Bayesian prior impact assessment for dynamical systems described by ordinary differential equations

Damian N. Mingo[a], Jack S. Hale[a], Christophe Ley[b]

[a]*Institute of Computational Engineering, Department of Engineering, Faculty of Science, Technology and Medicine, University of Luxembourg, Luxembourg.*
[b]*Department of Mathematics, Faculty of Science, Technology and Medicine, University of Luxembourg, Luxembourg.*

Abstract

In this study we extend the use of the Wasserstein Impact Measure (WIM) to the problem of assessing prior impact in Bayesian models governed by systems of ordinary differential equations (ODEs) with moderate (5 to 10) parametric dimension. First, we utilise algorithms from computational optimal transport to compute the WIM in moderate parametric dimensions. Second, we propose a new prior scaled Wasserstein Impact Measure (sWIM) measure which gives a relative sense of distance, easing with interpretation of the WIM for understanding the impact of the prior on the resulting inference. We show numerical computation and interpretation of the WIM and sWIM for a Lotka-Volterra predator-prey model calibrated against the Hudson Bay Company dataset and a compartment epidemiological model calibrated against first-wave COVID-19 data from Luxembourg.

Keywords: Bayes theorem, prior impact, Wasserstein Impact Measure, dynamical systems, ordinary differential equations, Lotka-Volterra model, SEIR model

1. Introduction

A strength of Bayesian statistics is the ability to easily incorporate prior information, such as historical or expert knowledge, into a parameter inference problem. However, the inherent price to pay for this flexibility is that the choice of prior may have a strong impact on the subsequent parameter inference. Consequently a crucial component of Bayesian analysis is the choice and justification of a prior, along with understanding the impact of prior choice on the resulting parameter estimates and model output.

Bayesian methods are widely used for solving parameter identification problems of continuous time dynamical systems modelled by ordinary differential equatiosince (ODEs) [1]. Fields of study where ODEs are a prominent methodology for constructing models include epidemiology [2, 3], hydrology [4] and mechanics [5]. These models typically contain 5 to 10 unknown parameters that have a critical effect on the overall behaviour of the system, and so it is necessary to identify these parameters using observed data before the model can be put to use.

We now discuss some of the issues particular to solving parameter identification problems involving ODEs. Firstly, the observed data to calibrate the model is often sparse in the sense that there are only observations available at a limited number of points in time, and for each time point, usually only a single (noisy) observation. Additionally there are usually only observations on a limited subset of the system states. In some cases, we may only be able to observe a proximal function of the system states, for example a weighted sum. Due to this sparsity it is often the case that the data can only weakly constrain the parameters, leading to inference problems that are inherently

ill-posed. Aside from this notion of data sparsity, the parameters themselves are usually constrained *a priori* by fundamental physical considerations, e.g. positivity, or by a large body of expert knowledge from previous related studies on perceived 'sensible' values.

Because of these issues, Bayesian methods, precisely because of the inclusion of a prior, are a natural choice for parameter identification problems involving ODEs. Specifically, the inclusion of prior knowledge can turn a problem that is inherently ill-conditioned due to data sparsity, into a well-conditioned problem with a reasonable solution. Furthermore, the prior gives a way to directly encode information about physical constraints on the parameters and/or to incorporate historical data. Because of this priors used with ODEs are rarely of the non-informative or objective type [6, 7], with a preference for weakly informative [8, 9] or strongly informative priors [10, 11, 12]. We also remark that in these problems we are almost always working far from the regime where Bernstein–von–Mises type results may be applicable. Furthermore it has been shown in e.g. [13, 8, 9] that weakly and strongly informative priors can lead to reduced computational cost over non-informative priors when exploring the posterior distribution using Markov Chain Monte Carlo (MCMC). Due to this widespread use of informative priors, it is desirable to develop quantitative methods to assess how the choice of prior impacts posterior parameter estimates. For example, do the priors have similar impact or does one have a higher impact on the posterior?

Before proceeding to our contribution, we take a moment to review the existing literature on assessing prior impact in a general Bayesian model context - we are not aware of any studies that discuss quantitative prior impact

methods in the context of ordinary differential equation (ODE) systems. A common approach is to recalibrate the model with slightly different priors and judge how it impacts the resulting inference [14, 15, 16, 17]. Such an approach is qualitative as it tells us if different priors lead to different posterior statistics, but there is typically no quantitative measure of the difference between the resulting posteriors.

Quantitative measures of posterior difference can be obtained by calculating the discrepancy between distributions as measures of informativeness. Popular discrepancy measures include the Kullback–Leibler (KL) divergence, mean square error, Wasserstein distance and Hellinger distance. The KL divergence has been used to calculate the prior effective sample size (ESS) [18] as a measure of prior impact. The ESS is the number of observations with the same amount of information as the prior. This measure can suffer from being over-estimated for mixture and multimodal distributions. The mean square error has been used to determine the effective current sample size [19], which is the number of observations that have to be added or subtracted from the prior to obtain the same inference as a baseline prior. The effective current sample size might change if the mode or median is used instead of the mean. Jones et al. [20] used the Wasserstein distance to calculate the mean observed effective sample size (MOPESS).

Discrepancy measures have also been used for prior assessment outside of determining the effective sample. Tang et al. [21] used the KL divergence to quantify prior informativeness in hydrology. Ghaderinezhad et al. [22] introduced the Wasserstein Impact Measure (WIM) which captures the Wasserstein distance between two posterior distributions resulting from two

different priors. They calculated the WIM for univariate and bivariate parameters. Weiss [23] used the chi-square divergence as an interpretable measure of prior sensitivity. Roos et al. [24] applied the Hellinger distance to quantify the prior impact for hierarchical models. Gustafson [25] employed the L^2 norm to evaluate the sensitivity of prior perturbations to posterior expectations.

The above divergences and distances [26] have trade offs and are generally chosen based on a combination of convenience [24], their suitability for the problem at hand and the availability of a robust algorithm and software implementation for their calculation. The KL divergence and chi-square divergence are not symmetric, and are consequently not metrics/distances. The KL divergence is undefined if the intersection of the support of the distributions is the empty set. The Hellinger distance is bounded between zero and one and is a probability metric. The Wasserstein distance is symmetric, computable between discrete and continuous probability distributions, takes into account the geometry of the parameter space [27] and is a properly defined metric. For these reasons, we focus now on the Wasserstein metric.

1.1. Contributions

We make two main contributions in this paper. Firstly, we extend the WIM proposed in [22] to Bayesian models involving ODEs. This extension is enabled by recent advances in computational optimal transport [28, 29, 30] allowing for the efficient estimation of the Wasserstein distance in moderate dimensions.

Secondly, we propose a new prior impact measure, which we call the prior scaled Wasserstein Impact Measure (sWIM), that improves on the WIM by

endowing it with a relative sense of scale. To fix ideas, we describe the sWIM and WIM now in words. Given a baseline prior and a prior of interest, and the two induced posteriors, the sWIM scales the Wasserstein distance between the posteriors (the definition of the original WIM) by the distance between the priors. This scaling is inspired by the one developed in [31, 24] which gives a relative change of scale to the Hellinger distance when used as a prior impact measure for priors with perturbed parameters. We show that this sense of scale helps overcome some of the difficulties intepreting the WIM, particularly whether a WIM for a given problem is 'big' (impactful) or 'small' (not impactful). We also introduce the marginal sWIM as the sWIM for each parameter with a different prior from the baseline instead of for all parameters.

1.2. Outline

An outline of this paper is as follows. In Section 2 we give an overview the core components of the methodology for calculating the sWIM for ODE systems. Then in Section 3 demonstrate and discuss the methodology on a Lotka-Volterra predator-prey ODE model and a Susceptible-Exposed-Infected-Removed (SEIR) epidemiological model, before concluding in Section 4.

2. Methodology

In this section, we begin by introducing the basic notion of a Bayesian inference problem involving ODEs. We then discuss the Wasserstein distance from an optimal transport perspective [32] and then introduce the Sinkhorn algorithm that we use to calculate it. We then discuss the WIM and introduce the new sWIM prior impact measure.

2.1. Bayesian inference for ordinary differential equations

To set ideas we briefly review Bayesian inference in the context of a model involving the solution of an ODE.

Bayesian inference is based on Bayes' theorem which states that the posterior distribution is proportional to the product of the data likelihood and the prior distribution. In the subjective Bayesian framework, the prior reflects the practitioner's beliefs, expert knowledge or any other historical information one may have about the parameters of interest. The likelihood indicates how likely it is for a parameter value to have generated the data. For a given data set $y \in \mathbb{R}^n$ and parameter vector $\theta \in \mathbb{R}^p$, Bayes' theorem can be stated as

$$p(\theta|y) = \frac{f(y|\theta)p(\theta)}{p(y)}, \tag{1}$$

where $p(\theta|y)$ is the posterior distribution, $p(\theta)$ the prior distribution, $f(y|\theta)$ the likelihood function and $p(y)$ a normalising constant.

An ODE describes how the state $z(t)$ changes over time $t \in (0, T]$ and can be written in standard form as

$$\frac{\mathrm{d}z}{\mathrm{d}t} = F(t, z, \theta), \tag{2a}$$

$$\hat{z} = z(t = 0), \tag{2b}$$

where $\hat{z} \in \mathbb{R}$ is the initial condition and $F : (0, T] \times \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}$ a known function particular to the modelled system. As there is usually no closed-form solution of Eq. (2) we resort to numerical methods to find a solution, see e.g. [33] for details.

To link Eq. (2) with Eq. (1), without loss of generality, we assume the

following data generating model for y

$$y_i \mid \theta \sim N([G_z(\theta)]_i, \sigma^2), \quad i = 1, \ldots, n, \tag{3}$$

where $G_z(\theta) : \mathbb{R}^p \to \mathbb{R}^n$ requires the solution of Eq. (2) for z at a fixed θ and the subsequent evaluation of z at a set of n observation points in time $(t_1, t_2, \ldots, t_n)$ with each $t_i \in (0, T]$. With the additional specification of priors θ, the Bayesian posterior in Eq. (1) is fully defined and can be explored using standard techniques e.g. MCMC, see e.g. [34, 35].

2.2. Wasserstein distance

The definition of the p-Wasserstein distance (W_p) between two probability measures μ, ν defined on the spaces $\mathcal{X}$ and $\mathcal{Y}$ is

$$W_p(\mu, \nu) = \inf_{\pi \in U(\mu,\nu)} \left(\int_{\mathcal{X}\times\mathcal{Y}} ||x - y||^p d\pi(x, y) \right)^{1/p}, \quad p \geq 1, \tag{4}$$

where $U(\mu, \nu)$ is the set of joint probability measures on $\mathcal{X} \times \mathcal{Y}$ [32]. Calculating the Wasserstein distance becomes non-trivial in moderate to high dimensions due to the ill-posed nature of the squared Euclidean distance [36]. The Wasserstein distance is the minimum amount of work required to reconfigure the mass of one distribution into another [27].

2.3. Wasserstein impact measure

Let P_0 and P_1 be posteriors induced from a baseline prior p_0 and a prior of interest p_1, respectively, with all other factors defining the Bayesian problem

(likelihood, data etc.) kept the same. The WIM from [22] is defined as

$$\text{WIM}\left(p_0, p_1\right) = W_2\left(P_0, P_1\right). \tag{5}$$

In order to overcome the difficulty in interpreting the WIM, we take inspiration from [24] that scaled the Hellinger distance between posteriors by the distance between priors with applications to hierarchical Bayesian models. We thus propose to divide the Wasserstein distance (the original WIM) between the two posteriors by the Wasserstein distance between the two priors, resulting in

$$\text{sWIM}\left(p_0, p_1\right) = \frac{\text{WIM}\left(p_0, p_1\right)}{W_2\left(p_0, p_1\right)}. \tag{6}$$

The interpretation of our new measure of prior impact is similar to [24]: when sWIM < 1, the distance between posteriors is smaller compared to the distance between priors, while when sWIM > 1, the distance between posteriors is greater than the distance between priors. In addition, when sWIM $\simeq 1$, the differences in prior and posterior are similar and hence not much affected by the data (via the likelihood). The sWIM can also be interpreted as the relative change in the posteriors due to a change in priors. This interpretation is inspired by Roos & Held [31] for their chi-squared divergence for sensitivity analysis

$$\chi^2(P_0) = \frac{\mathbb{E}[P_0] - \mathbb{E}[P_1]}{\mathbb{E}[P_0]} \tag{7}$$

The posterior expectation of the baseline prior is $\mathbb{E}[P_0]$ while the posterior expectation of the prior in question is $\mathbb{E}[P_1]$. The $\chi^2(P_0)$ is interpreted as

the relative change in the posterior expectation due to perturbations in the baseline prior.

In summary, our approach to prior impact assessment involves:

1. Choose a baseline prior denoted as p_0 and a prior of interest denoted as p_1.
2. Obtain m samples from the baseline prior and m samples from the prior of interest.
3. Perform Bayesian inference on the two models to obtain m baseline posterior samples from P_0 and m posterior samples from the prior of interest P_1.
4. Compute the Wasserstein distances between the baseline prior samples and those of interest, as well as between the baseline posterior samples and the samples of interest.
5. Finally calculate the sWIM, marginal sWIM using Eq. (6) and interpret the resulting quantities.

2.4. Discrete optimal transport and Sinkhorn algorithm

For the practical calculation of the sWIM, we make use of techniques from discrete optimal transport. For a full introduction to we refer the reader to [30, 37], and we use similar notations to these two papers.

Consider two probability measures μ and ν approximated by m weighted discrete samples $X = \{x_1, x_2 \ldots, x_m\}$ and $Y = \{y_1, y_2 \ldots, y_m\}$, respectively

$$\mu = \sum_{i=1}^{m} a_i \delta(x_i), \qquad \nu = \sum_{j=1}^{m} b_j \delta(y_j), \tag{8}$$

where δ is the usual Dirac delta function, a is a vector of weights with elements $a_i > 0$ and b is a vector of weights with elements $b_j > 0$.

Let C_{ij} be a cost matrix, which is the pairwise metric between the elements of the samples X and Y. When the metric is chosen as the squared Euclidian distance, as in our case, the cost matrix is

$$C_{ij} = \|x_i - y_j\|^2, \quad i = 1, \ldots, m, \quad j = 1, \ldots, m. \tag{9}$$

For two matrices of the same size, A and B, the Frobenius inner product is

$$\langle A, B \rangle = \mathrm{Tr}(A^T B).$$

Then, the squared 2-Wasserstein distance between two discrete probability measures μ and ν is

$$W_2^2(\mu, \nu) = \min_{T \in \mathrm{U}(a,b)} \langle C, T \rangle \tag{10}$$

Eq. Eq. (10) is the primal formulation of the Wasserstein distance, where T is the joint probability and

$$\mathrm{U}(a, b) \stackrel{\mathrm{def}}{=} \{T \in \mathbb{R}_+^{m \times m} : T^T \mathbb{1}_m = a \text{ and } T\mathbb{1}_m = b\}.$$

contains all possible joint probabilities.

Cuturi [28] introduced entropic regularized optimal transport. This approach regularizes the optimal Wasserstein distance with an entropy term. The entropy $H(T)$ is defined as

$$H(T) = -\sum_{ij} T_{ij}(\log T_{ij} - 1).$$

The entropy regularized 2-Wasserstein distance is defined as

$$W^2_{2,\varepsilon}(\mu, \nu) = \min_{T \in \mathrm{U}(a,b)} \langle C, T \rangle - \varepsilon H(T), \tag{11}$$

where ε controls the strength of the penalty, when ε is zero, we recover the original problem. The entropy-regularised optimal transport problem is convex and has a unique solution [38]. Entropy based penalty is inspired by the Schrödinger problem (see [39, 40]). The entropy regularized primal and dual formulations are linear programming problems and can be solved with a simple iterative scaling algorithm known as the Sinkhorn algorithm [30, 37].

3. Examples

In this section, we illustrate the proposed prior impact assessment methodology on the Lotka-Volterra predator prey and the SEIR models calibrated against real-world data. By using different sets of priors, we aim to gain insights into how priors impact inference. Specifically, we seek to answer questions such as whether a prior is more informative than some baseline and, if so, whether its impact on posterior inference is small or large. To do this, we calculate the WIM and sWIM for different sets of priors as well as comparing the parameter estimates and graphical posterior predictive checks. The computational models are constructed using TensorFlow Probability [41] on JAX [42], the No-U-Turn sampler (NUTS) algorithm [35] is used to obtain posterior samples and Optimal Transport Tools (OTT) [29] is used to

calculate Wasserstein distances with the Sinkhorn algorithm. Complete code to reproduce the results is available in the supplementary material.

3.1. Lotka–Volterra model

The Lotka–Volterra model describes predator-prey dynamics in an environment

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \alpha u - \beta uv, \quad \hat{u} = u(t=0), \tag{12a}$$

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \delta uv - \gamma v, \quad \hat{v} = v(t=0), \tag{12b}$$

where $u(t) > 0$ represents the number of preys at time t, $v(t) > 0$ the number of predators at time t, $\alpha > 0$ is the prey birth rate, $\beta > 0$ links prey mortality to the number of preys and predators, $\delta > 0$ links the increase in predators to the number of predators and preys, and $\gamma > 0$ stands for the predator death rate. The initial prey state is $\hat{u}$ and the initial predator state is $\hat{v}$.

Let the parameter vector $\theta = (\alpha, \beta, \delta, \gamma, \hat{u}, \hat{v})^T \in \mathbb{R}^6$. The data generating model can be constructed as

$$\begin{aligned} z_i^1 \mid \theta &\sim \mathrm{LN}\left([G_u(\theta)]_i, \sigma_u^2\right), \quad \forall i = 1, \ldots, n, \\ z_i^2 \mid \theta &\sim \mathrm{LN}\left([G_v(\theta)]_i, \sigma_v^2\right), \quad \forall i = 1, \ldots, n, \\ y &= \left(z^1, z^2\right)^T \in \mathbb{R}^{2n}. \end{aligned}$$

where the operators G_u and G_v involve the solution of Eq. (12a) and Eq. (12b), respectively, at n time points distributed between 1845 and 1935, and LN is the log-normal distribution.

We consider hare-lynx data based on historical pelt records of the Hudson

Bay Company [43] which is available in numpyro [44]. The pelt records are used as a proxy for the populations of hare and lynx in the environment. Part of this data from 1900 to 1920 has been analyzed in [45] using a Bayesian approach, and the entire data has been used for model calibration using a frequentist approach in [46].

Since all involved parameters must be positive, the prior distributions on the parameters are defined as

$$\begin{aligned} \alpha, \gamma &\overset{\text{iid}}{\sim} \text{TN}(1.0, 0.5), \\ \beta, \delta &\overset{\text{iid}}{\sim} \text{TN}(0.05, 0.05), \\ \hat{u}, \hat{v} &\overset{\text{iid}}{\sim} \text{LN}(\ln(10), 1.0), \\ \sigma_u^2, \sigma_v^2 &\overset{\text{iid}}{\sim} \text{LN}(-1, 1), \end{aligned}$$

where TN is the truncated normal distribution. The priors are the same as those used in [45]. In order to obtain a variety of priors for our prior impact assessment, we perturb the prior distribution for the initial states $(\hat{u}, \hat{v})$ as well as for the error variances (σ_u^2, σ_v^2) to create four distinct sets of priors (p_0, p_1, p_2, p_3) as shown in Table 1.

Table 1: Priors used for the Lotka–Volterra model. Only the priors on the initial states $(\hat{u}, \hat{v})$ and on the error variances (σ_u^2, σ_v^2) are perturbed. The truncated normal (TN) and the log-normal (LN) distributions are used.

p_0 [45]	p_1	p_2	p_3
$\alpha, \gamma \overset{\text{iid}}{\sim} \text{TN}(1.0, 0.5)$	$\alpha, \gamma \overset{\text{iid}}{\sim} \text{TN}(1.0, 0.5)$	$\alpha, \gamma \overset{\text{iid}}{\sim} \text{TN}(1.0, 0.5)$	$\alpha, \gamma \overset{\text{iid}}{\sim} \text{TN}(1.0, 0.5)$
$\beta, \delta \overset{\text{iid}}{\sim} \text{TN}(0.05, 0.05)$	$\beta, \delta \overset{\text{iid}}{\sim} \text{TN}(0.05, 0.05)$	$\beta, \delta \overset{\text{iid}}{\sim} \text{TN}(0.05, 0.05)$	$\beta, \delta \overset{\text{iid}}{\sim} \text{TN}(0.05, 0.05)$
$\hat{u}, \hat{v} \overset{\text{iid}}{\sim} \text{LN}(\ln(10), 1.0)$	$\hat{u}, \hat{v} \overset{\text{iid}}{\sim} \text{LN}(\ln(2), 1.0)$	$\hat{u} \overset{\text{iid}}{\sim} \text{LN}(\ln(15), 1.0)$	$\hat{u} \overset{\text{iid}}{\sim} \text{LN}(\ln(15), 1.0)$
$\sigma_u^2, \sigma_v^2 \overset{\text{iid}}{\sim} \text{LN}(-1, 1)$	$\sigma_u^2, \sigma_v^2 \overset{\text{iid}}{\sim} \text{LN}(-1, 1)$	$\hat{v} \overset{\text{iid}}{\sim} \text{LN}(\ln(6), 1.0)$	$\hat{v} \overset{\text{iid}}{\sim} \text{LN}(\ln(6), 1.0)$
		$\sigma_u^2, \sigma_v^2 \overset{\text{iid}}{\sim} \text{LN}(2.0, 0.2)$	$\sigma_u^2, \sigma_v^2 \overset{\text{iid}}{\sim} \text{LN}(1.0, 0.1)$

We fitted four different models corresponding to the four sets of priors. We obtained the posterior samples using the preconditioned NUTS. A thousand samples were discarded as warm-up and three thousand used for inference. The marginal posterior distributions and pairwise correlations are displayed in Fig. 1. We can see high pairwise correlations between some parameters, e.g. in the couples $(\alpha, \hat{u}), (\alpha, \delta)$, and (α, γ). The choice of gradient-based algorithms such as NUTS is crucial to sample the entire posterior distribution, and we show the posterior mean estimates in Table 2. The prey birth rate α is more than one unit higher for the baseline than the other prior models. Also, the posterior mean estimates of the initial number of preys $\hat{\mu}$ and initial predators $\hat{\nu}$ are more than two units greater for each model compared to the baseline priors.

Now we calculate the WIM and the sWIM between various pairs of distributions. The results of the WIM are shown in Table 3 along with other Wasserstein-2 distances, while those related to the sWIM appear in Table 4. We first note that based on both WIM and sWIM the prior p_1 is closest to the baseline. This is in line with the posterior predictive check in Fig. 2(a). The sWIM for p_2 and p_3 are greater than one. This means that p_2 and p_3 have different impact on the posterior, in the sense that in each case the posteriors are further from the baseline posterior compared to the distances between priors. The results are consistent with the graphical posterior predictive check, which shows that the predictions are noticeably different from the baseline Fig. 2(a). In order to get more detailed information, we compute the marginal sWIM for the initial conditions and error variances and the results are given in Table 5. The marginal sWIM is greater than one

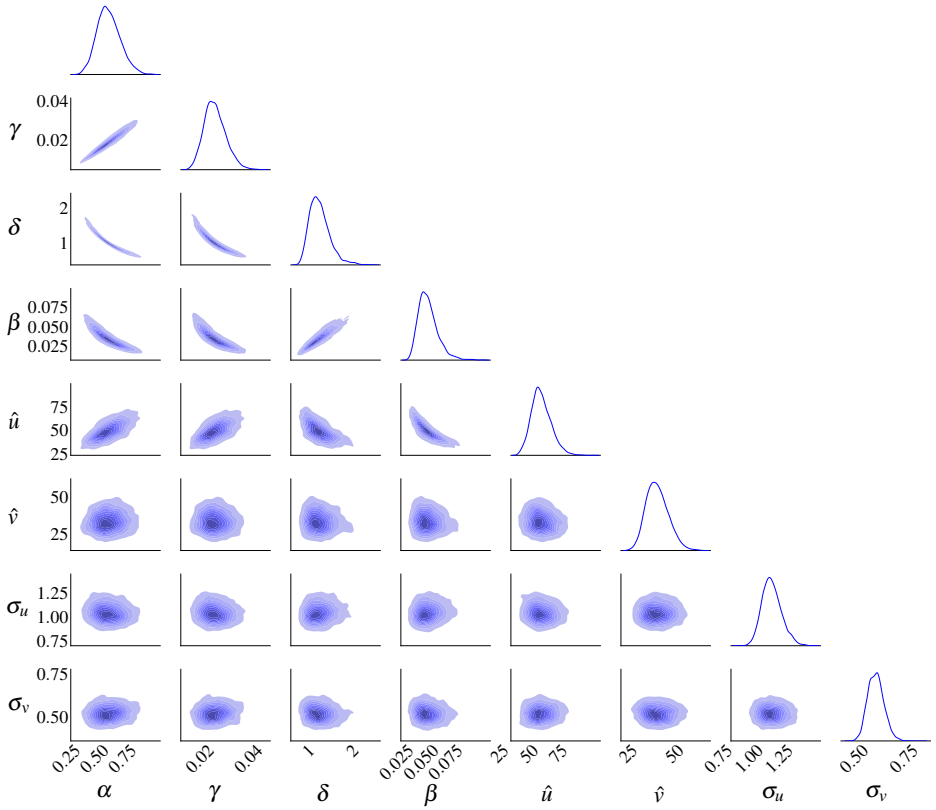

Figure 1: Plot of posterior distributions associated to p_0 in the Lotka-Volterra model, with posterior marginal distributions on the diagonal and bivariate distributions for outside the diagonal. One can observe high correlations between some pairs of parameters.

for the initial condition ($\hat{\mu}$) with priors p_1, p_2 and p_3. Thus, the priors have different impacts on the posterior number of hares compared to the baseline prior. This information is not immediately apparent in the posterior predictive plots Fig. 2(a), as p_2 and the baseline prior (p_0) appear indistinguishable. However, it becomes evident when looking at the posterior estimates(Table 2) as the initial number of hares $\hat{\mu}$ for the baseline prior is further away from the other priors. The marginal sWIM is below one for the initial number of lynxs (ν). This is consistent with the posterior estimates Table 2 where the

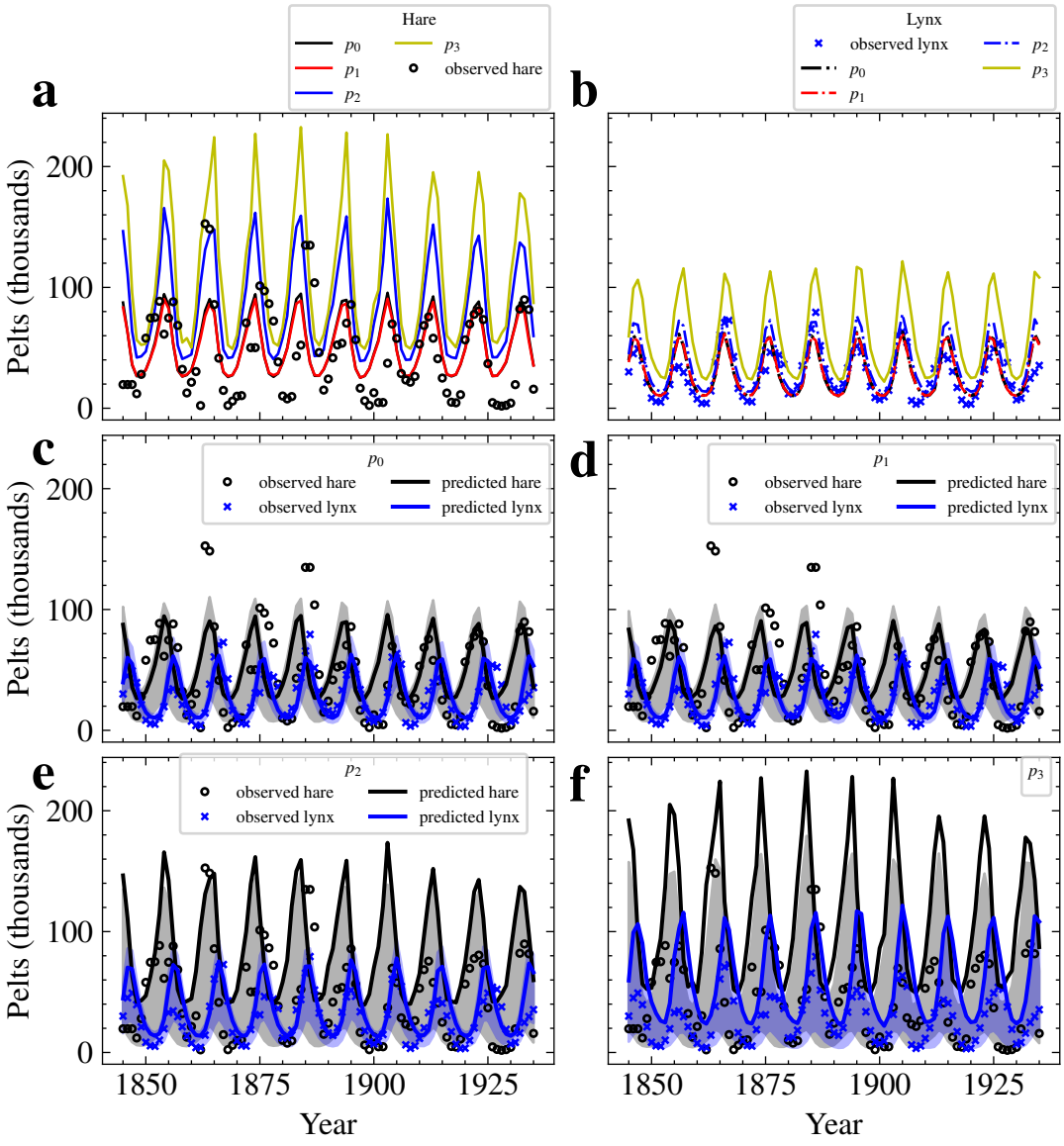

Figure 2: Graphical posterior predictive check for (a) Hare and (b) Lynx. The prior p_3 has a noticeable visual impact compared to p_0 and p_1. (c-f) Posterior predictive check for each prior with 25% and 75% quantiles.

difference between the baseline prior and the others is very small. Again, this is not evident from the posterior predictive plots Fig. 2(b), where p_3 is visually different from the other priors.

Table 2: Posterior mean estimates of the different models for the Lotka-Volterra model.

parameter	p_0	p_1	p_2	p_3
	Baseline			
α	0.825	0.462	0.557	0.633
γ	0.037	0.019	0.023	0.028
β	0.404	1.038	0.894	0.810
δ	0.013	0.036	0.030	0.028
$\hat{u}$	16.427	48.662	53.451	53.378
$\hat{v}$	27.818	33.392	30.843	25.689
σ_u	1.054	1.042	1.413	1.597
σ_v	0.775	0.519	0.866	1.331

Table 3: Wasserstein-2 distances between prior p_i and posterior P_i distributions in the Lotka-Wolterra model, $i = 0, 1, 2, 3$. The values in bold are the WIM between the baseline posterior and the three other posteriors.

Wasserstein 2-distance						
	prior			posterior		
Baseline	p_1	p_2	p_3	P_1	P_2	P_3
p_0	36.266	26.975	25.124	28.864	40.504	39.40
P_0	49.647	51.257	49.320	**34.137**	**39.973**	**40.674**

Table 4: Prior scaled WIM between the baseline posterior and the three other posteriors in the Lotka-Volterra model.

	posterior		
Baseline posterior	P_1	P_2	P_3
P_0	0.941	1.482	1.619

Table 5: Marginal prior scaled WIM between the baseline posterior and the three other posteriors in the Lotka-Volterra model, only for parameters whose priors change across models.

	posterior		
parameter	P_1	P_2	P_3
$\hat{\mu}$	2.457	4.515	4.507
$\hat{\nu}$	0.415	0.646	0.812
σ_u	-	0.052	0.250
σ_v	-	0.013	0.275

3.2. SEIR model

The Bayesian approach is widely used for parameter estimation in epidemiology. We illustrate our prior assessment technique on the Susceptible-Exposed-Infectious-Recovered (SEIR) model, a commonly used epidemiological dynamical model. The COVID-19 pandemic in Luxembourg has been analysed in a Bayesian setting in [3] considering various control measures such as lockdown. Here, we calibrate the SEIR model for the first wave of the COVID-19 pandemic in Luxembourg using publicly available data found in [47]. The first wave lasted from February to mid June 2020, a time frame that also similar studies determined as first wave [48]. Compared to the original study [3], we additionally perform posterior predictive checks and prior impact assessment using the proposed WIM and sWIM.

The SEIR model is defined by the following system of ODEs

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\eta S\frac{I}{N}, \quad S(t=0) = \hat{S} = \hat{N} - \hat{I} - \hat{E} - \hat{R}, \tag{13a}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \eta S\frac{I}{N} - \sigma E, \quad E(t=0) = \hat{E}, \tag{13b}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \sigma E - \rho I, \quad I(t=0) = \hat{I}, \tag{13c}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \rho I, \quad R(t=0) = \hat{R} = 0, \tag{13d}$$

$$I_R = \lambda\left(\frac{\mathrm{d}E}{\mathrm{d}t} + \frac{\mathrm{d}I}{\mathrm{d}t}\right), \tag{13e}$$

$$N = S + E + I + R, \tag{13f}$$

where S is the number of susceptible individuals, that is, people not immune to COVID-19, E the number of exposed individuals, meaning people who have been infected but are themselves not yet infectious, I the number of

infected and infectious individuals, R the number of recovered or deceased individuals, and I_R the rate of infection. N is the total number of individuals at any time which we assume to be constant.

The model has seven parameters, namely the transmission rate $\eta > 0$, the reciprocal of the incubation rate $\sigma > 0$, the recovery rate $\rho > 0$, the initial value for the number of infected individuals $\hat{I}$, the initial value for the number of exposed individuals $\hat{E}$, and a multiplicative correction for the rate of infection denoted by $\lambda \in (0, 1]$. This factor corrects for under-reporting and is the approach employed in [48].

We are interested in calibrating the model against the reported daily rate of infection I_R. To that end, we employ a count distribution, more precisely the negative binomial NB with dispersion parameter $\phi > 0$:

$$y_i \mid \theta \sim \mathrm{NB}([G_{I_R}(\theta)]_i, \phi), \quad i = 1, \ldots, n,$$

where $G_{I_R} : \mathbb{R}^p \to \mathbb{R}^n$ requires the solution of Eq. (13e) and its evaluation at n daily time points in the study period. We approximate I_R in Eq. (13e) with a first-order backwards finite differencing of Eqs. (13a) and (13b). Note that because $I_R \in \mathbb{R}$ we employ a non-standard negative binomial parametrization in TensorFlow Probability [41] which naturally extends to the real numbers.

For the prior impact assessment, we use five sets of priors where the first set of priors is the baseline prior to which other priors are compared. The baseline prior p_0 and is similar to the non-informative priors used in [48, 49]

except for the moments of the distributions:

$$\begin{aligned}
\eta &\sim \text{TN}(2, 1), \\
\rho &\sim \text{TN}(0.4, 0.5), \\
\sigma &\sim \text{TN}(0.4, 0.5), \\
\hat{I} &\sim \text{TN}(0, 1), \\
\hat{E} &\sim \text{TN}(0, 1), \\
\lambda &\sim \text{Beta}(1, 2), \\
\phi^{-1} &\sim \text{Exponential}(5).
\end{aligned}$$

These priors are truncated normal distributions for most of the parameters which need to be positive, or zero in the case of $\hat{I}$ and $\hat{E}$.

The remaining five sets of priors shown in Table 6 where only the dispersion parameter ϕ varies across the priors. This is because overdispersion is usually the issue when modelling count data. Hence, it is important to see how overdispersion impacts inference. For the fifth set of priors not shown in Table 6, the dispersion parameter $\phi^{-1} \sim \text{Exponential}(150)$ is such that we can make a statement when the prior has a high impact compared to the baseline prior. It is noteworthy that we also choose a Gamma distribution as prior p_3, of which the Exponential is a special case.

The marginal posterior distributions for the baseline prior p_0 are shown in Fig. 3. The posterior estimates for $\hat{I}$, $\hat{E}$ and ϕ have a higher standard deviation as shown by the density plots on the diagonal. Also, there is a noticeable correlation in the parameter pairs (λ, σ) and $(\lambda, \hat{I})$. The results of the posterior parameter estimates can be found in Table 7. The posterior

mean estimates of ϕ for the priors p_1 and p_4 lie below one. For the Gamma prior p_3, the posterior estimate of the dispersion parameter is closer to that of the baseline prior p_0.

Fig. 4 shows posterior predictive checks for each prior with 25% and 75% confidence bands. The Gamma prior p_3 here resembles the most to the baseline prior p_0 but with a lower overdispersion parameter as shown in Table 7. This indicates that the gamma prior might be an alternative to the exponential prior on the overdispersion parameter. When the gamma prior is applied, the prediction intervals are narrower and enclose more observations than the baseline prior. In the case of the exponential prior, the mean value increases with the overdispersion parameter. The task is to address overdispersion without affecting the mean. Additionally, it is worth noting that the gamma distribution becomes the exponential distribution when the shape parameter equals 1. The mean predicted daily number of cases is quite close to the observed number for all priors except for p_4. For this prior, the 25% and 75% quantiles are also the widest compared to the others (see Fig. 3c and f).

Table 6: Priors used for the SEIR model. Only the prior on the dispersion parameter is different since overdispersion is usually the main modelling concern for count data. We choose the baseline prior p_0 like in other studies [49, 48]. A fifth prior not shown was also included.

p_0 (Baseline prior)	p_1	p_2	p_3
$\eta \sim \text{TN}(2,1)$	$\eta \sim \text{TN}(2,1)$	$\eta \sim \text{TN}(2,1)$	$\eta \sim \text{TN}(2,1)$
$\rho \sim \text{TN}(0.6,0.5)$	$\rho \sim \text{TN}(0.6,0.5)$	$\rho \sim \text{TN}(0.6,0.5)$	$\rho \sim \text{TN}(0.6,0.5)$
$\sigma \sim \text{TN}(0.4,0.5)$	$\sigma \sim \text{TN}(0.4,0.5)$	$\sigma \sim \text{TN}(0.4,0.5)$	$\sigma \sim \text{TN}(0.4,0.5)$
$\hat{I}, \hat{E} \overset{\text{iid}}{\sim} \text{TN}(0,1)$	$\hat{I}, \hat{E} \overset{\text{iid}}{\sim} \text{TN}(0,1)$	$\hat{I}, \hat{E} \overset{\text{iid}}{\sim} \text{TN}(0,1)$	$\hat{I}, \hat{E} \overset{\text{iid}}{\sim} \text{TN}(0,1)$
$\lambda \sim \text{Beta}(1,2)$	$\lambda \sim \text{Beta}(1,2)$	$\lambda \sim \text{Beta}(1,2)$	$\lambda \sim \text{Beta}(1,2)$
$\phi^{-1} \sim \text{Exponential}(5)$	$\phi^{-1} \sim \text{Exponential}(42)$	$\phi^{-1} \sim \text{Exponential}(1)$	$\phi^{-1} \sim \text{Gamma}(16,16)$

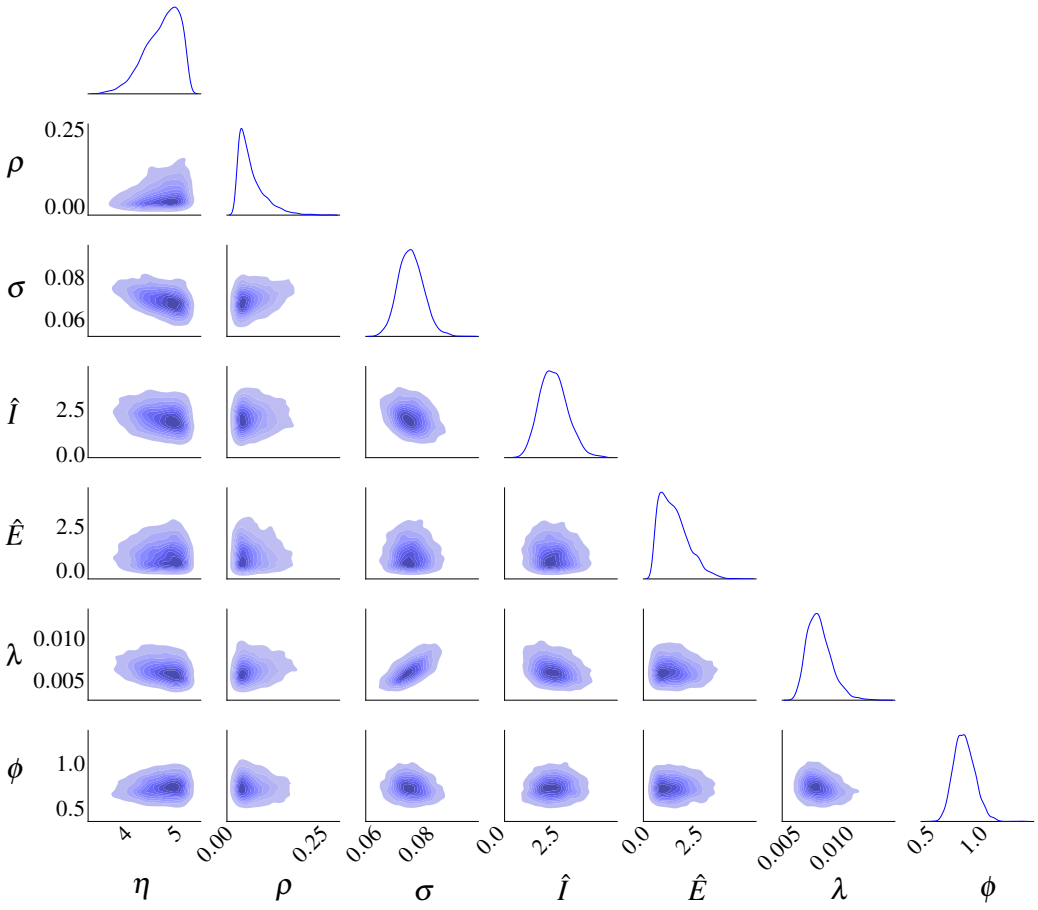

Figure 3: Plot of posterior distributions associated to p_0 in the SEIR model, with posterior marginal distributions on the diagonal and bivariate distributions for outside the diagonal. There is a high correlation between the parameters λ and σ.

Table 7: Posterior mean estimates of the different models for the SEIR model.

parameter	p_0	p_1	p_2	p_3	p_4
η	4.689	4.497	4.729	4.678	4.161
ρ	0.024	0.036	0.022	0.025	0.059
σ	0.066	0.069	0.066	0.066	0.076
$\hat{I}$	2.255	2.000	2.294	2.235	1.701
$\hat{E}$	0.891	0.908	0.901	0.886	0.881
λ	0.005	0.006	0.005	0.005	0.009
ϕ	1.856	0.734	2.426	1.674	0.315

Let us now discuss the findings based on WIM and sWIM. The results for the WIM are provided in Table 8. The WIM is highest for the prior p_4, which is totally in line with the posterior predictive check where the plot is clearly distinguishable from the other plots and hence in particular from that corresponding to p_0. The same holds for p_1, though with a smaller WIM and this is consistent with a lower peak in Fig. 4a. The sWIM is shown in Table 9. We can see that the sWIM is below 1 for p_2 and p_3 while above 1 for p_1 and p_4. This example illustrates that the gamma prior (p_3) has similar impact on the posterior as the baseline prior (p_0). For p_1 with sWIM of 1.276, the parameter estimates are still close to those of the baseline posterior, although the posterior predictive check shows higher predictions than observed. In the case of p_4 with sWIM of 1.762, the parameter estimates are further from the baseline posterior estimates, and the posterior predictive check shows higher predictions than observed.

The marginal sWIM was also computed for ϕ, and the results are in

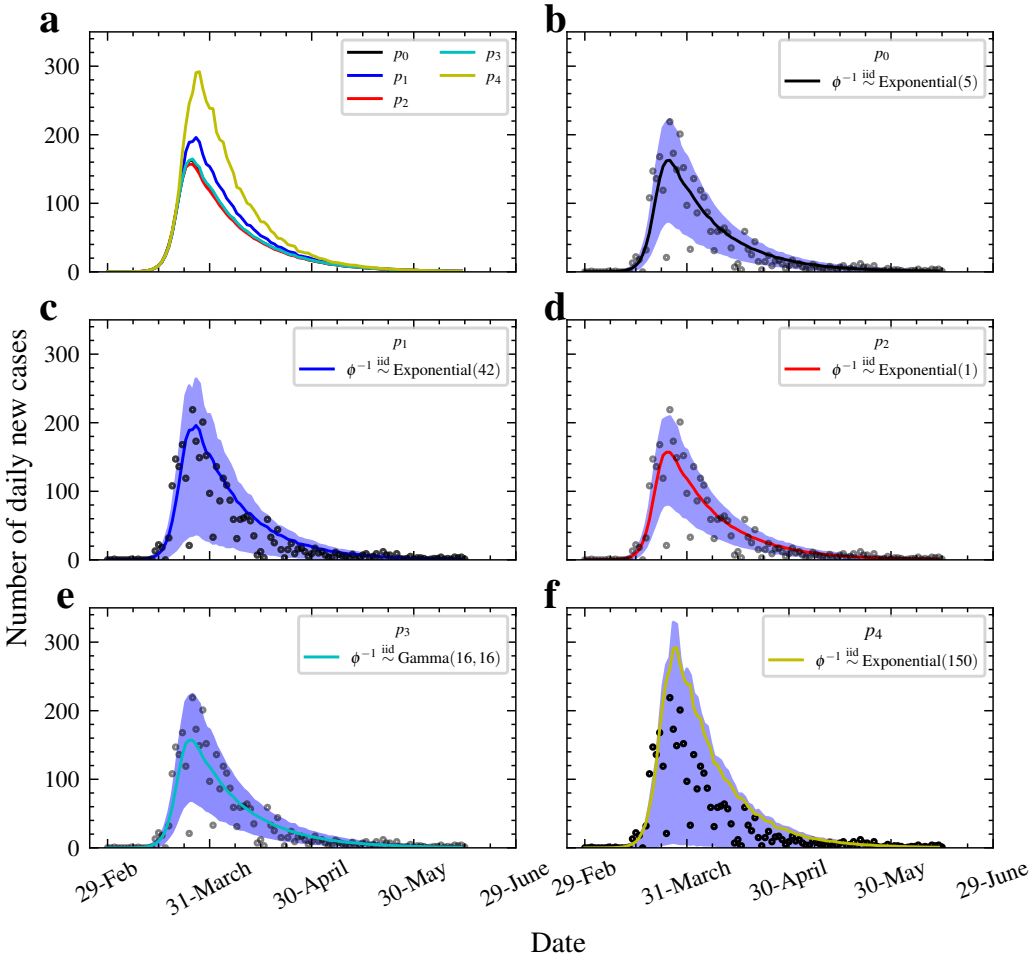

Figure 4: (a) Graphical posterior predictive check for all priors in the SEIR model, and (b-f) posterior predictive check for each prior with 25% and 75% quantiles. The Gamma(16,16) prior seems to be a better option since the posterior has less variability compared to the other. Moreover, most of the observed counts are in the 25% to 75% prediction bands unlike for other posteriors considered where the highest and lowest counts are outside, or the bands are wider like for Exponential(42). The prior p_4 has the largest predicted values and the predictions are further away from the observed values compared to the other priors.

Table 10. The marginal sWIM values are less than 1 for p_2 and p_3, which ties to parameter estimates and predictions closer to baseline.

Table 8: Wasserstein-2 distances between prior p_i and posterior P_i distributions in the SEIR model, $i = 0, 1, 2, 3$. The values in bold are the WIM between the baseline posterior and the four other posteriors.

Wasserstein-2 distance								
	prior				posterior			
Baseline	p_1	p_2	p_3	p_4	P_1	P_2	P_3	P_4
p_0	1.1055	1.650	1.401	1.109	3.299	4.240	3.811	1.109
P_0	3.991	3.747	3.644	4.000	**1.410**	**0.925**	**0.688**	**1.954**

Table 9: Prior scaled WIM between the baseline posterior and the four other posteriors in the SEIR model.

sWIM				
	posterior			
Baseline posterior	P_1	P_2	P_3	P_4
P_0	1.276	0.561	0.491	1.762

Table 10: Marginal prior scaled WIM between the baseline posterior and the four other posteriors in the SEIR model, only for parameters whose priors change across models.

sWIM				
	posterior			
parameter	P_1	P_2	P_3	P_4
ϕ	6.245	0.855	0.324	7.866

4. Conclusions

This study employs computational optimal transport to quantify the difference in posterior inference between different priors for dynamic systems modelled by ODEs. Using the Sinkhorn algorithm from computational optimal transport, we can rapidly compute the Wasserstein distance for multiparameter systems. Building on this, we have extended the WIM as prior impact assessment tool to ordinary differential equations. We have also introduced the sWIM, an interpretable impact measure. Like the WIM, it can be quickly calculated. When sWIM < 1, the prior in question has no greater impact than the baseline prior. When sWIM > 1, the prior in question has a higher impact than a baseline prior. It is also insightful to compute the marginal sWIM, meaning the sWIM for each parameter instead of the joint parameters. Our approach has been exemplified with the Lotka–Volterra predator-prey model and the SEIR for Covid-19. For both examples, we used real-world data. The results show that the difference in posterior mean estimates is close to zero when sWIM < 1. In addition, graphical posterior predictive checks show that predictions are closer to the baseline when sWIM < 1 and further when sWIM > 1. In future research, our goal is to decide by means of an extensive simulation study at what values of the sWIM one should label a prior as high or low impact relative to a baseline prior.

Glossary

KL Kullback–Leibler. 4, 5

MCMC Markov Chain Monte Carlo. 3, 8

MOPESS mean observed effective sample size. 4

NUTS No-U-Turn sampler. 12, 15

ODE ordinary differential equation. 4, 6, 7

ODEs ordinary differential equatiosince. 2, 3, 5, 6, 28

OTT Optimal Transport Tools. 12

SEIR Susceptible-Exposed-Infected-Removed. 6, 12, 20, 28

sWIM prior scaled Wasserstein Impact Measure. 1, 5, 6, 9, 10, 12, 15, 16, 25, 26, 27, 28

WIM Wasserstein Impact Measure. 1, 4, 5, 6, 9, 12, 15, 18, 19, 25, 28

Code and data availability

The full code to produce the results in this article is available at [50] and GitHub at BayesianODE-PriorImpactAssessment.

Credit contributor roles

DNM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing - review & editing. JSH: Conceptualization, Formal analysis, Funding acquisition, Resources, Methodology, Project administration, Software, Supervision, Validation, Writing - original draft, Writing - review & editing. CL: Conceptualization, Funding acquisition, Formal analysis, Methodology, Supervision, Writing - review & editing.

Competing interests

The authors declare no competing interests.

Acknowledgements

This work was funded under the Luxembourg National Research Fund under the PRIDE programme (PRIDE17/12252781).

The experiments were carried out using the HPC facilities of the University of Luxembourg [51] – see https://hpc.uni.lu.

References

[1] M. Girolami, Bayesian inference for differential equations, Theoretical Computer Science 408 (1) (2008) 4–16. doi:10.1016/j.tcs.2008.07.005.

[2] G. C. Gibson, N. G. Reich, D. Sheldon, Real-time mechanistic Bayesian forecasts of COVID-19 mortality (2020). doi:10.1101/2020.12.22.20248736.

[3] F. Kemp, D. Proverbio, A. Aalto, L. Mombaerts, A. Fouquier d'Hérouël, A. Husch, C. Ley, J. Gonçalves, A. Skupin, S. Magni, Modelling COVID-19 dynamics and potential for herd immunity by vaccination in Austria, Luxembourg and Sweden, Journal of Theoretical Biology 530 (2021) 110874. doi:10.1016/j.jtbi.2021.110874.

[4] D. Machac, P. Reichert, C. Albert, Emulation of dynamic simulators with application to hydrology, Journal of Computational Physics 313 (2016) 352–366. doi:https://doi.org/10.1016/j.jcp.2016.02.046.

[5] J. Awrejcewicz, Ordinary Differential Equations and Mechanical Systems, Springer International Publishing, Cham, 2014. doi:10.1007/978-3-319-07659-1.

[6] J. O. Berger, J. M. Bernardo, D. Sun, Overall objective priors, Bayesian Analysis 10 (1) (2015). doi:10.1214/14-BA915.

[7] M. Ghosh, Objective priors: An introduction for frequentists, Statistical Science 26 (2) (2011). doi:10.1214/10-STS338.

[8] A. Gelman, D. Simpson, M. Betancourt, The prior can often only be understood in the context of the likelihood, Entropy 19 (10) (2017). doi:10.3390/e19100555.

[9] A. Gelman, Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper), Bayesian Analysis 1 (3) (2006) 515–534. doi:10.1214/06-BA117A.

[10] G. C. Gibson, N. G. Reich, D. Sheldon, Real-time mechanistic Bayesian forecasts of COVID-19 mortality, The Annals of Applied Statistics 17 (3) (2023). doi:10.1214/22-AOAS1671.

[11] C.-C. Lai, C.-Y. Hsu, H.-H. Jen, A. M.-F. Yen, C.-C. Chan, H.-H. Chen, The Bayesian Susceptible-Exposed-Infected-Recovered model for the outbreak of COVID-19 on the Diamond Princess cruise ship, Stochastic Environmental Research and Risk Assessment 35 (7) (2021) 1319–1333. doi:10.1007/s00477-020-01968-w.

[12] B. Calderhead, M. Girolami, Statistical analysis of nonlinear dynamical

systems using differential geometric sampling methods, Interface Focus 1 (6) (2011) 821–835. doi:10.1098/rsfs.2011.0051.

[13] S. Golchi, Informative priors in Bayesian inference and computation, Statistical Analysis and Data Mining: The ASA Data Science Journal 12 (2) (2019) 45–55. doi:10.1002/sam.11371.

[14] A. M. Stefan, D. Katsimpokis, Q. F. Gronau, E.-J. Wagenmakers, Expert agreement in prior elicitation and its effects on Bayesian inference, Psychonomic Bulletin & Review 29 (5) (2022) 1776–1794. doi:10.3758/s13423-022-02074-4.

[15] C. Pedroza, W. Han, V. T. Thanh Truong, C. Green, J. E. Tyson, Performance of informative priors skeptical of large treatment effects in clinical trials: A simulation study, Statistical Methods in Medical Research 27 (1) (2018) 79–96. doi:10.1177/0962280215620828.

[16] H. Schmidli, S. Gsteiger, S. Roychoudhury, A. O'Hagan, D. Spiegelhalter, B. Neuenschwander, Robust meta-analytic-predictive priors in clinical trials with historical control information, Biometrics 70 (4) (2014) 1023–1032. doi:10.1111/biom.12242.

[17] D. Nur, D. Allingham, J. Rousseau, K. L. Mengersen, R. McVinish, Bayesian hidden Markov model for DNA sequence segmentation: A prior sensitivity analysis, Statistical Genetics & Statistical Genomics: Where Biology, Epistemology, Statistics, and Computation Collide 53 (5) (2009) 1873–1882. doi:10.1016/j.csda.2008.07.007.

[18] S. Morita, P. F. Thall, P. Müller, Determining the effective sample size of a parametric prior, Biometrics 64 (2) (2008) 595–602. doi:10.1111/j.1541-0420.2007.00888.x.

[19] M. Wiesenfarth, S. Calderazzo, Quantification of prior impact in terms of effective current sample size, Biometrics 76 (1) (2020) 326–336. doi:10.1111/biom.13124.

[20] D. E. Jones, R. N. Trangucci, Y. Chen, Quantifying observed prior impact, Bayesian Analysis 17 (3) (2022) 737–764. doi:10.1214/21-BA1271.

[21] Y. Tang, L. Marshall, A. Sharma, T. Smith, Tools for investigating the prior distribution in Bayesian hydrology, Journal of Hydrology 538 (2016) 551–562. doi:10.1016/j.jhydrol.2016.04.032.

[22] F. Ghaderinezhad, C. Ley, B. Serrien, The Wasserstein Impact Measure (WIM): A practical tool for quantifying prior impact in Bayesian statistics, Computational Statistics & Data Analysis 174 (2022) 107352. doi:10.1016/j.csda.2021.107352.

[23] R. Weiss, An approach to Bayesian sensitivity analysis, Journal of the Royal Statistical Society: Series B (Methodological) 58 (4) (1996) 739–750. doi:10.1111/j.2517-6161.1996.tb02112.x.

[24] M. Roos, T. G. Martins, L. Held, H. Rue, Sensitivity analysis for Bayesian hierarchical models, Bayesian Analysis 10 (2) (2015). doi:10.1214/14-BA909.

[25] P. Gustafson, Local sensitivity of inferences to prior marginals, Journal of the American Statistical Association 91 (434) (1996) 774–781. doi:10.1080/01621459.1996.10476945.

[26] S. M. Ali, S. D. Silvey, A general class of coefficients of divergence of one distribution from another, Journal of the Royal Statistical Society: Series B (Methodological) 28 (1) (1966) 131–142. doi:10.1111/j.2517-6161.1966.tb00626.x.

[27] V. M. Panaretos, Y. Zemel, Statistical aspects of Wasserstein distances, Annual Review of Statistics and Its Application 6 (2019) 405–431. doi:10.1146/annurev-statistics-030718-104938.

[28] M. Cuturi, Sinkhorn distances: Lightspeed computation of optimal transport, in: C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K. Weinberger (Eds.), Advances in Neural Information Processing Systems, Vol. 26, Curran Associates, Inc., 2013.
URL https://proceedings.neurips.cc/paper_files/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf

[29] M. Cuturi, L. Meng-Papaxanthos, Y. Tian, C. Bunne, G. Davis, O. Teboul, Optimal Transport Tools (OTT): A JAX toolbox for all things Wasserstein, arXiv preprint arXiv:2201.12324 (2022).

[30] M. Cuturi, A. Doucet, Fast computation of Wasserstein barycenters, in: E. P. Xing, T. Jebara (Eds.), Proceedings of the 31st International Conference on Machine Learning, Vol. 32 of Proceedings of Machine

Learning Research, PMLR, Bejing, China, 2014, pp. 685–693.
URL https://proceedings.mlr.press/v32/cuturi14.html

[31] M. Roos, L. Held, Sensitivity analysis in Bayesian generalized linear mixed models for binary data, Bayesian Statistics 6 (2) (2011) 259–278. doi:10.1214/11-BA609.

[32] C. Villani, Optimal Transport: Old and New, no. 338, Springer, 2009. doi:10.1007/978-3-540-71050-9.

[33] P. Kidger, J. Morrill, J. Foster, T. Lyons, Neural controlled differential equations for irregular time series, Advances in Neural Information Processing Systems 33 (2020) 6696–6707.
URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4a5876b450b45371f6cfe5047ac8cd45-Paper.pdf

[34] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, Bayesian Data Analysis, Third Edition, Texts in Statistical Science Series, CRC Press, Taylor and Francis Group, 2014. doi:10.1201/b16018.

[35] M. D. Hoffman, A. Gelman, The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo., Journal of Machine Learning Research 15 (1) (2014) 1593–1623.

[36] M. Cuturi, M. Klein, P. Ablin, Monge, Bregman and Occam: Interpretable Optimal Transport in High-Dimensions with Feature-Sparse Maps, Proceedings of the 40th International Conference on Machine

Learning 202 (2023).
URL https://proceedings.mlr.press/v202/cuturi23a.html

[37] M. Scetbon, M. Cuturi, G. Peyré, Low-rank Sinkhorn factorization, in: M. Meila, T. Zhang (Eds.), Proceedings of the 38th International Conference on Machine Learning, Vol. 139 of Proceedings of Machine Learning Research, PMLR, 2021, pp. 9344–9354.
URL https://proceedings.mlr.press/v139/scetbon21a.html

[38] G. Peyré, M. Cuturi, Computational optimal transport: With applications to data science, Foundations and Trends® in Machine Learning 11 (5-6) (2019) 355–607. doi:10.1561/2200000073.

[39] P. Ghosal, M. Nutz, E. Bernton, Stability of entropic optimal transport and Schrödinger bridges, Journal of Functional Analysis 283 (9) (2022) 109622. doi:10.1016/j.jfa.2022.109622.

[40] G. Peyré, M. Cuturi, et al., Computational optimal transport: With applications to data science, Foundations and Trends in Machine Learning 11 (5-6) (2019) 355–607. doi:10.1561/2200000073.

[41] J. V. Dillon, I. Langmore, D. Tran, E. Brevdo, S. Vasudevan, D. Moore, B. Patton, A. Alemi, M. Hoffman, R. A. Saurous, TensorFlow Distributions, arXiv (2017). doi:10.48550/ARXIV.1711.10604.

[42] R. Frostig, M. J. Johnson, C. Leary, Compiling machine learning programs via high-level tracing, Systems for Machine Learning 4 (9) (2018).
URL https://mlsys.org/Conferences/doc/2018/146.pdf

[43] Z. Zhibin, Y. Tao, Z. Li, Factors affecting hare–lynx dynamics in the classic time series of the Hudson Bay Company, Canada, Climate Research 34 (2) (2007) 83–89.
URL https://www.int-res.com/abstracts/cr/v34/n2/feature/

[44] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. A. Szerlip, P. Horsfall, N. D. Goodman, Pyro: Deep universal probabilistic programming, Journal of Machine Learning Research 20 (2019) 28:1–28:6.
URL http://jmlr.org/papers/v20/18-403.html

[45] B. Carpenter, Predator-Prey Population Dynamics: The Lotka-Volterra Model in Stan (2018).
URL https://mc-stan.org/users/documentation/case-studies/lotka-volterra-predator-prey.html

[46] Z. Xinyu, J. Cao, R. J. Carroll, On the selection of ordinary differential equation models with application to predator-prey dynamical models, Biometrics 71 (1) (2015) 131–138. doi:10.1111/biom.12243.

[47] E. Mathieu, H. Ritchie, L. Rodés-Guirao, C. Appel, C. Giattino, J. Hasell, B. MacDonald, S. Dattani, D. Beltekian, E. Ortiz-Ospina, M. Roser, Coronavirus pandemic (COVID-19) (2020).
URL https://covid.ourworldindata.org/data/owid-covid-data.csv

[48] L. Grinsztajn, E. Semenova, C. C. Margossian, J. Riou, Bayesian work-

flow for disease transmission modeling in Stan, Statistics in Medicine 40 (27) (2021) 6209–6234. doi:10.1002/sim.9164.

[49] R. E. Moore, C. Rosato, S. Maskell, Refining epidemiological forecasts with simple scoring rules, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 380 (2233) (2022) 20210305. doi:10.1098/rsta.2021.0305.

[50] D. N. Mingo, J. S. Hale, Wasserstein distance prior impact assessment for ODE models (2024). doi:10.5281/zenodo.11553775.

[51] S. Varrette, H. Cartiaux, S. Peter, E. Kieffer, T. Valette, A. Olloh, Management of an academic HPC & research computing facility: The ULHPC experience 2.0, in: Proc. of the 6th ACM High Performance Computing and Cluster Technologies Conf. (HPCCT 2022), Association for Computing Machinery (ACM), Fuzhou, China, 2022. doi:10.1145/3560442.3560445.