On the predictability of short-lived particulate matter around a cement plant in Kerman, Iran: machine learning analysis

F. Borhani[1] · M. Shafiepour Motlagh[1] · A. H. Ehsani[1] · Y. Rashidi[2] · S. Maddah[1] · S. M. Mousavi[3]

Abstract
One of the greatest environmental risks in the cement industry is particulate matter emission (i.e., $PM_{2.5}$ and PM_{10}). This paper aims to develop descriptive-analytical solutions for increasing the accuracy of predicting particulate matter emissions using resample data of Kerman cement plant. Photometer instruments DUST TRAK and BS-EN-12341 method were used to determine concentration of $PM_{2.5}$ and PM_{10}. Sampling was performed on 4 environmental stations of Kerman cement plant in the four seasons. In order to accurate assessment of particulate matter concentration, a new model was proposed to resample cement plant time series data using Pandas in Python. The effect of meteorological parameters including wind speed, relative humidity, air temperature and rainfall on the particulate matter concentration was investigated through statistical analysis. The results indicated that the maximum annual average of 24-h of $PM_{2.5}$ belonged to the east side (opposite the clinker depot) in 2019 (31.50 µg m^{-3}) and west side (in front of the mine) in 2020 (31.00 µg m^{-3}). Also, maximum annual average of 24-h of PM_{10} belonged to the west side (in front of the mine) in 2020 (121.00 µg m^{-3}) and east side (opposite the clinker depot) in 2020 (120.75 µg m^{-3}). The $PM_{2.5}$ and PM_{10} concentrations are more than the allowable limit. The results demonstrate that particulate matter concentration increases with increasing relative humidity and rainfall. Finally, the SARIMA model was used to predict the particulate matter concentration.

Keywords Particulate matter · Short-lived climate pollutants · Cement plant · SARIMA forecasting model · Machine learning analysis

Introduction

Human life depends on air (Nemerow et al. 2009; Borhani et al. 2017). In recent years, there are increasing concerns about air pollution, which is one of the leading environmental problems and environmental health (Borrego et al. 2006). The ambient air quality around industrial plants (including Cement plants, Bitumen insulation production plants, etc.) is affected by the dispersion of short-lived

climate pollutants and of other sources which can cause some impairment to human health, air pollution and climate change (Merenu et al. 2007; Borhani et al. 2016, 2021a; Cheraghi and Borhani 2016a, b; Hoveidi et al. 2017; Mousavi et al. 2017; Borhani and Noorpoor 2020; Mousavi and Falahatkar 2020; Maddah et al. 2022). The cement industry is one of the major and strategic industries in the country. The cement industry, as the basis of the country's development, has a key role in the construction of housing, dam construction projects, industrial plants, buildings, road development, etc. With the growth of cement industry, cement industries are one of the biggest causes of air pollution (Humphreys and Mahasenan 2002; Ali et al. 2011; Assegaf and Jayadipraja 2015). One of the most important pollutants released from the cement industry is particulate matter. The most important sources of particulate matter emissions in this industry are including crushing, dry grinding, raw materials transportation, rotary cement kilns operation, Clinker cooling process, packaging and vehicle movement (Chehregani 2004). So, short-lived climate
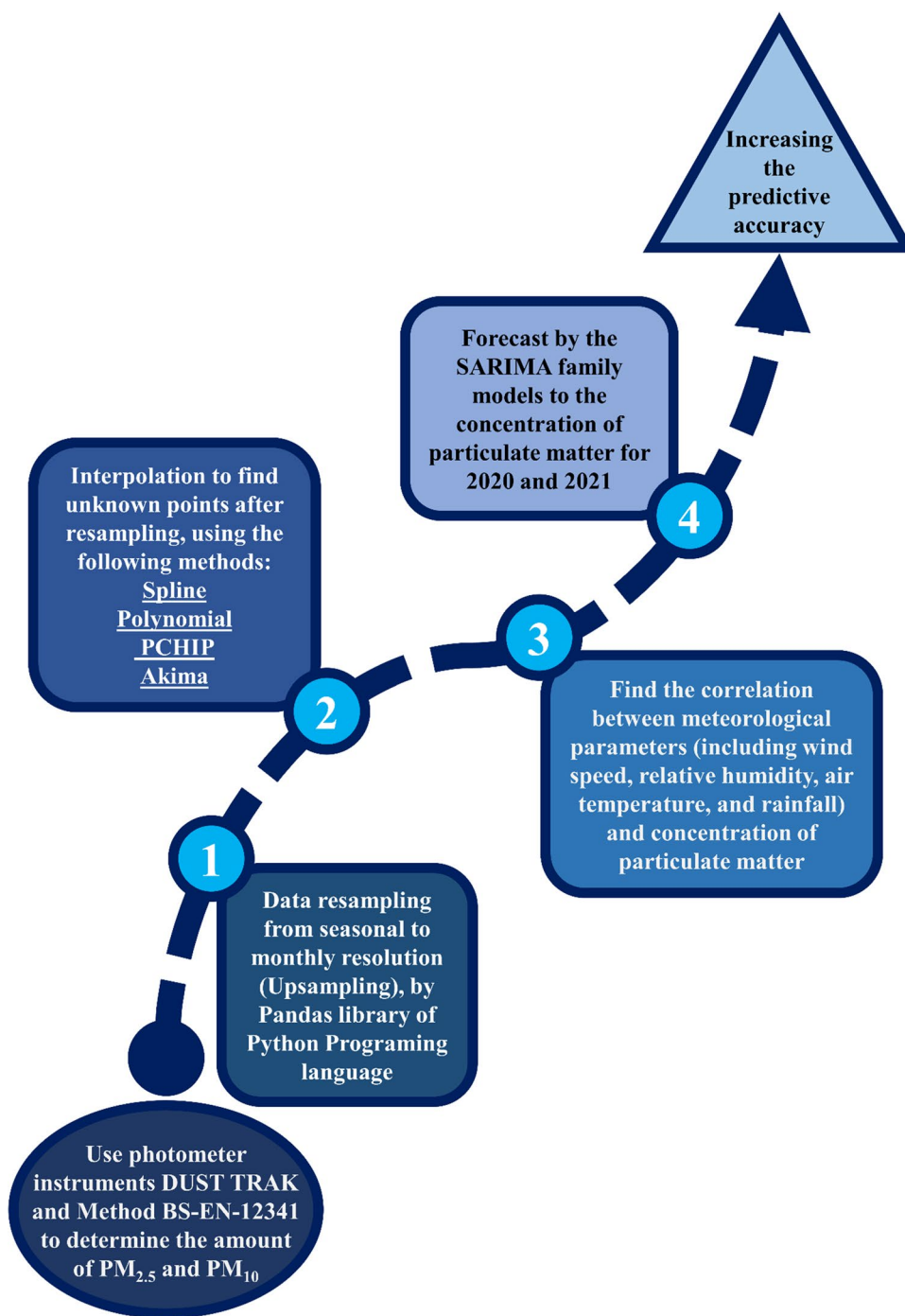
✉ F. Borhani
    fborhani78@ut.ac.ir

1   School of Environment, College of Engineering, University of Tehran, Tehran, Iran

2   Environmental Sciences Research Institute, Shahid Beheshti University, Tehran, Iran

3   Department of Environmental Planning and Design, Shahid Beheshti University, 1983969411 Tehran, Iran

Fig. 1 Flowchart of research
methodology

pollutants (with special emphasis paid to $PM_{2.5}$ and PM_{10}) monitoring and prediction studies are of great importance as they warn air pollutants emission systems and improve ambient air quality control regulations in an industrial plant (Baldasano et al. 2003; Özden et al. 2008).

Several studies assessed the environmental impacts of industrial plants productions on around environment. For example, Sharma and Pervez (2003) studied the changes

seasonal of PM_{10} and SPM levels in ambient air around a cement plant. Their results show that there is a positive correlation coefficient between PM_{10} and SPM. Agrawal and Khanam (1997) investigated the concentration of particulate matter around a cement plant. The results showed Concentrations of particulate matter often exceeded permissible limits up to 2 km SE (southeast) of the source. Dust fall rate was quite significant up to 5 km SE. The dust
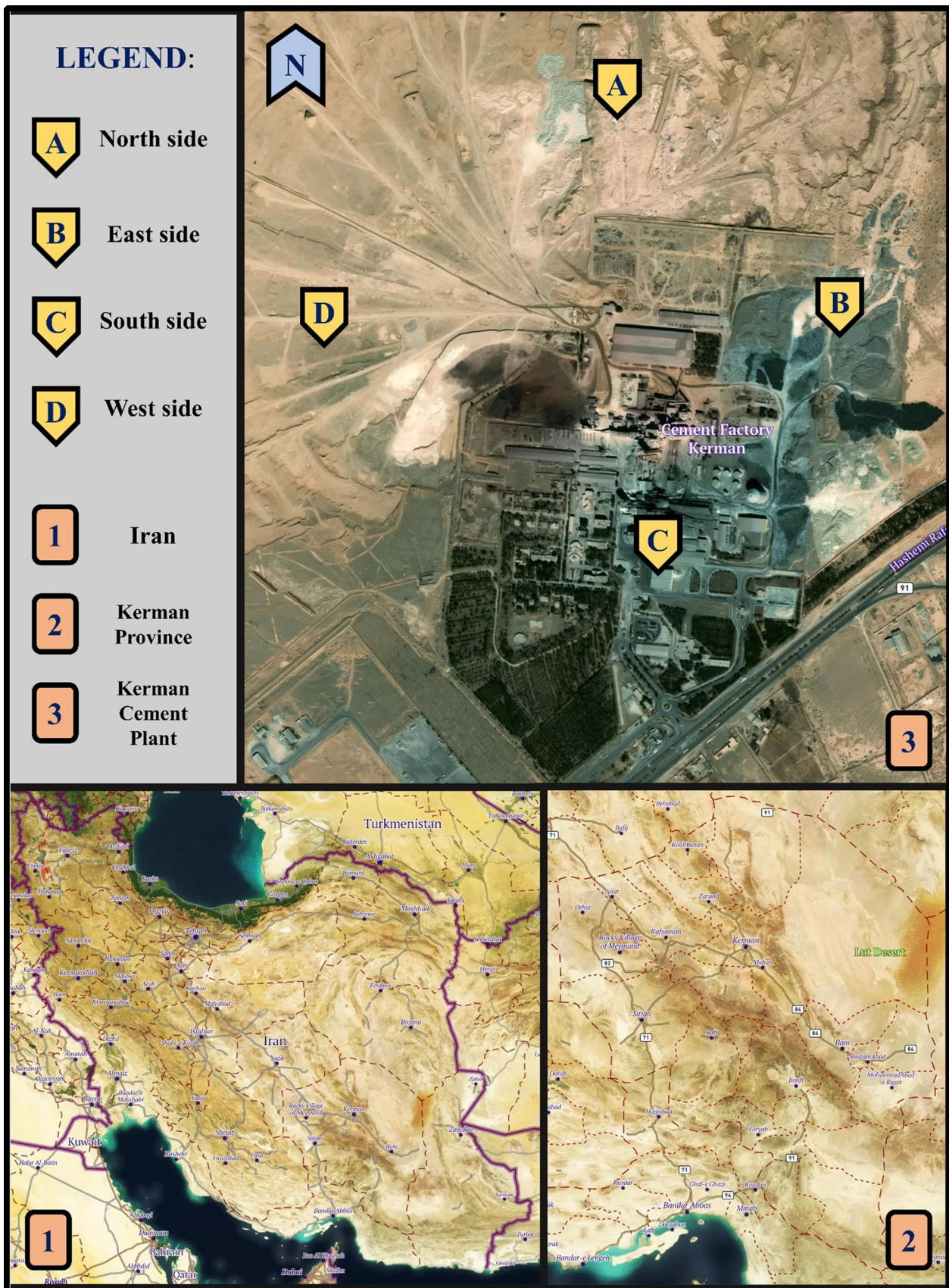
Fig. 2 Map of the study area and sampling stations in Kerman cement plant

Table 1 Location information of sampling in Kerman cement plant

No	Location	Coordinate point		Floor material
		Latitude	Longitude	
A	North side (in front of the waste depot)	$30°14'28.4^\varepsilon N$	$56°54'26.7^\varepsilon E$	Soil
B	East side (opposite the clinker depot)	$30°14'12.7^\varepsilon N$	$56°54'48.0^\varepsilon E$	Soil
C	South side (adjacent to the entrance door)	$30°13'55.6^\varepsilon N$	$56°54'30.7^\varepsilon E$	Concrete
D	West side (in front of the mine)	$30°14'12.9^\varepsilon N$	$56°53'59.2^\varepsilon E$	Soil

Table 2 The average concentrations of particulate matter for the period of 2016 to 2020 in Kerman cement plant

	Year	Particulate matter levels							
		Winter		Spring		Summer		Autumn	
		$PM_{2.5}$ (µg m^{-3})	PM_{10} (µg m^{-3})	$PM_{2.5}$ (µg m^{-3})	PM_{10} (µg m^{-3})	$PM_{2.5}$ (µg m^{-3})	PM_{10} (µg m^{-3})	$PM_{2.5}$ (µg m^{-3})	PM_{10} (µg m^{-3})
A	2016	14.00	38.00	16.00	43.00	19.00	41.00	21.00	43.00
	2017	23.00	48.00	21.00	45.00	18.00	42.00	16.00	38.00
	2018	19.00	42.00	21.00	45.00	24.00	53.00	21.00	51.00
	2019	32.00	82.00	24.00	57.00	19.00	48.00	20.00	61.00
	2020	30.00	97.00	27.00	84.00	20.00	75.00	17.00	69.00
B	2016	16.00	42.00	18.00	39.00	16.00	34.00	15.00	35.00
	2017	19.00	39.00	20.00	43.00	22.00	44.00	20.00	43.00
	2018	23.00	47.00	28.00	80.00	30.00	88.00	29.00	71.00
	2019	43.00	111.00	31.00	74.00	27.00	61.00	25.00	95.00
	2020	39.00	142.00	32.00	131.00	27.00	108.00	24.00	102.00
C	2016	22.00	46.00	13.00	37.00	17.00	35.00	19.00	38.00
	2017	15.00	36.00	18.00	47.00	15.00	36.00	11.00	33.00
	2018	13.00	36.00	16.00	40.00	18.00	45.00	16.00	40.00
	2019	29.00	76.00	18.00	42.00	15.00	37.00	14.00	40.00
	2020	22.00	65.00	24.00	68.00	18.00	69.00	14.00	58.00
D	2016	19.00	39.00	21.00	42.00	18.00	36.00	17.00	39.00
	2017	21.00	43.00	19.00	40.00	20.00	40.00	24.00	46.00
	2018	21.00	44.00	25.00	68.00	27.00	73.00	26.00	69.00
	2019	34.00	89.00	29.00	71.00	33.00	87.00	22.00	84.00
	2020	35.00	124.00	34.00	136.00	29.00	115.00	26.00	109.00

fall rate was highest in winter, while total suspended particles (TSP) levels were highest in summer. Mohebi and Baroutian (2006) measured the concentration of PM_{10} dispersion in Kerman cement plant and predict PM_{10} concentration using the Eulerian model, Gaussian dispersion model and neural Networks (ANNs) and compared with the measured data. Borhani and Noorpoor (2017) investigated the cancer risk assessment of the release of volatile organic compounds (VOCs) from the production of insulation bituminous. Also, they studied and modeled the environmental effects of the emission of NO_X and CO pollutants around the insulation bituminous plant (Borhani et al. 2019).

Further, several researchers have investigated the forecast of particulate matter (PM_{10} and $PM_{2.5}$) concentrations using machine learning techniques. Jeong et al. (2020) used a machine learning algorithm called random forest method to predict PM_{10} concentration using air parameters (i.e., wind speed, relative humidity and air temperature). Masood and Ahmad (2020) presented a model for evaluating particulate matter ($PM_{2.5}$) based on machine learning approaches such as support artificial neural networks (ANN) and vector machines (SVM). Vidnerová and Neruda (2021) modeled air pollution by machine learning methods including kernel methods, regularization networks, regularization networks with composite kernels and deep neural networks. Díaz-Robles et al. (2008) predicted particulate matter concentrations in Temuco city using Box–Jenkins time series (ARIMA) and multilinear regression (MLR) models. In another study, an ARMA/ARIMA time-series model

Table 3 Annual average of particulate matter and meteorological parameters recorded from 2016 to 2020 around the cement plant, in Kerman, Iran

Year	Particulate matter		Meteorological parameters			
	$PM_{2.5}$ ($\mu g\ m^{-3}$)	PM_{10} ($\mu g\ m^{-3}$)	T (°C)	WS (km/h)	RH (%)	PP (mm)
2016	17.56	39.19	18.00	11.90	21.10	53.58
2017	18.88	41.44	17.50	11.50	20.50	90.42
2018	22.31	55.75	17.80	11.70	20.20	73.65
2019	25.94	69.69	17.40	11.60	20.50	125.21
2020	26.13	97.00	16.80	11.40	30.60	205.50

was employed to predict the short-term series of the PM_{10} concentrations (Zhang et al. 2017).

The purpose of this research is to investigate the changes in particulate matter concentrations from 2016 to 2020 around the Kerman cement plant. Due to several constraints in measuring pollutants around the Kerman cement plant, measurements were taken only at seasonal intervals and only once in each season, at four sampling stations of the plant to provide relative coverage of the plant area both spatially and temporally. The number of measurements is not enough to achieve the goal of this study, which is to examine the changes in mean particulate matter concentrations around the Kerman cement plant and forecast for the next 1 year (2021). The novelty of this research consists in its sampling-resampling approach that combines the actual measurement of particulate matter with statistical analysis performed using Pandas libraries in Python programming language. Therefore, we investigate the employment of resampling strategies in imbalanced time series. Then, we use interpolation methods to fill missing values in series data using Pandas in Python. In the following, resulting correlations between particulate matter and meteorological parameters were investigated. Finally, the SARIMA time series analysis model is used for predicting particulate matter (PM) concentrations. Figure 1 shows a flowchart of the steps followed during the research.

Materials and methods

Study area

Kerman city

Kerman, the capital city of the largest province of Iran, is in the southeast of Iran. The city with an area of 238.80 km^2 is located at 30°17′30″ and 57°05′00″. Kerman is about 1755 m (5758 ft) above sea level, that makes it the third in elevation among provincial capitals of Iran. Kerman city has a moderate weather and the mean annual rainfall is 142 mm, the minimum and maximum temperatures are −7 and 39.6 °C, respectively, and the mean relative humidity is 36% (Kerman-met 2020). Kerman city is exposed to increasing levels of air pollution as a result of concentrated industrial activities and urbanization and transportation. Figure 2 shows the location of the city of Kerman in Iran.

Kerman cement plant

This plant is located in the southwest of Kerman and near Bandar Abbas city (Fig. 2). Kerman cement plant, established in 1965, is the first cement production plant in the southeast to produce different types of Portland cement type 2, type 5, Pozzolanic Portland cements, and oil well cement class G with a nominal capacity of 3600 tons of clinker per day according to the market demand and needs (Kcig 2020). Also, the Kerman cement plant is located approximately 17 km on the Kerman-Tehran highway. The presence of particulate matter in the exhaust fumes of cars traveling on the highway also imposes additional pollution on the area. Therefore, it can be termed as one of the most important sources of main air pollutants in Kerman.

Sampling

In the present study, the amount of environmental suspended particles in the air of the Kerman cement plant during the years 2016–2020 has been studied. A portable aerosol particulate monitor (Dust-Trak instrument, TSI Model 8534, Germany) based on the photometry and the Standard Methods of Examination (BS-EN-12341) with a measuring range of 0.001–150 mg/m^3 and accuracy of ± 0.001 mg/m^3 in the ambient air used for measuring the concentrations of particulate matter with aerodynamic diameters less than 2.5 μm ($PM_{2.5}$) and 10 μm (PM_{10}). Sampling was implemented in four corners of the Kerman cement plant which were located in short distance—a few hundred meters—from the origin of the Stacks in spring, summer, autumn and winter. In this method, the device is first calibrated and placed 1.5 m above the ground (De Nevers 2000) on the north side (in front of the waste depot), east side (opposite the clinker depot), south side (adjacent to the entrance door), and west side (in front of the mine) (Table 1). Then, the flow of the adjusting device (pump flow 0.0017 m^3/min and suction volume 0.102 m^3) and sampling time were done for 60 min. Due to the capability of the device, the average amount of particulate matter

(PM) in the air per unit volume was determined (McMurry 2000; Gokhale 2009). Meteorological data (including wind speed, relative humidity, air temperature and rainfall) data are recorded by Kerman Meteorological Station (Kerman-met 2020).

Resampling

Resampling involves changing the frequency of Kerman cement plant data's time series observations. Upsampling and downsampling are two types of resampling (Brownlee 2019). Here, we used of upsampling scenario using Pandas in Python. We increased the frequency of the samples, from seasonally to monthly. For this purpose, we assigned data of a season to the first month of that season and filled the other 2 months with NaN values.

Interpolation

Then, our aim is to investigate the accuracy of four interpolation techniques to fill missing data from Kerman cement plant data. Data interpolation is a statistical method of estimation that finds new data based on the range of a discrete set of existing, neighboring values (Dan et al. 2020). Here, we applied the spline method, polynomial, PCHIP (piecewise cubic Hermite interpolating polynomial) and Akima interpolation method (Chapra and Canale 1988; Cheney and

Fig. 3 The auto-correlation function (ACF) and partial auto-correlation function (PACF) plots for prediction of PM$_{2.5}$ concentrations, **a** north side, **b** east side, **c** south side, **d** west side ▶

Kincaid 2012). The spline and polynomial methods can take different orders to perform the best interpolation, so in this study, orders 1, 2, and 3 were used for both methods. They are the simplest and the most common type of interpolation (Jung and Chong 2017).

Prediction

To predict the data, we used seasonal autoregressive integrated moving average (SARIMA or Seasonal ARIMA). SARIMA (p, d, q) models are based on autoregressive (AR) and moving average (MA) (Suhartono 2011). We used Partial Auto-Correlation Function (PACF) and Auto-Correlation Function (ACF) plots to find p, d and q parameters of the AR-I-MA model (Agrawal et al. 2017). Here, in the written code, we have defined a simple AR process and found its order using the ACF and PACF plots (Salvi 2019; Dettling 2013). Auto regressive process, a time series is said to be auto regressive when present amount of the time series can be obtained using previous amounts of the same time series, i.e., the present amount is weighted mean of the past amounts. The AR process of an order p can be written as,

$$y_t = C + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t \qquad (1)$$

Table 4 Spearman's rank correlations between particulate matter and meteorological parameters data from 2016 to 2020 around the cement plant, in Kerman, Iran

Parameters	PM$_{2.5}$ (µg m^{-3})	PM$_{10}$ (µg m^{-3})	T (°C)	WS (km/h)	RH (%)	PP (mm)
PM$_{2.5}$ (µg m^{-3})	1					
PM$_{10}$ (µg m^{-3})	0.9076	1				
T (°C)	−0.7499*	−0.8854*	1			
WS (km/h)	−0.6046*	−0.6716*	0.9075	1		
RH (%)	0.5262	0.8338	−0.8325*	−0.6008*	1	
PP (mm)	0.8035	0.9480	−0.982*	−0.8206*	0.8809	1

*p-value < 0.05

Table 5 Residual mean squared error (RMSE) values of competing Interpolation methods

Method	PM$_{2.5}$				PM$_{10}$			
	A	B	C	D	A	B	C	D
Spline (1)	4.24	4.75	3.28*	3.11	7.55	4.22*	19.89	10.33*
Spline (2)	3.55	4.74	5.43	3.63	6.63	36.30	18.97	12.81
Spline (3)	4.45	4.99	3.87	3.90	11.99	12.25	20.88	12.30
Polynomial (1)	1.86*	3.80*	3.86	4.31	6.56	20.44	18.22	11.09
Polynomial (2)	3.30	4.70	3.92	6.45	7.00	10.75	19.00	36.92
Polynomial (3)	3.38	4.09	4.04	1.31	6.60	15.81	19.72	30.69
PCHIP	4.37	5.47	4.17	5.53	6.39	8.67	16.80*	17.12
Akima	3.38	4.91	3.96	0.55*	5.95*	14.38	18.36	29.63

*Minimum residual mean squared error
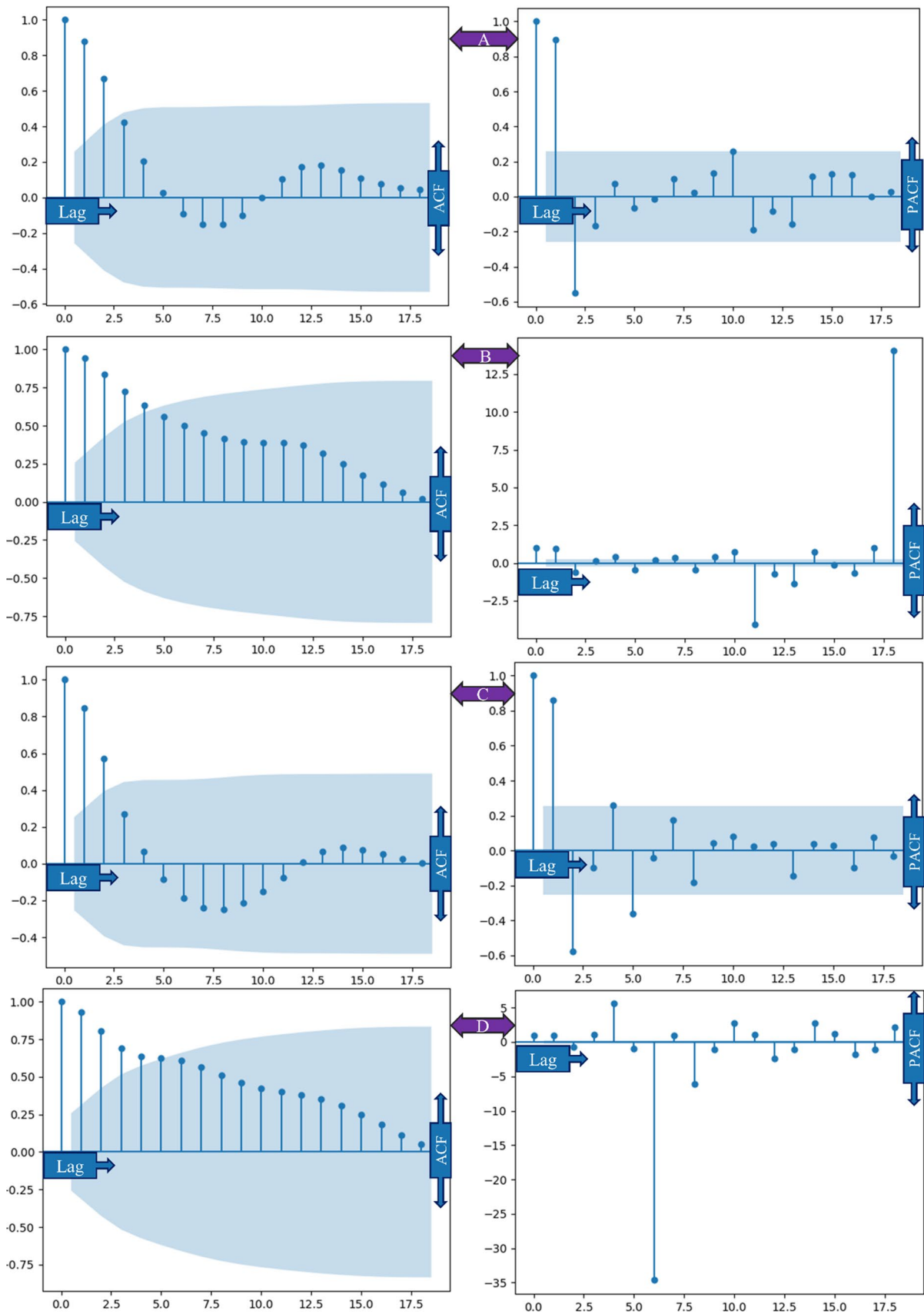
$$\varepsilon_t = y_t - y_{t-1} \tag{2}$$

where C is intercept, $\varnothing_i$ ($i = 1, 2... p$) is auto-regressive model parameters, y_t is current time-series value, y_{t-1}, $y_{t-2} \cdots y_{t-p}$ is past values and ε_t is random error. The Root Mean Square Error (RMSE) is a good measure of accuracy, but only to compare forecasting errors of different models or model configurations for a particular variable and not between variables, as it is scale-dependent. RMSE was given by the following equation (Beckerman et al. 2013; Elavarasan et al. 2018; Goap et al. 2018).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{3}$$

where N denotes the number of datum points in the set, y_i is actual value and $\hat{y}_i$ is model predicted value. Finally, the best interpolation method was selected by minimizing the residual mean squared error (RMSE) using cross-validation.

Result and discussion

Sampling data analysis

The results of $PM_{2.5}$ and PM_{10} concentrations are presented in Table 2. Results showed that the maximum annual average of 24 h of particulate matter ($PM_{2.5}$) concentration belong to the east side (opposite the clinker depot) in 2019 (31.50 µg m^{-3}) and west side (in front of the mine) in 2020 (31.00 µg m^{-3}), which is more than the allowable limit (the quality standard of $PM_{2.5}$ is 25.00 µg m^{-3}). Also, the maximum annual average of 24-h of particulate matter (PM_{10}) concentration belong to the west side (in front of the mine) in 2020 (121.00 µg m^{-3}) and east side (opposite the clinker depot) in 2020 (120.75 µg m^{-3}), which is more than the allowable limit (50.00 µg m^{-3}) (WHO 2013). Alizadeh et al. (2010) estimated the average of particulate matter 380 µg m^{-3} in the Kerman cement plant. The results also showed that particulate matter concentration (PM_{10} and $PM_{2.5}$) varied in different units in the Kerman cement plant. The highest average concentrations of PM_{10} and $PM_{2.5}$ identified in the east side (opposite the clinker depot), west side (in front of the mine), north side (in front of the waste depot) and south side (adjacent to the entrance door), respectively. These findings are in agreement with previous studies (Abu-Allaban and Abu-Qudais 2011; Ahmad et al. 2013; Aghamolaie et al. 2015; Jayadipraja et al. 2016). The concentration of PM_{10} in 2018, 2019 and 2020 is higher than the standard limits, but in general, the concentration of $PM_{2.5}$ in the environment could be considered acceptable (Table 3). The highest

Fig. 4 The auto-correlation function (ACF) and partial auto-correlation function (PACF) plots for prediction of PM_{10} concentrations, **a** north side, **b** east side, **c** south side, **d** west side

concentrations of particulate matter (PM_{10} and $PM_{2.5}$) are obtained in winter (24.45 and 64.30 µ/m^3), followed by spring and summer but the lowest amounts in the autumn (Table 2). Our findings are similar to previous research (Leone et al. 2016; Shahri et al. 2019; Ciobanu et al. 2021; Borhani et al. 2022a).

Correlation analysis

Table 4 presents correlation between particulate matter concentration and meteorological parameters were tested using the Pearson coefficients (r). The value of P-value below 0.05 was considered statistically significant (Field and Miles 2009; Carslaw and Ropkins 2012; Borhani et al. 2021b). The results indicated that the particulate matter concentration increases with increasing relative humidity and rainfall (p-value > 0.05). The above-mentioned findings are in accordance with Wang et al. (2010) which means relative humidity increases will lead to increased particulate matter concentration. There is an inverse relationship between particulate matter and wind speed and temperature (p-value < 0.05). This finding is similar to the results obtained in the research of (Nazif et al. 2019; Akbal and Ünlü 2022; Bañuelos Gimeno et al. 2022; Borhani et al. 2022b). Also, a significant positive correlation was found between $PM_{2.5}$ and PM_{10}, which showed that $PM_{2.5}$ and PM_{10} have the same origin ($r = 0.907$).

Results of the machine learning techniques on python

The sampling data trend and their prediction trend for particulate matter concentration are shown in Figs. 5 and 6 for each location from 2016 up to 2021. Also, Figs. 5 and 6 show the $PM_{2.5}$ and PM_{10} concentrations predictions, in Kerman cement plant from 2020 to 2021, obtained using all interpolation methods. We have forecasted particulate matter concentrations for 2 years (2020 and 2021), using SARIMA statistical model. Table 5 presents the lowest RMSE values obtained from 8 interpolation methods. This table clearly shows that the spline order 1 interpolation method is relatively superior to other methods. The Akima's interpolation method has a high prediction accuracy for $PM_{2.5}$ concentrations in location D (RMSE = 0.55).

Residual mean squared error (RMSE) values of $PM_{2.5}$ concentrations are lower than PM_{10} concentrations in competing interpolation methods (Table 5). So, we conclude that forecasting the PM_{10} concentrations is more difficult
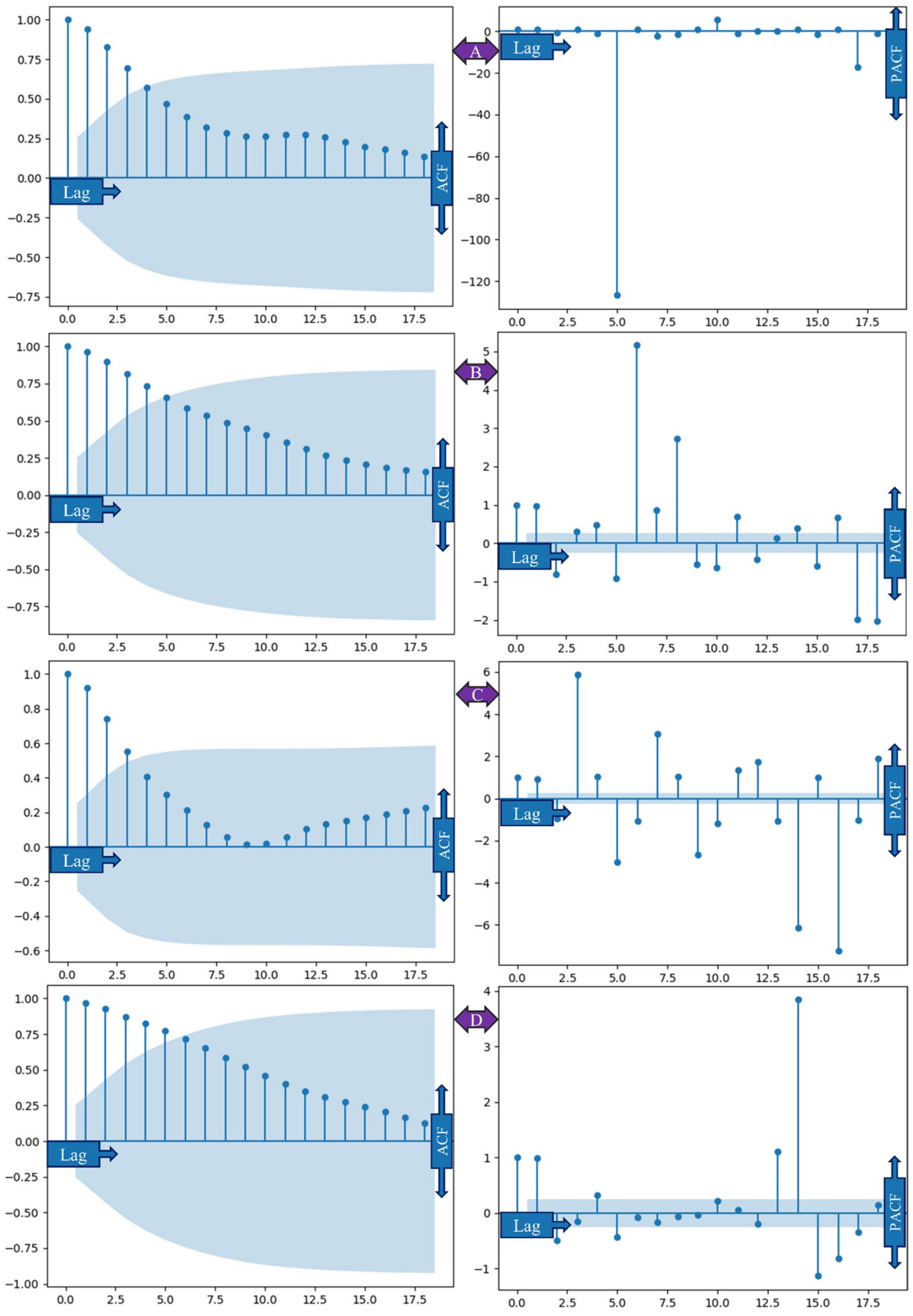
Table 6 Forecast monthly average concentrations of $PM_{2.5}$ and PM_{10} from January 2021 to December 2021 based on minimum RMSE values for each interpolation method

| Date	$PM_{2.5}$ (µg m^{-3})				PM_{10} (µg m^{-3})			
	A	B	C	D	A	B	C	D
2021/01	20.56	28.75	17.39	23.72	64.68	92.80	42.17	106.50
2021/02	20.69	28.41	17.41	23.64	63.87	90.22	43.09	105.70
2021/03	20.81	28.10	17.38	23.68	62.77	87.92	44.61	104.92
2021/04	20.92	27.82	17.31	23.80	61.53	85.90	45.61	104.15
2021/05	20.99	27.57	17.24	23.94	60.37	84.11	45.55	103.41
2021/06	21.04	27.33	17.19	24.10	59.43	82.53	44.66	102.68
2021/07	21.07	27.12	17.15	24.25	58.76	81.15	43.61	101.97
2021/08	21.07	26.93	17.14	24.38	58.31	79.93	43.01	101.27
2021/09	21.07	26.75	17.15	24.49	57.99	78.86	43.07	100.59
2021/10	21.06	26.59	17.17	24.57	57.75	77.92	43.56	99.92
2021/11	21.05	26.45	17.18	24.62	57.53	77.10	44.07	99.27
2021/12	21.04	26.31	17.20	24.66	57.29	76.38	44.27	98.64
Annual average	20.95	27.34	17.24	24.15	60.02	82.90	43.94	102.42
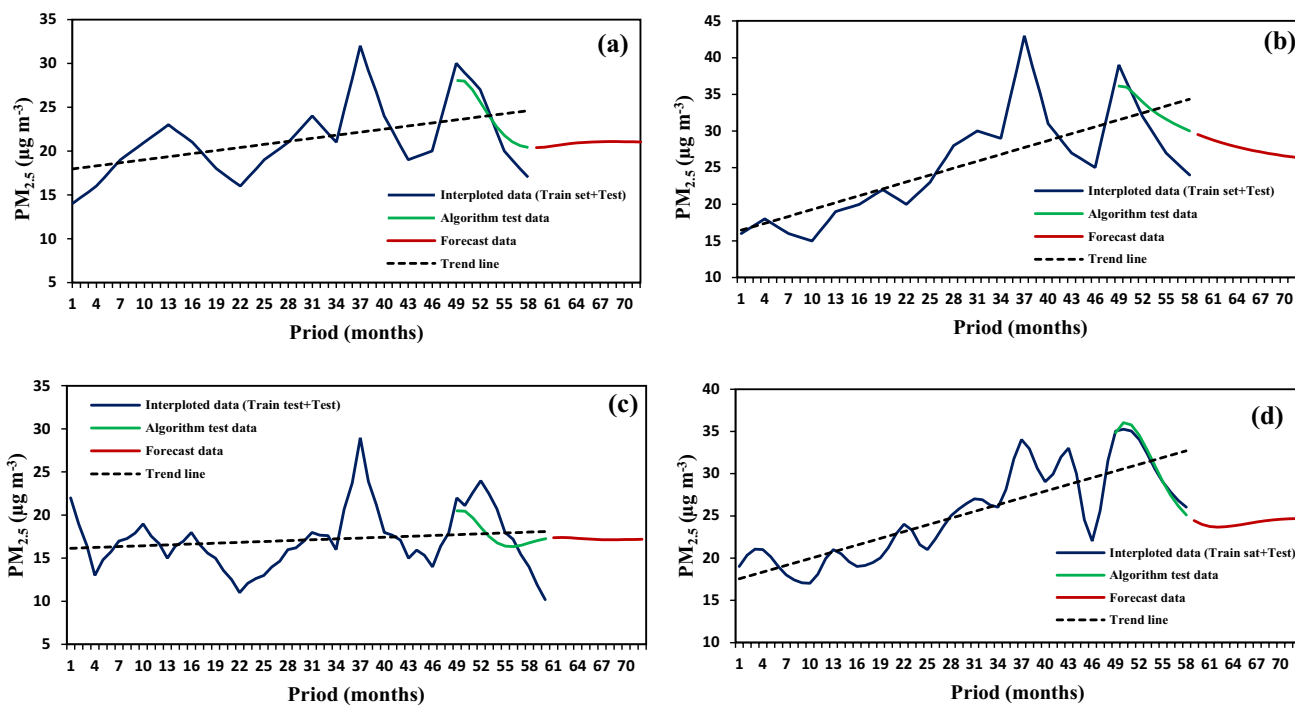
Fig. 5 Time-series analysis (Train set, Forecast set, Algorithm forecast, Test set and Algorithm test) model of $PM_{2.5}$ concentrations in sampling stations in Kerman cement plant from 2016 to 2020, **a** A, **b** B, **c** C, **d** D

than $PM_{2.5}$ concentrations. Figures 3 and 4 show the partial auto-correlation function (PACF) and auto-correlation function (ACF) plots for prediction particulate matter parameters (i.e., $PM_{2.5}$ and PM_{10}) from 2020 to 2021 in the Kerman cement plant. The shaded region represents the 95% confidence interval. Table 6 predicts future $PM_{2.5}$ and PM_{10} with 95% confidence intervals for Kerman cement plant from January 2021 to December 2021 on monthly

basis (Figs. 5 and 6). According to the forecast model, the concentrations of $PM_{2.5}$ and PM_{10} have decreased by about 14.20 and 25.44%, respectively, in 2021 compared to 2020 (Tables 3 and 6). maybe to some extent, this decrease in the concentrations of particulate matter can be attributed to the effects of the COVID-19 pandemic on the air pollutants concentration during the lockdown in Iran.
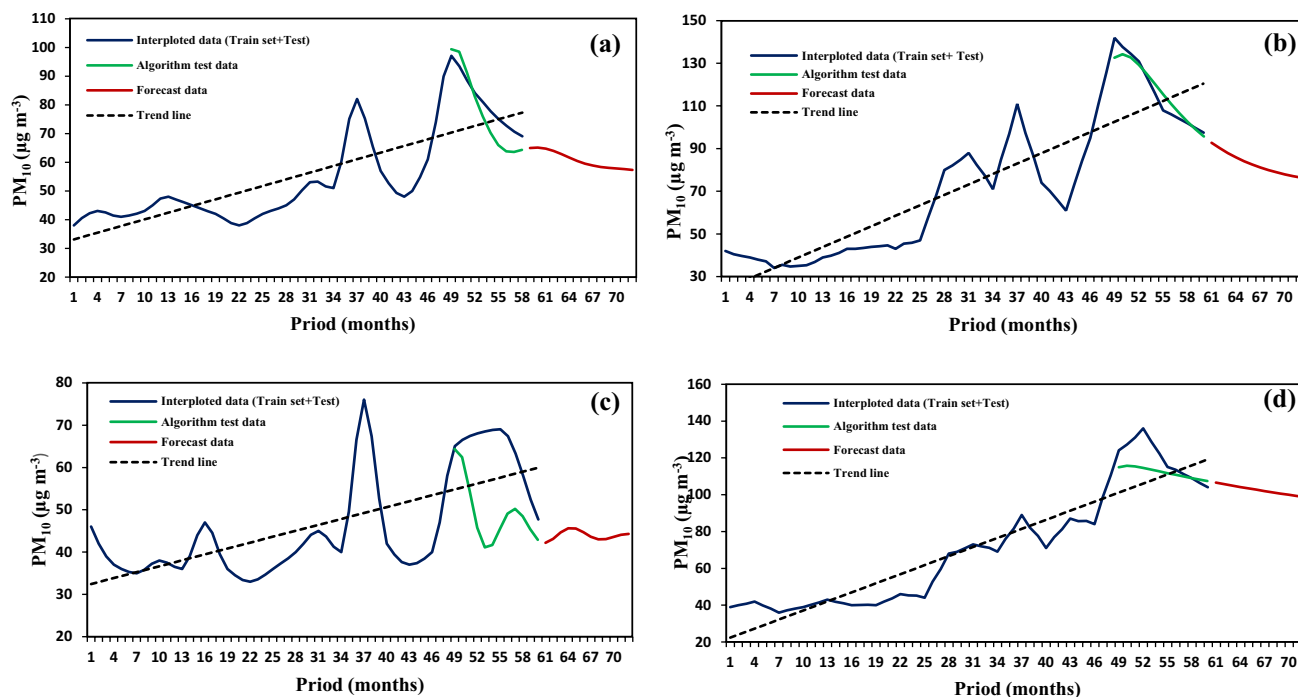
Fig. 6 Time-series analysis (Train set, Forecast set, Algorithm forecast, Test set and Algorithm test) model of PM_{10} concentrations in sampling stations in Kerman cement plant from 2016 to 2020, **a** A, **b** B, **c** C, **d** D

Conclusion

The paper presents the results of comparative analysis of the monthly trend of the concentration of PM_{10} and $PM_{2.5}$ pollutants in 5-year average time series (2016–2020) and predict future trends in the next year (2021) in the Kerman cement plant. We show how to define and then compute efficiently the marginal likelihood of a SARIMA model with missing observations. We also showed how to predict and interpolate missing observations and obtained the mean squared error of the estimate. Therefore, different methods of sampling, resampling, interpolation and prediction on Python were used. The results indicated that the maximum annual average of 24-h of $PM_{2.5}$ belonged to the east side (opposite the clinker depot) in 2019 (31.50 $\mu g\ m^{-3}$) and west side (in front of the mine) in 2020 (31.00 $\mu g\ m^{-3}$). Also, maximum annual average of 24-h of PM_{10} belonged to the west side (in front of the mine) in 2020 (121.00 $\mu g\ m^{-3}$) and east side (opposite the clinker depot) in 2020 (120.75 $\mu g\ m^{-3}$). According to the forecast model, the concentrations of $PM_{2.5}$ and PM_{10} have decreased by about 14.20 and 25.44%, respectively, in 2021 compared to 2020. In the next step, the correlations of $PM_{2.5}$ and PM_{10} and with meteorological parameters (i.e., temperature, wind speed, relative humidity and rainfall) were examined using Spearman's rank correlations. The results

showed that $PM_{2.5}$ and PM_{10} have a positive and significant correlation with relative humidity and rainfall and a negative correlation with temperature and wind speed. Based on our predictions, it is recommended that Kerman cement plant officials and administrators must make suitable and good decisions at the proper time for equipment and materials supply to control and reduce the effects of pollution by the production cement units because the ambient air quality around the Kerman cement plant is affected by the dispersion of industrial short-lived climate pollutants emissions, meteorological conditions, and other sources which can cause some impairment to air pollution and human health.

Author contribution FB was involved in the conception or design of the work, data collection, data analysis and interpretation, drafting the article and final approval of the version to be published; MSM contributed to the conception or design of the work, critical revision of the article and final approval of the version to be published; AHE helped in the data collection and final approval of the version to be published; YR was involved in the data collection and final approval

of the version to be published; SM was involved in the data collection and final approval of the version to be published; and SMM contributed to the data collection and final approval of the version to be published.

Data availability The database analyzed during the present study is available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval Not applicable.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication The authors declare that they agree with the publication of this paper in this journal.

References

Abu-Allaban M, Abu-Qudais H (2011) Impact assessment of ambient air quality by cement industry: a case study in Jordan. Aerosol Air Qual Res 11(7):802–810. https://doi.org/10.4209/aaqr.2011.07.0090

Aghamolaie I, Lashkaripour GR, Ghafoori M (2015) Assessment of air pollution from cement industry. Iran Occupational Health 12(2):9–92. http://ioh.iums.ac.ir/article-1-1314-en.html

Agrawal M, Khanam N (1997) Variations in concentrations of particulate matter around a cement factory. Indian J Environ Health 39(2):97–102

Agrawal KP, Garg S, Sharma S, Patel P, Bhatnagar A (2017) Fusion of statistical and machine learning approaches for time series prediction using earth observation data. Int J Comput Sci Eng 14(3):255–266. https://doi.org/10.1504/IJCSE.2017.084159

Ahmad W, Nisa S, Mohammad N, Hussain RAHIB (2013) Assessment of particulate matter (PM_{10} & $PM_{2.5}$) and associated health problems in different areas of cement industry, Hattar, Haripur. J Sci Technol 37:7–15

Akbal Y, Ünlü KD (2022) A deep learning approach to model daily particular matter of Ankara: key features and forecasting. Int J Environ Sci Technol 19(7):5911–5927. https://doi.org/10.1007/s13762-021-03730-3

Ali MB, Saidur R, Hossain MS (2011) A review on emission analysis in cement industries. Renew Sustain Energy Rev 15(5):2252–2261. https://doi.org/10.1016/j.rser.2011.02.014

Alizadehdakhel A, Ghavidel A, Panahandeh M (2010) CFD modeling of particulate matter dispersion from Kerman cement plant. Iran J Health and Environ 3(1):67–74. http://ijhe.tums.ac.ir/article-1-136-en.html

Assegaf AH, Jayadipraja EA (2015) Pemodelan Dispersi CO Dari Cerobong Pabrik Semen Tonasa dengan Menggunakan Model AERMOD. In: Seminar Nasional Fisika Makassar.

Baldasano JM, Valera E, Jiménez P (2003) Air quality data from large cities. Sci Total Environ 307(1–3):141–165. https://doi.org/10.1016/S0048-9697(02)00537-5

Bañuelos Gimeno J, Blanco A, Díaz J, Linares C, López JA, Navas MA, Sánchez-Martínez G, Luna Y, Hervella B, Belda F, Culqui DR (2022) Air pollution and meteorological variables' effects on COVID-19 first and second waves in Spain. Int J Environ Sci Technol, pp 1–14. https://doi.org/10.1007/s13762-022-04190-z

Beckerman BS, Jerrett M, Martin RV, van Donkelaar A, Ross Z, Burnett RT (2013) Application of the deletion/substitution/addition algorithm to selecting land use regression models for interpolating air pollution measurements in California. Atmos Environ 77:172–177. https://doi.org/10.1016/j.atmosenv.2013.04.024

Borhani F, Mirmohammadi M, Aslemand A (2017) Experimental study of benzene, toluene, ethylbenzene and xylene (BTEX) concentrations in the air pollution of Tehran, Iran. J Res Environ Health 3(2):105–115. https://doi.org/10.22038/jreh.2017.23688.1151

Borhani F, Zahed F, Noorpoor A (2019) Modeling and evaluating the contribution of NO_X and CO pollutants emitted in the insulation Bituminous units (Isogam) exhaust flue gas on the around area (Case study: Delijan City). New Sci Technol 1(2):91–100

Borhani F, Motlagh MS, Stohl A, Rashidi Y, Ehsani AH (2021a) Changes in short-lived climate pollutants during the COVID-19 pandemic in Tehran, Iran. Environ Monit Assessment 193(6):1–12. https://doi.org/10.1007/s10661-021-09096-w

Borhani F, Shafiepour Motlagh M, Stohl A, Rashidi Y, Ehsani AH (2021b) Tropospheric Ozone in Tehran, Iran, during the last 20 years. Environ Geochem Health, pp 1–23. https://doi.org/10.1007/s10653-021-01117-4

Borhani F, Shafiepour Motlagh M, Rashidi Y, Ehsani AH (2022a) Estimation of short-lived climate forced sulfur dioxide in Tehran, Iran, using machine learning analysis. Stoch Env Res Risk Assess, pp 1–14. https://doi.org/10.1007/s00477-021-02167-x

Borhani F, Shafiepour Motlagh M, Ehsani AH, Rashidi Y (2022b) Evaluation of short-lived atmospheric fine particles in Tehran, Iran. Arab J Geosci 15(16):1–10. https://doi.org/10.1007/s12517-022-10667-5

Borhani F, Noorpoor A (2017) Cancer risk assessment Benzene, Toluene, Ethylbenzene and Xylene (BTEX) in the production of insulation bituminous. Environ Energy Econ Res 1(3):311–320. https://doi.org/10.22097/eeer.2017.90292.1010

Borhani F, Noorpoor A (2020) Measurement of air pollution emissions from chimneys of production units moisture insulation (Isogam) Delijan. J Environ Sci Technol 21(12):57–71. https://doi.org/10.22034/jest.2020.25934.3488

Borhani F, Noorpoor A, Khalili K (2016) measuring and evaluation of non-hydrocarbon air pollutants emitted in the production of insulation bituminous (Isogam) exhaust flue gas. Education p. 335–343.

Borrego C, Tchepel O, Costa AM, Martins H, Ferreira J, Miranda AI (2006) Traffic-related particulate air pollution exposure in urban areas. Atmos Environ 40(37):7205–7214. https://doi.org/10.1016/j.atmosenv.2006.06.020

Brownlee J(2019) Introduction to time series forecasting with python. Jason Brownlee.

Carslaw DC, Ropkins K (2012) Openair—an R package for air quality data analysis. Environ Model Softw 27:52–61. https://doi.org/10.1016/j.envsoft.2011.09.008

Chapra SC, Canale RP (1988) Numerical methods for engineers. McGraw-Hill Inc., New York

Chehregani H (2004) Environmental engineering in cement industry, Tehran. Hazegh Publications, 1383:135–278

Cheney EW, Kincaid DR (2012) Numerical mathematics and computing. Cengage Learning, Boston

Cheraghi A, Borhani F (2016a) Assessing the effects of air pollution on four methods of pavement by using four methods of multi-criteria decision in Iran. J Environ Sci Stud 1(1):59–71

Cheraghi A, Borhani F (2016b) Evaluation of environmental and sustainable development of four pavements in Iran by four method of multi-criteria analysis. J Environ Sci Stud 1(2):51–62

Ciobanu C, Istrate IA, Tudor P, Voicu G (2021) Dust emission monitoring in cement plant mills: a case study in Romania. Int J Environ Res Public Health 18(17):9096. https://doi.org/10.3390/ijerph18179096

Dan EL, Dînşoreanu M, Mureşan RC (2020) Accuracy of six interpolation methods applied on pupil diameter data. In 2020 IEEE international conference on automation, quality and testing, robotics (AQTR) (pp. 1–5). IEEE, New York. https://doi.org/10.1109/AQTR49680.2020.9129915

De Nevers N (2000) Air pollution control engineering, 2nd edn. McGraw Hill, International Edition

Dettling M (2013) Applied time series analysis. Zurich: Zurich University of Applied Sciences, pp. 203.

Díaz-Robles LA, Ortega JC, Fu JS, Reed GD, Chow JC, Watson JG, Moncada-Herrera JA (2008) A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. Atmos Environ 42(35):8331–8340. https://doi.org/10.1016/j.atmosenv.2008.07.020

Elavarasan D, Vincent DR, Sharma V, Zomaya AY, Srinivasan K (2018) Forecasting yield by integrating agrarian factors and machine learning models: a survey. Comput Electron Agric 155:257–282. https://doi.org/10.1016/j.compag.2018.10.024

Field AP, Miles J (2009) Discovering statistics using SPSS (and sex and drugs and rock'n'roll). London: Sage.

Goap A, Sharma D, Shukla AK, Krishna CR (2018) An IoT based smart irrigation management system using Machine learning and open-source technologies. Comput Electron Agric 155:41–49. https://doi.org/10.1016/j.compag.2018.09.040

Gokhale S (2009) Air pollution sampling and analysis. QIP, Indian Institute of Technology-Guwahati, Assam, India, 47p.

Hoveidi H, Aslemand A, Borhani F, Naghadeh SF (2017) Emission and health costs estimation for air pollutants from municipal solid waste management scenarios, case study: NOx and SOx pollutants, Urmia, Iran. J Environ Treatment Tech 5(1):59–64

Humphreys K, Mahasenan M (2002) Towards a sustainable cement industry. Substudy 8: climate change. An Independent Study Commissioned to Battelle by World Business Council for Sustainable Development. https://www.osti.gov/etdeweb/biblio/20269589

Jayadipraja EA, Daud A, Assegaf AH (2016) Air pollution and lung capacity of people living around the cement industry. Public Health Indonesia 2(2):76–83

Jeong Y, Youn Y, Cho S, Kim S, Huh M, Lee Y (2020) Prediction of Daily PM_{10} concentration for Air Korea stations using artificial intelligence with LDAPS weather data, MODIS AOD, and Chinese air quality data. Korean J Remote Sensing 36(4):573–586. https://doi.org/10.7780/kjrs.2020.36.4.7

Jung I, Chong K (2017) Interpolation and spatial matching method of various public data for building an integrated database. WIT Trans Built Environ 176:307–318

Kerman cement industries group, Kcig (2020) https://kcig.ir/

Kerman meteorological administrative, Kerman-met (2020) http://kerman-met.ir/

Leone V, Cervone G, Iovino P (2016) Impact assessment of PM_{10} cement plants emissions on urban air quality using the SCIPUFF dispersion model. Environ Monit Assess 188(9):1–12. https://doi.org/10.1007/s10661-016-5519-5

Maddah S, Bidhendi GN, Borhani F, Taleizadeh AA (2022) Resilient-sustainable supplier selection considering health-safety-environment performance indices: a case study in automobile industry. https://doi.org/10.21203/rs.3.rs-2046543/v1

Masood A, Ahmad K (2020) A model for particulate matter ($PM_{2.5}$) prediction for Delhi based on machine learning approaches. Proc Comput Sci 167:2101–2110. https://doi.org/10.1016/j.procs.2020.03.258

McMurry PH (2000) A review of atmospheric aerosol measurements. Atmos Environ 34(12–14):1959–1999. https://doi.org/10.1016/S1352-2310(99)00455-0

Merenu IA, Mojiminiyi F, Njoku CN, Ibrahim M (2007) The effect of chronic cement dust exposure on lung function of cement factory workers in Sokoto, Nigeria. Af J Biomed Res 10(2). https://doi.org/10.4314/ajbr.v10i2.50617

Mohebi A, Baroutian S (2006) A detailed investigation of particulate dispersion from Kerman cement plant. Iran J Chem Eng 3(3):65–74. http://www.ijche.com/article_15223.html

Mousavi SM, Falahatkar S (2020) Spatiotemporal distribution patterns of atmospheric methane using GOSAT data in Iran. Environ, Dev and Sustain. 22(5):4191–4207. https://doi.org/10.1007/s10668-019-00378-5

Mousavi SM, Falahatkar S, Farajzadeh M (2017) Monitoring of monthly and seasonal methane amplitude in Iran using GOSAT data. Phys Geogr Res Q 49(2):327–340. https://doi.org/10.22059/jphgr.2017.62848

Nazif A, Mohammed NI, Malakahmad A, Abualqumboz MS (2019) Multivariate analysis of monsoon seasonal variation and prediction of particulate matter episode using regression and hybrid models. Int J Environ Sci Technol 16(6):2587–2600. https://doi.org/10.1007/s13762-018-1905-6

Nemerow NL, Agardy FJ, Sullivan PJ, Salvato JA (2009) Environmental engineering: water, wastewater, soil and groundwater treatment and remediation. Wiley, New York

Özden Ö, Döğeroğlu T, Kara S (2008) Assessment of ambient air quality in Eskişehir, Turkey. Environ Int 34(5):678–687. https://doi.org/10.1016/j.envint.2007.12.016

Salvi, J. (2019). Significance of ACF and PACF plots in time series analysis. Towards Data Science, 27. https://towardsdatascience.com/significance-of-acf-and-pacf-plots-in-time-series-analysis-2fa11a5d10a8

Shahri E, Velayatzadeh M, Sayadi MH (2019) Evaluation of particulate matter $PM_{2.5}$ and PM_{10} (Case study: Khash cement company, Sistan and Baluchestan). J Air Pollut Health 4(4):221–226. https://doi.org/10.18502/japh.v4i4.2196

Sharma R, Pervez S (2003) Seasonal variation of PM10 and SPM levels in ambient air Around a cement plant. J Sci Ind Res. 62:827–833

Suhartono S (2011) Time series forecasting by using seasonal autoregressive integrated moving average: subset, multiplicative or additive model. J Math Stat 7:20–27

Vidnerová P, Neruda R (2021) Air pollution modelling by machine learning methods. Modelling 2(4):659–674. https://doi.org/10.3390/modelling2040035

Wang Z, Chen L, Tao J, Zhang Y, Su L (2010) Satellite-based estimation of regional particulate matter (PM) in Beijing using vertical-and-RH correcting method. Remote Sens Environ 114(1):50–63. https://doi.org/10.1016/j.rse.2009.08.009

WHO, World Health Organization (2013) Health Effects of Particulate Matter: Policy implications for countries in eastern Europe, Caucasus and central Asia.

Zhang H, Zhang S, Wang P, Qin Y, Wang H (2017) Forecasting of particulate matter time series using wavelet analysis and wavelet-ARMA/ARIMA model in Taiyuan, China. J Air Waste Manag Assoc 67(7):776–788. https://doi.org/10.1080/10962247.2017.1292968