Chapter 9
Fuzzy Forecasting with Fractal Analysis for the Time Series of Environmental Pollution

Wang-Kun Chen and Ping Wang

Abstract. Environmental pollution, which is complicated for forecasting, is a phenomenon related to the environmental parameters. There are many studies about the calculations of concentration variation on pollution time series. A new framework of prediction methodology using the concept of fuzzy time series with fractal analysis (FTFA) was introduced. The FTFA uses the concept of turbulence structure with the fractal dimension analysis to estimate the relationship by fuzzy time series. The candidate indexes of each pattern can be selected from the most important factors by fractal dimension analysis with autocorrelation and cross correlation. Based on the given approach, the relationship between the environmental parameters and the pollution concentration can be evaluated. The proposed methodology can also serve as a basis for the future development of environmental time series prediction. For this reason, the management of environmental quality can be upgraded because of the improvement of pollution forecasting.

Keywords: Fuzzy Theory, Fractal Analysis, Environmental Pollution, Time Series.

1 Time Series of Environmental Phenomenon and Its Physical Nature

Environmental pollution is a kind of natural phenomenon which could be explained by the turbulence structure of fluid dynamics. The environmental properties such as pollution concentration, wind velocity, ocean current and thermal diffusion are all the outcome of natural turbulence. The air pollution is a typical phenomenon caused by pollutants emitted into the atmosphere and diffused with the eddy.

Wang-Kun Chen
Department of Environment and Property Management,
Jinwen University of Science and Technology
email: wangkun@just.edu.tw

Ping Wang
Department of Civil Engineering, Chinyun University
email: pwang@cyu.edu.tw

W. Pedrycz & S.-M. Chen (Eds.): Time Series Analysis, Model. & Applications, ISRL 47, pp. 199–213.
DOI: 10.1007/978-3-642-33439-9_9

The environmental phenomenon can be described by its physical properties. The presence of “eddies” in the environment leads to the complexity and variation of the outcome of observation. The length scale of “eddy” differs to several orders of magnitude. For example, typhoon is one of those large eddies exists in the atmosphere, and the sea breeze is the smaller eddy caused by the air-sea exchange. Thus it is very difficult to explain the difference by traditional methods. A better insight into the characteristics of turbulence with these “eddies” will be helpful in understanding the nature of these environmental phenomenon.

The modern principle of fluid dynamics and the theory of fractal analysis are suitable tools to investigate the properties of environmental events. The irregularity and the randomness are the most important characteristics of turbulent flow. These characteristics make it impossible to explain the environmental events using deterministic approach, except invoking statistical methods. The environmental phenomenon can be investigated by long term and large area monitoring. However, the complexity of the time series-based observation has made the interpretation more difficult. The models used to describe environmental turbulence should be able to simulate the non-linear and non-stationery properties of time series. Therefore, the recently developed tool, fractal analysis, can be employed to meet the needs.

Environmental pollution is a phenomenon resulting from the presence of turbulence, characterized by non-linear, randomness, irregularity, and chaos. Thus, turbulence is a complex environmental phenomenon that is difficult to predict precisely through mathematical modeling.

2 Interpretation of Pollution Time Series by Fractal Analysis

Since there are so many different eddies with different scales, the concept of fractal analysis become useful to understand the behavior of environmental turbulence. Fractal analysis, which expresses the complexity using the fractal dimension, is a contemporary method to describe the natural phenomenon. It applies the nontraditional mathematics in analyzing the environmental problem and has been used in the analysis of the scale dependence environmental phenomenon such as rainfall (Olsson & Niemczynowicz,1994,1996), air pollutant concentration (Lee,2002, Lee et al, 2003, Lee & Lin, 2008; Lee et al,2006a,), and earthquake (Lee et al,2006b).

Mandelbrot has defined fractal as a special class of subsets of a complete metric space (Mandelbrot,1982). The fractal dimension, D_F, which is deduced from the scaling rule, is the key concept of fractal analysis. The complexity of environmental phenomenon is due to a change with the variation of turbulence eddies in scale. So it is possible to have many types of fractal dimension, D_F, in an environmental system. These fractal dimensions can be explained in terms of measure of the complexity. Comparing with the change with scale in turbulence, so it is necessary to deduce a scaling system to represent the “patterns of complexity”

Here a simple model is proposed for estimation of pollution concentration influenced by environmental turbulence. Generally, the scaling rule or fractal dimension, D_F, can be represented by two terms, N and ε. The term N is the number of pieces and ε is the scale used to get new pieces. The relationship can be written as:

$$N \propto \varepsilon^{-D_F} \tag{1}$$

which can be further formulated in the form of a scaling rule:

$$N = A\ \varepsilon^{-D_F} \tag{2}$$

where A is a certain constant. By taking the logarithm of both sides of (2), the variable D_F becomes the ratio of the log of "the number of new parts (N)" to the log of "scale (ε)":

$$D_F = \log N/\log \varepsilon. \tag{3}$$

The scaling rule of fractal dimension helps us explain the variation of pollution time series in the fuzzy time series prediction with fractal analysis.

In turbulence, the attribute of correlated variable helps to characterize the phenomenon. The analysis starts from the average of products, which are computed in the following way. (Tennekes and Lumley, 1972)~

$$\begin{aligned}\overline{\tilde{u}_i\tilde{u}_j} &= \overline{(U_i+u_i)(U_j+u_j)} \\ &= U_iU_j + \overline{u_iu_j} + \overline{U_iu_j} + \overline{U_ju_i} \\ &= U_iU_j + \overline{u_iu_j}\end{aligned} \tag{4}$$

The terms consisting of a product of a mean value and a fluctuation vanish if they are averaged, because the mean value is a mere coefficient as far as averaging is concerned, and the average of a fluctuation quantity becomes zero.

If $\overline{u_iu_j} \neq 0$, u_i and u_j are said to be a correlated; if $\overline{u_iu_j} = 0$, the two variable are uncorrelated. Figure 1 illustrates the concept of correlated fluctuating variable. The correlation coefficient C_{ij}, is defined by

$$C_{ij} \equiv \frac{\overline{u_iu_j}}{\sqrt{\overline{u_i}^2 \cdot \overline{u_j}^2}}, \tag{5}$$

where C_{ij} is a measure for the degree of correlation between two variable u_i and u_j. If $C_{ij} = \pm 1$, the correlation is said to be perfect, and could be chosen as the best predictor for forecasting.

In the analysis of turbulence dimension, the "standard deviation" of "root mean square (RMS)" amplitude is defined. For a turbulent flow field, a characteristic velocity, or "velocity scale", might be defined as the mean RMS velocity taken across the flow field at that position. In this way velocity scale could be used as a precise definition in dimensional analysis.

If the evolution of fluctuating function (t) is to be described, it is necessary to know that the value of u at different time is related. The question could be answered by considering a joint density for u(t) and u'(t). The time difference, or time lag, in the property time series is defined by $\tau = t' - t$. The correlation $\overline{u(t)u(t')}$ at two different times is called the autocorrelation, and the correlation between u and v is called the cross-correlation.

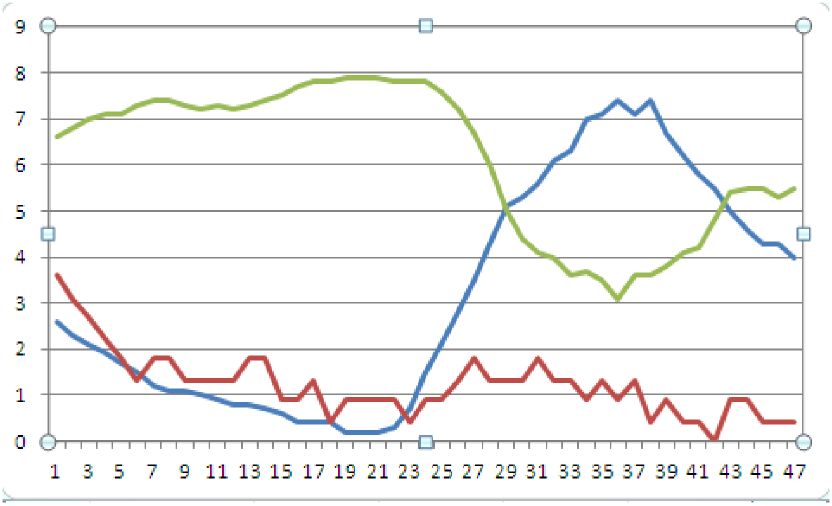

Fig. 1 The example of correlated variable and uncorrelated variable. The green and blue line has a negative correlation; however, the red line is uncorrelated to the two variables.

A tensor R to deal with the correlation between different location x and x+r is given by

$$R_{ij} \equiv \overline{u_i(x,t) \cdot u_i(x+t)}, \tag{6}$$

3 Representation of Environmental Phenomenon by Fuzzy Time Series

After interpretation of environmental phenomenon in terms of turbulence scale and fractal dimension analysis, we return to the problem of predicting the time series value of pollution. Time series is frequently applied to the prediction of environmental events. An example of using time series for pollutant concentration is given by

$$X = \{x_t \quad t = 1, \ldots\ldots, N\} \tag{7}$$

where t is time index and N is the total number of observations. For example, the time series of ozone concentration from a continuous monitoring station, the instantaneous wind velocity at the meteorological observation station is considered as an event for a time series.

The prediction method of fuzzy time series with fractal analysis is a scheme revised from Chen's study of fuzzy time series (Chen, 1996, 2002)(Chen and Hsu, 2004, 2008)(Chen and Hwang, 2000)(Song and Chrisson, 1993, 1994, 2003). The concept of fuzzy time series has been applied to the prediction of pollutant concentration by Chen (Chen, 2011), which produced good results. However, it is still unable to describe the non-linear characteristic of the environmental system.

On the other hand, the database is not extensive enough to generate a complete inference engine to simulate all the possible variation of time series.

In this study, due to an addition of 1440 data sets to the database for generating the inference engine, along with the fractal analysis of turbulence, more insight is given than before into the behavior of pollution concentration in the environment.

The prediction method of fuzzy time series with fractal analysis (FTFA) can be implemented by the following steps: (1) Define the interval. (2) Get the statistical distribution of concentration in each interval. (3) Define each fuzzy set Ai based on the re-divided intervals u_i derived in step 2. (4) Establish fuzzy logical relationship based on the fuzzified concentration. (5) Use the high-order difference to determine the upward or downward trend. (6) Find the appropriate predictors by fractal analysis.

Let U = {u_1, u_2, u_3,……,u_n}, where U is the universe of discourse. Fuzzy set A, in the universe of discourse U, is defined as follows:

$$A_i = fA_1(u_1)/u_1 + fA_2(u_2)/u_2 + \cdots + fA_n(u_n)/u_n \quad , \qquad (8)$$

where f_A is the membership function of the fuzzy set A, $f_A : U \rightarrow [0,1]$, $f_A(u_i)$ indicates the grade of membership of u_i in the fuzzy set A, $f_A(u_i) \in [0,1]$, and $1 \leqq i \leqq n$.

Define F(t) as the fuzzy time series of X(t) (t = ……, 0, 1, 2, ……), and X(t) (t = ……, 0, 1, 2, ……) is the universe of discourse in X (t). In order to extract the knowledge from the time series database, assume there exists a fuzzy relationship R (t, t-1) such that

$$F(t) = F(t-1) \cdot R(t, t-1) \qquad (9)$$

Where, R(t, t-1) denotes the fuzzy relationship between F(t) and F(t-1). If fuzzy set F(t-1)=A_i, and F(t)=A_j, the fuzzy relationship is called the first order fuzzy time series.

More hidden relationships could be found in the time series database. If F(t) is caused by F(t-1), F(t-2), F(t-3) ……,and F(t-n), then there is a high-order fuzzy time series which can be represented by

$$F(t-n), \ldots\ldots, F(t-2), F(t-1) \rightarrow F(t). \qquad (10)$$

The fuzzy interval of pollution time series can be an equal-length interval (ELI) or an un-equal length interval (ULI). ULI is the improved model of ELI by adjusting the length of each interval in the universe of discourse This is called the multi-step fuzzy time series for the forecasting of concentration (Chen, 2008) (Chen, 2010). The proposed method of FTFA is presented as follows:

Step 1: Define the interval

Let U be the universe of discourse, U = [D_{min} - D_1, D_{max} + D_2], where D_{min} and D_{max} denote the minimum and maximum concentration.

Step 2: Form the statistical distribution of concentration in each interval..

Sort the interval based on the number of concentration data falling into each interval in a descending sequence.

Step 3 : Define each fuzzy set Ai based on the re-divided intervals u_i derived in step 2, and fuzzify the historical concentration.

The interval with no data distributed was discarded. The interval with more data was divided into more sub-intervals. The idea behind the determination of interval and sub-interval is to divide the interval containing a higher number of historical concentration data into more sub-intervals to improve the accuracy of predict.

Step 4: Establish fuzzy logical relationship based on the fuzzified concentration.

If the fuzzified concentration of month i and i+1 are A_j and A_k, respectively, then construct the fuzzy logical relationship "$A_j \rightarrow A_k$", where A_j and A_k are called the current state and the next state of the concentration.

If the fuzzified concentration of month i is A_j and the fuzzy logical relationship is shown as:$A_j \rightarrow A_{k1(x1)}$, $A_{k2(x2)} \ldots A_{kp(xp)}$, then the estimated concentration of month i is calculated as

$$C(i) = \frac{X_1 \times m_{k1} + X_2 \times m_{k2} + \cdots + X_P \times m_{kp}}{X_1 + X_2 + \cdots + X_p} \tag{11}$$

where X_i denotes the number of fuzzy logical relationships "A_j A_k" in the fuzzy logical relationship group, $1 \leqq i \leqq p$, $m_{k1}, m_{k2}, \ldots\ldots$ and m_{kp} are the mid point of the intervals $u_{k1}, u_{k2}, \ldots\ldots$ and u_{kp} respectively, and the maximum membership values of $A_{k1}, A_{k2}, \ldots\ldots$ and A_{kp} occur at intervals $u_{k1}, u_{k2}, \ldots\ldots$ and u_{kp}, respectively.

Step 5: Use the high-order difference to determine the upward or downward trend.

The difference of the second order difference between any two neighboring time segments of the historical concentration can be used for forecasting the trend. The second order difference is calculated by the equation: $Y_n = Y_{n-1} - Y_{n-2}$.

The α-cut value determines the fuzzified concentration in the interval. It is quite usual to use the triangle function and chose the value of α-cut equal to 0.5 for estimation. Another important factor is the value of high-order difference; it will dominate the trend of concentration variation.

Step 6: Selecting the appropriate predictor by fractal analysis

There are many factors which may influence the concentration variation of time series. To improve the accuracy of prediction, it is better to incorporate these factors into a more advanced model. The properties of pollution, which are influenced by many factors, can be described in terms of an appropriate function and α-cut value. The triangular function, trapezoidal function, or Gaussian membership function are all possible choices.

The autocorrelation and cross correlation in the knowledge space phase of fractal dimension analysis help us find the best predictor. The autocorrelation coefficient between X_t and $X_{t-\tau}$ is calculated as follows

$$C_{auto}(X_t, X_{t-\tau}) = \frac{E(X_t X_{t-\tau} - E(X_t)^2)}{E\{[X_t - E(X_t)^2]\}} \tag{12}$$

Where, τ is the time lag of the two time segments X_t and $X_{t-\tau}$. The autocorrelation coefficient helps us define the fractal dimension of the environmental system.

The cross correlation help us know the relationship between two properties. It is calculated as

$$C_{cross}(X_t, Y_t) = \frac{E(X_t Y_t - E(X_t)^2)}{E\{[X_t - E(X_t)^2]\}} \tag{13}$$

The fractal dimension is determined by finding the maximum and minimum values of these two correlation coefficients in the time plot.

4 Statistical Pattern Recognition of Environmental Concentration in Space-Time Series

The space-time series could be described by the average values and fluctuating quantities such as U and $\overline{uv}$. It is also important to know how fluctuations are related to the adjacent fluctuations in time or space next to each other. The statistical pattern recognition helps us examine how fluctuations are distributed around an average value in the space-time series. Some statistical properties are introduced for the purpose of pattern recognition of environmental concentration time series such as probability density function and its Fourier transform, the autocorrelation and its Fourier transform, etc.

A steady time series is statistically stable, calculated by the mathematical function. The probability density function $B(\tilde{u})$ is defined by

$$B(\tilde{u})\Delta(\tilde{u}) = \lim_{T\to\infty} \frac{1}{T} \sum (\Delta t), \tag{14}$$

where $\tilde{u}$ denotes the fluctuation value.

$$B(\tilde{u}) \geq 0, \int_{-\infty}^{\infty} B(\tilde{u}) d\tilde{u} = 1. \tag{15}$$

The mean values of the various powers of $\tilde{u}$ are called moments. The means value is the first moment defined by

$$U = \int_{-\infty}^{\infty} \tilde{u} B(\tilde{u}) d\tilde{u}, \tag{16}$$

The variance, or the mean square departure, σ^2, from the mean value U is the second moment, which is defined by

$$\begin{aligned} \sigma^2 &= \overline{u^2} \\ &= \int_{-\infty}^{\infty} u^2 B(\tilde{u}) d\tilde{u}, \\ &= \int_{-\infty}^{\infty} u^2 B(u) du \end{aligned} \tag{17}$$

where σ is the standard deviation of root mean square(rms) amplitude.

The third moment, skewness (S), K, helps us discriminate the symmetric and anti symmetric parts of the time series. It is defined by

$$\overline{u^3} = \int_{-\infty}^{\infty} u^3 B(u) du \,, \tag{18}$$

and the value of skewness is

$$S = \frac{\overline{u^3}}{\sigma^3} \,. \tag{19}$$

The fourth moment, kurtosis or flatness factor K is represented as

$$\begin{aligned} K &= \frac{\overline{u^4}}{\overline{\overline{\sigma^4}}} \\ &= -\int_{-\infty}^{\infty} e^{-iku} B(u) du \end{aligned} \tag{20}$$

The other important measurement is the Fourier transform of B(u), which is defined as

$$\begin{aligned} \phi(K) &= \int_{-\infty}^{\infty} e^{-iku} B(u) du, B(u) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iku} \phi(K) dK \end{aligned} \tag{21}$$

The Fourier transform of B(u) is the characteristic function which is more convenient to see the pattern of flow property and behavior of pollution concentration.

In order to examine the feasibility of this model, the numerical experiment was conducted. The experimental data for time series analysis were acquired from the observed air quality data of Taiwan Environmental Protection Administration. Different pollutant concentration such as carbon monoxide, sulfur dioxide, nitrogen oxide, $PM_{2.5}$, and ozone were used. Data for analyzed were obtained from the year Sep, 2010, to AUG, 2011.

The results of the pattern of space-time series was analyzed by the autocorrelation to know the fractal dimension for different pollutants.

Figure 2 is the autocorrelation coefficient time series plot of ozone concentration time series. The amplitude of autocorrelation coefficient gradually decreases with time lag. There are totally five peaks in five days (120hours) of monitoring results, which means c(t) is correlated with itself every twenty four hours. The maximum value, which is near 1, is close to the first hours. This result reveals that the best predictor for ozone concentration is the time segment prior to the time of prediction.

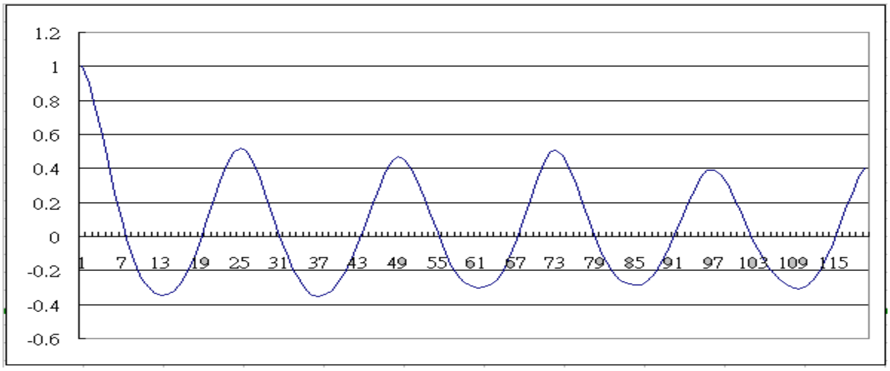

Fig. 2 Autocorrelation coefficient of ozone concentration time series

Figure 3 is the autocorrelation coefficient time series plot of nitrogen dioxide concentration. The results exhibit a wave-like trend of decreasing amplitude. The reason why the trend of nitrogen dioxide is the same as that of ozone is that they are all photochemical pollutants. Their formations are mostly governed by solar radiation. Therefore, all the graphical results show the same daily cycle.

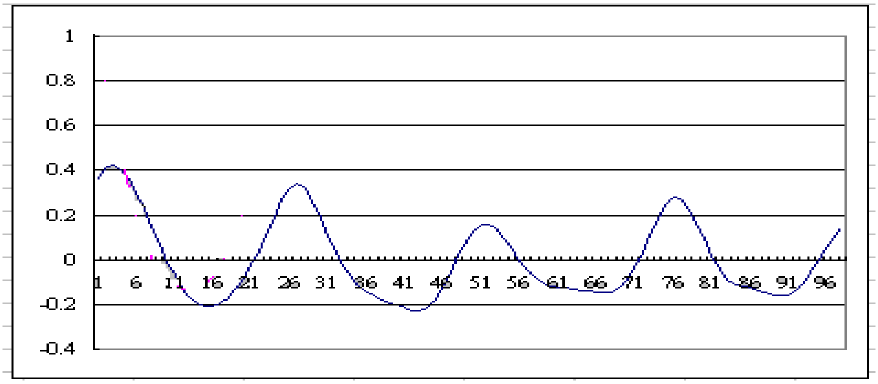

Fig. 3 Autocorrelation coefficients of nitrogen dioxide concentration time series

Figure 4 shows the plot of autocorrelation coefficient time series of carbon monoxide concentration. Two peaks, including a higher peak and a lower one, are found in one day. This represents two possible sources of carbon monoxide emission in the morning and in the evening. The time lag of twenty four hours has the highest value and the one of twelve hours has the second highest value. It is expected that the autocorrelation coefficient will approaches zero as the time lag approaches infinity.

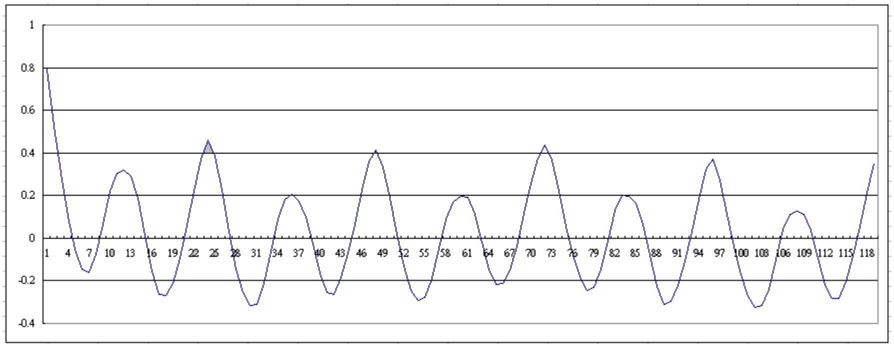

Fig. 4 Autocorrelation coefficients of carbon monoxide concentration time series

Figure 5 includes the autocorrelation coefficient time series plot of PM-10 concentration. The irregular trend fails to indicate significant correlation between u(t) and time lag. An assumption can be made that the correlation is significant only up to 8-hour time lag corresponding to the least coefficient of 0.2.

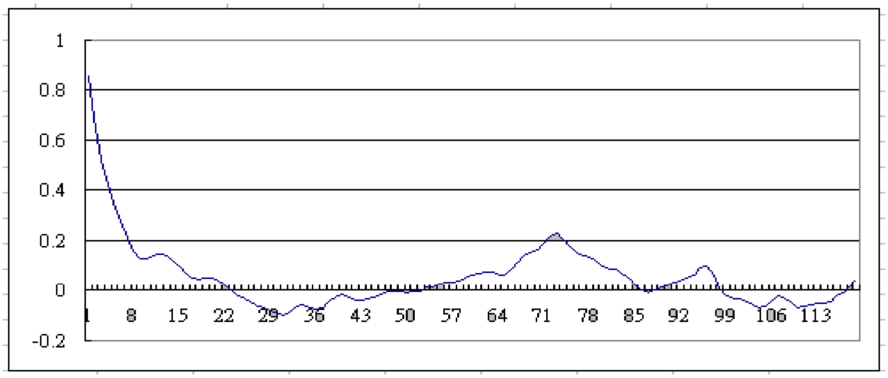

Fig. 5 Autocorrelation coefficients of PM-10 concentration time series

Figure 6 presents the autocorrelation coefficient of Non-methane hydrocarbon concentration. The figure reveals a very sharp decrease in the autocorrelation coefficient, which means that there are many factors which influence the variation of non-methane hydrocarbon concentration. It is more difficult to predict the concentration of non-methane hydrocarbon in time series.

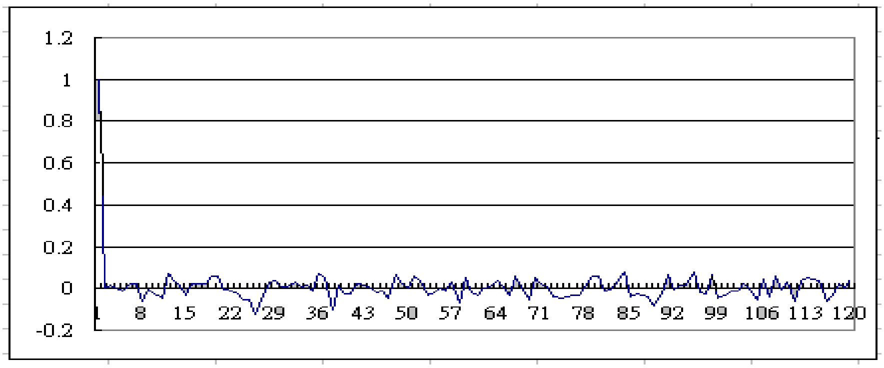

Fig. 6 Autocorrelation coefficients of Non-methane hydrocarbon concentration time series

5 Prediction of Environmental Pollution by Fuzzy Forecasting

Once understood space-time series pattern by turbulence theory and fractal dimension analysis, an attempt is made to apply fuzzy forecasting approach for prediction. The FTFA method, including a variety of schemes with different fuzzy intervals, multi-step fuzzy time series and the high-order fuzzy time series, is used to predict pollution concentration. These methods are used to determine the trend of data by adjusting the length of each interval in the universe of discourse.

The Mean Square Error was calculated as follows:

$$MSE = \frac{\sum_{1}^{m} (C_{obs} - C_{est})^2}{m} \tag{22}$$

where C_{obs} denotes the actual particulate concentration of time step I, C_{est} denotes the forecasting concentration, and m denotes the historical data.

The FTFA model is compared with other forecasting methods:

(1) The linear regression model: This model uses a linear trend over time to estimate the concentration:

$$C = a\,X + b \tag{23}$$

where C is the estimated concentration at specific time X.

(2) The autoregressive model: The autoregressive model uses the previous data to estimate the concentration as follows:

$$C = r_1\,C_{t-1} + \varepsilon \tag{24}$$

where C denotes the regression result at time t, C_{t-1} denotes the concentration at time t-1, r_1 denotes the regression coefficient , and ε denotes the predicting error.

The auto-regression model can be modified by the two time step estimation as follows:

$$C = r_1\,C_{t-1} + r_2\,C_{t-2} + \varepsilon \tag{25}$$

where r_1 and C_{t-1} are defined as before, C_{t-2} denotes the concentration at time t-2, and r_2 denotes the regression coefficient.

The concentration predicted by the above FTFA method is listed in Table 1and graphically shown in Fig 7. As shown in table 1, the mean square error from the FTFA model is least, exhibiting its superiority over other models in predicting concentration.

Table 1 Example of the observed concentration and fuzzy concentration C_{obs} : observed concentration; Int : concentration intervals C_{fuzzy} : fuzzified concentration

time	C_{obs}	Int	C_{fuzzy}
T_1	46.63	[46,47]	47
T_2	51.35	[51,52]	51
T_3	63.33	[63,64]	63
T4	58.77	[58,59]	59

The fuzzy logical relationship is listed in Table 2. For example, the following fifth-order fuzzy logical relationship: A_{17}, A_{17}, A_{16}, A_{16}, $A_{15}\rightarrow$A14, where the fuzzy logical relationship denotes the fuzzified concentration.

Table 2 Fuzzy logical relationship

Number of steps	fuzzy logical relationship
One step	$A_1\rightarrow A_{13}$
Two steps	$A_{17},A_{13}\rightarrow A_{20}$
Three steps	$A_{12},A_{15},A_{16}\rightarrow A_{19}$
Four steps	$A_{15},A_{15},A_{16}, A_{16}, \rightarrow A_{18}$
Five steps	$A_{17},A_{17},A_{16}, A_{16}, A_{15}, \rightarrow A_{14}$

Each interval is equally divided into four subintervals, where the points at 0.25 and 0.75 are used as bases to make forward or backward prediction. From the fuzzy logical relationship described above, the forecasted concentration can be determined. The results of several different prediction methods are shown in Table 3.

The value of mean square error for multi-steps fuzzy model (MSF) is the smallest among all the prediction methods, as shown in Table 3. It indicates that the proposed method is better than other models based on intervals with 10, 20 and 30 equal spaces. The prediction results of traditional auto-regressive model and linear egression model are both worse than those of the FTFA model. The reason is that the traditional statistical method and pattern recognition are either parametric or non-parametric models, but the high-order fuzzy time series recognize the pattern in other ways.

There are usually some data with unknown pattern in our forecasting procedure. The pattern recognition of concentration prediction is important in forecasting the particulate concentration, therefore more advanced tool has to be used.

An example is taken to compare the "patterned" and "un-patterned" data in this study. The method has the capability to catch the pattern of the concentration variation in the atmosphere. Also, the mean square error from the forecasted results of FTFA models was lower than that from the linear regression model and autoregressive model. Since the prediction of pollution concentration also involves hourly, daily, concentration, a more sophisticated analysis should take space and time into account in predicting variation of concentration.

Table 3 includes the result of comparison of different forecasting method.

Table 3 Comparison of the results by different prediction method Note: $\hat{C}$: actual concentration A :(10) equal interval fuzzy model B: (20) equal interval fuzzy model C: (30) equal interval fuzzy model D: multi steps fuzzy model E: simple linear model F: AR(1) model G: AR(2) model σ: standard deviation MSE: mean square error

statistical Property	$\hat{C}$	forecasted concentration						
		A	B	C	D	E	F	G
mean	7049	7109	7064	7055	7055	7100	7321	7230
σ	4022	4168	4055	4033	4233	8965	5655	4699
MSE		827	143	38	34	16323	3266	6803

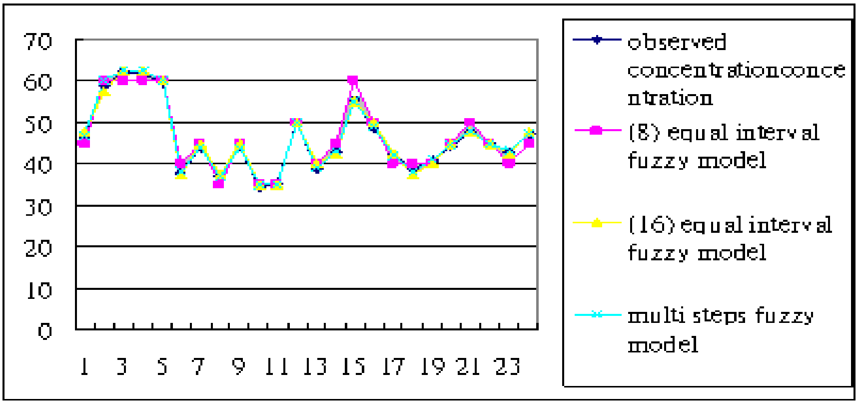

Fig. 7 Forecasted results with different interval

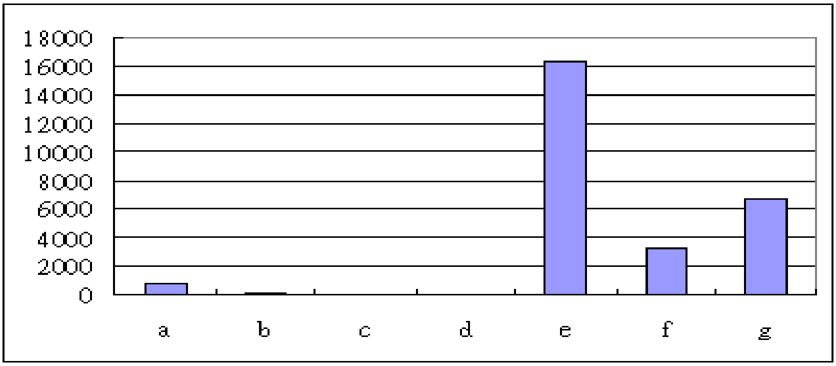

Fig. 8 Comparison of mean square error by different forecasting methods A : (10) equal interval fuzzy model B: (20) equal interval fuzzy model C: (30) equal interval fuzzy model D: multi steps fuzzy model E: simple linear model F: AR(1) model G: AR(2) model

6 Conclusions

In this paper, an attempt has been made to predict pollution concentration by two methods, i.e. multi-step fuzzy time series (MSFT) and different interval fuzzy time series (DIFT). The MFT method was implemented by adjusting the length of each interval in the universe of discourse and using the "second order difference" of concentration to predict the variation of concentration. The characteristics of pattern recognition of these two methods were discussed. The predicted results from those data with known pattern were better than those with unknown pattern.

By comparing the results, it is shown that the proposed MSFT method produces the smallest mean square error among the seven predicting methods. That is, these methods give higher accuracy than traditional fuzzy time series, linear regression model, and auto-regressive model. Accordingly, the proposed methods are obviously a better choice in predicting pollution concentration.

References

Box, G., Jenkins, G.: Time series analysis: forecasting and control, rev. ed. Holden-Day, Oakland (1976)

Hsu, C.C., Chen, S.M.: A New Mothod for Forcasting Enrollments Based on High-Order Fuzzy Time Series. In: Proceeding of the 2003 Joint Conference on AI Fuzzy System and Grey System, Taipei, Taiwan, Republic of China (2003)

Lee, C.K.: Multifractal characteristics in air pollutant concentration time series. Water Air Soil Poll. 135, 389–409 (2002)

Lee, C.K., Ho, D.S., Yu, C.C., Wang, C.-C.: Fractal analysis of temporal variation of air pollutant concentration by box counting. Environmental Modeling & Software 18(2003), 243–251 (2003)

Lee, C.K., Lin, S.C.: Chaos in Air Pollutant Concentration (APC) Time Series. Aerosol and Air Quality Research 8(4), 381–391 (2008)

Lee, C.K., Yu, C.C., Wang, C.C., Hwang, R.D., Yu, G.K.: Scaling characteristics in ozone concentration time series (OCTS). Chemosphere 62(2006), 934–946 (2006a)

Lee, C.K., Yu, C.C., Wang, C.C., Hwang, R.D., Yu, G.K.: Scaling characteristics in aftershock sequence of earthquake. Physica A 371(2006), 692–702 (2006b)

Olsson, J., Niemczynowicz, J.: On the possible use of fractal theory in rainfall applications. Wat. Sci. Tech (1994)

Olsson, J., Niemczynowicz, J.: Multifractal analysis of daily spatial rainfall distributions. J. Hydrol. 187, 29–43 (1996)

Huarng, J.R., Chen, S.M., Lee, C.H.: Handling Forcasting Problem Using Fuzzy Time Series. Fuzzy Sets and Systems 100(2), 217–229 (1998)

Huang, K.: Heuaistic Models of Fuzzy Time Series for Forcasting. Fuzzy Sets and System 123(3), 369–386 (2001)

Huang, K.: Effective Length of Intervals to Improve Forcasting in Fuzzy Time Series. Fuzzy Sets and Systems 123, 387–394 (2001)

Zadeh, L.A.: Fuzzy sets. Information and Control 8(3), 338–353 (1965)

Zadeh, L.A.: Fuzzy logic and approximate reasoning. Synthese 30, 407–428 (1975)

Mandelbrot, B.B.: The fractal Geometry of Nature. Freeman, New York (1982)

Song, Q., Chissom, B.S.: Fuzzy Time Series and Its Models. Fuzzy Sets and Systems 54(3), 269–277 (1993)

Song, Q., Chissom, B.S.: Forcasting Enrollments with Fuzzy Time Series – Part I. Fuzzy Sets and Systems 54(1), 1–9 (1993)

Song, Q., Chissom, B.S.: Forcasting Enrollments with Fuzzy Time series Part II. Fuzzy Sets and Systems 62(1), 1–8 (1994)

Chen, S.M.: Forcasting Enrollments Based on Fuzzy Time Series. Fuzzy Sets and Systems 81(3), 311–319 (1996)

Chen, S.M.: Forecasting Enrollments Based on High-Order Fuzzy time Series. Cybernetics and System, An International Journal 33(1), 1–16 (2002)

Chen, S.M., Hsu, C.C.: A New Approach for Handling Forcasting Problem Using High-Order Fuzzy Time series. Intelligent Automation and Soft Computing 14(1), 29–43 (2008)

Chen, S.M., Hwang, J.R.: Temperature Predicting Using Fuzzy Time Series. IEEE Transaction on System, Man, and Cybernetics- Part B: Cybernetics 30(2), 263–275 (2000)

Chen, S.M., Hsu, C.C.: A New Method to Forecast Enrollments Using Fuzzy Time Series. International Journal of Applied Science and Engineering 2(3), 234–244 (2004)

Tennekes, H., Lumley, J.L.: A First Course in Turbulence. MIT Press, Cambridge (1972)

Chen, W.K.: Environmental applications of granular computing and intelligent systems. In: Pedrycz, W., Chen, S.-M. (eds.) Granular Computing and Intelligent Systems- Design with Information Granules of High Order and Higher Type. Springer (2011)