5-2011

A Statistical model for long-term forecasting of strong sand dust storms

Siqi Tan
University of Nevada, Las Vegas

A STATISTICAL MODEL FOR LONG-TERM FORECASTING OF STRONG SAND DUST STORMS

by

Siqi Tan

Bachelor of Science
Capital Normal University, Beijing
2003

A thesis submitted in partial fulfillment
of the requirements for the

Master of Science in Mathematical Sciences
Department of Mathematical Sciences
College of Sciences

Graduate College
University of Nevada, Las Vegas
May 2011

THE GRADUATE COLLEGE

We recommend the thesis prepared under our supervision by

Siqi Tan

entitled

A Statistical Model for Long-Term Forecasting of Strong Sand Dust Storms

be accepted in partial fulfillment of the requirements for the degree of

Master of Science in Mathematical Sciences

Chih-Hsiang Ho, Committee Chair

Amei Amei, Committee Member

Kaushik Ghosh, Committee Member

Sheniz Moonie, Graduate Faculty Representative

Ronald Smith, Ph. D., Vice President for Research and Graduate Studies
and Dean of the Graduate College

May 2011

ABSTRACT

A Statistical Model for Long-Term Forecasting of Strong Sand Dust Storms

by

Siqi Tan

Dr. Chih-Hsiang Ho, Examination Committee Chair
Professor of Mathematical Sciences
University of Nevada, Las Vegas

Dust elevated into the atmosphere by dust storms has numerous environmental consequences. These include contributing to climate change; modifying local weather conditions; producing chemical and biological changes in the oceans; and affecting soil formation, surface water, groundwater quality, crop growth, and survival (Goudie and Middleton, 1992). Societal impacts include disruptions to air, road and rail traffic; interruption of radio services; the myriad effects of static-electricity generation; property damage; and health effects on humans and animals (Warner, 2004).

In this thesis, we extend the idea of empirical recurrence rate (ERR), developed by Ho (2008), to model the temporal trend of the sand-dust storms in northern China. Specifically, we show that the ERR time series has the following characteristics: (1) it is a potent surrogate for a point process; (2) it is created to take advantage of the well-developed and powerful time series modeling tools; and (3) it can produce reliable forecasts, capable of retrieving the corresponding mean numbers of strong sand-dust storms.

ACKNOWLEDGEMENTS

A few words of thanks seem inadequate to express my appreciation for those who have helped me over these three years. First of all, I would like to say thank you to my advisor, Dr. Chih-Hsiang Ho, his encouragement and guidance are invaluable, and his excellence in both research and teaching will always be a great example to me.

I am indebted to many individuals who have directly and indirectly supported me in this work, including these respectful committee members: Dr. Amei Amei, Dr. Kaushik Ghosh, Dr. Sheniz Moonie, for their their positive inputs and mentoring during my graduate studies.

Last but not the least, I am eternally grateful to my family for their love, support and understanding, which gave me the strength to finish.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

A dust storm is said to occur when the horizontal visibility is less than 1000m, and when the dust is being circulated into the atmosphere within sight of the observer. In spite of this international standard, some researchers develop their own definition with respect to different areas. Tao *et al* (2002) give the following criteria particularly used in Inner Mongolia, China:

Dust storm – at least three stations reporting with horizontal visibility of less than 1000m and an average wind speed of 10.8 to 20.7 m/s;

Strong dust storm – at least three stations reporting with horizontal visibility of less than 500m and an average wind speed of 17.2 to 24.4 m/s;

Very strong dust storm – at least one station reporting with horizontal visibility of less than 50m and an average wind speed of 20.8 m/s or greater.

Dust storms can cause numerous environmental consequences. These include contributing to climate change; modifying local weather conditions; producing chemical and biological changes in the oceans; and affecting soil formation, surface water, groundwater quality, crop growth and survival. Societal impacts include disruptions to air, land and rail traffic; interruption of radio services; the myriad effects of static-electricity generation; property damage; and health effects on humans and animals. Although commonly viewed as an ecological evil, dust storm has a positive effect of neutralizing acid rain. Chinese scientists discovered that sand and dust rich in calcium carbonate with a pH indicator between seven and eight can increase the acid-base indicator of rainfall in northern China, which can effectively alleviate the harmful effects of acid rain.

Climate is generally regarded to be an important factor influencing the occurrences of dust storms. This indicates factors such as wind, relative humidity, air temperature, precipitation and dryness index. This thesis focuses on the northern China area from 1954 to 2002 and the sand storms that mainly originated from the following regions: Hexi Corridor of Gansu Province and Alxa Plateau, southern rim of South Xinjiang Basin, and central Inner Mongolia. The features of sand storms' frequency variations during the past 50 years are as follows: a fluctuating increase during 1960s-1970s and a fluctuating decrease during 1980s-1990s. After 2000, activities of the sun began a new round of weak trend, which weakened the warm trend of climate. Consequently, the intensity of the surface heat-field in the Tibetan plateau was weak and the air temperature of the northern Xinjiang, Hexi corridor and Ningxia region was abnormally low. All these changes have made sand storms enter a new active period in northern China. (Thomas T. Warner, 2004; Zhou and Zhang, 2003; Yang et al, 2007; Zhang et al, 2002)

By developing an empirical recurrent rate (ERR) time series, this thesis presents a new treatment to smooth the point process. The ERR is computed sequentially and cumulatively at equidistant time intervals during the observation period. Once we establish the ERRs, we explore the possibility of using the linear stochastic model ARIMA model to develop reliable and robust forecasts, appropriately designed simulations could help us to have a general idea about the real data, and give some hints for finding the final model.

To sum up, definition of ERR, ARIMA model and relevant time series theories and method are introduced in Chapter 2, we perform the simulation in Chapter 3. Chapter 4 uses the sand storm data to build the model and discuss the sensitivity of deleting the burn in period. Chapter 5 will be the conclusion of this study.

CHAPTER 2

THEORIES AND METHODS

2.1 Empirical Recurrence Rates

Let $t_1, \ldots t_n$ be the times of occurrences of n sand storms during an observation period $(t_0, 0)$, where 0 = present time. Then we can generate a series $\{z_l\}$ based upon the counting data sequentially at equidistant time intervals $t_o + h,\ t_0 + 2h,\ \ldots,\ t_0 + lh,\ \ldots,\ t_0 + Nh$ (= 0 = present time). If t_0 is viewed as the time-origin and h as the time-step, then $\{z_l\}$ can be regarded as the observation at time $t = t_0 + lh$, for the sand storms to be modeled. Therefore, a time series of the empirical recurrence rates (ERR) is proposed and is defined as follows:

$$z_l = n_l / lh \text{ = total number of sand storms in } (t_0, t_0 + lh)/lh,$$

where n_l is the cumulative number of sand storms, l =1, 2, …, *N*. Note that z_l evolves over time and it is simply the MLE of the mean, if the underlying process observed in $(t_0,\ t_0 + lh)$ is a homogeneous Poisson process. If we start at time T, the value z_{T+k}, $k \geq 1$ is needed to be predicted based on the sample observation $(z_1, \ldots, z_T)$ of an ERR time series. In a regression situation, let X denote the time index, z the response values, and then use the fitted regression model to obtain z_{T+k}. However, a regression model assumes that the observations are independent and this is not a reasonable assumption for a process that evolves over time. Thus the ARIMA model is introduced.

2.2 ARIMA Models

Classical regression is often insufficient for explaining all of the interesting dynamics of a time series. It is developed for the static case. Namely, the regression only allows the dependent variable to be influenced by current values of the independent variables. Besides, the regression may not capture the additional structure such as presented in a random walk process.

The acronym ARIMA, stands for autoregressive integrated moving average. The original key reference is from Box and Jenkins (1970). The basic processes of the Box–Jenkins ARIMA (*p*,*d*,*q*) model may consists of the following: the autoregressive process, the integrated process, and the moving average process. The autoregressive model is analogous to the regression model, based on the idea that the current value of the series X_t . The current value is a linear combination of the *p* most recent past values of itself plus an “innovation” term W_t that incorporates everything new in the series at time *t* that is not explained by the past values. An autoregressive model of order *p*, is of the form:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + W_t$$

where $t = 1, 2, ..., N$, X_t is mean-zero stationary, $\phi_1, \dots, \phi_p$ are called the autoregressive coefficients for an pth order process, W_t is Gaussian white noise series with mean zero and variance σ^2, independent of X_{t-1}, $X_{t-2}, \dots$, X_{t-p} for every t.

A moving average (MA) process of order *q* is a linear combination of the current white noise term and the q most recent past white noise terms Z_t and is defined by

$$X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$$

where $t = 1, 2, \ldots, N$, X_t is mean-zero stationary time series, Z_t is Gaussian white noise with mean zero and variance σ^2. $\theta_1, \ldots, \theta_q$ are called the MA parameters of the model.

A general autoregressive moving average (ARMA) model, ARMA (p,q), is given by:

$$X_t - \phi_1 X_1 - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q} .$$

A time series X_t is said to follow an integrated autoregressive moving average model (ARIMA) if the dth difference $Y_t = \nabla^d X_t$ is a stationary ARMA process. If Y_t follows an ARMA (p,q) model, we say that X_t is an ARIMA (p,d,q) process. In constructing ARIMA model we go through 3 stages: identification, estimation, and diagnostic checking. In the identification stage, preliminary estimates for q, p and d are obtained using the plots of the sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF). Sometimes identification is done by an auto fit procedure – fitting many different possible model structures and orders and using a goodness-of-fit statistic to select the best model. The second stage is to estimate the coefficients of the model. In this step, we adopt the maximum likelihood estimation method. The last stage is model diagnostic checking. In the ARIMA modeling, it is important to perform diagnostic checking on the residuals of the fitted model. This usually consists of a group of tests including tests for normality using the residuals.

Moreover, it is necessary to test that all the model parameters are statistically significant. The fitting process is often guided by the principle of parsimony, by which the best model is one that has the fewest parameters among all models that fit the data. (Cryer and Chan, 2008; Box and Jenkins, 1976; Shumway and Stoffer, 2005)

2.3 Data Pretreatment

2.3.1 Data Splitting

Cross-validation is the statistical practice of partitioning a sample of data into subsets so that the analysis is initially performed on a single subset, while the other subset are retained for subsequent use in confirming and validating the initial analysis. For a large enough data set, it can be partitioned into two sets: training sample used to develop a model and prediction set used to evaluate the reasonableness and predictive ability of the model.

2.3.2 Data Transformation

The one important condition for ARMA (ARIMA) model is obtaining a stationary time series (mean=0), which needs appropriate transformations for different types of data.

(a) Box-Cox Transformation

For a given value of the parameter λ, and positive observations $Y_1, Y_2, Y_3, ...Y_n$, the transformation is defined:

$$f_\lambda(y) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(y), & \lambda = 0 \end{cases}$$

The power transformations are useful when the variability of the data increases or decreases with the level. By suitable choice of λ, the variability can often be made nearly constant, which is a requirement for stationarity. For example, $\lambda = ½$ produces a square root transformation useful for Poisson-like data, and $\lambda = -1$ corresponds to a reciprocal transformation.

(b) Differencing

Differencing a time series can remove trends, whether these trends are stochastic, as in a random walk, or deterministic, as in the case of a linear trend. By subtracting each data point in a series from its predecessor, the first order difference is defined:

$$\nabla X_t = X_t - X_{t-1} = (1-B)X_t$$

where B is the backward shift operator. A series Yt is said to be integrated of order d if:

$$\text{Yt} = \nabla^d X_t = (1-B)^d X_t$$

By introducing the lag-d differencing operator ∇_d, we can eliminate seasonality and trend of period d:

$$\nabla_d X_t = X_t - X_{t-d} = (1-B^d)X_t$$

For example, differencing at lag 12 will remove the seasonal effect in a monthly time series.

If the data suggest nonstationarity, then it is necessary to perform a power transformation or differencing to produce a new series that is more compatible with the assumption of stationarity. Appropriate numbers of differencing will generate a series with rapidly decaying sample ACF, and then the differenced data set can be fitted by a low-order ARMA process. This means that the fitted parameters will be well away from the boundary of the allowable parameter set. Therefore, after every differencing, we check the plots of the sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) to see where the ACF/PACF "cuts off" at the bounds $\pm 1.96/\sqrt{n}$. If the sample ACF has very few significant spikes at very small lags and cuts off drastically or dies down very quickly, we get a stationary series. If the sample ACF dies slowly, we should do further differencing. (Brockwell and Davis., 2002).

(c) Subtracting the Mean

The term ARMA model is used in the program ITSM2000 (Brockwell and Davis., 2002) to denote a zero-mean ARMA process. Therefore, the sample mean of the data should be small before modeling. Once the apparent deviations from stationarity of the data have been removed, we subtract the sample mean of the transformed data from each observation. The search for a fitted ARMA model for a transformed mean-corrected data set then follows.

2.4 Model Diagnostics

2.4.1 The Sample ACF of the Residuals

For large n, the sample autocorrelations of an independent and identically distributed (iid) sequence $Y_1,\ldots,\ Y_n$ with finite variance are approximately iid with distribution N(0, 1/n). We can therefore test whether or not the observed residuals are consistent with iid noise by examining the sample autocorrelations of the residuals and rejecting the iid noise hypothesis if more than two or three out of 40 fall outside the bounds $\pm 1.96/\sqrt{n}$ or if one falls far outside the bounds (Brockwell and Davis., 2002).

2.4.2 Tests for Randomness of the Residuals

In addition to looking at residual correlations at individual lags, it is useful to have a test that takes into account their magnitudes as a group. Ljung and Box (1978) proposed the statistic used to test the overall independence based on a few of lags. The definition of Ljung-Box test is as follows.

H_0 : The sequence data are iid

H_a : The sequence data are not iid

The test statistic is: $\hat{Q}(\hat{r}) = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}\hat{r}_k^2$,

where $\hat{r}_k = \sum_{l=k+1}^{n}\hat{a}_l\hat{a}_{l-k} \Big/ \sum_{l=1}^{n}\hat{a}_l^2$, the estimated autocorrelation at lag k ,

$n =$ sample size,

m = number of lags being tested,

and $\hat{a}_1,\ldots,\hat{a}_n$ are the residuals after a model has been fitted to a series $z_1,\ldots,z_n$. If no model is being fitted, then $\hat{a}_1,\ldots,\hat{a}_n$ are the "mean corrected" vectors of $z_1,\ldots,z_n$.

The chi-square distribution for $\hat{Q}(\hat{r})$ is based on a limit theorem as $n \to \infty$, in other words, the statistic $\hat{Q}(\hat{r})$ has a finite sample distribution that is close to χ^2_{m-p-q} for large n, if the correct ARMA(*p*,*q*) model is estimated. Thus, a general "portmanteau" test would reject the ARMA(*p*,*q*) model if the observed value of $\hat{Q}(\hat{r})$ exceeded an appropriate critical value in a chi-square distribution with $m - p - q$ degrees of freedom at level α . (Brockwell and Davis, 2002).

2.4.3 AIC, BIC and AICC Statistics

Many time series models are introduced along with the respective diagnostic checking procedures. Through the utilization of diagnostic checking methods, it is hoped that the researcher should be able to grasp the relative merits of these models, hence, answering the question "Which model describes the data best?" Thus, the model diagnostic checking is often used together with model selection criteria such as the Akaike information criterion (AIC), or the bias-corrected version of the AIC statistic (AICC) and the Bayesian information criterion (BIC). Akaike first introduced AIC statistic in 1974, and the BIC statistic was proposed by Schwarz in 1978. These two approaches actually complement each other. Each information statistic is defined as follows:

$$AIC_{p,q} = N \log \hat{\sigma}^2_\varepsilon + 2r$$

$$AICC_{p,q} = N \log \hat{\sigma}^2_\varepsilon + 2rN/(N - r - 1)$$

$$BIC_{p,q} = N \log \hat{\sigma}^2_\varepsilon + r \log N$$

where $\hat{\sigma}^2_\varepsilon$ is the maximum likelihood estimator of σ^2_ε, and $r = p + q + 1$ is the number of parameters estimated in the model, including a constant term. The second term in all three equations is a penalty for increasing r. Hence, if we want to minimize the values of these criteria, we should minimize the number of parameters. Therefore, the best model is the model adequately describes data and has fewest parameters. (Li, 2003)

2.5 Forecasting

One of the primary objectives of building a model for a time series is to be able to forecast the values for that series at future times. The forecasting function $z_t = f(z_{t-1}, \ldots, z_1) + a_t$ has the minimum mean square error. The first part of the above equation $f(z_{t-1,} \ldots, z_1)$ is a function of the past values of the series and it should be determined by the data. The second part a_t, called noise part, is a sequence of independent and identically distributed (iid) variables. Predictions will be achieved by forecasting the residuals and then inverting the transformations adopted to arrive at forecasts of the original series. Also, we will see which model is the best fitting model by comparing the forecasted values with the prediction set. Then, we will combine the training sample and the prediction set as a full data set to forecast sand storms for the future based on the same techniques as before. Note that the cumulated mean numbers inverted from the forecasted ERRs should be non- decreasing, and should sometimes be adjusted accordingly. (Ho, 2008.)

2.6 Subset Model Checking

In the ITSM2000 package, the coefficients of models are given with the ratio of each estimate to 1.96 times its standard error, if it is a causal model (P85, Brockwell et al., 2002). The denominator (1.96×standard error) is the critical value (at level 0.05) for the coefficient. Thus, if the ratio is less than 1 in absolute value, we may conclude (at level 0.05) that the corresponding coefficient in the model may be zero. After dropping the non-significant coefficients, a subset model comes up, which requires additional model selection process.

CHAPTER 3

SIMULATION

The number of strong sand storms that occurred in the northern China during March 1954 to April 2002, are obtained from the paper published by Zhou and Zhang (2003). In this time period, there were 908 sand storms occurred in 578 months (1954 March—2002 April). By graphing the raw data (Figure 3.1), we see that the time series has a lot of zeroes, and clearly it is not a Poisson process. The ordinary ARIMA modeling techniques can not handle the series with many zeros, since the stationarity may be difficult to achieve. Therefore, proper smoothing should be implemented. Before we perform the real data analysis, it is always good to explore the applicability of the proposed technique based on the simulated data. Here we adopt the following simulation method: We randomly select one year from the raw data, repeat it 17 times, and treat it as a whole data set of 17 years. Note that due to seasonality, most of the sand storms occurred during the months of March, April and May with no storms for the rest of the year. Our selected year, 1996, has 2 sand storms in April and 14 in May. After converting it to an ERR time series, we use the technique described in Section 2.3.1 to split the data into two sets: training sample and prediction set. In this case, our training sample is the whole data set excluding the last 2 years (24 lags) which is the prediction set (Figure 3.2). Our modeling approach based on the simulated data will be addressed in detail below.

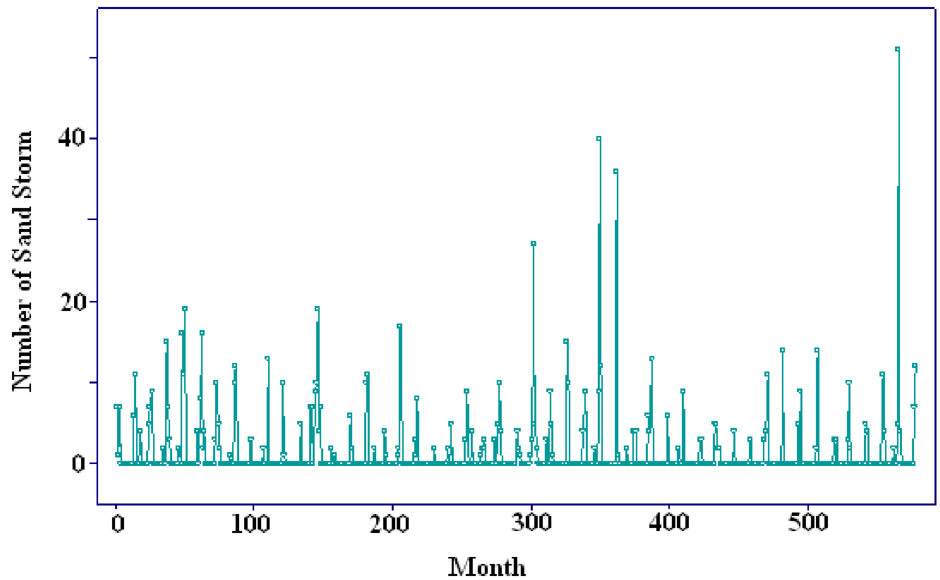

Figure 3.1 Plot of strong sand storms in northern China between March 1954 and April 2002.

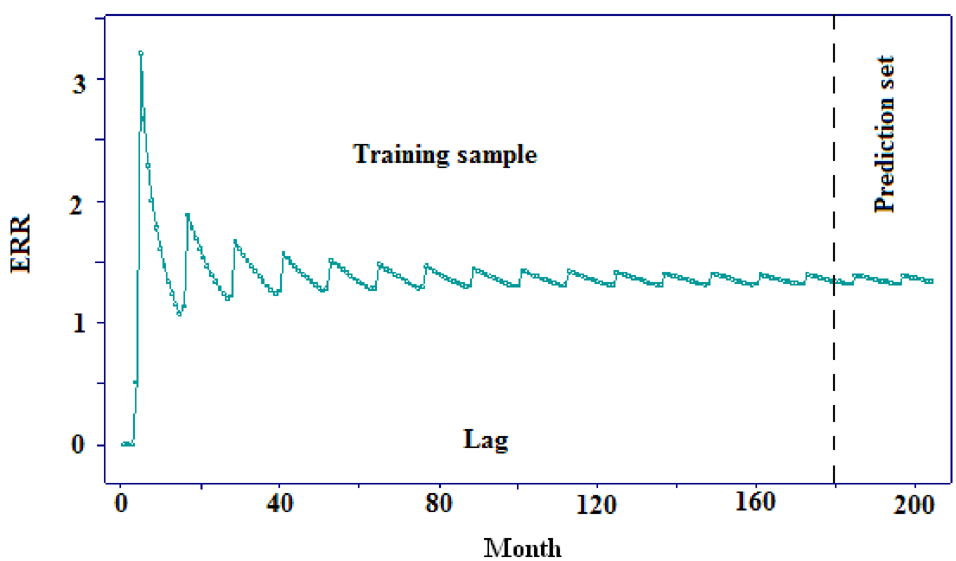

Figure 3.2 Training sample and prediction set of data set with h = 1 month.

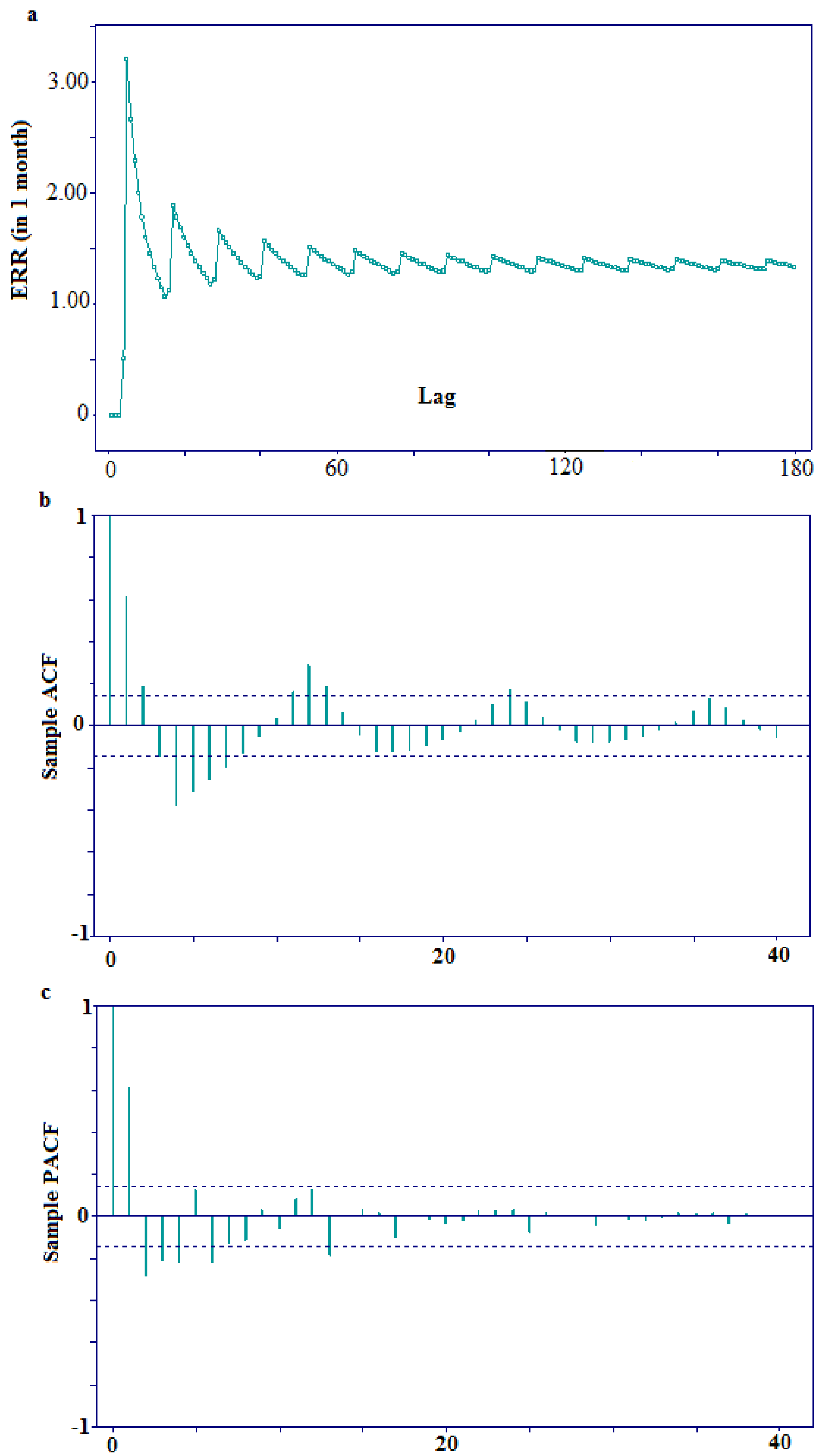

Figure 3.3 a, ERR plot of the training sample; b, Sample ACF; c, Sample PACF

To model the ERR with h = 1 month, we will use the software ITSM2000 to analyze the data. The plot of the training sample (Figure 3.3a), and its sample ACF and PACF (Figure 3.3b,c) show nonstationarity and periodicity. Therefore, the data pretreatment, Box-Cox transformation and differencing, described in Section 2.3 will be explored to remove the trend and the seasonality. First, we use the Box-Cox transformation to stabilize the variance before differencing. A combination of λ=1.5 for the Box-Cox transformation (Figure 3.4) and a lag 12 differencing indicate some improvements in achieving stationarity (Figure 3.5). Furthermore, the remaining linear trend has been removed by taking additional difference at lag 1. The ACF/PACF plots (Figure 3.6) indicate that the stationarity is by and large attained. Thus, we subtract the sample mean of the transformed data from each observation to generate a series to which we then fit a zero-mean stationary model. An MA (1) is considered based on our initial model search. The estimated (MLE) model is:

$X_t = Z_t - 0.1456\, Z_{t-1}$.
Estimated WN Variance = 0.048960
Standard Error of MA Coefficient = 0.088909

Note that X_t represents a twice-differenced stationary mean-corrected time series and the error term Z_t represents a white noise process.

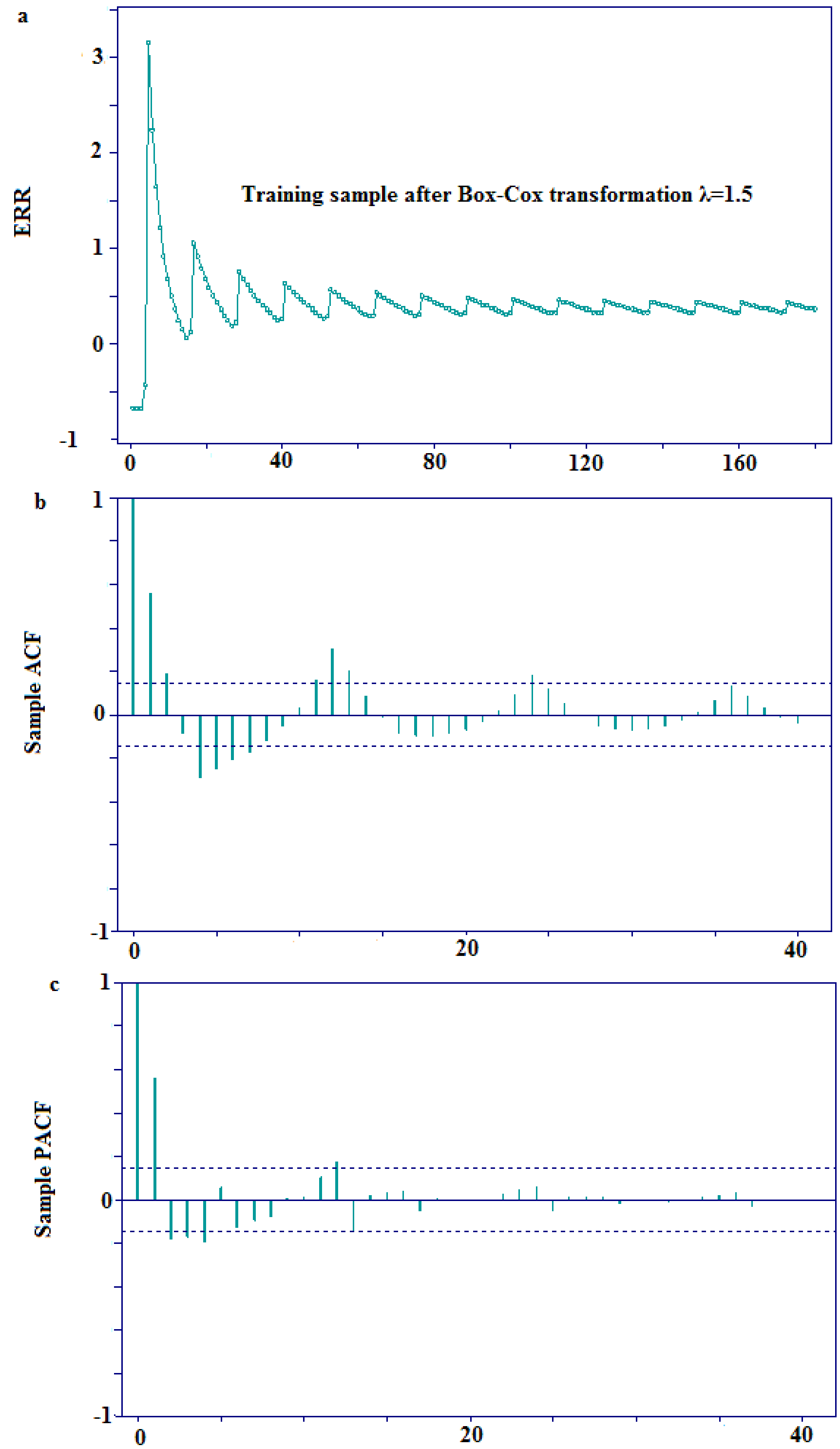

Figure 3.4 a, Time plot after Box-Cox transformation at λ=1.5; b, Sample ACF; c, Sample PACF.

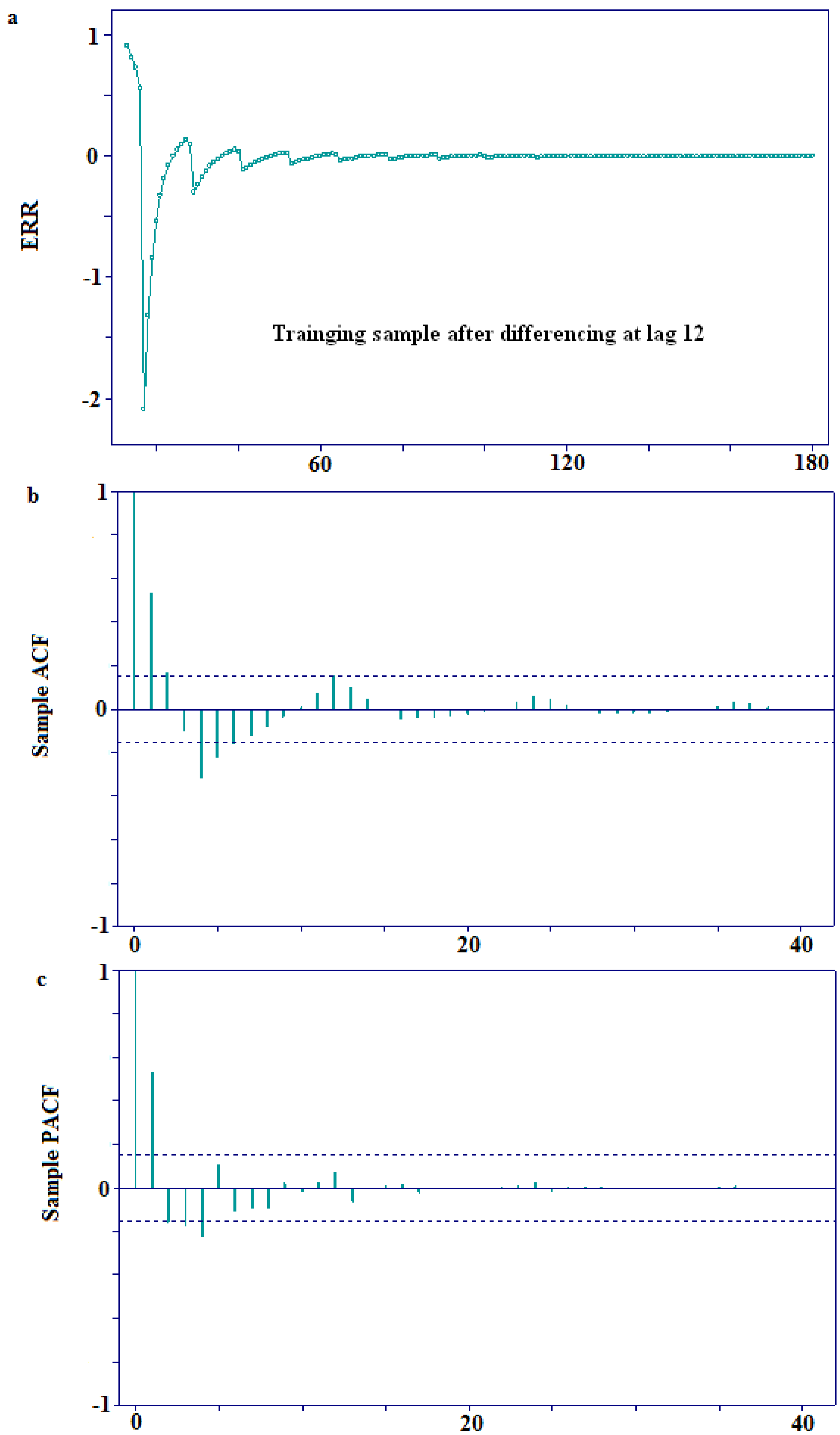

Figure 3.5 a, Time plot of after differencing at lag 12; b, Sample ACF; c, Sample PACF.

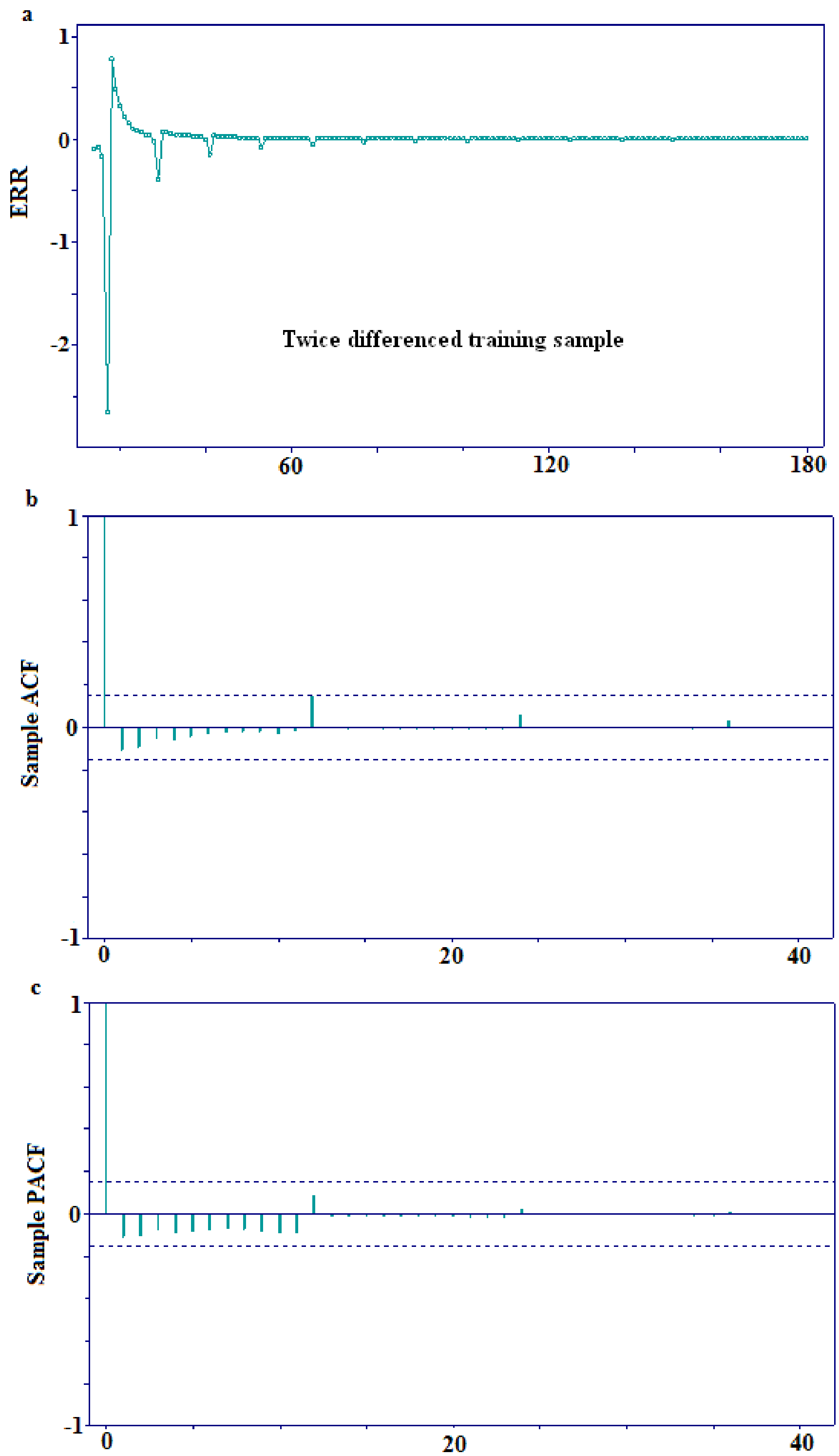

Figure 3 6 a, Time plot of after further differencing at lag 1; b, Sample ACF; c, Sample PACF.

The AICC statistic is -25.778570, the plots of sample ACF/PACF of the residuals are shown in Figure 3.7. Also, the Ljung-Box test is not significant (p-value =0.94258), indicating that the residuals are approximately white noise. Some evidence of the validity of the fitted model can be obtained through an examination of the actual predictive capability of the selected model. Thus, for the purpose of model validation, we produce Figures 3.8 and 3.9 to compare the 24 forecasts with the prediction set. By visual inspection of Figure 3.9, we conclude that selected model is not seriously biased and gives an appropriate indication of the predictive ability of the model. We also present a more rigorous model validation based on a set of the estimated mean numbers (Table 3.1). Note that several mean numbers are invalid and need adjustment because the mean function should be non-decreasing.

In summary, it predicts a mean number of 14 sand storms for the month of April for both years, which is close to the actual number of events. Hence, the model validation results are successful and the proposed modeling technique can be repeated using the real data set, to be addressed in the next chapter.

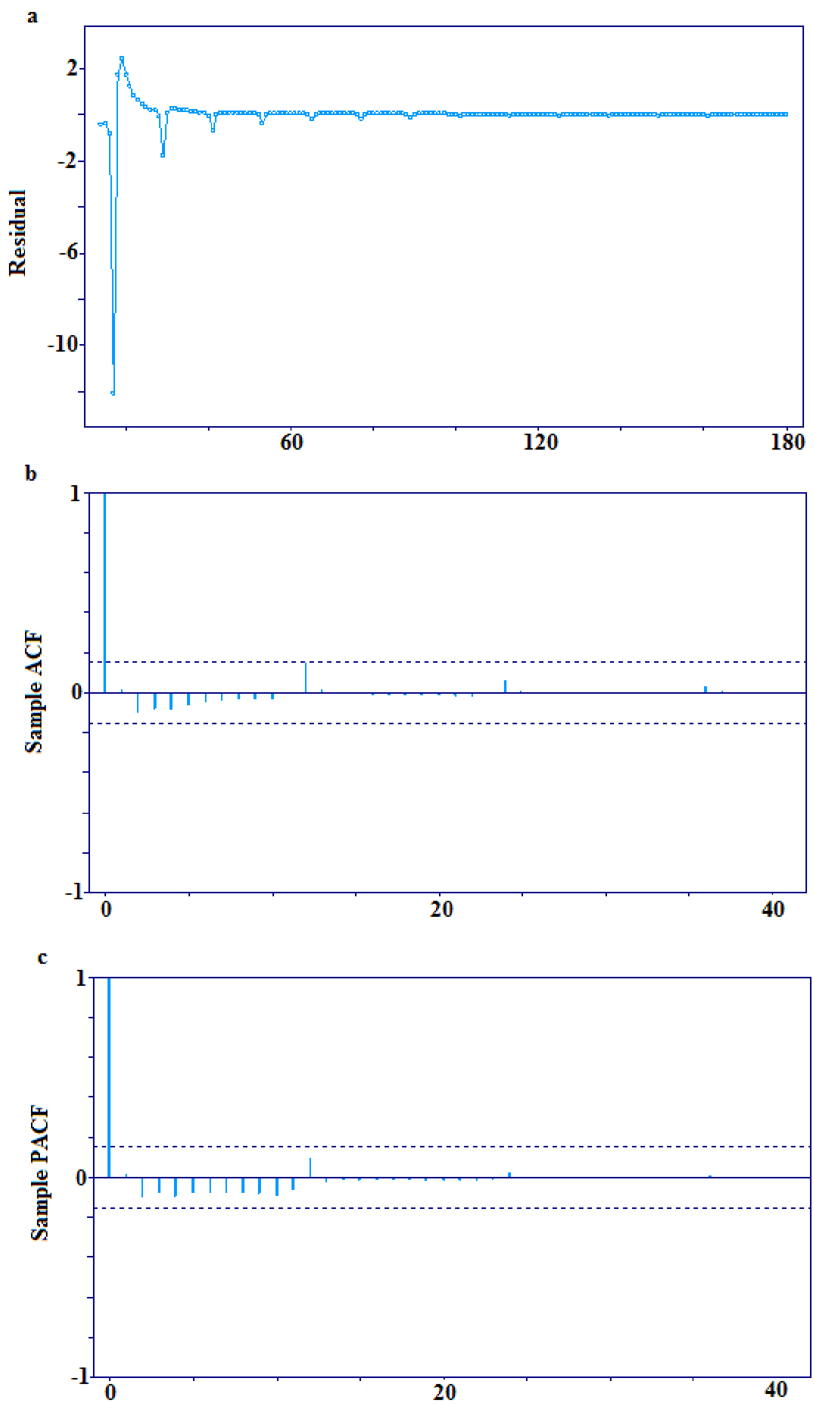

Figure 3.7 Diagnostics for the MA (1). a, Residual plot; b, Residual ACF; c, Residual PACF.

Table 3.1. The numerical values of the actual ERRs and mean numbers in the prediction set, and the predicted ERRs using the MA (1) with their counterparts (the corresponding mean values derived from the predicted ERRs)

Predicted month	Monthly ERR		Mean number	
	Actual	Prediction	Actual	Prediction
1	1.325967	1.31980	0	-1.1162 adjust to 0
2	1.318681	1.30723	0	-0.96794 adjust to 0
3	1.311475	1.29471	0	-0.97141 adjust to 0
4	1.315217	1.29395	2	1.15639
5	1.383784	1.36316	14	13.89017
6	1.376344	1.35043	0	-0.9537 adjust to 0
7	1.368984	1.33776	0	-0.95551 adjust to 0
8	1.361702	1.32513	0	-0.9609 adjust to 0
9	1.354497	1.31257	0	-0.96079 adjust to 0
10	1.347368	1.30005	0	-0.96607 adjust to 0
11	1.340314	1.28758	0	-0.96949 adjust to 0
12	1.333333	1.27515	0	-0.97468 adjust to 0
13	1.326425	1.25646	0	-2.12643 adjust to 0
14	1.319588	1.23867	0	-1.98132 adjust to 0
15	1.312821	1.22088	0	-1.99911 adjust to 0
16	1.316327	1.21516	2	0.17984
17	1.380711	1.28168	14	13.3218
18	1.373737	1.26370	0	-1.99068 adjust to 0
19	1.366834	1.24572	0	-2.00866 adjust to 0
20	1.36	1.22772	0	-2.03028 adjust to 0
21	1.353234	1.20971	0	-2.0501 adjust to 0
22	1.346535	1.19168	0	-2.07175 adjust to 0
23	1.339901	1.17362	0	-2.09524 adjust to 0
24	1.333333	1.15554	0	-2.11694 adjust to 0

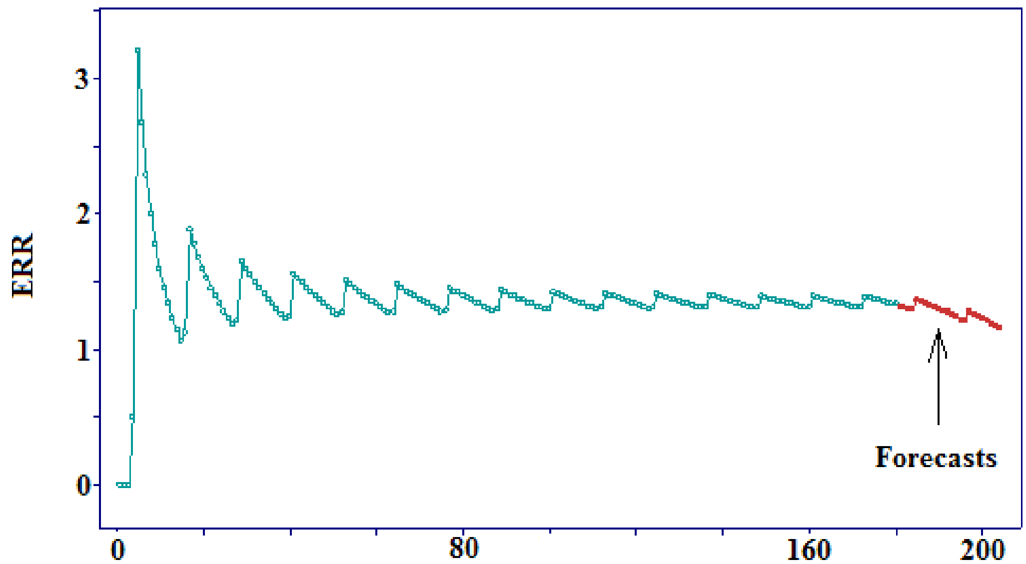

Figure 3.8 The predicted ERR of the simulation the year 1996 for 24 months.

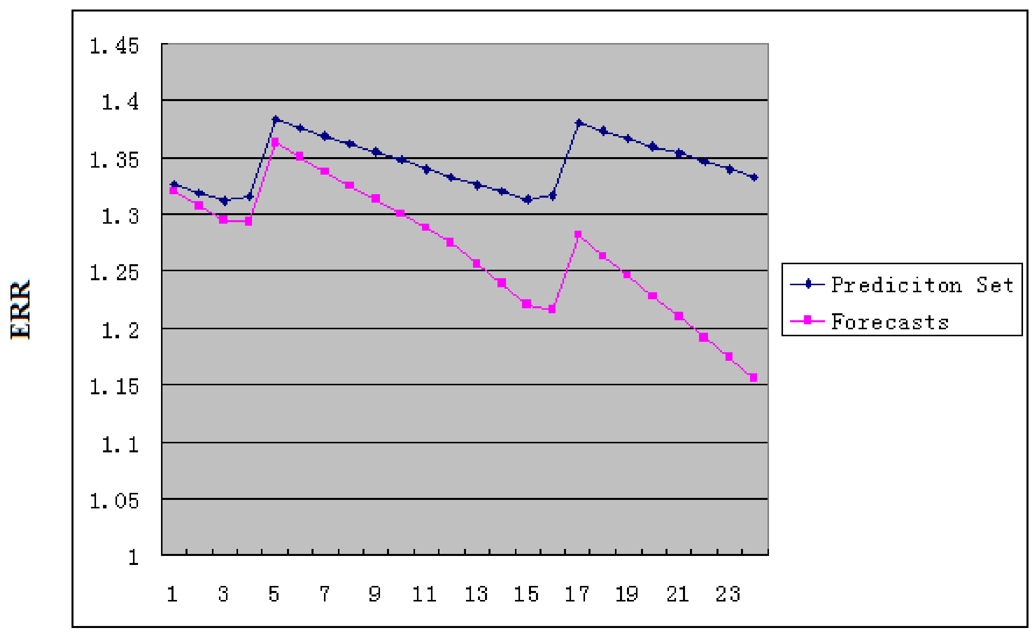

Figure 3.9 Plot of the ERR comparison of the prediction set and forecasts.

CHAPTER 4

APPLICATION

4.1 ARIMA modeling without ERR conversion

By graphing the raw data (Figure 3.1), we see that the time series has a lot of zeroes. Before introducing the idea of the ERR, we first try to model the raw data using the ARIMA technique. After taking the Box-Cox transformation and a differencing at lag l2 and 1 (Figure 4.1), we see the routine transformation methods hardly make a stationary time series for the sand storm data. Therefore, we fail to find a model by using the ARIMA techniques.

4.2 ARIMA modeling using non-cumulative ERR

Next, by adopting the ERR method, we calculate the ERRs for each of the 55 years respectively and put them together (Figure 4.2a). This time series has less number of zeros than the raw data. Because the ERRs are calculated independently for each year, we can see that there exists a huge fluctuation in the series. The stationarity is still hard to achieve by the common transformations (Figure 4.3b, c).

4.3 ARIMA modeling using ERR

We apply the idea of ERR to smooth the whole raw data, which is described in Chapter 2. First, we see a large variation at the beginning of the series from Figure 4.3. Since the initial search for the whole ERR time series was not successful, we then trim the data set from the beginning of the data, and it appears if we drop the raw data about 13 months and recalculate the ERRs, the modeling process becomes much simpler.

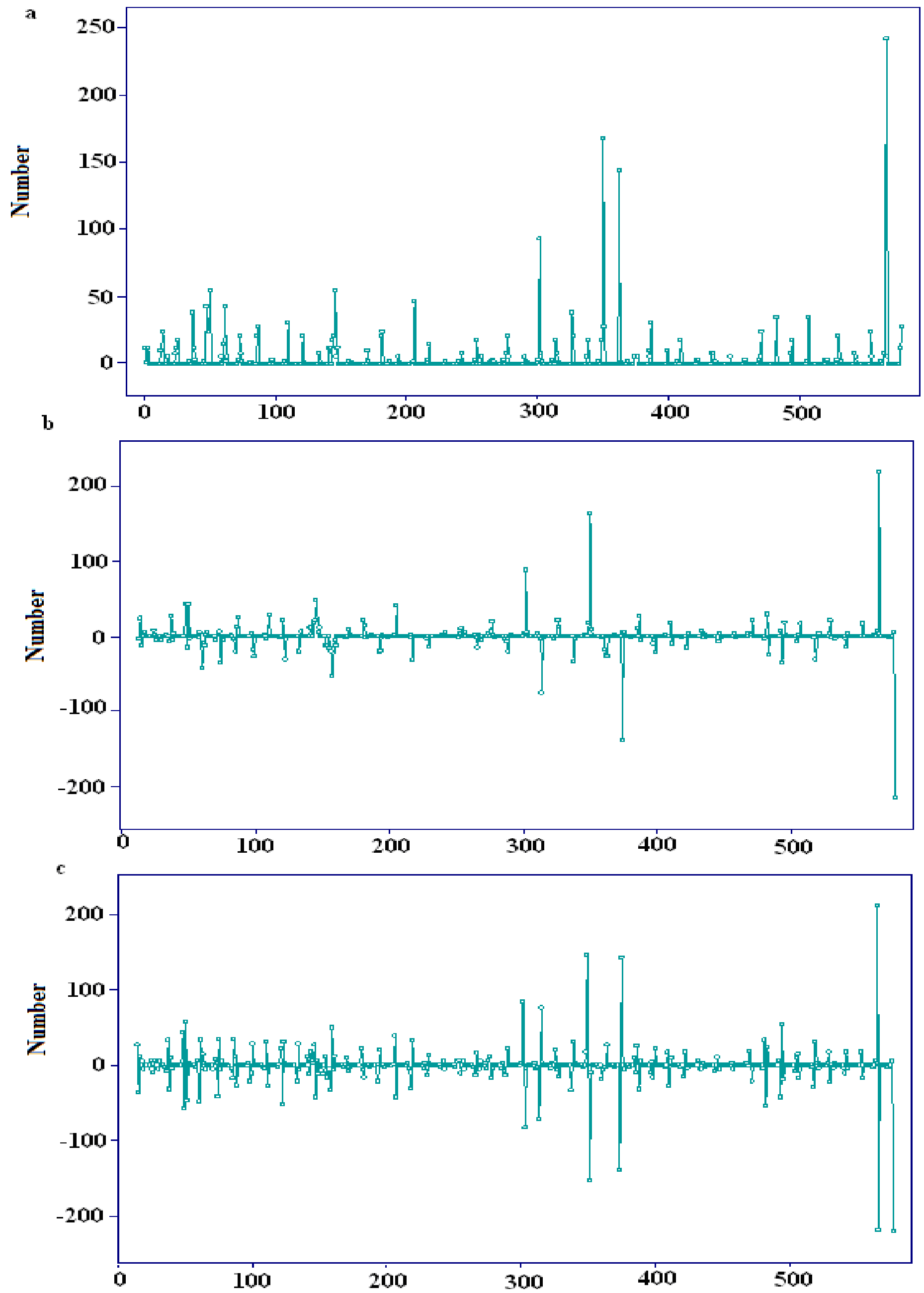

Figure 4.1 a. Time plot after Box-Cox transformation at λ=1.5; b. Time plot after differencing at lag 12; c. Time plot after further differencing at lag 1.

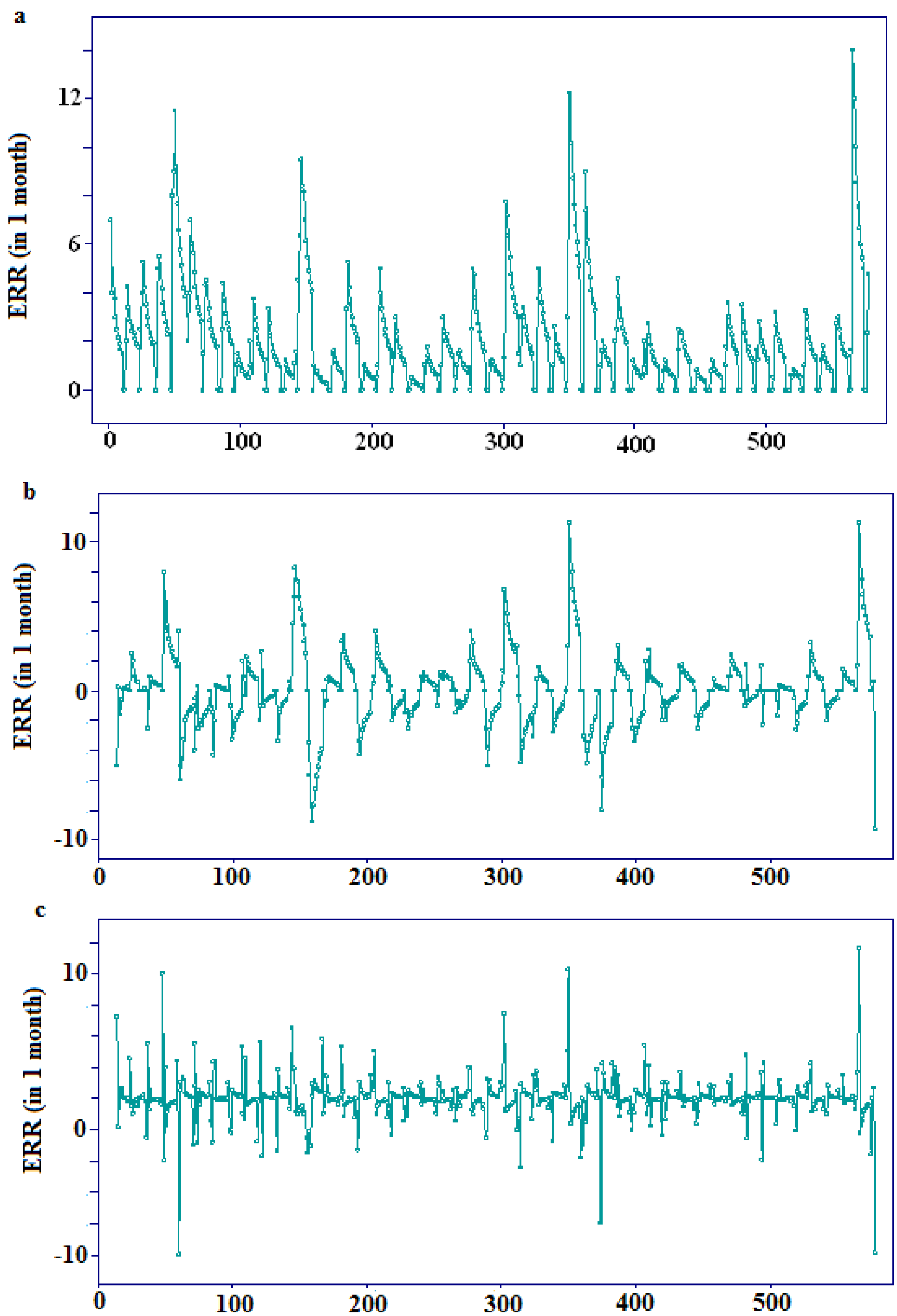

Figure 4.2 a**.** Time plot of the non-cumulative ERR. b. Time plot after differencing at lag 12. c. Time plot after further differencing at lag 1.

Therefore, a modified data set, which includes 565 months of sand storms (April 1955- April 2002), will be used to demonstrate our modeling techniques. Consequently, the training sample (537 lags, Figure 4.4a) consists of all the data set excluding the last 28 ERRs (January 2000 – April 2002), which is the prediction set.

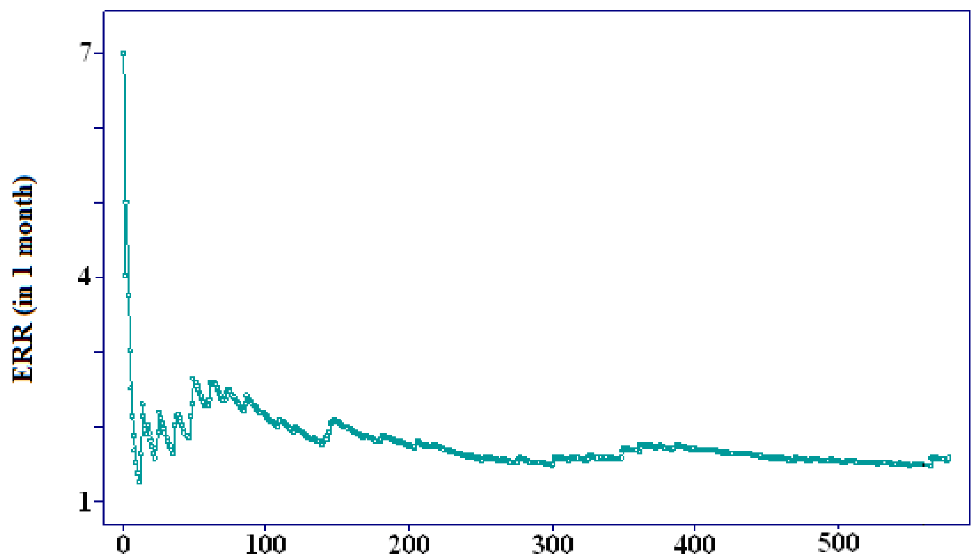

Figure 4.3 ERR plot with whole data set from March 1954 – April 2002

4.3.1 Model search for the training sample

The plots of the sample ACF and PACF (Figure 4.4b, c) show that the spikes are slowly decaying. This indicates nonstationary behavior and seasonality. Thus differencing is applied. Since the data has evident increasing or decreasing variability, we consider the Box-Cox transformation to stabilize the variability. After the λ=1.5 Box-Cox transformation, we see the trend still exists (Figure 4.5). Then we take the differencing operator ∇ on the training sample at lag 12, as the data is collected monthly. Figure 4.6

tells us the series has not reached stationarity yet.

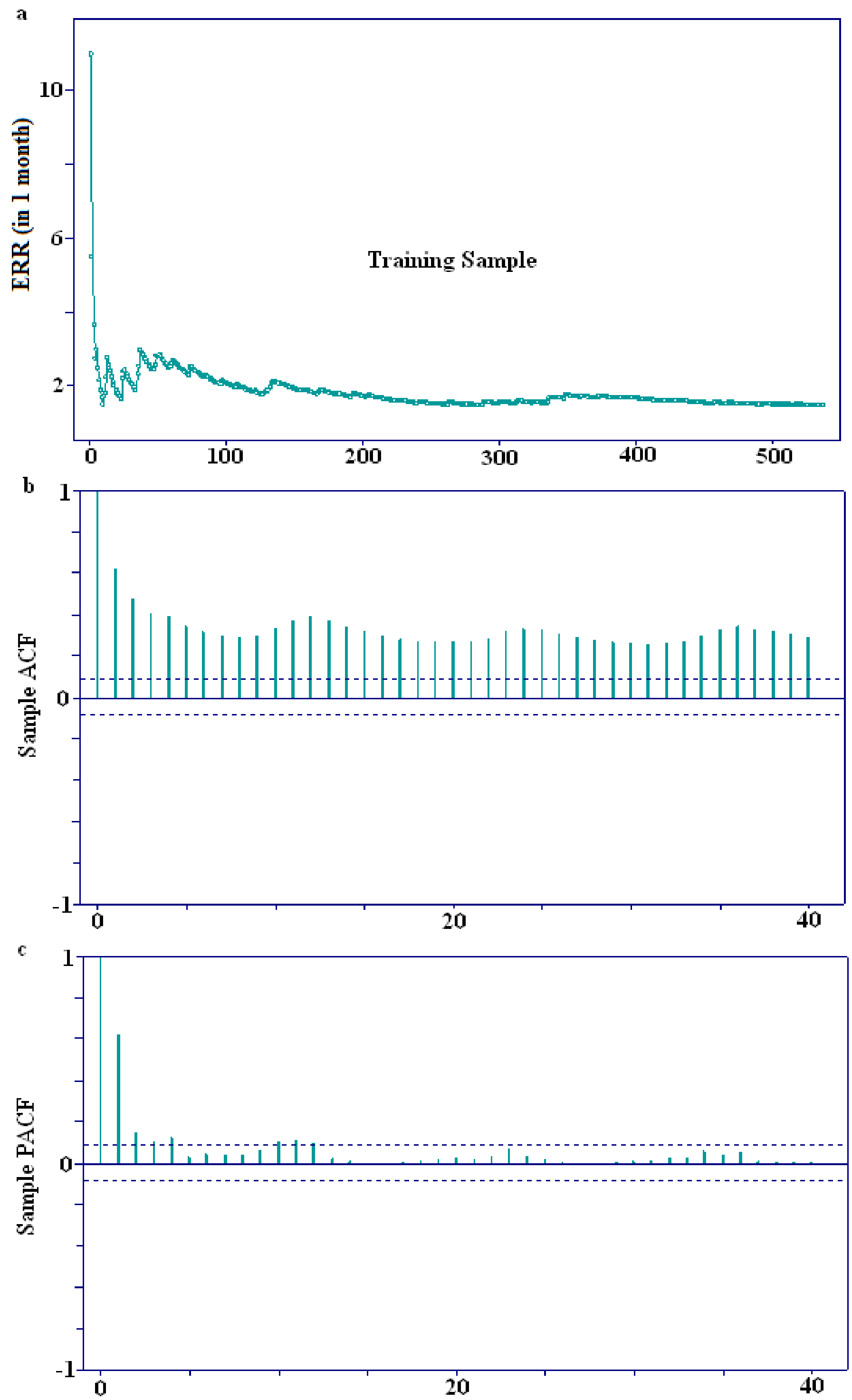

Figure 4.4 a, Time-plot of the ERR based on the 'trimmed' data set; b, Sample ACF; c, Sample PACF.

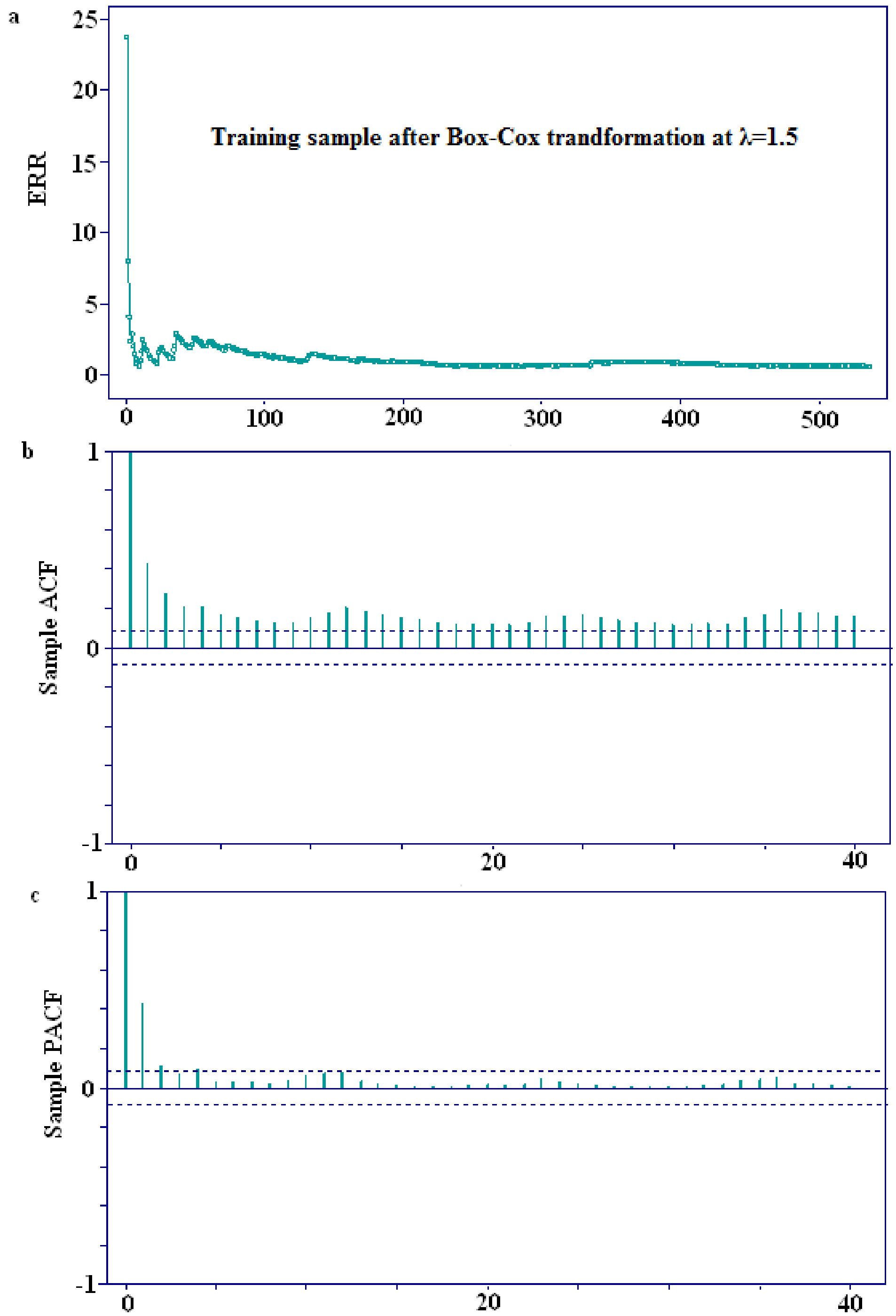

Figure 4.5 a, Time-plot after Box-Cox transformation at λ=1.5 ; b, Sample ACF; c, Sample PACF.

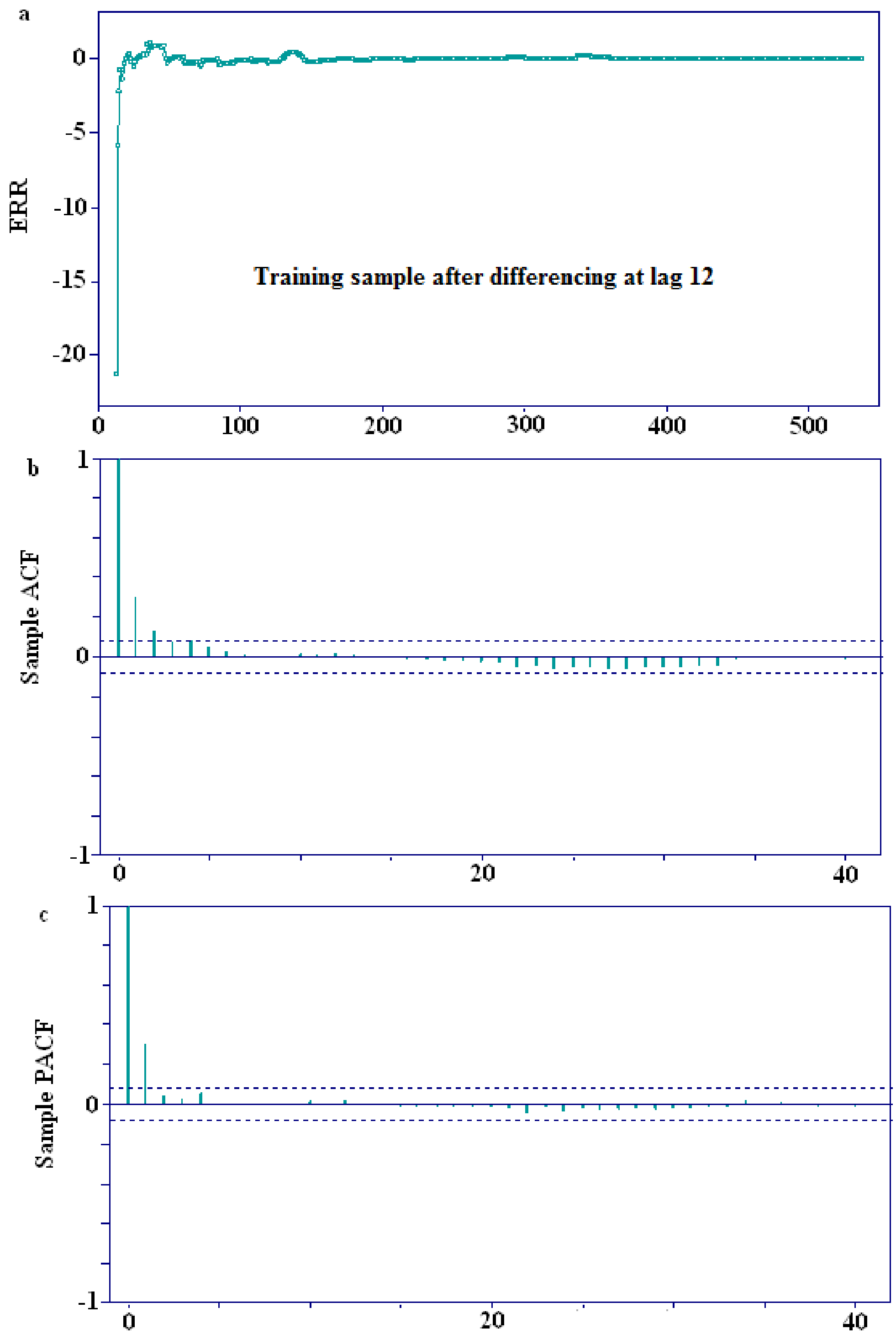

Figure 4.6 a, Time-plot after differencing at lag 12; b, Sample ACF; c, Sample PACF.

So we do further differencing at lag 1, Then we subtract the sample mean from each observation of the differenced series to generate a stationary zero-mean time series

(Figure 4.7). The sample ACF and PACF suggest an ARMA(1,1). Therefore, our initial model estimated using MLE is:

$X_t = 0.9950\ X_{t-1} + Z_t + 0.2688\ Z_{t-1}$

Estimated WN Variance = 0.286035

Standard Error of AR Coefficients: 0.000363

Standard Error of MA Coefficients: 0.050929

Note that X_t represents a twice-differenced stationary mean-corrected time series and the error term Z_t represents a white noise process.

The AICC statistic is 842.380, and the Ljung - Box test is significant (p-value =0), indicating that the residuals are not approximately white noise. The plots of the residual, ACF and PACF of residual are shown in Figure 4.8. For the ARMA (1,1), we can calculate the ratios (estimated coefficient)/(1.96×standard error), the ratios of each coefficient are:

Ratio of AR coefficient= 1.008444

Ratio of MA coefficient= -0.451766

We see the absolute value of ratio of MA coefficient is less than 1, so we keep the AR coefficient, which indicates a subset AR (1) model, the MLE of AR(1) is

$X_t = 0.9966\ X_{t-1} + Z_t$

Estimated WN Variance = 0.298764

Standard Error of AR Coefficients: 0.003595

Subset model AR (1) is considered to fit the time series $\nabla\nabla_{12}z_t$, ITSM2000 produces AICC value and a set of diagnostic plots such as residual sample ACF and sample PACF, as shown in Figure 4.9. The AICC statistic is 863.020, and the Ljung - Box test is not significant (p-value = 0.15533), indicating that the residuals are approximately white noise. The prediction calculation is performed using the ITSM2000. Table 4.1 shows the numerical values of the actual ERRs and mean numbers in the prediction set, and the predicted ERRs using the AR (1) (Subset of ARMA(1,1)) with their counterparts. Figure 4.10 shows the comparison between the actual ERRs and predicted ERRs.

4.3.2 Full-Data Forecasting

We already saw the possibility of the ERR modeling with the training sample, now we focus on the ARIMA model which is using the full data set (including the prediction set). This training sample consists of 565 lags from April, 1955 to April, 2002 (Figure 4.11a.). As before, we still take the Box-Cox transformation at λ=1.5 (Figure 4.12). Next we difference the time series at lag 12 (Figure 4.13). Then we model this stationary series with an ARMA(1,1), the initial model estimated in MLE is:

$X_t = 0.9968\ X_{t-1} + 0.6301Z_{t-1} + Z_t$

Estimated WN Variance = 0.383546.

Standard Error of AR Coefficients: 0.003398.

Standard Error of MA Coefficients: 0.033046.

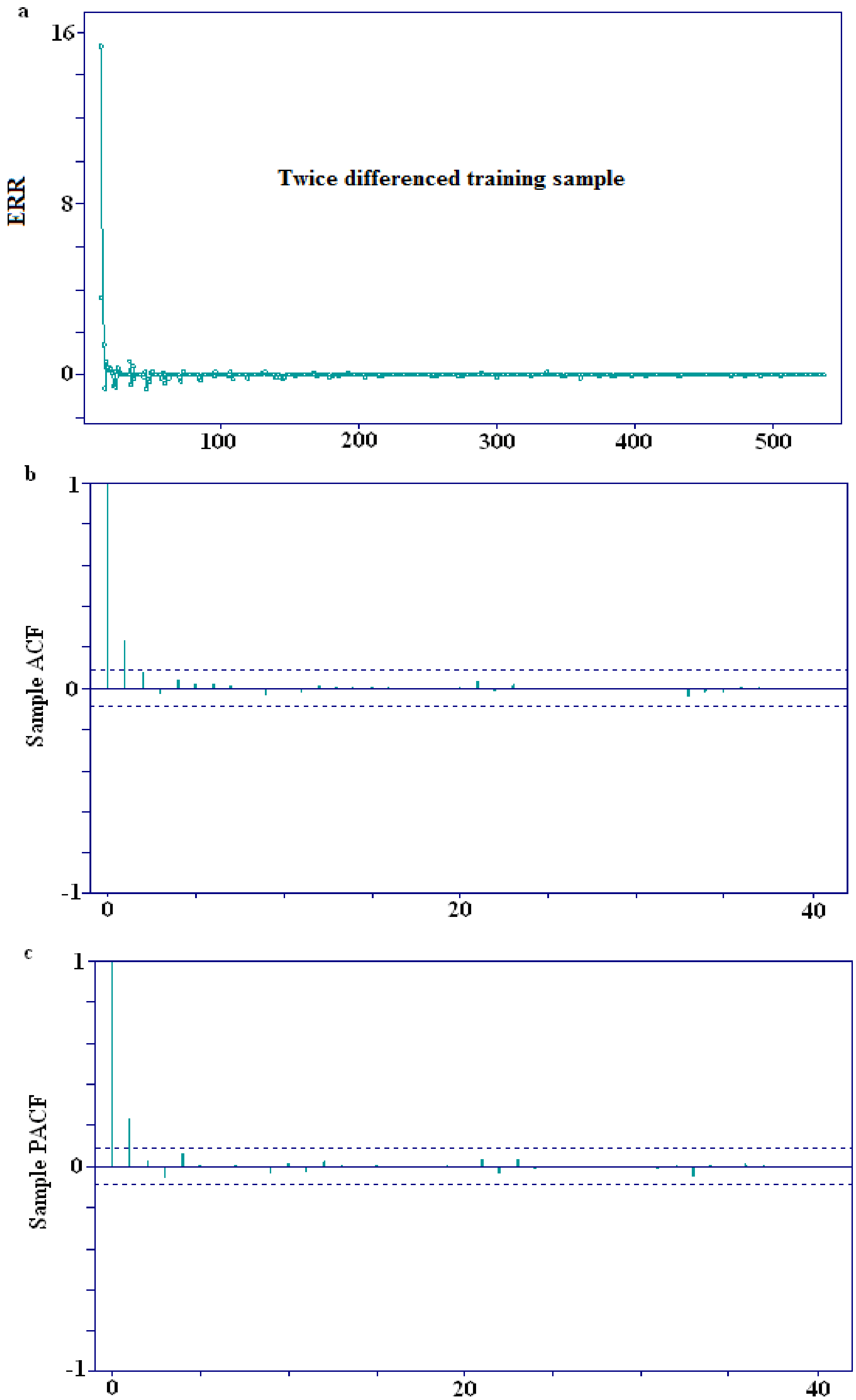

Figure 4.7 a, Time-plot after differencing at lag 1; b, Sample ACF; c, Sample PACF.

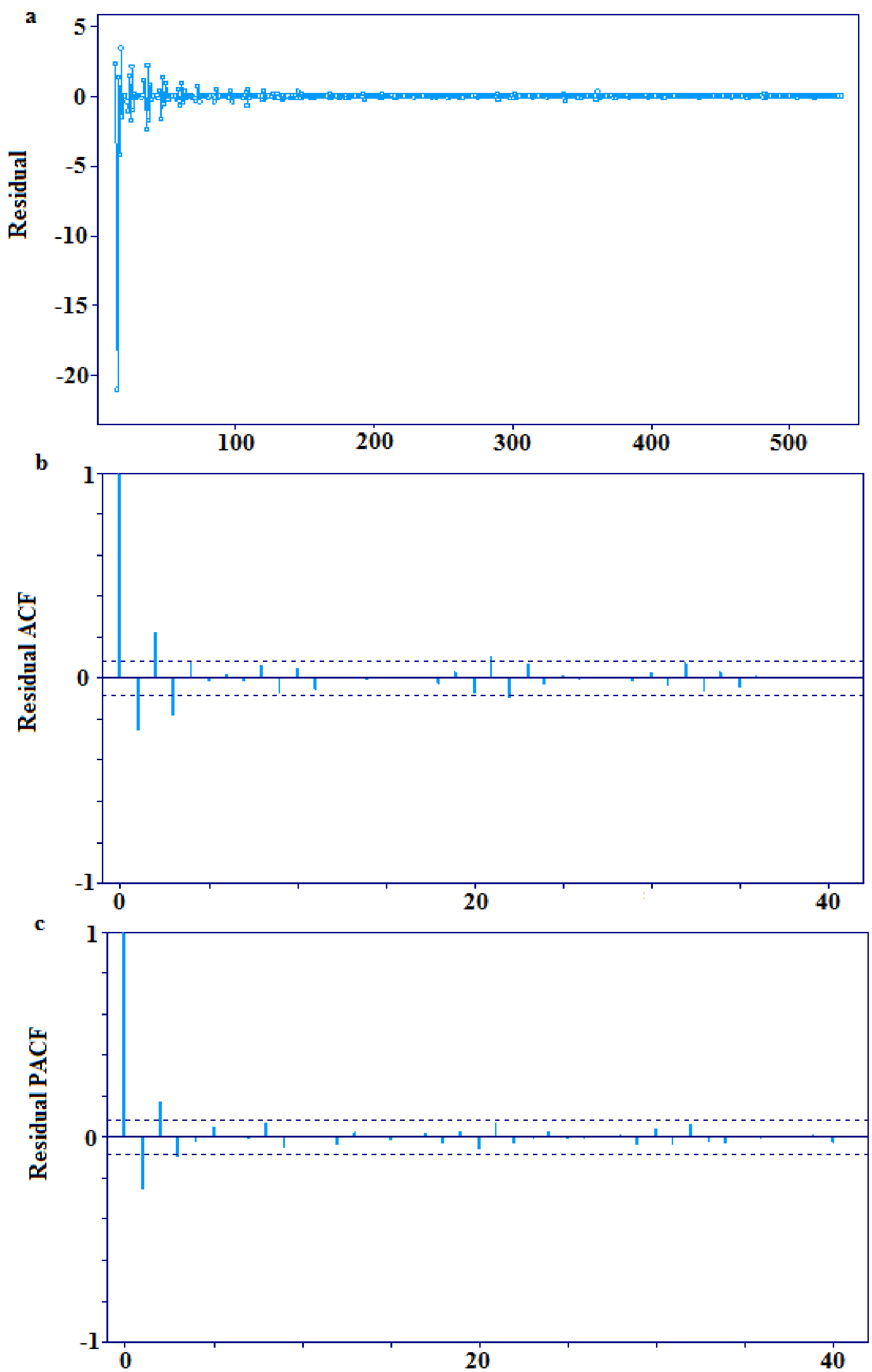

Figure 4.8 Diagnostics for the ARMA (1,1) model. a, Residual plot; b, Residual ACF; c, Residual PACF.

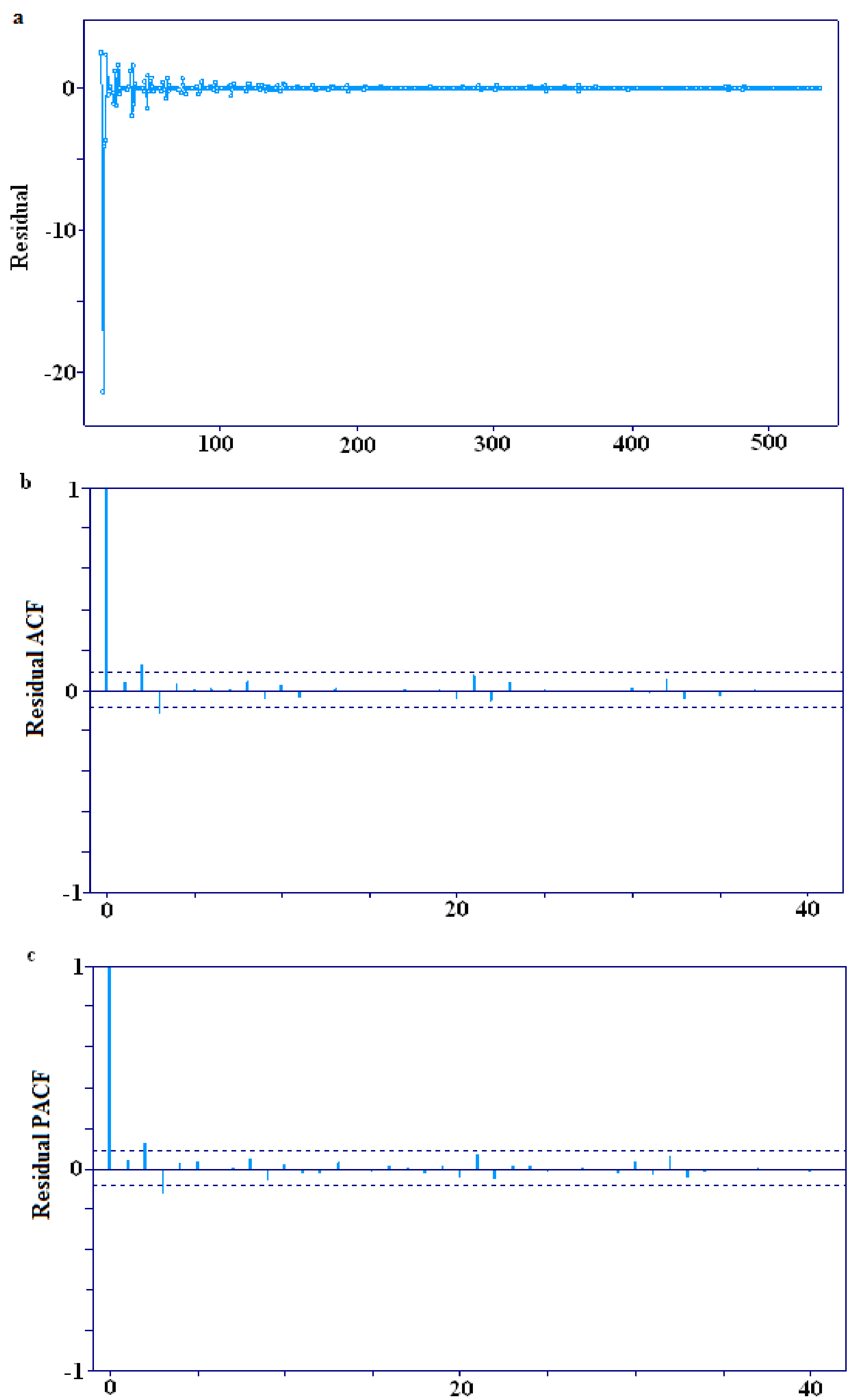

Figure 4.9 Diagnostics for the subset AR (1) model a, Residual plot; b, Residual ACF; c, Residual PACF.

Table 4.1. The numerical values of the actual ERRs and mean numbers in the prediction set, and the predicted ERRs using the subset AR (1) with their counterparts (the corresponding mean values derived from the predicted ERRs)

Predicted month	Monthly ERR		Mean number	
	Actual	Prediction	Actual	Prediction
January 2000	1.470260223	1.47037	0	0.05906
February 2000	1.467532468	1.46787	0	0.12287
March 2000	1.464814815	1.4655	0	0.19044
April 2000	1.482439926	1.47274	11	5.36786
May 2000	1.487084871	1.47818	4	4.4049
June 2000	1.484346225	1.47613	0	0.37323
July 2000	1.481617647	1.47421	0	0.44125
August 2000	1.478899083	1.47241	0	0.50401
September 2000	1.476190476	1.47074	0	0.57228
October 2000	1.473491773	1.46918	0	0.6299
November 2000	1.47080292	1.46775	0	0.69841
December 2000	1.471766849	1.46644	2	0.76166
January 2001	1.469090909	1.46536	0	0.88432
February 2001	1.466424682	1.46452	0	1.0126
March 2001	1.472826087	1.46390	5	1.13034
April 2001	1.56238698	1.47303	51	6.38497
May 2001	1.566787004	1.48043	4	5.46163
June 2001	1.563963964	1.48047	0	1.50199
July 2001	1.561151079	1.48074	0	1.626
August 2001	1.558348294	1.48123	0	1.74485
September 2001	1.555555556	1.48196	0	1.8747
October 2001	1.552772809	1.48291	0	1.99401
November 2001	1.55	1.48408	0	2.11354
December 2001	1.547237077	1.48548	0	2.23868
January 2002	1.544483986	1.48722	0	2.42334
February 2002	1.541740675	1.48929	0	2.60295
March 2002	1.55141844	1.49169	7	2.78289
April 2002	1.569911504	1.50383	12	8.03515

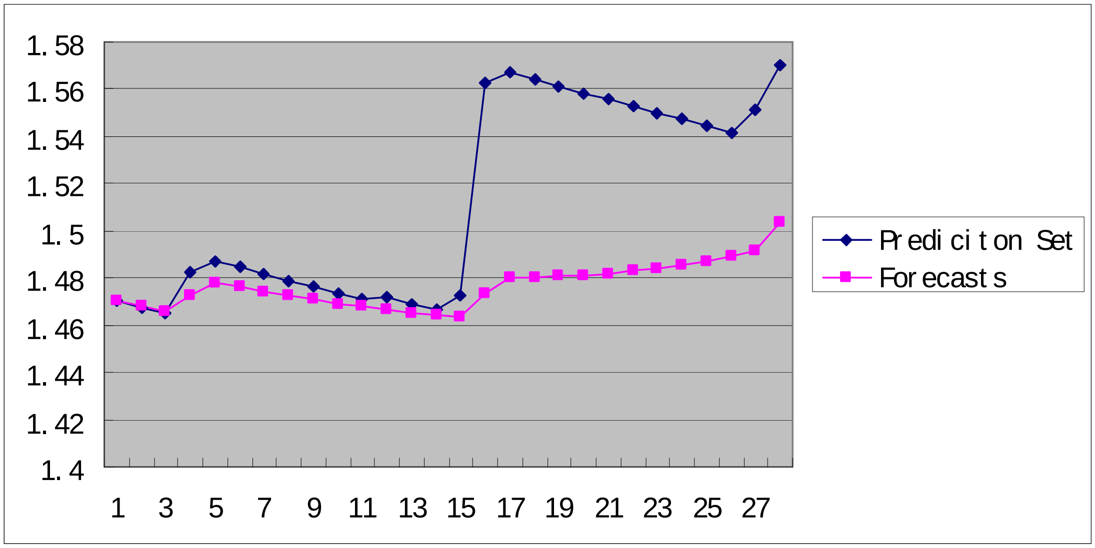

Figure 4.10 Comparison of 28 forecasted ERRs with the prediction set and actual ones

The AICC statistic is 1051.99, and the Ljung - Box test is significant (p-value =0), indicating that the residuals are not approximately white noise. The plots of the residual, residual of ACF and PACF are shown in Figure 4.14. For the ARMA (1,1), we can calculate the ratios (estimated coefficient)/(1.96×standard error), the ratios of each coefficient are:

Ratio of AR coefficient= 1.324450

Ratio of MA coefficient= -0.436091

We see the absolute value of ratio of MA coefficient is less than 1, so we keep the AR coefficient, which indicates a subset AR (1) model, the MLE of AR(1) is:

$X_t = 0.9968\ X_{t-1} + Z_{t.}$

Estimated WN Variance = 0. 465074.

Standard Error of AR Coefficients: 0. 003394.

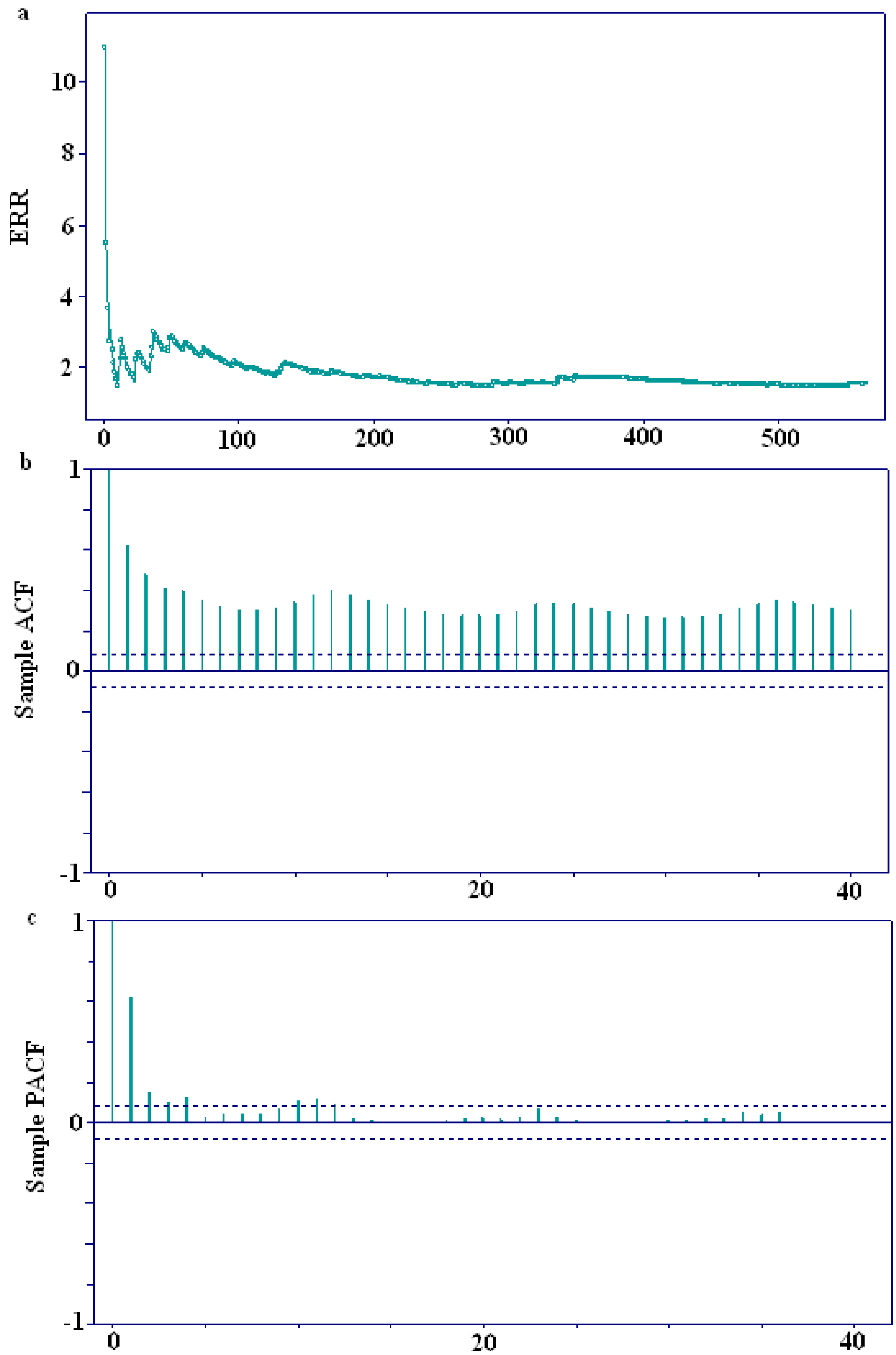

Figure 4.11 a, Time plot of full data set; b, Sample ACF; c, Sample PACF.

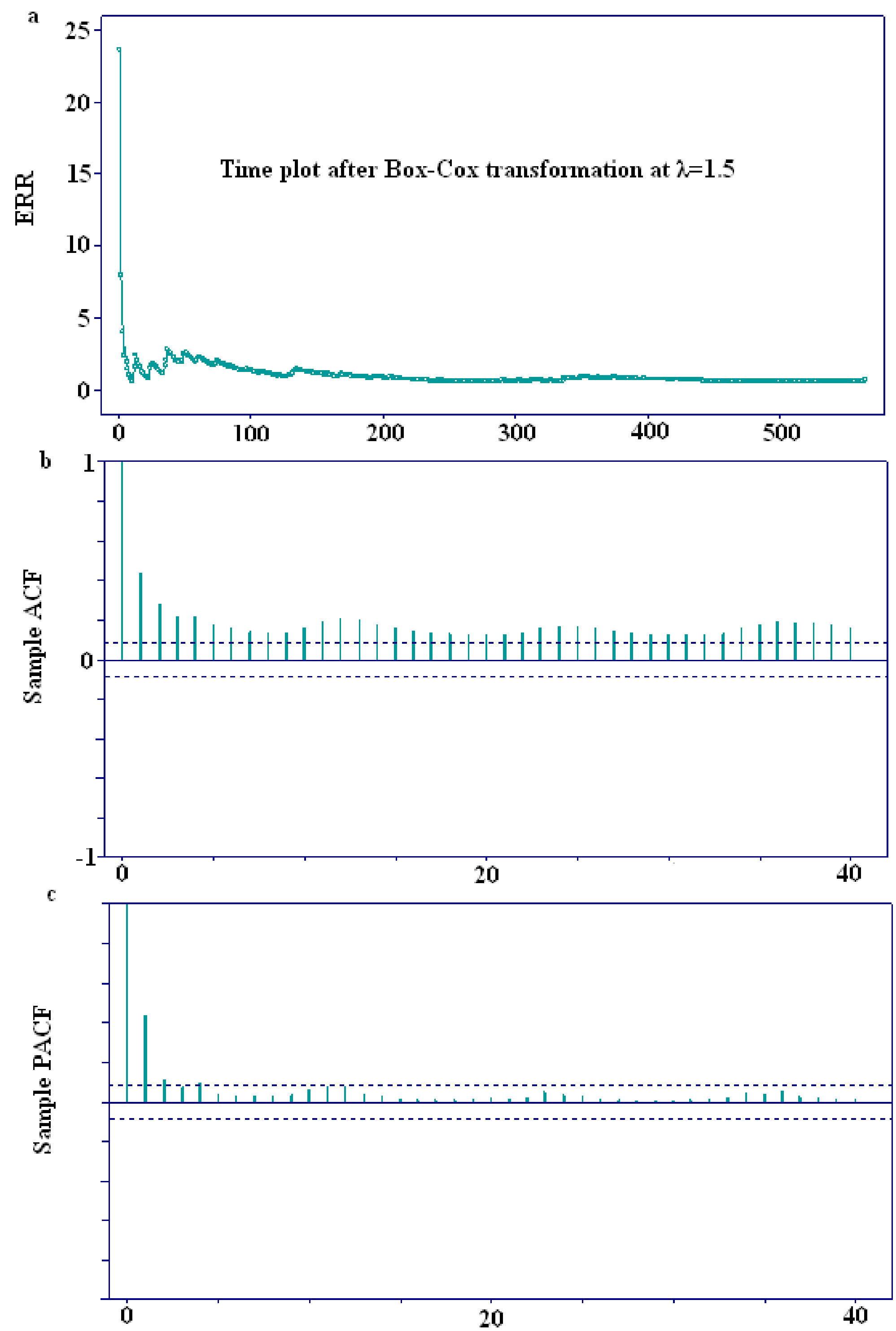

Figure 4.12 a, Time plot after Box-Cox transformation at λ=1.5 ; b, Sample ACF; c, Sample PACF.

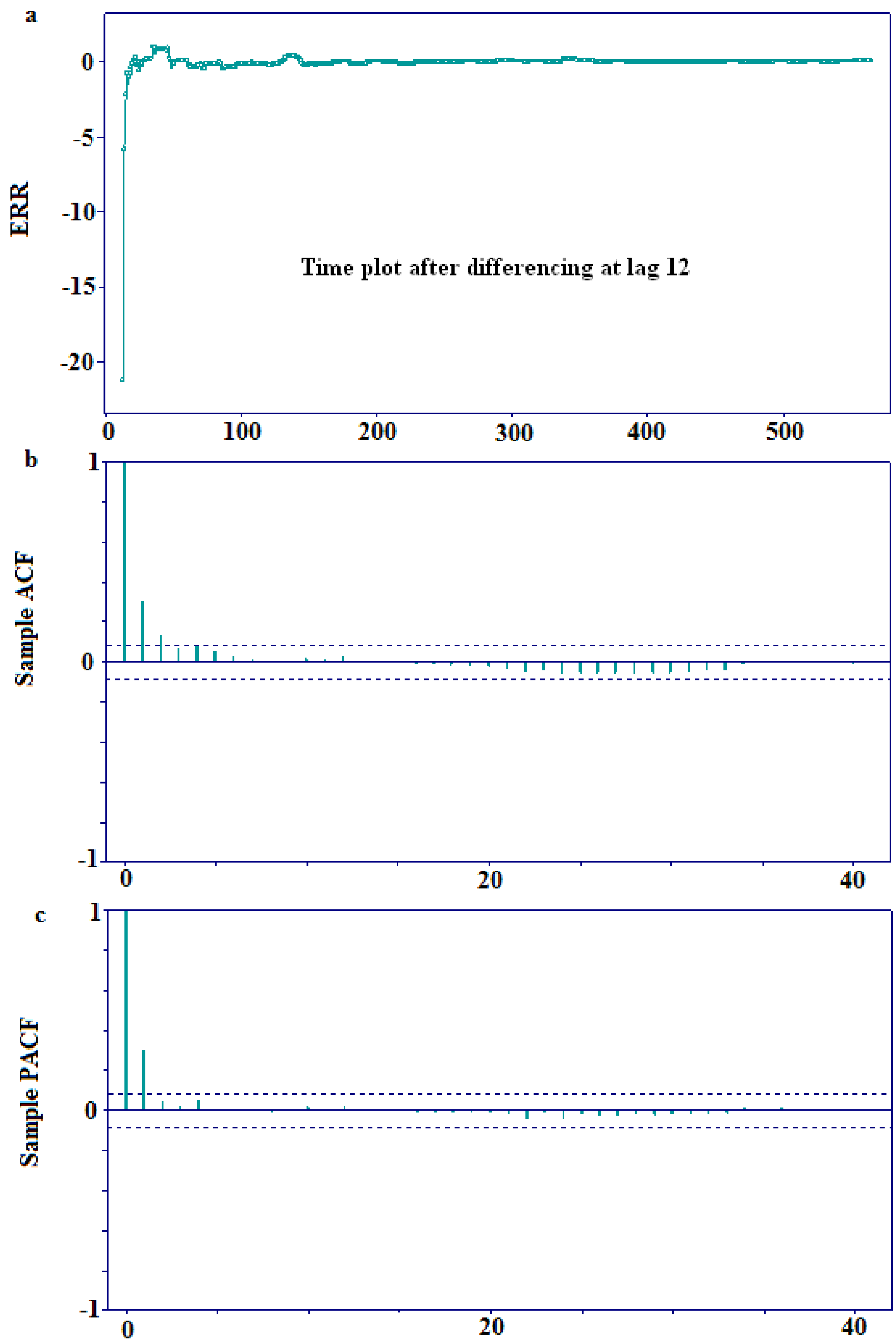

Figure 4.13 a, Time plot after differencing at lag 12; b, Sample ACF; c, Sample PACF.

The AICC statistic is 1155.07, and the Ljung - Box test is not significant (p-value =0.75132). The plots of the residual, residual of ACF and PACF are shown in Figure 4.15. Table 4.2 shows the 24 forecasts values of actual ERRs and mean numbers.

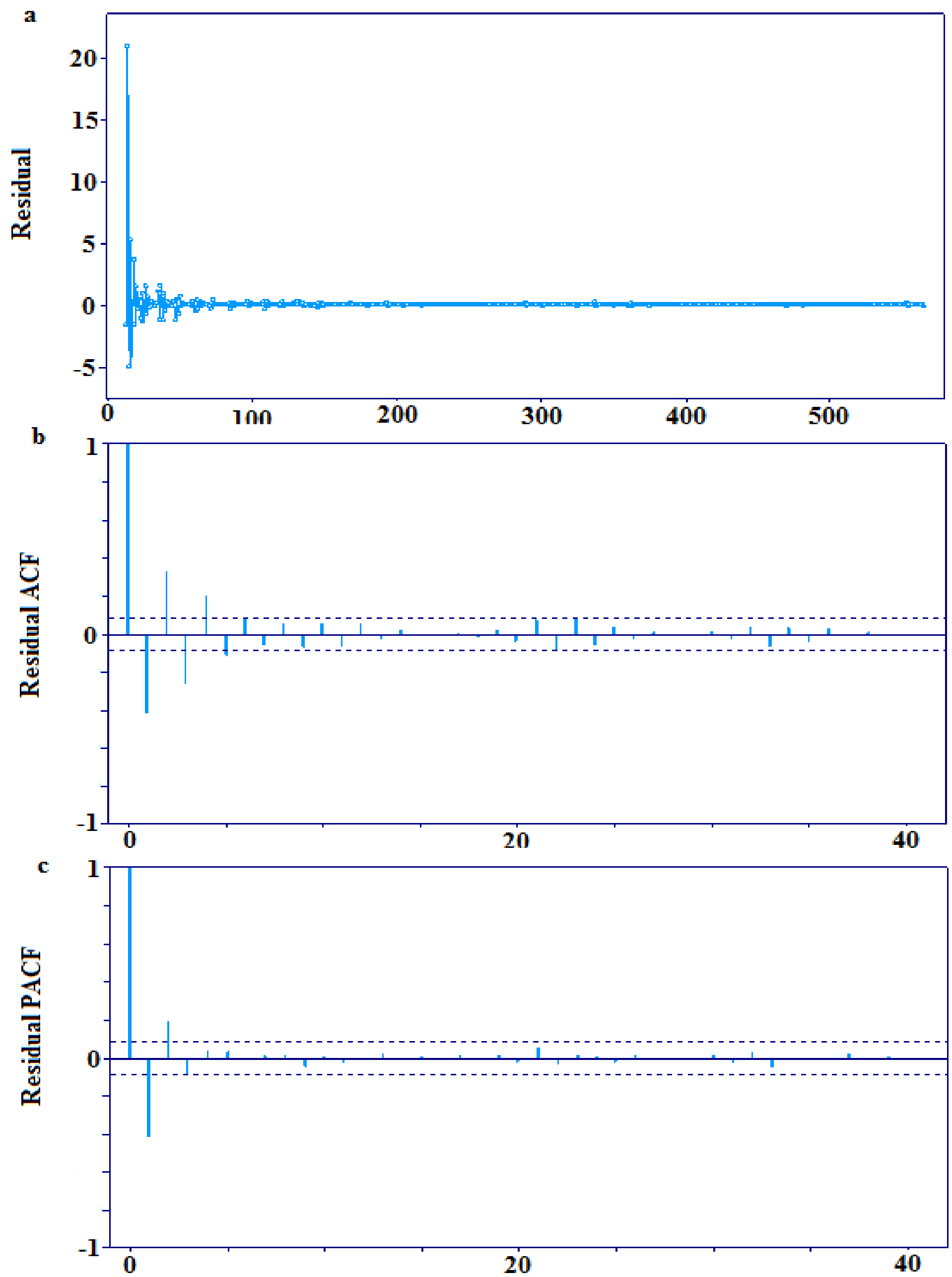

Figure 4.14 Diagnostics for the ARMA (1,1) model. a, Residual plot; b, Residual ACF; c, Residual PACF.

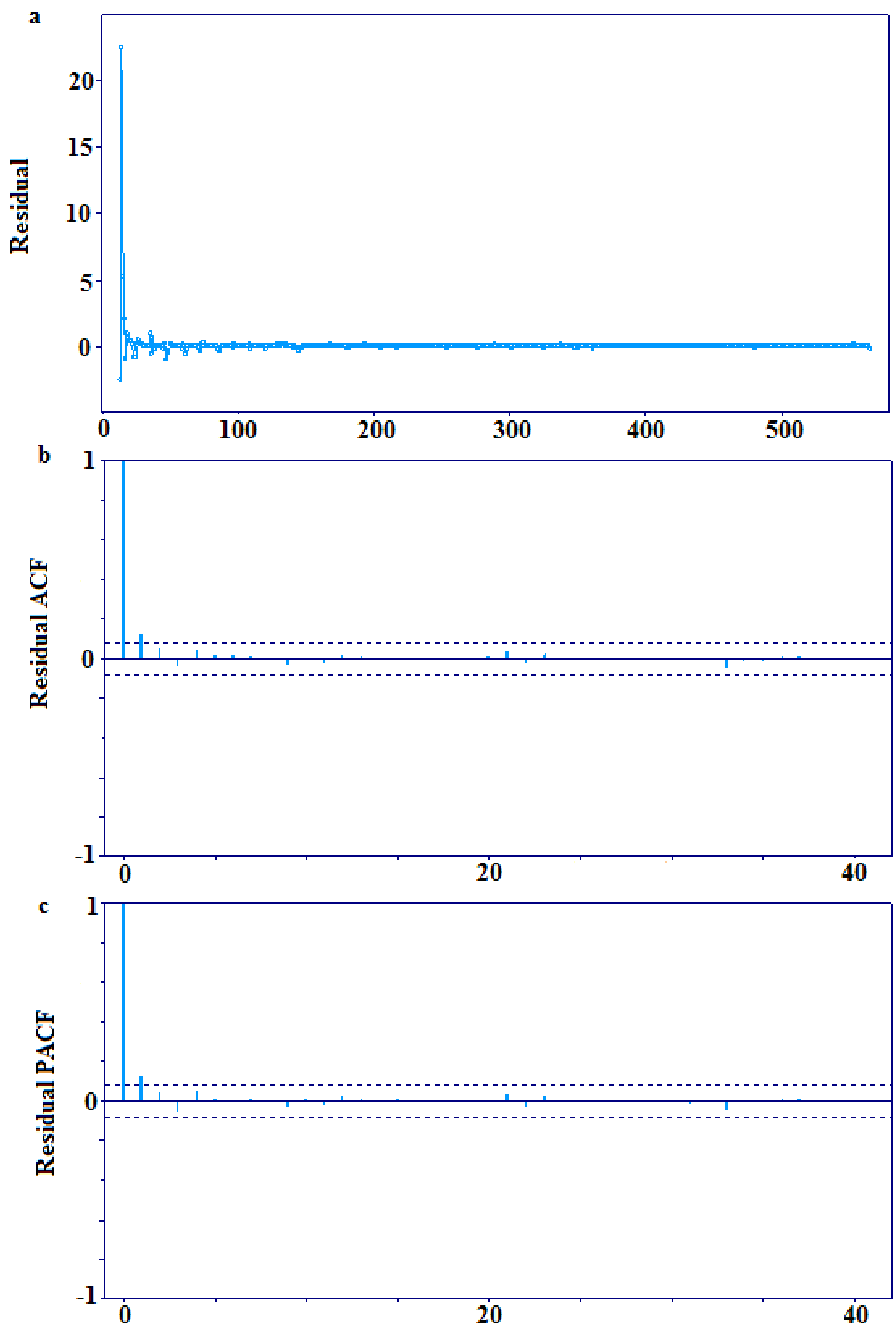

Figure 4.15 Diagnostics for the subset AR (1) model. a, Residual plot; b, Residual ACF; c, Residual PACF.

Table 4.2 The numerical values of the predicted ERRs and mean numbers (the corresponding mean values derived from the predicted ERRs)

Predicted month	Monthly ERR	Mean number
	Prediction	Prediction
May 2002	1.57409	3.93494
June 2002	1.57105	-0.14959 adjust to 0
July 2002	1.56803	-0.14431 adjust to0
August 2002	1.56502	-0.14466 adjust to 0
August 2002	1.56202	-0.14498 adjust to 0
September 2002	1.55903	-0.14527 adjust to 0
October 2002	1.55605	-0.14553 adjust to 0
November 2002	1.55309	-0.14003 adjust to 0
December 2002	1.55013	-0.14595 adjust to 0
January 2003	1.54718	-0.14612 adjust to 0
February 2003	1.55663	6.99038
March 2003	1.57488	12.08688
April 2003	1.57884	3.86376
May 2003	1.57561	-0.29133 adjust to 0
June 2003	1.57239	-0.29199 adjust to 0
July 2003	1.56918	-0.29262 adjust to 0
August 2003	1.56598	-0.29322 adjust to 0
September 2003	1.56279	-0.29379 adjust to 0
October 2003	1.55961	-0.29433 adjust to 0
November 2003	1.55644	-0.29484 adjust to 0
December 2003	1.55328	-0.29532 adjust to 0
January 2004	1.55013	-0.29577 adjust to 0
February 2004	1.55937	6.98325
March 2004	1.57741	12.18493

4.4 The role of prediction set

We use the prediction set to facilitate our model selection process. It is quite common that several ARIMA models may fit a training sample equally well. However, the quality of forecasting varies among candidate models. Therefore, our prediction set plays an important role in filtering out the model with poor prediction. Recall that we have a large variation at the beginning of the series, a common practice is to delete a small subset of ERR at the beginning of the time series, which is called burn in period.

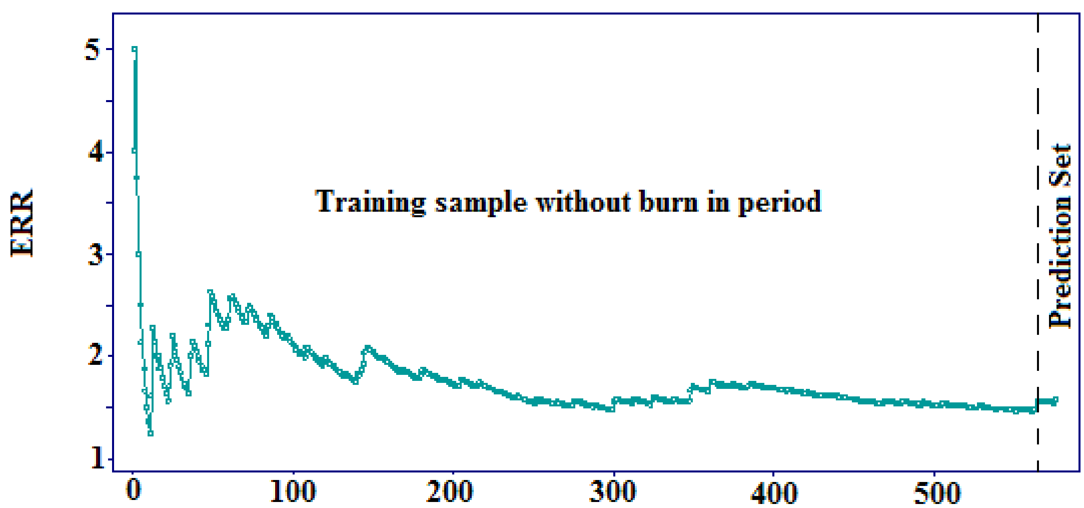

Figure 4.16 ERR plot with training sample and prediction set

In order to demonstrate this point, we apply the same techniques in 4.3 to following time series: (1) converting the whole data set (March 1954 – April 2002) to an ERR time series; (2) delete the first ERR; (3) let the first 549 ERRs be the training sample and keep the last 28 lags as the prediction set. Figure 4.16 shows the plot of the training sample and prediction set. To model the training sample, we use the same transformations: first the Box-Cox transformation at λ=1.5, next differencing at lag 12 and 1. An AR(1) model is considered. The MLE of the AR(1) is:

$X_t = -0.1833\ X_{t-1} + Z_t$

Estimated WN Variance = 0.029588

Standard Error of AR Coefficients: 0.053076

Note that X_t represents a twice differenced stationary mean-corrected time series and the error term Z_t represents a white noise process.

The AICC statistic is -361.796, and the Ljung - Box test is not significant (p-value = 0.26756), indicating that the residuals are approximately white noise. The plots of the ACF and PACF residuals are shown in Figure 4.17. Figure 4.18 shows the comparison between the actual ERRs and predicted ERRs using the model 1, which is not encouraging, even though, the model passes all the diagnostics. Figure 4.19 shows the comparisons of the results with the prediction set for model 2. Model 1(4.3.1) and 2(4.4) are defined in Table 4.3. Clearly, the forecasts from the model 1 appear to be more realistic in showing the seasonality of the sand storms occurred in northern China area.

Table 4.3. The two ARIMA models with different data sets.

MLE Model	Data Set
1. $X_t = 0.9966\ X_{t-1} + Z_t$	Without the first 13 months of raw data.
2. $X_t = -0.1833\ X_{t-1} + Z_t$	Without the burn in period of the first ERR.

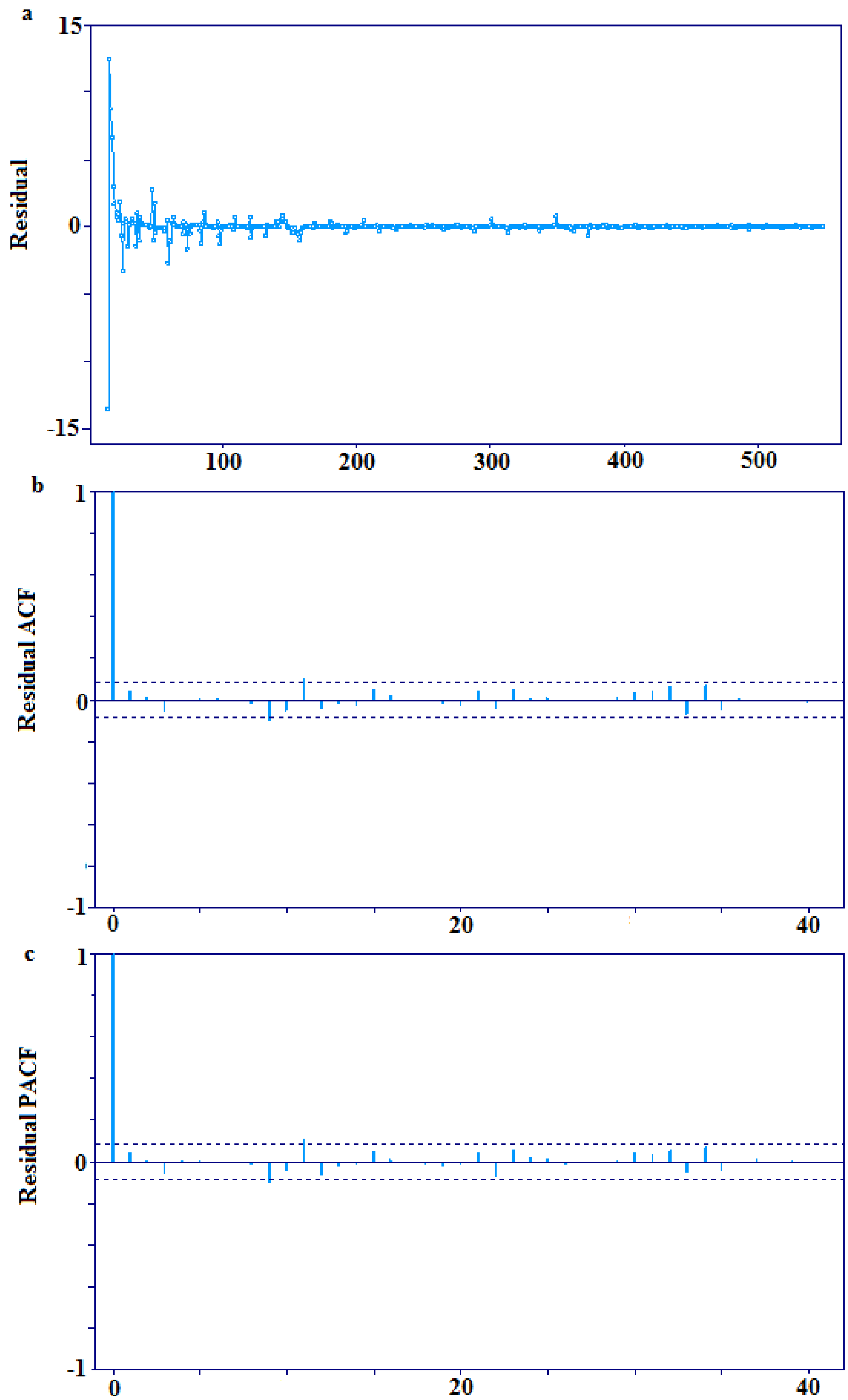

Figure 4.17 Diagnostics for the AR(1) model. a, Residual plot; b, Residual ACF; c, Residual PACF.

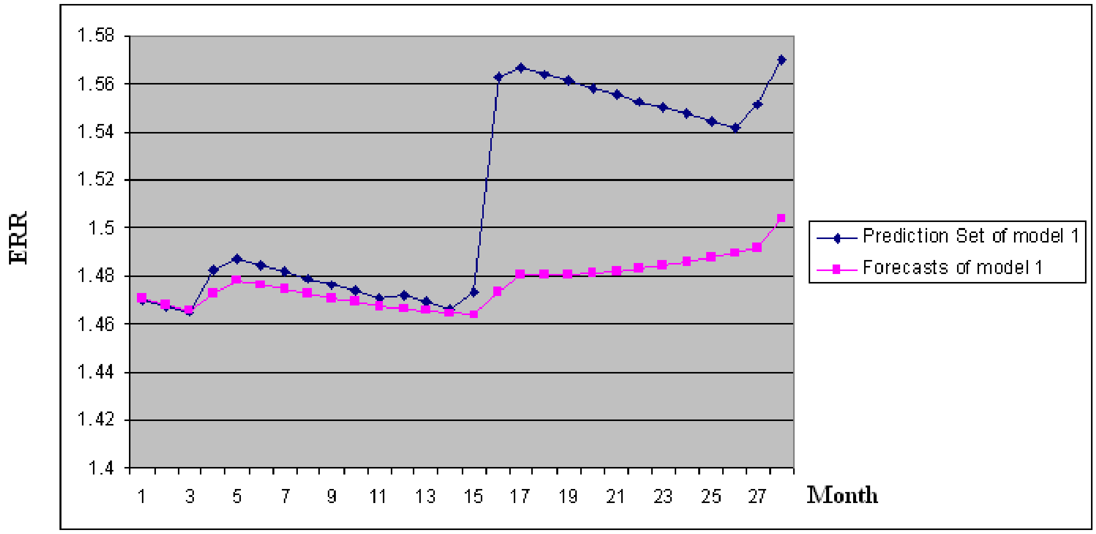

Figure 4.18 Comparison of 28 forecasted ERRs using model 1 with the prediction set

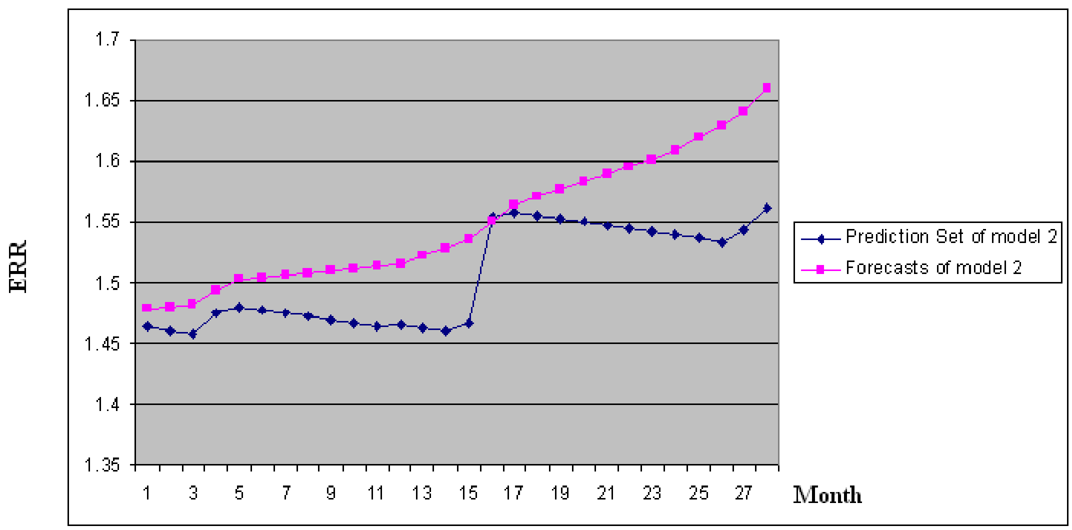

Figure 4.19 The comparisons of the results with the prediction set for model 2

CHAPTER 5

CONCLUSIONS

In the thesis, we are able to find a suitable ARIMA model for the sand dust storms which occurred in northern China from March 1954 to April 2002 by using the empirical recurrence rates (ERR). ERR is calculated sequentially at equidistant time intervals. We adopt this method which is used to smooth the data that fits into the ARIMA. Clearly, the number of the sand dust storms does not follow a Poisson process. Through the ERR, we build a linking bridge between the classical time series and a point process.

Before fitting the real data into ARIMA model, we examine the applicability of the proposed technique based on the simulated data: We randomly select a year of the raw data, repeat it 17 times and calculate the 180 ERRs, use first 15 years as our training sample, and the last two years as the prediction set. An MA(1) predicts very well. Next we apply this technique into the real data analysis. After the initial model search, we find that it is difficult to fit the ARIMA with the ERRs based on the whole data set, so we drop the raw data about 13 months and recalculate the ERRs. The training sample is the rest of the data set excluding the last 28 months, which is the prediction set. Before modeling, we must make sure that the ARMA process is stationary. First we take the Box-Cox transformation at λ=1.5 then differencing twice at lag 12 and 1 respectively. A subset AR(1) of ARMA(1,1) passes the randomness test for residuals and has all the residual ACF lags falling within the bounds $\pm 1.96/\sqrt{n}$, it predicts well in the short term of 12 months compare to the prediction set. Next, we also find a subset AR(1) of

ARMA(1,1) using the full data set. It successfully predicts a seasonal trend in the following two years. In addition, we try several other methods to analyze the data. The model we find for the data which cut off burn in period does not predict very well. We hardly find a suitable model for the raw data without ERR conversion or non-cumulative ERR time series, since they have trouble in making the stationary time series.

From our work we see that the ERR is an effective way to analyze the meteorological data. The sand storm data display a special seasonality, with the maximum frequencies in March, April and May. The rest of the years are zeros. After converting to an ERR time series, we can use the classical ARIMA techniques to model and predict sand dust storms. Likewise, the application of ARIMA models for sand storms will further facilitate the research in the areas of biology, economies, social science, etc.

BIBLIOGRAPHY

[1] Box, G..E.P. and Jenkins, G.M. (1976) *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

[2] Brockwell, P.J. and Davis, R.A. (2002). *Introduction to Time Series and Forecasting*. 2nd Edition. Springer-Verlag, New York.

[3] Cowpertwait, P.S.P and Metcalfe, A.V. (2009). *Introductory Time Series with R.* Springer, New York.

[4] Cryer, J.D and Chan, K.S. (2008). *Time Series Analysis With Applications in R.* 2nd Edition. Springer, New York.

[5] Goudie, A. S., and N. J. Middleton (1992). The changing frequency of dust storms through time. *Climatic Change*, **20**, 197–225.

[6] Ho, C.-H. (2008). Empirical recurrent rate time series for volcanism: application to Avachinsky volcano, Russia. *Volcanol Geotherm Res*, 173, 15-25.

[7] Li,W.K. (2004). *Diagnostic Checks in Time Series.* Chapman & Hall/CRC, Florida

[8] Shumway, R.H. and Stoffer, D.S. (2006). *Time Series Analysis and Its Applications With R Examples.* Springer, New York.

[9] Tao, G., L. Jingtao, Y. Xiao, K. Ling, F. Yida, and H. Yinghua (2002). Objective pattern discrimination model for dust storm forecasting. *Meteor. Appl*., 9, 55–62.

[10] Warner, T.T. (2004). *Desert Meteorology*. Cambridge University Press, New York

[11] Yang B, Brauning W, Zhang Z.Y, Dong Z.B, Esper J. (2007). Dust storm frequency and its relation to climate changes in Northern China during the past 1000 years. *Atmospheric Environment*, 41, 9288–9299

[12] Zhang, Q.Y, Zhao, X.Y, Zhang, Y and Li, L. (2002). Preliminary study on sand-dust storm disaster and countermeasures in China. *Chinese Geographical Science*, 12(1), 9-13.

[13]Zhou, Z.J and Zhang G.C. (2003). Typical strong sand storm events in the Northern China. *Chinese Science Bulletin*, 48(11), pp 1224-1228.

APPENDIX

DATA

Table 1 Number of the sand storms in northern China (March1954-April 2000)

Year	Month	Count	Year	Month	Count
1954	March	7	1957	January	0
1954	April	1	1957	February	0
1954	May	7	1957	March	15
1954	June	0	1957	April	7
1954	July	0	1957	May	3
1954	August	0	1957	June	0
1954	September	0	1957	July	0
1954	October	0	1957	August	0
1954	November	0	1957	September	0
1954	December	0	1957	October	0
1955	January	0	1957	November	0
1955	February	0	1957	December	2
1955	March	6	1958	January	0
1955	April	11	1958	February	16
1955	May	0	1958	March	11
1955	June	0	1958	April	19
1955	July	0	1958	May	0
1955	August	4	1958	June	0
1955	September	0	1958	July	0
1955	October	0	1958	August	0
1955	November	0	1958	September	0
1955	December	0	1958	October	0
1956	January	0	1958	November	0
1956	February	5	1958	December	0
1956	March	7	1959	January	4
1956	April	9	1959	February	0
1956	May	0	1959	March	8
1956	June	0	1959	April	16
1956	July	0	1959	May	2
1956	August	0	1959	June	4
1956	September	0	1959	July	0
1956	October	0	1959	August	0
1956	November	0	1959	September	0
1956	December	2	1959	October	0

Year	Month	Count	Year	Month	Count
1959	November	0	1963	February	0
1959	December	0	1963	March	0
1960	January	0	1963	April	13
1960	February	3	1963	May	0
1960	March	10	1963	June	0
1960	April	5	1963	July	0
1960	May	2	1963	August	0
1960	June	0	1963	September	0
1960	July	0	1963	October	0
1960	August	0	1963	November	0
1960	September	0	1963	December	0
1960	October	0	1964	January	0
1960	November	0	1964	February	0
1960	December	1	1964	March	10
1961	January	0	1964	April	1
1961	February	0	1964	May	0
1961	March	0	1964	June	0
1961	April	10	1964	July	0
1961	May	12	1964	August	0
1961	June	0	1964	September	0
1961	July	0	1964	October	0
1961	August	0	1964	November	0
1961	September	0	1964	December	0
1961	October	0	1965	January	0
1961	November	0	1965	February	0
1962	January	0	1965	March	0
1962	February	0	1965	April	5
1962	March	3	1965	May	0
1962	April	3	1965	June	0
1962	May	0	1965	July	0
1962	June	0	1965	August	0
1962	July	0	1965	September	0
1962	August	0	1965	October	0
1962	September	0	1965	November	7
1962	October	0	1965	December	7
1962	November	0	1966	January	0
1962	December	0	1966	February	9
1963	January	2	1966	March	10

Year	Month	Count	Year	Month	Count
1966	April	19	1969	June	0
1966	May	4	1969	July	2
1966	June	7	1969	August	0
1966	July	0	1969	September	1
1966	August	0	1969	October	0
1966	September	0	1969	November	0
1966	October	0	1969	December	0
1966	November	0	1970	January	0
1966	December	0	1970	February	0
1967	January	0	1970	March	0
1967	February	2	1970	April	1
1967	March	0	1970	May	0
1967	April	1	1970	June	0
1967	May	0	1970	July	0
1967	June	0	1970	August	0
1967	July	0	1970	September	0
1967	August	0	1970	October	0
1967	September	0	1970	November	0
1967	October	0	1970	December	0
1967	November	0	1971	January	0
1967	December	0	1971	February	1
1968	January	0	1971	March	2
1968	February	0	1971	April	17
1968	March	0	1971	May	0
1968	April	6	1971	June	0
1968	May	2	1971	July	0
1968	June	0	1971	August	0
1968	July	0	1971	September	0
1968	August	0	1971	October	0
1968	September	0	1971	November	0
1968	October	0	1971	December	0
1968	November	0	1972	January	0
1968	December	0	1972	February	3
1969	January	0	1972	March	1
1969	February	0	1972	April	8
1969	March	10	1972	May	0
1969	April	11	1972	June	0
1969	May	0	1972	July	0

Year	Month	Count	Year	Month	Count
1972	August	0	1975	October	0
1972	September	0	1975	November	0
1972	October	0	1975	December	0
1972	November	0	1976	January	0
1972	December	0	1976	February	1
1973	January	0	1976	March	2
1973	February	0	1976	April	3
1973	March	0	1976	May	2
1973	April	2	1976	June	0
1973	May	0	1976	July	0
1973	June	0	1976	August	0
1973	July	0	1976	September	0
1973	August	0	1976	October	0
1973	September	0	1976	November	0
1973	October	0	1976	December	3
1973	November	0	1977	January	0
1973	December	0	1977	February	5
1974	January	0	1977	March	10
1974	February	2	1977	April	4
1974	March	0	1977	May	0
1974	April	5	1977	June	0
1974	May	0	1977	July	0
1974	June	0	1977	August	0
1974	July	0	1977	September	0
1974	August	0	1977	October	0
1974	September	0	1977	November	0
1974	October	0	1977	December	0
1974	November	0	1978	January	0
1974	December	0	1978	February	0
1975	January	0	1978	March	0
1975	February	0	1978	April	4
1975	March	3	1978	May	2
1975	April	9	1978	June	1
1975	May	0	1978	July	0
1975	June	0	1978	August	0
1975	July	4	1978	September	0
1975	August	0	1978	October	0
1975	September	0	1978	November	0

Year	Month	Count	Year	Month	Count
1978	December	0	1982	February	0
1979	January	0	1982	March	0
1979	February	1	1982	April	4
1979	March	3	1982	May	9
1979	April	27	1982	June	0
1979	May	5	1982	July	0
1979	June	2	1982	August	0
1979	July	0	1982	September	0
1979	August	0	1982	October	0
1979	September	0	1982	November	0
1979	October	0	1982	December	2
1979	November	0	1983	January	0
1979	December	0	1983	February	0
1980	January	3	1983	March	9
1980	February	0	1983	April	40
1980	March	0	1983	May	12
1980	April	9	1983	June	0
1980	May	5	1983	July	0
1980	June	1	1983	August	0
1980	July	0	1983	September	0
1980	August	0	1983	October	0
1980	September	0	1983	November	0
1980	October	0	1983	December	0
1980	November	0	1984	January	0
1980	December	0	1984	February	0
1981	January	0	1984	March	0
1981	February	0	1984	April	36
1981	March	0	1984	May	1
1981	April	15	1984	June	0
1981	May	10	1984	July	0
1981	June	0	1984	August	0
1981	July	0	1984	September	0
1981	August	0	1984	October	0
1981	September	0	1984	November	2
1981	October	0	1984	December	0
1981	November	0	1985	January	0
1981	December	0	1985	February	0
1982	January	0	1985	March	0

Year	Month	Count	Year	Month	Count
1985	April	4	1988	June	0
1985	May	4	1988	July	0
1985	June	4	1988	August	0
1985	July	0	1988	September	0
1985	August	0	1988	October	0
1985	September	0	1988	November	0
1985	October	0	1988	December	0
1985	November	0	1989	January	0
1985	December	0	1989	February	0
1986	January	0	1989	March	0
1986	February	0	1989	April	3
1986	March	6	1989	May	3
1986	April	4	1989	June	0
1986	May	13	1989	July	0
1986	June	0	1989	August	0
1986	July	0	1989	September	0
1986	August	0	1989	October	0
1986	September	0	1989	November	0
1986	October	0	1989	December	0
1986	November	0	1990	January	0
1986	December	0	1990	February	0
1987	January	0	1990	March	5
1987	February	0	1990	April	5
1987	March	0	1990	May	2
1987	April	0	1990	June	2
1987	May	6	1990	July	0
1987	June	0	1990	August	0
1987	July	0	1990	September	0
1987	August	0	1990	October	0
1987	September	0	1990	November	0
1987	October	0	1990	December	0
1987	November	0	1991	January	0
1987	December	0	1991	February	0
1988	January	2	1991	March	0
1988	February	0	1991	April	0
1988	March	0	1991	May	4
1988	April	9	1991	June	0
1988	May	0	1991	July	0

Year	Month	Count	Year	Month	Count
1991	August	0	1994	October	0
1991	September	0	1994	November	0
1991	October	0	1994	December	0
1991	November	0	1995	January	0
1991	December	0	1995	February	0
1992	January	0	1995	March	5
1992	February	0	1995	April	0
1992	March	0	1995	May	9
1992	April	3	1995	June	0
1992	May	3	1995	July	0
1992	June	0	1995	August	0
1992	July	0	1995	September	0
1992	August	0	1995	October	0
1992	September	0	1995	November	0
1992	October	0	1995	December	0
1992	November	0	1996	January	0
1992	December	0	1996	February	0
1993	January	0	1996	March	0
1993	February	0	1996	April	2
1993	March	3	1996	May	14
1993	April	4	1996	June	0
1993	May	11	1996	July	0
1993	June	0	1996	August	0
1993	July	0	1996	September	0
1993	August	0	1996	October	0
1993	September	0	1996	November	0
1993	October	0	1996	December	0
1993	November	0	1997	January	0
1993	December	0	1997	February	0
1994	January	0	1997	March	0
1994	February	0	1997	April	0
1994	March	0	1997	May	3
1994	April	14	1997	June	0
1994	May	0	1997	July	3
1994	June	0	1997	August	0
1994	July	0	1997	September	0
1994	August	0	1997	October	0
1994	September	0	1997	November	0

Year	Month	Count	Year	Month	Count
1997	December	0	2000	September	0
1998	January	0	2000	October	0
1998	February	0	2000	November	0
1998	March	3	2000	December	2
1998	April	10	2001	January	0
1998	May	2	2001	February	0
1998	June	0	2001	March	5
1998	July	0	2001	April	51
1998	August	0	2001	May	4
1998	September	0	2001	June	0
1998	October	0	2001	July	0
1998	November	0	2001	August	0
1998	December	0	2001	September	0
1999	January	0	2001	October	0
1999	February	0	2001	November	0
1999	March	0	2001	December	0
1999	April	5	2002	January	0
1999	May	4	2002	February	0
1999	June	0	2002	March	7
1999	July	0	2002	April	12
1999	August	0			
1999	September	0			
1999	October	0			
1999	November	0			
1999	December	0			
2000	January	0			
2000	February	0			
2000	March	0			
2000	April	11			
2000	May	4			
2000	June	0			
2000	July	0			
2000	August	0			

VITA

Graduate College
University of Nevada, Las Vegas

Siqi Tan

Degree:

Bachelor of Science in Biological Science and Technology, 2003
Capital Normal University, Beijing

Thesis Title: A Statistical Model for Long-Term Forecasting of Strong Sand Dust Storms

Thesis Examination Committee:

Chairperson, Chih-Hsiang Ho, Ph.D.
Committee Member, Amei Amei, Ph.D.
Committee Member, Kaushik Ghosh, Ph.D.
Graduate Faculty Representative, Sheniz Moonie, Ph.D.