MATEMATIKA, MJIAM, Volume 42, Number 2, 155–171

Air Quality Analysis Using Non-Informative and Weakly Informative Priors for Regression Coefficients in the BSTS Model

[1,2]Aeshah Ali Mohammed, [1]Mohd Aftar Abu Bakar*, [3,4]Mohd Mahayaudin Mansor and [1]Noratiqah Mohd Ariff

[1]Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.
[2]Faculty of Science, University of Benghazi, AL Marje, Libya.
[3]School of Mathematical Sciences, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.
[4]Institute for Big Data Analytics and Artificial Intelligence, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.
*Corresponding author: aftar@ukm.edu.my

Article history
Received: 28 March 2025
Received in revised form: 4 February 2026
Accepted: 8 May 2026
Published on line: 20 July 2026

Abstract Understanding the complex dynamics of air quality is crucial for reducing environmental pollution and protecting public health. This study aims to investigate the impact of different prior specifications in Bayesian structural time series (BSTS) modelling on the accuracy of predicting air pollution levels (PM2.5 and PM10) across various monitoring stations. The study examines time series data from 5 July 2017 to 30 June 2019. Two priors are compared: non-informative prior regression coefficients (NIP) and weakly informative prior regression coefficients (WIP). The analysis focuses on forecast accuracy metrics, including observed and predicted standard deviations, relative goodness of fit, and R-squared. The results indicate that both approaches indicate that PM10 is more predictable than PM2.5 at all stations, but WIP has been seen to produce lower MAPEs and RMSEs, as well as tighter credible intervals that reduce forecast uncertainty in most stations. For instance, at station CA13A, for PM2.5, WIP lowers MAPE by 0.34% (from 10.68 to 10.34%) and reduces RMSE by 0.017 (0.391 to 0.374), while R^2 remains stable at 0.85, and relative gof improves slightly from 0.226 to 0.221. For PM10, Relative gof improves slightly from 0.249 to 0.246. Posterior distributions of coefficients ($\boldsymbol{\beta}$) under both prior specifications identified ambient temperature and humidity as significant predictors across stations, confirming that the choice of prior does not drastically change which covariates are important. The WIP specification acts towards a more balanced model through enhancing the inclusion of less dominant predictors. Despite minor discrepancies in coefficient estimates, WIP models provided nuanced advantages in model fit and explanatory power. These findings underscore the importance of prior specifications in Bayesian modelling, offering valuable insights for air quality management and forecasting.

Keywords Air quality; Bayesian structural time series; Non-Informative Prior (NIP); Weakly Informative Prior (WIP).

Mathematics Subject Classification 62M10.

1 Introduction

Air quality is a global issue with significant environmental and public health impacts, originating from industrial emissions, vehicle traffic, and natural events like wildfires [1]. Long-term exposure to fine particulate matter is linked to respiratory illness, heart disease, and premature death [2-3]. Many countries, including Malaysia, have implemented air quality monitoring systems to reduce health risks [4].

Researchers have employed many models to enhance the understanding and prediction of air quality. Studies like Abu Bakar, Che Ilias, and Abd Rahman have developed various models for predicting air pollution levels in Peninsular Malaysia. Bakar's Network Temporal Convolution (TCN) model outperformed LSTM in accuracy [5]. Mun's MLP model performed better than ARIMA in central Malaysia [6]. Bakar's multivariate LSTM model outperformed the LSTM and ARIMA models in PM10 forecasting [7]. Zheng et al. [8] found long-term relationships between climate parameters and air quality fluctuations. Lastly, Ariff et al. [9] used a hybrid k-means clustering-LSTM model to estimate daily PM10 levels from 60 Malaysian air quality monitoring sites with shorter training times. Though some level of accuracy has been achieved, in all cases, the methodologies have had difficulties in probability theory to assess uncertainties and long-term trends.

Bayesian Structural Time Series (BSTS) modelling stands out as a promising alternative method. It captures temporal dynamics of trends, seasonality and autoregression as part of its probabilistic model forecasts that take into account uncertainties explicitly [10]. Besides such features, the framework incorporates spike-and-slab regression and Bayesian model averaging, making it easy to find critical predictors and enhance interpretability [11, 12, 13].

There have been other previous applications of the BSTS in environmental studies, including our recent research on Malaysian air quality [14], where they proved effective in capturing temporal dynamics and integrating meteorological predictors. However, previous applications have been primarily limited to predictive accuracy, with little emphasis on one of the critical methodological questions whether the choice of prior distribution affect forecasting performance in BSTS, whether a weakly informative prior yield statistically and practically better forecasts than a non-informative prior for the Malaysian air-quality series, or whether covariate selection, which meteorological covariates retain high posterior inclusion probability under both priors.

Therein lies the gap that has been filled by this study through systematic comparisons of the two prior regimes: non-informative priors (NIP) and weakly informative priors (WIP). NIP specifies very few assumptions, letting the data dominate the inference process, whereas WIP applies some weak regularization that prevents the estimates from being very wiggly but does not allow the model to overfit [15-16].

To date, this is the first empirical study on the sensitivity of priors within BSTS forecasting of Malaysian air-quality data. Its tropical climate, monsoon-driven seasonal dominance, and episodic haze events from those in temperate regions make Malaysia a unique case study subject. The monitoring network covers rural, urban, suburban, and industrial areas, offering a varied environment for model evaluation.

There are three objectives of this study: to assess differences in forecast skill between NIP and WIP, to assess covariate inclusion probabilities under both priors, and to provide methodological and applied guidance to environmental forecasters. In so doing, it contributes to the methodological development of Bayesian forecasting as well as to policy-relevant understanding

of air quality dynamics in Malaysia.

The remainder of this paper is organized as follows: The Bayesian structural time-series approach is explained in the methodology section. The subsequent section presents the results of an application to air quality data, emphasizing the impact of non-informative and weakly informative priors on regression coefficients. Lastly, the final section contains the study's conclusions.

2 Methods

Scott and Varian (2014) introduced the Bayesian structural time series (BSTS) model to forecast economic time series. The model integrates the Kalman filter, spike-and-slab regression, and Bayesian model averaging to improve predictive performance.

2.1 Structural Time Series

A structural time series model requires specifying equations linking the observed time series $\boldsymbol{y}_i$ to a vector of latent state variables, $\boldsymbol{\alpha}_t$. The observation error $\boldsymbol{\varepsilon}_t$ follows a normal distribution with mean $\mathbf{0}$ and variance $\mathbf{H}_t$. The vector $\mathbf{Z}_t$ is fixed and has dimension $m \times 1$. This is given by

$$\mathbf{y}_t = \mathbf{Z}_t^T \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{H}_t). \tag{1}$$

The evolution of the latent state is described as

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{R}_t \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t). \tag{2}$$

Here, $\boldsymbol{\alpha}_{t+1}$ represents the state variables at time $t+1$. The matrix $\mathbf{T}_t$ is a square transition, $\mathbf{R}_t$ is a rectangular control matrix, and the vector $\boldsymbol{\eta}_t$ is a normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{Q}_t$ [17].

A latent state model typically includes a local linear trend, seasonality, and regression [15]. A general BSTS formulation is

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\tau}_t + \boldsymbol{\beta}^T \mathbf{X}_t + \boldsymbol{\varepsilon}_t \tag{3}$$

,

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\delta}_{t-1} + \mathbf{u}_t, \tag{4}$$

$$\boldsymbol{\delta}_t = \boldsymbol{\delta}_{t-1} + \mathbf{v}_t, \tag{5}$$

$$\boldsymbol{\tau}_t = -\sum_{s=1}^{S-1} \boldsymbol{\tau}_{t-s} + \mathbf{w}_t. \tag{6}$$

In these equations, $\boldsymbol{\mu}_t$ and $\boldsymbol{\delta}_t$ represent the level and slope at time t, $\boldsymbol{\mu}_{t-1}$ and $\boldsymbol{\delta}_{t-1}$ represent the level and slope of the time series at time $t-1$.

The seasonal component $\boldsymbol{\tau}_t$ is comprises a series of S dummy variables with dynamic coefficients constrained to have zero expectation over an S-season period. In this study, we modelled daily data with an $S = 52$ and a season duration of 7 days. The noise terms $\boldsymbol{\varepsilon}_t, \mathbf{u}_t, \mathbf{v}_t$ and $\mathbf{w}_t$ are independent Gaussian normal distributions with variances $(\sigma_\varepsilon^2, \sigma_u^2, \sigma_v^2, \sigma_w^2)$. The vector $\mathbf{X}_t$

contains explanatory variables (meteorological predictors). Their coefficients $\boldsymbol{\beta}$ follow a spike and slab prior.

Prior distributions, such as normal and gamma distributions, are used to initialize initial states and variance parameters. The Kalman filter estimates unobserved state variables through initialization, prediction, and update stages, while the Kalman Smoother combines filtered and backward estimates [18].

2.2 Spike-and-Slab Regression and Spike and Slab Prior Structure

Spike and slab regression, proposed by George and McCulloch (1997) and Madigan and Raftery (1994), provides a Bayesian variable selection framework. It reduces the number of variables and incorporates prior beliefs about predictor relevance.

The prior on the regression coefficients ($\boldsymbol{\beta}$) is decomposed into three components:

- The subset of non-zero coefficients ($\boldsymbol{\beta}_\gamma$).
- The variance (σ^2).
- The inclusion vector (γ).

The joint prior is

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_\varepsilon^2) = p(\boldsymbol{\beta}_\gamma | \boldsymbol{\gamma}, \sigma_\varepsilon^2) p(\sigma_\varepsilon^2 | \boldsymbol{\gamma}) p(\boldsymbol{\gamma}). \tag{7}$$

Prior to Spike

The spike $p(\boldsymbol{\gamma})$ is typically modelled using a product of independent Bernoulli distributions [17]. Each predictor is either included ($\gamma = 1$) or excluded ($\gamma = 0$), with

$$\boldsymbol{\gamma} \sim \prod_{k=1}^{K} \pi_k^{\gamma k} (1 - \pi_k)^{1 - \gamma k}. \tag{8}$$

If $\gamma = 1$, the corresponding coefficient $\boldsymbol{\beta}$ is non-zero, and if $\gamma = 0$, $\boldsymbol{\beta}$ is zero. Analysts may set $\pi_K = 0$ or $\pi_K = 1$ for predictors that must be excluded or included, respectively [19].

Prior to Slab

The slab portion uses a conjugate normal-inverse gamma prior. This choice allows computation efficiency due to conjugacy with the normal likelihood function and supports shrinkage for weaker predictors [17]. This distribution is parameterized by the mean vector $\mathbf{b}$ (often set to 0), the precision matrix ($\boldsymbol{\Omega}$), and hyperparameters for the inverse gamma distribution (df and SS). The prior coefficients $\boldsymbol{\beta}$ given the error variance σ_ε^2 is expressed as

$$\boldsymbol{\beta}_\gamma | \sigma_\varepsilon^2, \boldsymbol{\gamma} \sim N\left(\mathbf{b}_\gamma, \sigma_\varepsilon^2 (\boldsymbol{\Omega}_\gamma^{-1})^{-1}\right). \tag{9}$$

Zellner's g-prior is often used, which sets the precision matrix to

$$\boldsymbol{\Omega}^{-1} = K\left(W\mathbf{X}^T\mathbf{X} + (1 - W)\mathrm{diag}(\mathbf{X}^T\mathbf{X})\right)/n,$$

to ensure propriety [20].

Error Variance Prior

The error variance, σ_ε^2 is assumed to follow the inverse gamma distribution, where

$$\frac{1}{\sigma_\varepsilon^2}|\boldsymbol{\gamma} \sim \text{Ga}\left(\frac{df}{2}, \frac{SS}{2}\right). \tag{10}$$

The inverse Gamma distribution is commonly used as a conjugate prior to the variance parameter in Bayesian inference, particularly for continuous data. This prior stabilizes estimation and improves robustness, particularly for noisy data [21].

In brief, the spike-and-slab prior provides considerable flexibility in representing prior beliefs through parameters such as π_K, $\mathbf{b}$, $\boldsymbol{\Omega}^{-1}, SS$, and df. To simplify, prior information can be captured using the expected model size, expected R^2, and sample size df, or by default values. Our study proceeds with a comparative analysis, employing BSTS with non-informative priors and weakly informative priors. Non-informative values ($R^2 = 0.01$, $df = 0.001$, and $\pi_K = 0.3$) foster a more conservative estimation approach with broader, credible intervals. In contrast, weakly informative priors are set with $R^2 = 0.5$, $df = 0.01$, and $\pi_K = 0.5$; these defaults allow data to influence posterior estimates significantly while providing some regularization [10,17,22].

2.3 Bayesian model averaging

Bayesian inference with spike-and-slab priors enables efficient model averaging. By sampling from the posterior distribution, various parameter sets for forecasting $\hat{y}_{t+h}$ are obtained. Repeating this process estimates the posterior distribution.

3 Results and Discussion

This study used Bayesian structural time series to analyze air pollutant readings from four monitoring stations in Malaysia. Hourly PM10 or PM2.5 measurements from July 5th, 2017, to June 30th, 2019, and calculated daily averages. The selected data period spans two full years to comprehensively capture seasonal variations, long-term trends, and potential environmental events that influence PM10 and PM2.5 levels. This timeframe ensures high-quality, reliable data from monitoring stations while avoiding the impact of unusual events, such as COVID-19, that could introduce bias or outliers into the analysis.

The model incorporated ambient temperature, solar radiation, humidity, wind speed, and wind direction as exogenous variables. A Monte Carlo Markov Chain with 2000 iterations was used to simulate posterior distributions and assess parameter-estimation uncertainty. The findings are presented in the tables below.

The analysis began by specifying a prior distribution $\rho(\theta)$ over the model parameters θ before observing any data. This prior distribution encapsulates our beliefs or uncertainties about the parameters prior to data collection. The selection of the prior is crucial as it impacts posterior inference. In prior elicitation, prior distributions are defined and assigned based on existing knowledge, beliefs, or available information about the model parameters. For instance, the Gamma distribution (a, b) is often used for setting priors on variance parameters like σ^2 [10]. The choice of the Gamma distribution (df, SS) further illustrates this concept, with df

representing the shape parameter and SS representing the scale parameter. The parameter SS/df serves as a prior estimate of σ^2, with df denoting the weight or prior sample size assigned to this estimate. Additionally, in some cases, such as seasonal and local linear trend models, the sample variance s_y^2. The target series y is utilized to scale the prior distribution appropriately. This scaling aids in setting a reasonable prior scale based on the observed data characteristics. Although scaling by s_y^2 may slightly deviate from strict Bayesian principles, but it allows us to establish meaningful priors that align with the underlying characteristics of the data. In summary, effective prior elicitation involves defining informative prior distributions that incorporate prior knowledge and observed data characteristics. By using the prior sum of squares, a guess of the variance can be derived (df/SS) for the estimation process by using the default prior for the variance σ^2, $1/\sigma^2 \sim \left(10^{-2}, 10^{-2}s_y^2\right)$. This approach allows for the specification of the prior distribution for variables such as PM2.5 and PM10 (see Table 1).

Table 1: The Standard Deviation Prior Distribution for Each Station

Stations		Components	Prior guess	Prior df	Initial value	Upper limit
CA13A	PM2.5	σ_u (level)	0.0041	0.01	0.0041	0.41
		σ_v (slope)	0.0041	0.01	0.0041	0.41
		σ_w (seasonal)	0.0041	0.01	0.0041	0.41
	PM10	σ_u (level)	0.0034	0.01	0.0034	0.34
		σ_v (slope)	0.0034	0.01	0.0034	0.34
		σ_w (seasonal)	0.0034	0.01	0.0034	0.34
CA20B	PM2.5	σ_u (level)	0.0031	0.01	0.0031	0.31
		σ_v (slope)	0.0031	0.01	0.0031	0.31
		σ_w (seasonal)	0.0031	0.01	0.0031	0.31
	PM10	σ_u (level)	0.0029	0.01	0.0029	0.29
		σ_v (slope)	0.0029	0.01	0.0029	0.29
		σ_w (seasonal)	0.0029	0.01	0.0029	0.29
CA46D	PM2.5	σ_u (level)	0.0045	0.01	0.0045	0.45
		σ_v (slope)	0.0045	0.01	0.0045	0.45
		σ_w (seasonal)	0.0045	0.01	0.0045	0.45
	PM10	σ_u (level)	0.0038	0.01	0.0038	0.38
		σ_v (slope)	0.0038	0.01	0.0038	0.38
		σ_w (seasonal)	0.0038	0.01	0.0038	0.38
CA05K	PM2.5	σ_u (level)	0.0038	0.01	0.0038	0.38
		σ_v (slope)	0.0038	0.01	0.0038	0.38
		σ_w (seasonal)	0.0038	0.01	0.0038	0.38
	PM10	σ_u (level)	0.0030	0.01	0.0030	0.30
		σ_v (slope)	0.0030	0.01	0.0030	0.30
		σ_w (seasonal)	0.0030	0.01	0.0030	0.30

The term prior assumption refers to the initial estimate for the standard deviation of a component before fitting a model. The "prior degrees of freedom" parameter represents the prior sample size or the degree of certainty associated with the prior guess. By default, the prior degrees of freedom are set to 0.01, indicating a relatively low level of certainty or diffuse prior distribution.

Setting model parameters starting points to prior assumption values is crucial in Bayesian inference. By initializing parameters like σ_u, σ_v and σ_w. With assumed values before observing data, we incorporate our prior beliefs into the model. This strategy enhances convergence and stability by guiding the estimation algorithm towards regions likely to produce realistic results. It also improves computational efficiency by reducing the need for extensive parameter space exploration during fitting. As the model learns from data, these starting points are a foundation for continuously updated posterior estimates, reflecting the interplay between prior beliefs and empirical evidence.

The model fitting process uses a boundary calculated as (prior assumption/prior degrees of freedom) to prevent unrealistic or unstable results. This ensures reliable and interpretable results by truncating the estimated value if it exceeds the boundary.

Table 1 shown the estimates of level (σ_u), slope (σ_v), and seasonal (σ_w) parameters for each monitoring station and pollutant components from the model. Noted that the time series components have the same standard deviation value before observing any data; this simplifies the initialization process and provides a convenient starting point for estimation. The prior degrees of freedom are the same (0.01) in all stations, accommodating a wider range of potential values. The starting point can be updated and refined during the estimation process as the model learns from the data. The boundary is set relatively high for all stations, indicating that the model will not allow the components to exceed these limits during estimation.

The specifications for the prior distribution of the starting point of the component in Table 2 are included in the model specifications that describe the state vector's prior distribution at time 1. It utilizes a Normal prior distribution.

Table 2 displays the starting point and prior distributions for various components of a state space model that will be utilized to simulate air quality data from different monitoring stations. The parameter μ specifies the mean of the normal prior, and sigma specifies the standard deviation. The parameter starting point is the starting point for the level in the state space model. Noted that "starting point" is equivalent to μ. In some cases, the starting point can be used as the prior mean (μ) because it is a good initial value for the prior distribution. This assumption assumes that the Starting Point is a reasonable estimate of the true underlying parameter. It's possible that in the specific context we are referring to, the initial value is used as the prior mean (μ) for simplicity and as an initial assumption [10].

Bayesian structural time series modelling involves using two types of priors to guide the estimation of parameters and structure the regression components. In Table 3, the non-informative prior (NIP) assumes a prior was configured with a very low expected R^2 value of 0.01, indicating little belief in the covariates' ability to explain variation in the outcome. Additionally, the prior degrees of freedom were set to a near-zero value (0.001), further flattening the prior distribution. Covariate inclusion probabilities were uniformly low (0.3), reflecting weak prior assumptions about the importance of any specific predictor, and the absence of an upper limit for the error term's standard deviation signifies an openness to a broader range of possibilities. On the other hand, the weakly informative prior (WIP) assumes that all potential independent variables have equal inclusion probabilities ($\pi_K = 0.5$), to ensure a balanced consideration of their relevance in the model. This approach incorporates a moderate prior belief ($R^2 = 0.5$) regarding the overall variation explained and a 1% shrinkage parameter ($df = 0.01$) for a cautious approach. Additionally, the upper limit for the error term's standard deviation prevents the priors from being overly diffuse.

Table 2: Specification for the Prior Distribution of the Starting Point for Each Station

Stations		Initial Components	μ	Sigma	Initial Value
CA13A	PM2.5	σ_u (level)	3.759	0.41	3.759
		σ_v (slope)	-0.0017	0.41	-0.0017
		σ_w (seasonal)	0	0.41	0
	PM10	σ_u (level)	3.871	0.34	41.24
		σ_v (slope)	-0.0015	0.34	-0.0015
		σ_w (seasonal)	0	0.34	0
CA20B	PM2.5	σ_u (level)	3.160	0.31	3.160
		σ_v (slope)	0.0002	0.31	0.0002
		σ_w (seasonal)	0	0.31	0
	PM10	σ_u (level)	3.509	0.29	3.509
		σ_v (slope)	0.0001	0.29	0.0001
		σ_w (seasonal)	0	0.29	0
CA46D	PM2.5	σ_u (level)	2.376	0.45	2.376
		σ_v (slope)	0.0004	0.45	0.0004
		σ_w (seasonal)	0	0.45	0
	PM10	σ_u (level)	2.848	0.38	2.848
		σ_v (slope)	0.001	0.38	0.001
		σ_w (seasonal)	0	0.38	0
CA05K	PM2.5	σ_u (level)	3.324	0.38	3.324
		σ_v (slope)	-0.0015	0.38	-0.0015
		σ_w (seasonal)	0	0.38	0
	PM10	σ_u (level)	3.606	0.30	3.606
		σ_v (slope)	-0.001	0.30	-0.001
		σ_w (seasonal)	0	0.30	0

Table 3 describes the configuration of the spike-and-slab prior under the non-informative prior setting for models and the weakly informative prior. In the non-informative prior, the prior inclusion probability for the models was uniformly set to 0.3 to reflect weak prior belief in the relevance of individual predictors. The prior mean (μ) was fixed at 0, implying no assumed effect size, and the prior degrees of freedom (prior *df*) were set at 0.001 to indicate minimal prior confidence. The (sigma guess) values, which serve as initial estimates of the coefficient standard deviation, varied by station and pollutant depending on the empirical variability observed in the data. The upper limit for sigma was set to infinity, thereby removing any constraint on the spread of the parameter, consistent with non-informative prior configurations. This configuration is to be such that the data prevails in inference.

In this section, we also set the spike and slab priors for the regression component, using a weakly informative prior (WIP). The WIP typically generates moderate assumptions about the model structure and relevance. Thus, the expected R^2 was increased to 0.5 as an assumption that covariates have a meaningful, although not overwhelming, explanatory role. Prior degrees of freedom were raised to 0.01 to favour regularisation against the NIP, yet offer substantial flexibility. Also, all covariates were given a moderate level of probability for inclusion into the model (0.5), which balances prior belief and model adaptability. This design is particularly

Table 3: Prior Settings Summary for Regression Coefficients Under NIP and WIP

Stations	Pollutant	Model	Prior inc. prob.	μ	Sigma guess	Prior df	Sigma boundary
CA13A	PM2.5	NIP	0.3	0	0.41	0.001	Inf
		WIP	0.833	0	0.291	0.01	0.495
	PM10	NIP	0.3	0	0.24	0.001	Inf
		WIP	0.833	0	0.243	0.01	0.413
CA20B	PM2.5	NIP	0.3	0	0.30	0.001	Inf
		WIP	0.833	0	0.216	0.01	0.367
	PM10	NIP	0.3	0	0.28	0.001	Inf
		WIP	0.833	0	0.203	0.01	0.345
CA46D	PM2.5	NIP	0.3	0	0.44	0.001	Inf
		WIP	0.833	0	0.318	0.01	0.539
	PM10	NIP	0.3	0	0.32	0.001	Inf
		WIP	0.833	0	0.269	0.01	0.458
CA05K	PM2.5	NIP	0.3	0	0.38	0.001	Inf
		WIP	0.833	0	0.271	0.01	0.459
	PM10	NIP	0.3	0	0.29	0.001	Inf
		WIP	0.833	0	0.209	0.01	0.356

useful in situations where the data-generating process is known about but not entirely certain. Table 3 shows that the upper bound of the error term standard deviation ensures that the priors are not overly diffuse. Noted that the expected model size is 5. Therefore, all six regression coefficients have an equal inclusion probability of $5/6 = 0.833$ in all stations. The sigma boundary parameter sets the upper limit for the standard deviation of the error term. Any values higher than this limit will be truncated at this value to prevent the prior from being too diffuse and allowing for unrealistic or uninformative values. The sigma. guess parameter represents the prior guess for the error term's standard deviation (σ_ε) in the model. The degrees of freedom (df) parameter control the shape of the gamma distribution for the error variance (σ_ε). In this case, degrees of freedom (df) are set to 0.01, indicating a diffuse prior distribution.

The decision to set the prior mean (μ) of regression coefficients to 0, suggesting that the predictors are not significant on the response variable in Table 3, does not imply a clear assertion about the true effects of these predictors. Instead, this is a matter of conservative modelling, where strong prior beliefs about the direction or size of effects of predictors are deliberately avoided. With the prior distribution centred around zero, the modelling process can remain wider open to the influences of observed data. This has been very valuable for those instances when not much prior information is available or when the effects of predictors are quite uncertain. Under a Bayesian framework, the prior specification with ($\mu = 0$) serves as a neutral starting point, enabling the model to adapt and update its beliefs based on the observed data [10].

In Table 3, all stations have the same mean vector of 0 has been included for all stations, as they contain five identical variables, namely ambient temperature, solar radiation, humidity, wind speed, and wind direction per station. These prior values provide a starting point for estimating the posterior distribution of the regression coefficients and error variance.

The data was divided as follows, 80% training data and 20% testing data. Training data included observations between July 5th, 2017 and February 4, 2019, covering a sufficient range of time to capture seasonal variation as well as long-term trends. The remaining 20% (February 5th to June 30th, 2019) was reserved for testing so that performance can be evaluated on unseen data. The split was also made to maximize the amount of training data while retaining samples for reliable testing.

It is against these actuals that we ran the model data for performance evaluation after the training process was complete, as shown in Figure 1 and Figure 2. Each subfigure shows in-sample fitting and out-of-sample forecasting for PM2.5 and PM10 time series data by the BSTS model. In-sample fitting relates the values observed during the training period with the fitted values produced by the model, and how well the model captures some temporal patterns and underlying structure is then assessed. Out-of-sample model evaluation involves prediction-testing the pure difference between actual test values and model forecasts, which indicates the generalization capability of the model. The shaded grey region indicates the 95% credible range for the forecast. Error measures such as MAPE and RMSE quantify accuracy. are calculated based on out-of-sample data to quantify accuracy.

Comparing Figures 1 and 2, the performances of the different BSTS models brought about using weakly informative priors (WIP) are estimated to be better than non-informative priors (NIP). Both approaches indicate that PM10 is more predictable than PM2.5 at all stations, but WIP has been seen to produce lower MAPEs and RMSEs as well as tighter credible intervals that reduce forecast uncertainty at CA13A, respectively. NIP, on the other hand, is more likely to underestimate some peak events occurring within sites and gives less reliable predictions within those sites having greater variability, such as CA46D and CA05K.

Overall, forecasts based on WIP are more stable and accurate, indicating that prior information in the model enhances the site-specific pollution dynamics. The consistently lower performance of forecasts for PM2.5 is clearly lower, judging against forecasts for PM10 and can be attributed to both measurement and source-related causes. Although PM10 measurements rely heavily on more stable meteorological drivers, such as wind transport and dust episodes, PM2.5 is more localized and highly varied in its sources: traffic emissions, combustion processes, and industrial activity. The huge spike of these sources' forms extremely abrupt, sometimes episodic, pollution concentrations that the time-series models find quite difficult to capture, especially when not all relevant covariates are observed. Furthermore, PM2.5 measurements are very prone to error owing to their dependent measurements, which require much finer monitoring equipment, as well as the small sizes of the particles that are monitored, further complicating prediction. In conditions such as this kind of volatility, priors take on a very important role. Non-informative priors cause the model to depend at most solely on the observed data and thus may be quickly swayed in responses to noise or episodic events and inflate forecast errors. Weakly informative priors, however, infuse a moderate level of regularization into the process, which will help to stabilize coefficient estimation and dampen sudden fluctuations. This would improve the model's generalisation capacity while at the same time reducing uncertainties regarding volatile episodes since site-specific dynamics would still be captured. But exactly what 'extreme' means is shown by CA05K; here, even WIP may not suffice, and further complements like stronger priors, hierarchical structures, or additional covariates (e.g., traffic volume, industrial output, or biomass burning indicators) may be warranted. For further information on model evaluation, Table 4 compares forecast accuracy between Bayesian

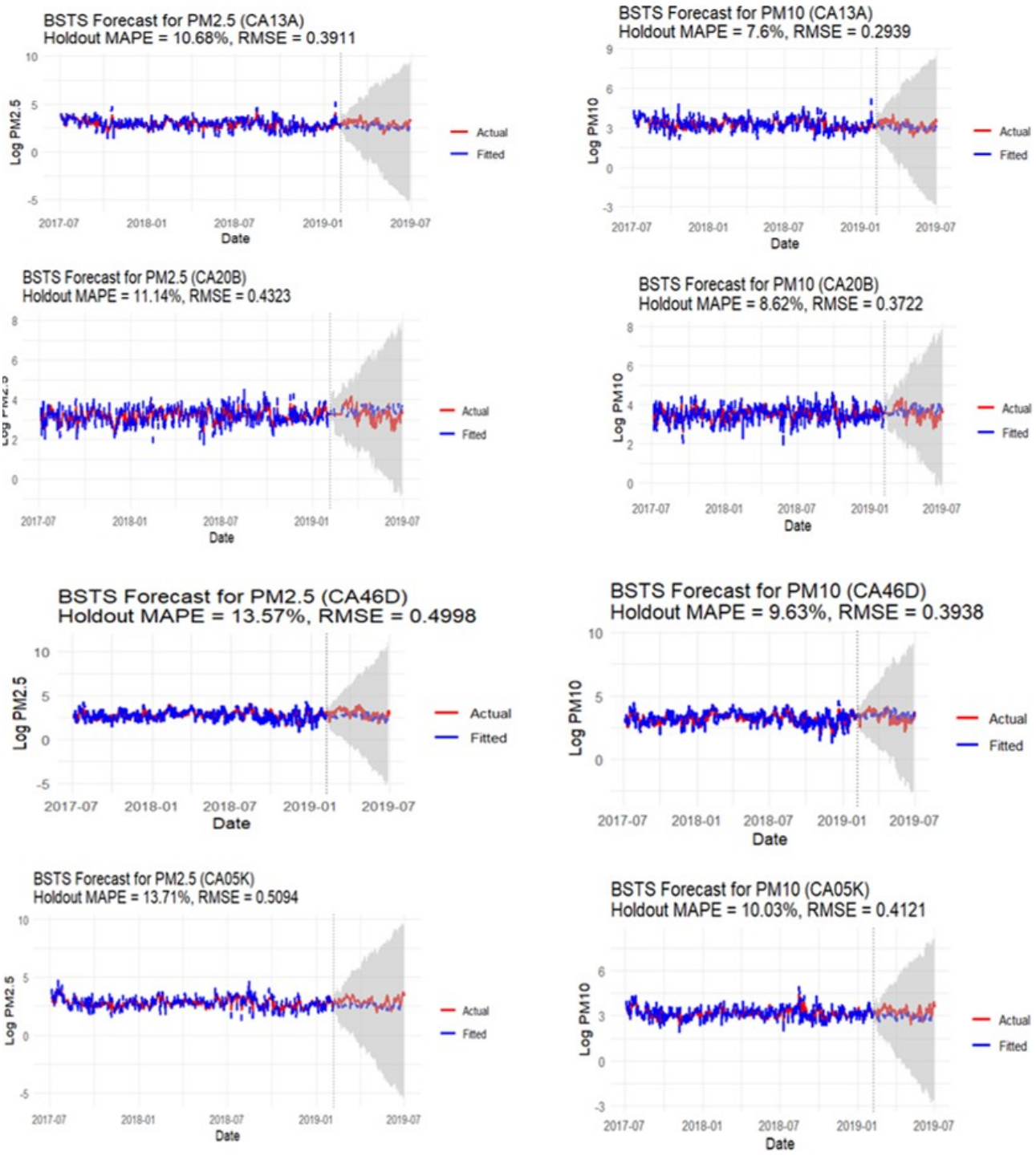

Figure 1: Comparison of Actual with Fitted Values Based on The BSTS Model with Non-Informative Prior (NIP)

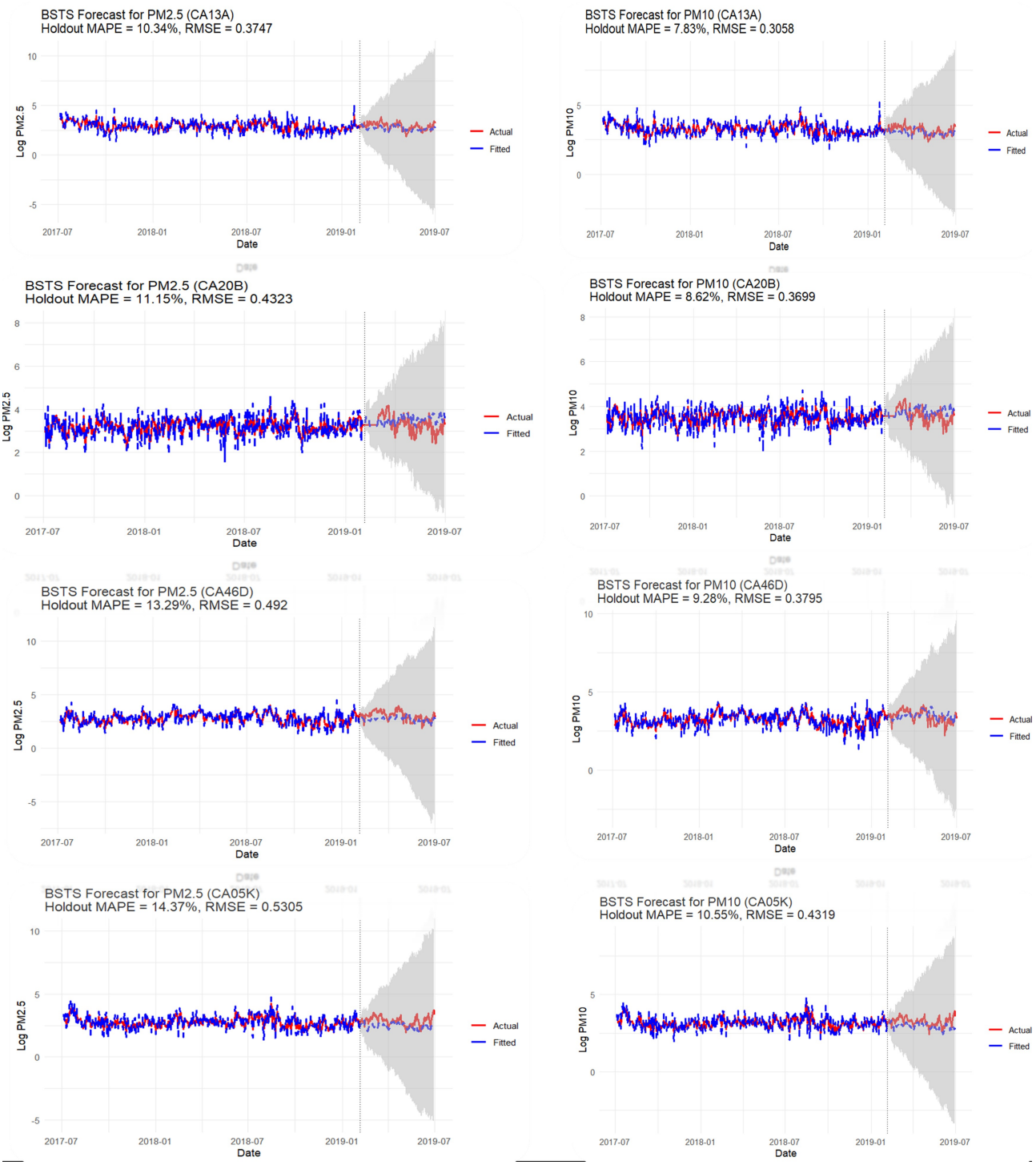

Figure 2: Comparison of Actual with Fitted Values Based on The BSTS Model with Weakly Informative Prior (WIP)

structural time series models using a non-informative prior (NIP) and a weakly informative prior (WIP) for regression coefficients.

Table 4: Core Forecast Accuracy by Using WIP and NIP for Regression Coefficients from February to July

Station	Pollutant	Model	MAPE	RMSE	Pred. sd	Obs. sd	Rel. gof	R^2
CA13A	PM2.5	NIP	10.68%	0.391	0.280	0.358	0.226	0.85
		WIP	10.34%	0.374	0.278	0.358	0.221	0.85
	PM10	NIP	7.60%	0.294	0.237	0.324	0.249	0.83
		WIP	7.83%	0.306	0.237	0.324	0.246	0.83
CA20B	PM2.5	NIP	11.14%	0.432	0.234	0.358	0.345	0.68
		WIP	11.15%	0.432	0.233	0.358	0.347	0.70
	PM10	NIP	8.62%	0.372	0.222	0.315	0.352	0.67
		WIP	8.62%	0.369	0.222	0.315	0.347	0.69
CA46D	PM2.5	NIP	13.57%	0.499	0.263	0.439	0.206	0.90
		WIP	13.29%	0.492	0.262	0.439	0.120	0.90
	PM10	NIP	9.63%	0.394	0.222	0.382	0.324	0.86
		WIP	9.28%	0.379	0.228	0.382	0.282	0.87
CA05K	PM2.5	NIP	13.71%	0.509	0.264	0.355	0.104	0.88
		WIP	14.37%	0.530	0.257	0.355	0.148	0.90
	PM10	NIP	10.03%	0.412	0.210	0.275	0.158	0.84
		WIP	10.55%	0.432	0.206	0.275	0.173	0.86

Table 4 reports that in general, WIP does better than NIP in terms of model performance, which we see out of key metrics like observed standard deviation, prediction standard deviation, R-squared and relative goodness of fit. In the BSTS framework, we use Harvey's [23] relative goodness of fit, which is an adapted version of the traditional R-squared for time series models. This stat compares the sum of squared prediction errors to a baseline random walk with constant drift and will be negative if the model does worse than the baseline.

At CA13A for PM2.5, WIP lowers MAPE by 0.34% (from 10.68 to 10.34%) and reduces RMSE by 0.017 (0.391 to 0.374), while R^2 remains stable at 0.85 and relative gof improves slightly from 0.226 to 0.221. For CA13A-PM10, prediction sd stays the same, 0.237, and remains constant at 0.83. At CA20B, WIP has minor effects for PM2.5, prediction Sd decreases slightly from 0.234 to 0.233 with MAPE unchanged, while R^2 increases by 0.02. For PM10, RMSE decreases marginally from 0.371 to 0.369, and increases by 0.02% (0.67 $\rightarrow$ 0.69). At CA46D, WIP reduces MAPE by 0.28% (13.57 to 13.29%) and RMSE by 0.007 (0.499 to 0.492) for PM2.5, while for PM10, prediction Sd rises slightly from 0.222 to 0.228 and R^2 increases by 0.01%. At CA05K for PM2.5, WIP shows mixed effects; MAPE increases (13.71% $\rightarrow$ 14.37%), RMSE increases (0.509 $\rightarrow$ 0.530), prediction sd decreases slightly (0.264 $\rightarrow$ 0.257), and R^2 rises from 0.88 to 0.90. For PM10, MAPE increases (10.03% $\rightarrow$ 10.55%), RMSE rises (0.412 $\rightarrow$ 0.432), prediction sd reduces slightly (0.210 $\rightarrow$ 0.206), and R^2 increases from 0.84 to 0.86.

Slight progressive improvements under weakly informative priors (WIP) may be understood through the processes of prior shrinkage that stabilize coefficient estimation in the BSTS setup. Due to the introduction of soft constraints, WIP avoids overfitting the local noise while still allowing the data to drive posterior estimates. This has been reflected in the low MAPE reductions, combined with moderate but systematic R^2 increases across most stations. The one exception of CA05K likely captures site-specific dynamics wherein the shrinkage that was applied weakly may have curtailed the model flexibility, thus marginally adding to the forecast error. Still, the overall pattern suggests that WIP indeed enhances generalization and thus provides more robust forecasts against NIP, albeit with less pronounced effect sizes.

The choice between WIP and NIP models would depend on the data and forecasting objectives. It is the WIP models that work best in datasets exhibiting moderate levels of variability, where such models help to strike a balance between tracking the underlying trend and avoiding overfitting. In contrast, NIP models are designed for datasets that exhibit significant fluctuations, where capturing extreme variations requires flexibility. Thus, WIP outperforms in accounting for variability and forecasting accuracy in most stations and pollutants, hence being the more reliable choice for an enhanced forecast.

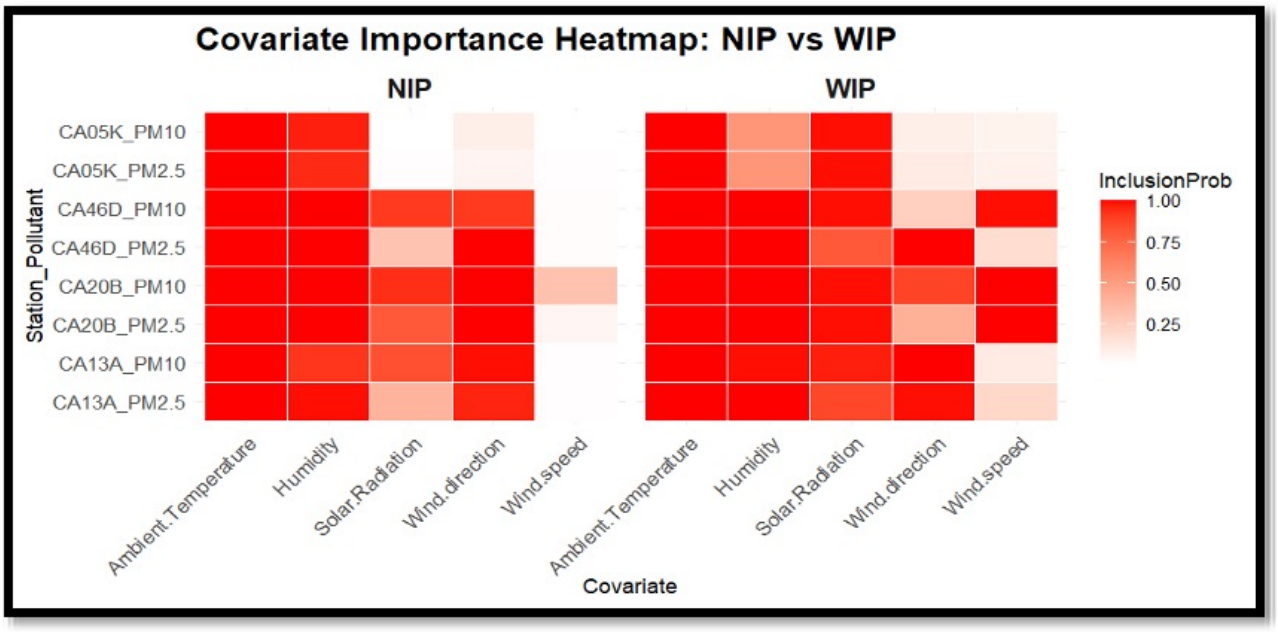

Figure 3: Covariate Importance Heat-map of the Models in the Case of Noninformative Prior (NIP), and Weakly Informative Prior (WIP)

Figure 3 illustrates how environmental factors affect PM2.5 and PM10 air pollution at various monitoring stations. They demonstrate the importance of predictors and the distinctions between non-informative and weakly informative priors in Bayesian structural time series models. The comparison between the NIP and WIP models reveals notable differences in coefficient estimates for PM2.5 and PM10 across various stations.

Comparative representation, as shown via Figure 3 of NIP and WIP conditional prior models, exhibited both stable and greatly differing shifts in covariate importances. Temperature remains the most dominant predictor of all stations and pollutants in Malaysia (inclusion prob-

ability = 1), indicating that it has a fairly strong and consistent impact regardless of the choice of prior.

Humidity did show some strong and fairly stable importance, but it was even more emphasized under WIP, especially at the CA05K site, where inclusion probabilities showed a steep increase. Solar radiation, having had weak or moderate importance under NIP, became really important in the WIP models at several stations, hence indicating that the weakly informative prior brought more impetus to include secondary covariates. Similarly, for wind direction and wind speed, under the NIP prior, often considered negligible, they become moderately important under the WIP model, particularly at CA05B and CA46D. Overall, while both prior settings would attest to the crucial role played by temperature and humidity, confirming that the choice of prior does not drastically change which covariates are important, the WIP specification acts towards a more balanced model through enhancing the inclusion of less dominant predictors.

This study has several limitations. First, the series length is relatively short, limiting the model's ability to fully capture long-term trends or rare events. Second, the results are regionally representative, focusing only on selected stations in Malaysia, and may not generalize to areas with different pollution sources or meteorology. Third, unmeasured emission events, such as industrial accidents, open burning, or sudden traffic surges, are not included in the covariates, which can reduce forecast accuracy and increase uncertainty. Also, to evaluate the adequacy of the model, we carried out post-estimation diagnostics on the BSTS residuals. The residual autocorrelation function (ACF) confirmed that no significant autocorrelation could be found across lags, suggesting that most of the temporal dependencies were well captured by the models in the data.

4 Conclusion

This study used Bayesian structural time series (BSTS) analysis to explore air quality dynamics and discover insights into the relationship between environmental factors and particulate matter concentrations. The study focused on PM2.5 and PM10 pollutants across four monitoring stations in Malaysia, from July 5th, 2017, to June 30th, 2019. By incorporating explanatory variables such as ambient temperature, solar radiation, humidity, wind speed, and wind direction, we were able to gain a nuanced understanding of pollutant behaviour. The use of Bayesian methods allowed for estimating uncertainty in model parameters, providing a more realistic representation of air quality dynamics.

The results indicated that BSTS models successfully predicted air pollution levels. Models with weakly informative priors (WIP) exhibited slightly better performance in terms of MAPEs and RMSEs than non-informative priors (NIP). Both approaches indicate that PM10 is more predictable than PM2.5 at all stations. For instance, at station CA13A, for PM2.5, WIP lowers MAPE by 0.34% and reduces RMSE by 0.017, and relative gof improves slightly from 0.226 to 0.221. For PM10, Relative. gof improves slightly from 0.249 to 0.246. Despite discrepancies between forecasted and out-of-sample observed values due to inherent uncertainties in air quality modelling, WIP models aligned more closely with observed trends, capturing key variability more effectively. Consistent environmental predictors, particularly ambient temperature, humidity, emerged as key drivers affecting both PM2.5 and PM10 levels across most stations, as shown in Figure 3.

The study highlights the importance of prior specifications in Bayesian modelling for air pollution forecasting and management. Future research could refine prior specifications, explore additional environmental variables, and assess the BSTS model's robustness across different regions and pollutant types, contributing to more accurate air quality predictions and aiding policymakers in mitigating pollution impacts.

Acknowledgements

The authors would like to acknowledge the Ministry of Higher Education Malaysia (FRGS/1/2024/ICT06/UKM/02/2) and the Universiti Kebangsaan Malaysia for the allocation of research grants and infrastructures used in this study.

References

[1] World Health Organization (WH0). *Air Pollution*, World Health Organization, London, 2021.

[2] United States Environmental Protection Agency (US EPA). *Particulate Matter (PM) Pollution*, 2021.

[3] Chowdhury, S. and Dey, S. Cause-specific premature death from ambient $PM_{2.5}$ exposure in India, *Environment International*, 91 (2016), 283–290.

[4] Department of Environment Malaysia. *Air Quality Monitoring in Malaysia*, 2023.

[5] Bakar, M. A., Ariff, N. M., Bakar, S. A., Chi, G. P. and Rajendran, R. Peramalan kualiti udara menggunakan kaedah pembelajaran mendalam rangkaian perlingkaran temporal (TCN), *Sains Malaysiana*, 51(11) (2022), 3785–3793.

[6] Cho, K. M., Abd Rahman, N. H. and Che Ilias, I. S. Performance of Levenberg–Marquardt neural network algorithm in air quality forecasting, *Sains Malaysiana*, 51(8) (2022), 2645–2654.

[7] Bakar, M. A. A., Ariff, N. M., Nadzir, M. S. M., Wen, O. L. and Suris, F. N. A. Prediction of multivariate air quality time series data using long short-term memory network, *Malaysian Journal of Fundamental and Applied Sciences*, 18(1) (2022), 52–59.

[8] Zheng, Y., et al. Assessing the impacts of climate variables on long-term air quality trends in Peninsular Malaysia, *Science of the Total Environment*, 901 (2023), 166430.

[9] Ariff, N. M., Bakar, M. A. A. and Lim, H. Y. Prediction of PM_{10} concentration in Malaysia using k-means clustering and LSTM hybrid model, *Atmosphere*, 14(5) (2023), 853.

[10] Scott, S. L. and Varian, H. R. Predicting the present with Bayesian structural time series, *International Journal of Mathematical Modelling and Numerical Optimisation*, 5(1–2) (2014), 4–23.

[11] George, E. and McCulloch, R. Approaches for Bayesian variable selection, *Statistica Sinica*, 7(2) (1997), 339–373.

[12] Durbin, J. and Koopman, S. J. A simple and efficient simulation smoother for state space time series analysis, *Biometrika*, 89(3) (2002), 603–615.

[13] Madigan, D. and Raftery, A. E. Model selection and accounting for model uncertainty in graphical models using Occam's window, *Journal of the American Statistical Association*, 89(428) (1994), 1535–1546.

[14] Mohammed, A., Bakar, M. A. A., Mansor, M. M. and Ariff, N. M. Modelling Malaysia air quality data using Bayesian structural time series models, *Sains Malaysiana*, 53(11) (2024), 3817–3829.

[15] Jeffreys, H. An invariant form for the prior probability in estimation problems, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186(1007) (1946), 453–461.

[16] Gelman, A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper), *Bayesian Analysis*, 1(3) (2006), 515–534.

[17] Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N. and Scott, S. L. Inferring causal impact using Bayesian structural time-series models, *The Annals of Applied Statistics*, 9(1) (2015), 247–274.

[18] Kalman, R. E. A new approach to linear filtering and prediction problems, *ASME Journal of Basic Engineering*, 82(1) (1960), 35–45.

[19] Scott, J. G. and Berger, J. O. Bayes and empirical Bayes multiplicity adjustment in the variable-selection problem, *The Annals of Statistics*, 38(5) (2010), 2587–2619.

[20] Zellner, A. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In Goel, P. K. and Zellner, A. (Eds.), *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, North-Holland Publishing Co., Amsterdam, 1986, 233–243.

[21] Diaconis, P. and Ylvisaker, D. Conjugate priors for exponential families, *The Annals of Statistics*, 7 (1979), 269–281.

[22] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. *Bayesian Data Analysis*, 3rd ed., Chapman and Hall/CRC, New York, 2013.

[23] Harvey, A. C. *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, 1989.