

# Bayesian comparison of stochastic models of dispersion

Martin T. Brolly<sup>1,†</sup>, James R. Maddison<sup>1</sup>, Aretha L. Teckentrup<sup>1</sup> and Jacques Vanneste<sup>1</sup>

<sup>1</sup>School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh, King's Buildings, Edinburgh EH9 3FD, UK

(Received 6 January 2022; revised 14 April 2022; accepted 24 May 2022)

Stochastic models of varying complexity have been proposed to describe the dispersion of particles in turbulent flows, from simple Brownian motion to complex temporally and spatially correlated models. A method is needed to compare competing models, accounting for the difficulty in estimating the additional parameters that more complex models typically introduce. We employ a data-driven method, Bayesian model comparison, which assigns probabilities to competing models based on their ability to explain observed data. We focus on the comparison between the Brownian and Langevin dynamics for particles in two-dimensional isotropic turbulence, with data that consist of sequences of particle positions obtained from simulated Lagrangian trajectories. We show that, while on sufficiently large time scales the models are indistinguishable, there is a range of time scales on which the Langevin model outperforms the Brownian model. While our set-up is highly idealised, the methodology developed is applicable to more complex flows and models of particle dynamics.

**Key words:** dispersion, machine learning, turbulent mixing

## 1. Introduction

Since Taylor introduced the notion of turbulent diffusion in the 1920s (Taylor 1922), a wide variety of stochastic models have been proposed to represent the dynamics of particles in turbulent flows (e.g. Thomson 1987; Rodean 1996; Majda & Kramer 1999; Berloff & McWilliams 2002). The Brownian dynamics used by Taylor models Lagrangian velocities as white noise processes and is a good approximation only on sufficiently long time scales. More complex models incorporate temporal and/or spatial correlation (e.g. Griffa

† Email address for correspondence: [m.brolly@ed.ac.uk](mailto:m.brolly@ed.ac.uk)

1996; Pasquero, Provenzale & Babiano 2001; Lilly *et al.* 2017). For example, Langevin dynamics incorporates autocorrelation in Lagrangian velocities by representing them as Ornstein–Uhlenbeck processes (Uhlenbeck & Ornstein 1930). It is in general unclear when such additional complexity leads to improved predictions rather than to overfitting. Given the increased difficulty and cost of implementing more complex models, a method for comparing the performance of competing stochastic models for particle dynamics is needed.

To this end, we propose a data-driven approach: we apply Bayesian model comparison (BMC) (Kass & Raftery 1995; Jaynes 2003; MacKay 2003), which assigns probabilities to competing models based on their ability to explain observed data. We focus on the comparison between the Brownian and Langevin models for particles in two-dimensional homogeneous isotropic turbulence, with data that consist of sequences of particle positions obtained from simulated Lagrangian trajectories. While this set-up is highly idealised, the methodology developed is applicable to more complex flows and models of particle dynamics.

Model comparison is complicated by two issues: (i) proposed models typically contain a number of parameters whose values are uncertain, and (ii) a measure of model suitability is required, balancing accuracy and complexity. The natural language for this problem is then that of decision theory (see e.g. Bernardo & Smith (1994) and Robert (2007) for an overview of decision problems under uncertainty); however, several philosophical issues therein, such as the choice of utility function and its subjectivity, can be avoided by adopting the ready-made approach of BMC. BMC and the related technique Bayesian model averaging are gaining popularity in many applied fields (Min, Simonis & Hense 2007; Mann 2011; Carson *et al.* 2018; Mark *et al.* 2018). In this paper, we demonstrate the potential of BMC by comparing the Brownian and Langevin models of dispersion in two-dimensional turbulence. This provides a simple illustration of the BMC methodology while addressing a problem of interest: dispersion in two-dimensional turbulence has received much attention as a paradigm for transport and mixing in stratified, planetary-scale geophysical flows (Provenzale, Babiano & Villone 1995), and can be modelled with stochastic processes (e.g. Pasquero *et al.* 2001; Lilly *et al.* 2017).

The paper is structured as follows. We introduce the Brownian and Langevin models in § 2 and review the BMC method in § 3. In § 4 we show how this method can be applied to discrete particle trajectory data; we also show results of a test case, where the data are generated by the Langevin model itself. In § 5 we apply BMC to data from direct numerical simulations of two-dimensional turbulence. In § 6 we give our conclusions on the method.

## 2. Models and data

### 2.1. Brownian and Langevin models

The models of interest are the Brownian model, which for passive particles in homogeneous and isotropic turbulence is given by

$$d\mathbf{X} = \sqrt{2\kappa} d\mathbf{W}, \quad (2.1)$$

with  $\kappa > 0$ , and the Langevin model, which, under the same conditions, is given by

$$d\mathbf{X} = \mathbf{U} dt, \quad (2.2a)$$

$$d\mathbf{U} = -\gamma \mathbf{U} dt + \gamma \sqrt{2k} d\mathbf{W}, \quad (2.2b)$$

with  $\gamma, k > 0$ , and where, in both cases,  $\mathbf{W}$  is a vector composed of independent Brownian motions. We denote the models by  $\mathcal{M}_B(\kappa)$  and  $\mathcal{M}_L(\gamma, k)$ .

We note some important characteristics of the two models. The Brownian model involves particle position,  $X$ , as its only component, which evolves as a scaled  $d$ -dimensional Brownian motion, where  $d$  is the number of spatial dimensions. This implies that particle velocity evolves as a white noise process. The model has one parameter, the diffusivity  $\kappa$ . The validity of (2.1) is typically justified by arguments involving strong assumptions of scale separation between mean flows and small-scale fluctuations which rarely hold in applications (Majda & Kramer 1999; Berloff & McWilliams 2002).

The Langevin model, by contrast, involves two components, particle position and particle velocity,  $(X, U)$ . The velocity component evolves according to a mean-zero Ornstein–Uhlenbeck process, and position results from time integration of this velocity. The model has two parameters,  $\gamma$  and  $k$ , where  $\gamma^{-1}$  is a Lagrangian velocity decorrelation time and  $k$  characterises the strength of Gaussian velocity fluctuations. The Brownian and Langevin models are the first two members of a hierarchy of Markovian models involving an increasing number of time derivatives of the position (Berloff & McWilliams 2002).

In practice, the Brownian model is favoured over the Langevin model for its simplicity as well as for the practical virtue of having a smaller, more easily explored, one-dimensional parameter space. Note that if these models are to be implemented in the limit of continuous concentrations of particles then it is their corresponding Fokker–Planck equations which must be solved – this means solving partial differential equations in  $d + 1$  or  $2d + 1$  dimensions, respectively.

Both the Brownian and Langevin model can be extended to account for spatial anisotropy, inhomogeneity and the presence of a mean flow, at the cost of increasing the dimension of their parameter spaces; full details are given in Berloff & McWilliams (2002). Brownian and Langevin dynamics underlie the so-called random displacement and random flight models used for dispersion in the atmospheric boundary layer (Esler & Ramli 2017), and have been applied to the simulation of ocean transport, as models of mixing in the horizontal (Berloff & McWilliams 2002), vertical (Onink, van Sebille & Laufkötter 2022) and on neutral surfaces (Reijnders, Deleersnijder & van Sebille 2022). Ying, Maddison & Vanneste (2019) showed how Bayesian parameter inference can be applied to the Brownian model in the inhomogeneous setting using Lagrangian trajectory data. We restrict attention to isotropic turbulence in this work for simplicity, noting that the methods demonstrated below are equally applicable in the more general case.

## 2.2. Data

For our comparison we consider trajectory data of the form

$$\{(X_0^{(p)}, \dots, X_{N_\tau}^{(p)}) : p \in \{1, \dots, N_p\}\}, \quad (2.3)$$

where  $X_n^{(p)}$  is the position of particle  $p$  at time  $t = n\tau$ . In words, we observe the positions of a set of  $N_p$  particles at discrete time intervals of length  $\tau$ , which we refer to as the sampling time. The performance of the models depends crucially on  $\tau$ . Since both models are homogeneous in space, we can rewrite the observations as the set of displacements

$$\Delta X_\tau = \{(\Delta X_0^{(p)}, \dots, \Delta X_{N_\tau-1}^{(p)}) : p \in \{1, \dots, N_p\}\}, \quad (2.4)$$

where  $\Delta X_n^{(p)} = X_{n+1}^{(p)} - X_n^{(p)}$ .

In §4 we consider the case that the trajectory data are generated by Langevin dynamics, while in §5 we compare the Brownian and Langevin models given data

from direct numerical simulations of a forced–dissipative model of stationary, isotropic two-dimensional turbulence.

### 3. Methods

In this work we appeal to the Bayesian interpretation of probability and statistics. This means that probabilities reflect levels of plausibility in light of all available information. In particular, we deal with uncertainty in both the parameters of each model and the models themselves by assigning probabilities to them. We outline this procedure in §§ 3.1 and 3.2.

#### 3.1. Parameter inference

The goal of parameter inference is to infer the values of the parameters  $\theta \in \Theta$  of a statistical model, say  $\mathcal{M}(\theta)$ , given observational data  $\mathcal{D}$ . A model is characterised completely by its likelihood function  $p(\cdot|\mathcal{M}(\theta))$ , which denotes the probability (density) of observations under  $\mathcal{M}(\theta)$ . Bayesian inference requires the specification of one’s belief prior to observations through a prior distribution  $p(\theta|\mathcal{M})$ . One can then invoke Bayes’ theorem, (3.1), to update this belief in light of the observations. This results in a posterior distribution

$$\overbrace{p(\theta|\mathcal{D}, \mathcal{M})}^{\text{Posterior}} = \frac{\overbrace{p(\mathcal{D}|\mathcal{M}(\theta))}^{\text{Likelihood}} \overbrace{p(\theta|\mathcal{M})}^{\text{Prior}}}{\underbrace{p(\mathcal{D}|\mathcal{M})}_{\text{Evidence}}}, \quad (3.1)$$

which denotes the probability (density) of each  $\theta \in \Theta$  given observations and prior knowledge (Jeffreys 1983). The posterior fully describes the uncertainty in the inferred parameters, in our case  $\theta = \kappa$  or  $\theta = (k, \gamma)$ . In applications where point estimates of the parameters are required, these can be taken as e.g. the mean or mode of the posterior.

#### 3.2. Model inference

Beyond parameter inference we can also make inferences when the model itself,  $\mathcal{M}$ , is considered unknown. However, in order to meaningfully assign probabilities to models we must assume that the set of models under consideration,  $M = \{\mathcal{M}_i\}_{i=1}^{N_M}$ , includes all plausibly true models. That is, for any  $\mathcal{M}^* \notin M$ ,  $p(\mathcal{M}^*) = 0$ . This is known as the  $\mathcal{M}$ -closed regime (see chapter 6 of Bernardo & Smith (1994) or Clyde & Iversen 2013). In situations where all models under consideration are known to be false this assumption appears dubious; however, we note that the same fallacy is committed in Bayesian parameter inference when we assign probabilities to the parameters of a parametric model which we know is imperfect, i.e. false. In the  $\mathcal{M}$ -closed regime one assigns prior probabilities to models such that  $\sum_{i=1}^{N_M} p(\mathcal{M}_i) = 1$ . This allows us to again invoke Bayes’ theorem in the form

$$p(\mathcal{M}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathcal{M})p(\mathcal{M})}{p(\mathcal{D})}. \quad (3.2)$$

If  $\mathcal{M}_i$  is parametric with parameters  $\theta_i \in \Theta_i$ ,  $p(\mathcal{D}|\mathcal{M}_i)$  is given by

$$p(\mathcal{D}|\mathcal{M}_i) = \int_{\Theta_i} p(\mathcal{D}|\mathcal{M}_i(\theta_i))p(\theta_i|\mathcal{M}_i) d\theta_i, \quad (3.3)$$

which is known as the model evidence (or marginal likelihood, or model likelihood) of  $\mathcal{M}_i$ .

An important property of the evidence is that it accounts for parameter uncertainty. Considering the likelihood as a score of model performance given some fixed parameter values, the evidence can be viewed as an expectation of that score with respect to the prior measure on parameters. In this way the evidence favours models where observations are highly probable for the range of parameter values considered plausible *a priori*. In particular, this means that a model with many parameters which achieves a very high value of the likelihood only for a narrow range of parameter values which could not be predicted *a priori* is not likely to attain a higher value of the evidence than a model with fewer parameters whose values are better constrained by prior information. This apparent penalty is usually quantified by the so-called Occam (or Ockham) factor, named in reference to Occam’s razor,

$$\text{Occam}_i = p(\mathcal{D}|\mathcal{M}_i)/p(\mathcal{D}|\mathcal{M}_i(\theta_i^*)) \in [0, 1], \tag{3.4}$$

where  $\theta_i^*$  is the posterior mode of  $\theta_i$  (Jaynes 2003; MacKay 2003).

Given two models,  $\{\mathcal{M}_0, \mathcal{M}_1\}$ , a test statistic for the hypotheses

$$\left. \begin{aligned} \mathcal{H}_0 : & \mathcal{M}_0 \text{ is the true model,} \\ \mathcal{H}_1 : & \mathcal{M}_1 \text{ is the true model,} \end{aligned} \right\} \tag{3.5}$$

is given by the Bayes factor (Kass & Raftery 1995),

$$K_{1,0} = \frac{p(\mathcal{D}|\mathcal{M}_1)}{p(\mathcal{D}|\mathcal{M}_0)}, \tag{3.6}$$

where a large value of  $K_{1,0}$  represents statistical evidence against  $\mathcal{H}_0$ .

The log evidence is exactly equal to the log score (Gneiting & Raftery 2007), also known as the ignorance score (Bernardo 1979; Bröcker & Smith 2007), for probabilistic forecasts. Therefore, the log Bayes factor can be understood as a difference of scores for probabilistic models. Merits of the log score have been appreciated since at least the 1950s (Good 1952), including its intimate connection with information theory (Roulston & Smith 2002; Du 2021). This interpretation of the Bayes factor does not rely on the assumption of the  $\mathcal{M}$ -closed regime. In what follows we use the Bayes factor to compare the Brownian and Langevin models.

A useful approximation for the evidence (3.3) is given by Laplace’s method: a Gaussian approximation of the unnormalised posterior,  $p_u(\theta) = p(\mathcal{D}|\mathcal{M}(\theta))p(\theta|\mathcal{M})$ , is obtained from a quadratic expansion of  $\ln p_u$  about the posterior mode  $\theta^*$ ,

$$\ln(p_u(\theta)) \approx \ln(p_u(\theta^*)) - \frac{1}{2}(\theta - \theta^*)^T \mathbf{J}(\theta - \theta^*), \tag{3.7}$$

where

$$\mathbf{J}_{ij} = - \left. \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln p_u(\theta) \right|_{\theta=\theta^*}. \tag{3.8}$$

Taking an exponential of (3.7) we recognise that we have approximated  $p_u(\theta)$  with the probability density function (up to a known normalisation) of a Gaussian random variable with mean  $\theta^*$  and covariance  $\mathbf{J}^{-1}$ , so (3.3) becomes

$$\underbrace{p(\mathcal{D}|\mathcal{M}_i)}_{\text{Evidence}} \approx \underbrace{p(\mathcal{D}|\mathcal{M}_i(\theta_i^*))}_{\text{Maximum likelihood}} \times \underbrace{p(\theta_i^*|\mathcal{M}_i) (\det(\mathbf{J}/2\pi))^{-1/2}}_{\text{Occam factor}}. \tag{3.9}$$

This approximation is accurate for a large number of data points  $N_p \times N_\tau$  where a Bernstein–von Mises theorem can be shown to hold, guaranteeing asymptotic normality of the posterior measure (van der Vaart 1998).

We highlight that a model’s evidence is sensitive to the prior distribution on the parameters,  $p(\theta|\mathcal{M})$ . This is entirely in the spirit of Bayesian statistics in that a parametric model accompanied by the prior uncertainty on its parameters constitutes a single, complete hypothesis for explaining observations. The evidence for a model is less when the mass of prior probability on parameters is less concentrated on those values for which the likelihood is largest.

#### 4. Results

In this section we provide details on how BMC can be performed for the Brownian and Langevin model and consider data generated by the Langevin model. We derive the likelihood function for each model, discuss prior distributions for parameters and the practicalities of inference calculations.

Before we compute the Bayes factor for the Langevin and Brownian models  $\mathcal{M}_L$  and  $\mathcal{M}_B$ , we infer the parameters of both models using a range of datasets with varying sampling time,  $\tau$ , to establish when each model is sampling-time consistent – we say a model is sampling-time consistent when inferred parameter values are stable over a range of  $\tau$ . We emphasise that sampling-time consistency does not imply a model is good, but is certainly a desirable property when one wishes to use a model for extrapolation, e.g. for unobserved values of  $\tau$ .

Justifications for the Brownian model apply formally only in the large-time limit; we are, therefore, interested in establishing a minimum time scale for the sampling-time consistency of the Brownian model, and further establishing whether the Langevin model, given that it includes time correlation, is sampling-time consistent on shorter time scales.

Note that in the large-time limit, that is, for  $t \gg \gamma^{-1}$ , the Langevin dynamics is asymptotically diffusive: for  $\gamma \rightarrow \infty$ , the Langevin equations (2.2) reduce to (Pavliotis 2014)

$$dX = \sqrt{2k} dW. \tag{4.1}$$

This fact is important when comparing the models, and we return to it later.

##### 4.1. Likelihoods

We can derive explicit expressions for the probability of data of the form of  $\Delta\mathcal{X}_\tau$  under  $\mathcal{M}_B(\kappa)$  and  $\mathcal{M}_L(\gamma, k)$  by using their transition probabilities. The position increments for  $\mathcal{M}_B(\kappa)$  satisfy

$$X(t + \tau) - X(t) \sim \mathcal{N}(0, 2\kappa\tau\mathbb{I}), \tag{4.2}$$

where  $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$  is the  $d$ -dimensional Gaussian distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\mathbf{C}$ , and  $\mathbb{I}$  is the  $d \times d$  identity matrix. Further, distinct increments are independent under  $\mathcal{M}_B(\kappa)$ . Therefore, the desired probability is

$$p(\Delta\mathcal{X}_\tau|\mathcal{M}_B(\kappa)) = \prod_{p=1}^{N_p} \prod_{n=0}^{N_\tau-1} \prod_{i=1}^d \rho_{\mathcal{N}}(\Delta X_{n,i}^{(p)}; 0, 2\kappa\tau), \tag{4.3}$$

where  $i$  indexes spatial dimension and  $\rho_{\mathcal{N}}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{C})$  is the probability density at  $\mathbf{x}$  of the Gaussian distribution  $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ .



The corresponding likelihood for the Langevin model is shown in [Appendix A](#) to be

$$p(\Delta\mathcal{X}_\tau|\mathcal{M}_L(\gamma, k)) = \prod_{p=1}^{N_p} \prod_{i=1}^d \rho_{\mathcal{N}}((\Delta X_{0,i}^{(p)}, \dots, \Delta X_{N_\tau-1,i}^{(p)})^T; \mathbf{0}, \mathbf{S}), \quad (4.4)$$

where  $\mathbf{S}$  is the symmetric Toeplitz matrix with

$$S_{ij} = \begin{cases} 2k\tau(1 - \varphi(\gamma\tau)) & \text{if } m = 0 \\ k\gamma\tau^2\varphi^2(\gamma\tau)\exp(-(m-1)\gamma\tau) & \text{if } m > 0 \end{cases}, \quad (4.5)$$

$$\varphi(x) = \frac{1 - e^{-x}}{x}, \quad (4.6)$$

and  $m = |i - j|$ .

#### 4.2. Prior distributions

It is necessary, both for parameter and model inference, to specify a prior distribution for each of the parameters,  $\kappa$ ,  $\gamma$  and  $k$ . For a given flow we can appeal to scaling considerations to assign a prior mean to each parameter, derived from characteristic scales. Once such prior means are prescribed, the maximum entropy principle, along with positivity and independence of the parameters, motivates a choice of corresponding exponential distributions as priors (Jaynes 2003; Cover & Thomas 2006). That is, for a parameter  $\theta > 0$  with prior mean  $\mu$ , the distribution with maximum entropy is the exponential distribution  $\text{Exp}(\lambda)$  with rate  $\lambda = 1/\mu$ . We use this prescription for our choice of prior.

#### 4.3. Inference numerics

The computations we perform for Bayesian parameter inference are: (i) an optimisation procedure to find the posterior mode,  $\theta^*$ , and (ii) a single evaluation of the Hessian of the log-posterior distribution at  $\theta^*$ ,  $-\mathbf{J}$  in (3.8), which we can use to estimate the posterior variance by a Gaussian approximation as in (3.7). We have analytical expressions for the likelihood and prior for both models, so we can easily evaluate the negative log unnormalised posterior,  $f(\theta) = -\ln p_u(\theta|\mathcal{D})$ , in each case; we find  $\theta^*$  by minimising  $f(\theta)$  using the SciPy function `optimize.minimize()`.

In the case of the Brownian model derivatives of  $f(\theta)$  are easily derived analytically, so we use the L-BFGS-B routine which exploits gradient information and allows for the specification of lower bound constraints to enforce positivity (Zhu *et al.* 1997). In the case of the Langevin model calculation of derivatives of the posterior is non-trivial because the likelihood (4.4) is a complicated function. For this reason we use the gradient-free Nelder-Mead (Nelder & Mead 1965) routine rather than L-BFGS-B. We evaluate  $\mathbf{J}^{-1}$  approximately using a fourth-order central difference approximation for the log likelihood.

No further computations are required for BMC if the Laplace's method approximation for the evidence in (3.9) is used.

#### 4.4. Test case: Langevin data

As a test case and to build intuition, we first consider trajectory data generated by the Langevin model with  $d = 3$ . In this case, one of the two candidate models is the

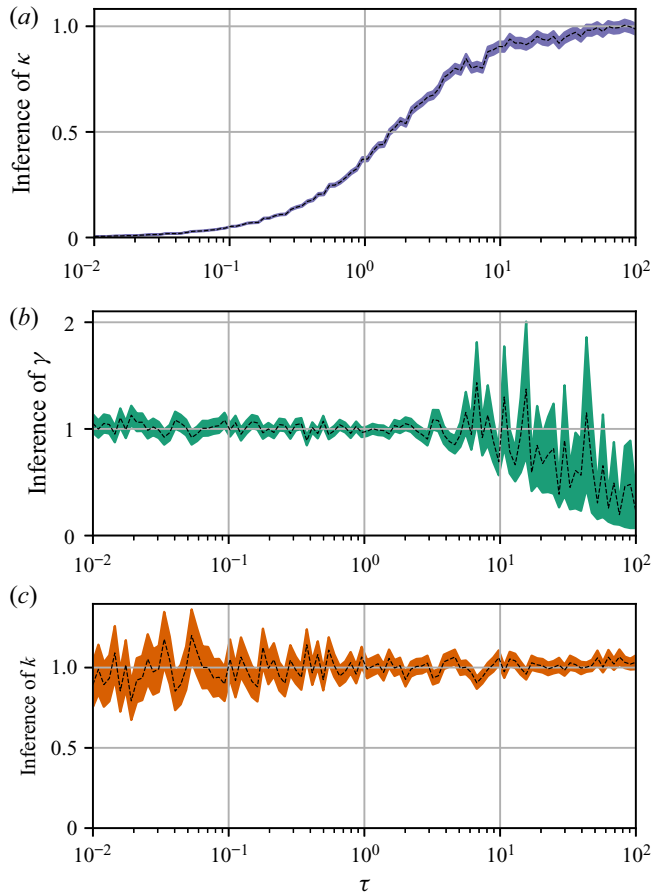


Figure 1. Parameter inference for the Brownian and Langevin models as a function of observation interval,  $\tau$ , for data from the Langevin model in three spatial dimensions. Dashed lines indicate posterior mode estimates,  $\theta^* = \kappa$  (a),  $\gamma^*$  (b) and  $k^*$  (c); shaded areas show  $\theta^* \pm \text{SD}(\theta|\Delta\mathcal{X}_\tau)$ . Each inference is made with a fixed volume of data:  $N_p = 100$  and  $N_\tau = 10$ .

true model. We generate the data by simulating the Langevin equation (2.2) exactly, drawing initial velocities from the stationary distribution  $U|\mathcal{M}_L(\gamma, k) \sim \mathcal{N}(\mathbf{0}, \gamma k \mathbb{I})$ , and using the transition probabilities (A1); velocity data are discarded to construct the dataset of position increments  $\Delta\mathcal{X}_\tau$ .

We set  $\gamma = k = 1$ , fix  $N_p = 100$  and  $N_\tau = 10$  and perform Bayesian parameter inference and model comparison with a series of independently generated datasets with  $\tau \in [10^{-2}, 10^2]$ . We set fixed priors  $\gamma, k, \kappa \sim \text{Exp}(1)$ .

Figure 1 shows the results of the parameter inference. Note that both Langevin parameters are well identified until, at sufficiently large  $\tau$ , the error of the posterior mode estimate of  $\gamma$  grows along with the posterior standard deviation of  $\gamma$ . This is a manifestation of the diffusive limit (4.1) of the Langevin dynamics, wherein  $X(t + \Delta t) - X(t) \sim \mathcal{N}(\mathbf{0}, 2k\Delta t \mathbb{I})$  is independent of  $\gamma$ . Unsurprisingly, then,  $\Delta\mathcal{X}_\tau$  is less informative about  $\gamma$  when  $\tau$  is large.

The diffusivity  $\kappa$  of the Brownian model is sampling-time consistent only when  $\tau$  is sufficiently large, i.e. in the diffusive limit of the Langevin dynamics, when  $\kappa \approx k$ .



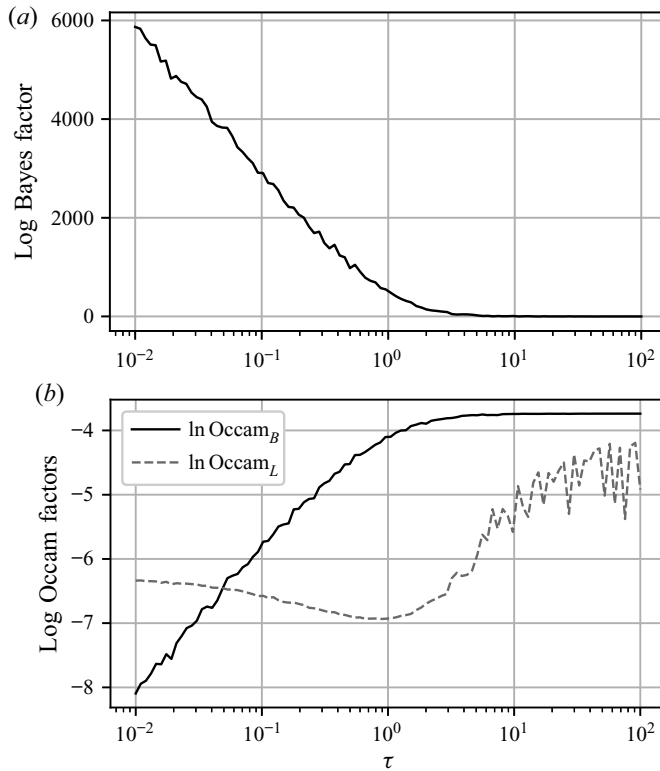


Figure 2. Log Bayes factors,  $\ln K_{L,B}$ , and corresponding log Occam factors, as a function of  $\tau$ , given the same data used for [figure 1](#).

The inaccuracy of posterior mode estimates of  $\kappa$  at small  $\tau$  is expected as it is known that the inference of diffusivities from discrete trajectory data is sensitive to sampling time (Cotter & Pavliotis 2009). We note that  $\gamma^{-1} = 1$  is the decorrelation time for this data so that the time scales at which this limiting behaviour is observed,  $\tau \gtrsim 10$ , are indeed large.

Note that the posterior mode estimates of  $\gamma$  eventually decay to zero as  $\tau$  increases; since, as observed,  $\Delta \mathcal{X}_\tau$  becomes less informative about  $\gamma$  with increasing  $\tau$ , the contribution of the prior information to the posterior becomes dominant over the contribution from the likelihood – the consequence of this is that the posterior mode tends to the prior mode, which is zero since we take  $\gamma \sim \text{Exp}(1)$ .

[Figure 2](#) shows the log Bayes factors found using Laplace’s method for the evidence. We see that, for a significant range of  $\tau$ , the Langevin model is preferred, indicated by large positive values of  $\ln K_{L,B}$ , but its dominance diminishes as  $\tau$  increases until the diffusive limit is reached, at which point values of  $|\ln K_{L,B}| < 1$  are typical, indicating insubstantial preference for either model.

Also shown in [figure 2](#) are the corresponding log Occam factors. Occam factors for the Brownian model are approximately constant once  $\tau$  is sufficiently large, while the Occam factors for the Langevin model increase at large  $\tau$ , in line with a broadening posterior. This is indicative of decreased sensitivity to choice of parameters, specifically  $\gamma$ , whose value becomes less critical for explaining the dynamics on large time scales.

domain =  $[0, 2\pi]^2$ ; resolution =  $1024 \times 1024$  grid points; timestep size =  $2.5 \times 10^{-4}$ ;  
 $k_F = 64$ ;  $A_F = 8.9 \times 10^8$ ;  $A_{lsf} = 1$ .

Table 1. Flow configuration parameter values for simulations of the 2-D turbulence model.

## 5. Application to two-dimensional turbulence

In this section we report an application of BMC to particle trajectories in a model of stationary, isotropic two-dimensional (2-D) turbulence.

### 5.1. Forced–dissipative model

We consider a forced–dissipative model of isotropic 2-D turbulence in an incompressible fluid governed by the vorticity equation (Vallis 2017)

$$\frac{\partial \zeta}{\partial t} + (\mathbf{u} \cdot \nabla) \zeta = F + D, \quad (5.1)$$

where  $\zeta$  is the vertical vorticity and  $F$  and  $D$  represent forcing and dissipation, respectively. The particular forcing used is an additive homogeneous and isotropic white Gaussian noise concentrated in a specified range of wavenumber centred about a forcing wavenumber,  $k_F$ . In particular, following Scott (2007), we have that, at each timestep, the Fourier transform of  $F$  satisfies

$$\text{Re}(\hat{F}(\mathbf{k})) \stackrel{d}{=} \text{Im}(\hat{F}(\mathbf{k})) \sim \mathcal{N}\left(0, \frac{A_F \mathcal{F}_F(|\mathbf{k}|)}{2\pi|\mathbf{k}|}\right), \quad (5.2)$$

where  $A_F$  is the forcing amplitude, and  $\mathcal{F}_F(|\mathbf{k}|) = 1$  for  $||\mathbf{k}| - k_F| \leq 2$  and  $\mathcal{F}_F = 0$  otherwise.

Two dissipation mechanisms are included: (i) small-scale dissipation implemented with a scale-selective exponential cutoff filter (see Arbic & Flierl (2003) for details and justification), and (ii) large-scale friction (aka hypodiffusion), so that total dissipation is given by

$$D = A_{lsf} \psi + \text{ssd}, \quad (5.3)$$

where  $\text{ssd}$  denotes the small-scale dissipation.

Equation (5.1) is solved in a periodic domain,  $[0, 2\pi]^2$ , using a standard pseudospectral solver, at a resolution of  $1024 \times 1024$  gridpoints, with the third-order Adams–Bashforth timestepping scheme. The complete set of flow configuration parameter values for our simulations are given in table 1.

The model is initialised with a random Gaussian field with prescribed mean energy spectrum and is run until the total energy,

$$E(t) := \frac{1}{2} \int |\mathbf{u}(\mathbf{x}, t)|^2 \, dx \, dy, \quad (5.4)$$

appears to reach a statistically stationary state; this amounted to a spin-up time of approximately 6800 eddy turnover times, where the eddy turnover time is estimated by

$$\tau_\zeta = 2\pi/\sqrt{Z}, \quad (5.5)$$

and  $Z$  is the total enstrophy,

$$Z := \frac{1}{2} \int \zeta^2 \, dx \, dy. \quad (5.6)$$

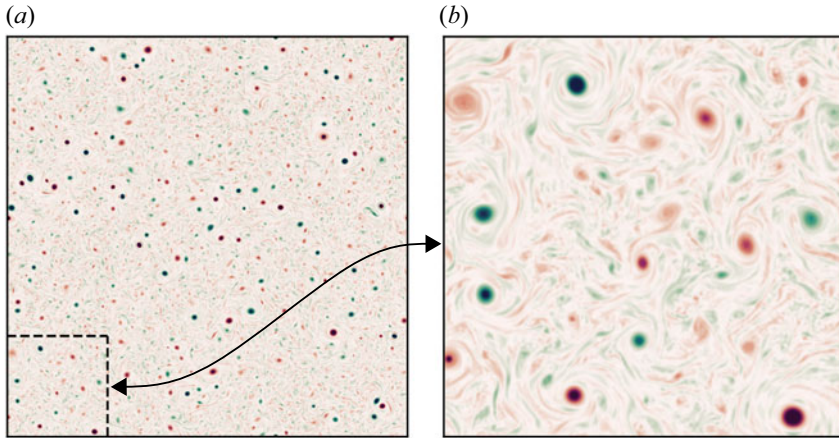


Figure 3. Snapshot of the vorticity field in the forced–dissipative model at stationarity showing  $x, y \in [0, 2\pi]$  (a) and  $x, y \in [0, \pi/2]$  (b).

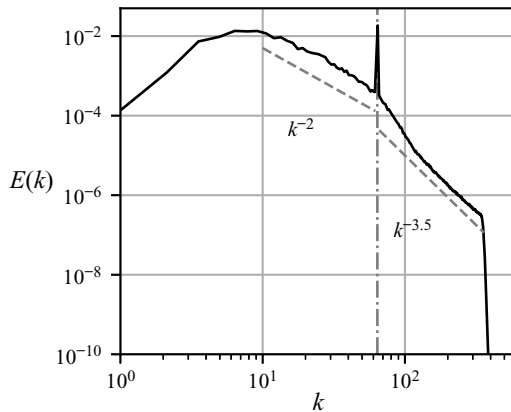


Figure 4. Snapshot of the isotropic energy spectrum in the forced–dissipative model at stationarity.

Figure 3 shows a snapshot of the vorticity field at the end of the spin-up process. Enstrophy is concentrated in a population of coherent vortices whose scale is set by the forcing scale,  $k_F^{-1}$ . Figure 4 shows the isotropic energy spectrum calculated at the same instant. A power law of approximately  $k^{-2}$  is observed at wavenumbers between the peak wavenumber at  $k \approx 6$  and the forcing wavenumber at  $k \approx 64$  (indicated with a vertical line). A second power law of approximately  $k^{-3.5}$  is seen at wavenumbers between the forcing scale and the dissipation range. Large-scale friction prevents the indefinite accumulation of energy at the largest scales, while continued forcing prevents energy from concentrating exclusively around a peak wavenumber at late times, and, by inputting enstrophy at a moderate scale, sustains a lively population of vortices.

## 5.2. Particle numerics

After spin-up, a set of 1000 passive tracer particles are evolved in the flow of the forced–dissipative model for approximately another 6800 eddy turnover times; this is done using bilinear interpolation of the velocity field and the fourth-order Runge–Kutta

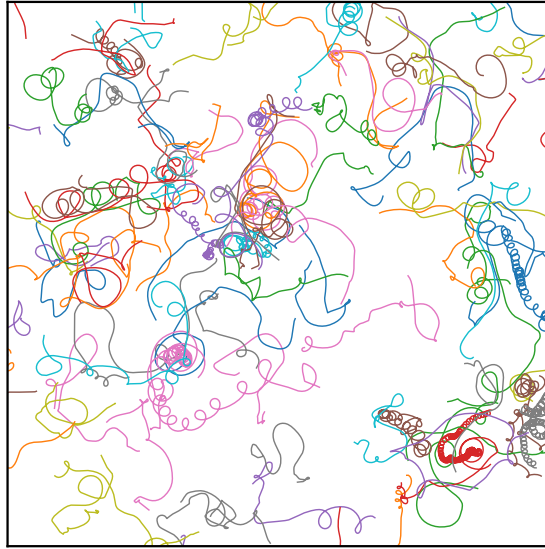


Figure 5. Trajectories of 100 passive particles advected in the forced–dissipative model, shown as recorded over a period of  $100\tau_c$  with a different colour for each trajectory.

timestepping scheme. Particles are seeded at initial positions chosen uniformly at random in the domain. Figure 5 shows a subset of the trajectory data generated. Some particles follow highly oscillatory paths, while others do not, depending on whether they are seeded in the interior of a coherent vortex or in the background turbulence.

### 5.3. Diagnostics

To illustrate the dynamics that we parameterise with the stochastic models we show two diagnostics commonly used in Lagrangian analyses (Pasquero *et al.* 2001; van Sebille *et al.* 2018), namely, the Lagrangian velocity autocovariance function (LVAF), defined in the isotropic case as

$$r(\tau) = \langle U^{(p)}(t + \tau)U^{(p)}(t) \rangle, \quad (5.7)$$

where the angled brackets denote the average over  $t$  and particles  $p$ , and the absolute diffusivity

$$\kappa_{abs}(\tau) = \frac{\langle (\Delta X^{(p)}(\tau))^2 \rangle}{2\tau}. \quad (5.8)$$

Figure 6 shows the LVAF as estimated from the simulated particle trajectory data. The corresponding LVAF of the Brownian model is a delta function at zero, since velocity is implicitly represented as a white noise process, while the LVAF of the Langevin model, which represents particle velocity as an Ornstein–Uhlenbeck process, is

$$r_{OU}(\tau; \gamma, k) = k\gamma \exp(-\gamma|\tau|). \quad (5.9)$$

In contrast, the estimated LVAF of the forced–dissipative model not only shows finite decorrelation time but is noticeably sub-exponential.

Figure 7 shows the estimated absolute diffusivity. In line with the asymptotic laws described in Taylor (1922) the absolute diffusivity is linear at small  $\tau$  and constant at

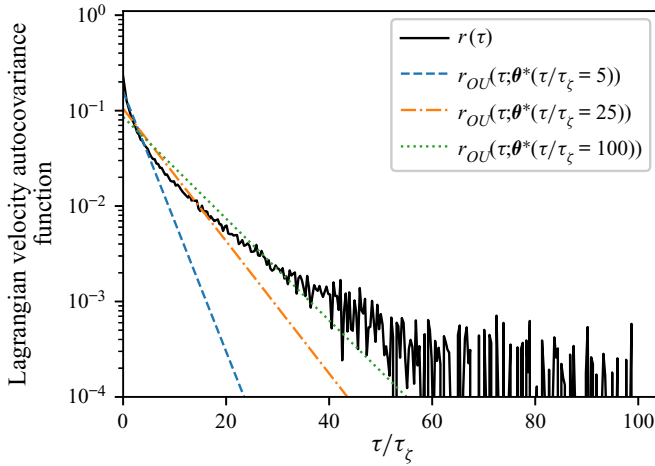


Figure 6. LAVAF  $r(\tau)$  for the forced–dissipative model, as estimated from the full set of 1000 simulated particle trajectories. The LAVAF of the Langevin model  $r_{OU}(\tau)$  is also shown using posterior mode estimates (discussed below)  $\theta^* = (\gamma^*, k^*)$  derived from datasets with  $\tau = (5, 25, 100)\tau_\zeta$ , respectively (see figure 8).

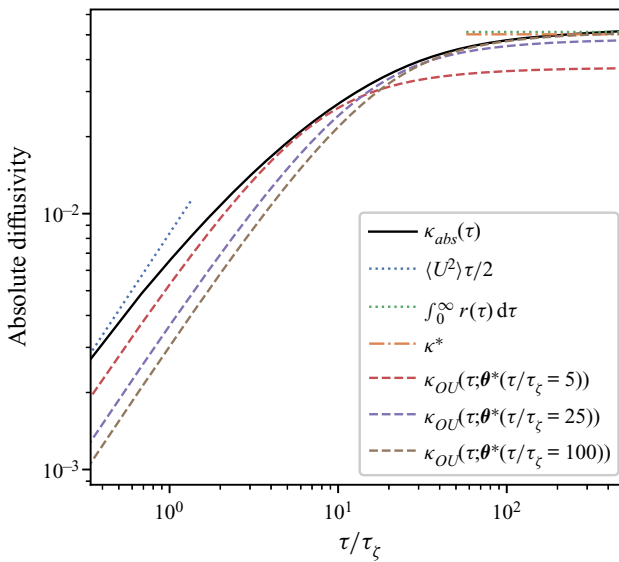


Figure 7. Absolute diffusivity,  $\kappa_{abs}(\tau)$ , for the forced–dissipative model, as estimated from the full set of 1000 simulated particle trajectories. A posterior mode estimate  $\kappa^*$  is shown, along with two asymptotic laws:  $\kappa_{abs}(\tau) = \text{linear}$  (ballistic regime), and  $\kappa_{abs}(\tau) = \text{const.}$  (diffusive regime).

large  $\tau$ , corresponding to the ballistic and diffusive regimes, respectively. The absolute diffusivity of the Brownian model is constant, while that of the Langevin model is

$$\kappa_{OU}(\tau) = k(1 - \varphi(\gamma\tau)), \tag{5.10}$$

which is linear at small  $\tau$  and constant at large  $\tau$ .

Qualitatively, from comparing these diagnostics with those of the stochastic models it is clear that, on sufficiently large times (in the diffusive regime), the Brownian model is valid; in particular, the LAVAF is well approximated by a delta function at large times,

and correspondingly, the absolute diffusivity is constant. On time scales shorter than the diffusive regime the LVAF of the observed trajectories may be better approximated by that of the Langevin model; however, the quality of this approximation is in general unclear *a priori*. It could be tempting to estimate  $\gamma$  by fitting the LVAF, using e.g. a least-squares method, but this approach would not correctly deal with uncertainty in parameters.

#### 5.4. Parameter inference and BMC

We now apply the parameter inference and BMC procedures demonstrated in the test case of § 4.4. By subsampling the results of our particle simulations we generate datasets with  $N_p = 1000$ ,  $N_\tau = 25$ , and a set of sampling times  $\tau$  in the range  $[\tau_\zeta, 250\tau_\zeta]$ .

Prior means for the parameters are derived from  $\tau_\zeta$  and the root-mean-square velocity  $u_{rms} = \sqrt{2E}$ , where  $E$  is the mean energy: as discussed in § 4.2 we take

$$\mathbb{E}[\kappa] = \mathbb{E}[k] = u_{rms}^2 \tau_\zeta, \tag{5.11a}$$

$$\mathbb{E}[\gamma] = \tau_\zeta^{-1}, \tag{5.11b}$$

and use the corresponding exponential distributions as priors.

The results of the parameter inference are shown in figure 8. The Brownian model is sampling-time consistent for  $\tau \gtrsim 150\tau_\zeta$ , with a posterior mode that differs by 40% from the scaling estimate used as prior mean. The long time required for sampling-time consistency is in line with the expected validity of the Brownian model in the long-time limit. In this limit  $\kappa^*$  agrees very well with Taylor’s (1922) theoretical prediction,  $\kappa = \int_0^\infty r(\tau) d\tau$  (see figure 7).

The Langevin model is roughly sampling-time consistent from much smaller values of  $\tau$ , say  $\tau \gtrsim 50\tau_\zeta$ . This suggests that there is a range of sampling times, roughly  $50\tau_\zeta \lesssim \tau \lesssim 150\tau_\zeta$ , where the Langevin model is potentially useful but the Brownian model is not. The BMC analysis below sheds further light on this. However, there is noticeable decay in the posterior mode estimates of  $\gamma$  with increasing  $\tau$  – this is likely a reflection of the sub-exponential nature of the true LVAF. In figure 6 we plot the Langevin LVAF given parameters inferred with data of various  $\tau$ , where the decay in estimates of  $\gamma$  corresponds to a shallowing of the Langevin LVAF. In figure 7 we plot the absolute diffusivity of the Langevin model  $\kappa_{OU}$  using the same parameter estimates as in figure 6. The absolute diffusivity is best approximated at a time scale matching the sampling time of the data. The posterior mode of  $\gamma$ , when roughly stable, is almost an order of magnitude smaller than the scaling estimate in (5.11), indicating that particle dynamics decorrelate slower than might be predicted by a naive dimensional analysis based on the enstrophy alone. In particular, the inferred value of  $\gamma$  corresponds to a decorrelation time of approximately 8 eddy turnover times. As in the test case of § 4.4 the posterior standard deviation of  $\gamma$  grows with  $\tau$  as the diffusive limit is reached and the particle dynamics becomes insensitive to  $\gamma$ . It is interesting to note that the Langevin diffusivity  $k$  is estimated consistently for sampling times much shorter than those required to estimate the Brownian diffusivity  $\kappa$  even though their values are identical when the Brownian model represents the dispersion well. This suggests that carrying out Bayesian inference of the Langevin model might provide a means to estimate the Brownian diffusivity when data are not available over the long, diffusive time scales that are required *a priori*. This may generalise to other flows only when the Langevin model is a reasonable approximation – inference of  $k$  is unlikely to be sampling-time consistent if inference of  $\gamma$  is not, for example, due to the LVAF of the flow of interest being very far from exponential. We emphasise that the inference results just described are largely insensitive to specification of the prior.



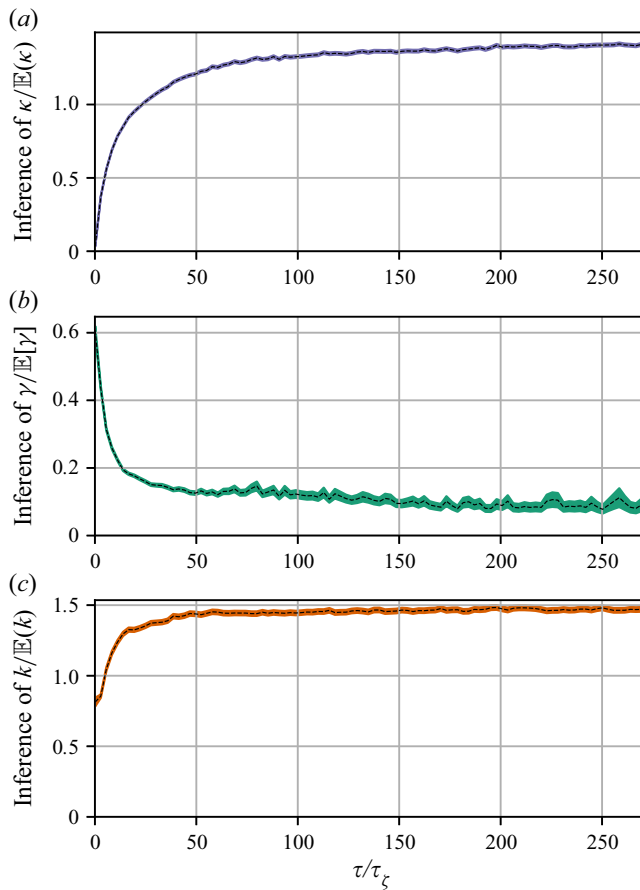


Figure 8. Parameter inference for the Brownian and Langevin models as a function of observation interval,  $\tau$ , for data from the two-dimensional turbulence model. Dashed lines indicate posterior mode estimates,  $\theta^*$ , normalised with respect to prior means, and shaded areas are  $\theta^* \pm \text{SD}(\theta|\Delta\mathcal{X}_\tau)$ . Each inference is made with a fixed volume of data:  $N_p = 1000$  and  $N_\tau = 25$ .

The results of the BMC for the turbulence model data are shown in figure 9. The picture is similar to that in the test case of § 4.4, in that the Bayes factor favours the Langevin model for shorter time scales, but with diminishing strength as  $\tau$  is increased, until, at time scales corresponding to convergence of the Brownian diffusivity, the value of the log Bayes factor becomes small enough that the two models cannot be meaningfully discriminated.

Also shown in figure 9 are the corresponding log Occam factors. For  $\tau$  large enough that the Brownian model is sampling-time consistent, its Occam factor is approximately constant and larger than that of the Langevin model. As in the test case in § 4.4, the Occam factor for the Langevin model increases towards that of the Brownian model at large  $\tau$  when the particle dynamics is sufficiently decorrelated that the likelihood is less sensitive to the value of  $\gamma$ , albeit more slowly, owing to the more slowly decaying LVAF of the turbulent dynamics. The difference in log Occam factors is much smaller than the difference in the corresponding maximum log likelihoods for all but the largest values of  $\tau$ , which explains why the Bayes factor mainly favours the Langevin model.

In summary, these results indicate that, while the Brownian model is adequate on sufficiently large time scales ( $\tau \gtrsim 150\tau_\zeta$ ), the Langevin model can explain better the

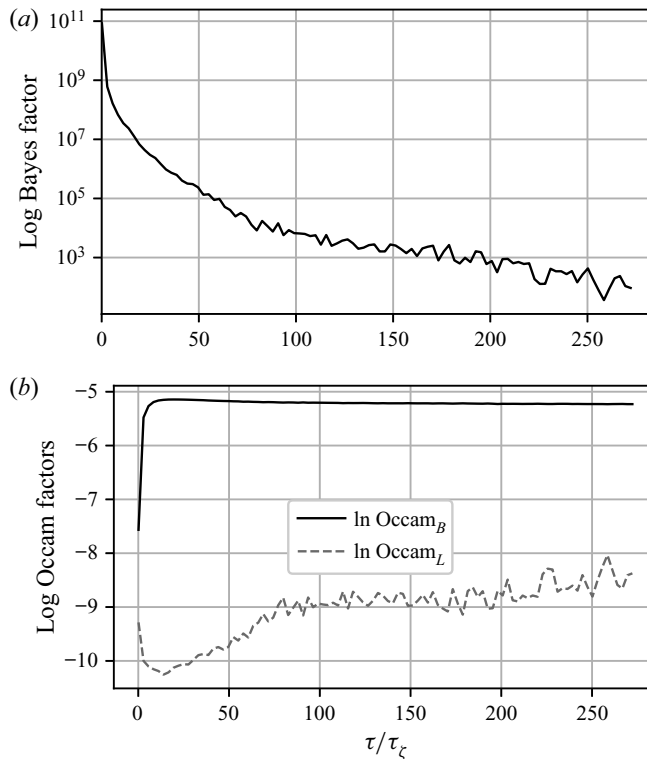


Figure 9. Log Bayes factors,  $\ln K_{L,B}$ , and corresponding log Occam factors, as a function of  $\tau$ , given the same data used for figure 8.

dynamics of tracer particles in the turbulence model of § 5.1 on shorter time scales ( $50\tau_\zeta \lesssim \tau \lesssim 150\tau_\zeta$ ). On time scales  $\tau \gtrsim 150\tau_\zeta$  the two models are indistinguishable in their performance, so that in this regime the Brownian model should be favoured in practice as a more parsimonious description.

## 6. Conclusions

We have demonstrated the application of BMC to a problem of interest in fluid dynamics, and shown that we can compare the performance of competing stochastic models of particle dynamics given discrete trajectory data alone while accounting for parameter uncertainty. In particular, we found that the Langevin model is preferred over the Brownian model for describing the particle dynamics in a model of two-dimensional turbulence on a range of time scales, but that on sufficiently large time scales the two models perform equally well.

The broad conclusion of the BMC, then, is that the additional complexity of the Langevin model, associated with the presence of an additional parameter, is justified: its better capability to explain the data, as quantified by the maximum likelihood, overwhelms the penalty for complexity quantified by the Occam factor. We stress, however, that this conclusion does not take into account the computational cost involved if the models are used for predictions.

The application of the BMC method to other problems is limited by the feasibility of the calculation of the model evidence. Specifically, BMC inherits the usual challenges

of Bayesian and likelihood-based inference procedures, namely that the likelihood can be intractable or expensive to compute for complex models – the Brownian and Langevin models considered here, as linear stochastic differential equations, are very simple examples whose likelihoods could be computed analytically – alternative models which are nonlinear, have higher dimension or have more complicated correlation structure will likely have intractable likelihoods. For example, for spatially inhomogeneous flows, such as in the atmosphere or oceans, nonlinear models arise with spatially varying (and hence high-dimensional) parameters. Fortunately, the collection of methods referred to as approximate Bayesian computation have been developed to deal with this problem. For example, Carson *et al.* (2018) used the SMC<sup>2</sup> (‘sequential Monte Carlo squared’) algorithm to compare SDE models of glacial–interglacial cycles with intractable likelihoods.

There is the further issue of performing the integration required to obtain the evidence as in (3.3). When the posterior is sufficiently Gaussian like, i.e. peaked around a single mode, Laplace’s method can be very accurate (Kass & Raftery 1995) as well as cheap, however, this requires (at least an approximation to) the Hessian of the log posterior at its mode. Aside from Laplace’s method, Krog & Lomholt (2017) and Thapa *et al.* (2018) have implemented the nested sampling algorithm of Skilling (2004) to calculate the evidence in similar analyses, while the line of work by Hannart *et al.* (2016), Carrassi *et al.* (2017) and Metref *et al.* (2019) has sought to perform model evidence estimation using ensemble-based data assimilation methods originally designed for state estimation in the context of incomplete, noisy observations of high-dimensional dynamical systems.

While BMC inevitably comes with computational challenges in complex problems, there are many cases where it can feasibly be applied, and, where it cannot, it should serve as a useful theoretical starting point, with alternative methods measured by how well their conclusions agree with those of BMC.

**Funding.** M.B. was supported by the MAC-MIGS Centre for Doctoral Training under EPSRC grant EP/S023291/1. J.V. was supported by the EPSRC Programme Grant EP/R045046/1: Probing Multiscale Complex Multiphase Flows with Positrons for Engineering and Biomedical Applications (PI: Professor M. Barigou, University of Birmingham).

**Declaration of interests.** The authors report no conflict of interest.

**Data availability statement.** The code required to reproduce the results herein is available at [doi.org/10.5281/zenodo.5820320](https://doi.org/10.5281/zenodo.5820320) and the trajectory data generated with the 2-D turbulence model are available at [doi.org/10.7488/ds/3267](https://doi.org/10.7488/ds/3267).

#### Author ORCIDs.

-  Martin T. Brolly <https://orcid.org/0000-0002-8508-2847>;
-  James R. Maddison <https://orcid.org/0000-0001-5742-4363>;
-  Aretha L. Teckentrup <https://orcid.org/0000-0001-5089-0476>;
-  Jacques Vanneste <https://orcid.org/0000-0002-0319-589X>.

## Appendix A. Langevin likelihood for position observations

To derive  $p(\Delta\mathcal{X}_\tau|\mathcal{M}_L(\gamma, k))$  we simplify notation by recognising that all particles are independent under  $\mathcal{M}_L$  and that the dynamics in each spatial dimension are independent. We therefore need only calculate  $p(\Delta\mathcal{X}_\tau|\mathcal{M}_L(\gamma, k))$  in the one-dimensional, single-particle case. We proceed by: (i) showing that the joint process of particle position and velocity is an order-one vector autoregressive process, or VAR(1) process, and hence, has a Gaussian likelihood, (ii) calculating the mean and covariance for a sequence

of joint position–velocity observations and (iii) marginalising this likelihood to find  $p(\Delta\mathcal{X}_\tau|\mathcal{M}_L(\gamma, k))$ .

It can be shown that for the one-dimensional Langevin equation

$$Y_n|Y_{n-1} \sim \mathcal{N}\left(\begin{pmatrix} U_n\varphi(\gamma\tau)\tau \\ U_n e^{-\gamma\tau} \end{pmatrix}, \mathbf{C}\right), \tag{A1}$$

where  $Y_n := (\Delta X_n, U_{n+1})^T$  and

$$C_{11} = 2k\tau(1 - 2\varphi(\gamma\tau) + \varphi(2\gamma\tau)), \tag{A2a}$$

$$C_{12} = C_{21} = k(\varphi(\gamma\tau)\gamma\tau)^2, \tag{A2b}$$

$$C_{22} = 2k\gamma^2\tau\varphi(2\gamma\tau). \tag{A2c}$$

This follows from the well-known solution of the Ornstein–Uhlenbeck process,

$$U(t) = U(0)\exp(-\gamma t) + \gamma\sqrt{2k} \int_0^t \exp(-\gamma(t-t')) dW(t') \tag{A3}$$

and the corresponding solution for the position,

$$X(t) = X(0) + \int_0^t U(t') dt' \tag{A4a}$$

$$= X(0) + U(0)\varphi(\gamma t)t - \sqrt{2k} \int_0^t \exp(-\gamma(t-t')) dW(t') + \sqrt{2k} W(t). \tag{A4b}$$

Therefore, we can write the Langevin model in the time-discretised form

$$Y_n = \mathbf{A}Y_{n-1} + \boldsymbol{\varepsilon}_n, \tag{A5}$$

where

$$\mathbf{A} = \begin{pmatrix} 0 & \varphi(\gamma\tau)\tau \\ 0 & e^{-\gamma\tau} \end{pmatrix}, \tag{A6}$$

and  $\boldsymbol{\varepsilon}_n$  is a mean-zero white noise process with covariance matrix  $\mathbf{C}$ .

The discrete process (A5) has the form of a VAR(1) process. Furthermore,  $Y_n$  is stationary with mean and stationary variance

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} 2k\tau(1 - \varphi(\gamma\tau)) & k\varphi(\gamma\tau)\gamma\tau \\ k\varphi(\gamma\tau)\gamma\tau & k\gamma \end{pmatrix}. \tag{A7a,b}$$

To see this, note that the marginal distribution of  $U(t)$  at any time is given by the stationary distribution of the Ornstein–Uhlenbeck process,

$$U(t) \sim \mathcal{N}(0, k\gamma), \tag{A8}$$

which gives  $\mu_2$  and  $V_{22}$ . Using (A1), (A8) and Lemma B.1 given in Appendix B yields  $\mu_1$  and  $V_{11}$ . Finally,  $V_{12} = V_{21}$  can be calculated using (A1) and the law of total covariance – specifically

$$\text{Cov}(\Delta X_n, U_{n+1}) = \mathbb{E}[\text{Cov}(\Delta X_n, U_{n+1}|U_n)] + \text{Cov}(\mathbb{E}[\Delta X_n|U_n], \mathbb{E}[U_{n+1}|U_n]) \tag{A9a}$$

$$= C_{12} + \text{Cov}(U_n\varphi(\gamma\tau)\tau, U_n e^{-\gamma\tau}) \tag{A9b}$$

$$= C_{12} + \varphi(\gamma\tau)\tau e^{-\gamma\tau} \text{Var}(U_n) \tag{A9c}$$

$$= k\varphi(\gamma\tau)\gamma\tau, \tag{A9d}$$

recalling that  $\text{Var}(U_n) = k\gamma$ .

The autocovariance of  $Y_n$  is defined as

$$\mathbf{G}(m) := \mathbb{E}[(Y_n - \boldsymbol{\mu})(Y_{n-m} - \boldsymbol{\mu})^T], \tag{A10}$$

where  $m \in \mathbb{Z}$ . Notice that  $\mathbf{G}(0)$  is the stationary variance of  $Y_n$ . Postmultiplying (A5) by  $Y_{n-m}^T$  and taking expectations gives

$$\mathbb{E}[Y_n Y_{n-m}^T] = \mathbf{A} \mathbb{E}[Y_{n-1} Y_{n-m}^T] + \mathbb{E}[\boldsymbol{\varepsilon}_n Y_{n-m}^T]. \tag{A11}$$

Thus, for  $m > 0$ , since  $Y_{n-m}$  is independent of  $\boldsymbol{\varepsilon}_n$ ,

$$\mathbf{G}(m) = \mathbf{A} \mathbf{G}(m-1). \tag{A12}$$

Therefore,  $\mathbf{G}(m)$  can be calculated recursively for  $m > 0$  as

$$\mathbf{G}(m) = \mathbf{A}^m \mathbf{G}(0) \tag{A13a}$$

$$= \mathbf{A}^m \mathbf{V}. \tag{A13b}$$

Note that (A12) is an instance of a Yule–Walker equation (Lütkepohl 2007, pp. 26–27).

Thus, the joint distribution of a sequence of observations  $\{Y_n : n \in \{0, \dots, N_\tau - 1\}\}$  is given by

$$\begin{pmatrix} Y_0 \\ Y_1 \\ \vdots \\ Y_{N_\tau-1} \end{pmatrix} \sim \mathcal{N} \left( \mathbf{0}, \begin{pmatrix} \mathbf{G}(0) & \mathbf{G}(1) & \dots & \mathbf{G}(N_\tau - 1) \\ \mathbf{G}(1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{G}(1) \\ \mathbf{G}(N_\tau - 1) & \mathbf{G}(1) & \mathbf{G}(0) & \mathbf{G}(0) \end{pmatrix} \right). \tag{A14}$$

Marginalising (A14) for the distribution of  $(\Delta X_0, \dots, \Delta X_{N_\tau-1})^T$  we find

$$\begin{pmatrix} \Delta X_0 \\ \Delta X_1 \\ \vdots \\ \Delta X_{N_\tau-1} \end{pmatrix} \sim \mathcal{N} \left( \mathbf{0}, \begin{pmatrix} \mathbf{G}_{11}(0) & \mathbf{G}_{11}(1) & \dots & \mathbf{G}_{11}(N_\tau - 1) \\ \mathbf{G}_{11}(1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{G}_{11}(1) \\ \mathbf{G}_{11}(N_\tau - 1) & \mathbf{G}_{11}(1) & \mathbf{G}_{11}(1) & \mathbf{G}_{11}(0) \end{pmatrix} \right). \tag{A15}$$

Using (A7a,b) and (A12) it is easy to see that for  $m \geq 1$

$$\mathbf{G}(m) = \begin{pmatrix} k\gamma\tau^2\varphi^2(\gamma\tau) \exp(-(m-1)\gamma\tau) & k\gamma\tau\varphi(\gamma\tau) \exp(-(m-1)\gamma\tau) \\ k\gamma\tau\varphi(\gamma\tau) \exp(-m\gamma\tau) & k\gamma \exp(-m\gamma\tau) \end{pmatrix}. \tag{A16}$$

Hence, in particular,

$$\mathbf{G}_{11}(m) = k\gamma\tau^2\varphi^2(\gamma\tau) \exp(-(m-1)\gamma\tau). \tag{A17}$$

The likelihood  $p(\Delta X_0, \dots, \Delta X_{N_\tau-1})$  is determined by (A15) and (A17).

### Appendix B

LEMMA B.1. *If  $X \sim \mathcal{N}(\mu, \sigma^2)$  and  $Y|X \sim \mathcal{N}(aX, \tau^2)$ , then  $Y \sim \mathcal{N}(a\mu, a^2\sigma^2 + \tau^2)$ .*

*Proof.* Consider the moment generating function of  $Y$ ,  $M_Y(t)$ . Recall that for a Gaussian random variable such as  $X$  the moment generating function is

$$M_X(t) := \mathbb{E}_X[\exp(Xt)] = \exp(\mu t + \sigma^2 t^2/2); \quad (\text{B1})$$

similarly, since  $aX \sim \mathcal{N}(a\mu, a^2\sigma^2)$ ,

$$M_{aX}(t) := \mathbb{E}_X[\exp(aXt)] = \exp(a\mu t + a^2\sigma^2 t^2/2). \quad (\text{B2})$$

Now,

$$M_Y(t) = \mathbb{E}_Y[\exp(Yt)] \quad (\text{B3a})$$

$$= \mathbb{E}_X[\mathbb{E}_{Y|X}[\exp(Yt)]] \quad (\text{B3b})$$

$$= \mathbb{E}_X[\exp(aXt + \tau^2 t^2/2)] \quad (\text{B3c})$$

$$= \exp(\tau^2 t^2/2) \mathbb{E}_X[\exp(aXt)] = \exp((a\mu)t + (a^2\sigma^2 + \tau^2)t^2/2), \quad (\text{B3d})$$

which we can recognise as the moment generating function of a Gaussian random variable with mean  $a\mu$  and variance  $a^2\sigma^2 + \tau^2$ . ■

#### REFERENCES

- ARBIC, B.K. & FLIERL, G.R. 2003 Coherent vortices and kinetic energy ribbons in asymptotic, quasi two-dimensional  $f$ -plane turbulence. *Phys. Fluids* **15** (8), 2177–2189.
- BERLOFF, P.S. & MCWILLIAMS, J.C. 2002 Material transport in oceanic gyres. Part II: hierarchy of stochastic models. *J. Phys. Oceanogr.* **32** (3), 797–830.
- BERNARDO, J.M. 1979 Expected information as expected utility. *Ann. Stat.* **7** (3), 686–690.
- BERNARDO, J.M. & SMITH, A.F.M. 1994 *Bayesian Theory*. Wiley.
- BRÖCKER, J. & SMITH, L.A. 2007 Scoring probabilistic forecasts: the importance of being proper. *Weather Forecast.* **22** (2), 382–388.
- CARRASSI, A., BOCQUET, M., HANNART, A. & GHIL, M. 2017 Estimating model evidence using data assimilation. *Q. J. R. Meteorol. Soc.* **143** (703), 866–880.
- CARSON, J., CRUCIFIX, M., PRESTON, S. & WILKINSON, R.D. 2018 Bayesian model selection for the glacial–interglacial cycle. *J. R. Stat. Soc. C: Appl.* **67** (1), 25–54.
- CLYDE, M. & IVERSEN, E.S. 2013 Bayesian model averaging in the M-open framework. In *Bayesian Theory and Applications* (ed. P. Damien, P. Dellaportas, N.G. Polson, & D.A. Stephens). Oxford University Press.
- COTTER, C.J. & PAVLIOTIS, G.A. 2009 Estimating eddy diffusivities from noisy Lagrangian observations. *Commun. Math. Sci.* **7** (4), 805–838.
- COVER, T.M. & THOMAS, J.A. 2006 *Elements of Information Theory*. John Wiley and Sons, Inc.
- DU, H. 2021 Beyond strictly proper scoring rules: the importance of being local. *Weather Forecast.* **36** (2), 457–468.
- ESLER, J.G. & RAMLI, H.M. 2017 Shear dispersion in the turbulent atmospheric boundary layer. *Q. J. R. Meteorol. Soc.* **143** (705), 1721–1733.
- GNEITING, T. & RAFTERY, A.E. 2007 Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **102** (477), 359–378.
- GOOD, I.J. 1952 Rational decisions. *J. R. Stat. Soc. B: Met.* **14** (1), 107–114.
- GRIFFA, A. 1996 *Applications of Stochastic Particle Models to Oceanographic Problems*, pp. 113–140. Birkhäuser.
- HANNART, A., CARRASSI, A., BOCQUET, M., GHIL, M., NAVEAU, P., PULIDO, M., RUIZ, J. & TANDEO, P. 2016 DADA: data assimilation for the detection and attribution of weather and climate-related events. *Clim. Change* **136** (2), 155–174.
- JAYNES, E.T. 2003 *Probability Theory: The Logic of Science*. Cambridge University Press.
- JEFFREYS, H. 1983 *Theory of Probability*. Clarendon.
- KASS, R.E. & RAFTERY, A.E. 1995 Bayes factors. *J. Am. Stat. Assoc.* **90** (430), 773–795.
- KROG, J. & LOMHOLT, M.A. 2017 Bayesian inference with information content model check for Langevin equations. *Phys. Rev. E* **96** (6), 062106.
- LILLY, J.M., SYKULSKI, A.M., EARLY, J.J. & OLHEDE, S.C. 2017 Fractional Brownian motion, the Matérn process, and stochastic modeling of turbulent dispersion. *Nonlinear Process. Geophys.* **24** (3), 481–514.



## Bayesian comparison of stochastic models of dispersion

- LÜTKEPOHL, H. 2007 *New Introduction to Multiple Time Series Analysis*. Springer.
- MACKAY, D.J.C. 2003 *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.
- MAJDA, A.J. & KRAMER, P.R. 1999 Simplified models for turbulent diffusion: theory, numerical modelling, and physical phenomena. *Phys. Rep.* **314** (4), 237–574.
- MANN, R.P. 2011 Bayesian inference for identifying interaction rules in moving animal groups. *PLoS One* **6** (8), e22827.
- MARK, C., METZNER, C., LAUTSCHAM, L., STRISSEL, P.L., STRICK, R. & FABRY, B. 2018 Bayesian model selection for complex dynamic systems. *Nat. Commun.* **9** (1), 1803.
- METREF, S., HANNART, A., RUIZ, J., BOCQUET, M., CARRASSI, A. & GHIL, M. 2019 Estimating model evidence using ensemble-based data assimilation with localization – the model selection problem. *Q. J. R. Meteorol. Soc.* **145** (721), 1571–1588.
- MIN, S.-K., SIMONIS, D. & HENSE, A. 2007 Probabilistic climate change predictions applying Bayesian model averaging. *Phil. Trans. R. Soc. Lond. A* **365** (1857), 2103–2116.
- NELDER, J.A. & MEAD, R. 1965 A simplex method for function minimization. *Comput. J.* **7** (4), 308–313.
- ONINK, V., VAN SEBILLE, E. & LAUFKÖTTER, C. 2022 Empirical Lagrangian parametrization for wind-driven mixing of buoyant particles at the ocean surface. *Geosci. Model Dev.* **15** (5), 1995–2012.
- PASQUERO, C., PROVENZALE, A. & BABIANO, A. 2001 Parameterization of dispersion in two-dimensional turbulence. *J. Fluid Mech.* **439**, 279–303.
- PAVLIOTIS, G.A. 2014 *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations*. Texts in Applied Mathematics, vol. 60. Springer.
- PROVENZALE, A., BABIANO, A. & VILLONE, B. 1995 Single-particle trajectories in two-dimensional turbulence. *Chaos, Solitons Fractals* **5** (10), 2055–2071.
- REIJNDERS, D., DELEERSNIJDER, E. & VAN SEBILLE, E. 2022 Simulating Lagrangian subgrid-scale dispersion on neutral surfaces in the ocean. *J. Adv. Model. Earth Syst.* **14** (2), e2021MS002850.
- ROBERT, C.P. 2007 *The Bayesian choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd edn. Springer.
- RODEAN, H.C. 1996 *Stochastic Lagrangian Models of Turbulent Diffusion*. Springer.
- ROULSTON, M.S. & SMITH, L.A. 2002 Evaluating probabilistic forecasts using information theory. *Mon. Weath. Rev.* **130** (6), 1653–1660.
- SCOTT, R.K. 2007 Nonrobustness of the two-dimensional turbulent inverse cascade. *Phys. Rev. E* **75**, 046301.
- VAN SEBILLE, E. *et al.* 2018 Lagrangian ocean analysis: fundamentals and practices. *Ocean Model.* **121**, 49–75.
- SKILLING, J. 2004 Nested sampling. *AIP Conf. Proc.* **735** (1), 395–405.
- TAYLOR, G.I. 1922 Diffusion by continuous movements. *Proc. Lond. Math. Soc.* **s2-20** (1), 196–212.
- THAPA, S., LOMHOLT, M.A., KROG, J., CHERSTVY, A.G. & METZLER, R. 2018 Bayesian analysis of single-particle tracking data using the nested-sampling algorithm: maximum-likelihood model selection applied to stochastic-diffusivity data. *Phys. Chem. Chem. Phys.* **20** (46), 29018–29037.
- THOMSON, D.J. 1987 Criteria for the selection of stochastic models of particle trajectories in turbulent flows. *J. Fluid Mech.* **180**, 529–556.
- UHLENBECK, G.E. & ORNSTEIN, L.S. 1930 On the theory of the Brownian motion. *Phys. Rev.* **36**, 823–841.
- VAN DER VAART, A.W. 1998 *Asymptotic Statistics*. Cambridge University Press.
- VALLIS, G.K. 2017 *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*, 2nd edn. Cambridge University Press.
- YING, Y.K., MADDISON, J.R. & VANNESTE, J. 2019 Bayesian inference of ocean diffusivity from Lagrangian trajectory data. *Ocean Model.* **140**, 101401.
- ZHU, C., BYRD, R.H., LU, P. & NOCEDAL, J. 1997 Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.* **23** (4), 550–560.