



Article

# Comparison of Frequentist and Bayesian Generalized Additive Models for Assessing the Association between Daily Exposure to Fine Particles and Respiratory Mortality: A Simulation Study

Xin Fang <sup>1</sup>, Bo Fang <sup>2</sup>, Chunfang Wang <sup>2</sup>, Tian Xia <sup>3</sup>, Matteo Bottai <sup>1</sup>, Fang Fang <sup>4</sup> and Yang Cao <sup>5,\*</sup> 

<sup>1</sup> Unit of Biostatistics, Institute of Environmental Medicine, Karolinska Institutet, Stockholm 17177, Sweden; fx63783985@hotmail.com (X.F.); matteo.bottai@ki.se (M.B.)

<sup>2</sup> Division of Vital Statistics, Shanghai Municipal Center for Disease Control and Prevention, Shanghai 200336, China; fangbo@scdc.sh.cn (B.F.); wangchunfang@scdc.sh.cn (C.W.)

<sup>3</sup> Institute of Health Information, Shanghai Municipal Center for Disease Control and Prevention, Shanghai 200336, China; xiatian@scdc.sh.cn

<sup>4</sup> Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden; fang.fang@ki.se

<sup>5</sup> Clinical Epidemiology and Biostatistics, School of Medical Sciences, Örebro University, Örebro 70182, Sweden

\* Correspondence: yang.cao@oru.se; Tel.: +46-19-602-62-36

Received: 1 January 2019; Accepted: 26 February 2019; Published: 1 March 2019



**Abstract:** *Objective:* To compare the performance of frequentist and Bayesian generalized additive models (GAMs) in terms of accuracy and precision for assessing the association between daily exposure to fine particles and respiratory mortality using simulated data based on a real time-series study. *Methods:* In our study, we examined the estimates from a fully Bayesian GAM using simulated data based on a genuine time-series study on fine particles with a diameter of 2.5  $\mu\text{m}$  or less (PM<sub>2.5</sub>) and respiratory deaths conducted in Shanghai, China. The simulation was performed by multiplying the observed daily death with a random error. The underlying priors for Bayesian analysis are estimated using the real world time-series data. We also examined the sensitivity of Bayesian GAM to the choice of priors and to true parameter. *Results:* The frequentist GAM and Bayesian GAM show similar means and variances of the estimates of the parameters of interest. However, the estimates from Bayesian GAM show relatively more fluctuation, which to some extent reflects the uncertainty inherent in Bayesian estimation. *Conclusions:* Although computationally intensive, Bayesian GAM would be a better solution to avoid potentially over-confident inferences. With the increasing computing power of computers and statistical packages available, fully Bayesian methods for decision making may become more widely applied in the future.

**Keywords:** Bayesian statistics; generalized additive model; time-series analysis; fine particulate matter; respiratory mortality

## 1. Introduction

With accelerated urban development and modernization, air pollution is worsening and its impact on human health has been the main research topic in developing countries [1]. Air pollutants include gaseous pollutants and particulate matter (PM). Studies have shown that PMs with an aerodynamic diameter of 2.5  $\mu\text{m}$  or less (referred to as PM<sub>2.5</sub> or fine particles) have a greater impact on human health [2,3]. The impact of both long term (chronic) and short term (acute) exposure to PMs has been

widely studied, with the former being most often estimated from cohort studies [4] and the latter from ecological time-series studies [5]. However, cohort studies are expensive and time-consuming to implement due to the long follow-up period required for collecting individual-level pollution and disease data [6]. This has led to the use of spatiotemporal ecological study design in this field [7,8], which takes advantage of routinely collected data [5] such as data from the air quality monitoring (AQM) stations in Beijing and the Causes of Death Registry (CDR) of the Chinese Centers for Disease Control and Prevention [9]. Although causal inference is an important problem in time-series studies due to the difficulty of selecting an appropriate regression model that can be well fitted to the data, they contribute to and independently corroborate the body of evidence provided by cohort studies [10].

The semiparametric Poisson regression has been widely used for time-series analyses of air pollution and health, which uses daily mortality or morbidity counts as the outcome, linear terms to measure the percentage of increase in the outcome associated with elevations in air pollution levels, and smooth functions of time and weather variables to adjust for time-varying confounders [11]. Generalized linear models (GLMs) with parametric splines (e.g., natural cubic splines) [12] or generalized additive models (GAMs) with nonparametric splines (e.g., smoothing splines or locally weighted smoothers or (LOESS) [13] are used to estimate effects associated with exposure to air pollution while accounting for smooth fluctuations in outcome that confound the estimated effects of pollution. GAM and GLM can be applied in similar situations, but they serve different analytic purposes. GLM emphasizes estimation and inference for the parameters of the model, while GAM focuses on exploring data non-parametrically. GAM is more suitable for exploring the data and visualizing the relationship between the dependent variable and the independent variables [14].

The basic form of GAM applied in air pollution and health studies may be expressed using Equation (1) [11]:

$$\log(E(Y_t)) = \beta_0 + \beta_l X_{t-l} + S(t) + S(\text{weather}) + \boldsymbol{\phi} \cdot \text{DOW}_t \quad (1)$$

where  $Y_t$  is count of daily mortality or morbidity,  $\beta_0$  denotes the intercept,  $t$  indicates calendar day,  $X_t$  are daily concentrations of the studied air pollutant, i.e.,  $\text{PM}_{2.5}$  in our study,  $l$  is the lag time of the pollution exposure (which is generally restricted to one to seven days for acute effects),  $S(t)$  denotes a smooth function of a covariate (calendar day or meteorological variables such as temperature and humidity). The smooth functions are usually constructed using LOESS, smoothing splines or natural cubic splines.  $\boldsymbol{\phi}$  is the vector of the regression coefficients associated with vector  $\text{DOW}_t$  (indicating the 7 days of a week) for the  $t$ th day.  $\beta_l$  is the parameter of interest describing the change in the logarithm of the average mortality count over population per unit of change in  $X_{t-l}$ , which is generally interpreted as the percentage of increase in mortality for every 10 units or a standard deviation (SD) or interquartile range (IQR) of increase in ambient concentrations of the studied air pollutant at lag  $l$ . The reason that the model is called semiparametric is that it assumes a linear relation with  $X_{t-l}$  and unknown functional relations with time and weather variables.

The conventional algorithm for fitting GAM (hereinafter called frequentist GAM) is the backfitting algorithm [15] and the corresponding robust estimation method has also been developed [16]. A disadvantage of backfitting is that it is difficult to integrate with the estimation of the degree of smoothness of the model terms, so that in practice the user must set these, or select between a modest set of pre-defined smoothing levels. Although the degree of smoothness can be estimated as part of model fitting using generalized cross-validation or by restricted maximum likelihood when the smooth components are represented using smoothing spline, it carries a computational cost [17]. The computationally efficient approaches such as fully Bayesian method have thus been developed in recent years [18].

Although there are some applications [19–23] of Bayesian GAM analyses in recent years, few of them compared the performance of frequentist and Bayesian GAMs in terms of accuracy and precision. In our study, we examined the estimates from a fully Bayesian GAM using simulated data based

on underlying ‘true’ parameters from a genuine time-series study on PM<sub>2.5</sub> and respiratory deaths conducted in Shanghai, China.

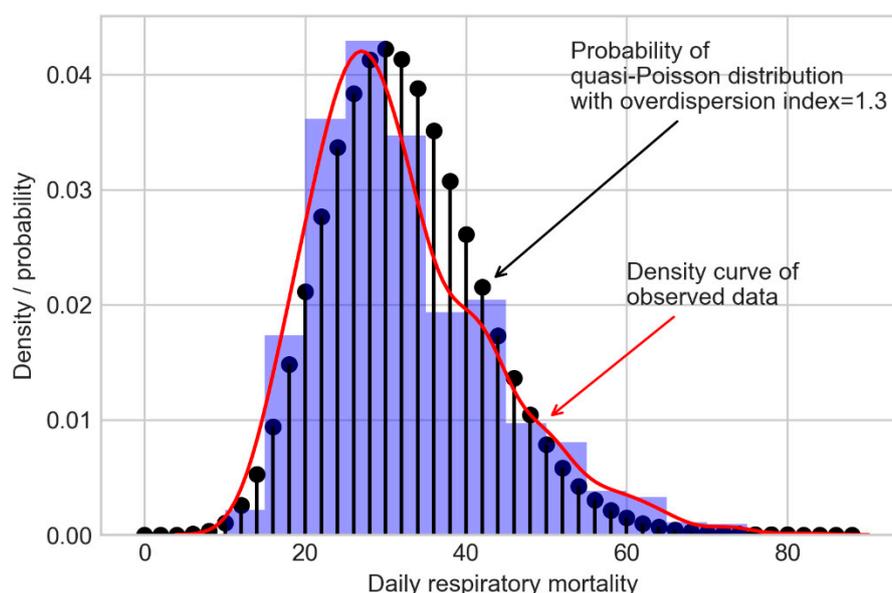
## 2. Methods

### 2.1. Real World Data

The data of our study were from the Swedish-Chinese joint project Assessing Variability in the Short-term Impact of Air Pollution and Extreme Weather Conditions on Non-accidental Deaths in Shanghai during 2012–2014. The study was approved by the Ethical Review Committee of the Shanghai Municipal Center for Disease Control and Prevention (SCDC), China (approval number: SCDC2016-08).

Briefly, daily average PM<sub>2.5</sub> concentrations between 1 January 2012 and 31 December 2014 were obtained from one fixed air quality monitor of the Shanghai Meteorological Bureau and one monitor from the U.S. Consulate General in Shanghai, China. Daily mortality data during the corresponding time period for all the permanent residents in Shanghai were obtained from the Causes of Death Registry of the SCDC. The causes of death were coded according to the International Classification of Diseases Codes, version 10 (ICD-10). Deaths for respiratory diseases (ICD-10 codes J00–J99) were retrieved. Citywide daily meteorological data including temperature, relative humidity, barometric pressure, wind speed, precipitation, and sunshine time were retrieved from the Shanghai Meteorological Bureau.

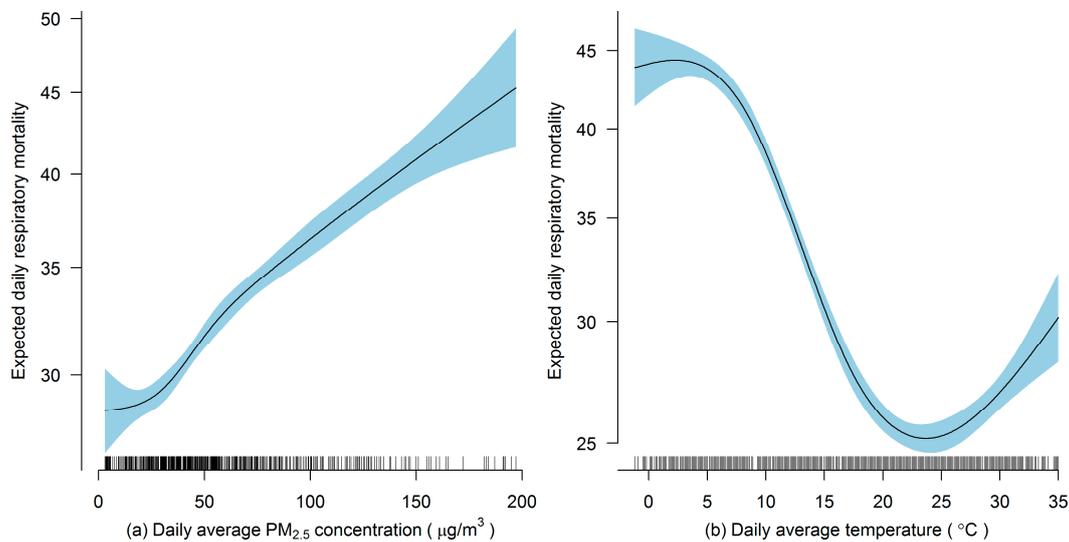
The distribution of the observed daily respiratory mortality and the theoretical quasi Poisson distribution with the same mean and an overdispersion index = 1.3 are shown in Figure 1. The observed data showed a few deviations from the theoretical distribution. Therefore, a Poisson regression model is suitable for our data.



**Figure 1.** Distribution of observed daily respiratory mortality (the red curve) and theoretical distribution quasi-Poisson distribution (the black vertical lines, mean = 32, overdispersion index = 1.3).

To illustrate the relationship between respiratory mortality and PM<sub>2.5</sub> pollution and meteorological variables, we used the daily respiratory mortality and the corresponding average PM<sub>2.5</sub> concentration and the average temperature in lag 1 as an example (Figure 2). In general, daily respiratory mortality was positively related to PM<sub>2.5</sub> concentration and an approximately linear relationship was found. A similar relationship was found for PM<sub>2.5</sub> concentrations at other lags. However, an apparent reversed sigmoid relationship was found between daily respiratory mortality and daily average temperature.

A nonlinear relationship was also found for other meteorological variables. Therefore, a GAM with linear components for PM<sub>2.5</sub> concentrations and nonlinear components for weather conditions and time trend is suitable for our study.



**Figure 2.** Smoothing plots of (a) daily respiratory mortality against daily average PM<sub>2.5</sub> concentration and (b) daily average temperature.

### 2.2. Fully Bayesian Generalized Additive Model

A fully Bayesian approach for modeling and inference within GAM requires prior assumption for unknown function  $S(t)$ . Several alternatives have been recently proposed to specify smoothness prior for continuous covariates or time trends, such as random walk priors or more generally autoregressive priors [18,24], Bayesian P(enalized)-splines [25], and Bayesian smoothing splines [26]. Bayesian P-splines were used in our study because it has the advantage of allowing for simultaneous estimation of smooth functions and smoothing parameters. Moreover, it can easily be extended to more complex formulations [25]. The method assumes that an unknown smooth function  $S_j(x_j)$  of a covariate  $x_j$  can be approximated by a polynomial spline of degree  $\nu$  defined on a set of equally spaced knots  $x_j^{min} = \zeta_0 < \zeta_1 < \dots < \zeta_{d-1} < \zeta_d = x_j^{max}$  within the domain of  $x_j$ . Such a spline can be written in terms of a linear combination of  $M_j(= d + \nu)$  B-spline basis functions  $B_m$ , i.e.,

$$S_j(x_j) = \sum_{m=1}^{M_j} \beta_{j,m} B_m(x_j) \tag{2}$$

and  $\beta_j = (\beta_{j,1}, \dots, \beta_{j,M_j})'$  corresponds to the vector of unknown regression coefficients. The functions  $B_m$  are only positive with an area spanned by  $\nu + 2$  knots, which is essential for the construction of the smoothness penalty for P-splines. The estimation of  $S_j(x_j)$  is thus reduced to the estimation of the vector of unknown regression coefficients  $\beta_j$  from the data. To choose the number of knots, a moderately large number of equally spaced knots (around 10 knots per time interval) as suggested by Eilers and Marx [27] is flexible enough to capture the variability of the data. In the Bayesian approach, penalized splines are introduced by replacing the different penalties with their stochastic analogues, i.e., the first or second order random walk priors for the regression coefficients. First order random walk prior for equidistant knots is given by [28]:

$$\beta_{j,m} = \beta_{j,m-1} + u_{j,m}, \quad m = 2, \dots, M_j \tag{3}$$

and a second order random walk by:

$$\beta_{j,m} = 2\beta_{j,m-1} - \beta_{j,m-2} + u_{j,m}, \quad m = 3, \dots, M_j \quad (4)$$

with Gaussian errors  $u_{j,m} \sim N(0, \tau_j^2)$  and diffuse priors  $p(\beta_{j,1}) \propto \text{const}$ , or  $p(\beta_{j,1})$  and  $p(\beta_{j,2}) \propto \text{const}$ , for initial values, respectively.

When we define the unknown function evaluation  $S_j$  as the matrix product of a design matrix  $\psi_j$  and a vector of unknown parameters  $\beta_j$ , i.e.,

$$S_j = \psi_j \beta_j, \quad j = 1, \dots, p \quad (5)$$

then we obtain the predictor in Equation (1) as

$$\log(E(Y_t)) = \beta_0 + \beta_l X_{t-l} + \sum_{j=1}^p \psi_j \beta_j + \phi \cdot \text{DOW}. \quad (6)$$

Depending on the above parameterization of the model, the posterior for fully Bayesian inference is given by:

$$p(\beta_0, \beta_l, \beta_1, \dots, \beta_p, \tau_1^2, \dots, \tau_p^2, \phi | y) \propto L(y, \beta_0, \beta_l, \beta_1, \dots, \beta_p, \phi) \prod_{j=1}^p (p(\beta_j | \tau_j^2) p(\tau_j^2)) \quad (7)$$

where  $L$  denotes the likelihood which is the product of individual likelihood contributions.

In the fully Bayesian approach, parameter estimates are obtained by drawing random samples from the posterior Equation (7) via Markov Chain Monte Carlo (MCMC) simulation techniques. More details about the fully Bayesian inference can be found in Fahrmeir and Lang [18] and Brezger and Lang [29].

### 2.3. Estimation of True Parameters

Although we focused on the effect of PM<sub>2.5</sub> concentrations in day lag 1 in the study, to exclude the effects from other lag days, the 'true' parameters were estimated using the frequentist distributed lag GAM instead of the basic model aforementioned. The distributed lag model associates health outcomes in a given day with PM<sub>2.5</sub> concentrations in several earlier days by replacing  $\beta_l X_{t-l}$  in (1) with:

$$\sum_{l=1}^L \beta_l X_{t-l} \text{ or } \theta \sum_{l=1}^L \eta_l X_{t-l}, \quad \sum_{l=1}^L \eta_l = 1 \quad (8)$$

where  $\theta$  measures the cumulative effect of PM<sub>2.5</sub> during the days, and  $\eta_l$  measures the contribution of the lagged exposure  $X_{t-l}$  to  $\theta$  [30,31]. To reduce the computational work, we introduced only one smoothness term i.e., seasonal trend in the model. The nonlinear effects from meteorological variables were examined using a categorical variable for six synoptic weather types (SWTs) [32]. The final model may be expressed as:

$$\log(E(Y_t)) = \beta_0 + \beta_1 \cdot \text{lag}1_t + \beta_2 \cdot \text{lag}2_t + \dots + \beta_7 \cdot \text{lag}7_t + S(t) + \phi_W \cdot W_t + \phi \cdot \text{DOW}_t \quad (9)$$

where  $\beta_l$  ( $l = 1, \dots, 7$ ) approximate to the percentage increase in deaths associated with the corresponding explanatory variables, i.e.,  $\text{lag}1_t - \text{lag}7_t$ , the daily PM<sub>2.5</sub> concentrations of lag 1–7 days. Frequentist  $S(t)$  was realized using cubic B-splines with in total 18 knots (6 knots per year).  $W_t$  denotes the vector of the six SWTs, i.e., hot dry, warm humid, cold dry, moderately dry, moderately humid, and cold humid weather types. The SWTs are defined using 15 routinely monitored meteorological variables through principal component analysis and K-means clustering method. The details of the methods were published elsewhere [33].

In total, 35,135 respiratory deaths occurred during the study period between 1 January 2012 and 31 December 2014 in Shanghai. Mean and median daily deaths were 32 and 30, respectively.

The descriptive statistics of daily respiratory deaths, ambient PM<sub>2.5</sub> concentrations, and major meteorological variables are shown in Table 1.

The parametric coefficients of the covariates derived from the frequentist GAM are shown in Table 2.

**Table 1.** Descriptive statistics of daily respiratory deaths, ambient PM<sub>2.5</sub> concentrations, and major meteorological variables.

Variable	<i>n</i>	Mean	SD	Min	<i>P</i> <sub>25</sub>	<i>P</i> <sub>50</sub>	<i>P</i> <sub>75</sub>	Max
Respiratory deaths	1096	32	11	12	24	30	38	82
log(Respiratory deaths)	1096	3.41	0.34	2.48	3.18	3.40	3.64	4.41
PM <sub>2.5</sub> (µg/m <sup>3</sup> )	1091	55.0	38.6	3.0	29.4	45.5	68.7	447.5
Temperature (°C)	1096	17.2	9.0	−1.2	8.8	18.2	24.3	35.0
Barometric Pressure (kPa)	1096	101.6	0.9	99.5	100.8	101.6	102.3	103.8
Relative Humidity (%)	1096	70.3	12.6	30	62	72	80	98
Wind speed (m/s)	1096	2.8	1.0	0.6	2.1	2.7	3.4	8.6
Precipitation (mm)	1096	3.3	10.4	0	0	0	1.1	195.3
Sunshine (h)	1096	4.7	4.0	0	0	4.8	8.2	12.9

SD, standard deviation; *P*<sub>*x*</sub>, the *x*th percentile.

**Table 2.** Parametric coefficients of the covariates in Equation (9).

Coefficient	Estimate	Standard Error	<i>z</i> Value	<i>Pr</i> (>  <i>z</i>  )
$\beta_0$	3.9666423	0.1214331	32.665	2e-16
$\beta_1$	0.0049014	0.0016902	2.900	0.00373
$\beta_2$	−0.0020338	0.0018396	−1.106	0.26892
$\beta_3$	0.0024222	0.0018156	1.334	0.18219
$\beta_4$	−0.0000745	0.0018219	−0.041	0.96738
$\beta_5$	0.0004227	0.0018279	0.231	0.81712
$\beta_6$	0.0006673	0.0018274	0.365	0.71500
$\beta_7$	−0.0004166	0.0018230	−0.229	0.81922
$\phi_{W1}$	Reference			
$\phi_{W2}$	−0.0612972	0.0220647	−2.778	0.00547
$\phi_{W3}$	−0.0780534	0.0320905	−2.432	0.01500
$\phi_{W4}$	−0.0494911	0.0267129	−1.853	0.06392
$\phi_{W5}$	−0.0085775	0.0274555	−0.312	0.75473
$\phi_{W6}$	−0.0603464	0.0319503	−1.889	0.05892
$\phi_0$	Reference			
$\phi_1$	−0.0094768	0.0202340	−0.468	0.63953
$\phi_2$	0.0010382	0.0201531	0.052	0.95891
$\phi_3$	−0.0140684	0.0202318	−0.695	0.48683
$\phi_4$	−0.0144893	0.0203167	−0.713	0.47574
$\phi_5$	−0.0104156	0.0202757	−0.514	0.60746
$\phi_6$	−0.0186630	0.0203238	−0.918	0.35847

In our study, we focused on the acute effect of PM<sub>2.5</sub> concentrations on respiratory mortality, i.e.,  $\beta_1$  of lag 1 day in Equation (9). It is 0.0049 and statistically significant (corresponding to 5.37 excessive deaths per day or 1960 excessive deaths per year). The standard error (SE) of  $\beta_1$  is 0.0017. The unit of PM<sub>2.5</sub> concentrations is 10 µg/m<sup>3</sup> in the analysis, therefore  $\beta_1$  can be explained as that per 10 µg/m<sup>3</sup> increase in PM<sub>2.5</sub> concentration of lag 1 day is associated with approximately 0.49% increase in daily respiratory deaths. The result is consistent with the findings from two meta-analyses [34,35] and a large observational study performed in China [36], where the corresponding percent increase in respiratory mortality ranged from 0.29% to 0.75%.

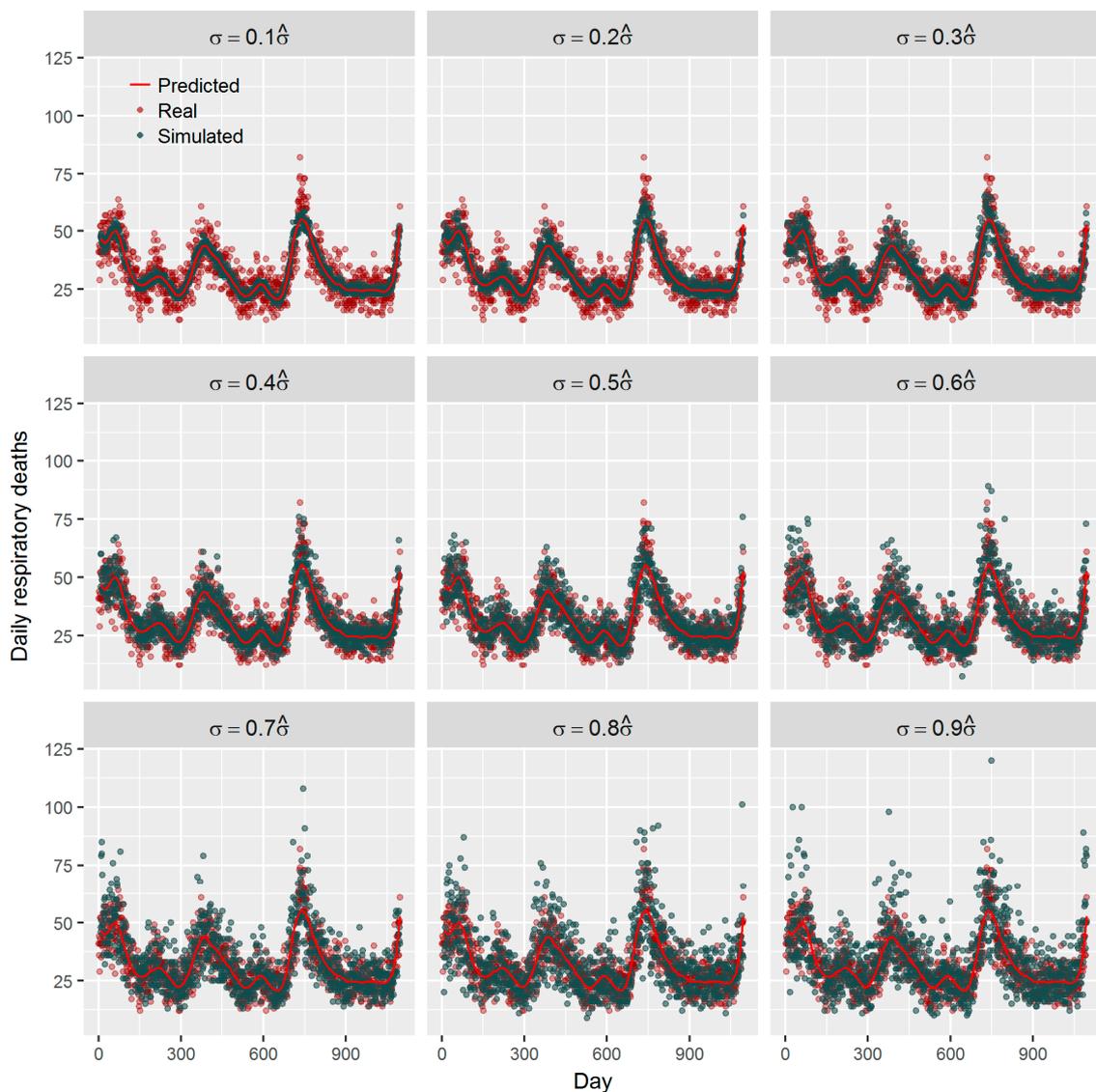
#### 2.4. Simulation

To assess the estimation of the fully Bayesian GAM approach compared with the realistic situations, we conducted a simulation study. In our simulation, the estimates in Table 2 derived

from the real world data were used as ‘true’ parameters. We used the predicted daily deaths  $\hat{Y}_t$  based on Equation (9) as the mean daily deaths, then simulated daily death  $Y'_t$  by multiplying  $\hat{Y}_t$  by a random error  $e^\varepsilon$ :

$$Y'_t = \hat{Y}_t \cdot e^\varepsilon, \varepsilon \sim N(0, \sigma^2) \tag{10}$$

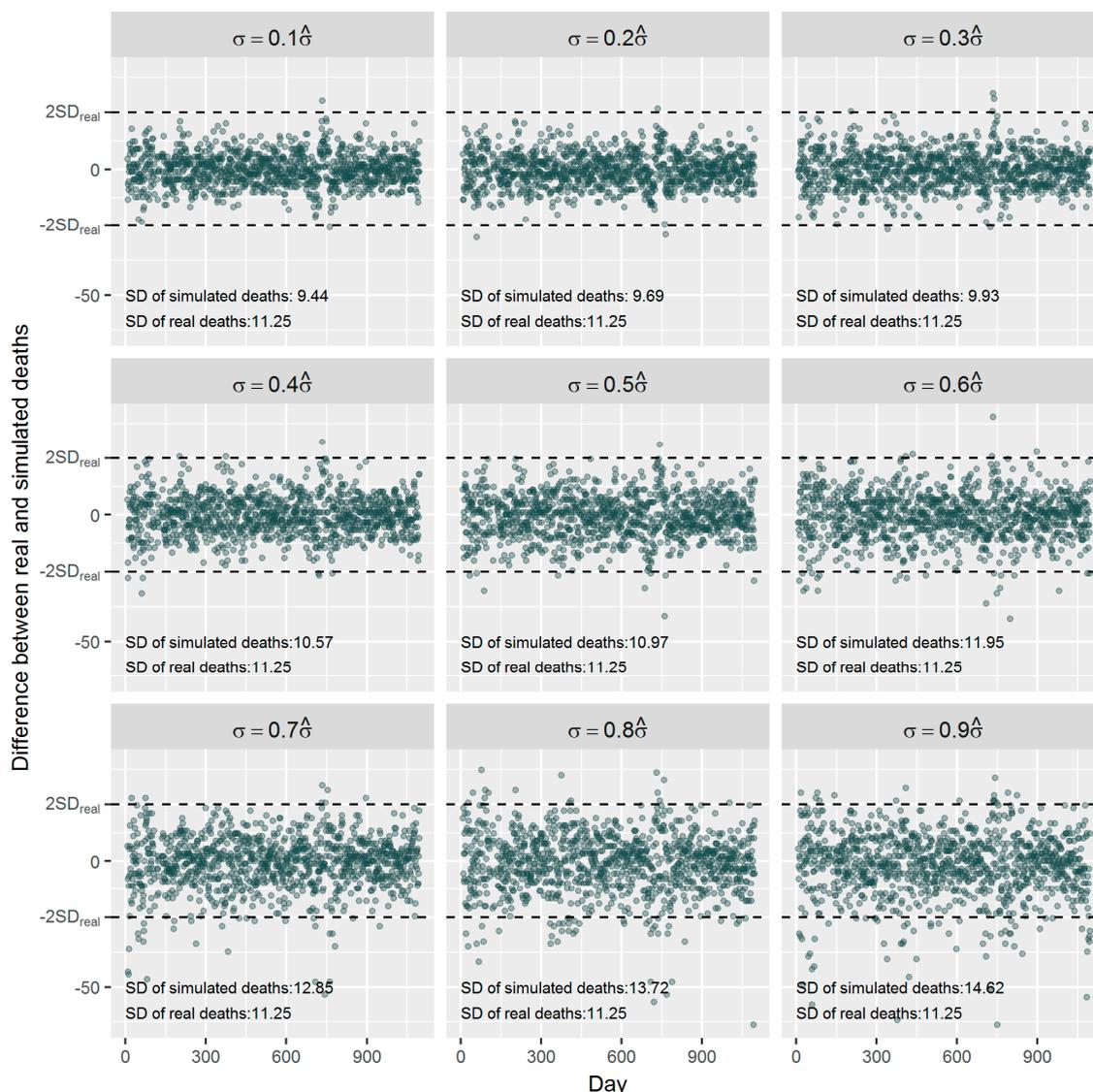
where  $\varepsilon$  follows a distribution from exponential family and we applied normal distribution here. The simulation framework ensures that the same concavity will exist between the simulated mortalities and the observed ones. By changing  $\sigma$  we may introduce different ‘noise’ in mortality to simulate the effects of unobserved confounders. In our simulation, the changing of the  $\sigma$  was achieved by multiplying  $\hat{\sigma}$ , the SD of logarithmic daily deaths, by a factor  $\gamma$ , i.e.,  $\sigma = \gamma\hat{\sigma}$ . By selecting different random seeds, we may generate different time-series using random number generator in any statistical software. Figure 3 shows nine simulated time-series of daily respiratory deaths for  $\gamma = 0.1, 0.2, \dots, 0.9$ . We can see when  $\gamma$  is equal to 0.4 or 0.5 the simulated data are most approximate to the real world data.



**Figure 3.** Examples of simulated time-series of daily respiratory deaths with different random noises.

Figure 4 shows the difference between the simulated data and the real world data. Almost all the differences were within the range of  $\pm 2SD$  of the real data when  $\gamma$  is between 0.1 and 0.9. The larger  $\gamma$  the larger difference presents, which means more ‘noise’ being included in the simulated data. When

$\gamma$  is equal to 0.5 the difference between SDs of the simulated data and the real world data (10.97 vs. 11.25) is the smallest. Therefore, we used the simulated data with  $\gamma \geq 0.5$  in the following analyses to include relatively larger uncertainty for our estimation.

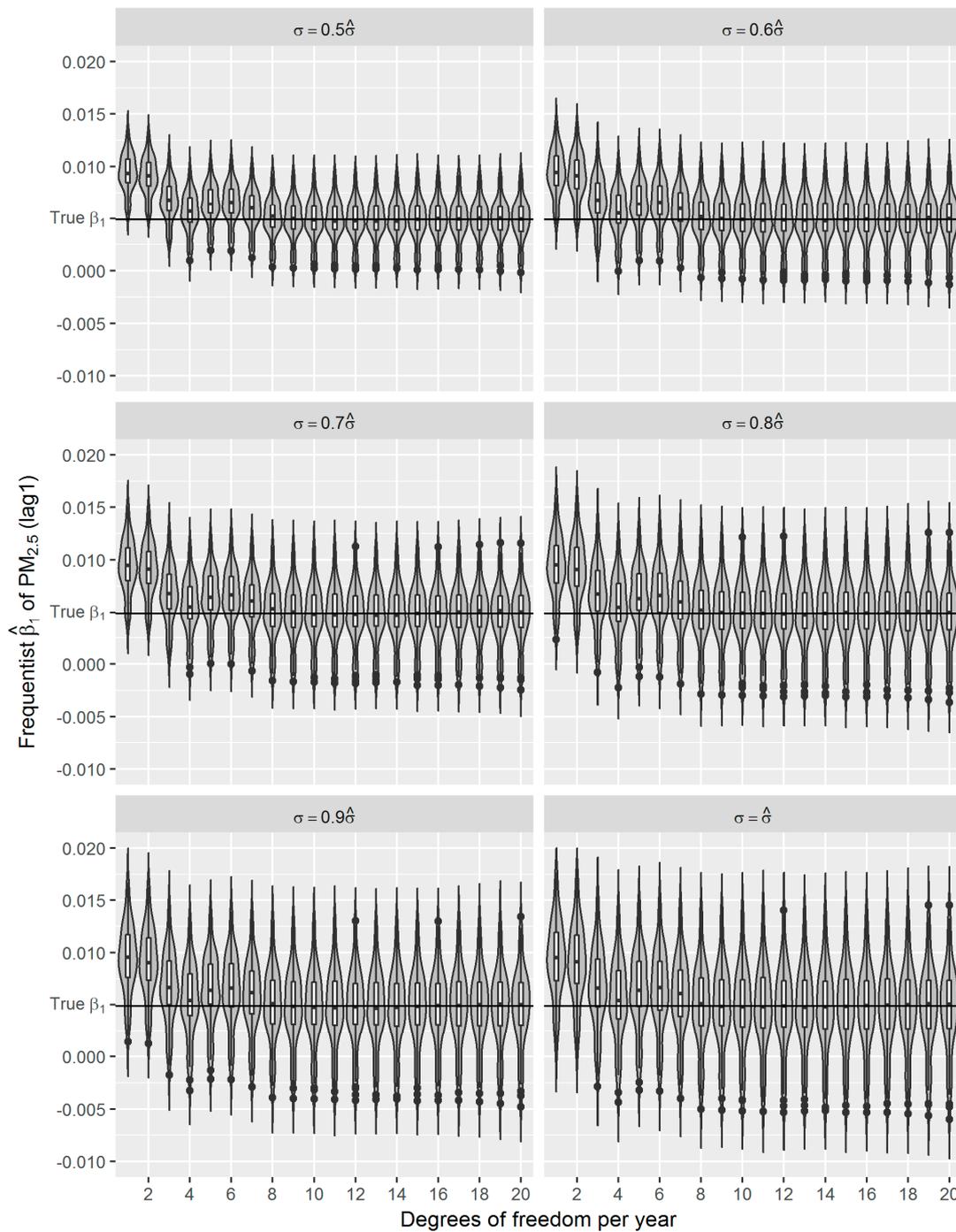


**Figure 4.** Examples of the difference between the real and the simulated daily respiratory deaths with different random noises.

### 3. Results

#### 3.1. Comparison of the Results from Frequentist GAM and Bayesian GAM

First, we set  $\gamma = 0.5, 0.6, \dots, 1.0$  to generate six sets of simulated respiratory mortality data, where each set included 2000 time-series. Then we run the frequentist GAMs using simulated daily mortality as the dependent variable. We set the degrees of freedom ( $df$ ) for  $S(t)$  from 1 to 20 per year in our models. For each  $df$ , we ran the frequentist GAM using 100 simulated time-series. In total, 12,000  $\hat{\beta}_1$ s (from 6  $\gamma$ s  $\times$  20  $df$ s  $\times$  100 time-series) were derived. Distributions of the estimated  $\hat{\beta}_1$ s against the true  $\beta_1$  are shown using a violin plot combined with a box plot in Figure 5.



**Figure 5.** Distributions of frequentist  $\hat{\beta}_1$ s for simulated data with different random noises ( $\sigma = 0.5 - 1.0\hat{\sigma}$ ) and degrees of freedom (1–20 per year) for  $S(t)$ ; the true  $\beta_1 = 0.0049$ .

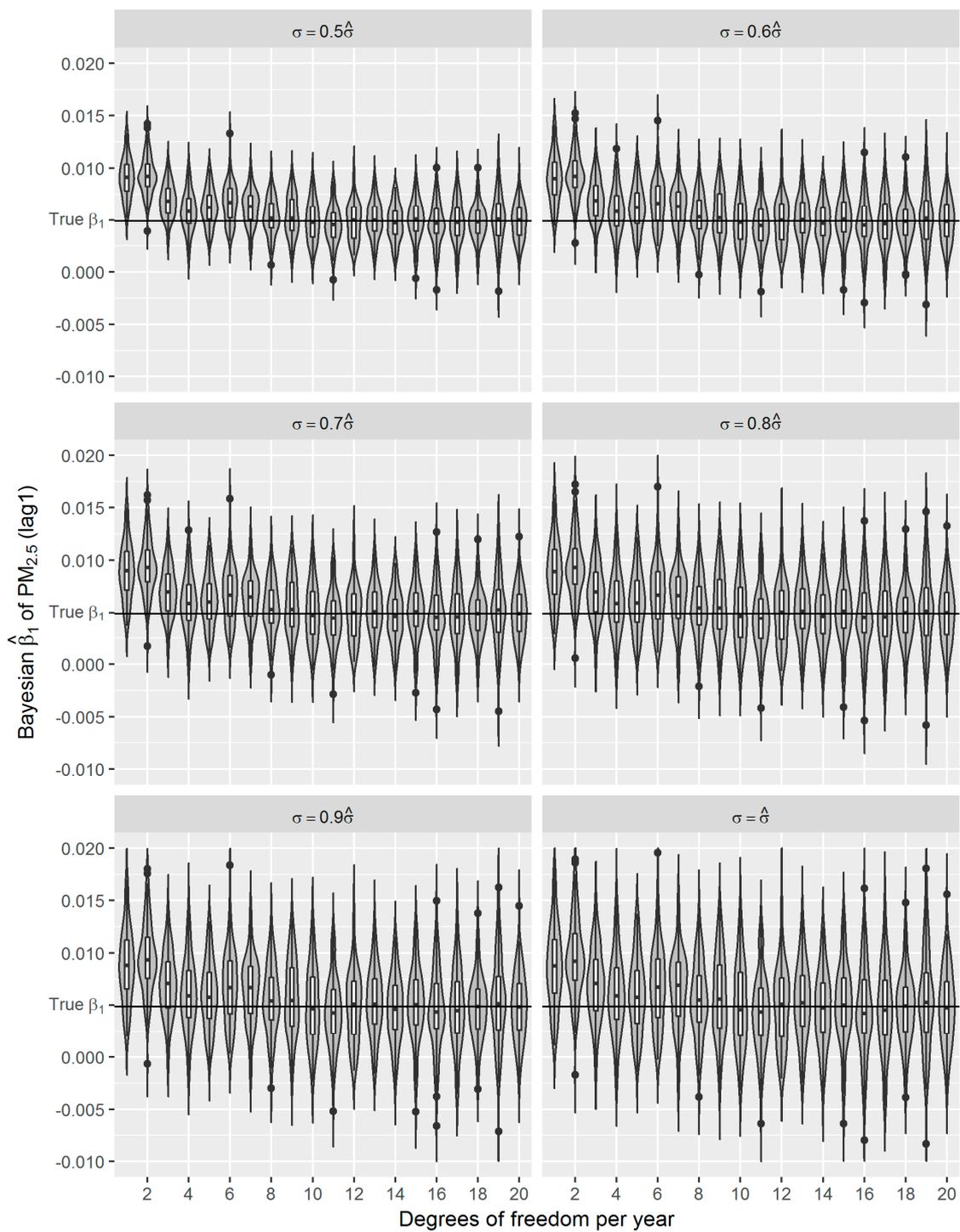
In general, although the variance of  $\hat{\beta}_1$ s tends to increase as the  $df$  for  $S(t)$  is increased, the decrease in bias is far more dramatic, i.e., the greater  $df$  the closer mean of  $\hat{\beta}_1$ s approach to the true  $\beta_1$ . The bias in the estimates is only serious for  $df \leq 4$  per year. With sufficient  $df$  to represent the smoothness of the nonparametric nonlinear trend leads to an asymptotically unbiased estimate of  $\hat{\beta}_1$  [37]. The apparent decrease in the bias of  $\hat{\beta}_1$ s with increasing  $df$  is explained by Daniels et al. [38], Rice [39] and Speckman [40]. As we expected, the increased noise in the simulated data results in a larger variance of  $\hat{\beta}_1$ s (Figure 5 and Table 3), which is approximate to  $2\gamma \times SE(\beta_1)$  or  $2\gamma \times 0.0018$ .

**Table 3.** Mean (SD) of estimated  $\hat{\beta}_1$ s derived from frequentist GAMs.

<i>df</i>	$\sigma = 0.5\hat{\sigma}$	$\sigma = 0.6\hat{\sigma}$	$\sigma = 0.7\hat{\sigma}$	$\sigma = 0.8\hat{\sigma}$	$\sigma = 0.9\hat{\sigma}$	$\sigma = \hat{\sigma}$
1	0.0095 (0.0018)	0.0095 (0.0022)	0.0094 (0.0026)	0.0094 (0.0029)	0.0094 (0.0033)	0.0094 (0.0037)
2	0.0091 (0.0018)	0.0091 (0.0022)	0.0091 (0.0025)	0.0091 (0.0029)	0.0091 (0.0033)	0.0090 (0.0037)
3	0.0068 (0.0019)	0.0068 (0.0022)	0.0068 (0.0026)	0.0067 (0.0030)	0.0067 (0.0034)	0.0067 (0.0038)
4	0.0056 (0.0019)	0.0056 (0.0023)	0.0056 (0.0026)	0.0056 (0.0030)	0.0056 (0.0034)	0.0055 (0.0038)
5	0.0064 (0.0019)	0.0064 (0.0022)	0.0064 (0.0026)	0.0064 (0.0030)	0.0064 (0.0034)	0.0064 (0.0038)
6	0.0065 (0.0019)	0.0065 (0.0023)	0.0065 (0.0027)	0.0065 (0.0031)	0.0065 (0.0034)	0.0064 (0.0038)
7	0.0059 (0.0019)	0.0059 (0.0023)	0.0058 (0.0027)	0.0058 (0.0031)	0.0058 (0.0035)	0.0058 (0.0039)
8	0.0051 (0.0020)	0.0051 (0.0023)	0.0050 (0.0027)	0.0050 (0.0032)	0.0050 (0.0035)	0.0050 (0.0040)
9	0.0049 (0.0020)	0.0049 (0.0024)	0.0049 (0.0028)	0.0049 (0.0032)	0.0049 (0.0036)	0.0048 (0.0040)
10	0.0049 (0.0020)	0.0049 (0.0024)	0.0048 (0.0028)	0.0048 (0.0032)	0.0048 (0.0036)	0.0048 (0.0040)
11	0.0049 (0.0020)	0.0048 (0.0024)	0.0048 (0.0028)	0.0048 (0.0032)	0.0048 (0.0036)	0.0048 (0.0040)
12	0.0048 (0.0020)	0.0048 (0.0024)	0.0048 (0.0028)	0.0048 (0.0032)	0.0047 (0.0036)	0.0047 (0.0041)
13	0.0048 (0.0020)	0.0048 (0.0024)	0.0048 (0.0028)	0.0048 (0.0032)	0.0047 (0.0036)	0.0047 (0.0040)
14	0.0049 (0.0020)	0.0048 (0.0024)	0.0048 (0.0028)	0.0048 (0.0032)	0.0048 (0.0036)	0.0047 (0.0041)
15	0.0049 (0.0020)	0.0049 (0.0024)	0.0048 (0.0028)	0.0048 (0.0032)	0.0048 (0.0036)	0.0048 (0.0041)
16	0.0049 (0.0020)	0.0049 (0.0024)	0.0048 (0.0028)	0.0048 (0.0032)	0.0048 (0.0036)	0.0048 (0.0041)
17	0.0049 (0.0020)	0.0049 (0.0024)	0.0048 (0.0028)	0.0048 (0.0032)	0.0048 (0.0036)	0.0048 (0.0041)
18	0.0049 (0.0020)	0.0048 (0.0024)	0.0048 (0.0028)	0.0048 (0.0033)	0.0048 (0.0037)	0.0047 (0.0041)
19	0.0049 (0.0020)	0.0048 (0.0024)	0.0048 (0.0028)	0.0048 (0.0033)	0.0048 (0.0037)	0.0047 (0.0041)
20	0.0049 (0.0020)	0.0048 (0.0025)	0.0048 (0.0029)	0.0048 (0.0033)	0.0048 (0.0037)	0.0047 (0.0042)

SD, standard deviation; GAM, generalized additive model.

The Bayesian GAM analyses were conducted in software SAS 9.4 and R version 3.41 with MCMCpack package [41]. We used 1000 burn-in iterations and 2000 iterations after burn-in for Markov chains. Although 10,000 iterations after burn-in are usually suggested, we did not find noticeable differences in the estimates between 10,000 iterations and 2000 iterations. The latter reduced however the computation time significantly. For comparison purpose, we also used *dfs* ranged from 1 to 20 per year for the Bayesian P-splines to control for smoothness. By default we used non-informative uniform priors for the coefficients and the derived estimates are quite similar to those derived from frequentist GAMs. Although the Bayesian  $\hat{\beta}_1$ s appear more fluctuated around the true  $\beta_1$  (Figure 6), their SDs are comparable to those of their frequentist counterparts (Table 4).



**Figure 6.** Distributions of Bayesian  $\hat{\beta}_1$ s for simulated data with different noises ( $\sigma = 0.5 - 1.0\hat{\delta}$ ) and degrees of freedom (1–20 per year) for  $S(t)$ ; the true  $\beta_1 = 0.0049$ .

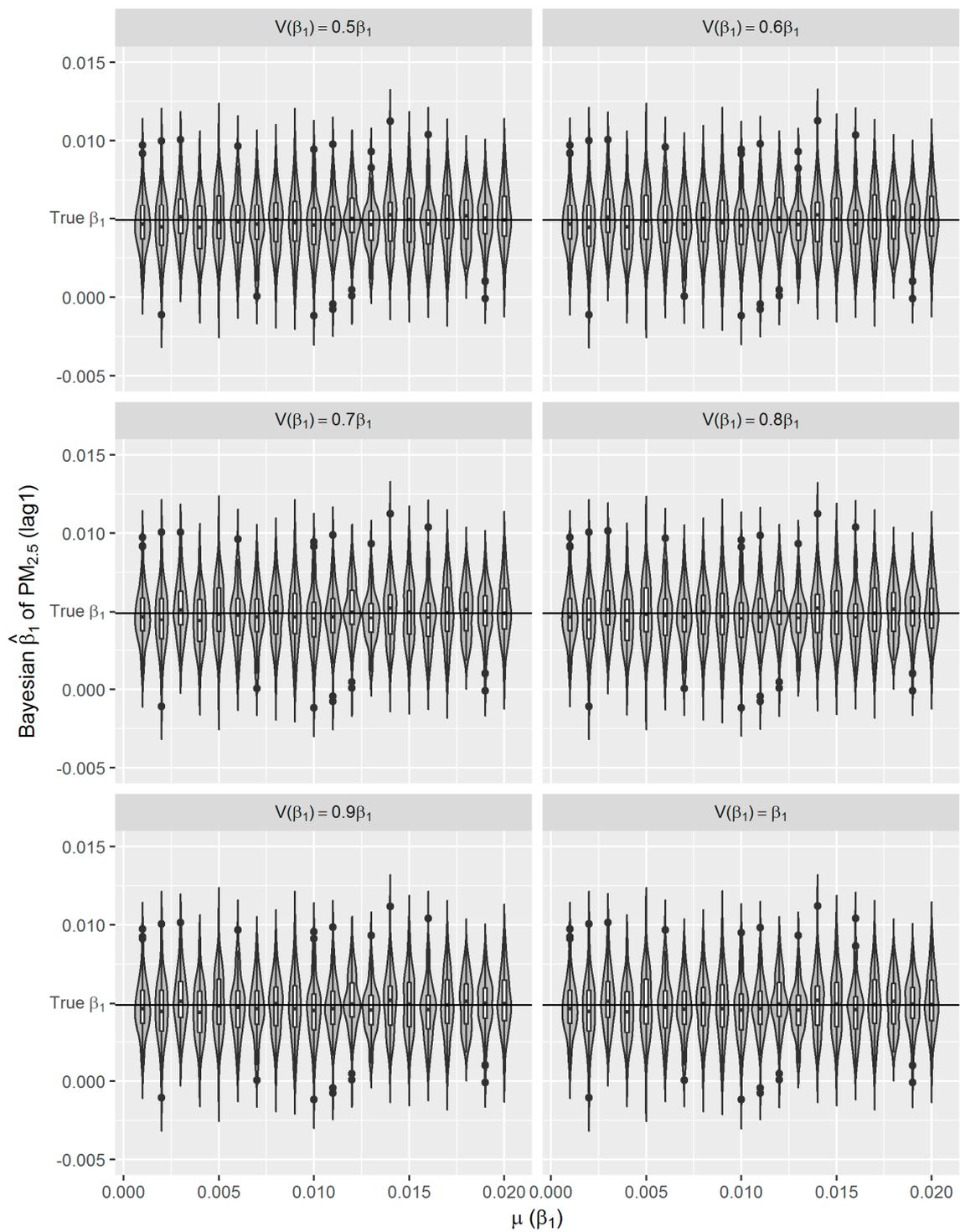
**Table 4.** Mean (SD) of estimated  $\hat{\beta}_1$ s derived from Bayesian GAMs.

<i>df</i>	$\sigma = 0.5\hat{\sigma}$	$\sigma = 0.6\hat{\sigma}$	$\sigma = 0.7\hat{\sigma}$	$\sigma = 0.8\hat{\sigma}$	$\sigma = 0.9\hat{\sigma}$	$\sigma = \hat{\sigma}$
1	0.0091 (0.0019)	0.0090 (0.0023)	0.0090 (0.0027)	0.0089 (0.0031)	0.0088 (0.0035)	0.0087 (0.0039)
2	0.0093 (0.0019)	0.0094 (0.0022)	0.0094 (0.0026)	0.0094 (0.0030)	0.0095 (0.0034)	0.0095 (0.0038)
3	0.0069 (0.0016)	0.0069 (0.0020)	0.0069 (0.0023)	0.0070 (0.0027)	0.0070 (0.0031)	0.0070 (0.0034)
4	0.0059 (0.0019)	0.0059 (0.0023)	0.0060 (0.0026)	0.0060 (0.0030)	0.0060 (0.0034)	0.0061 (0.0038)
5	0.0062 (0.0018)	0.0062 (0.0021)	0.0062 (0.0025)	0.0061 (0.0028)	0.0061 (0.0032)	0.0060 (0.0036)
6	0.0066 (0.0019)	0.0067 (0.0023)	0.0067 (0.0027)	0.0067 (0.0031)	0.0068 (0.0035)	0.0068 (0.0039)
7	0.0062 (0.0017)	0.0062 (0.0021)	0.0063 (0.0024)	0.0064 (0.0028)	0.0064 (0.0032)	0.0065 (0.0036)
8	0.0053 (0.0018)	0.0054 (0.0021)	0.0054 (0.0025)	0.0055 (0.0029)	0.0056 (0.0032)	0.0056 (0.0036)
9	0.0054 (0.0020)	0.0055 (0.0024)	0.0056 (0.0028)	0.0057 (0.0031)	0.0058 (0.0035)	0.0058 (0.0039)
10	0.0049 (0.0019)	0.0049 (0.0023)	0.0049 (0.0028)	0.0050 (0.0032)	0.0050 (0.0036)	0.0050 (0.0040)
11	0.0046 (0.0018)	0.0045 (0.0022)	0.0045 (0.0026)	0.0044 (0.0029)	0.0043 (0.0033)	0.0043 (0.0037)
12	0.0049 (0.0018)	0.0049 (0.0022)	0.0049 (0.0026)	0.0049 (0.0030)	0.0049 (0.0034)	0.0049 (0.0038)
13	0.0051 (0.0018)	0.0052 (0.0022)	0.0052 (0.0025)	0.0053 (0.0029)	0.0053 (0.0032)	0.0054 (0.0036)
14	0.0047 (0.0017)	0.0047 (0.0021)	0.0047 (0.0024)	0.0047 (0.0028)	0.0046 (0.0032)	0.0046 (0.0036)
15	0.0050 (0.0019)	0.0050 (0.0023)	0.0050 (0.0027)	0.0051 (0.0031)	0.0051 (0.0035)	0.0051 (0.0039)
16	0.0047 (0.0021)	0.0047 (0.0025)	0.0046 (0.0030)	0.0046 (0.0034)	0.0045 (0.0039)	0.0045 (0.0044)
17	0.0048 (0.0020)	0.0048 (0.0025)	0.0048 (0.0029)	0.0048 (0.0033)	0.0048 (0.0037)	0.0047 (0.0042)
18	0.0048 (0.0017)	0.0048 (0.0020)	0.0048 (0.0024)	0.0048 (0.0027)	0.0047 (0.0031)	0.0048 (0.0034)
19	0.0052 (0.0023)	0.0052 (0.0028)	0.0053 (0.0032)	0.0054 (0.0037)	0.0054 (0.0042)	0.0055 (0.0047)
20	0.0049 (0.0018)	0.0049 (0.0021)	0.0049 (0.0025)	0.0049 (0.0028)	0.0049 (0.0032)	0.0049 (0.0036)

SD, standard deviation; GAM, generalized additive model.

### 3.2. Sensitivity of Bayesian GAM to Choice of Prior Mean and Variance

One of the most important (however also controversial) features of Bayesian method is that they can integrate prior knowledge and observed data in their inference. The posterior is a compromise between prior and likelihood. In the second simulation, we investigated the impact of informative priors rather than non-informative uniform prior on the posterior  $\hat{\beta}_1$ . We simulated the time-series of daily respiratory mortality with a fixed  $\sigma = 0.5\hat{\sigma}$  and true  $\beta_1 = 0.0049$ . In total, 12,000 time-series were generated. For Bayesian GAM analyses, we used a normal prior for  $\beta_1$ , and set the varied prior mean  $\mu(\beta_1)$  ranging from 0.001 to 0.020 by 0.001, and varied prior variance  $V(\beta_1)$  equal to  $\gamma\beta_1$ , where  $\gamma = 0.5, 0.6, \dots, 1.0$ . For each combination of  $\mu(\beta_1)$  and  $V(\beta_1)$ , we did 100 Bayesian analyses. To get unbiased estimates with fewer computation task, we set the *df* for splines to eight per year. The distribution of Bayesian estimates ( $\hat{\beta}_1$ s) are shown in Figure 7. The mean of  $\hat{\beta}_1$ s is fluctuated but closely around the true  $\beta_1$  for different  $\mu(\beta_1)$ . These is no noticeable difference among the means of  $\hat{\beta}_1$ s derived from different  $V(\beta_1)$  (Figure 7 and Table 5). The SD of  $\hat{\beta}_1$ s is not sensitive to the  $V(\beta_1)$  (Table 5).



**Figure 7.** Distributions of Bayesian  $\hat{\beta}_1$ s from simulated data with  $\sigma = 0.5\hat{\sigma}$ , the true  $\beta_1 = 0.0049$ ; in Bayesian GAM analyses  $df = 8$  for  $S(t)$ ,  $V(\beta_1)$  equal to  $\gamma\beta_1$ , where  $\gamma = 0.5, 0.6, \dots, 1.0$ , and  $\mu(\beta_1)$  ranging from 0.001 to 0.020 by 0.001.

**Table 5.** Mean (SD) of estimated  $\hat{\beta}_1$ s derived from Bayesian GAMs with different priors.

$\mu(\beta_1)$	$\text{Var}(\beta_1) = 0.5\beta_1$	$\text{Var}(\beta_1) = 0.6\beta_1$	$\text{Var}(\beta_1) = 0.7\beta_1$	$\text{Var}(\beta_1) = 0.8\beta_1$	$\text{Var}(\beta_1) = 0.9\beta_1$	$\text{Var}(\beta_1) = \beta_1$
0.001	0.0048 (0.0018)	0.0048 (0.0018)	0.0048 (0.0018)	0.0048 (0.0018)	0.0048 (0.0018)	0.0048 (0.0018)
0.002	0.0046 (0.0020)	0.0046 (0.0020)	0.0046 (0.0020)	0.0046 (0.0020)	0.0046 (0.0020)	0.0046 (0.0020)
0.003	0.0052 (0.0018)	0.0052 (0.0018)	0.0052 (0.0018)	0.0052 (0.0018)	0.0052 (0.0018)	0.0052 (0.0018)
0.004	0.0045 (0.0018)	0.0045 (0.0018)	0.0045 (0.0018)	0.0045 (0.0018)	0.0045 (0.0018)	0.0045 (0.0018)
0.005	0.0049 (0.0020)	0.0050 (0.0020)	0.0050 (0.0020)	0.0049 (0.0020)	0.0049 (0.0020)	0.0049 (0.0020)
0.006	0.0048 (0.0018)	0.0048 (0.0018)	0.0048 (0.0018)	0.0048 (0.0018)	0.0048 (0.0018)	0.0048 (0.0018)
0.007	0.0047 (0.0018)	0.0047 (0.0018)	0.0047 (0.0018)	0.0047 (0.0018)	0.0047 (0.0018)	0.0047 (0.0018)
0.008	0.0048 (0.0018)	0.0048 (0.0018)	0.0048 (0.0018)	0.0048 (0.0018)	0.0048 (0.0018)	0.0048 (0.0018)
0.009	0.0048 (0.0020)	0.0048 (0.0020)	0.0048 (0.0020)	0.0048 (0.0020)	0.0048 (0.0020)	0.0048 (0.0020)
0.010	0.0046 (0.0020)	0.0046 (0.0020)	0.0046 (0.0020)	0.0046 (0.0020)	0.0046 (0.0020)	0.0046 (0.0020)
0.011	0.0047 (0.0018)	0.0047 (0.0018)	0.0047 (0.0018)	0.0047 (0.0018)	0.0047 (0.0018)	0.0047 (0.0018)
0.012	0.0051 (0.0017)	0.0051 (0.0017)	0.0051 (0.0017)	0.0051 (0.0017)	0.0051 (0.0017)	0.0051 (0.0017)
0.013	0.0046 (0.0016)	0.0046 (0.0016)	0.0046 (0.0016)	0.0046 (0.0016)	0.0046 (0.0016)	0.0046 (0.0016)
0.014	0.0050 (0.0019)	0.0050 (0.0019)	0.0050 (0.0019)	0.0050 (0.0019)	0.0050 (0.0019)	0.0050 (0.0019)
0.015	0.0050 (0.0020)	0.0050 (0.0020)	0.0050 (0.0020)	0.0050 (0.0020)	0.0050 (0.0020)	0.0050 (0.0020)
0.016	0.0046 (0.0018)	0.0046 (0.0018)	0.0046 (0.0018)	0.0046 (0.0018)	0.0046 (0.0018)	0.0046 (0.0018)
0.017	0.0049 (0.0020)	0.0049 (0.0020)	0.0049 (0.0020)	0.0049 (0.0020)	0.0049 (0.0020)	0.0049 (0.0020)
0.018	0.0050 (0.0017)	0.0050 (0.0017)	0.0050 (0.0017)	0.0050 (0.0017)	0.0050 (0.0017)	0.0050 (0.0017)
0.019	0.0049 (0.0016)	0.0049 (0.0016)	0.0049 (0.0016)	0.0049 (0.0016)	0.0049 (0.0016)	0.0049 (0.0016)
0.020	0.0051 (0.0018)	0.0051 (0.0018)	0.0051 (0.0018)	0.0051 (0.0018)	0.0051 (0.0018)	0.0051 (0.0018)

SD, standard deviation; GAM, generalized additive model.

### 3.3. Sensitivity of Bayesian GAM to True Parameter

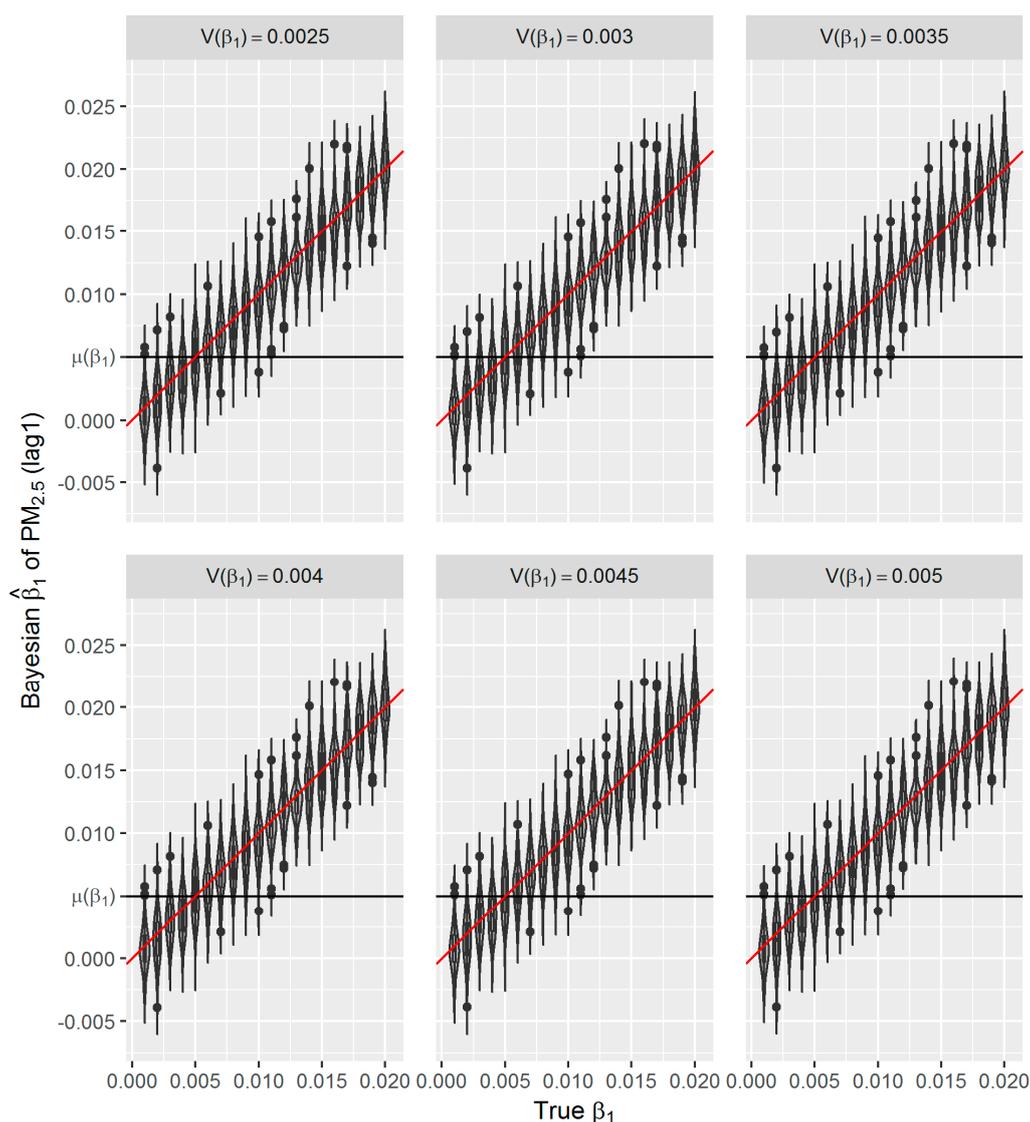
In our third simulation study, we artificially set the  $\sigma = 0.5\hat{\sigma}$  and ‘true’  $\beta_1$  ranged from 0.001 to 0.020 to generate 20 sets of simulated daily respiratory deaths, while kept the other coefficients in Equation (9) unchanged.

In Bayesian estimation, we used a normal prior for  $\beta_1$  with a fixed mean  $\mu(\beta_1) = 0.005$  but varied  $V(\beta_1) = 0.5, 0.6, \dots, 1.0$  times of  $\mu(\beta_1)$ , i.e., 0.0025, 0.003, 0.0035, 0.004, 0.0045 and 0.005. For each combination of  $\beta_1$  and  $V(\beta_1)$ , we generated 100 simulated time-series. Therefore, for each prior  $\beta_1$ , we did 600 Bayesian GAM analyses ( $100 \times 6$  variances). The estimates by the 20 true  $\beta_1$ s and 6 varied variances were shown in Figure 8 and Table 6.

**Table 6.** Mean (SD) of estimated  $\hat{\beta}_1$ s derived from Bayesian GAMs of different true  $\beta_1$ s.

True $\beta_1$	$\text{Var}(\beta_1) = 0.5\beta_1$	$\text{Var}(\beta_1) = 0.6\beta_1$	$\text{Var}(\beta_1) = 0.7\beta_1$	$\text{Var}(\beta_1) = 0.8\beta_1$	$\text{Var}(\beta_1) = 0.9\beta_1$	$\text{Var}(\beta_1) = \beta_1$
0.001	0.0008 (0.0018)	0.0008 (0.0018)	0.0008 (0.0018)	0.0008 (0.0018)	0.0008 (0.0018)	0.0008 (0.0018)
0.002	0.0016 (0.0020)	0.0016 (0.0020)	0.0016 (0.0020)	0.0016 (0.0020)	0.0016 (0.0020)	0.0016 (0.0020)
0.003	0.0032 (0.0018)	0.0032 (0.0018)	0.0032 (0.0018)	0.0032 (0.0018)	0.0032 (0.0018)	0.0032 (0.0018)
0.004	0.0035 (0.0018)	0.0035 (0.0018)	0.0035 (0.0018)	0.0035 (0.0018)	0.0035 (0.0018)	0.0035 (0.0018)
0.005	0.0049 (0.0020)	0.0050 (0.0020)	0.0050 (0.0020)	0.0049 (0.0020)	0.0049 (0.0020)	0.0049 (0.0020)
0.006	0.0058 (0.0018)	0.0058 (0.0018)	0.0058 (0.0018)	0.0058 (0.0018)	0.0058 (0.0018)	0.0058 (0.0018)
0.007	0.0067 (0.0018)	0.0067 (0.0018)	0.0067 (0.0018)	0.0067 (0.0018)	0.0067 (0.0018)	0.0067 (0.0018)
0.008	0.0078 (0.0018)	0.0078 (0.0018)	0.0078 (0.0018)	0.0078 (0.0018)	0.0078 (0.0018)	0.0078 (0.0018)
0.009	0.0088 (0.0020)	0.0088 (0.0020)	0.0088 (0.0020)	0.0088 (0.0020)	0.0088 (0.0020)	0.0088 (0.0020)
0.010	0.0096 (0.0020)	0.0096 (0.0020)	0.0096 (0.0020)	0.0096 (0.0020)	0.0096 (0.0020)	0.0096 (0.0020)
0.011	0.0107 (0.0018)	0.0107 (0.0018)	0.0107 (0.0018)	0.0107 (0.0018)	0.0107 (0.0018)	0.0107 (0.0018)
0.012	0.0121 (0.0017)	0.0121 (0.0017)	0.0121 (0.0017)	0.0121 (0.0017)	0.0121 (0.0017)	0.0121 (0.0017)
0.013	0.0125 (0.0016)	0.0125 (0.0016)	0.0126 (0.0016)	0.0126 (0.0016)	0.0126 (0.0016)	0.0126 (0.0016)
0.014	0.0140 (0.0019)	0.0140 (0.0019)	0.0140 (0.0019)	0.0140 (0.0019)	0.0140 (0.0019)	0.0140 (0.0019)
0.015	0.0150 (0.0020)	0.0150 (0.0020)	0.0150 (0.0020)	0.0150 (0.0020)	0.0150 (0.0020)	0.0150 (0.0020)
0.016	0.0156 (0.0019)	0.0156 (0.0019)	0.0156 (0.0019)	0.0156 (0.0019)	0.0156 (0.0019)	0.0156 (0.0019)
0.017	0.0170 (0.0020)	0.0170 (0.0020)	0.0170 (0.0020)	0.0170 (0.0020)	0.0170 (0.0020)	0.0170 (0.0020)
0.018	0.0180 (0.0017)	0.0180 (0.0017)	0.0180 (0.0017)	0.0180 (0.0017)	0.0180 (0.0017)	0.0180 (0.0017)
0.019	0.0189 (0.0017)	0.0189 (0.0017)	0.0189 (0.0017)	0.0189 (0.0017)	0.0189 (0.0017)	0.0190 (0.0017)
0.020	0.0201 (0.0019)	0.0201 (0.0019)	0.0202 (0.0019)	0.0202 (0.0019)	0.0202 (0.0019)	0.0202 (0.0019)

SD, standard deviation; GAM, generalized additive model.



**Figure 8.** Distributions of Bayesian  $\hat{\beta}_1$ s from simulated data with  $\sigma = 0.5\hat{\sigma}$ , true  $\beta_1 = 0.001$  to  $0.02$  by  $0.001$ ; in Bayesian GAM analyses  $df = 8$  for  $S(t)$ ,  $\mu(\beta_1) = 0.005$  and  $V(\beta_1) = 0.0025, 0.003, 0.0035, 0.004, 0.0045$  and  $0.005$ .

We can see that the mean of the estimated  $\hat{\beta}_1$ s is only sensitive to the underlying true  $\beta_1$  and is almost not affected by the prior  $\mu(\beta_1)$ . Due to the small coefficients, the difference between means and SDs of the estimated  $\hat{\beta}_1$ s can only be seen in the fifth or sixth decimal digit. The bivariate linear regression analysis between the true  $\beta_1$ s and the Bayesian  $\hat{\beta}_1$ s in Table 6 reveals that all the coefficients are almost equal to 1 and all the R are larger than 0.99, which indicates the high precision and accuracy of the Bayesian estimates.

#### 4. Discussion

In the presented study, we evaluated the performance of frequentist and fully Bayesian GAM approaches in a time-series study on the relationship between daily exposure to  $PM_{2.5}$  and respiratory mortality. According to our estimates, per  $10 \mu\text{g}/\text{m}^3$  increase in  $PM_{2.5}$  concentration of lag 1 day is associated with an approximately 0.49% increase in daily respiratory deaths in Shanghai between 2012 and 2014, which is consistent with the results from other studies conducted in China [9,36]. Using the estimated effect as true parameter, we compared the frequentist GAM and Bayesian GAM based on simulation. Both frequentist GAM and Bayesian GAM show the similar mean estimates

of the interested parameters. However, the estimates from frequentist GAM showed relatively less fluctuation (Figures 5 and 6), which to some extent reflects the over-confident inferences embedded in this method. Regarding the accuracy and precision of the estimates, both methods gave mean estimates close to the true parameter with comparable confidence intervals. It means that Bayesian GAM might be an ideal alternative to the conventional frequentist GAM. Our simulation study also indicated that when the underlying parameter was true, the informative normal priors had no noticeable influence on the Bayesian estimate (Figure 7), which was only sensitive to the underlying true parameter (Figure 8). The reason might be the large number of data that we have and the posterior is dominated by the data rather than the prior.

As a flexible extension of GLM introduced by Hastie and Tibshirani [42], GAM can estimate both linear trends from parametric components and nonlinear trend from any general nonparametric components during the fitting. It has been widely used in time-series studies on air pollution and health effects, controlling for daily variations in meteorological conditions and seasonal trends. The original GAM fitting method estimated the smooth components of the model using non-parametric smoothers, such as smoothing splines or LOESS, via the backfitting algorithm [42]. By iteratively smoothing partial residuals, backfitting provides a general module to estimate the  $S_j$  terms that are capable of using a wide variety of smoothing methods. The computational cost issue of full spline method has been addressed recently by using Markov random fields to find sparse representations of the smooths, which can be viewed as an empirical Bayesian method [43]. An alternative approach with particular advantages in high dimensional settings is to use boosting, which typically requires bootstrapping for uncertainty quantification [44,45].

Although frequentist GAM gives a rich family of models that have been widely applied, a crucial problem with GAM is the choice of the number and the position of the knots, in terms of analytical tractability. A small number of knots may result in insufficient flexibility to capture the variability of the data while a large one may lead to overfitting. P-splines approach makes a more parsimonious parameterization possible, which is of particular advantage in a Bayesian framework where inference is based on MCMC techniques [25]. By taking into account the complete likelihood surface rather than plugging in the maximum likelihood estimate of the covariance structure, the approach provides the posterior distributions of the quantities of interest, such as 'the true parameter has a probability of 0.95 of falling in a 95% credible interval (CI)', which is more interpretable. Unlike frequentist methods, which provided one estimate for each model parameter, Bayesian methods may provide, for each parameter, a sample of thousands of MCMC estimates from the simulated posterior distribution of the parameter. The reported posterior mean and posterior distribution are the corresponding summaries of the marginal posterior distribution of the parameter.

Due to recent developments in MCMC algorithms, software, and hardware, we can now use MCMC methods to analyze complex models that would have been impossible only a few decades ago [46]. However, the Bayesian GAM is only available in a few R packages such as *gammSlice* [47], *R2BayesX* [48], and *spikeSlabGAM* [49], which limits its application in a variety of scientific fields. In *gammSlice* Pham et al. used the slicing sampling method [50] for GAM fitting and inference within a Bayesian framework [47]. *R2BayesX* supports similar models to those in *gammSlice*. Scheipl used spike-and-slab type prior distributions on the spline coefficients [49]. Especially, Klein et al. proposed a general class of Bayesian GAM for count data within the GAM framework for location, scale, and shape where semiparametric predictors can be specified for several parameters of a count data distribution [21].

There are some limitations to our study. First, we did not impose any structure on the relationship of the coefficients of the lagged PM<sub>2.5</sub> concentrations with each other. However, multicollinearity among the lagged independent variables often arises, leading to high variance of the coefficient estimates. We plan to address this problem by constraining the  $\beta_l$  to be a simple function of the lag number using a structured finite distributed lag model in the future [51]. This problem can also be

addressed by a time-scale model which estimates the association between daily smooth variations of pollution and health outcomes by replacing  $\beta_l X_{t-l}$  in Equation (1) with:

$$\sum_{k=1}^K \beta_k W_{kt} \quad (11)$$

where  $W_{1t}, \dots, W_{kt}, \dots, W_{Kt}$  is a set of predictors obtained by applying a wavelet analysis or Fourier analysis to  $X_t$  to satisfy orthogonality, such that  $\sum_{k=1}^K W_{kt} = X_t$  [52–54]. The parameter  $\beta_k$  measures the logarithmic relative rate of the health outcome for increasing air pollution at time scale  $k$ . Time scales of interest include short-term variations within several days and long-term variations within one to two months because it is believed that any effects are dominated by seasonal confounding beyond two months [54]. Other tools such as auto-regressive moving average model or vector auto-regressive model could be also appropriate [55]. Second, our simulation framework did not address the issue of measurement error in the covariates. Because such error can in some situations attenuate the estimated effects, it may be useful in the future to employ a more elaborate simulation framework to incorporate the measurement error. Third, our model did not model interaction between  $PM_{2.5}$  and other covariates yet. The concentration-response relationship between  $PM_{2.5}$  and respiratory deaths was assumed to be approximately linear, but this might not be true in very low or very high  $PM_{2.5}$  concentrations. The two limitations can be addressed using varying coefficient model (VCM), where linear  $\beta_l X_{t-l}$  is replaced by  $S(X_{t-l})z_l$ . Estimation of VCM poses no further difficulties since only the  $\beta_l$  in Equation (7) has to be redefined by  $S(X_{t-l})$  multiplying with  $Z_l$ . Furthermore, we acknowledge that our study is based on simulated data, and the difference in the results between the frequentist and Bayesian methods is small and might be biased towards a Bayesian approach because of the predefined parameters. Therefore, the evidence to advocate the Bayesian approach over the frequentist one has yet been strong, and a confirmation study using real-world data is needed to address these issues in the future.

## 5. Conclusions

Our simulation study indicates that fully Bayesian GAM may generate an accurate and precise estimations as conventional frequentist GAM does while revealing potential uncertainty that frequentist GAM could not detect. Although computationally intensive, Bayesian GAM would be a better solution to avoid over-confident inferences potentially seen in a frequentist GAM. With the increasing computing power and available statistical packages, fully Bayesian methods may see wider applications in decision-making processes.

**Author Contributions:** Conceptualization, Y.C. and X.F.; Funding acquisition, Y.C.; Data curation, X.F. and B.F.; Formal analysis, X.F.; Methodology, Y.C. and X.F.; Project administration, Y.C., C.W., and T.X.; Resources, C.W., T.X., and B.F.; Supervision, Y.C., F.F., M.B.; Validation, Y.C. and X.F.; Visualization, X.F.; Writing—original draft, X.F. and Y.C.; Writing—review & editing, X.F., F.F., M.B., B.F., C.W., and T.X.

**Funding:** Xing Fang and Yang Cao received funding from the Karolinska Institutet Strategic Grant (Project no. C62400032), the Junior Faculty Grant of the Institute of Environmental Medicine, Karolinska Institutet (Project no. C62412022), and the Joint China-Sweden Mobility Grant of the Swedish Foundation for International Cooperation Research and Education (Grant no. CH2015-6145). The funding bodies were not involved in the design of the study, data collection and analysis, interpretation of results, writing the manuscript.

**Acknowledgments:** We thank the Shanghai Municipal Center for Disease Control and Prevention team for their cooperation.

**Ethics Approval:** The study was approved by the Ethical Review Committee of the Shanghai Municipal Center for Disease Control and Prevention (SCDC), China (approval number: SCDC2016-08).

**Availability of Data and Materials:** Data sharing is not applicable to this article as datasets were simulated during the current study. All the parameters for generating the simulated datasets were provided in Tables 1 and 2.

**Conflicts of Interest:** The authors declare that they have no conflict of interests.

## Abbreviations

AQM	Air quality monitoring
CDR	Causes of Death Registry
CrI	Credible interval
<i>df</i>	Degrees of freedom
DOW	Day of week
GAM	Generalized additive model
GLM	Generalized linear model
ICD-10	International Classification of Diseases Codes, version 10
IQR	Interquartile range
LOESS	locally weighted smoothers
MCMC	Markov Chain Monte Carlo
PM	Particulate matter
PM <sub>2.5</sub>	Fine particles with a diameter of 2.5 µm or less
SCDC	Shanghai Municipal Center for Disease Control and Prevention
SD	Standard deviation
SE	Standard error
SWT	Synoptic weather types
VCM	Varying coefficient model

## References

1. Cohen, A.J.; Anderson, H.R.; Ostro, B.; Pandey, K.D.; Krzyzanowski, M.; Kunzli, N.; Gutschmidt, K.; Pope, A.; Romieu, I.; Samet, J.M.; et al. The global burden of disease due to outdoor air pollution. *J. Toxicol. Environ. Health A* **2005**, *68*, 1301–1307. [[CrossRef](#)] [[PubMed](#)]
2. Franklin, M.; Koutrakis, P.; Schwartz, J. The role of particle composition on the association between pm2.5 and mortality. *Epidemiology* **2008**, *19*, 680–689. [[CrossRef](#)] [[PubMed](#)]
3. Englert, N. Fine particles and human health—A review of epidemiological studies. *Toxicol. Lett.* **2004**, *149*, 235–242. [[CrossRef](#)] [[PubMed](#)]
4. Cesaroni, G.; Forastiere, F.; Stafoggia, M.; Andersen, Z.J.; Badaloni, C.; Beelen, R.; Caracciolo, B.; de Faire, U.; Erbel, R.; Eriksen, K.T.; et al. Long term exposure to ambient air pollution and incidence of acute coronary events: Prospective cohort study and meta-analysis in 11 european cohorts from the escape project. *BMJ* **2014**, *348*, f7412. [[CrossRef](#)] [[PubMed](#)]
5. Shang, Y.; Sun, Z.; Cao, J.; Wang, X.; Zhong, L.; Bi, X.; Li, H.; Liu, W.; Zhu, T.; Huang, W. Systematic review of chinese studies of short-term exposure to air pollution and daily mortality. *Environ. Int.* **2013**, *54*, 100–111. [[CrossRef](#)] [[PubMed](#)]
6. Mann, C.J. Observational research methods. Research design ii: Cohort, cross sectional, and case-control studies. *Emerg. Med. J.* **2003**, *20*, 54–60. [[CrossRef](#)] [[PubMed](#)]
7. Rushworth, A.; Lee, D.; Mitchell, R. A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in greater london. *Spat. Spatiotemporal Epidemiol.* **2014**, *10*, 29–38. [[CrossRef](#)] [[PubMed](#)]
8. Lee, D.; Shaddick, G. Spatial modeling of air pollution in studies of its short-term health effects. *Biometrics* **2010**, *66*, 1238–1246. [[CrossRef](#)] [[PubMed](#)]
9. Yang, Y.; Cao, Y.; Li, W.; Li, R.; Wang, M.; Wu, Z.; Xu, Q. Multi-site time series analysis of acute effects of multiple air pollutants on respiratory mortality: A population-based study in Beijing, China. *Sci. Total Environ.* **2015**, *508*, 178–187. [[CrossRef](#)] [[PubMed](#)]
10. Pannullo, F.; Lee, D.; Waclawski, E.; Leyland, A.H. How robust are the estimated effects of air pollution on health? Accounting for model uncertainty using bayesian model averaging. *Spat. Spatio-Temporal Epidemiol.* **2016**, *18*, 53–62. [[CrossRef](#)] [[PubMed](#)]
11. Dominici, F. Time-series analysis of air pollution and mortality: A statistical review. *Res. Rep. Health Eff. Inst.* **2004**, *123*, 3–27, discussion 29–33.
12. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*, 2nd ed.; Monograph on Statistics and Applied Probability; Chapman & Hall: New York, NY, USA, 1989.

13. Hastie, T.J.; Tibshirani, R.J. *Generalized Additive Models*; Monographs on Statistics and Applied Probability; Chapman & Hall: London, UK, 1990.
14. Wood, S.N. *Generalized Additive Models: An Introduction with R*; CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2017.
15. Binder, H.; Tutz, G. A comparison of methods for the fitting of generalized additive models. *Stat. Comput.* **2008**, *18*, 87–99. [[CrossRef](#)]
16. Wong, R.K.W.; Yao, F.; Lee, T.C.M. Robust estimation for generalized additive models. *J. Comput. Graph. Stat.* **2014**, *23*, 270–289. [[CrossRef](#)]
17. Kim, Y.J.; Gu, C. Smoothing spline gaussian regression: More scalable computation via efficient approximation. *J. R. Stat. Soc. B* **2004**, *66*, 337–356. [[CrossRef](#)]
18. Fahrmeir, L.; Lang, S. Bayesian inference for generalized additive mixed models based on markov random field priors. *J. R. Stat. Soc. C-Appl.* **2001**, *50*, 201–220. [[CrossRef](#)]
19. Osei, F.B.; Duker, A.A.; Stein, A. Bayesian structured additive regression modeling of epidemic data: Application to cholera. *BMC Med. Res. Methodol.* **2012**, *12*, 118. [[CrossRef](#)] [[PubMed](#)]
20. Bové, D.S.; Held, L.; Kauermann, G. Objective bayesian model selection in generalized additive models with penalized splines. *J. Comput. Graph. Stat.* **2015**, *24*, 394–415. [[CrossRef](#)]
21. Klein, N.; Kneib, T.; Lang, S. Bayesian generalized additive models for location, scale, and shape for zero-inflated and overdispersed count data. *J. Am. Stat. Assoc.* **2015**, *110*, 405–419. [[CrossRef](#)]
22. Cengiz, M.A.; Terzi, Y. Comparing models of the effect of air pollutants on hospital admissions and symptoms for chronic obstructive pulmonary disease. *Cent. Eur. J. Public Health* **2012**, *20*, 282–286. [[CrossRef](#)] [[PubMed](#)]
23. Mamouridis, V. Additive Mixed Models Applied to the Study of Red Shrimp Landings: Comparison between Frequentist and Bayesian Perspectives. Master's Thesis, Universidade de Santiago de Compostela, Santiago de Compostela, Spain, 14 January 2011.
24. Fahrmeir, L.; Lang, S. Bayesian semiparametric regression analysis of multicategorical time-space data. *Ann. Inst. Stat. Math.* **2001**, *53*, 11–30. [[CrossRef](#)]
25. Lang, S.; Brezger, A. Bayesian p-splines. *J. Comput. Graph. Stat.* **2004**, *13*, 183–212. [[CrossRef](#)]
26. Hastie, T.; Tibshirani, R. Bayesian backfitting. *Stat. Sci.* **2000**, *15*, 196–213.
27. Eilers, P.H.C.; Marx, B.D. Flexible smoothing with b-splines and penalties. *Stat. Sci.* **1996**, *11*, 89–102. [[CrossRef](#)]
28. Fahrmeir, L.; Kneib, T.; Lang, S. Penalized structured additive regression for space-time data: A bayesian perspective. *Stat. Sin.* **2004**, *14*, 731–761.
29. Brezger, A.; Lang, S. Generalized structured additive regression based on bayesian p-splines. *Comput. Stat. Data Anal.* **2006**, *50*, 967–991. [[CrossRef](#)]
30. Zanobetti, A.; Schwartz, J.; Samoli, E.; Gryparis, A.; Touloumi, G.; Atkinson, R.; Le Tertre, A.; Bobros, J.; Celko, M.; Goren, A.; et al. The temporal pattern of mortality responses to air pollution: A multicity assessment of mortality displacement. *Epidemiology* **2002**, *13*, 87–93. [[CrossRef](#)] [[PubMed](#)]
31. Almon, S. The distributed lag between capital appropriations and expenditures. *Econom. J. Econom. Soc.* **1965**, 178–196. [[CrossRef](#)]
32. Vanos, J.K.; Cakmak, S.; Bristow, C.; Brion, V.; Tremblay, N.; Martin, S.L.; Sheridan, S.S. Synoptic weather typing applied to air pollution mortality among the elderly in 10 canadian cities. *Environ. Res.* **2013**, *126*, 66–75. [[CrossRef](#)] [[PubMed](#)]
33. Fang, X.; Fang, B.; Wang, C.; Xia, T.; Bottai, M.; Fang, F.; Cao, Y. Relationship between fine particulate matter, weather condition and daily non-accidental mortality in shanghai, china: A bayesian approach. *PLoS ONE* **2017**, *12*, e0187933. [[CrossRef](#)] [[PubMed](#)]
34. Ren, M.; Fang, X.; Li, M.; Sun, S.; Pei, L.; Xu, Q.; Ye, X.; Cao, Y. Concentration-response relationship between pm2. 5 and daily respiratory deaths in china: A systematic review and metaregression analysis of time-series studies. *BioMed Res. Int.* **2017**, 2017. [[CrossRef](#)] [[PubMed](#)]
35. Lu, F.; Xu, D.; Cheng, Y.; Dong, S.; Guo, C.; Jiang, X.; Zheng, X. Systematic review and meta-analysis of the adverse health effects of ambient PM2.5 and PM10 pollution in the chinese population. *Environ. Res.* **2015**, *136*, 196–204. [[CrossRef](#)] [[PubMed](#)]
36. Chen, R.; Yin, P.; Meng, X.; Liu, C.; Wang, L.; Xu, X.; Ross, J.A.; Tse, L.A.; Zhao, Z.; Kan, H.; et al. Fine particulate air pollution and daily mortality: A nationwide analysis in 272 chinese cities. *Am. J. Respir. Crit. Care Med.* **2017**, *196*, 73–81. [[CrossRef](#)] [[PubMed](#)]

37. Dominici, F.; McDermott, A.; Hastie, T.J. Improved semiparametric time series models of air pollution and mortality. *J. Am. Stat. Assoc.* **2004**, *99*, 938–948. [[CrossRef](#)]
38. Daniels, M.J.; Dominici, F.; Zeger, S. Underestimation of standard errors in multi-site time series studies. *Epidemiology* **2004**, *15*, 57–62. [[CrossRef](#)] [[PubMed](#)]
39. Rice, J. Convergence rates for partially splined models. *Stat. Probab. Lett.* **1986**, *4*, 203–208. [[CrossRef](#)]
40. Speckman, P. Kernel smoothing in partial linear models. *J. R. Stat. Soc. Ser. B (Methodol.)* **1988**, *50*, 413–436. [[CrossRef](#)]
41. Martin, A.D.; Quinn, K.M.; Park, J.H. Mcmcpack: Markov chain monte carlo in r. *J. Stat. Softw.* **2011**, *42*, 1–21. [[CrossRef](#)]
42. Hastie, T.J.; Tibshirani, R.J. *Generalized Additive Models*; CRC Press: Boca Raton, FL, USA, 1990; Volume 43.
43. Rue, H.; Martino, S.; Chopin, N. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *J. R. Stat. Soc. B* **2009**, *71*, 319–392. [[CrossRef](#)]
44. Schmid, M.; Hothorn, T. Boosting additive models using component-wise p-splines. *Comput. Stat. Data Anal.* **2008**, *53*, 298–311. [[CrossRef](#)]
45. Mayr, A.; Fenske, N.; Hofner, B.; Kneib, T.; Schmid, M. Generalized additive models for location, scale and shape for high dimensional data—A flexible approach based on boosting. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2012**, *61*, 403–427. [[CrossRef](#)]
46. Kruschke, J. *Doing Bayesian Data Analysis: A Tutorial With r, Jags, and Stan*; Academic Press: Amsterdam, The Netherlands, 2014.
47. Pham, T.H.; Wand, M.P. *Generalized Additive Mixed Model Analysis via Gammslice*; Wiley: New York, NY, USA, 2015.
48. Umlauf, N.; Kneib, T.; Heinzl, F.; Lang, S.; Zeileis, A. R2bayesx: Estimate structured additive regression models with bayesx. In *R Package Version 0.1-2*; The Comprehensive R Archive Network: Vienna, Austria, 2013.
49. Scheipl, F. Spikeslabgam: Bayesian variable selection, model choice and regularization for generalized additive mixed models in r. *J. Stat. Softw.* **2011**, *43*, 1–24. [[CrossRef](#)]
50. Neal, R.M. Slice sampling. *Ann. Stat.* **2003**, *31*, 705–741. [[CrossRef](#)]
51. Zanobetti, A.; Wand, M.P.; Schwartz, J.; Ryan, L.M. Generalized additive distributed lag models: Quantifying mortality displacement. *Biostatistics* **2000**, *1*, 279–292. [[CrossRef](#)] [[PubMed](#)]
52. Valari, M.; Martinelli, L.; Chatignoux, E.; Crooks, J.; Garcia, V. Time scale effects in acute association between air-pollution and mortality. *Geophys. Res. Lett.* **2011**, *38*. [[CrossRef](#)]
53. Zeger, S.L.; Dominici, F.; Samet, J. Harvesting-resistant estimates of air pollution effects on mortality. *Epidemiology* **1999**, *10*, 171–175. [[CrossRef](#)] [[PubMed](#)]
54. Dominici, F.; McDermott, A.; Zeger, S.L.; Samet, J.M. Airborne particulate matter and mortality: Timescale effects in four us cities. *Am. J. Epidemiol.* **2003**, *157*, 1055–1065. [[CrossRef](#)] [[PubMed](#)]
55. de Souza, J.B.; Reisen, V.A.; Franco, G.C.; Ispany, M.; Bondon, P.; Santos, J.M. Generalized additive models with principal component analysis: An application to time series of respiratory disease and air pollution data. *J. R. Stat. Soc. C-Appl.* **2018**, *67*, 453–480. [[CrossRef](#)]

