypothesis tests. The two One Inflated Beta distributions shown seem to be useful with the distribution with positive density for values near one being preferred to that where the density is of the portion less than zero being concentrated more towards the middle of the support. Further, work should be done to evaluate other parametrizations and possibly other distributions and their influence on the posterior hypothesis probabilities.

More complex models and models with larger parameter spaces should be included in future work where numeric solvers more than likely need to be employed to solve such complex fractional differential equations and systems. Many real world applications of fractional calculus and differential equations exist where Bayesian methods and estimation is important. These range from viscoelastic diffusion in complex fluids [30], anomalous diffusion [31], fractional order control problems [32], biological systems [33], and a lot more [18]. Determining the accuracy of the numeric solver and its impact on inferences should be examined to ensure that the choice of numeric solver does not unduly influence any parameter inferences or predictive distributions.

The significance of this work is two fold. Firstly, it establishes the viability of using the Bayesian paradigm for both estimation and hypothesis testing when applied to fractional differential equations. Secondly, the work is pioneering because it brings together fractional partial differential equations and statistical techniques for the first time.

References

[1] W. Hundsdorfer, J. Verwer, *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*, Springer-Verlag, Berlin, Heidelberg (2003).

[2] R.B. Bird, W.E. Stewart, E.N. Lightfoot, *Transport Phenomena*, John Wiley and Sons, Inc., New York (2002).

[3] T.E. Graedel, P.J. Crutzen, *Atmosphere, Climate and Change*, Henry Holt and Company (1997).

[4] K. Aziz, A. Settari, *Petroleum Reservoir Simulation*, Applied Science Publishers Ltd., London (1979).

[5] I. Glassman, R.A. Yetter, *Combustion*, Academic Press, London, San Diego, Burlington (2008).

[6] I. Ali , N.A. Malik, A realistic transport model with pressure-dependent parameters for gas flow in tight porous media with application to determining shale rock properties, *Transport in Porous Media*, **124**, No 3 (2018), 723-742.

[7] M. Weiss, M. Elsner, F. Kartberg, T. Nilsson, Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells, *Biophysical J.*, **87** (2004), 3518-3524.

[8] W. Chen, H. Sun, X. Zhang, D. Korosak, Anomalous diffusion modeling by fractal and fractional derivatives, *Comp. Math. Appl.* **59**, No 5 (2010), 1754-1758.

[9] R. Hilfer, Experimental evidence for fractional time evolution in glass forming materials, *Chemical Physics*, **284** (2002), 399-408.

[10] K. Diethelm, N.J. Ford, Analysis of fractional differential equations, *J. Math. Anal. Appl.*, **265** (2002), 229-248.

[11] R. Hilfer, *Applications of Fractional Calculus in Physics*, World Scientific Publishing Company, Singapore (2000).

[12] A.A. Kilbas, H.M. Srivastava, J.J. Trujillo, *Theory and Applications of Fractional Differential Equations*, North-Holland, Elsevier Science Publisher (2006).

[13] R.L. Magin, *Fractional Calculus in Bioengineering*, Begell House Publishers (2006).

[14] T.A. Atabong, M.O. Oyesanya, Nonlinear analysis of a fractional reaction diffusion model for tumour invasion, *J. of Mathematics and Computer Sci. Research* **4**, No 3 (2011), 136-158.

[15] T. Sandev, R. Metzler, Z. Tomovski, Fractional diffusion equation with a generalized Riemann-Liouville time fractional derivative, *J. Phys. A: Math. Theor.*, **44** (2011), Art. 255203, 21 pp.

[16] I. Ali, N.A. Malik, Hilfer fractional advection-diffusion equations with power-law initial condition: A numerical study using variational iteration method, *Computers and Math. with Appl.*, **68**, No 10 (2014), 1161-1179.

[17] N.A. Malik, R. Ghanam, S. Al-Homidan, Sensitivity of fluid transport through porous media to the fractional dimension in a fractional Darcy equation model, *Canadian J. Phys.*, **93** (2015), 1–19.

[18] H.G. Sun, Y. Zhang, D. Baleanu, W. Chen, Y.Q. Chen, A new collection of real world applications of fractional calculus in science and engineering, *Comm. Nonlin. Science and Numer. Simul.*, **64** (2018), 213-231.

[19] M. Caputo, Diffusion of fluids in porous media with memory, *Geothermics*, **28** (1999), 113-130.

[20] S. Havlin, D. Ben-Avraham, Diffusion in disordered media, *Advances in Phys.*, **51** (2002), 187-292.

[21] Y. Luchko, A. Punzi, Modeling anomalous heat transport in geothermal reservoirs via fractional diffusion equations, *Int. J. Geomath.*, **1** (2011), 257-276.

[22] R. Metzler, J. Klafter, The random walk's guide to anomalous diffusion: a fractional dynamics approach, *Phys. Reports*, **339** (2000), 1-77.

[23] M. Caputo, Diffusion with space memory modelled with distributed order space fractional differential equations, *Annals of Geophys.*, **2**, No 46 (2003), 223-234.

[24] N.A. Malik, I. Ali, B. Chanane, Numerical solutions of non-linear fractional transport models in unconventional hydrocarbon reservoirs using variational iteration method, *Proc. 5th Intern. Conf. on Porous Media and its Applications in Science and Engineering*, Kona Hawaii (2014).

[25] J.O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd Ed., Springer, New York (1985).

[26] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, *Bayesian Data Analysis*, 3rd Ed., CRC Press, Boca Raton, FL (2013).

[27] L. Jing, N. Ou N, Bayesian inference using intermediate distribution based on coarse multiscale model for time fractional diffusion equations, *Multiscale Model. Simul.*, **16**, No 1 (2018), 327–355.

[28] W.R. Gilks, S. Richardson, D.J. Spiegelhaler, *Markov Chain Monte Carlo in Practice*, Chapman & Hall, CRC Press, Boca Raton (1996).

[29] J. Albert, *Bayesian Computation with R*, 2nd Ed., Springer, New York (2009).

[30] J.F. Reverey, J.H. Jeon, H. Bao, M. Leippe, R. Metzler, C. Selhuber-Unkel, Superdiffusion dominates intracellular particle motion in the supercrowded space of pathogenic acanthamoeba castellanii, *Sci. Rep.*, **5** (2015), Art. 11690.

[31] W. Deng, X. Wu, W. Wang, Mean exit time and escape probability for the anomalous processes with the tempered power-law waiting times, *EPL* **117** (2017), Art. 10009.

[32] P.D. Mandic, T.B. Scekara, M.P. Lazarevic, Dominant pole placement with fractional order PID controllers: D-decomposition approach, *ISA Trans.* **67** (2017), 76–86.

[33] C.M.A. Pinto, A.R.M. Carvalho, Fractional order model for HIV dynamics, *J. Comput. Appl. Math.*, **312** (2017), 240–256.

[34] T. Bayes, R. Price, An essay towards solving a Problem in the Doctrine of Chance (By the late Rev. Mr. Bayes, Communicated by Mr. Price in a letter to John Canton), A.M.F.R.S. *Philos. Trans. of Royal Soc. London*, **53** (1763), 370-418.

[35] D. Wackerly, W. Mendenhall, R.L. Scheaffer, *Mathematical Statistics with Applications*, 7th Ed., Thomson Brooks/Cole, Belmont, CA (2008).

[36] G. Casella, R.L. Berger, *Statistical Inference*, 2nd Ed., Duxbury, Belmont, CA (2002).

[37] R.E. Kass, L. Wasserman, A reference Bayesian test for nested hypothesis and its relationship to the Schwarz criterion, *J. of the Amer. Stat. Assoc.*, **90**, No 431 (1995), 928-934.

[38] J.A. Hoeting, D. Madigan, A.E. Raftery, C.T. Volinsky, Bayesian model averaging: A tutorial, *Statistical Sci.*, **14**, No 4 (1999), 382-417.

[39] R. Ospina, SLP. Ferrari, A general class of zero-or-one inflated beta regression models, *Computat. Stat.* & *Data Analysis*, **56**, No 6 (2012), 1609-1623.