Wavelet spectrum analysis of PM10 data in Bangkok, Thailand

P Anusasananan
Department of Physics, Faculty of Science, Ramkhamhaeng University, Bangkok 10240, Thailand

E-mail: panatcha02@hotmail.com

Abstract. PM10 is one of the key factors which influences the air quality in metropolitan areas throughout the world. In this research we investigate the variation period of PM10 concentration and its temporal patterns from 2009 to 2017 in Bangkok, Thailand using wavelet spectrum analysis. We also utilize the cross wavelet transform to study the potential relation between PM10 and the temperature. From the wavelet power spectrum of PM10, we can distinguish two dominant bands, one in the period between 8-16 months and the other between 1-8 months. The first oscillation is obviously related to natural annual periodicities that the high power occurs from December 2011 to February 2015. The PM10 concentrations are high in winter and low in the rainy season. The second band between 1-8 months is the transient pattern with very high concentration of PM10 occurs only in the year 2013. Wavelet spectrum of the temperature is similar in pattern for it shows the strong annual signal and the lowest temperature band between 4-8 months. The cross wavelet transform (XWT) power spectra between PM10 and the temperature show significant common power, in the 8-16 month band from November 2009 to June 2016, and in the 1-8 month band around November 2012 to February 2014. The wavelet transform coherence (WTC) spectra show that PM10 and the temperature co-varied out of phase during all observed time intervals. This indicates the high seasonal dependence between PM10 and the temperature. Knowledge of the variation period of PM10 concentration and its evolution feature in Bangkok area will help the authorities to better prepare for public health and environmental hazard from the next PM10 pollution.

1. Introduction
Particulate matter (PM10) is one of the primary air pollutants affecting the air quality in metropolitan areas throughout the world. The main sources of PM10 are vehicular traffic, combustion and agriculture, domestic fuel burning, natural dust and salt, and industrial activities [1, 2]. These particles have a direct impact on human health via inhalation. According to World Health Organization (WHO), in 2012 ambient air pollution contributed to 6.7% of all deaths worldwide. There are several research studies devoted to the analysis and prediction of the concentration of PM10 in advance. In this research, we study the PM10 concentration using wavelet transform. The simplest method to analyze a PM10 time series would be to compute statistics such as the mean, the maximum and the minimum for different time periods (days, months and years) and see if they change by using regression analysis. Another method to extract periodic features in the PM10 data is to utilize a wavelet transform (WT). Wavelets can keep track of time and frequency information. Jean Morlet introduced the concept of
wavelets as a new tool to study seismic signals in 1982 [3]. Recently, wavelet analysis has been used for studies in many fields, i.e., atmospheric climate, geology, biology, electrocardiogram data, temperature variation and global warming, or the relationship between two time series [4, 5]. The wavelet transform introduces a useful representation of a function in the time-frequency domain. Mathematically, a wavelet is a function which a zero average that can be defined as follows:

\[ \int_{-\infty}^{\infty} \psi(t) dt = 0 \]

Wavelets can be created from the function called the Mother wavelet. One particular wavelet used in this research, the Morlet, is defined as

\[ \psi(t) = \pi^{-1/4} \exp(\text{i} \omega_0 t) \exp(-\frac{t^2}{2}), \]

where \( \omega_0 \) is dimensionless frequency and \( t \) is dimensionless time.

The cross wavelet transform (XWT) method is a technique that characterizes the interaction between the wavelet transform of two individual time-series. The WT method can be extended to the analysis of the interaction between time-series. The XWT of two time-series \( W_{\text{est}} \) is the product of \( W_{\omega_0} \) with the complex conjugate of \( W_{\tau} \). Moreover, the XWT also allows us to access the continuous relative phase of two time series for each of the main frequencies.

Autoregressive Integrated Moving Average (ARIMA) model, also known as Box Jenkins model, is widely used in time series forecasting because of its flexibility in representing different time series, i.e., pure autoregressive (AR), pure moving average (MA) and combined AR and MA (ARMA) series. The ARIMA model is usually denoted as ARIMA \((p, q, d)\). Here \( p \) is the number of autoregressive orders that specify which previous values from the series are used to predict current values. The order of differencing, \( d \) is applied to the series before estimating model [6, 7].

2. Methodology

2.1 Study area and data

The monthly PM10 and temperature data used in this study were taken from a meteorological station in Bangkok, Thailand. Bangkok, the capital city of Thailand, is located at 13.75° latitude and 100.50° longitude and it is situated at the elevation 1.5 meters above sea level. Bangkok has a population of 8 million, making it the biggest city in Thailand. The city occupies 1,568.7 square kilometers in the Chao Phraya basin. The amount of PM10 was collected and recorded every day from the year 2009 to 2018, and a total of more than 108 data points were analysed. The data is collected by TEOM 1405 monitor which is furnished with software for personal computers. In all 108 data collected only 5 data are missing because of power failure so we used the average values instead. We first calculated the monthly mean of PM10 using the data from 2009 to 2018. Then, we used wavelet analysis to identify PM10 trend series with significant multi-temporal scales and the cross wavelet transform to investigate the relation between PM10 and temperature. Finally, we predict the PM10 using the data from 2009 to 2017 as the training set and the data from 2018 to verify the predicted values. We also predict the PM10 levels for the year 2019 in advance.

2.2 Wavelet analysis

The Morlet wavelet transform of a time series \((x_s)\) is defined as the convolution of the series with a set of “wavelet daughters” generated by the mother wavelet by translation in time by \( \tau \) and scaling by \( s \):

\[ \text{Wave} (\tau, s) = \sum \frac{1}{s^{1/2}} x_s \psi^* \left( \frac{t-\tau}{s} \right), \]

with * denoting the complex conjugate.

The position of the particular daughter wavelet in the time domain is determined by the localizing time parameter \( \tau \) being shifted by a time increment of \( dt \). The choice of the set of scales \( s \) determines the wavelet coverage of the series in the frequency domain.

The local amplitude of any periodic component of the time series under investigation, and how it evolves with time, can then be retrieved from the modulus of its wavelet transform.

\[ \text{Ampl}(\tau, s) = \frac{1}{s^{1/2}} |\text{Wave} (\tau, s)|. \]
The square of the amplitude has an interpretation as time frequency (or time period) wavelet energy density, and is called the wavelet power spectrum.

\[ \text{Power}(\tau,s) = \frac{1}{\tau} |\text{Wave}(\tau,s)|^2 \]

The XWT spectrum is defined by:

\[ W_{XY}(s,t) = W_X(s,t)W_Y^*(s,t), \]

where \(^*\) denotes the complex conjugate of \(W_Y(s,t)\).

### 2.3 Prediction of PM10

In forecasting with ARIMA, the model is fitted using the training data set (from 2009 to 2017) and then forecast the fitted model over the validation period (2018). The procedure in the experiment consists of the following steps: 1) Preprocessing step which includes data clearing, such as identification of the potential errors in data sets, handling missing values, and removal of noises or other unexpected results that could appear during the acquisition process. 2) Use the wavelet transform to decompose the data for the training set. 3) After obtaining the wavelet decomposition, we select the information from each level of decomposition for building the model. 4) In the training phase we design predictive models for each of the decomposed components of the original series. In the test phase the developed forecasting models are used to predict future values for each component. There are some software available for the applications of the wavelet transform. All of our calculation in this work use the R programming language.

### 3. Results and Discussion

![Time series of monthly PM10 and temperature in Bangkok.](image)

**Figure 1.** Time series of monthly PM10 and temperature in Bangkok.

The PM10 concentration and the temperature measured in Bangkok in the period 2009-2017 are shown in figure 1. It is obvious from figure that the PM10 concentration and the temperature time series shows the periodicity in one-year period. The monthly average of PM10 concentration is 20-60 \(\mu g/m^3\), and the yearly average over this period is 41.9 \(\mu g/m^3\). The monthly temperature is between 24-32°C, with the mean of 28.95°C. The maximum PM10 concentration is 131 \(\mu g/m^3\) in December 2013.
and 111 μg/m$^3$ in January 2014, while the temperature is lowest in December 2013 (24.2°C) and in January 2014 (24.0°C). This phenomenon suggests that PM10 concentration and temperature move in opposite directions. When temperature decreases, the PM10 increases, and vice versa.

3.1 Trend of PM10 and temperature

To study the PM10 concentration and temperature in a one year period, we calculated the average PM10 and average temperature in each month over the period of nine years from 2009 to 2017, as shown in figure 2.

![Figure 2](image)

Figure 2. The average PM10 and temperature in each month from 2009 to 2017.

It is evident from this graph that the trend of PM10 and temperature values move in opposite directions. The PM10 is maximum in January and linearly decreases until mid-March. From here, the PM10 concentration is nearly stable from April until mid-September, when its concentration increases linearly again to reach the maximum in December. In the opposite way, the temperature is lowest in January and then increases to the maximum in mid-April. The temperature continues to drop and reaches the lowest value in December. Whenever the temperature increases, the PM10 concentration decreases, and when the temperature decreases, the PM10 concentration increases.

3.2 Wavelet Power Spectrum

The wavelet power spectrum of monthly PM10 concentration in Bangkok over a period of nine years (January 2009 to December 2017) is presented in figure 3. The horizontal axis represents time in months and the vertical axis represents period in months. The colors in the figure stand for the structure of PM10 concentration variety, with the power range from weak (blue shades) to strong (red shades). The absolute value squared gives information on the relative power at a certain period and in a certain month. The WT power spectra of PM10 in figure 3 are evaluated using nine years of data and over a period from 2 to 32 months. As we can see from this figure, there are two dominant frequencies for PM10 fluctuations, one in the period of 8-16 months and another one for 4-8 months. The first oscillation is obviously connected with natural annual periodicities that the high power occurs between 2012 and 2015. The second oscillation occurs during almost the entire year in 2013, the same as
observed data that show high PM10 concentration in 2013. The wavelet power spectrum of monthly temperature in Bangkok between January 2009 and December 2017 is also presented in figure 3. As we can see, the pattern looks similar to the wavelet power spectrum of PM10, suggesting that there is a correlation between temperature and PM10. The first band, 8-16 months, occurs from 2010 to 2011 and from 2012 to 2016. The second band, 4-8 months, occurs from 2013 to 2014, at the same time as PM10. There are no other research to study PM10 concentrations using wavelet analysis in Thailand. However, Yija Liang, et al. study the temporal pattern of PM2.5 and its association with influenza in Beijing, China. Their results from Wavelet power spectrum analysis of monthly PM2.5 concentrations from 2008 to 2013 also show the periodic oscillation around 12-14 month band, the same as our results [8].

![Figure 3](image3.png)

**Figure 3.** The wavelet power spectrum of PM10 (above) and temperature (below).

### 3.3 Cross wavelet transform (XWT) analysis

The XWT between PM10 concentration and temperature is shown in figure 4. The XWT finds regions with high common power. The cross wavelet power spectra between PM10 and temperature shows significant common power in two significant bands. The first band covers 8-16 months between November 2009 to June 2016 and the second band between 1-8 month around November 2012 to February 2014.

![Figure 4](image4.png)

**Figure 4.** XWT between PM10 and temperature time series.
3.4 Wavelet Transform coherence (WTC)

To explore and illustrate the co-movement between PM10 concentration and temperature at different time-frequency (period) spaces, the WTC approach is applied. The WTC indicates the regions in time-frequency space where the PM10 and temperature time series co-vary. The phase difference between the two time series is defined by arrows. The orientation of the arrows indicate the level and type of correlation. Wavelet coherence between PM10 concentrations is shown in figure 5. The colored shading represents the wavelet squared coherence. The thick black line represents the 95% significant level. The arrow denotes the phase relationship between PM10 and temperature. The arrows are pointing slightly to the left, which indicates that PM10 and temperature are out of phase (anti-phase). When the temperature increases, the PM10 decreases, and vice versa. The high common powers are shown in the period band of 8-16 months between the entire observation period with the temperature is leading, and 4-8 months between mid-2012 to mid-2014 with the PM10 is leading.

![Figure 5. WTC between PM10 and temperature time series.](image)

3.5 Forecasting with ARIMA model

First, we examine the PM10 data and we can see that it’s very strong seasonal dependent as we have

![Figure 6. Wavelet decomposition of PM10 time series.](image)
already seen from the wavelet power spectrum. They do not appear to be any outliers and there are no missing values therefore no data cleaning is required. Next, we will decompose the time series for estimates of trend, seasonal, and random components using the decompose function in R. The results are shown in figure 6.

After the decomposition, we tested stationarity of the time series by running the Augmented Dickey-Fuller Test using the `adf.test` function from the time series R package [9]. The results show that the p-value is less than 0.01, therefore we reject the null in favour of the alternative hypothesis that the time series is stationary. We concluded that the time series is non-stationary. Now, we have to transfer the time series from non-stationary state to a stationary state using first-order differencing for such transformation. Differencing a series will remove trends. Next, fit an ARIMA model using `arima()` function or `auto.arima()` function. The forecast package allows the user to explicitly specify the order of the model using the `arima()` function, or automatically generate a set of optimal (p, d, q) using `auto.arima()`. This function searches through combinations of order parameters and picks the set that optimizes model fit criteria. Finally we can plot a forecast of the time series using the forecast function, again from the forecast R package, with a 95% confidence interval. The dark blue line in figure 7 shows the fit provided by the model.

![Figure 7. The fit on seasonal time series.](image)

The comparison between the forecast PM10 and the actual PM10 from January 2018 to November 2018 (the recent PM10 data we have until now) is shown in figure 8. The forecast values are very close to the actual values with the correlation coefficient equal to 0.9284. In January, February and March 2018, the forecast values are less than the actual values. On these three months most of the Thai farmers will prepare their rice fields for the next rice-planting season by burning the left over straw to eliminate the remaining hulls of the harvested rice. Unfortunately, the burning impacted the air quality of the region during this burning season causing the high PM10 concentration across Thailand. In addition, the calculation of predicted values are based on the past PM10 on every months. These may caused the forecast values to be lower from January to March and nearly equal from April to November. Actually we also forecast up to another 13 months ahead from December 2018 until December 2019 although we do not have the Actual values to compare. Longer term forecast will usually have more uncertainty, as the model will regress future forecast values on previously predicted values.
4. Conclusion
In this work, we studied the variation period of PM10 and its relation with the temperature by using wavelet spectrum analysis and also forecast the PM10 levels using ARIMA model. The PM10 concentration is high in December and January, while the temperature is low during these months. The cross wavelet analysis indicates that PM10 and temperature are co-varied out of phase. Our study also shows that during this nine year period, the level of PM10 is surprising high in December 2013 and January 2014. This should be seriously concerned since it is tripled than normal average values and it can make very harmful to our health. We also noticed that on these two months the temperatures were also lower than normal. So, clearly, it is important to pay attention to both the high PM10 levels and the unusual decrease in temperature. The calculation using ARIMA model shows that the predicted values and the actual values of PM10 agree quite well in the year 2018 with the correlation coefficient equal to 0.9284. The authority must pay attention to both the high PM10 levels in December to March every years, the transient high PM10 in this period and the forecast of PM10 level to better inform the public as a precautionary measure.

Acknowledgments
The author is grateful to the pollution control department for PM10 data and to the Department of Physics, Faculty of Science, Ramkhamhaeng University, Bangkok, for their support.
References

[1] Du Y, Xu X, Chu M, Guo Y and Wang J 2016 J. Thorac. Dis. 8 E8
[2] Vallero D A 2008 Fundamentals of Air Pollution (North Carolina: Academic Press)
[3] Morlet J, Arens G, Fourgeau E and Glard D 1982 Geophysics 47 203
[4] Torrence C and Compo G P 1998 Bull. Am. Meteorol. Soc. 79 61
[5] Bloomfield D S, McAteer R T J, Lites B W and Judge P G 2004 Astrophys. J. 617 623
[6] Montgomery D C, Jennings C L and Kulah M 2008 Introduction to Time Series Analysis and Forecasting (New Jersey: John Wiley & Sons)
[7] Shumway R H and Stoffer D S 2011 Time Series Analysis and Its Applications (New York: Springer)
[8] Liang Y, Fang L, Pan H, Zhang K, Kan H, Brook J R and Sun Q 2014 Environ. Health 13 102
[9] Tilman M D 2016 The Book of R (San Francisco: No Starch Press)