



The impact of El Niño Southern Oscillation on space time PM₁₀ levels in Peninsular Malaysia: the hierarchical spatio-temporal autoregressive models approach

Maizatul F. Zulkifli¹ · Rossita M. Yunus¹

Received: 10 November 2020 / Accepted: 25 January 2022 / Published online: 12 February 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract

Hierarchical spatio-temporal autoregressive models are useful to understand the impact of predictors on a spatio-temporal-dependent variable. This study aims to fit the model to monthly PM₁₀ concentration using potential predictors from 33 monitoring stations within Peninsular Malaysia from 2006 to 2015 and predict the space–time data spatially and temporally. Using Monte Carlo Markov Chain (MCMC), spatial predictions are obtained based on the posterior and predictive distributions of the model. The posterior distribution of the model that is without covariates exhibits a strong temporal correlation between successive months and also a strong spatial correlation with an effective range of 300 km. Spatio-temporal models were fitted to the data with a sine term, a cosine term, and a lagged El Niño Southern Oscillation (ENSO) index as predictors. Of the 33 monitoring sites, 8 were selected randomly for validation sets. The predictions and forecasts are validated using the root mean square error (RMSE), the mean absolute error (MAE), and the predictive model choice criteria (PMCC). The model with a sine term and a cosine term as predictors produces a reasonable RMSE, MAE, and PMCC of 7.23, 5.91, and 114.54, respectively. It is lower compared to those of the other models. The coverage percentage of the forecast 5–95 percentile range is 89.2% implying good prediction results. The results also show that none of the ENSO indices has a significant impact on the spatial distribution of the PM₁₀ concentration.

1 Introduction

Particulate matter PM is a mixture of fine particles and liquid droplets suspended in the air, consisting of some components such as reactive gases, metals, and carbon core particles (Olague 2017) that are related to mortality beyond certain concentration level. The particle size range 10 microns or smaller (PM₁₀) can cause damage to the upper respiratory tract (El Morabet 2019). High concentrations of PM₁₀ in the atmosphere are responsible for adverse health issues (Al-Hemoud et al. 2018; Asl et al. 2018; Feng et al. 2019; Yin et al. 2019), reduction in visibility (Jeensorn et al. 2018; Won et al. 2020), and damage to materials. Since the

2000s, researchers have paid increasing attention to the role of the geographical locations (Manga and Awang 2018), seasonality (Juneng et al. 2011), source of pollutants (Alifa et al. 2020), and meteorological or climatological conditions in modeling the PM₁₀ level in Malaysia. Due to the trend and spatial variability, the PM₁₀ concentration in some parts of Malaysia has exceeded the Recommended Malaysian Air Quality Guidelines (RMAQG) of 150 µg per cubic meter (µg/m³) for daily mean concentration. In this paper, we use the spatio-temporal modeling of the PM₁₀ concentration with global climatological predictors to understand the trend and spatial variability of PM₁₀ levels.

For the statistical modeling, we require attributes that are supposed relevant in conjunction with the atmospheric fields related to the transportation of the particles after emission from its primary origin (Wie and Moon 2017). Thus, attributes related to seasonality may play a role in modeling the PM₁₀ levels. In modelling seasonality, researchers fit the meteorological data with a harmonic seasonal model using the sine and cosine functions as predictors in the model (eg. Yunus et al. 2017; Hasan and Dunn 2012). A few studies have found an association between various global

Responsible Editor: Emilia Kyung Jin.

✉ Rossita M. Yunus
rossita@um.edu.my
Maizatul F. Zulkifli
fashahah.mai@gmail.com

¹ Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

climatological variables and PM_{10} concentration (eg. in Wie and Moon 2017; Sentian et al. 2018; Hassan et al. 2020; Kim et al. 2013; Kim et al. 2019) using well-known statistical methods such as the Pearson or Spearman's correlation coefficient, multiple linear regression, and lead-lag correlation analysis. However, most researchers do not take account of both spatial and temporal properties simultaneously in the methods and models.

The seasonal peaks in the concentration of PM_{10} in Malaysia occur between every June and September and coincide with the southwest monsoon season (Juneng et al. 2009; Noor et al. 2015; Yusof et al. 2009). The strength of the monsoon season depends on the local and sea thermal contrast, with preconditioning by the monsoon air temperatures over land play an important role. The El Niño Southern Oscillation (ENSO) can modulate Malaysia's rainfall regime, with El Niño (La Niña) events corresponding to low (high) rainfall seasons (Chen et al. 2002; Singhrattna et al. 2005). The El Niño events are concurrent with rainfall deficits and dry weather in Malaysia and other countries in Southeast Asia (Sum 2018). During the period of El Niño events of 1997–1998 and 2014–2015, the transportation of dust particles emitted from massive biomass burning in Sumatra and Kalimantan has led to a sharp increase in the levels of PM_{10} concentration in Malaysia (Sentian et al. 2018). As reported, the high levels of PM_{10} coincide with the El Niño events in the country (Shaadan et al. 2015; Sentian et al. 2018).

To investigate the impact of global climatological variables on the level of PM_{10} concentrations in Peninsular Malaysia, five ENSO indicators, namely NINO12, NINO3, NINO 34, NINO4, Southern Oscillation Indicator (SOI), and ENSO precipitation index (ESPI), are considered for the analysis. The NINO12, NINO3, NINO34, and NINO4 indices are the monthly mean sea surface temperature (SST) anomalies averaged over the region (0° – 10° S, 90° – 80° W), (5° N– 5° S, 150° – 90° W), (5° N– 5° S, 170° – 120° W) and the (5° N– 5° S, 160° E– 150° W) in the Pacific, respectively. SOI is the normalized difference in surface pressure between Tahiti and Darwin. It represents the strength of trade winds that are associated with the flow from high- to low-pressure regions. An ENSO rainfall-based index, known as the ENSO precipitation index (ESPI), is based on rainfall anomalies average measured in two rectangular areas, one in the eastern tropical Pacific (10° N– 10° S and 160° E– 100° W) and the other over the Maritime Continent (10° S– 10° N and 90° E– 150° E). It is reported that ESPI is well correlated with sea surface temperature and pressure indices (e.g. NINO 34 and SOI) (Curtis and Adler 2000). Correlations in between NINO12, NINO3, NINO34, and NINO4 indices were reported moderate to strong (Rehman et al. 2012).

The data we are modeling is inherently both spatial and temporal in nature, consisting of PM_{10} levels measurements taken over time at various locations within Peninsular

Malaysia. An overview of spatio-temporal data and modeling of spatio-temporal data can be found in Cressie and Wikle (2011) and Banerjee et al. (2015). Some relevant and recent works in spatio-temporal modeling includes Benth and Šaltytė (2011), Nowak et al. (2018), with specific applications to PM_{10} data given in Al-Awadhi and Al-Awadhi (2006), Cocchi et al. (2007) and Pollice and Lasinio (2009). Most current works undertaken in spatio-temporal adopt a Bayesian approach, as the hierarchical nature of the models naturally suitable to this framework. The spatio-temporal model that we use to model the PM_{10} spatio-temporal data is the hierarchical Bayesian autoregressive model proposed by Sahu (2012) to space–time environmental data. It is known as the space–time autoregressive model under the Bayesian hierarchical setup. The model consists of the autoregressive term, the regression term, the Gaussian spatially correlated error term, and the Gaussian non-spatial error term. The model appears to fit better various space–time environmental data (Camaletti et al. 2011; Mukhopadhyay 2019; Manga and Awang 2018) than the other models such as the simple linear regression, the Bayesian linear regression, the Bayesian kriging-based model, and the Gaussian process model, that ignore both space and time simultaneously. Manga and Awang (2018) have considered regional factors for the prediction of the PM_{10} level using the spatio-temporal model. With a different aim, we consider lagged global factors in our work.

The objective of the research in writing this paper is to quantify the impact of ENSO indicators on the space–time PM_{10} concentration levels in Peninsular Malaysia using the hierarchical spatio-temporal model approach. The contents of this paper are structured as follows. The following section explains the data used in this study. In Sect. 3, we describe the hierarchical spatio-temporal models considered for the analysis. Section 4 gives the results of fitting models on the space–time PM_{10} data and the diagnostic validation to evaluate how well the model fits the data. The last section presents the conclusions of this study.

2 Data

Monthly average of PM_{10} levels from 33 stations in Peninsular Malaysia (see Fig. 1) recorded from January 2006 to December 2015 are studied (data obtained from the Department of Environmental Malaysia, available on the Malaysia Open Data Portal website <https://www.data.gov.my/>).

The distribution of average monthly PM_{10} for the 33 stations varies across stations (see Fig. 2). The monthly average of PM_{10} concentrations ranges from the minimum of $15.68 \mu\text{g}/\text{m}^3$ in Tanjung Malim (in the northern) to the maximum of $134.15 \mu\text{g}/\text{m}^3$ in Pelabuhan Klang (in the central region of Peninsular Malaysia). The central zone has a

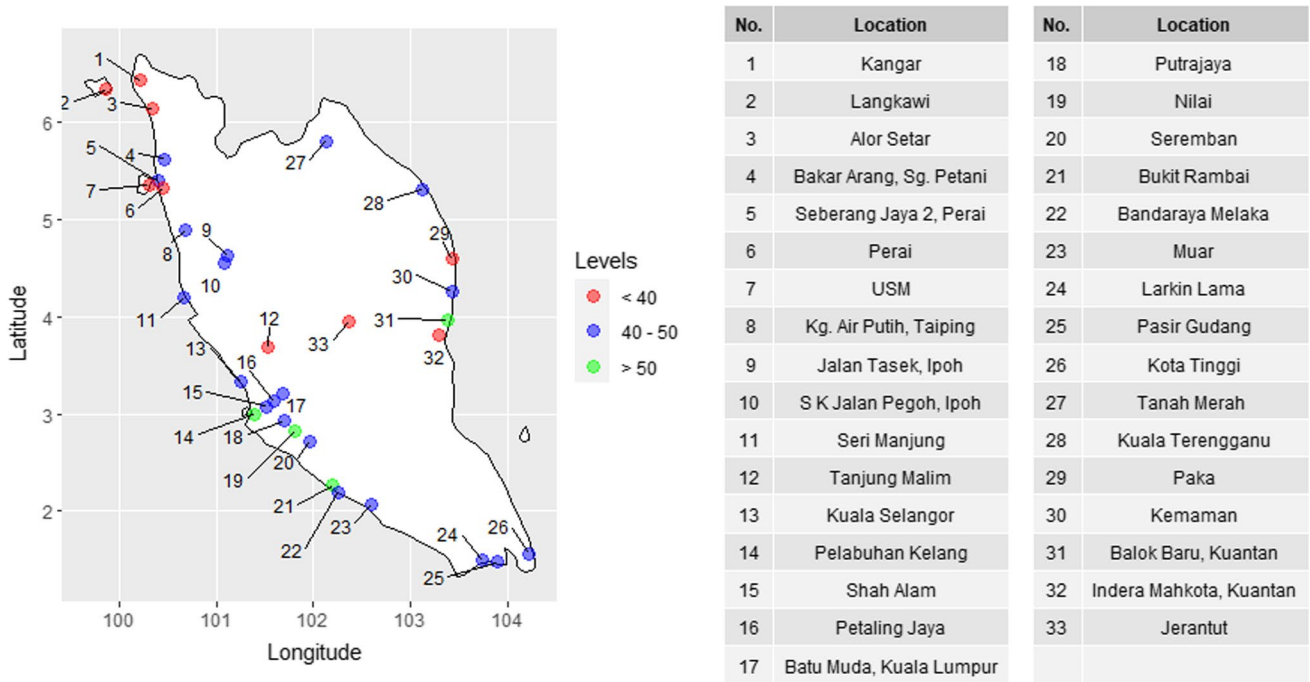


Fig. 1 Spatial distribution of the median of the monthly average of PM₁₀ levels taken over ten years for thirty-three studied stations

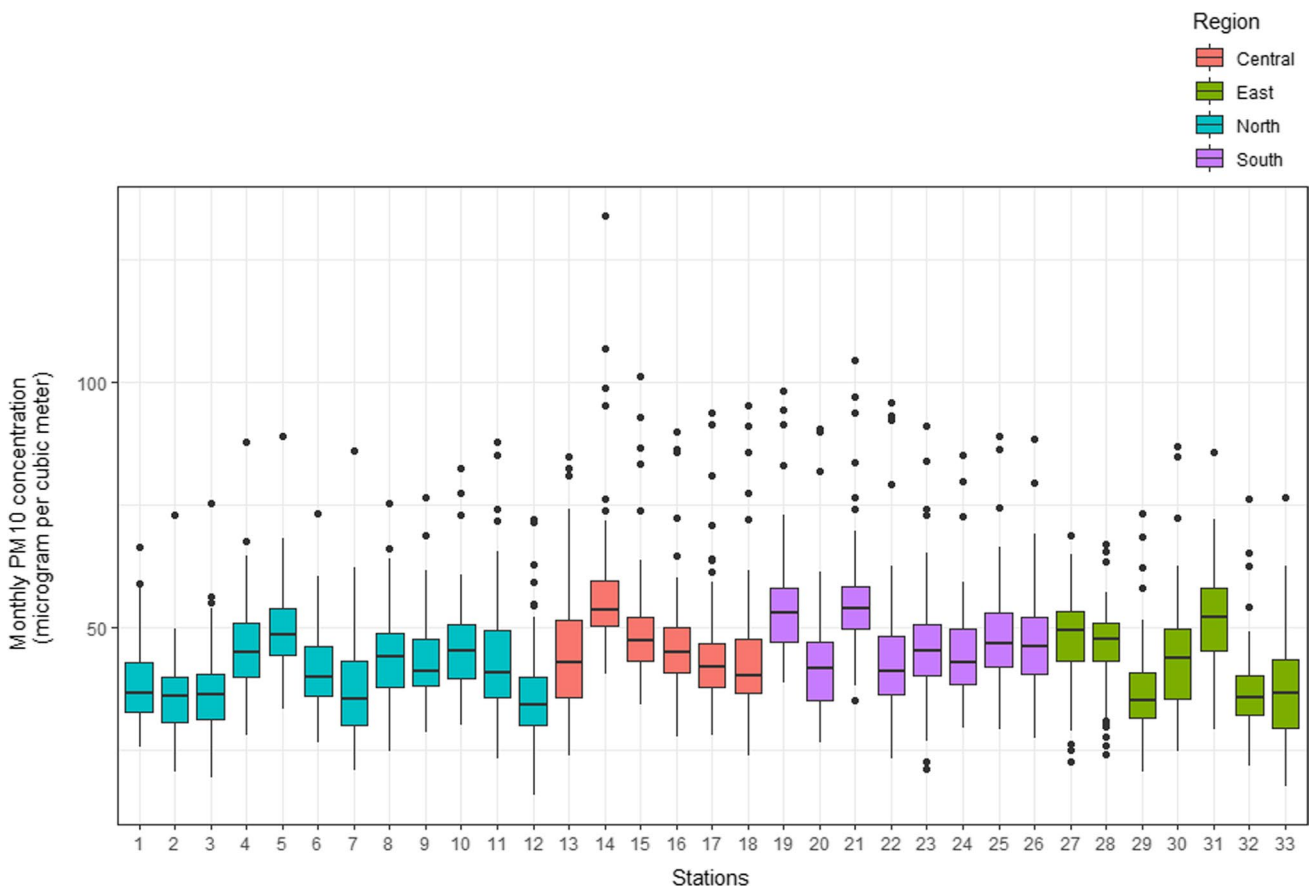


Fig. 2 Boxplots of monthly average PM₁₀ concentration levels for all studied stations

broader range of monthly PM_{10} concentrations than in other places. The median of the monthly PM_{10} levels of about 61% of the stations ranged between 40 and 50 $\mu g/m^3$, 12% is above 50 $\mu g/m^3$, and 27% is below 40 $\mu g/m^3$ (see Fig. 1).

The annual PM_{10} values range between 40.67 and 50.06 $\mu g/m^3$ (see Fig. 3) and below the annual limit value (50 $\mu g/m^3$) set by Malaysia Air Quality Guideline from 1994 to 2014, and slightly exceeds the limit for the year 2015. However, they are above the annual air quality guideline set by European and national legislation (40 $\mu g/m^3$) and WHO

(20 $\mu g/m^3$). The variation in the monthly average concentration of PM_{10} across time for all stations shows a cyclical pattern. Generally, the monthly PM_{10} concentrations were higher from June to September than in the other months.

The boxplots for the monthly PM_{10} concentration distribution (Fig. 4) show the average monthly PM_{10} levels peak from every May to September. This peak period coincides with the southwest monsoon (dry season), which is characterized by low rainfall and less cloud. Rainfall is received in most parts of Malaysia through the northeast monsoon (wet season) from

Fig. 3 Annual and monthly average PM_{10} concentration temporal trend

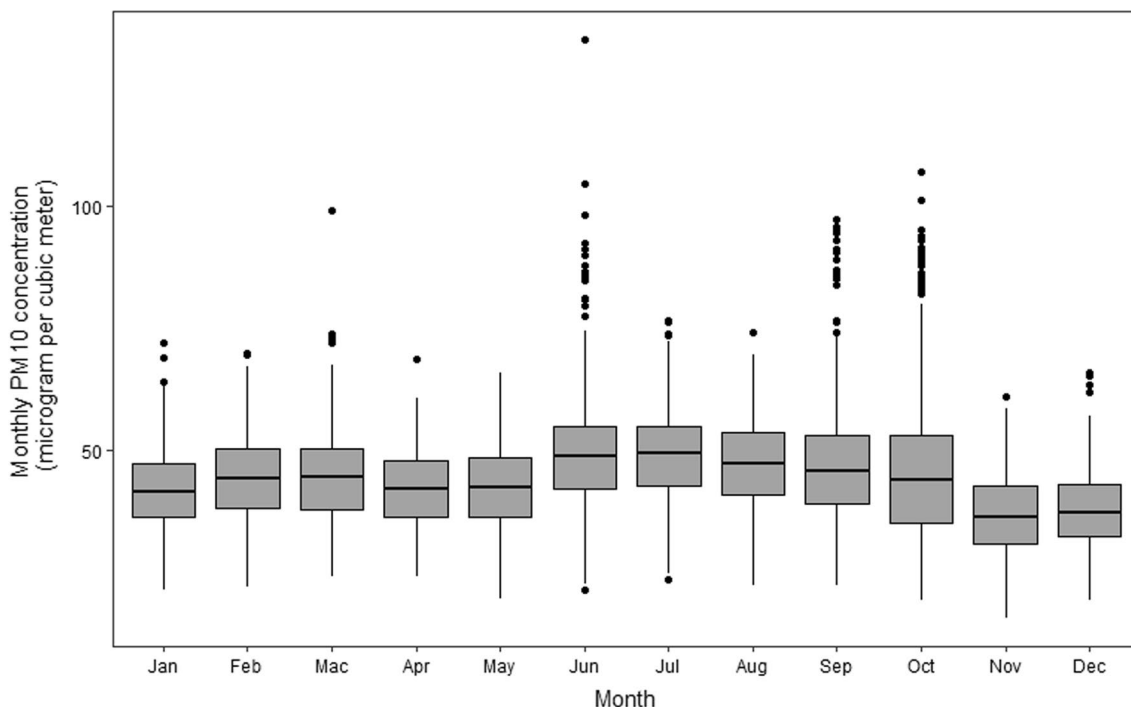
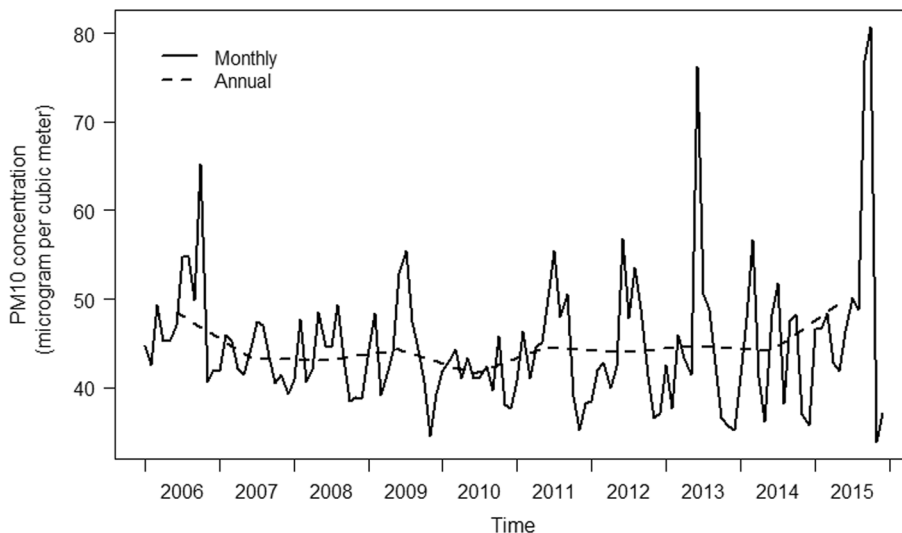


Fig. 4 Boxplots of the monthly average PM_{10} concentration in all stations from January 2006 to December 2015

every October to December. Peninsular Malaysia is experiencing lower PM₁₀ levels during the northeast monsoon across the months of the year. The ENSO can modulate Malaysia’s rainfall regime, with El Niño (La Niña) events corresponding to low (high) rainfall seasons (Chen et al. 2002; Singhrattna et al. 2005).

NINO12, NINO3, NINO34, and NINO4 indices are the average sea surface temperature (SST) anomaly in bounded regions across the Pacific. Since the region has large variability on El Niño time scales, thus is used by some authors to understand its impact on PM₁₀ levels. The monthly NINOs is available on the <https://psl.noaa.gov/enso/data.html> website. The ESPI is recorded based on rainfall anomalies in two rectangular areas, one in the eastern tropical Pacific (10° N–10° S and 160° E–100° W) and the other over the Maritime Continent (10° S–10° N and 90° E–150° E). The monthly ESPI data collected are obtained from http://eagle1.umd.edu/GPCP_ICDR/Data/ESPI.txt website. SOI, the standardized fluctuations in the air pressure difference between Tahiti and Darwin (Troup 1965) is associated with the strength of Pacific trade winds. Sustained negative values of SOI indicates El Niño events associated with warmer in the surface waters in the Equatorial Pacific Ocean) and positive values indicate the La Niña episodes (cooler ocean temperature in the Equatorial Pacific). The SOI data are available on https://psl.noaa.gov/gcos_wgsp/Timeseries/Data/soi.long.data website.

Lagged values of the climatological variables, the value of the climatological variables of the previous month is used to model the PM₁₀ concentration of a given month.

3 Statistical models

3.1 Hierarchical Bayesian autoregressive spatio-temporal model

The hierarchical Bayesian autoregressive spatio-temporal model will be used to analyze the PM₁₀ levels in the study region.

Let $Z(s_i, t)$ be the log-transformed of the monthly average of PM₁₀ values at location s_i for time t , where the spatial reference s_i is a two-dimensional vector represents the latitude–longitude pair, $i = 1, \dots, n$, and $t = 1, \dots, T$. Here, n is the total number of monitoring sites, and T is the total number of months throughout the study.

The hierarchical model of the first level is given by

$$Z(s_i, t) = O(s_i, t) + \epsilon(s_i, t), \tag{1}$$

where $O(s_i, t)$ represents the true underlying spatio-temporal and $\epsilon(s_i, t)$ represents the error term that is normal $N(0, \sigma_\epsilon^2)$. Here, σ_ϵ^2 is the unknown pure error variance or nugget effect that describes the variation in space and $\epsilon(s_1, t), \epsilon(s_2, t), \dots, \epsilon(s_n, t)$ are independently and identically distributed.

For the second level of the hierarchical model, following Sahu et al. (2007), the true spatial temporal process $O(s_i, t)$ with autoregressive terms is specified by:

$$O(s_i, t) = \rho O(s_i, t - 1) + \mathbf{x}(s_i, t)' \boldsymbol{\beta} + \eta(s_i, t), \tag{2}$$

where $\boldsymbol{\beta}$ is a p dimensional vector of regression terms associated with covariates $\mathbf{x}(s_i, t) = (x_1(s_i, t), x_2(s_i, t), \dots, x_p(s_i, t))'$. The model specifies the first-order autoregressive term $\rho O(s_i, t - 1)$ to describe dependence on past values. Temporal correlation parameter denoted by ρ describes the strength of temporal dependence of PM₁₀ levels between successive months. The spatial temporal random effect is denoted by $\eta(s_i, t)$ and $\eta(s_i, t = 1), \eta(s_i, t = 2), \dots, \eta(s_i, t = T)$ are independent (in time) and $\eta(s_i, t)$ follows a Gaussian Process (GP) assumption, thus $\eta(s_i, t) \sim N(0, \boldsymbol{\Sigma}_\eta)$, where $\boldsymbol{\Sigma}_\eta = \sigma_\eta^2 \mathbf{S}_\eta$ for $t = 1, 2, \dots, T$. Here, σ_η^2 denotes the site invariant spatial variance and \mathbf{S}_η is the spatial correlation matrix.

In spatial modeling, the correlation is driven by the distance between locations. That is, two units that are located near each other are more highly correlated. The spatial correlation matrix \mathbf{S}_η is constructed from a positive definite function (p.d.f). A flexible and popular p.d.f. is the Matern correlation function. This function is indexed by the scale and smoothness parameters denoted by ϕ and ν (Handcock and Stein 1993), respectively. Following several authors, this study chooses the most used form of the Matern class of function which is the exponential covariance function. In the exponential case, when the degree of spatial smoothing $\nu = 0.5$. the covariance, κ depends on the distance between site s_i and s_j and ϕ and it is given by

$$\kappa(\|s_i - s_j\|, \phi) = \exp(-\phi \|s_i - s_j\|), \tag{3}$$

where ϕ denotes the rate of decay of the correlation as the distance $\|s_i - s_j\|$ increases. The initial condition of the autoregressive model $O(s_i, 0)$ is $O_0 \sim N(\mu, \sigma^2 \mathbf{S}_0)$ where \mathbf{S}_0 is the correlation matrix obtained using the Matern correlation function in Eq. (3).

Then, the specification of the model is completed by assigning proper prior distributions for all the parameters of this model denoted by $\boldsymbol{\theta} = (\sigma_\epsilon^2, \sigma_\eta^2, \boldsymbol{\beta}, \rho, \phi, \nu)$. The prior distribution for the parameters $\boldsymbol{\beta}$ and ρ is specified through normal distributions $N(0, 10^4)$. In this study, inverse Gamma distribution $IG \sim (a, b)$ is assumed for the variance parameters σ_ϵ^2 and σ_η^2 with hyper-parameters $a = 2$ and $b = 1$. The prior used for the spatial decay parameter ϕ is Gamma prior (2, 1) and random-walk Metropolis–Hastings is adopted for sampling ϕ from its full conditional distribution.

In the Bayesian framework, inference on model parameters is based on posterior distribution of the process and

parameters given the data. The posterior distribution can be obtained by using Bayes theorem;

$$P(\theta|Z(s, t)) \propto l(\theta|Z(s, t))\pi(\theta), \tag{4}$$

where $l(\theta|Z(s, t))$ is the likelihood function and $\pi(\theta)$ denotes the prior distribution of θ .

However, because of the complexity of the probability distribution, Markov Chain Monte Carlo (MCMC) techniques are usually used to approximate the posterior distribution. The technique used for this study is Gibbs sampling where the approximation of posterior distribution is done by simulating sequence of samples from the full conditional distributions of the parameters in the model. Furthermore, prediction of monthly PM₁₀ levels at any unobserved location s_0 and at a future time point t' can be obtained based on Sahu (2012) from the posterior predictive distribution of $Z(s_0, t')$ given z :

$$P(Z(s_0, t')|z) = \int \left\{ \frac{P(Z(s_0, t')|O(s_0, t'), \sigma_\epsilon^2)P(O(s_0, t')|\theta, \mathbf{O}, z^*) \cdot P(\theta, \mathbf{O}, z^*|z)}{P(\theta, \mathbf{O}, z^*|z)} d\mathbf{O} d\theta dz^* \right\}, \tag{5}$$

where $Z(s_0, t')$ follows the normal distribution $N(O(s_0, t'), \sigma_\epsilon^2)$ and $\theta = (\beta, \rho, \sigma_\epsilon^2, \sigma_\eta^2, \phi)$, z^* and z denote the vectors of missing and observed data, respectively. The integral in Eq. (5) is also evaluated by MCMC method and R package spTimer (Bakar and Sahu 2015) is used to obtain the posterior density estimates and predictions.

3.2 Models for the monthly average PM₁₀ and model assessment

Many studies have shown a high correlation between any two global climatological variables. To quantify the impact of the global climatological variable on monthly average of PM₁₀ levels, we consider seven hierarchical spatio-temporal autoregressive models. The linear predictor for.

- Model 1, $x(s_i, t) \beta = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12)$,
- Model 2, $x(s_i, t) \beta = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12) + \beta_3 \text{NINO}_{12-i}$,
- Model 3, $x(s_i, t) \beta = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12) + \beta_3 \text{NINO}_{3-i}$,
- Model 4, $x(s_i, t) \beta = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12) + \beta_3 \text{NINO}_{34-i}$,
- Model 5, $x(s_i, t) \beta = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12) + \beta_3 \text{NINO}_{4-i}$,
- Model 6, $x(s_i, t) \beta = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12) + \beta_3 \text{SOI}_{-i}$,
- Model 7, $x(s_i, t) \beta = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12) + \beta_3 \text{ESPI}_{-i}$, where $m =$ month of the year ($1 =$ January, $2 =$ February, ..., and so on, $t = 12(i - 1) + m$, $i = 1$ (first year of studied data), $i = 2$ (second year of studied data), ..., so on. Model 1 is the simplest with only sine and cosine terms as predictors and is nested within the other six models. Models 2, 3, ..., 7 consider the lagged values from one of the ENSO indices, respectively, in addition to harmonic sine and cosine terms.

We use the root means square error (RMSE), mean absolute error (MAE), and predictive model choice criteria (PMCC) to determine the best fitted model. The RMSE and the MAE measure the difference between the fitted spatio-temporal model at the testing sites (s_1, \dots, s_{n_0}) (the posterior mean of the predictive distribution), and the observed space–time PM₁₀ concentration data at the validation sites (s_{n_0+1}, \dots, s_n). The RMSE and MAE are calculated as follows: $RMSE = \sqrt{\frac{1}{m} \sum_{j=n_0+1}^n \sum_{t=1}^T (Z(s_j, t) - \hat{Z}(s_j, t))^2}$ and $MAE = \frac{1}{m} \sum_{j=n_0+1}^n \sum_{t=1}^T |Z(s_j, t) - \hat{Z}(s_j, t)|$, where m is the total number of observations from the validation sites and $\hat{Z}(s_j, t)$ denotes the posterior mean of the predictive distribution. Smaller values of RMSE and MAE indicate a superior model.

The predictive model choice criteria (PMCC) is used to compare the prediction capability (Gelfand and Ghosh 1998) and it is given by,

$$PMCC = \sum_{i=1}^n \sum_{l=1}^r \sum_{t=1}^{T_i} \left\{ E(Z(s_i, t)_{rep} - z(s_i, t))^2 + Var(Z(s_i, t)_{rep}) \right\},$$

where $Z(s_i, t)_{rep}$ is a future replicate of the data $z(s_i, t)$. The first term of PMCC assesses the goodness of fit while the second term is a penalty term for complexity of the model. This value is computed and estimated by sampling the posterior predictive distribution that is given in the Eq. (5). The model with the smallest value of PMCC has better prediction capability.

4 Results and discussion

Summaries of the posterior distribution displayed in Table 1 are for the model parameters associated with sine, cosine, and ENSO index. The parameter estimates were obtained just by fitting the model to the data of the testing sites in the period 2006–2015. The Gibbs sampler is run for 5000 iterations to obtain the results. The 1000 initial iterations run as burn-in and the others for making inference on the model parameters. Each of the models is run with the same set of parameters and hyper-parameters. For validation purposes, the RMSE, MAE, and PMCC are calculated to choose the superior model.

Model 1 (without ENSO indices) has the smallest value of PMCC. Amongst the models with ENSO indices, Model 7 (with ESPI) has the lowest RMSE, Model 4 (with NINO34) has the lowest MAE, and Model 6 (with SOI) has the lowest PMCC. However, observing the 95% credible interval, the ENSO indices are not statistically significant in these models, while sine and cosine are significant for almost all models. From the posterior distribution of the model parameters

Table 1 Summaries of the posterior distributions of model parameters

Model \emptyset				Model 1			
RMSE	7.2195			RMSE	7.2253		
MAE	5.8981			MAE	5.9017		
PMCC	145.98			PMCC	144.54		
	Mean	Median	95% Credible interval		Mean	Median	95% Credible interval
Intercept	0.547	0.547	(0.473, 0.623)	Intercept	0.549	0.55	(0.475, 0.623)
				Sine	0.026	0.026	(0.003, 0.050)
				Cosine	- 0.029	- 0.029	(- 0.054, - 0.006)
Model 2				Model 3			
RMSE	7.2502			RMSE	7.2307		
MAE	5.9293			MAE	5.9071		
PMCC	145.40			PMCC	146.04		
	Mean	Median	95% Credible interval		Mean	Median	95% Credible interval
Intercept	0.192	0.196	(- 0.233, 0.618)	Intercept	0.692	0.696	(0.208, 1.168)
Sine	- 0.012	- 0.011	(- 0.061, 0.040)	Sine	0.032	0.032	(0.000, 0.064)
Cosine	- 0.050	- 0.051	(- 0.084, - 0.015)	Cosine	- 0.030	- 0.029	(- 0.054, - 0.005)
Lagged NINO12	0.015	0.015	(- 0.003, 0.034)	Lagged NINO3	- 0.006	- 0.006	(- 0.024, 0.013)
Model 4				Model 5			
RMSE	7.2285			RMSE	7.2335		
MAE	5.9034			MAE	5.9064		
PMCC	146.03			PMCC	146.45		
	Mean	Median	95% Credible interval		Mean	Median	95% Credible interval
Intercept	0.779	0.784	(0.289, 1.268)	Intercept	0.870	0.88	(0.185, 1.539)
Sine	0.029	0.029	(0.004, 0.054)	Sine	0.024	0.025	(- 0.001, 0.049)
Cosine	- 0.033	- 0.033	(- 0.059, 0.007)	Cosine	- 0.033	- 0.032	(- 0.059, - 0.007)
Lagged NINO34	- 0.009	- 0.009	(- 0.027, 0.01)	Lagged NINO4	- 0.011	- 0.012	(- 0.035, 0.013)
Model 6				Model 7			
RMSE	7.2383			RMSE	7.2268		
MAE	5.9231			MAE	5.9060		
PMCC	144.79			PMCC	144.98		
	Mean	Median	95% Credible interval		Mean	Median	95% Credible interval
Intercept	0.550	0.55	(0.477, 0.624)	Intercept	0.547	0.547	(0.474, 0.621)
Sine	0.026	0.027	(0.002, 0.051)	Sine	0.026	0.026	(0.001, 0.050)
Cosine	- 0.027	- 0.026	(- 0.051, 0.003)	Cosine	- 0.029	- 0.029	(- 0.053, - 0.005)
Lagged SOI	- 0.010	- 0.01	(- 0.026, 0.005)	Lagged ESPI	- 0.003	- 0.003	(- 0.020, 0.014)

corresponding to the predictor variables, the ENSO indices are insignificant in each model. Therefore, there is not enough evidence in concluding ENSO indices are important predictors and have a significant impact on the space-time PM₁₀ concentrations.

We observe the convergence of the model parameters using the trace-plot given in Fig. 5. The result is satisfactory. As the number of iterations increases, the results showed

that the increment does not improve the convergence of the chain.

All regression coefficients of Model 1 are statistically significant as the 95% credible intervals do not contain zero (see Table 1). The estimate of the spatial variance parameter σ_{η}^2 (0.0396) is substantially higher than that of the error (non-spatial) variance parameter σ_{ϵ}^2 (0.0057). This result shows that the ability of the model to explain the variability

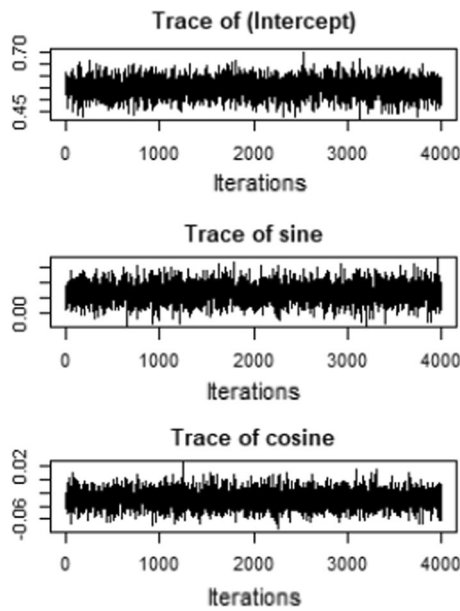
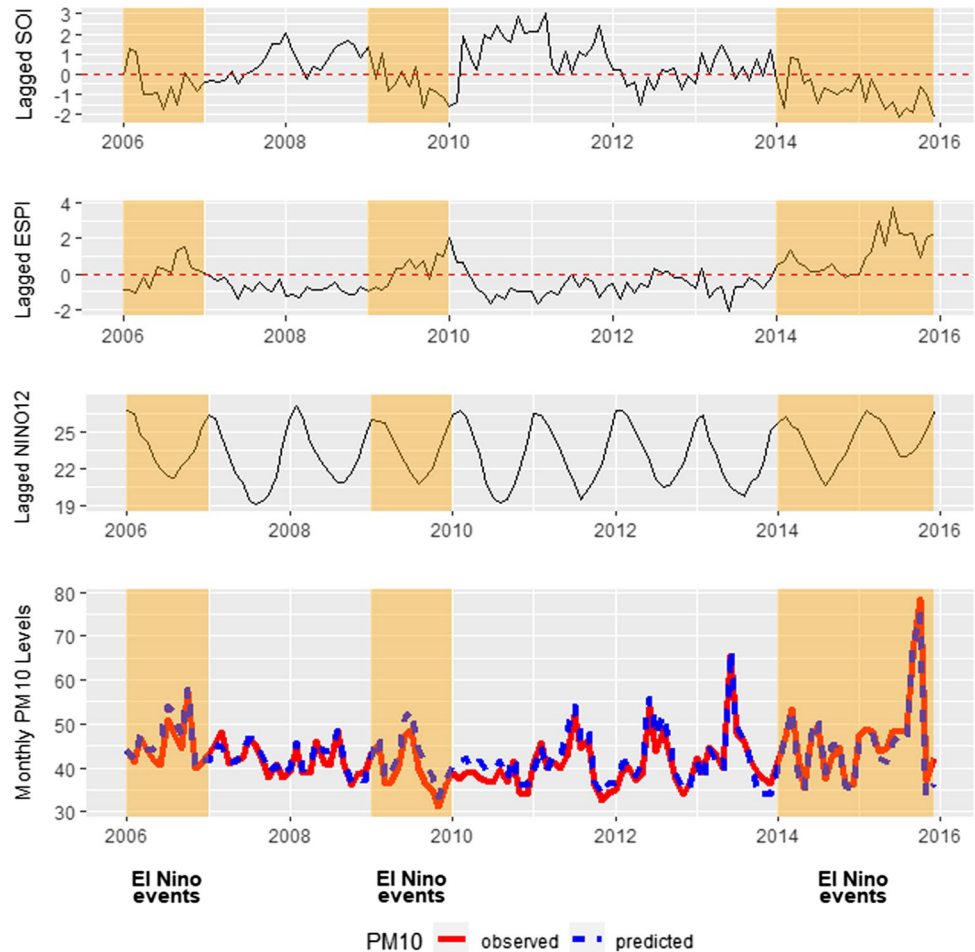


Fig. 5 MCMC trace plots of the model parameters

in the data is good. The estimate of the autoregressive temporal correlation parameter ρ is 0.8536. This estimate implies a strong correlation of PM_{10} levels between successive months. The 0.0073 posterior mean value of parameter ϕ shows that spatial correlation falls to zero at a distance of approximately 410 km, which is determined by the relationship $\exp(-\phi d_0) \approx 0.05$ (Sahu 2012), where d_0 denotes the effective range. Thus, the spatial correlation in PM_{10} levels between locations covering a distance of 410 km and beyond this distance, spatial correlation is negligible.

Figure 6 presents the line graph of the general temporal trend of lagged SOI, ESPI, NINO12, and PM_{10} levels averaged over all the validating sites. The line graph shows a non-linear trend with a fluctuating pattern throughout the study period from 2006 to 2015. The predicted values of PM_{10} generated using the predictive output of the spatio-temporal model (dotted blue line in Fig. 6) seem close to the observed values of PM_{10} recorded at the monitoring stations (solid red line). El Niño events occurred during 2006–2007 and 2009–2010 but were weaker than 2014–2016 (CPC 2019). The figure suggests the peaks in the PM_{10} levels curve concurrent with the El Niño episodes. The nonregular periodic pattern of ENSO indicators during the El Niño

Fig. 6 The general temporal trend of monthly PM_{10} levels averaged over the validating sites and associated ENSO indices



events coincides with high PM₁₀ levels. The PM₁₀ concentration levels usually maximum every May to September and coincide with the southwest monsoon winds season with a low wind scale, often below 7 m/s. However, the El Niño events of 2014–2015 prolonged the high levels of PM10 to the highest peak in October 2015.

We also compute the coverage percentage of the forecast 5–95 percentile range which we find to be 89.2% implying good prediction results since 89.2% of observations are covered by the predictions. The coverage percentage is the percentage of observations that the 5–95 percentile range of predicted future values contains the true outcomes. A good prediction method, say, for the 5–95 percentile range should have coverage close to 90%.

Figure 7 displays the forecast map of PM₁₀ in Peninsular Malaysia in October 2015. We perform the spatial interpolation (kriging) to estimate the PM₁₀ concentrations at the unmonitored locations. Both *fields* and *MBA* R packages (Nychka et al. 2015; Lee et al. 1997) were used in the study to obtain the forecast map. Figure 7 provides a graphical representation in visualizing PM₁₀ values in all areas for a certain period. From the spatial variation in the forecasts observed in Fig. 7, the PM₁₀ levels are lower on the north and east coasts than on the south and west coasts of Peninsular Malaysia during the El Niño event. The mean of the monthly PM₁₀ levels is above 80 μg/m³ in the central region.

Based on the posterior distribution of the model parameters, given in Table 1, none of the ENSO indices is significant and has an impact on the spatial distribution of the PM₁₀ concentration. Therefore, to quantify the effect of the ENSO indices, we apply the autoregressive seasonal model on the PM₁₀ concentration averaged over all the stations. By this means, we ignore the spatial effect. We consider the following models:

$$\text{Model 1A, } \log\text{PM10}_t = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12)$$

$$\text{Model 2A, } \log\text{PM10}_t = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12) + \beta_3 \text{NINO12}_{t-1}$$

$$\text{Model 3A, } \log\text{PM10}_t = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12) + \beta_3 \text{NINO3}_{t-1}$$

$$\text{Model 4A, } \log\text{PM10}_t = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12) + \beta_3 \text{NINO34}_{t-1}$$

$$\text{Model 5A, } \log\text{PM10}_t = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12) + \beta_3 \text{NINO4}_{t-1}$$

$$\text{Model 6A, } \log\text{PM10}_t = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12) + \beta_3 \text{SOI}_{t-1}$$

$$\text{Model 7A, } \log\text{PM10}_t = \beta_1 \sin(2\pi m/12) + \beta_2 \cos(2\pi m/12) + \beta_3 \text{ESPI}_{t-1}$$

We then perform a robust MM estimation procedure on the model parameters to obtain statistical results that are more reliable. Table 2 shows the summary statistics for all models. Based on the *t*-test and the Akaike Information Criteria (AIC), Model 2A is the best model compared to the other models. Thus, NINO12 is the most dominant ENSO indices that can affect the (temporal) distribution of PM₁₀ concentration in Peninsular Malaysia. The other ENSO indices that were significant at the 5% level of significance using the *t* test are SOI (Model 6A) and ESPI (Model 7A), and the AIC values are slightly larger than of Model 2A.

5 Conclusion

We have studied the impact of ENSO indices on the space–time monthly PM₁₀ levels data recorded in 33 sites across Peninsular Malaysia from 2006 to 2015 using the Bayesian hierarchical autoregressive spatio-temporal statistical model. The ENSO indices that we have considered in this study are the NINO12, NINO3, NINO34, NINO4, SOI, and ESPI indices. We have fitted the space–time PM₁₀ data with several models, each with a sine term, a cosine term and one of the ENSO indices as predictors. Based on the 95% credible interval of the posterior distribution of the model parameters associated with the predictors, none of the ENSO indices is associated with the space–time PM₁₀ levels of Peninsular Malaysia in this study. We have concluded that ENSO indices has no association with the spatial distribution of the PM₁₀ concentration.

However, the sine and cosine terms are useful for predicting the PM₁₀ concentrations. Besides, based on the RMSE, MAE and PMCC obtained for the validation set, the model with sine and cosine only is preferred over the other seasonal models with ENSO indices as a predictor. Based on the autoregressive temporal correlation parameter, there is a strong temporal correlation between successive months. On the basis of the spatial correlation parameter, the PM₁₀ levels are correlated spatially as far as 410 km, and past this distance, spatial correlation is negligible. The prediction map of PM₁₀ concentration shows a reasonable prediction of the PM₁₀ levels. The central region of Peninsular Malaysia recorded the highest PM₁₀ concentration levels than in the other places in October 2015. The model with sine and cosine functions as predictors is useful to predict and forecast the PM₁₀ levels in the locations and time points where

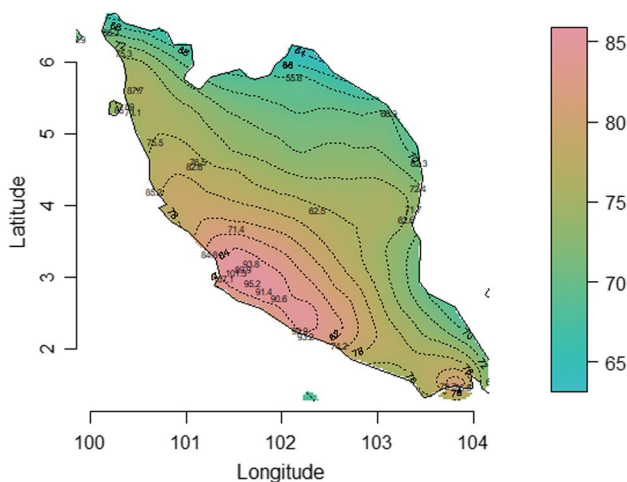


Fig. 7 Forecast map for PM₁₀ in Peninsular Malaysia on October 2015

Table 2 The autoregressive models analysis of the PM₁₀ concentration data

Model ØA			Model 1A		
AIC: – 115.196			AIC: – 132.778		
	Coefficient	<i>t</i> test		Coefficient	<i>t</i> test
Intercept	3.7608	334.4704*	Intercept	3.7602	393.9306*
			Sine	– 0.006	– 0.4458
			Cosine	– 0.0852	– 6.3094*
Model 2A			Model 3A		
AIC: – 139.999			AIC: – 134.899		
	Coefficient	<i>t</i> test		Coefficient	<i>t</i> test
Intercept	3.2213	14.1204*	Intercept	3.5112	13.3508*
Sine	– 0.0662	– 2.4184*	Sine	– 0.0183	– 1.0303
Cosine	– 0.1162	– 6.1323*	Cosine	– 0.0827	– 6.0972*
Lagged NINO12	0.0232	2.3691*	Lagged NINO3	0.0096	0.9506
Model 4A			Model 5A		
AIC: – 132.067			AIC: – 132.718		
	Coefficient	<i>t</i> test		Coefficient	<i>t</i> test
Intercept	3.6776	13.6366*	Intercept	3.5577	9.6291*
Sine	– 0.0076	– 0.538	Sine	– 0.0056	– 0.4073
Cosine	– 0.0835	– 5.8408*	Cosine	– 0.0823	– 5.7813*
Lagged NINO34	0.0031	0.3077	Lagged NINO4	0.0071	0.5497
Model 6A			Model 7A		
AIC: – 134.67			AIC: – 135.542		
	Coefficient	<i>t</i> test		Coefficient	<i>t</i> test
Intercept	3.7625	399.9294*	Intercept	3.7635	395.0141*
Sine	– 0.0067	– 0.5082	Sine	– 0.0051	– 0.383
Cosine	– 0.0805	– 6.0212*	Cosine	– 0.0832	– 6.2412*
Lagged SOI	– 0.015	– 1.8287**	Lagged ESPI	0.0163	1.7735**

** $p < 0.05$, *** $0.10 < p < 0.05$. No superscript means not significant

the data are not being collected. In comparison to some other regression models, we have considered both the spatial and temporal impact simultaneously using a single model. The prediction and forecasting help in the line of action to take in air quality planning, surveillance, intervention and also health funding allocation.

None of the ENSO indices has shown a significant impact on the spatial distribution of the PM₁₀ concentration. Among all the ENSO indices, NINO12, SOI and ESPI show a significant association with the temporal distribution of the PM₁₀ concentrations averaged over all stations in Malaysia. The result of this study provides a piece of valuable information on identifying suitable models, methods, and approaches to quantify the relationship between ENSO indices and PM₁₀ concentration. To study the impact of the global climatological variable on the spatial distribution of the PM₁₀ in the country seems unfeasible using the spatio-temporal model. Fitting the ordinary autoregressive model to the

PM₁₀ averaged over all the stations does not take into consideration the spatial variability. To identify which ENSO indices are significant, we have to fit the model to the PM₁₀ concentrations at each location separately. By this means, each location site in the country has its model. However, an effortless analysis of data should be much more preferred using a single model.

Acknowledgements This work was supported by the Ministry of Higher Education Malaysia [FRGS/1/2018/STG06/UM/02/12].

Data availability The data that support the findings of this study are provided by Department of Environmental Malaysia and are openly available in Malaysia Open Data Portal at https://www.data.gov.my/data/ms_MY/dataset/?q=kualiti+udara&sort=title_string+asc.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with animals or human participants by the author.

References

- Al-Awadhi FA, Al-Awadhi SA (2006) Spatial-temporal model for ambient air pollutants in the state of Kuwait. *Environmetrics* 17(7):739–752
- Al-Hemoud A, Al-Dousari A, Al-Shatti A, Al-Khayat A, Behbehani W, Malak M (2018) Health impact assessment associated with exposure to PM₁₀ and dust storms in Kuwait. *Atmosphere* 9(1):6
- Alifa M, Bolster D, Mead MI, Latif MT, Crippa P (2020) The influence of meteorology and emissions on the spatio-temporal variability of PM₁₀ in Malaysia. *Atmos Res* 246:105107
- Asl FB, Leili M, Vaziri Y, Arian SS, Cristaldi A, Conti GO, Ferrante M (2018) Health impacts quantification of ambient air pollutants using AirQ model approach in Hamadan. *Iran Environ Res* 161:114–121
- Bakar KS, Sahu SK (2015) spTimer: Spatio-temporal Bayesian modeling using R. *J Stat Softw* 63(15):1–32
- Banerjee S, Carlin BP, Gelfand AE (2015) Hierarchical modeling and analysis for spatial data, 2nd edn. Chapman and Hall/CRC, Boca Raton
- Benth JS, Šaltytė L (2011) Spatial-temporal model for wind speed in Lithuania. *J Appl Stat* 38(6):1151–1168
- Cameletti M, Ignaccolo R, Bande S (2011) Comparing spatio-temporal models for particulate matter in Piemonte. *Environmetrics* 22(8):985–996
- Chen J, Carlson BE, Del Genio AD (2002) Evidence for strengthening of the tropical general circulation in the 1990s. *Science* 295:838–841
- Cocchi D, Greco F, Trivisano C (2007) Hierarchical space-time modeling of PM₁₀ pollution. *Atmos Environ* 41(3):532–542
- CPC (2019) Historical El Niño/La Niña episodes (1950–present). United States Climate Prediction Center. 1 February 2019. https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php Accessed 28 June 2021.
- Cressie N, Wikle CK (2011) Statistics for Spatio-Temporal data. Wiley, Hoboken
- Curtis S, Adler (2000) ENSO indices based on patterns of satellite-derived precipitation. *J Clim* 13(15):2786–2793
- El Morabet R (2019) Effects of outdoor air pollution on human health. Reference module in earth systems and environmental sciences. Elsevier, Amsterdam, pp 278–286
- Feng W, Li H, Wang S et al (2019) Short-term PM₁₀ and emergency department admissions for selective cardiovascular and respiratory diseases in Beijing, China. *Sci Total Environ* 657:213–221
- Gelfand AE, Ghosh SK (1998) Model choice: a minimum posterior predictive loss approach. *Biometrika* 85(1):1–11
- Handcock MS, Stein ML (1993) A Bayesian analysis of kriging. *Technometrics* 35(4):403–410
- Hasan MM, Dunn PK (2012) Understanding the effect of climatology on monthly rainfall amounts in Australia using Tweedie GLMs. *Int J Climatol* 32:1006–1017
- Hassan H, Latif MT, Juneng L, Amil N, Khan MF, Yik DJ, Abdullah NA (2020) Interaction of PM₁₀ concentrations with local and synoptic meteorological conditions at different temporal scales. *Atmos Res* 241:104975. <https://doi.org/10.1016/j.atmosres.2020.104975>
- Jeensorn T, Apichartwiwat P, Jinsart W (2018) PM₁₀ and PM_{2.5} from haze smog and visibility effect in Chiang Mai province Thailand. *Appl Environ Res* 40(3):1–10
- Juneng L, Latif MT, Tangang FT, Mansor H (2009) Spatio-temporal characteristics of PM₁₀ concentration across Malaysia. *Atmos Environ* 43(30):4584–4594
- Juneng L, Latif MT, Tangang F (2011) Factors influencing the variations of PM₁₀ aerosol dust in Klang Valley, Malaysia during the Summer. *Atmos Environ* 45(26):4370–4378
- Kim J, Zhou W, Cheung HN, Chow CH (2013) Variability and risk analysis of Hong Kong air quality based on Monsoon and El Niño conditions. *Adv Atmos Sci* 30(2):280–290
- Kim J, Kim M, Ho C, Park RJ, Kim MJ, Lim J, Kim S, Song C (2019) Possible link between arctic sea ice and January PM₁₀ concentrations in South Korea. *Atmosphere* 10(10):619
- Lee S, Wolberg G, Shin S (1997) Scattered data interpolation with multilevel B-splines. *IEEE Trans Visual Comput Graphics* 3(3):228–244
- Manga E, Awang N (2018) Bayesian autoregressive spatiotemporal model of PM₁₀ concentrations across Peninsular Malaysia. *Stoch Env Res Risk Assess* 32(12):3409–3419
- Mukhopadhyay S, Ogutu JO, Bartzke G, Dublin HT, Piepho HP (2019) Modelling spatio-temporal variation in sparse rainfall data using a hierarchical Bayesian regression model. *J Agric Biol Environ Stat* 24(2):369–393
- Noor NM, Yahaya AS, Ramli NA, Luca FA, Abdullah MMA, Sandu AV (2015) Variation of air pollutant (particulate matter—PM₁₀) in Peninsular Malaysia Study in the southwest coast of Peninsular Malaysia. *Rev Chim* 66(9):1443–1447
- Nowak G, Welsh A, O'Neill T, Feng L (2018) Spatio-temporal modelling of rainfall in the Murray-Darling Basin. *J Hydrol* 557:522–538
- Nychka D, Furrer R, Paige J, Sain S (2015) Fields: tools for spatial data [R package version 8]. <https://doi.org/10.5065/D6W957CT>
- Olague EP (2017) Particulate matter and surface deposition. Atmospheric Impacts of the oil and gas industry. Elsevier, pp 47–53
- Pollice A, Lasinio GJ (2009) Spatiotemporal analysis of the PM₁₀ concentration over the Taranto area. *Environ Monit Assess* 162(1–4):177–190
- Rehman SU, Nasir H, Zia SS, Ansari WA, Saleem K, Tayyab N (2012) The impact of Indian Ocean high pressure system on rainfall and stream flow. *Nucleus* 49(2):89–99
- Sahu SK (2012) Hierarchical Bayesian models for space-time air pollution data. Handbook of statistics, time series analysis: methods and applications, handbook of statistic. Elsevier, pp 477–495
- Sahu SK, Gelfand AE, Holland DM (2007) High-resolution space-time ozone modeling for assessing trends. *J Am Stat Assoc* 102(480):1221–1234
- Sentian J, Jemain MA, Gabda D, Franky H, Wui JCH (2018) Long-term trends and potential associated sources of particulate matter (PM₁₀) pollution in Malaysia. *WIT Trans Ecol Environ* 230:607–618
- Shaadan N, Jemain AA, Latif MT, Deni SM (2015) Anomaly detection and assessment of PM₁₀ functional data at several locations in the Klang Valley, Malaysia. *Atmos Pollut Res* 6(2):354–375
- Singhtrattna N, Rajagopalan B, Kumar KK, Clark M (2005) Interannual and interdecadal variability of Thailand summer monsoon season. *J Clim* 18(11):1697–1708
- Sum LP (2018) El Niño—A review of scientific understanding and the impacts of 1997/98 event in Malaysia. Academy of Sciences Malaysia, Kuala Lumpur
- Troup AJ (1965) The Southern oscillation. *Q J R Meteorol Soc* 91:490–506
- Wie J, Moon B (2017) ENSO-related PM₁₀ variability on the Korean Peninsula. *Atmos Environ* 167:426–433
- Won WS, Oh R, Lee W, Kim KY, Ku S, Su PC, Yoon YJ (2020) Impact of fine particulate matter on visibility at Incheon international

- airport, South Korea. *Aerosol Air Qual Res* 20:1048–1061. <https://doi.org/10.4209/aaqr.2019.03.0106>
- Yin F, Ma Y, Zhao X, Lv Q, Liu Y, Li X, Zhang T (2019) Analysis of the effect of PM₁₀ on hand, foot and mouth disease in a basin terrain city. *Sci Rep* 9(1):3233
- Yunus RM, Hasan MM, Razak NA, Zubairi YZ, Dunn PK (2017) Modelling daily rainfall with climatological predictors: poisson-gamma generalized linear modelling approach. *Int J Climatol* 37(3):1391–1399
- Yusof NFFM, Ramli NA, Yahaya AS, Sansuddin N, Ghazali NA, Madhoun WA (2009) Monsoonal differences and probability distribution of PM₁₀ concentration. *Environ Monit Assess* 163(1–4):655–667

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.