Article

Bayesian Noise Modelling for State Estimation of the Spread of COVID-19 in Saudi Arabia with Extended Kalman Filters

Lamia Alyami [1,2]**, Deepak Kumar Panda** [1] **and Saptarshi Das** [1,3,*]

[1] Centre for Environmental Mathematics, Faculty of Environment, Science and Economy, University of Exeter, Penryn Campus, Penryn TR10 9FE, UK; la424@exeter.ac.uk (L.A.); dp457@exeter.ac.uk (D.K.P.)
[2] Department of Mathematics, College of Science, Najran University, Najran 11001, Saudi Arabia
[3] Institute for Data Science and Artificial Intelligence, University of Exeter, Exeter EX4 4QE, UK
[*] Correspondence: s.das3@exeter.ac.uk

Abstract: The epistemic uncertainty in coronavirus disease (COVID-19) model-based predictions using complex noisy data greatly affects the accuracy of pandemic trend and state estimations. Quantifying the uncertainty of COVID-19 trends caused by different unobserved hidden variables is needed to evaluate the accuracy of the predictions for complex compartmental epidemiological models. A new approach for estimating the measurement noise covariance from real COVID-19 pandemic data has been presented based on the marginal likelihood (Bayesian evidence) for Bayesian model selection of the stochastic part of the Extended Kalman filter (EKF), with a sixth-order nonlinear epidemic model, known as the SEIQRD (Susceptible–Exposed–Infected–Quarantined–Recovered–Dead) compartmental model. This study presents a method for testing the noise covariance in cases of dependence or independence between the infected and death errors, to better understand their impact on the predictive accuracy and reliability of EKF statistical models. The proposed approach is able to reduce the error in the quantity of interest compared to the arbitrarily chosen values in the EKF estimation.

Keywords: Bayesian model selection; nested sampling; skew-normal distributions; Bayesian evidence; extended Kalman filter (EKF)

1. Introduction

During the COVID-19 pandemic, there have been various mathematical modelling approaches used to forecast the behaviour of the pandemic, including different stochastic and deterministic compartmental models, in order to achieve greater predictive accuracy. Despite all these efforts, there is a large uncertainty associated with the nature of COVID-19's spreading dynamics. This uncertainty could be due to modelling errors; hidden or unknown factors; incomplete, missing or delayed reporting of data; and other unknown disturbances. In our previous work [1], we fitted a deterministic model called SEIQRD (Susceptible–Exposed– Infected–Quarantined–Recovered–Dead) using the COVID-19 data from Saudi Arabia to predict the evolution of the number of active cases and fatal cases. We used the nested sampling algorithm [2] to fit the epidemiological model's parameters to obtain the multivariate posterior mean of the parameters for simulation. However, we observed a difference between the mean of the deterministic model and the measured data. This difference lies within the modelled uncertainty or error bounds derived from the posterior samples. This has motivated us to further focus on modelling the noise (i.e., the aleatoric uncertainty) to fit this difference or the noise to the best possible probability distribution, which may not necessarily be symmetric in nature. Therefore, in this paper, we aim to identify the distribution that best describes the noise between a smooth deterministic model and the real-world fluctuations in COVID-19 data. To achieve this, we have fitted the parameters of the noise or residuals (i.e., the difference between real data and deterministic

SEIQRD model output), which may help to frame a Kalman filtering (KF) problem with greater precision.

Previously, in [1], for the state estimation results using the EKF with pre-assumed system error covariance (Ξ) and measurement error covariance (Ω) matrices, the noise distributions were empirical since the accurate values of the noise covariances are generally unknown in many applications. Several methods have been proposed to solve the issue of estimating noise covariances, such as using the least-squares methods, as described in [3–5]. Hence, we aim to address this gap from a different perspective by fitting the parameters of the noise distribution to obtain an accurate measurement of the covariance error (Ω) to be used in the EKF state estimation algorithm. This mainly aims to characterize the error of COVID-19 data to better describe the aleatoric uncertainty in addition to the previously tackled epistemic uncertainty and their propagation in state estimation using a nonlinear compartmental pandemic model, called SEIQRD with reinfection. Subsequently, we have proposed a method to resolve an open scientific question from the stochastic data modelling perspective; i.e., within Kalman filtering framework, can we assume that the data-fitting exercise adapts to normality assumptions, or is it strictly applicable to the normal distribution only? We also investigate the consequences of these assumptions of normality and correlation between errors in different variables or using skewed distributions with correlations, etc., in the Kalman filtering framework. In order to address these open questions, we will consider Gaussian and non-Gaussian distributions to represent the errors and fit the residuals between the fitted deterministic model response and real COVID-19 spread data in Saudi Arabia, using generalized Bayesian inference under the Bayesian model selection framework. For linear systems with non-Gaussian noise, the KF remains an optimal state estimator, as described in [6]. This refutes the claims made in much of the published literature, indicating that the KF is not optimal unless the noise is assumed to be Gaussian. Furthermore, the KF in a Gaussian approximation framework only estimates the mean and covariance of the states, which may neglect other valuable information of the underlying distributions, as discussed in [7]. Since many real-world state estimation problems exhibit non-Gaussian patterns, especially when the data are highly skewed or heavy-tailed, it is necessary to consider asymmetric distributions instead of assuming normality.

In recent years, there has been a great deal of interest in skewed distributions and their properties, as demonstrated by [8], who discussed these distributions in detail along with their various applications. Skewed distributions may be necessary to better describe the behaviour of COVID-19 modelling errors in the presence of asymmetric measurement noise. Kalman filtering has also been used in a few studies to estimate the COVID-19 outbreak. Zhu et al. [9] considered a stochastic SEIR model based on EKF to analyse the spread of COVID-19 and quantify the uncertainty with good accuracy as compared to the deterministic model. The ensemble Kalman filter (EnKF) in [10] has been used to estimate the time-varying model parameters with good performance close to the reported data. Developing the EKF to analyse the dynamics of COVID-19 based on maximum likelihood estimation (MLE) for the model parameters has been presented in [11]. Tracking the reproduction number ($\mathcal{R}_0$) with confidence bounds based on the KF has been introduced in [12]. In addition, some literature has compared the performance of COVID-19 prediction with different types of KFs, e.g., fractional-order EKF with an SEIR model [13], switching KF with time-series models [14], cubature KF with the SEIRRPV model (Susceptible–Exposed–Infected–Recovered from exposure–Recovered from infection–Passed away–Vaccinated) in [15] and Quadratic KF with SEIR/ARIMA (AutoRegressive Integrated Moving Average) models [16]. The novelty of this work as opposed to the existing Kalman filtering works on COVID-19 data is that it reports a thorough benchmarking of the state estimation performances using four alternative formations of the EKF with and without correlated noise and skewed distribution based on the residual of the deterministic model and real data. We do not use the traditional COVID-19 data-driven ARMA family of time series models or other machine learning methods such as neural networks for pandemic modelling since

they are mostly black-box models which are difficult to interpret. In contrast, the simple nonlinear deterministic compartmental model is well-studied in the context of epidemic modelling in different scales, and each of the measured or estimated states have meaningful interpretation as opposed to black-box machine learning models. Although recurrent neural networks (RNNs) and other non-parametric time-domain models have been used for the prediction of COVID-19 data, state estimation using compartmental models is a different research direction which has not been addressed in most the contemporary machine learning applications for COVID-19. One of the key motivations here is to estimate the hidden state variables of the compartmental model such as the susceptible, exposed, quarantined, and recovered populations, which is not possible with traditional time-series models as opposed to differential equation-based mechanistic models.

The rest of this paper is organized as follows. In Section 2, we provide a brief review of the KF prediction and the deterministic/stochastic model. In Section 3, the method is outlined. The investigation and the discussion of model comparison results are presented in Section 4, and Section 5 concludes the paper with future research directions.

2. Kalman Filtering and the Pandemic Model

we recall the state space model equations without giving the detailed derivation of the EKF algorithm as follows:

$$
\begin{aligned}
X_{t+1} &= f(X_t) + \xi_t, \quad \text{System equation,} \\
y_t &= HX_t + \omega_t, \quad \text{Measurement equation,}
\end{aligned}
\tag{1}
$$

Here, f is the nonlinear function; ξ_t is the system noise which is assumed to be Gaussian with a zero mean and covariance matrix Ξ; y_t is the measurement sequences for the observed data; H is the measurement matrix, with measurement noise ω_t which is Gaussian-distributed with a zero mean and covariance matrix Ω. The deterministic SEIQRD model in [1] divides the population into six classes: susceptible $S(t)$, exposed $E(t)$, infected $I(t)$, quarantined $Q(t)$, recovered $R(t)$ and dead $D(t)$. The compartments were modelled using the system of nonlinear differential equations, as illustrated in the first block in Figure 1. The EKF estimation based on the deterministic SEIQRD model for the active cases and death cases is updated with the new data in this study, as presented in Figure 2. The data are collected from the Saudi Health Ministry available online [17] with daily measurements from 15 March 2020 to 19 October 2022. The variations between the posterior mean (the output of SEIQRD model) and reported data are clear in both observed cases, as shown for active cases in Figure 2a and for death cases in Figure 2b, along with the EKF estimation, which is very close to the reported data.
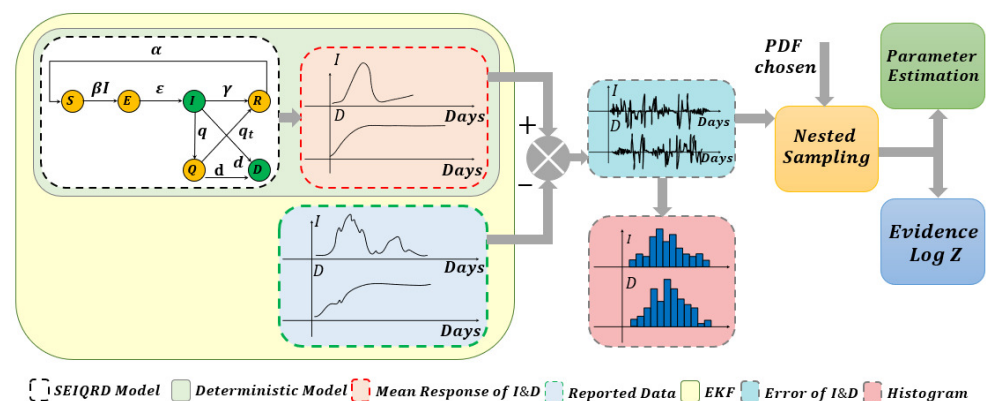

**Figure 1.** COVID-19 data analysis pipeline with deterministic and stochastic compartments.

Now, let us convert this model from the deterministic case to a stochastic case by adding the residuals between the mean posterior response of states I and D. This will help to capture the randomness in the dynamics of disease spreading model that have been

studied in, e.g., [18–20] in order to develop more realistic models matching with the real pandemic data. Then, the deterministic SEIQRD system becomes a stochastic model after adding the noise terms as follows:

$$
\begin{aligned}
\frac{dS}{dt} &= -\beta IS + \alpha R, \\
\frac{dE}{dt} &= \beta IS - \epsilon E, \\
\frac{dI}{dt} &= \epsilon E - \gamma I - qI - dI + w_I, \\
\frac{dQ}{dt} &= qI - q_t Q - dQ, \\
\frac{dR}{dt} &= \gamma I + q_t Q - \alpha R, \\
\frac{dD}{dt} &= dI + dQ + w_D.
\end{aligned}
\tag{2}
$$
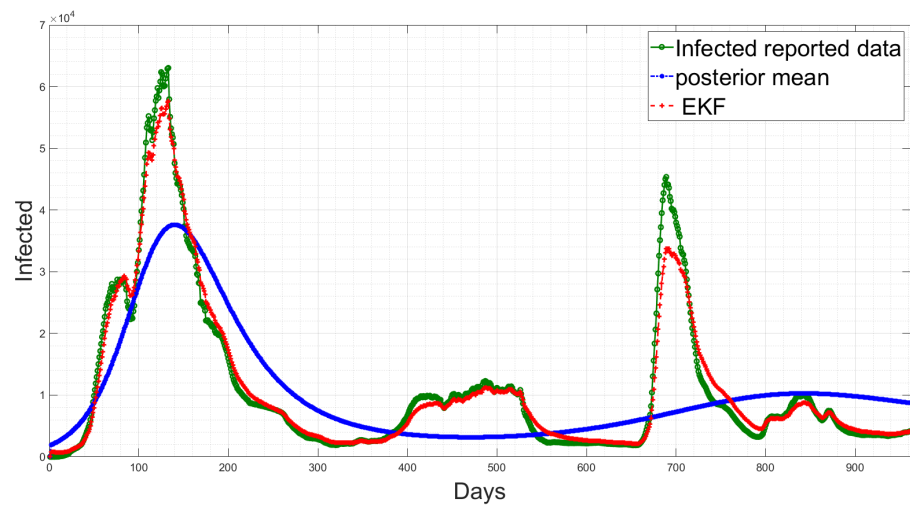
(**a**)
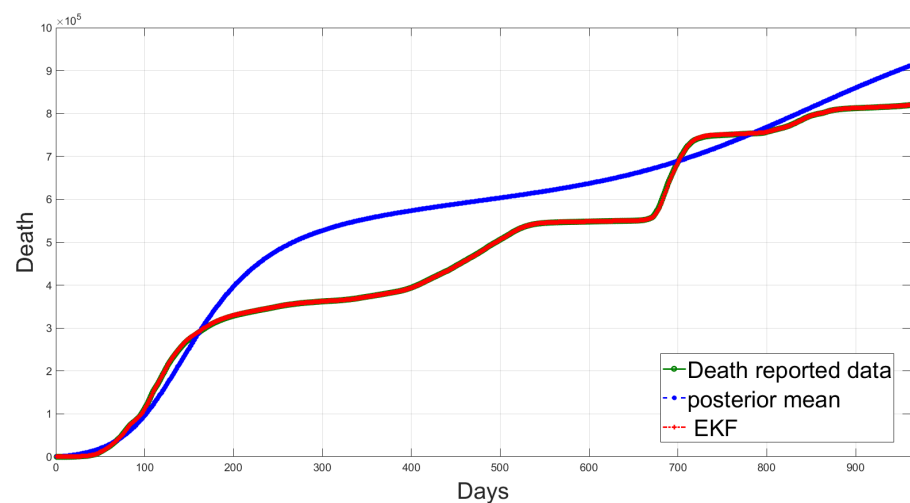
(**b**)

Figure 2. The estimation results of the active cases and death cases using the deterministic model and the EKF. (**a**) Comparison of active cases between the reported data, posterior mean and the proposed EKF; (**b**) Comparison of death cases between the reported data, posterior mean and the proposed EKF.

The unknown model parameters $\{\beta, \gamma, \epsilon, q, q_t, \alpha$ and $d\}$ are non-negative and defined as the infection rate, recovery rate, incubation rate, quarantine rate, quarantine period, reinfection rate and death rate, respectively. Additionally, w_I and w_D are the noise terms that we investigated in this work which are the residuals between the reported data and the fitted deterministic model reported in our earlier work [1]. We can now express the infected and death residuals, i.e., w_I and w_D as:

$$w_I = y_I - \hat{y}_I, \quad w_D = y_D - \hat{y}_D. \tag{3}$$

We take into account that w_I and w_D may be independent or may also be correlated while aiming to estimate their trends with the EKF using different noise distributions. Writing these components in a vector format yields:

$$W = \begin{bmatrix} w_I & w_D \end{bmatrix}^T. \tag{4}$$

Next, as shown in Figure 1, y_I and y_D present the binned frequencies of the error signals since the errors are first binned by turning them into a histogram and the differences in the bins are compared using the new likelihood function. Here, $\hat{y}$ is the binned version of the response of the deterministic SEIQRD model. We examine the residuals W with two different distributions. Firstly, we examine a multivariate normal ($\mathcal{MVN}$) distribution in two cases with correlation and without correlation between the parameters in the covariance matrix. Secondly, we examine the multivariate skew-normal ($\mathcal{MSN}$) distribution in two cases with correlation and without correlation between the parameters in the covariance matrix. The dependence between the two components of the error is modelled by the matrix Σ.

Multivariate Distributions

Now, consider $W \sim \mathcal{MVN}\,(0, \Sigma)$ with zero mean vector and covariance matrix Σ with strictly positive diagonal elements as follows:

$$\Sigma = \begin{bmatrix} \hat{\sigma}_1 & \hat{\sigma}_{12} \\ \hat{\sigma}_{21} & \hat{\sigma}_2 \end{bmatrix}. \tag{5}$$

The likelihood of the $\mathcal{MVN}$ distribution can be written as:

$$L(\theta) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} \prod_{i=1}^{n} (y_j - \hat{y}_j(\theta))^T \Sigma^{-1} (y_j - \hat{y}_j(\theta)) \right\}, \tag{6}$$

where $n = 32$ is the number of bins for the residual data which is the square root of the number of samples chosen as a standard approach for selecting the number of bins. Then, these differences in the bin count are compared in the likelihood function as shown in Equation (6). Taking the log-likelihood function $\ell(\theta)$ for Equation (6) where individual bins of the histogram are summed up yields:

$$\ell(\theta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} \left\{ \sum_{j=1}^{n} (y_j - \hat{y}_j(\theta))^T \Sigma^{-1} (y_j - \hat{y}_j(\theta)) \right\}, \tag{7}$$

where $\theta = (\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_{12})$.

The second distribution that we fit the residuals on is the $\mathcal{MSN}$ distribution proposed in [21]. The n-dimensional random vector y-distributed $\mathcal{MSN}$ distribution with skewness vector α has a probability density function (pdf) given by:

$$f(y) = 2\phi_n(y)\Phi\left(\alpha^T y\right), \quad y \in \mathbb{R}^n, \quad \alpha \in \mathbb{R}^n, \tag{8}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and the standard normal cumulative distribution function (cdf) of the n-dimensional standard normal distribution $\mathcal{N}(0, I)$, respectively. When $\alpha = 0$, the Equation (8) reduces to $\mathcal{N}_n(0, I_n)$.

Now consider $W \sim \mathcal{MSN}\,(0, \Sigma, \alpha)$ with zero mean vector, covariance matrix Σ and skewness vector α and following [22]. The log-likelihood function $\ell(\theta)$ of the Equation (8) can be written as:

$$\ell(\theta) = -\frac{n}{2}\log|\Sigma| - \frac{n}{2}\,\mathrm{tr}\left(\Sigma^{-1}V\right) + \sum_{j}^{n}\log[\Phi\left(\alpha^T\Sigma^{\frac{-1}{2}}(y_j - \hat{y}_j(\theta))\right)], \tag{9}$$

where

$$V = n^{-1}\sum_{j}^{n}(y_j - \hat{y}_j(\theta))(y_i - \hat{y}_j(\theta))^{\mathrm{T}}. \tag{10}$$

The components of the skewness vector α are expressed as:

$$\alpha = \begin{bmatrix} \hat{\alpha}_1 & \hat{\alpha}_2 \end{bmatrix}^T \tag{11}$$

and $\theta = (\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_{12}, \hat{\alpha}_1, \hat{\alpha}_2)$.

3. Methodology for Model Parameters and Uncertainty Estimation

In Bayesian inference methods, the Bayesian evidence or marginal likelihood ($\mathcal{Z}$) is commonly used for model selection and hypothesis testing, as discussed in various studies, e.g., [23–25]. A recent survey [26] extensively explored the computation of marginal likelihood using alternative methods. Many methods have been suggested to compute the marginal likelihood and normalizing constants, as summarized in [27–30]. Among these methods, we used the nested sampling algorithm proposed in [2] due to its simplicity and efficiency in approximating the marginal likelihood for even large-scale Bayesian inference problems, as shown by [31], in comparison to other methods. The nested sampling and efficient Bayesian inference algorithms have been successfully applied in various fields, including cosmology [32,33], epidemiology [1,34], spatio-temporal inference problems in geophysics [35], material science [36] and other such fields where Bayesian inferences are traditionally applied to fit data with mechanistic models.

Nested sampling is considered to be a generic and efficient Bayesian inference engine that provides the posterior samples for the unknown parameters as a by-product while calculating the marginal likelihood or Bayesian evidence $\mathcal{Z}$. Nested sampling aims to integrate the product of likelihood function $\mathcal{L}(\theta)$ and prior probability $\pi(\theta)$ over the entire prior parameter space. In Bayesian inference, the posterior probability distribution of the model parameters is normalized by the constant Bayesian evidence $\mathcal{Z}$ and can be expressed as:

$$p(\theta) = \frac{\mathcal{L}(\theta)\pi(\theta)}{\mathcal{Z}}. \tag{12}$$

The denominator in Equation (12) is known as the Bayesian evidence and is utilized for model comparison to choose the best model or probability distributions where the parameters have to be estimated. The Bayesian evidence $\mathcal{Z}$ in this study for the residuals W and models $M(\theta)$ with parameters θ is given by:

$$\mathcal{Z} = p(W|M(\theta)) = \int \mathcal{L}(\theta) \cdot \pi(M(\theta))d\theta. \tag{13}$$

The main advantage of nested sampling can perform well with high dimensional integration, which is sampled from the prior range towards higher likelihood regions as well as from the multimodal likelihood shape [37]. Once the estimates of the measurement error of Equation (3) are computed by the aforementioned distributions, we will be able to produce the posterior samples and the marginal likelihood $\mathcal{Z}$ for each of them. The number

of sampled points and the range of the uninformative prior are shown explicitly with numerical values in Section 4.2. Figure 3a,b show the frequency of the infected error and the frequency of the death error, which are the differences between the deterministic model fit and the real data, where the distributions are constructed by the kernel density estimate (KDE). The number of bins for both the infected and death error data has been selected using the square root of the number of reported data points, which we also use while calculating the likelihood function. Hence, for accurate characterization of the errors, we need to estimate the parameters of the probability distributions. The Bayesian evidence calculated from the nested sampling helps in selecting the right model for the error distribution, which in turn leads to more accurate state estimation using the EKF. Additionally, the nested sampling provides the posterior samples to estimate the covariance matrices for modelling the correlations between the two error variables, i.e., the infected and death errors.

(a)

(b)
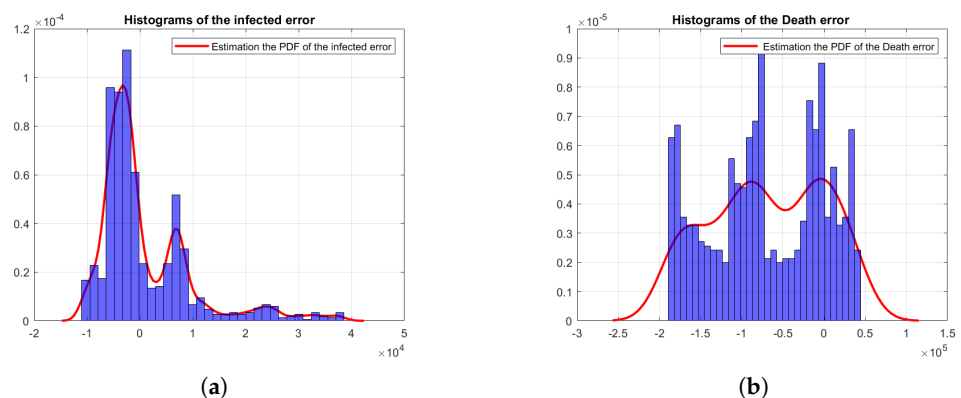
Figure 3. Histogram of the infected and death errors of the output of Equation (3). (**a**) Histogram of the w_I; (**b**) histogram of the w_D.

## 4. Results and Discussions

4.1. Bayesian Inference Results and Error Distributions

In the Bayesian inference framework, choosing the appropriate prior range is not a trivial task in the absence of expert knowledge of the range of possible values of each unknown parameter. In this context, we assumed the uninformative prior of the underlying parameters of the noise distributions to be uniform distributions with specified bounds. We considered the models that assume the quantity of interest is symmetrically and asymmetrically distributed. Since skewness is the key parameter in the skewed distributions, choosing the range of the skewness parameter should be selected carefully in order to obtain the right posterior inference. A discussion about the sensitivity of prior distribution in the asymmetric distributions can be found in [38–40]. We examine the skewness index of the residuals of the infected and death cases by using the sample coefficient of skewness, which is an unbiased estimator for the third central moment obtained by dividing the third central moment over the cubed standard deviation as follows:

$$\text{skewness} = \mathbb{E}\left[(y - \mu)^3\right]/\sigma^3 \tag{14}$$

where μ presents the sample mean and the σ is the sample standard deviation [41,42]. The skewness values calculated using Formula (14) are $w_I = 1.84$ and $w_D = -0.1325$, which indicate that the residual data have departed from normality. Even with the small deviations from normality in death cases where all cases are different from 0, it is significant to be considered, as pointed out in [43]. Selecting the skewness range has been debated in much of the statistics literature; e.g., Ref. [44] describes that if the skewness is greater than 1 in the absolute value, then the data are highly skewed, while skewness between 0.5 and 1 is moderately skewed and skewness between 0 and 0.5 is fairly symmetrical. In addition, Ref. [45] suggested the range of skewness is ± 1.5 and ± 2, as also described

in [46]. To evaluate the magnitude of skewness of the quantity of interest (the residuals), Figure 4 detects the skewness between samples fitted by the skew-normal distribution versus those fitted by a Gaussian distribution using a QQ-plot. Subsequently, from the distributional view of the residual data in Figure 4, the dispersion from normality is clearly identified. However, the prior boundaries in this study are chosen based on the associated data, shown in Table 1, where the prior assigns the same range across different models to make the comparison fair and consistent. Furthermore, two different model comparison approaches are illustrated in the next Section 4.2.
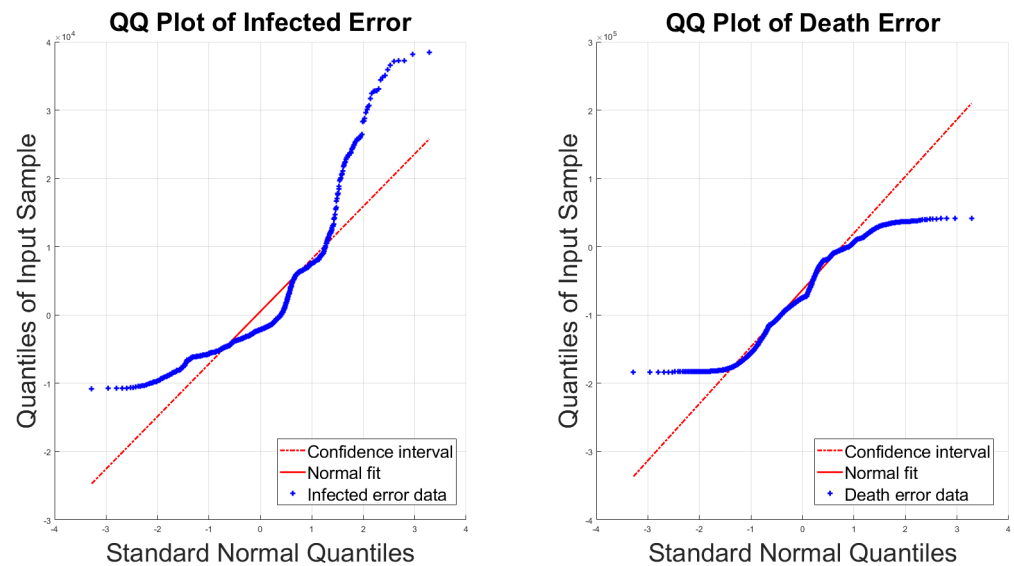
Figure 4. Quantile–Quantile plot of COVID-19 error predictions for infected and death residuals.

4.2. Assumptions on the Error Distributions

In this section, we aim to calculate the evidence of the aforementioned distributions with four different models:

- The $\mathcal{MVN}$ distribution with a full covariance matrix Σ between I and D;
- The $\mathcal{MVN}$ distribution with no correlation between I and D, i.e., diagonal covariance matrix Σ;
- The $\mathcal{MSN}$ distribution with a full covariance matrix, Σ, between I and D;
- The $\mathcal{MSN}$ distribution with no correlation between I and D, i.e., diagonal covariance matrix, Σ.

Conventionally, the first assumption is denoted by Model 1, the second assumption by Model 2, the third assumption by Model 3 and the last assumption by Model 4.

4.3. Sampling from the Posterior of the Unknown Noise Parameters Using Nested Sampling

A set of samples are drawn from the prior distributions for the above statistical models and the prior range is decreased iteratively as the live points navigate towards higher likelihood regions. Thus, during each iteration, the lowest likelihood value is identified and replaced with a new point randomly drawn from the prior distribution subject to the constraint that the newly sampled point must have a likelihood greater than the discarded point. This process continues until the difference in the evidence calculated from the samples gathered by all live points become smaller than a specified threshold ($\Delta\log(\mathcal{Z}) < 0.001$). At this point, the nested sampling algorithm terminates and returns the final estimate of the log-evidence $\log(\mathcal{Z})$, along with its confidence interval. The posterior distribution of the unknown noise parameters is calculated as a by-product of the evidence calculation process. The posterior represents the multivariate probability distribution of the unknown model parameters, given the observed data, and can be used for model comparison, prediction and inference.

For Model 1, the posterior distribution is shown in Figure 5, which shows the correlated samples under the $\mathcal{MVN}$ distribution. The main diagonal in Figure 5 shows the 1D marginal likelihood and provides the maximum a posteriori (MAP) estimate for the parameters $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\sigma}_{12}$. The upper triangular part in Figure 5 shows the scatter plots of the parameters coloured at each point by the log-likelihood values to analyse the correlation between the parameters. Moreover, the lower triangular part shows the marginal bivariate posterior distributions in 2D, estimated by the Gaussian KDE. For the posterior distribution of Model 2, the covariance matrix is diagonal under the $\mathcal{MVN}$ distribution is shown in Figure 6. For the $\mathcal{MSN}$ distribution, the posterior of Model 3 is shown in Figure 7, and the posterior of Model 4 is shown in Figure 8. Furthermore, Figures 5–8 all have similar interpretations in terms of the upper/lower diagonal parts with additional skewness parameters $\hat{\alpha}_1$ and $\hat{\alpha}_2$ in the $\mathcal{MSN}$ distribution cases.
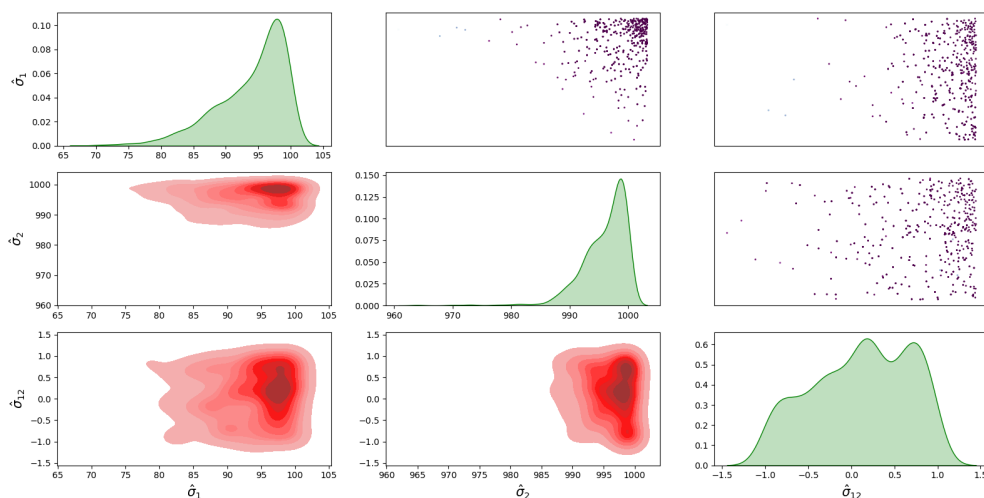

Figure 5. Posterior distribution of Model 1 for the infected and death errors with the correlated covariance matrix.
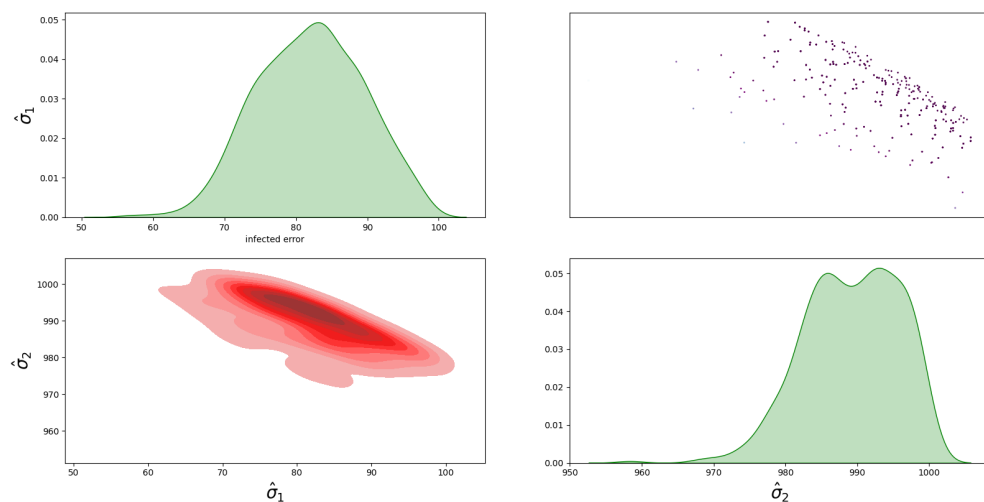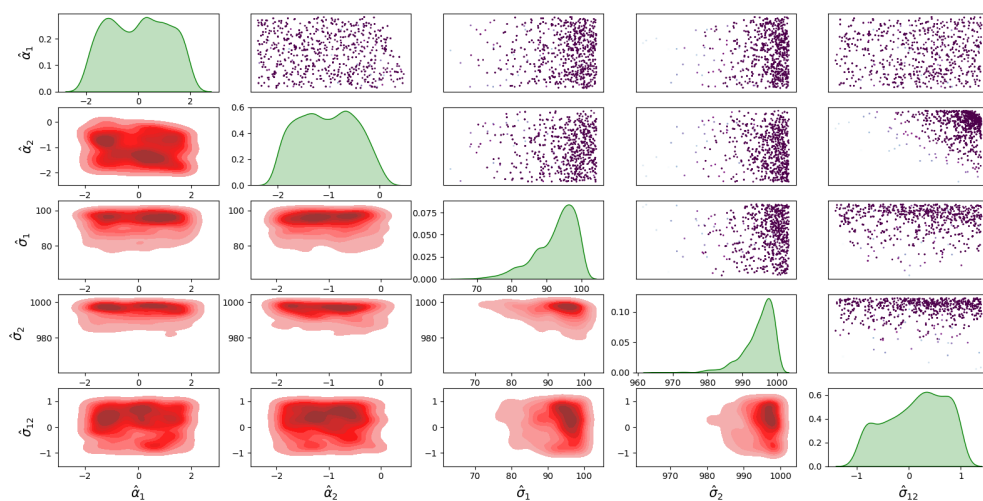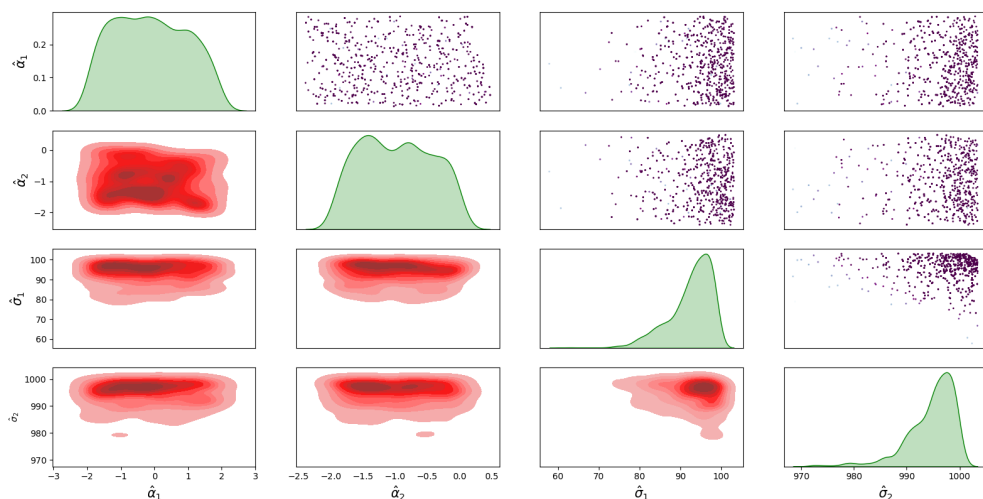

Figure 6. Posterior distribution of Model 2 for the infected and death errors with no correlation in the covariance matrix.

Figure 7. Posterior distribution of Model 3 for the infected and death errors with correlation in the covariance matrix.

Figure 8. Posterior distribution of Model 4 for the infected and death errors with no correlation in the covariance matrix.

Table 1. The uninformative prior distribution for the proposed models.

Model	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\sigma}_{12}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\mathcal{MVN}$ with correlation	$\mathcal{U}[100,1000]$	$\mathcal{U}[100,1000]$	$\mathcal{U}[-1,1]$	- [†]	- [†]
$\mathcal{MVN}$ without correlation	$\mathcal{U}[100,1000]$	$\mathcal{U}[100,1000]$	- [†]	- [†]	- [†]
$\mathcal{MSN}$ with correlation	$\mathcal{U}[100,1000]$	$\mathcal{U}[100,1000]$	$\mathcal{U}[-1,1]$	$\mathcal{U}[-2,2]$	$\mathcal{U}[-2,2]$
$\mathcal{MSN}$ without correlation	$\mathcal{U}[100,1000]$	$\mathcal{U}[100,1000]$	- [†]	$\mathcal{U}[-2,2]$	$\mathcal{U}[-2,2]$

[†] does not exist for the given distribution.

4.4. Bayesian Model Comparison for Selecting the Best Noise Distribution

Table 2 shows the summary of the number of live points (N_{live}), which is a set of points in the prior range which are used to calculate the evidence while navigating towards the high likelihood regions and explore the shrinking prior space efficiently, along with the total number of the likelihood function evaluations (N_{like}) and the log-evidence ($\log \mathcal{Z}$) for each chosen noise distribution obtained by run MultiNest sampling [47] from the multimodal likelihood for all the generated samples. The average number of the N_{live} iterations is proportional to the number of dimensions of the corresponding model, and in this study, we chose $N_{\text{live}} = 20 \times$ dimensions for the unknown PDF parameters. The N_{live} can be user-defined to make a trade-off between the computational cost and the accuracy of the

calculated evidence and posterior distribution. The tolerance between successive iterations is chosen as $\Delta \log(\mathcal{Z}) < 0.001$. A higher $\log \mathcal{Z}$ value is associated with the best-fitted distribution, which is the model $\mathcal{MVN}$ with correlation, followed by the $\mathcal{MSN}$ with correlation, which reflects a linear relationship between the infected sample and the death sample. Despite the tiny difference in the $\log \mathcal{Z}$ values between Model 1 and Model 3, we can justify that the $\mathcal{MVN}$ model is a better choice over the $\mathcal{MVN}$ model since the skewness index is fairly large for death cases and can be in an approximately normal scope for our data. However, for highly skewed data, it is likely that the statistical model selection based on the associated $\log \mathcal{Z}$ values will tend towards the asymmetric distributions. Now, we calculate the mean posterior of these fitted statistical models in the measurement covariance matrix Ω of the EKF in order to optimize the state estimates of the hidden or unmeasured variables, which has been shown in the Section 4.5.

The comparison between two different models can be evaluated using the Bayes factor, as suggested in the literature, e.g., [48–50], to measure the relative strength of evidence for one model over another. The Bayes factor ρ provides the posterior odds, which is the ratio of the marginal likelihoods multiplied by the model prior $P(\text{Model})$ given by the observed data W. Jeffrey's scale [51] provides a qualitative interpretation of the Bayes factor values that can be used to assess the strength of the evidence of one model with respect to another model. Thus, the degree or strength of evidence can be determined by Jeffrey's scale. It suggests an index to distinguish between the models, with $1 < \rho < 2.5$ being substantial, $2.5 < \rho < 5$ being strong and $\rho > 5$ being decisive, as reported in [32]. According to Table 2, the higher $\log \mathcal{Z}$ is associated with Model 1, then computing the Bayes factor for this model versus other models can be presented as:

$$\rho_1 = \log\left[\frac{P(\text{Model-1} \mid W)}{P(\text{Model-2} \mid W)}\right] = \log\frac{P(W \mid \text{Model-1}) \cdot P(\text{Model-1})}{P(W \mid \text{Model-2}) \cdot P(\text{Model-2})} = \log\left[\frac{\mathcal{Z}_{\text{Model-1}}}{\mathcal{Z}_{\text{Model-2}}}\right] \quad (15)$$

$$= \log[\mathcal{Z}_{\text{Model-1}}] - \log[\mathcal{Z}_{\text{Model-2}}] = 3.7253, \quad \text{strong evidence.}$$

Table 2. Bayesian evidence for each error distribution along with the number of live points N_{live} and the number of likelihood function evaluations N_{like}.

Model	N_{live}	$\log \mathcal{Z}$	N_{like}
Model-1: $\mathcal{MVN}$ with correlation	60	-263.215 ± 0.329	715
Model-2: $\mathcal{MVN}$ without correlation	40	-266.940 ± 0.403	463
Model-3: $\mathcal{MSN}$ with correlation	100	-263.546 ± 0.265	1239
Model-4: $\mathcal{MSN}$ without correlation	80	-263.548 ± 0.297	989

This provides evidence that the data are more likely to be under Model 1 than Model 2. Now,

$$\rho_2 = \log\left[\frac{P(\text{Model-1} \mid W)}{P(\text{Model-3} \mid W)}\right] == \log\frac{P(W \mid \text{Model-1}) \cdot P(\text{Model-1})}{P(W \mid \text{Model-3}) \cdot P(\text{Model-3})} = \log\left[\frac{\mathcal{Z}_{\text{Model-1}}}{\mathcal{Z}_{\text{Model-3}}}\right] \quad (16)$$

$$= \log[\mathcal{Z}_{\text{Model-1}}] - \log[\mathcal{Z}_{\text{Model-3}}] = 0.331, \quad \text{weak evidence.}$$

In addition, for Model 4, we obtain:

$$\rho_3 = \log\left[\frac{P(\text{Model-1} \mid W)}{P(\text{Model-4} \mid W)}\right] == \log\frac{P(W \mid \text{Model-1}) \cdot P(\text{Model-1})}{P(W \mid \text{Model-4}) \cdot P(\text{Model-4})} = \log\left[\frac{\mathcal{Z}_{\text{Model-1}}}{\mathcal{Z}_{\text{Model-4}}}\right] \quad (17)$$

$$= \log[\mathcal{Z}_{\text{Model-1}}] - \log[\mathcal{Z}_{\text{Model-4}}] = 0.333, \quad \text{weak evidence.}$$

From the comparative results obtained from ρ_2 and ρ_3, it suggests that Model 3 and Model 4 are equally likely, since their $\log \mathcal{Z}$ values are close to each other.

4.5. State Estimation Using Extended Kalman Filters with the Estimated Noise Distributions

The aim of this subsection is to show the fine-tuning of the EKF with different measurement error covariances Ω and distributions given from the mean posterior of the samples, as shown in Table 3, as compared to the value in the last row, i.e., the EKF5, which is a pre-assumed noise covariance reported in [1]. The root mean squared error (RMSE) for each case between the performance of EKFs and the reported data is presented using:

$$
\begin{aligned}
\text{RMSE} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_{\text{reported }I,D} - X_{\text{EKFs},I,D}\right)^2}, \\
&= \sqrt{\frac{1}{n}\sum_{i=1}^{n}e_i^2}.
\end{aligned}
\tag{18}
$$

Table 3 shows that the EKF2 using $\mathcal{MVN}$ with no correlation has a lower RMSE, as compared to the other EKF variant-based RMSE values. This indicates that the EKF2 predictions are closer to the actual values and have the best accuracy among those models. Furthermore, from Table 3, we observe that EKF1, EKF2, EKF3 and EKF4 all are able to reduce the prediction errors as compared to EKF5, which uses the pre-assumed covariance reported in [1], for the infected cases, while there are slight differences in the death cases. Consequently, using more accurate noise covariance in the stochastic model will enhance the EKF's prediction performance and provide further optimal results in EKF-based state estimation. Furthermore, our methodology focuses on noise model validation using Bayesian evidence and the RMSE score between the predicted and reported infected/death cases. Fitting the deterministic component of the EKF was reported previously in [1], and here, we report the estimated parameters with longer-term historic data. The RMSE has been used as a standard metric to quantify the prediction results in pandemic modelling. However, it is worth mentioning that all the results shown in this paper are subjected to the structure of the underlying nonlinear differential equation, i.e., the structure of the chosen SEIQRD model with reinfection. Since there are several epidemic models reported in the literature with different levels of complexity, if the deterministic model is changed to fit the raw data and generate residuals while including more compartments, geospatial distributions or age/social groups, the ranking of the EKF's performance may be different than the results reported in this paper. The simulations of EKFs have been performed for the number of active cases, as shown in Figure 9, and for the death cases in Figure 10, along with the four hidden states (Susceptible–Exposed–Recovered–Quarantine) in Figure 11. Moreover, we visualize the differences between the individual EKF's performance in the semi-log plot for the infected cases in Figure 12 and for the death cases in Figure 13.

Table 3. Root Mean Square Errors (RMSEs) of the proposed EKFs with different noise distributions and covariance matrices.

EKF	Covariance Matrices		Infected RMSE	Death RMSE
EKF1 $\mathcal{MVN}$ with correlation	$\Xi = 500 \times I_{6\times6}$,	$\Omega = \begin{bmatrix} 93.81 & 0.1001 \\ 0.1001 & 995.91 \end{bmatrix}$	67.1907	26.9871
EKF2 $\mathcal{MVN}$ without correlation	$\Xi = 500 \times I_{6\times6}$,	$\Omega = \text{diag}[82.017, 989.20]$	61.2482	26.6587
EKF3 $\mathcal{MSN}$ with correlation	$\Xi = 500 \times I_{6\times6}$,	$\Omega = \begin{bmatrix} 92.53 & 0.13469 \\ 0.13469 & 994.59 \end{bmatrix}$	66.5684	26.9649
EKF4 $\mathcal{MSN}$ without correlation	$\Xi = 500 \times I_{6\times6}$,	$\Omega = \text{diag}[93.17, 994.56]$	66.8042	26.9032
EKF5	$\Xi = 500 \times I_{6\times6}$,	$\Omega = \text{diag}[100, 1000]$, Ref. [1]	70.1258	27.0861
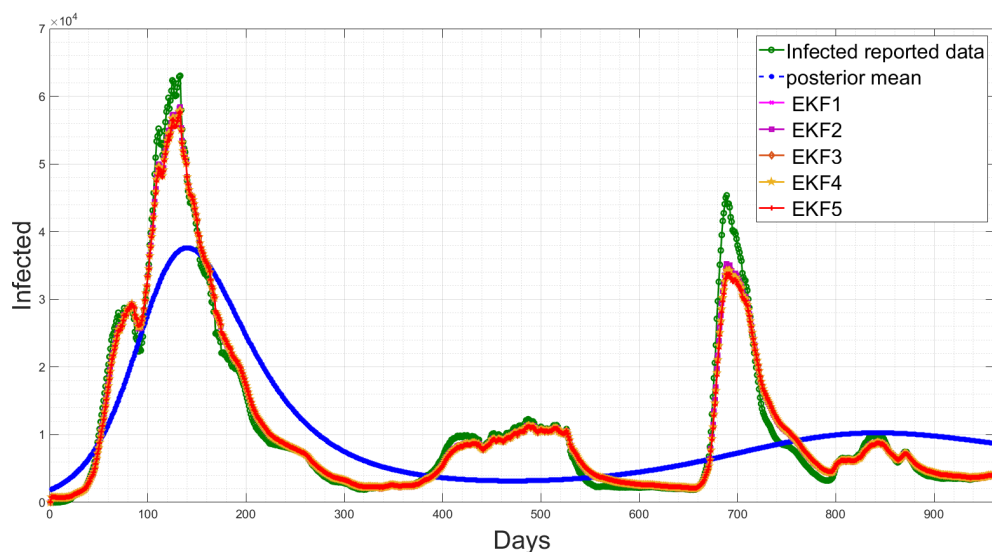
Figure 9. Temporal comparison of the EKFs' performances for the active cases along with the reported data and posterior mean response.
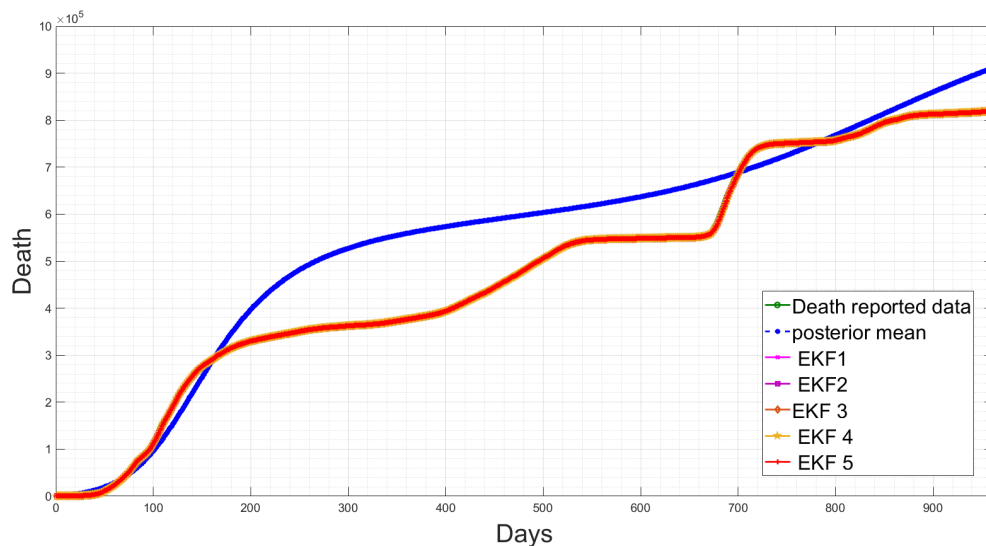

Figure 10. Temporal comparison of the EKFs' performances for the death cases along with the reported data and posterior mean response.

The mean posterior estimates of the SEIQRD model parameters for the long-term Saudi Arabia COVID-19 data are reported in Table 4, which is an improved version over [1] with the most updated data.

Table 4. Mean Posterior estimates of the SEIQRD model parameters for the long-term Saudi Arabia COVID-19 Data.

Parameter	Parameter Value		
	Description	Mean	Standard Deviation
β	infection rate	3×10^{-8}	4.4×10^{-9}
α	reinfection rate	0.0028	1.4×10^{-5}
ϵ	incubation period	0.0353	0.00361
q	quarantine rate	0.9593	0.0033
qt	quarantine period	0.5939	0.0521
d	death rate	4×10^{-3}	2.7×10^{-5}
γ	recovery rate	0.9586	0.0031
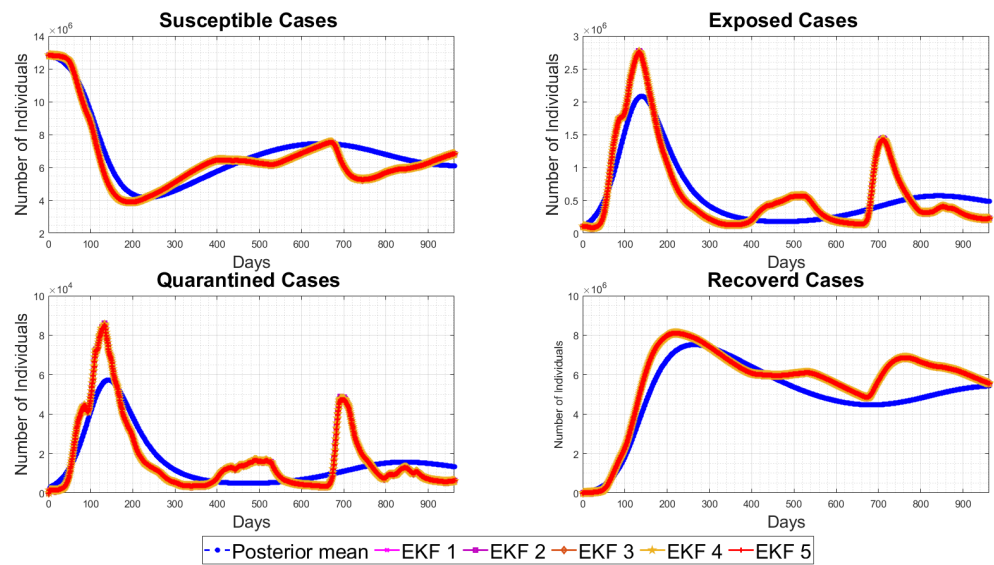
Figure 11. EKFs' performances for the unobserved hidden states, i.e., Susceptible, Exposed, Quarantined, Recovered cases, as compared to the posterior mean response.
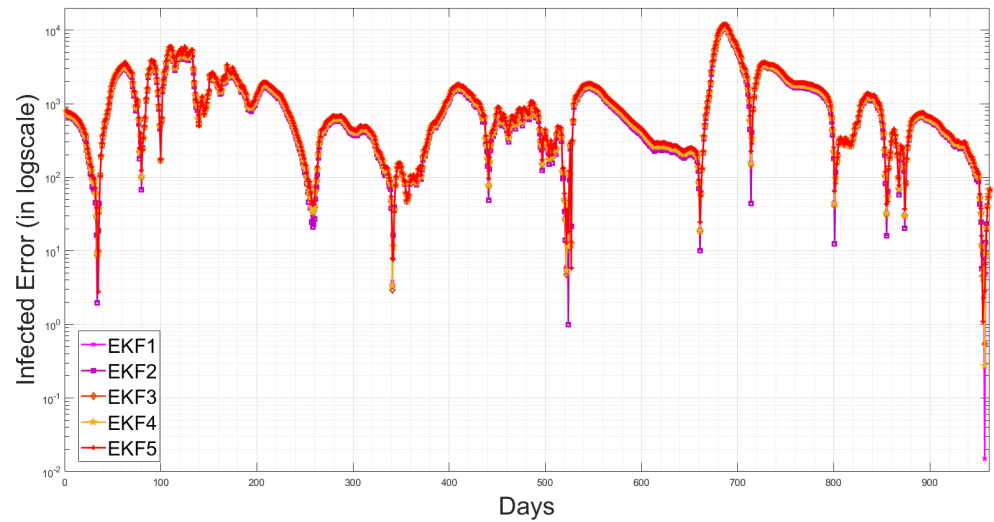
Figure 12. Semi-log plot of the infected cases for the proposed EKFs.
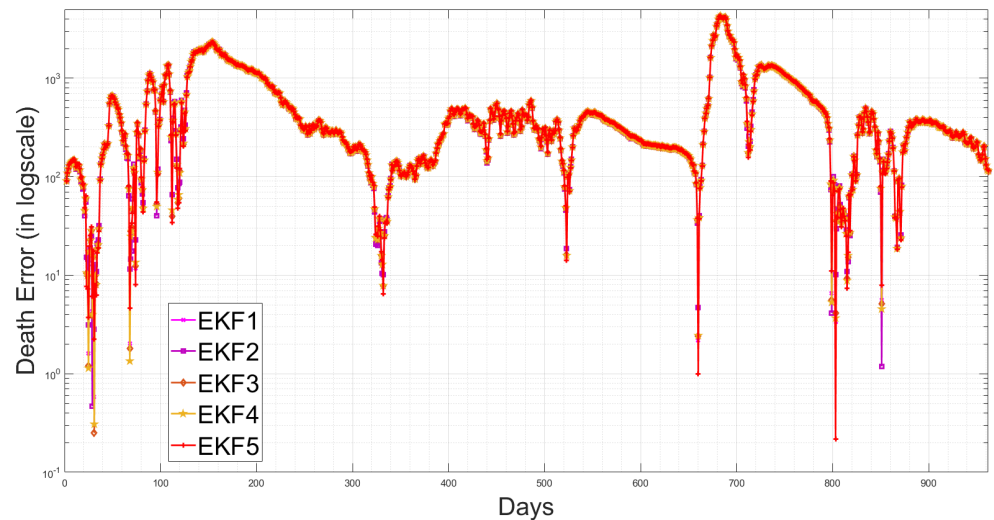
Figure 13. Semi-log plot of the death cases for the proposed EKFs.

5. Conclusions

This paper presents a new method for modelling the uncertainty in the epidemiological SEIQRD model considering dependence and independence in the observation noise. The method is based on the Bayesian inference method called nested sampling, and we provide an extensive analysis to quantify the uncertainty of the noise distributions and feed it back into the Kalman filtering framework for state estimation. We examined the full covariance as well as the diagonal case for each model since it is unknown which model is the best for state estimation. We have shown through simulation studies a comparison between the multivariate normal distribution and the multivariate skew normal distributions for the observation error. We have compared our method using five alternative formulations of the EKF with different noise distributions and manually tuned noise covariance matrices in order to make a fair comparison between different methods. Our main objective here is to obtain accurate state estimation for unmeasurable or hidden states in the epidemiological model and not just forecasting measurements, unlike black-box, non-parametric time series models.

5.1. Main Contributions and Limitations

The main contribution of the proposed approach is to fine-tune the EKFs, including the nonlinear epidemiological SEIQRD model with reinfection, using different structures of the measurement error distribution and covariance Ω in a more systematic way as compared to the previous work in [1]. The previous work in [1] assumed that Ω can take an arbitrary value and can be manually tuned, and despite that, we obtained decent prediction performance using the EKF. In this paper, we further fine-tune all the noise parameters in EKF based on the Bayesian evidence criterion and we conclude that all the examined four noise distributions are better than the pre-assumed or hand-tuned noise covariance, based on the RMSE criterion of the predictions of two measured variables-infected and death cases. Moreover, this generalized method can be performed regardless of which type of EKF is used including being agnostic on the chosen deterministic compartmental model structure. The second major contribution of this paper is to estimate the hidden states of COVID-19 dynamics from the chosen SEIQRD model, i.e., the susceptible, exposed quarantine and recovered cases, along with their respective uncertainty bounds, which is a non-trivial task, using standard time-series models. Such estimation is crucial to supporting decision making for pandemic management by the government. However, in this study, we do not tune the system covariance Ξ due to the high computational burden since our goal is to match the realistic measurement error in I and D.

There have been numerous studies on pandemic modelling based on a single country or even smaller regional data in different states or counties. However, parameter estimation, comparing the deterministic and stochastic parts of models and designing state estimation algorithms using nonlinear models are not straightforward tasks. Therefore, our study was based on only Saudi Arabian data, and we plan to extend it to other countries in future works. Although many countries have published COVID-19 data, mostly their cumulative death records, data on actives cases (different from daily new cases and confirmed and cumulative infected data) which is traditionally used in pandemic models are very hard to obtain. Moreover, the range of estimated parameters changes with the infrastructure of different countries, which motivated us to develop the methodology for a single country initially and extend it in the future by modifying the prior range. However, for better reproducibility, we report the estimated SEIQRD model parameters in Table 4 before noise modelling in the EKF algorithm.

5.2. Future Work

Several extensions of the reported work are possible:

- This work can be further extended with other, more complex nonlinear epidemiological compartmental models under the EKF framework with more experimental evidence.
- Another possible future direction is to include more model comparisons within different skewness distribution families using similar approaches such as the class of

skewness distributions presented in [52]. However, the challenge with this class is dealing with the extra parameters as well as prior range selection, which should be carried out carefully as there is a lack of interpretability in the literature which needs further investigation.

Overall, there is still a risk of a global pandemic like the one experienced during COVID-19, which is assumed in the SEIQRD model, and adopting various protective measures to prevent the transmission can significantly minimise that risk. It is fundamental to stay informed using recent scientific research and to continue to follow the guidance of public health organizations in order to protect ourselves and others. We are currently in the stage of living with COVID-19 after the substantial effort of developing successful treatments and effective vaccines, although there may be different variants of the virus that have not been discovered yet. Forecasting and developing predictive state estimation strategies for potential future pandemics is a challenging task for any government and has not been discussed well in the contemporary scientific literature since it requires interdisciplinary research involving epidemiology, public health, medicine, applied mathematics and statistics. In [53], the authors discussed some future pandemic threats associated with changes in the biosphere, such as climate change, loss of biodiversity and deforestation. However, this pandemic made the world vigilant and more prepared for other global crises. Consequently, we can analyse other potential future pandemics which are related to infectious diseases and pathogens that could threaten public health. Investigation into possible future pandemics may be classified into a few categories as follows:

- Unknown microbes or microorganisms that arrive from space, which are called panspermia, as discussed in [54,55]; however, there is a lack of evidence for this hypothesis and microbial data collection in space is limited due to the high cost, the complexity of the experimental setup and the high level of risk involved. As there are no clear findings of microorganisms coming from asteroids, comets or spacecraft, the evidence of possible infection is rather low. COVID-19 data can help in improving modelling of unknown pandemics from outer space.

- The outbreak of experimental microbes in scientific laboratories, due to accidents or poor management practices, can result in infections such as Brucella abortus, which can cause foetal death in pregnant women. Moreover, the origin of COVID-19 is still unknown, and whether it emerged through natural spillover, trans-species migration or a laboratory accident is still uncertain. However, a laboratory accident cannot be excluded as a potential risk for similar or even larger future pandemics, as discussed in [56,57].

- Future pandemics may also be caused by anthropogenic roots such as political conflicts or wars between countries, continents and specific genotypes, which can be modelled and controlled using the COVID-19 data as a test scenario [58,59]. This may also include manipulating the birth rate and other constants between the model compartments related to the health system's infrastructure, such as the hospital capacity, quarantine period and reinfection rate in a country or region.

We have learnt a lesson from the COVID-19 pandemic, which had the largest number of deaths in recent history. We developed robust combat policies and controlled public fear and uncertainty affecting quality of life, the economy and education to name a few amongst many social aspects [60]. Most modern epidemiological studies in humans, animals and plants utilize a wide range of mechanistic compartmental models as well as real-world data to accurately characterize the evolution of an epidemic through estimation algorithms and to assess risk levels for possible future pandemics which may cause wider-scale disruptions in daily life. Therefore, our modelling and state estimation results using the Saudi Arabian COVID-19 data can be seen as a generalized framework for systematically estimating the deterministic and stochastic components of a pandemic within an extended Kalman filtering framework utilizing the SEIQRD and even more complex compartmental models.

Abbreviations

The following abbreviations are used in this manuscript:

KF	Kalman filter
EKF	Extended Kalman filter
MVN	Multivariate normal distribution
MSN	Multivariate skew normal distribution
PDF	Probability density function
CDF	Cumulative density function
KDE	Kernel density estimate
MAP	Maximum a posteriori estimate
RMSE	Root Mean Squared Error

References

1. Alyami, L.; Das, S. State Estimation of the Spread of COVID-19 in Saudi Arabia using Extended Kalman Filter. In Proceedings of the 2022 Sensor Signal Processing for Defence Conference (SSPD), London, UK, 13–14 September 2022; pp. 1–5.
2. Skilling, J. Nested sampling for general Bayesian computation. *Bayesian Anal.* **2006**, *1*, 833–859. [CrossRef]
3. Odelson, B.J.; Lutz, A.; Rawlings, J.B. The autocovariance least-squares method for estimating covariances: Application to model-based control of chemical reactors. *IEEE Trans. Control. Syst. Technol.* **2006**, *14*, 532–540. [CrossRef]
4. Odelson, B.J.; Rajamani, M.R.; Rawlings, J.B. A new autocovariance least-squares method for estimating noise covariances. *Automatica* **2006**, *42*, 303–308. [CrossRef]
5. Åkesson, B.M.; Jørgensen, J.B.; Poulsen, N.K.; Jørgensen, S.B. A generalized autocovariance least-squares method for Kalman filter tuning. *J. Process. Control* **2008**, *18*, 769–779. [CrossRef]
6. Simon, D. *Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches*; John Wiley & Sons: Hoboken, NJ, USA, 2006.
7. Julier, S.J. Skewed approach to filtering. In Proceedings of the Signal and Data Processing of Small Targets, Orlando, FL, USA, 14–16 April 1998; Volume 3373, pp. 271–282.
8. Genton, M.G. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2004.
9. Zhu, X.; Gao, B.; Zhong, Y.; Gu, C.; Choi, K.S. Extended Kalman filter based on stochastic epidemiological model for COVID-19 modelling. *Comput. Biol. Med.* **2021**, *137*, 104810. [CrossRef] [PubMed]
10. Lal, R.; Huang, W.; Li, Z. An application of the ensemble Kalman filter in epidemiological modelling. *PLoS ONE* **2021**, *16*, e0256227. [CrossRef] [PubMed]
11. Song, J.; Xie, H.; Gao, B.; Zhong, Y.; Gu, C.; Choi, K.S. Maximum likelihood-based extended Kalman filter for COVID-19 prediction. *Chaos Solitons Fractals* **2021**, *146*, 110922. [CrossRef]
12. Arroyo-Marioli, F.; Bullano, F.; Kucinskas, S.; Rondón-Moreno, C. Tracking R of COVID-19: A new real-time estimation using the Kalman filter. *PLoS ONE* **2021**, *16*, e0244474. [CrossRef]
13. Bansal, R.; Kumar, A.; Singh, A.K.; Kumar, S. Stochastic filtering based transmissibility estimation of novel coronavirus. *Digit. Signal Process.* **2021**, *112*, 103001. [CrossRef]
14. Zeng, X.; Ghanem, R. Dynamics identification and forecasting of COVID-19 by switching Kalman filters. *Comput. Mech.* **2020**, *66*, 1179–1193. [CrossRef] [PubMed]
15. Nanda, S.K.; Kumar, G.; Bhatia, V.; Singh, A.K. Kalman-based compartmental estimation for COVID-19 pandemic using advanced epidemic model. *Biomed. Signal Process. Control* **2023**, *84*, 104727. [CrossRef] [PubMed]

16. Marques, J.A.L.; Gois, F.N.B.; Xavier-Neto, J.; Fong, S.J. *Predictive Models for Decision Support in the COVID-19 Crisis*; Springer: Cham, Switzeralnd, 2021.

17. Saudi Ministry of Health. Available online: https://covid19.moh.gov.sa (accessed on 30 November 2022).

18. Zhang, Z.; Zeb, A.; Hussain, S.; Alzahrani, E. Dynamics of COVID-19 mathematical model with stochastic perturbation. *Adv. Differ. Eqs.* **2020**, *2020*, 451. [CrossRef]

19. Khan, T.; Zaman, G.; El-Khatib, Y. Modeling the dynamics of novel coronavirus (COVID-19) via stochastic epidemic model. *Results Phys.* **2021**, *24*, 104004. [CrossRef] [PubMed]

20. Hussain, S.; Madi, E.N.; Khan, H.; Etemad, S.; Rezapour, S.; Sitthiwirattham, T.; Patanarapeelert, N. Investigation of the stochastic modeling of COVID-19 with environmental noise from the analytical and numerical point of view. *Mathematics* **2021**, *9*, 3122. [CrossRef]

21. Azzalini, A.; Valle, A.D. The multivariate skew-normal distribution. *Biometrika* **1996**, *83*, 715–726. [CrossRef]

22. Azzalini, A.; Capitanio, A. Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **1999**, *61*, 579–602. [CrossRef]

23. Gelfand, A.E.; Dey, D.K. Bayesian model choice: Asymptotics and exact calculations. *J. R. Stat. Soc. Ser. B (Methodol.)* **1994**, *56*, 501–514. [CrossRef]

24. Chen, M.H.; Shao, Q.M. On Monte Carlo methods for estimating ratios of normalizing constants. *Ann. Stat.* **1997**, *25*, 1563–1594. [CrossRef]

25. Chib, S. Marginal likelihood from the Gibbs output. *J. Am. Stat. Assoc.* **1995**, *90*, 1313–1321. [CrossRef]

26. Llorente, F.; Martino, L.; Delgado, D.; Lopez-Santiago, J. Marginal likelihood computation for model selection and hypothesis testing: An extensive review. *SIAM Rev.* **2023**, *65*, 3–58. [CrossRef]

27. Bos, C.S. A comparison of marginal likelihood computation methods. In *Proceedings of the COMPSTAT: Proceedings in Computational Statistics*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 111–116.

28. Friel, N.; Wyse, J. Estimating the evidence—A review. *Stat. Neerl.* **2012**, *66*, 288–308. [CrossRef]

29. Robert, C.P.; Wraith, D. Computational methods for Bayesian model choice. In *AIP Conference Proceedings*; American Institute of Physics: College Park, MD, USA, 2009; Volume 1193, pp. 251–262.

30. Ardia, D.; Baştürk, N.; Hoogerheide, L.; Van Dijk, H.K. A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihood. *Comput. Stat. Data Anal.* **2012**, *56*, 3398–3414. [CrossRef]

31. Chopin, N.; Robert, C.P. Properties of nested sampling. *Biometrika* **2010**, *97*, 741–755. [CrossRef]

32. Mukherjee, P.; Parkinson, D.; Liddle, A.R. A nested sampling algorithm for cosmological model selection. *Astrophys. J.* **2006**, *638*, L51. [CrossRef]

33. Vegetti, S.; Koopmans, L.V. Bayesian strong gravitational-lens modelling on adaptive grids: Objective detection of mass substructure in Galaxies. *Mon. Not. R. Astron. Soc.* **2009**, *392*, 945–963. [CrossRef]

34. Li, Y.I.; Turk, G.; Rohrbach, P.B.; Pietzonka, P.; Kappler, J.; Singh, R.; Dolezal, J.; Ekeh, T.; Kikuchi, L.; Peterson, J.D.; et al. Efficient Bayesian inference of fully stochastic epidemiological models with applications to COVID-19. *R. Soc. Open Sci.* **2021**, *8*, 211065. [CrossRef]

35. Das, S.; Hobson, M.P.; Feroz, F.; Chen, X.; Phadke, S.; Goudswaard, J.; Hohl, D. Microseismic event detection in large heterogeneous velocity models using Bayesian multimodal nested sampling. *Data-Centric Eng.* **2021**, *2*, e1. [CrossRef]

36. Pártay, L.B.; Csányi, G.; Bernstein, N. Nested sampling for materials. *Eur. Phys. J. B* **2021**, *94*, 159. [CrossRef]

37. Buchner, J. A statistical test for nested sampling algorithms. *Stat. Comput.* **2016**, *26*, 383–392. [CrossRef]

38. Ferreira, J.T.; Steel, M.F. Model comparison of coordinate-free multivariate skewed distributions with an application to stochastic frontiers. *J. Econom.* **2007**, *137*, 641–673. [CrossRef]

39. Ferreira, J.T.; Steel, M.F. On describing multivariate skewed distributions: A directional approach. *Can. J. Stat.* **2006**, *34*, 411–429. [CrossRef]

40. Rubio, F.; Steel, M. Bayesian modelling of skewness and kurtosis with Two-Piece Scale and shape distributions. *Electron. J. Stat.* **2015**, *9*, 1884–1912. [CrossRef]

41. Joanes, D.N.; Gill, C.A. Comparing measures of sample skewness and kurtosis. *J. R. Stat. Soc. Ser. D (Stat.)* **1998**, *47*, 183–189. [CrossRef]

42. Serfling, R.J. *Approximation Theorems of Mathematical Statistics*; John Wiley & Sons: Hoboken, NJ, USA, 2009.

43. Demir, S. Comparison of normality tests in terms of sample sizes under different skewness and Kurtosis coefficients. *Int. J. Assess. Tools Educ.* **2022**, *9*, 397–409. [CrossRef]

44. Bulmer, M.G. *Principles of Statistics*; Courier Corporation: Gloucester, MA, USA, 1979.

45. Tabachnick, B.G.; Fidell, L.S.; Ullman, J.B. *Using Multivariate Statistics*; Pearson: Boston, MA, USA, 2013; Volume 6.

46. Field, A. *Discovering Statistics Using IBM SPSS Statistics*; Sage: Thousand Oaks, CA, USA, 2013.

47. Feroz, F.; Hobson, M.; Bridges, M. MultiNest: An efficient and robust Bayesian inference tool for cosmology and particle physics. *Mon. Not. R. Astron. Soc.* **2009**, *398*, 1601–1614. [CrossRef]

48. Verdinelli, I.; Wasserman, L. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *J. Am. Stat. Assoc.* **1995**, *90*, 614–618. [CrossRef]

49. Dickey, J.M. The weighted likelihood ratio, linear hypotheses on normal location parameters. *Ann. Math. Stat.* **1971**, *42*, 204–223. [CrossRef]

50. DiCiccio, T.J.; Kass, R.E.; Raftery, A.; Wasserman, L. Computing Bayes factors by combining simulation and asymptotic approximations. *J. Am. Stat. Assoc.* **1997**, *92*, 903–915. [CrossRef]
51. Jeffreys, H. *Theory of Probability*; Oxford University Press: Oxford, UK, 1961.
52. Gonzalez-Farias, G.; Dominguez-Molina, A.; Gupta, A.K. Additive properties of skew normal random vectors. *J. Stat. Plan. Inference* **2004**, *126*, 521–534. [CrossRef]
53. Thoradeniya, T.; Jayasinghe, S. COVID-19 and future pandemics: A global systems approach and relevance to SDGs. *Glob. Health* **2021**, *17*, 59. [CrossRef]
54. Wesson, P.S. Panspermia, past and present: Astrophysical and biophysical conditions for the dissemination of life in space. *Space Sci. Rev.* **2010**, *156*, 239–252. [CrossRef]
55. Steele, E.J.; Gorczynski, R.M.; Lindley, R.A.; Tokoro, G.; Temple, R.; Wickramasinghe, N.C. Origin of new emergent Coronavirus and Candida fungal diseases—Terrestrial or cosmic? *Adv. Genet.* **2020**, *106*, 75–100. [PubMed]
56. Bloom, J.D.; Chan, Y.A.; Baric, R.S.; Bjorkman, P.J.; Cobey, S.; Deverman, B.E.; Fisman, D.N.; Gupta, R.; Iwasaki, A.; Lipsitch, M.; et al. Investigate the origins of COVID-19. *Science* **2021**, *372*, 694. [CrossRef] [PubMed]
57. Maxmen, A.; Mallapaty, S. The COVID lab-leak hypothesis: What scientists do and don't know. *Nature* **2021**, *594*, 313–315. [CrossRef] [PubMed]
58. Merrin, W. Anthropocenic war: Coronavirus and total demobilization. *Digit. War* **2020**, *1*, 36–49. [CrossRef]
59. Lyon, R.F. The COVID-19 response has uncovered and increased our vulnerability to biological warfare. *Mil. Med.* **2021**, *186*, 193–196. [CrossRef]
60. Pan, X.; Ojcius, D.M.; Gao, T.; Li, Z.; Pan, C.; Pan, C. Lessons learned from the 2019-nCoV epidemic on prevention of future infectious diseases. *Microbes Infect.* **2020**, *22*, 86–91. [CrossRef]