TO: Randy Segawa
Environmental Program Manager I
Environmental Monitoring Branch

FROM: Jing Tao, Ph.D
Environmental Scientist
Environmental Monitoring Branch
916-324-4201

Original signed by

DATE: December 23, 2009

SUBJECT: TIME SERIES ANALYSIS AND FORECASTING OF VENTURA COUNTY NONFUMIGANT PESTICIDE VOLATILE ORGANIC COMPOUND OZONE SEASON EMISSIONS–2009 UPDATE

INTRODUCTION

The Department of Pesticide Regulation (DPR) is required to limit emissions of volatile organic compounds (VOCs) from pesticides in Ventura County during annual May–October ozone season. The maximum allowable annual Ventura County ozone season pesticide VOC emissions (VOC_{MAX}) are defined in regulation (Title 3, California Code of Regulations section 6452.2). VOC_{MAX} consists of two parts, the maximum allowable fumigant emissions (VOC_{FUM}) and projected nonfumigant pesticide emissions (VOC_{NF}). Fumigant use is allocated based on VOC_{FUM} using application method adjustment factors (Barry et al, 2007). To calculate VOC_{FUM} as defined in the regulation, DPR needs to forecast VOC_{NF} in advance of upcoming ozone season.

In a previous analysis of 1990-2007 VOC_{NF}, time series models yielded better predictions than the previous procedure of using VOC_{NF} from two years prior as a forecast for the current year (Frank, 2009). The time series model also accounted for essentially all of the autocorrelation structure in the VOC_{NF} data. Therefore, this method was recommended for annual VOC_{NF} forecasting after recalibrating with additional VOC data every year, which improves estimates of model parameters. This memorandum documents the time series analysis of monthly average VOC_{NF} data in Ventura County from 1990 to 2008, one year more than the pervious data. An updated time series model is developed to forecast 2009 and 2010 VOC_{NF}. The forecast for 2009 will be compared to actual VOC_{NF} after they are calculated in 2010. The forecast for 2010 will be used to calculate VOC_{FUM} for the 2010 ozone season.

EXPLORATORY DATA ANALYSIS

The monthly VOC_{NF} data over 1990–2008 is analyzed by the time series method. The time series plot (Figure 1[a]) shows that the data has seasonal variation over the time. The sample autocorrelation function (ACF) and partial autocorrelation function (PACF) are plotted in figure 1(b-c) and shows that the series is not stationary. The sample ACF slowly decays with the indicated seasonality of 12 (repeated pattern every 12 lags). The sample PACF shows significance at multiple lags (1, 2, 3, 6, 8, 11, and 12). Considering the monthly nature of the data, the period 12 is used to estimate the seasonality.

The time series model for the VOC_{NF} data has the formation:

$$X_t = m_t + s_t + y_t \quad (1)$$

Where X_t is the monthly VOC_{NF} over 1990-2008, m_t is the trend, s_t is the seasonal component with period 12 and mean zero ($\sum_{j=1}^{12} s_t = 0$), y_t is an ARIMA process and t is the year as time index.

The classical decomposition algorithm (CDA) method is used to estimate and remove the trend and the seasonal component $\{s_t\}$ from the data (Brockwell and Davis, 2002). Firstly, the season component is calculated and removed. Secondly, a straight line model is fitted to the deseasonalized series as the trend component $\{m_t\}$ of the series. Previous analysis showed a significant ($p<0.05$) downward trend in the annual VOC_{NF} since 1990 (Spurlock, 2009). The residuals $\{y_t\}$ of the linear regression model are then analyzed and fitted with ARIMA models. The best ARIMA model is chosen based on the lowest value of bias corrected Akaike information criterion (AICC). All the analysis uses statistical software package R. The AICC is defined as (Brockwell and Davis, 2002):

$$AICC = AIC - 2 * npars * (1 - \frac{n}{n-1-npars}) \quad (2)$$

Where n is the number of data, npars is the number of parameters in ARIMA models plus 1. Akaike information criterion (AIC) is given by R as:

$$AIC = -2 \times \log L + k \times edf \quad (3) \text{ (Sakamoto, et al., 1986)}$$

Where L is the likelihood and edf the equivalent degrees of freedom (i.e., the number of free parameters for usual parametric models) of fit.

BUILDING TIME SERIES MODELS

Seasonal Component of the VOC_{NF} Series

The estimated seasonal component $\{s_t\}$ during a year is plotted in Figure 2. The figure shows that VOC_{NF} increases from April to May every year, keeps constant for two months and then decreases in July. After July, the emission increases again and reaches the peak in October. This seasonal variation is consistent with the known high ozone season, May–October.

Trend in the VOC_{NF} Series

Since the data has a monthly pattern, the trend over years is plotted for each month respectively (Figure 3). This plot shows that only months January–April have obviously decreasing VOC_{NF} in recent years. May, July and August had reduced emissions before 2002–2005 but this trend is changed in the following five years. November shows an increasing trend over time while the remaing months do not display a trend.

After removing the seasonal component shown in Figure 2, the linear regression model was fitted to the deseasonalized series $\{m_t\}$:

$$m_t = 699791.6 - 337.1 \times t \quad (4)$$

Where t is the year associated with time index (in the form year + month/12 where month is an integer in the range 0 . . .,11. R^2 of the model is 0.07, which suggest that the regression model only accounts for 7% of the variation in the deseasonalized data. The residuals of this regression model are stable over the years (Figure 4). The Q-Q plot has tails at two ends, especially on the right end, which suggests that the residuals are not very normal. The sample ACF decays with a cosine shape, indicating periodicity still remains in the data after removing the seasonal component. The Box-Pierce test resulted in p-value = 0.034, indicating the residuals are not stationary series, which is also shown as the sample ACF out of the 95% confidence interval ($\pm 1.96/\sqrt{n}$) at lag 1 and 2. Therefore, ARIMA model is needed to fit the residuals after CDA. The sample PACF is out of the confidence interval only at lag 1, which suggests the order of autoregressive in ARIMA model may be 1.

ARIMA Model for the Residuals

The next step is to build ARIMA model to fit the residual data. The notation used to denote a specific seasonal ARIMA model is

$$\text{ARIMA}(p,d,q) \times (P,D,Q)_L$$

where:

p = order of nonseasonal autoregressive component
d = order of nonseasonal differencing
q = order of the nonseasonal moving average process
P = order of seasonal autoregressive component
D = order of seasonal differencing
Q = order of the seasonal moving average process
L = seasonal length

Only lag one in the sample PACF of Figure 4 is out of the significant interval, indicating an autoregressive component of order 1. Several ARIMA model are fitted and compared: AR(1), ARMA (1,1), ARIMA $(1,0,0) \times (0,1,1)_{12}$, ARIMA $(0,0,1) \times (0,1,1)_{12}$, ARIMA $(1,0,1) \times (0,1,1)_{12}$ and ARIMA $(2,0,0) \times (0,1,1)_{12}$. The Box-pierce test obtained large p-value, which indicated that the residuals of all the models obtain stationary (Table 1). The sample ACF and PACF plots also prove this with all the points within the interval (example shown in Figure 5). The last four ARIMA models attain the lower AICC values than AR(1) and ARMA(1,1). The models including nonseasonal moving average component shows less fitness than other models. ARIMA$(1,0,0) \times (0,1,1)_{12}$ will be used as prediction model because of its fewer parameters and the lowest AICC.

Table 1. Summary of Box-pierce test and AICC on ARIMA models.

Model	**Box-pierce test p-value**	**npar**	**AICC**
AR(1)	0.597	2	4,654.1
ARMA(1,1)	0.629	3	4,655.6
ARIMA $(1,0,0) \times (0,1,1)_{12}$	0.738	4	4,451.0
ARIMA $(0,0,1) \times (0,1,1)_{12}$	0.663	4	4,452.2
ARIMA $(1,0,1) \times (0,1,1)_{12}$	0.780	5	4,452.3
ARIMA $(2,0,0) \times (0,1,1)_{12}$	0.809	5	4,451.5

The ARIMA(1,0,0) × $(0,1,1)_{12}$ model is

$$(y_t - y_{t-12}) - \phi(y_{t-1} - y_{t-13}) = \delta + w_t + \theta_{S,1} w_{t-12} \quad (5)$$

Where δ is a constant, $\theta_{S,1}$ is the seasonal moving average coefficient, estimated as -0.807 with standard error 0.056, ϕ is the nonseasonal autoregressive coefficient, estimated as 0.180 with standard error 0.067, and w_t is a Gaussian white noise term assumed $N(0, \sigma_{wt}^2 = 47745581)$.

PREDICTION USING THE MODEL

With the estimates of three components, the time series model X_t for the VOC_{NF} data is built by the combination of the seasonality s_t (Figure 2), the trend m_t (4) and the ARIMA model (5) for y_t as (1). The prediction of VOC_{NF} in 2009 and 2010 using this model is shown in Figure 6. The time series of these two years present the similar pattern with pervious years. The estimates of total VOC emission from nonfumigant in Ventura are 290,143 lbs in 2009 and 287,876 lbs in 2010. The prediction data for ozone season in these two years are listed in Table 2.

Table 2. The prediction of VOC_{NF} monthly emission (lbs) in 2009 and 2010 ozone season.

Monthly Prediction	**Year**	
	2009	**2010**
May	25,804.0	25,587.5
June	30,046.9	29,830.3
July	20,431.8	20,215.1
August	28,764.4	28,547.7
September	37,476.7	37,260.0
October	39,266.7	39050.0
Total	**181,790.5**	**180,490.7**

CONCLUSION

The time series model was developed for the VOC emission of nonfumigant in Ventura from 1990 to 2008. The model is formed by three components: the trend m_t, the seasonal component s_t and ARIMA model. The linear regression model (4) for m_t has R^2 equal to 0.07, which suggests that the decreasing linear trend explains 3% of the time series variation over years. During each year, ozone season (May–October) shows high VOC emissions except for July. The VOC emissions of nonfumigants exhibited different trends in different months over 1990-2008. ARIMA (1,0,0) × $(0,1,1)_{12}$ was chosen for the residuals after deseasonalization and linear regression. The combination of three models was used to predict the VOC_{NF} in Ventura County for the next two years, which is consistent with the series from pervious years. The estimates of

VOC emissions from nonfumigants are 181,790.5 lbs in 2009 ozone season and 180,490.7 lbs in 2010 ozone season. This prediction will be compared to the 2009 VOC data calculated in 2010 and used to regulate VOC_{FUM} of 2010.

REFERENCE

Brockwell, P. and Davis, R. 2002. Introduction to Time Series and Forecasting. 2nd edition. Springer.

R Statistical Software, version 2.9.0, 2009.

Sakamoto, Y., Ishiguro, M., and Kitagawa G. 1986. *Akaike Information Criterion Statistics*. D. Reidel Publishing Company.

Spurlock, F. 2009. Time Series Analysis and Forcasting of Ventura County Nonfumigant Pesticide Volatile Organic Compound Emissions. July 16, 2009 Memorandum to Randy Segawa.

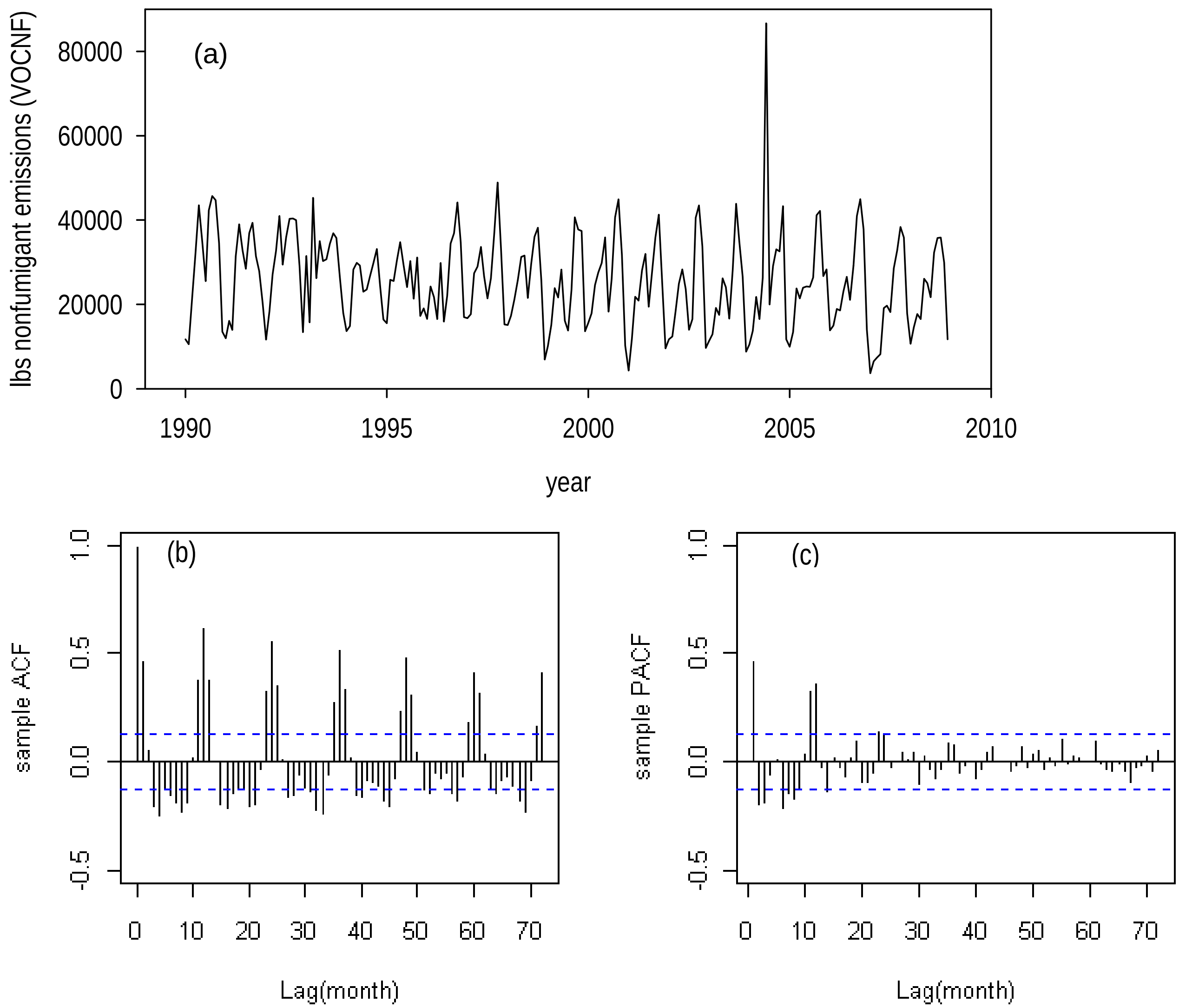


Figure 1. (a) Time series plot of monthly VOC_{NF} over 1990-2008, (b) The sample autocorrelation function (ACF) for the VOC_{NF} data, and. (c) The sample partial autocorrelation function (PACF) for the VOC_{NF} data.

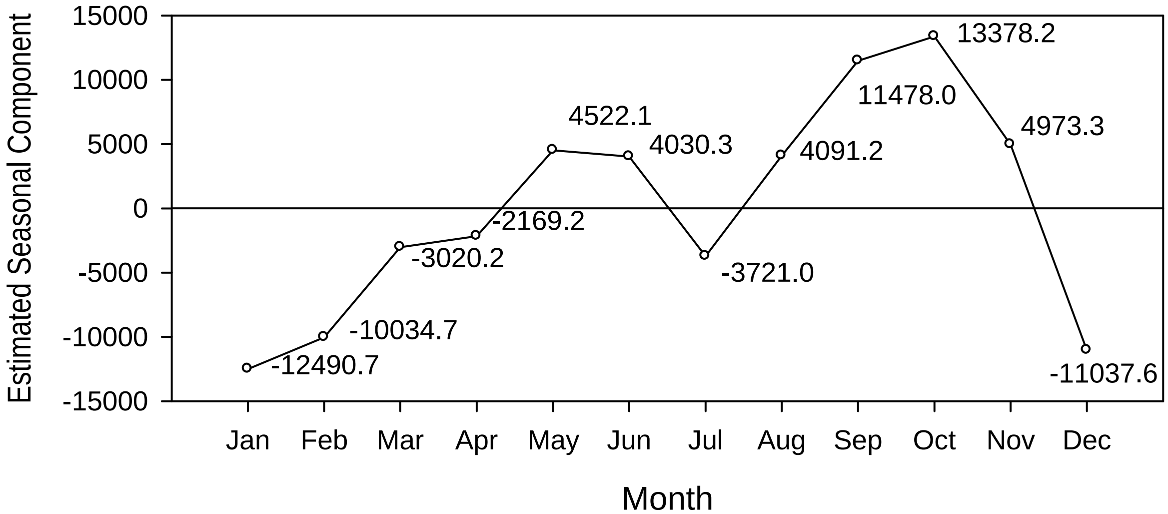


Figure 2. The estimate of seasonal component in the VOC_{NF} series.

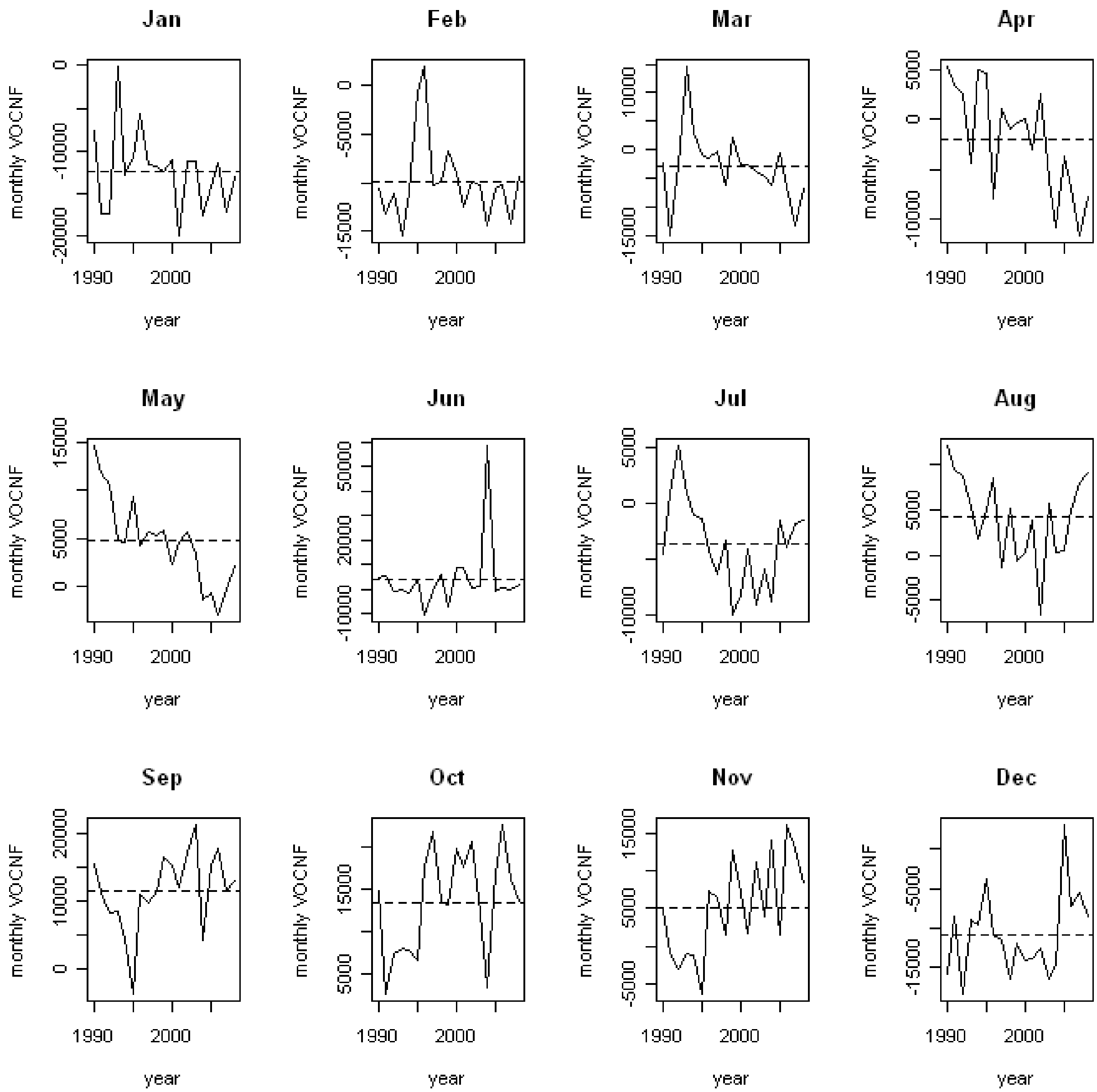


Figure 3. The trend of VOC_{NF} for each month over 1990–2008.

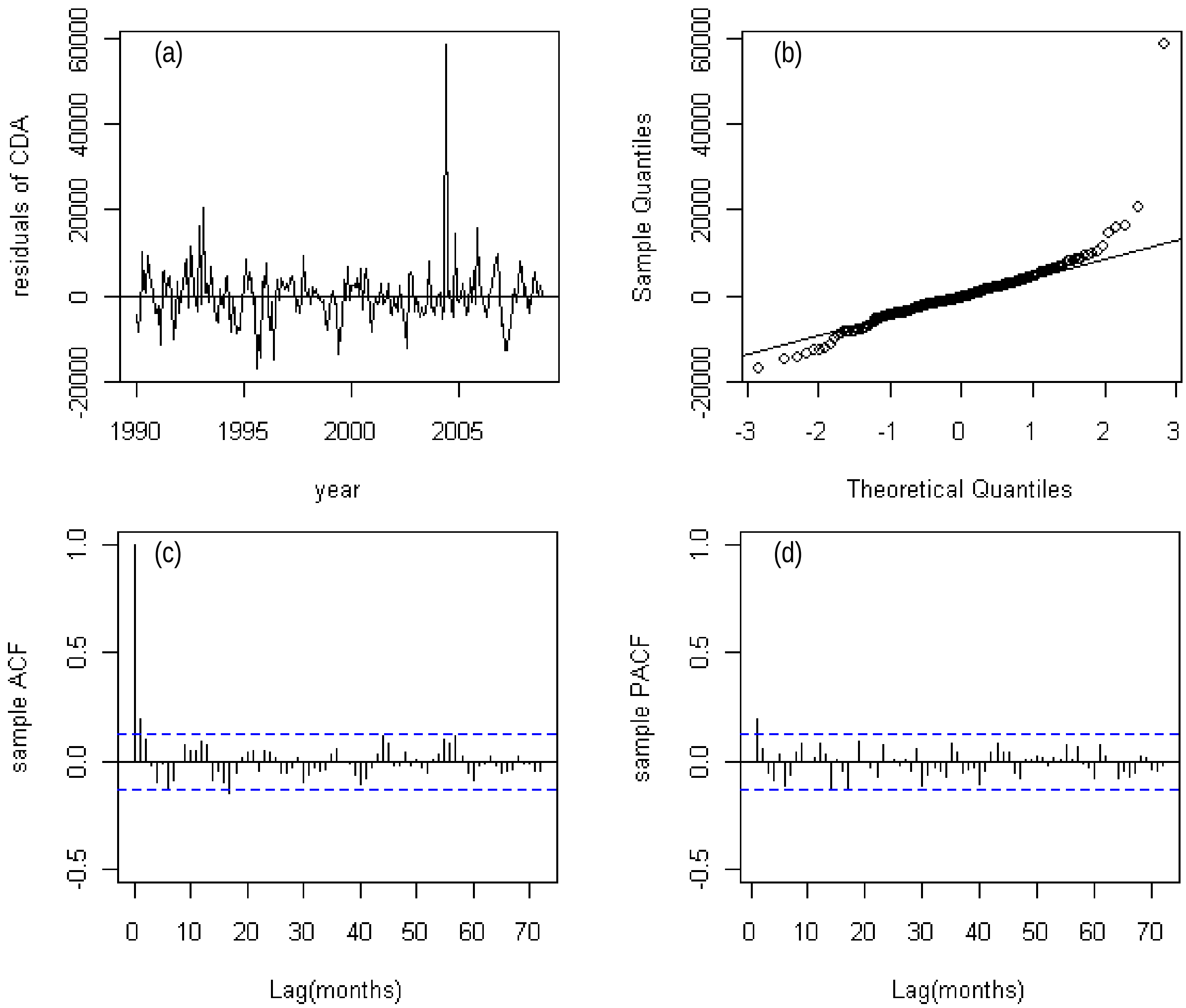
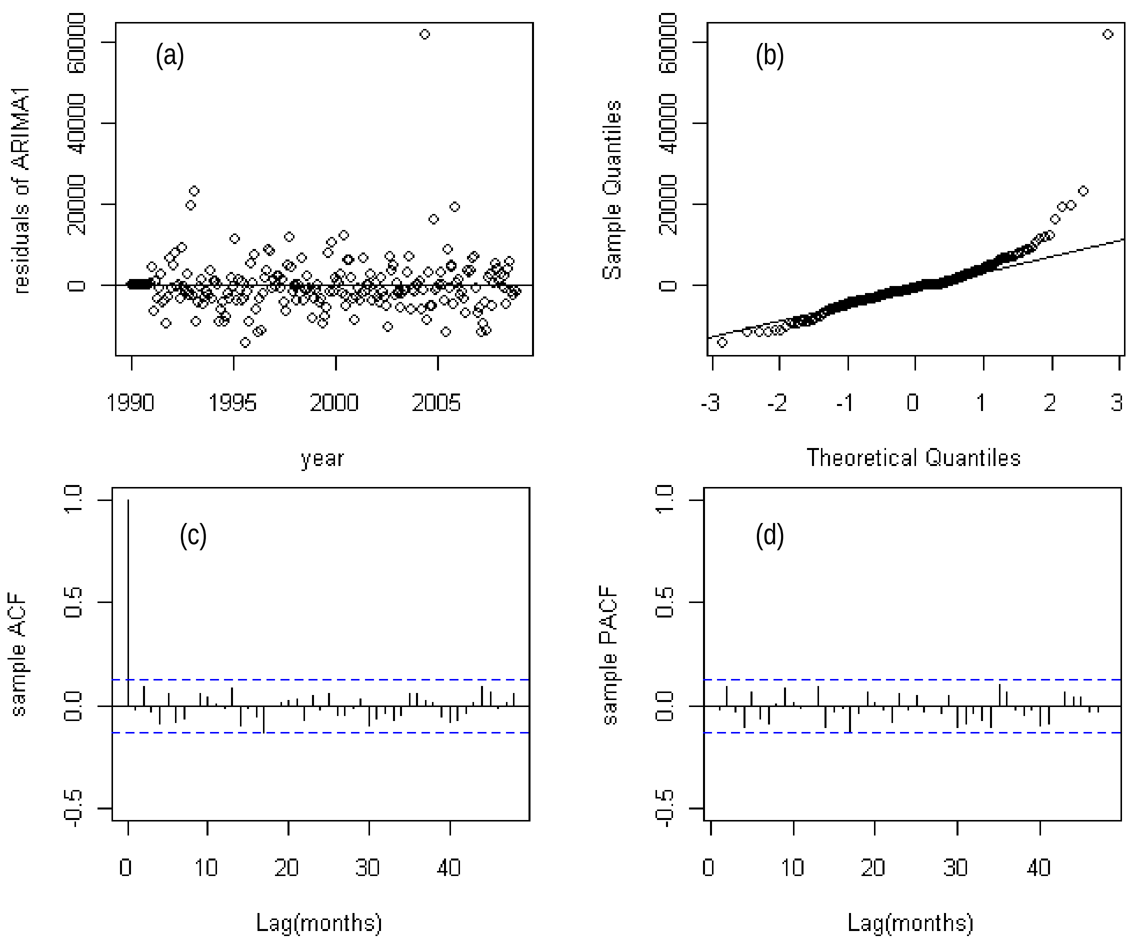


Figure 4. (a)Time series plot, (b) normal Q-Q plot, (c) sample ACF, and (d) sample PACF of the residuals obtained from the seasonal decomposition algorithm (CDA).



Fgiure 5. (a) Time series plot, (b) normal Q-Q plot, (c) sample ACF, and (d) sample PACF of the residuals obtained from ARIMA (1,0,0) × (0,1,1)$_{12}$.

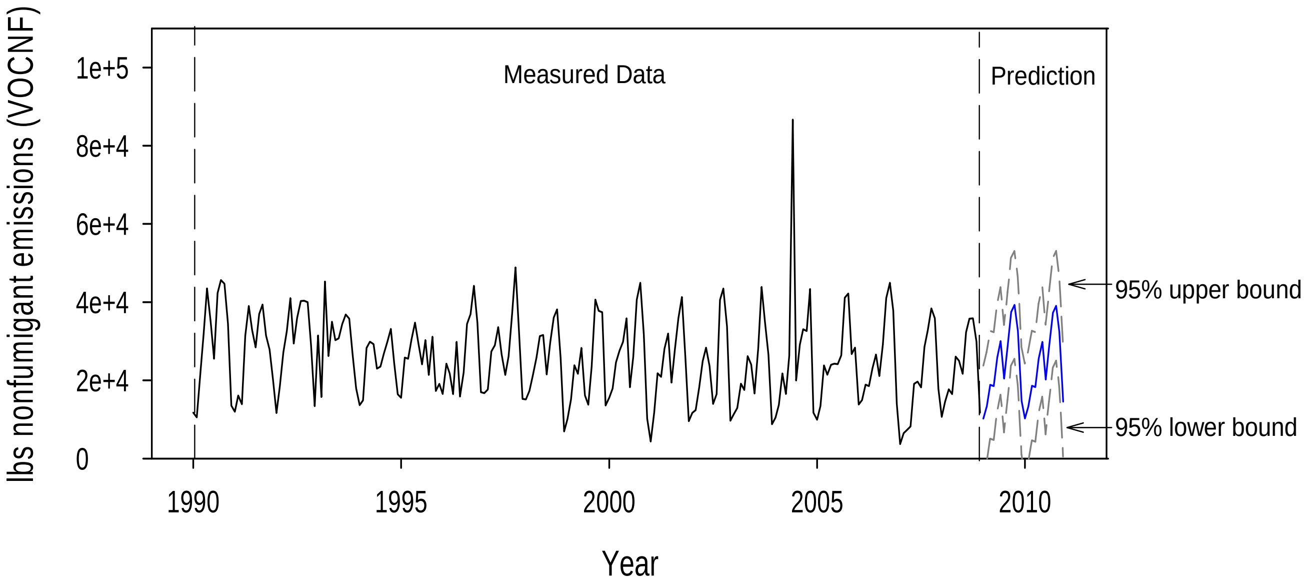


Figure 6. Time series plot of the monthly VOC_{NF} data over 1990 to 2008 and the time series model prediction in 2009 and 2010.