Spectral Analysis of Air Quality Index Using Haar Wavelet Transforms

Asad Ali Khan[1], Anil Kumar[2]

[1]Assistant Professor, Department of Physics, Moradabad Muslim Degree College, Moradabad-244001

[2]Professor, Department of Physics, Hindu College, Moradabad-244001

Affiliated to Mahatma Jyotiba Phule Rohilkhand University, Bareilly, Uttar Pradesh, INDIA

Abstract: ***Ambient air pollution is one of the big problems of today in perspective of environment and health degradation. The air quality of any place is measured in terms of the air quality index (AQI). The AQI of any place is decided by the highest value of one of the air pollutants out of the Particulate matter, Nitrogen dioxide, Ammonia, Sulphur Dioxide, Carbon Monoxide and Ozone of that place. Wavelet transform has become an important tool in the non-stationary and transient data/signal processing due to its time-frequency localization property. In wavelet transform as the scale increases, the resolution decreases, and a better estimate of the unknown trend of the data is obtained. We have selected Anand Vihar, Delhi, India as our study area and its daily average AQI data for one year before first lockdown (from March 25, 2019 to March 24, 2020) and during lockdown (March 25, 2020 to June 08, 2020) is taken as the raw data. The spectral analysis of this AQI data is performed by using Haar wavelet, level-7 with help of MATLAB wavelet toolbox. The highest scale approximation i.e. trend of data and histogram of the data are shown and discussed. Some statistical parameter like average, kurtosis, skewness and standard deviation are also determined and discussed.***

Keywords: AQI, approximation, detail, Haar, trend

1. Introduction

Ambient air pollution is defined by presence of particles and gases whose concentration becomes at a harmful level because of health-related problems and environmental degradation. Particulate matter (PM-2.5 and PM-10), Nitrogen dioxide (NO_2), Ammonia (NH_3), Sulphur Dioxide (SO_2), Carbon Monoxide (CO) and Ozone (O_3) are the main air pollutants. The combustion of petroleum products, industrialization, urbanization and anthropogenic activities are the main factors behind air pollution. The air pollution is measured in terms of Air Quality Index (AQI) based upon maximum value of any air pollutant out of PM-2.5, PM-10, NO_2, NH_3, SO_2, CO and O_3 [1, 2]. Short term exposure of poor air quality becomes the reason of sneezing and coughing, eye irritation, headaches, and dizziness, while long term exposure becomes the reason of severe health problems like cancer, heart disease, stroke, and respiratory diseases such as asthma, while [3]. On basis of health perspective, the AQI are categorize as follows: -

Table 1: AQI categories

S. No.	Category	AQI Range
1	Good	0 - 50
2	Moderate	51 - 100
3	Poor	101 - 200
4	Unhealthy	201 - 300
5	Severe	301 - 400
6	Hazardous	400 +

The Fourier transforms analyse stationary signals well but it is not capable to analyse the non-stationary and transient signals. To analyse the non-stationary and transient signals wavelet transforms is frequently being used due to its time-frequency localization property. Wavelets are a special kind of functions which exhibit oscillatory behaviour for a short period of time and then die out. For any two real numbers a and b, a wavelet function is defined as [4, 5]: -

$$\psi_{a,b} = \frac{1}{\sqrt{|a|}}\psi\left(\frac{x-b}{a}\right)$$

If we choose $a = 2^{-j}$ and $b/a = k$, we get discrete wavelets as follows: -

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$$

The wavelet transform of a signal captures the localized time frequency information of the signal. Wavelets satisfy the Heisenberg inequality, and hence, the time resolution becomes arbitrarily good at high frequency, while the frequency resolution becomes arbitrarily good at low frequencies. By wavelet transforms a signal is decomposed into approximation and detail. The approximation provides the average or trend of the signal, while the detail provides differential behaviour of the signal [6]. In wavelet analysis terms, the trend corresponds to the greatest scale value. As the scale increases, the resolution decreases, producing a better estimate of the unknown trend.

2. Basics of Spectral Analysis Using Wavelet Transforms

The spectral analysis of any data/signal using wavelet transforms depends upon the following basis ideas:

2.1 Multi-Resolution Analysis (MRA)

An MRA consists of a sequence $V_j : j \in \mathbb{Z}$ of closed subspaces of $L^2(\mathbb{R})$, a Lebesgue space of square integrable functions, satisfying the following properties [7, 8]: -

1) $V_{j+1} \subset V_j \quad : j \in \mathbb{Z}$
2) $\cap_{j\in\mathbb{Z}} V_j = \{0\}$, $\cup_{j\in\mathbb{Z}} W_j = L^2(\mathbb{R})$,
3) For every $L^2(\mathbb{R})$,

$$f(x) \in V_j \Rightarrow f\left(\frac{x}{2}\right) \in V_{j+1}, \quad \forall\ j \in \mathbb{Z}$$

4) There exists a function $\phi(x) \in V_0$ such that $\{\phi(x-k): k \in \mathbb{Z}\}$ is orthonormal basis of V_0.

The function $\phi\,(x)$ is called scaling function of given MRA and property 3 implies a dilation equation as follows: -

$$\phi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k\ \ \phi(2x-k)$$

where h_k is low pass filter and is defined as: -

$$h_k = \int_{-\infty}^{\infty} \phi\,(x)\phi\,(2x-k)\,dx$$

The wavelet function ψ is expressed as: -

$$\psi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} g_k\ \ \phi(2x-k)$$

where g_k is high pass filter and is defined as: -

$$g_k = \int_{-\infty}^{\infty} \psi\,(x)\phi\,(2x-k)\,dx$$

where, $g_k = (-1)^{k+1}h_{1-k}$. From conditions of multi-resolution analysis (MRA) and elementary functional analysis, each space V_{j-1} can be decomposed into subspace V_j and W_j such that every function f in V_{j-1} can be uniquely decomposed into $f = u + v$ with $u \in V_j$ and $v \in W_j$. We write this as follows: -

$$V_{j-1} = V_j \oplus W_j, \qquad \forall j \in \mathbb{Z}$$

If all such functions u and v are orthogonal ($\langle u, v\rangle = 0$), then W_j is the orthogonal complement of V_j in V_{j-1} ($V_j \perp W_j$) and the construction below will give the scaling function and mother wavelet of an orthonormal wavelet basis for $L^2((\mathbb{R})$ [9]. By MRA, the orthogonal decomposition of space $L^2(\mathbb{R})$ is as follows: -

$$L^2(\mathbb{R}) = \sum_j V_j = \sum_j \left(V_{j+p} \oplus \sum_{p=1}^{\infty} W_{j+p}\right)$$

2.2 Haar Wavelet

Haar discovered the simplest solution and at the same time open one of the routes leading to wavelets [10]. Haar begins with the function that is equal to 1 on $\left[0, \frac{1}{2}\right]$ and -1 on $\left[\frac{1}{2}, 1\right]$, and 0 outside the interval $[0, 1]$.

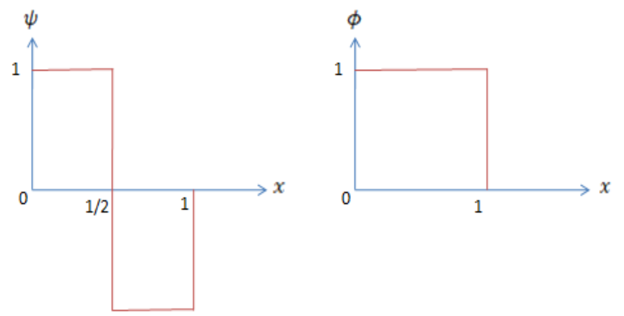

Figure 1: Haar wavelet and scaling function

Haar wavelet is constructed from the MRA generated by scaling function $\phi(x)=\chi_{[0,1]}(x)$. Since,

$\phi(x) = \phi(2x) + \phi(2x-1) = \chi_{\left[\frac{1}{2},1\right]} + \chi_{\left[0,\frac{1}{2}\right]}$

$\psi(x) = \phi(2x) - \phi(2x-1) = \chi_{\left[\frac{1}{2},1\right]} - \chi_{\left[0,\frac{1}{2}\right]}$

2.3 Data Analysis

We can approximate the data in space of square summable sequences $\ell^2(\mathbb{Z})$ as follows [11]: -

$$f[n] = \frac{1}{\sqrt{M}}\sum_k c[j+p,k]\phi_{j+p,k}[n] + \frac{1}{\sqrt{M}}\sum_{p=1}^{\infty}\sum_k w[j+p,k]\psi_{j+p,k}[n]$$

where $f[n]$, $\phi_{j,k}[n]$ and $\psi_{j,k}[n]$ are discrete functions defined in $[0, M-1]$, totally M points. We can simply take the inner product to obtain the wavelet coefficients,

$$c[j+p,k] = \frac{1}{\sqrt{M}}\sum_n f[n]\,\phi_{j+p,k}[n]$$

$$w[j+p,k] = \frac{1}{\sqrt{M}}\sum_n f[n]\,\psi_{j+p,k}[n]$$

where $c[j+p,k]$ and $w[j+p,k]$ are called approximation and detailed coefficients respectively. From property of scaling function,

$$\phi_{j,k}[n] = 2^{j/2}\phi[2^j n - k]$$
$$= \sum_{n'} h[n']\sqrt{2}\phi[2(2^j n - k) - n']$$

Let $n' = m - 2k$, we have $\phi_{j,k}[n] = \sum_m h[m - 2k]\sqrt{2}\phi[2^{(j+1)}n - m]$.

Now the approximation coefficient,

$$c[j+1,k] = \frac{1}{\sqrt{M}}\sum_n f[n]\,\phi_{j+1,k}[n]$$
$$= \frac{1}{\sqrt{M}}\sum_n f[n]\,2^{(j+1)/2}\phi[2^{(j+1)}n - k]$$

Therefore, we can write,

$$\frac{1}{\sqrt{M}}\sum_n f[n]\,2^{(j+1)/2}\sum_m h[m-2k]\ \sqrt{2}\phi[2^j n - m]$$
$$= \sum_m\ h[m-2k]\left(\frac{1}{\sqrt{M}}\sum_n f[n]\,2^{j/2}\,\phi[2^j n - m]\right)$$
$$= \sum_m\ h[m-2k]c[j,n]$$
$$= h[n'] * c[j,n]$$

where $k \geq 0$. Similarly, for the detail coefficients, we can write,

$w[j+1,k] = g[n'] * c[j,n]$, where $k \geq 0$.

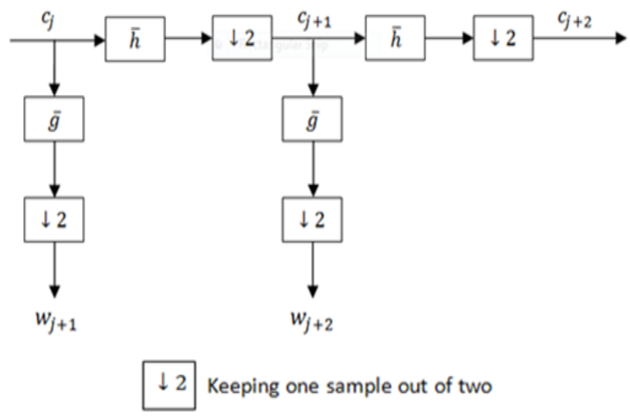

Figure 2: Pyramidal algorithm of data decomposition

Paper ID: SR24430181336 DOI: https://dx.doi.org/10.21275/SR24430181336

3. Study Area and Methodology

The average daily record of AQI of Anand Vihar, Delhi from March 25, 2019 to June 08, 2020 has been taken as the primary data. The data of AQI from March 25, 2019 to March 24, 2020 represents to one year data before the lockdown imposed, while from March 25, 2020 to June 08, 2020 to the lockdown period.

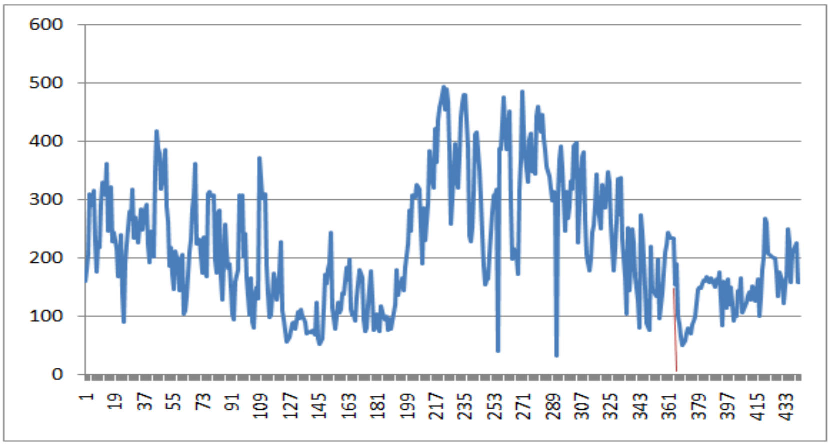

Figure 3: AQI from March 25, 2019 to June 08, 2020

Wavelet transform of the signal is performed using Haar wavelet up to level 7 using wavelet toolbox of software MATLAB. The trend of the AQI is shown by 7th order approximation. Approximation describes the average behaviour of signal and represents to the trend of signal. A further characterization of the data includes skewness and kurtosis [12]. Skewness is a measure of the lack of symmetry. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case. The skewness for a normal distribution is zero, and any symmetric data should have skewness near zero. The standard deviation indicates how the data points are spread out over a wide range of values.

4. Results and Discussion

The wavelet transform of AQI data from March 25, 2019 to June 08, 2020 is performed using Haar wavelet, level 7average because the trend of any data corresponds to the greatest scale value. The trend of any signal represents the slowest part of the signal.

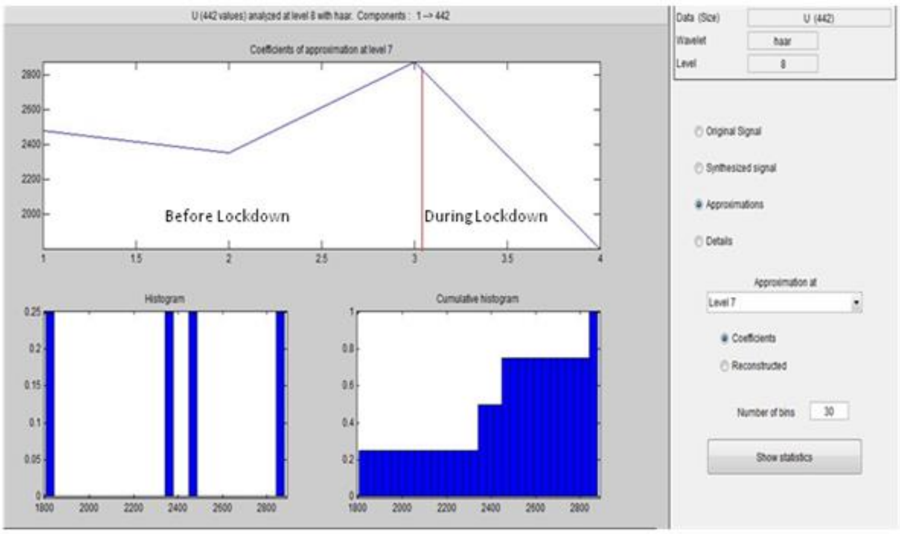

Figure 4: Trend of AQI

It is clear from the figure 4, that there is a decreased trend of AQI during the first lockdown. The histogram is an approximate representation of the distribution of numerical data. The statistical parameters for AQI are determined and enlisted as follows.

Table 2: Statistical Parameters of AQI

S. No.	Statistical Parameter	Before Lockdown	During Lockdown
1	Average	233.37978	146.28947
2	Kurtosis	-0.7302068	0.007787823
3	Skewness	0.322225784	0.236379653
4	Standard deviation	109.1409069	47.23919722

The average value of AQI is decreased very much due to lockdown imposed. The kurtosis which represents the peakedness of data is nearly equal to zero during the lockdown. The skewness of AQI data has low value during the lockdown which represents the more symmetry of data points. The standard deviation of AQI during the lockdown is also low, which indicates the spreading of data points is closer to mean value during lockdown.

5. Conclusion

In wavelet transforms the signal is decomposed into the approximation and detail, where approximation describes the average behaviour or trends of the signal. It is clear from wavelet analysis of AQI that during the first lockdown the trend of AQI is in the decreased mode. During the lockdown, the determined value of average, kurtosis, skewness and standard deviation indicates the decreased, less peaked, symmetric and less spreading of data to the mean value. On basis of above results, it is possible to conjecture that the wavelet analytical approach provides a simple and accurate framework for modelling the spectral and statistical behaviour of AQI variation.

References

[1] J. Flemming, R. Stern and R.J. Yamartino. “A New Air Quality Regime Classification Scheme for O_3, NO_2, SO_2 and PM10 Observations Sites.”, Atmospheric Environment, 39(2005), 6121–6129.

[2] C.Y. Mu, Y. Q. Tu and Y. Feng. “Effect Analysis of Meteorological Factors on the Inhalable Particle Matter Concentration of Atmosphere in Hami.” Meteorological Environmental Science, 34(2011), 75–79.

[3] A.J. Cohen. “Outdoor Air Pollution and Lung Cancer.” Environmental health perspectives, 108(2000), 743–750.

[4] J.C. Goswami and A.K. Chan. Fundamentals of Wavelets Theory, Algorithms and Applications, Willey-Interscience New York, 1999.

[5] J.P. Antoine. “Wavelet analysis: A new tool in Physics.” Wavelets in Physics, J.C. Van Den Berg, ed., 2004, pp. 9-21.

[6] S. Kumar, A. Kumar, R. Kumar, J.K. Pathak and M. Alam. “Spectral Analysis of Bio-chemical Oxygen Demand in River Water: An Analytical Approach of Discrete Wavelet Transforms.” American Journal of Mathematics and Statistics, 4.2 (2004), 107-112.

[7] S.G. Mallat. A Wavelet Tour of Signal Processing, Academic Press New York, 1998

[8] S.G. Mallat. “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation.” IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(1989), 674-693.

[9] I. Daubechies. “Ten Lectures on Wavelets.” CBS-NSF, Regional Conference in Applied Mathematics, SIAM Philadelphia, 61, 1992, pp. 278-285.

[10] E. Hernandez and G. Weiss, A First Course on Wavelets, CRC Press, New York, 1996.

[11] R.R. Coifman and M.V. Wicker Hauser. “Entropy Based Algorithms for Best Basis Selection.” IEEE Transactions on Information Theory, 38.2 (1992), 713-718.

[12] M. Rockinger and E. Jondeau. “Entropy Densities with an Application to Autoregressive Conditional Skewness and Kurtosis.” Journal of Econometrics, 106.1 (2002), 119-142.

Paper ID: SR24430181336 DOI: https://dx.doi.org/10.21275/SR24430181336