Mary-Ann Warmerdam
Director

Department of Pesticide Regulation

MEMORANDUM

Arnold Schwarzenegger
Governor

TO: Randy Segawa
Environmental Program Manager I
Environmental Monitoring Branch

FROM: Frank Spurlock, Ph.D
Research Scientist III
Environmental Monitoring Branch

Original signed by

DATE: July 16, 2009

SUBJECT: TIME SERIES ANALYSIS AND FORECASTING OF VENTURA COUNTY NONFUMIGANT PESTICIDE VOLATILE ORGANIC COMPOUND EMISSIONS

INTRODUCTION

The Department of Pesticide Regulation (DPR) is required to limit emissions of volatile organic compounds (VOCs) from pesticides in Ventura County during each 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). DPR limits emissions by restricting use of the highest VOC contributing pesticides: fumigants. These are methyl bromide, 1,3-dichloropropene, chloropicrin, metam-sodium, metam-potassium, dazomet, and sodium tetrathiocarbonate. DPR calculates the maximum allowable fumigant emissions (VOC_{FUM}) as the difference between VOC_{MAX} and projected nonfumigant pesticide emissions (VOC_{NF}) during the ozone season.

[1] $$VOC_{FUM} = VOC_{MAX} - VOC_{NF}$$

Fumigant use is then allocated based on VOC_{FUM} using application method adjustment factors as described in Barry et al. (2007). The procedure to calculate VOC_{FUM} as defined in regulation therefore requires DPR to develop an estimate of VOC_{NF} in advance of an upcoming ozone season.

Allowable fumigant emissions for the upcoming May–Oct ozone season are announced in mid-February of that year to allow growers and the Ventura County Agricultural Commissioner time to execute the permit process and plan for the year. However, the prior year's pesticide use report (data that are the basis for calculating historical VOC_{NF} are usually not available until late spring or early summer. Consequently VOC_{NF} forecasts for the upcoming year must be estimated from historical VOC_{NF} data from two years prior because these are the most recent available.

This memorandum documents exploratory analysis of the Ventura County historical VOC_{NF} data and development of a univariate Box-Jenkins, or ARIMA (autoregressive integrated moving average) time series model for forecasting future ozone season VOC_{NF} from historical VOC_{NF}. The model that was selected was one of several that were investigated, and the various modeling approaches are briefly summarized here. The ARIMA model was then compared to DPR's current approach of using VOC_{NF} from two years prior as a forecast for the current year. Mean forecast errors were 6.1 percent and 9.3 percent for the ARIMA and "two years prior" models, respectively, over the years 2000–2007.

More generally, this memorandum outlines a general framework for exploratory time series data analysis of emissions, model development, diagnostic and model validation procedures. These procedures may be adapted for DPR's future VOC_{NF} forecasting needs, such as in the San Joaquin Valley.

EXPLORATORY DATA ANALYSIS

There has been a significant ($p<0.05$) downward trend in historical *annual* VOC_{NF} since 1990 (Figure 1a). However, there are relatively few annual VOC_{NF} data for modeling, and those data possess no significant autocorrelation structure (Figure 2). Consequently the only obvious model appropriate for forecasting future *annual* VOC_{NF} from historical *annual* VOC_{NF} is a simple linear trend model. Such a trend model yields relatively poor estimates of VOC_{NF}, with consistent low bias estimates and high absolute error as measured by 2000–2007 data (Figure 1b). In contrast to the annual data, *monthly* historical VOC_{NF} data demonstrate a consistent periodicity (Figure 3). For example, VOC_{NF} in each year are lowest during winter months and greatest during mid- to late summer months. The presence of both nonseasonal and seasonal autocorrelation is evident from the spikes at lag = 1 and 12 in the autocorrelation function (ACF), the partial autocorrelation function (PACF), and spikes in the periodogram (Figures 4 and 5). Consequently, it may be possible to model the monthly VOC_{NF} using a time series model that accounts for the autocorrelation structure of the data. Monthly forecasts may then be aggregated by year and ozone season to yield more accurate VOC_{NF} forecasts than those obtained from the simple trend model.

A requirement for certain time series models including ARIMA is weak stationarity of the modeled time series $y_1, y_2, \ldots y_t$. A weakly stationary time series is a finite variance series with a mean that is constant over time [$E(y_t)$ = constant, independent of time t] and where the autocovariance function $Cov(y_t, y_{t+k})$ is independent of t and depends only on separation k (Shumway and Stoffer, 2006). Differencing is one of the most effective transformations for inducing time series stationarity (Shumway and Stoffer, 2006), and Bowerman and O'Connell (1987) provide practical diagnostics for evaluating stationarity of differenced time series through analysis of their ACF and PACF. Those diagnostics were used here to evaluate the ability of selected differencing procedures to induce stationarity.

The weak but significant downward trend in the 216 monthly VOC_{NF} data (Figure 3, $p<0.05$) and their persistent autocorrelation indicate nonstationarity (Bowerman and O'Connell, 1987). Several differencing transformations were evaluated to induce stationarity, including first regular differencing, first seasonal differencing, and first regular and first seasonal differencing. An analysis of the differenced series' ACFs and PACFs indicated that first seasonal differencing yielded a stationary series $z_t = y_t - y_{t-12}$.

BUILDING AND FITTING THE ARIMA MODELS

The next step in building the ARIMA model was to winnow down possibilities for the type (autoregressive or moving average) and order (0, 1, 2, . . .) of both the seasonal and nonseasonal components of the model. This consists of comparing certain characteristics of the ACF and PACF of the differenced series to theoretical characteristics of ACF and PACF for different ARIMA models. Operational guidelines for these comparisons are given in Bowerman and O'Connell (1987), and are similar to those described in other texts (e.g. Shumway and Stoffer, 2006).

At the nonseasonal level (lag ≤ 9), the ACF and PACF show identical behavior: both cutoff to very low values after lag = 2 (Figure 6). This is indicative of either a nonseasonal autoregressive component of order 2 or a moving average process of order 2. Bowerman and O'Connell recommend fitting both nonseasonal models and selecting the best fit model. At the seasonal level (lag = ML, L= seasonal length = 12, M= 1, 2, 3, . . .) the ACF cuts off after M=1 while the PACF dies down more gradually (Figure 6). This behavior is indicative of a moving average process of order 1 (Bowerman and O'Connell, 1987).

The notation used to denote a specific seasonal ARIMA model is ARIMA(p,d,q) x $(P,D,Q)_L$ where:

p = order of nonseasonal autogressive component,
d = order of nonseasonal differencing,
q = order of the nonseasonal moving average process,
P = order of seasonal autogressive component,
D = order of seasonal differencing,
Q = order of the seasonal moving average process, and
L = seasonal length.

While the exploratory analysis suggested either an ARIMA(2,0,0)x$(0,1,1)_{12}$ or an ARIMA(0,0,2)x$(0,1,1)_{12}$ model was appropriate for the VOC_{NF} data, several additional ARIMA(p,d,q)x$(P,D,Q)_{12}$ models were evaluated to test that conclusion. Models were compared based on several criteria, including:

- presence of significant nonseasonal or seasonal autocorrelation as shown by the residual ACF/PACF and/or spectral analysis of model residuals, including periodogram, cumulative periodogram and spectrum
- minimum value of the bias-corrected Akaike Information Criterion (AICc)
- statistical signficance of autoregressive and/or moving average model coefficients
- statistical signficance of the Ljung-Box chi-square statistics at seasonal lags L, 2L, 3L, and 4L
- correlation between fitted model values and the monthly 1990 – 2007 VOC_{NF} data

The AICc is calculated as (Shumway and Stoffer, 2006):

[2] $$AICc = \ln(\frac{RSS}{n}) + \frac{n+k}{n-k-2}$$

Where RSS is the model residual sum of squares, n = number of time series data after differencing, and k is the number of parameters in the model. Smaller values of AICc indicate a better model fit. The AICc statistic allows consistent comparison of models with different numbers of parameters and/or orders of differencing. The Ljung-Box chi-square statistic provides a joint test of whether all residuals below a specified lag are random white noise residuals. If the null hypothesis of random residuals is rejected ($p \leq \alpha$), the model is inadequate and should be rejected.

Based on the criteria above, the two best fit ARIMA models were those initially identified in the analysis of ACF and PACF of the differenced series z_t. The two models yielded very similar fits to the data; the ARIMA$(2,0,0)x(0,1,1)_{12}$ yielded an AICc of 18.17, while that of ARIMA$(0,0,2)x(0,1,1)_{12}$ was 18.16. The second nonseasonal autoregressive coefficient in the ARIMA$(2,0,0)x(0,1,1)_{12}$ model fit was not significant at the α= 0.05 level (p=0.07, Figure 7), but this model was retained because of the goodness of fit was nearly identical to the alternate ARIMA$(0,0,2)x(0,1,1)_{12}$ as measured by AICc, and removing the second order coefficient (fit not shown) yielded significant Ljung-Box statistics and a higher AICc. Both ARIMA models were therefore retained for further comparison in a validation exercise.

The residuals for both models were not significantly autocorrelated as indicated by the Ljung-Box statistic (Figures 7 and 8) and the ACFs, PACFs, and cumulative periodograms (Figures 7-10). While both ARIMA models described the temporal fluctuations of monthly VOC_{NF} well, the models did have a tendency to under-predict the absolute value of the annual maxima and minima, respectively (Figures 9b and 10b). Correlations between ARIMA-fitted and observed data were approximately 85 percent for both models. (Figures 9c and10c).

The equation of the multiplicative ARIMA(2,0,0)x(0,1,1)$_{12}$ model for the seasonally differenced series z_t (= y_t - y_{t-12}, y_t = VOC_{NF} in month *t*)

[3] $$z_t = \delta + w_t - \theta_{s,1} w_{t-12} + \phi_1 z_{t-1} + \phi_2 z_{t-2}$$

where δ is a constant, w_t is a Gaussian white noise term assumed $N[0,\sigma_{wt}^2]$, $\theta_{s,1}$ is the seasonal moving average coefficient, and ϕ_1 and ϕ_2 are the nonseasonal autoregressive coefficients.

The forecast equation for the multiplicative ARIMA(0,0,2)x(0,1,1)$_{12}$ model is:

[4] $$z_t = \delta + w_t - \theta_1 w_{t-1} - \theta_2 w_{t-2} - \theta_{s,1} w_{t-12} + \theta_1 \theta_{s,1} w_{t-13} + \theta_2 \theta_{s,1} w_{t-14}$$

where θ_1 and θ_2 are nonseasonal moving average coefficients. Note that the expected values of future w_t are zero, so those terms are ignored in [3] and [4] when forecasting future z_t.

MODEL COMPARISONS

Currently, Ventura County annual ozone season VOC_{NF} from two years prior are taken as the forecast for the current year (Neal et al., 2008). The two ARIMA models were compared to this “two year prior” forecast approach by fitting each model to the monthly VOC_{NF} series over the period *year*=1990 to *year*= *(*forecast year-2). This was performed eight times for each model using 2000 ≤ forecast year ≤ 2007. In each case the resultant fitted model was then used to forecast monthly VOC_{NF} for the forecast year. The May–October ozone season monthly VOC_{NF} forecasts were then summed to obtain the annual VOC_{NF} forecast for that forecast year.

Over the “validation” period 2000–2007, annual forecasts based on the two ARIMA models had a higher correlation with actual VOC_{NF} than the “year-2” forecasts (Figure 11), although none of the correlations were significant. This was not a surprise given the relatively narrow range in VOC_{NF}, the low sample size, and the high volatility of actual VOC_{NF}–especially relative to the ARIMA forecasts (Figure 11). Mean and median percent error were slightly lower for the year-2 forecast model, but this model also yielded the most extreme minimum and maximum percent error (Table 1). In terms of mean absolute percent error, both ARIMA models yielded more accurate forecasts than the year-2 model. However, the ARIMA(0,0,2)x(0,1,1)$_{12}$ provided slightly more accurate forecasts than ARIMA(2,0,0)x(0,1,1)$_{12}$. In addition, the ARIMA(0,0,2)x(0,1,1)$_{12}$ model was more robust than the ARIMA(2,0,0)x(0,1,1)$_{12}$ model in the sense that fits to the smaller “validation” datasets (e.g. 1990–1998, 1990–1999, 1990–2000 . . .) were consistently better based on significance of the fitted coefficients, the AICc and the Ljung-Box statistics.

TABLE 1. Comparison of "two year prior", ARIMA(2,0,0)x$(0,1,1)_{12}$ and ARIMA(0,0,2)x$(0,1,1)_{12}$ model forecasts over years 2000 – 2007. All ARIMA model forecasts for any given *year* are based solely on that respective ARIMA model as fitted to actual 1990 through actual (*year*-2) VOC_{NF} data.

	Ventura County Nonfumigant Ozone Season VOC Emissions (lbs)			
	actual and modeled emissions			
YEAR	actual	2 years prior	ARIMA(2,0,0)x$(0,1,1)_{12}$	ARIMA(0,0,2)x$(0,1,1)_{12}$
2000	**195552**	188227	178803	178332
2001	**184412**	160146	173201	172751
2002	**166441**	195552	173599	172622
2003	**173957**	184412	172004	171296
2004	**163690**	166441	166736	166243
2005	**182001**	173957	163867	163917
2006	**185780**	163690	162063	161888
2007	**156598**	182001	165145	163474
	percent difference = [(actual – modeled)/actual]			
		2 years prior	ARIMA(2,0,0)x$(0,1,1)_{12}$	ARIMA(0,0,2)x$(0,1,1)_{12}$
mean	---	-1.0%	3.4%	3.7%
median	---	1.0%	3.6%	3.9%
min	---	-17.5%	-5.5%	-4.4%
max	---	13.2%	12.8%	12.9%
	absolute percent difference			
		2 years prior	ARIMA(2,0,0)x$(0,1,1)_{12}$	ARIMA(0,0,2)x$(0,1,1)_{12}$
mean	---	9.3%	6.3%	6.1%
median	---	9.0%	5.8%	5.4%
min	---	1.7%	1.1%	1.5%
max	---	17.5%	12.8%	12.9%

OTHER TIME SERIES MODELING APPROACHES

Several other models were investigated for their ability to fit the historical monthly VOC_{NF} data. Based on the model fitting criteria discussed previously, none of these models yielded better fits than the two ARIMA models discussed in detail above. These other models are only generally described here and include:

1. A simple additive trend-seasonal decomposition model (Minitab, 2004).

[5] $$VOC_{NFi} = Tr_i + S_i + \varepsilon_i$$

The model assumes that emissions in any month i ($1 \le i \le 216$ for years 1990 – 2007) are the sum of three components: a simple linear trend, a seasonal component consisting of 12 monthly indices (January–December) that adjust each monthly value up or down from the trend line by a fixed amount, and a “white noise,” or random error term. The resultant model fit the general temporal pattern of seasonal VOC_{NF} fluctuations, but displayed two shortcomings. First, each calendar month’s adjustment was constant from year to year, so the model was unable to account for annual variations in each month’s excursions from the trend line. Secondly, the residuals displayed a high level of seasonal and non-seasonal autocorrelation, indicating the model is not accounting for all the structure in the data. This latter problem was addressed by subsequent modeling of the error term ε_i using ARIMA. This overall two-step modeling procedure did provide a relatively good fit of the data, but the additional complexity was a drawback.

2. Deterministic regression models

Two models were constructed using linear combinations of various trigonometric terms. The first model included the following explanatory variables: t, $\sin(n\pi\ mo/12)$ and $\cos(n\pi\ mo/12)$, where t = decimal time ($1990 \le t \le 2007$), mo = month ($1 \le mo \le 12$), and n is an integer ($1 \le n \le 6$). The second model also included two additional terms: $mo*\sin(n\pi\ mo/12)$, and $mo*\cos(n\pi\ mo/12)$. Stepwise multiple linear regression was initially used to identify significant explanatory variables related to the response variable $VOC_{NFi.}$ One motivation for this approach was based on Figure 5, where spectral analysis of the VOC_{NF} data showed multiple periodic components in the data. Although very different explanatory terms were selected for the two models, both models yielded nearly identical fits. Only the first model is described here. The initial stepwise multiple linear regression identified the following model:

[6] $$VOC_{NFi} = \varepsilon_i + a_1 t + a_2 \sin\left[\frac{2\pi\ mo}{12}\right] + a_3 \cos\left[\frac{2\pi\ mo}{12}\right] + a_4 \sin\left[\frac{3\pi\ mo}{12}\right] + a_5 \sin\left[\frac{4\pi\ mo}{12}\right]$$

This model had similar drawbacks to model 1 discussed above. Subsequent ARIMA modeling of the residuals ε_i did provide a fit of the data that was nearly as good as the ARIMA(0,0,2)x$(0,1,1)_{12}$ and ARIMA(2,0,0)x$(0,1,1)_{12}$ models based on AICc. However, the additional complexity of the two step modeling procedure was a drawback.

3. The Winters-Holt exponential smoothing model using trend, level and seasonal components. (Minitab, 2004).

This model requires selection of weighting factors α, γ, and δ for the trend, level and seasonal components, respectively. The weighting factors are restricted to values between 0 and 1, with values around 0.2 being typical for many applications. The approach taken here was to fit the model several times using different combinations of α, γ, and δ that covered the approximate range of 0.05 – 0.5 in 0.05 increments for each. The “best” combination of weighting factors was chosen based on the minimum mean absolute percent error (MAPE) between model fits and VOC_{NF}. this procedure yielded 0.3, 0.05 and 0.3 as the best fit α, γ, and δ, respectively, with MAPE = 22. However, the correlation between monthly VOC_{NF} and the “best” Winters-Holt model fits was r = 0.80, somewhat lower than the two ARIMA models that were evaluated ($r \cong 0.85$, Figures 7 and 8). In addition, the Winters-Holt exponential smoothing method is generally recommended for short- to medium-term forecasts (Minitab, 2004), whereas the requirements here are for longer range forecasts out to 2 years. Consequently VOC_{NF} forecasting may not be the best application for the Winters-Holt model.

CONCLUSION

Several univariate time series models were evaluated for their ability to describe historical Ventura VOC_{NF} and to forecast ozone season VOC_{NF} out to two years in the future. Based on the 1990–2007 Ventura monthly VOC_{NF}, the model with the best performance was an autoregressive integrated moving average model that included a second order nonseasonal moving average component, a first-order seasonal autoregressive component, and first order seasonal differencing. The adequacy of model performance was shown by

- a correlation of 0.85 between model fits and historical monthly VOC_{NF} data
- a lack of significant autocorrelation in model residuals at both the nonseasonal and seasonal levels based on autocorrelation, partial autocorrelation and spectral analysis
- the lowest value of mean percent error for 2000 – 2007 VOC_{NF} forecasts (6.1 percent)

The ARIMA(0,0,2)x$(0,1,1)_{12}$ forecasts are a clear improvement over our current procedure of using VOC_{NF} from 2 years prior as a forecast for the current year. This model should be used for our next forecast year. It is also apparent that the ARIMA(0,0,2)x$(0,1,1)_{12}$ model accounts for essentially all of the autocorrelation structure in the VOC_{NF} data. This suggests that any further improvement in forecasting accuracy will probably require development of models with additional explanatory variables.

For future forecasting efforts, the ARIMA model will need to be re-fitted to the most recent VOC_{NF} data to insure the best ARIMA parameter estimates are used. More generally, the underlying assumption of all the univariate time series models including the “two years prior”–is that past VOC_{NF} data are a good predictor of future emissions. This assumption has held reasonably well; the trend and seasonal components of the monthly data were relatively consistent over the period of 1990 to 2007. However, my recommendation is that the entire analysis be repeated periodically as the trend and seasonal components may very well change in the future due to regulatory actions, economic conditions, changes in land use and/or changes in cropping patterns.

REFERENCES

Barry, T., F. Spurlock and R. Segawa. 2007. Pesticide Volatile Organic Compound Emission Adjustments For Field Conditions And Estimated Volatile Organic Compound Reductions–Revised Estimates. September 29, 2007 memorandum to J. S. Sanders, Ph.D.

Bowerman, B.L. and R.T. O'Connell. 1987. Time Series Forecasting, Unified Concepts and Computer Implementation, Second Edition. Duxbury Press, Boston Mass.

Diggle, P.J. Time Series, A Biostatistical Introduction. Oxford Clarendon Press, 1990.

Minitab Statistical Software, ver. 14.20, 2004.

Neal, R., F. Spurlock and R. Segawa. 2008. Annual Report on Volatile Organic Chemical Emissions From Pesticides: Emissions for 1990–2007. Draft Report December 19, 2008.

Shumway, R.H. and D.S. Stoffer. 2006. Time Series Analysis and It is Applications, Second Edition. Springer, NY, NY.

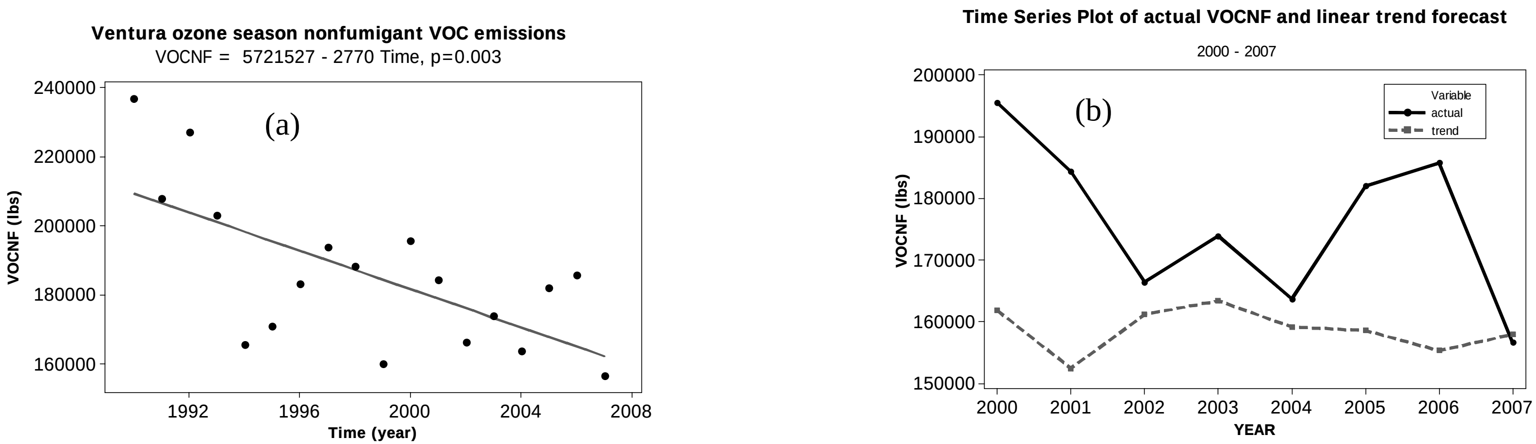


Figure 1. (a) Time series of Ventura ozone season nonfumigant emissions and fitted linear trend ($R^2 = 0.58$). **(b)** 2000 – 2007 actual and trend forecasted VOC_{NF}. Forecast datum for each year is predicted from regression of VOC_{NF} on year for time period of 1990 to year-2. For example, 2001 forecast VOC_{NF} calculated from regression of actual VOC_{NF} on year for $1990 \le year \le 1999$. The mean absolute percent error [(actual-forecast)/actual x 100] is approximately 10 percent over the period 2000 - 2007.

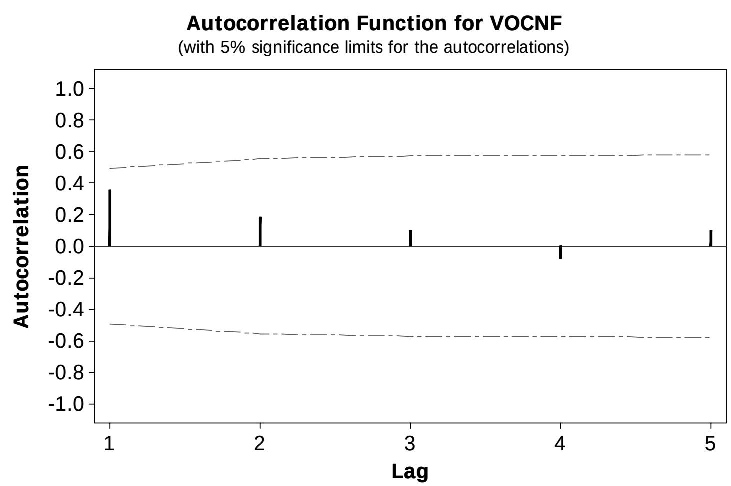


Figure 2. Autocorrelation function for 1990 – 2007 annual VOC_{NF} data. Autocorrelations that do not exceed the dotted lines are not significant.

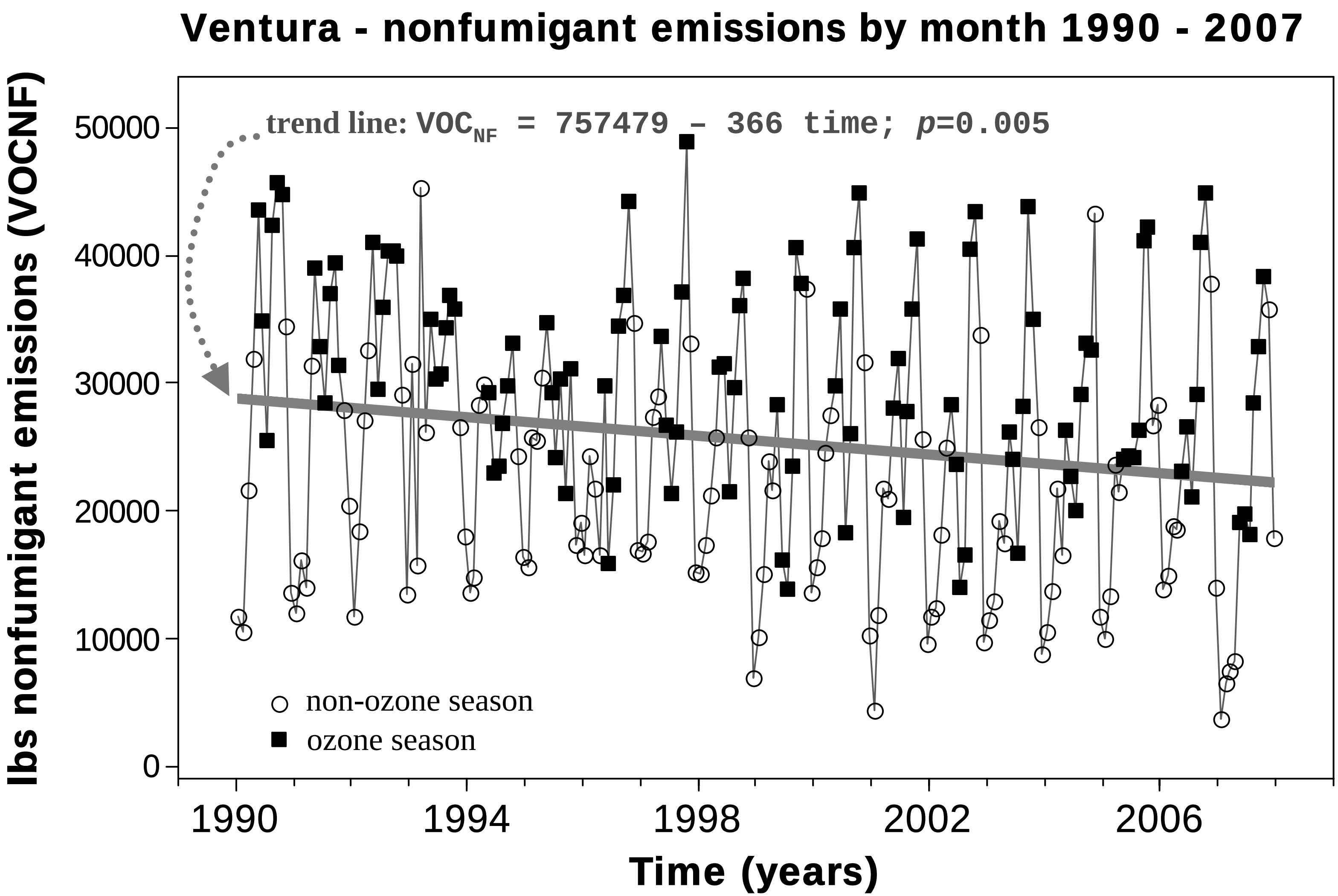
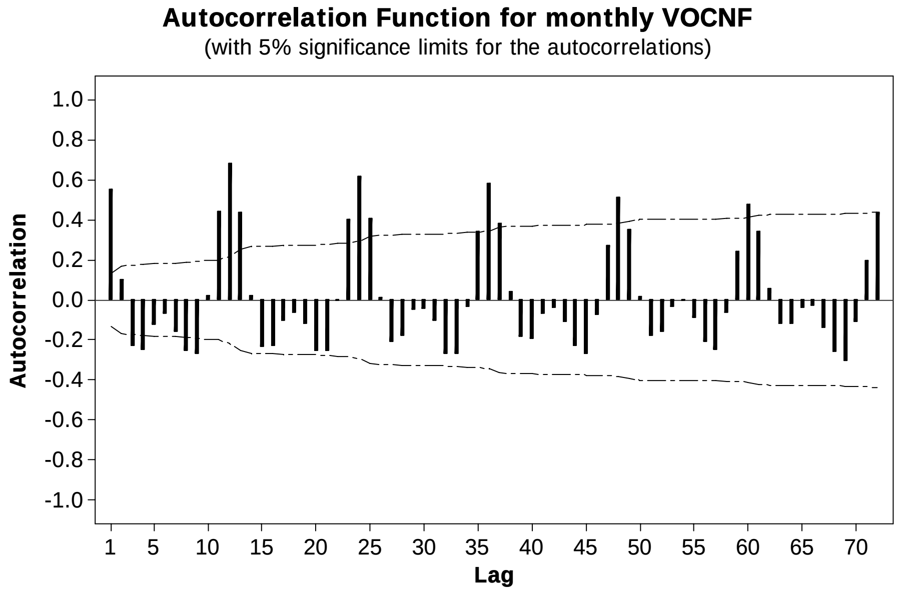


Figure 3. Time series of *monthly* Ventura ozone season nonfumigant emissions and fitted linear trend ($R^2 = 0.04$). Time (x-axis) is in decimal years.



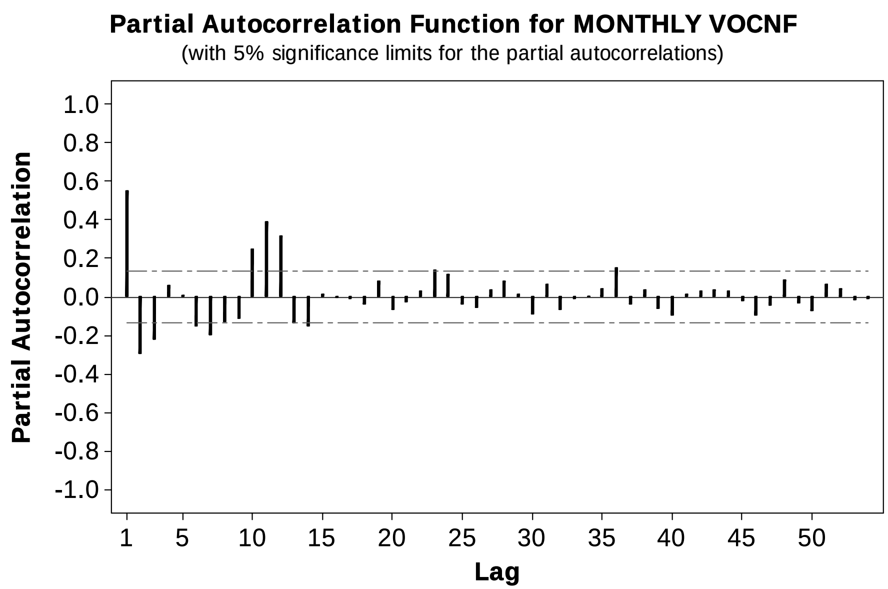


Figure 4. Autocorrelation function and partial autocorrelation function of *monthly* Ventura ozone season nonfumigant emission data. The partial autocorrelation at lag *k* represents the autocorrelation for lag *k* after effects of the intervening autocorrelations (for lags < *k*) are accounted for. Note strong autocorrelation at lag = 1 and also at seasonal lag L=12 months.

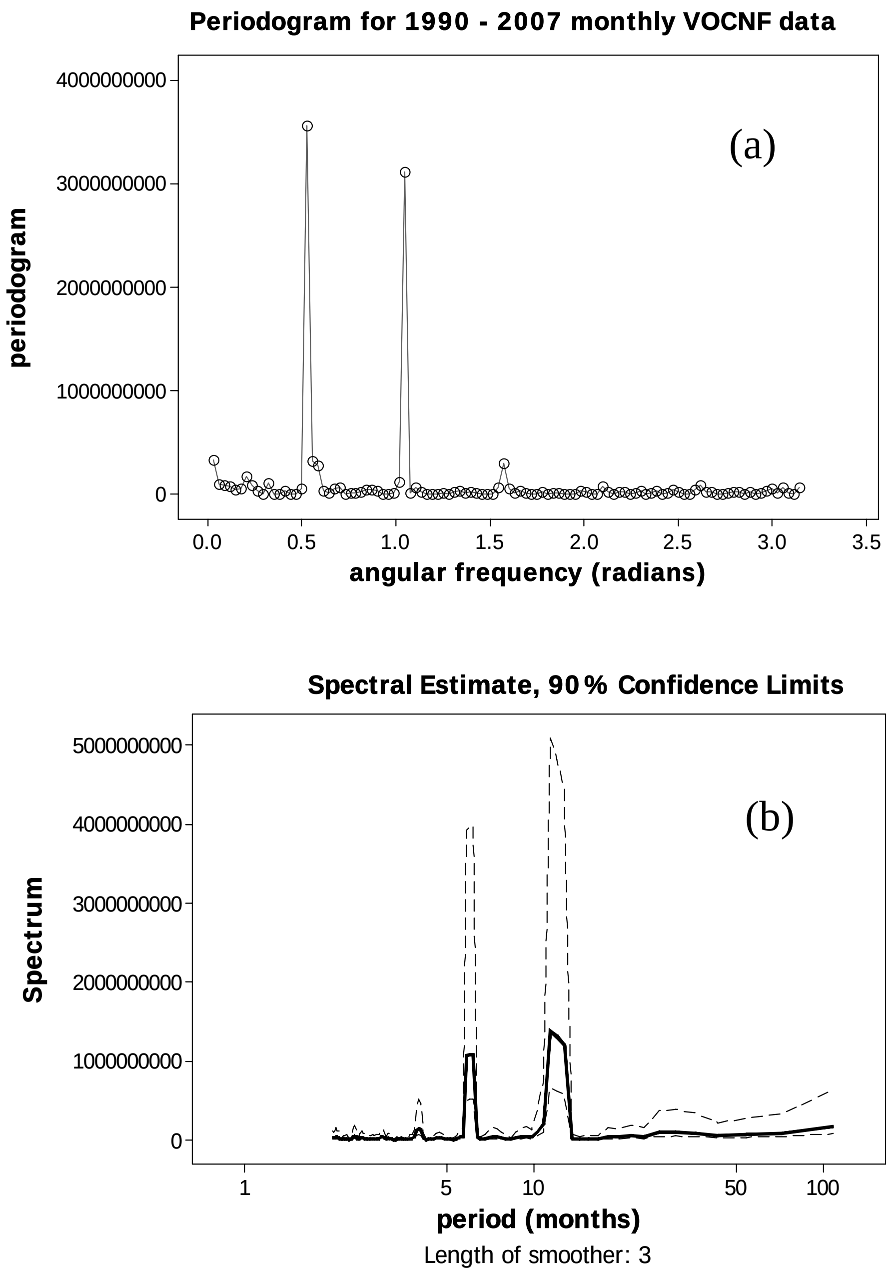
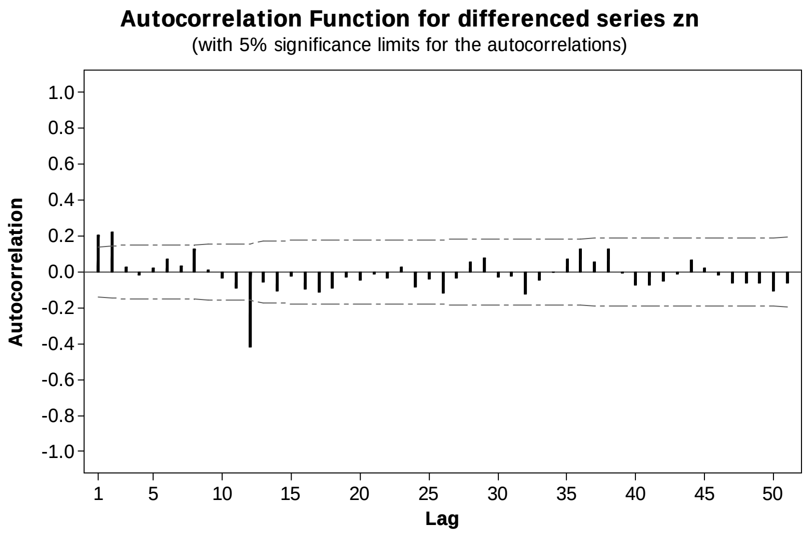


Figure 5. Spectral analysis of VOC_{NF} data. **(a)** periodogram showing high contribution to total VOC_{NF} variance at angular frequencies (ω) of approximately 0.52 and 1.05 radians. These angular frequencies correspond to frequencies f [= $\omega/(2\pi)$] of 0.083 and 0.167 month^{-1}. **(b)** Alternate presentation of data in (a) above: Spectral density function illustrating seasonal structure of data. The plot shows that maximum contributions to VOC_{NF} variance are due to seasonal components with periods of 12 and 6 months. These periods correspond to the frequencies of the observed peaks in (a) above (0.083 and 0.167 month^{-1},respectively). Analysis conducted with Minitab macro “spectral.mac”. Macro and documentation available on-line: *http://www.minitab.com/support/macros/default.aspx?action=all&id=34*



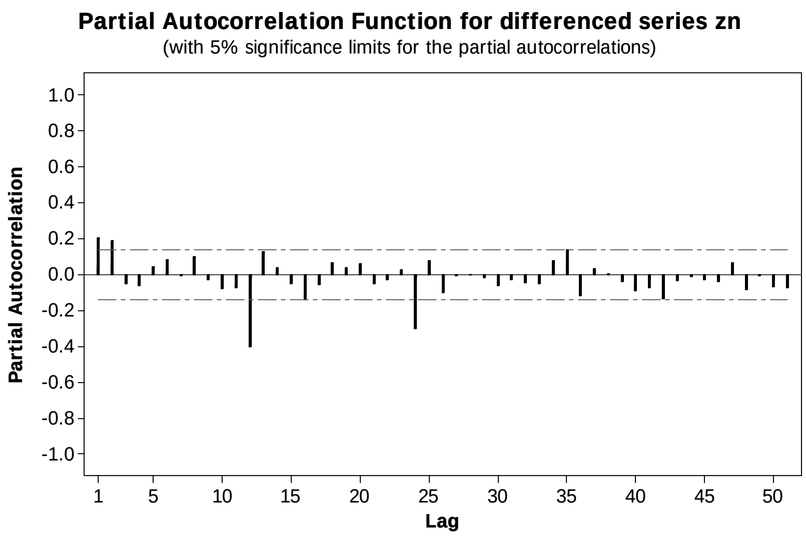


Figure 6. Autocorrelation function and partial autocorrelation function of seasonally differenced working series z_t ($=VOC_{NF,t} - VOC_{NF,\ t-12}$).

Figure 7. `ARIMA (2,0,0)x(0,1,1)`$_{12}$ `Model: VOCNF`

Final Estimates of Parameters

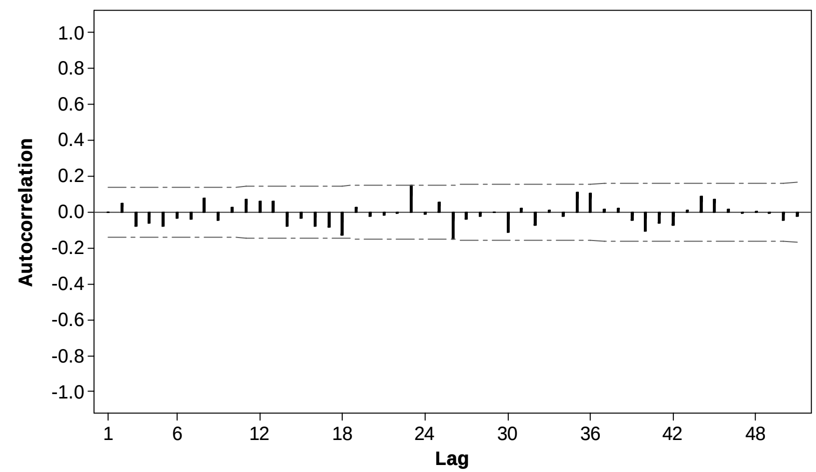
Type		Coef	SE Coef	T	P
AR	1	0.2577	0.0711	3.63	0.000
AR	2	0.1287	0.0703	1.83	0.069
SMA	12	0.8284	0.0458	18.11	0.000
Constant		-244.24	74.31	-3.29	0.001

Differencing: 0 regular, 1 seasonal of order 12
Number of observations: Original series 216,
after differencing 204
Residuals: SS = 5561170716 (backforecasts excluded)
MS = 27805854 DF = 200

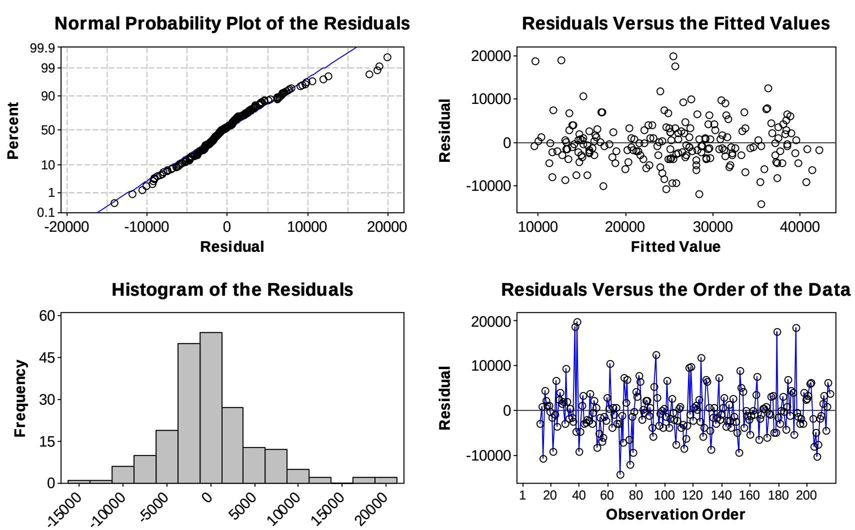
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	8.2	22.7	39.6	49.0
DF	8	20	32	44
P-Value	0.411	0.305	0.166	0.280

ACF of Residuals for ARIMA (2,0,0)x(0,1,1) fit of VOCNF
(with 5% significance limits for the autocorrelations)



Diagnostic Plots for ARIMA (2,0,0)x(0,1,1) Fit of 1990-2007 VOCnf



PACF of Residuals for ARIMA (2,0,0)x(0,1,1) fit to VOCNF
(with 5% significance limits for the partial autocorrelations)

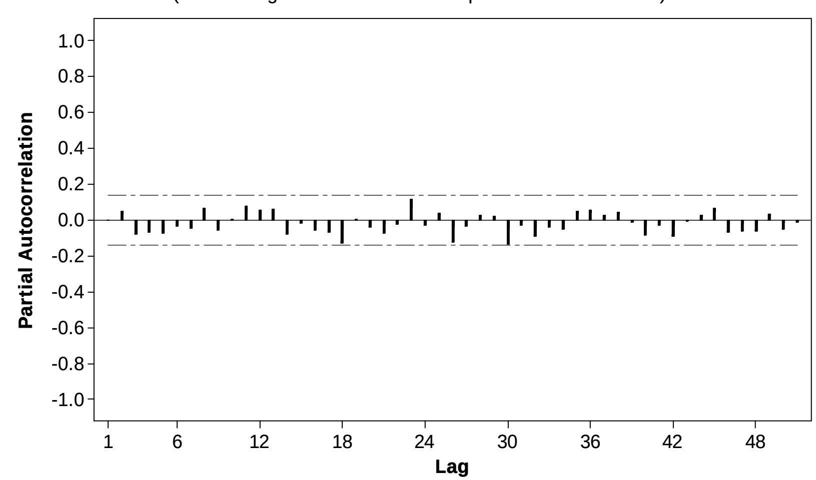


Figure 8. **ARIMA $(0,0,2)x(0,1,1)_{12}$ Model: VOCNF**

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
MA	1	-0.2581	0.0696	-3.71	0.000
MA	2	-0.2292	0.0692	-3.31	0.001
SMA	12	0.8242	0.0456	18.09	0.000
Constant		-405.2	118.2	-3.43	0.001

Differencing: 0 regular, 1 seasonal of order 12

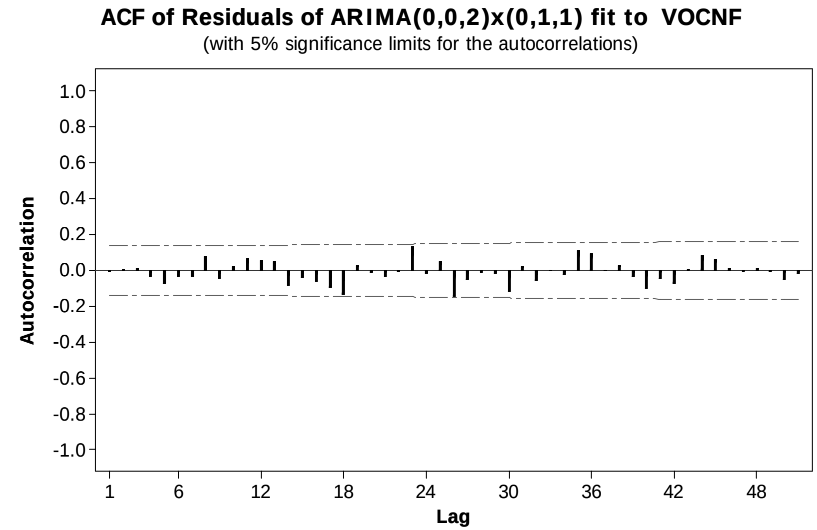
Number of observations: Original series 216, after differencing 204

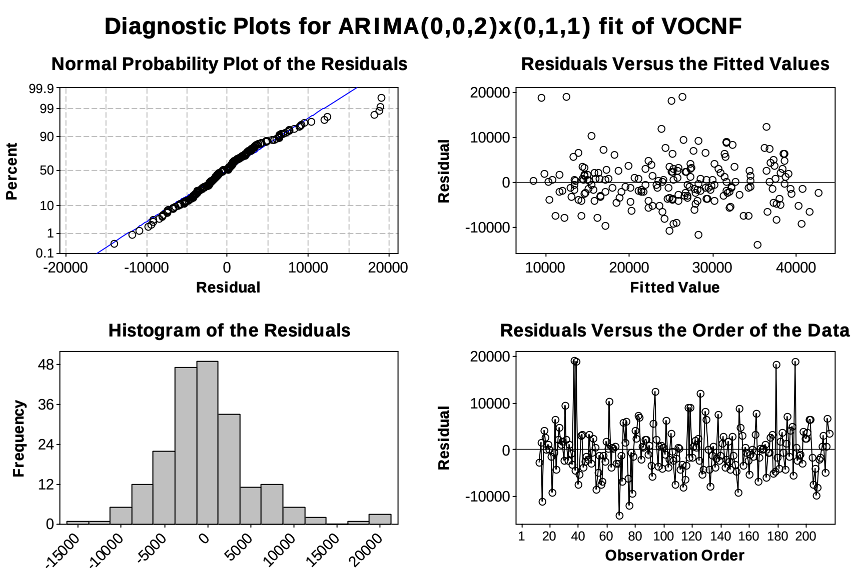
Residuals: SS = 5504544272 (backforecasts excluded)

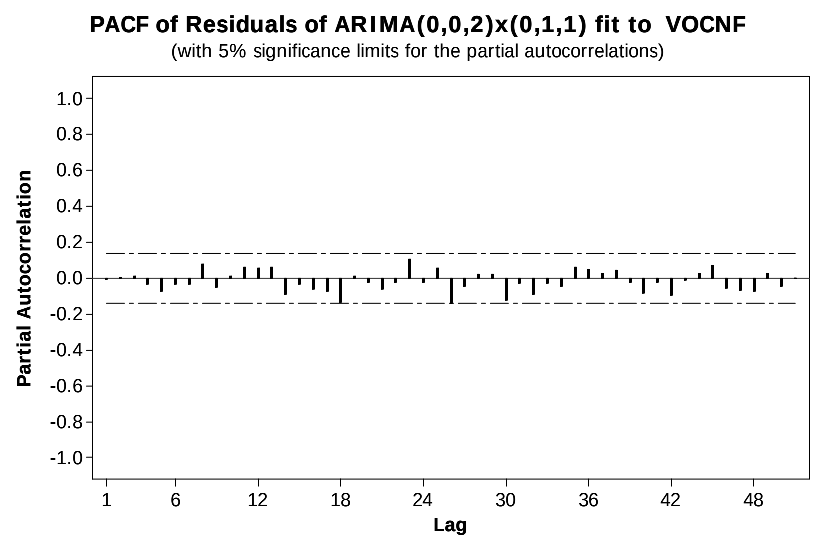
MS = 27522721 DF = 200

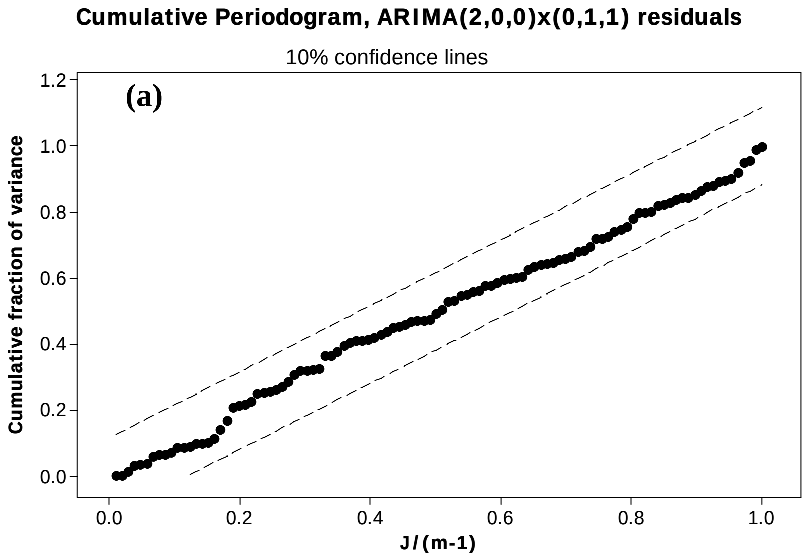
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

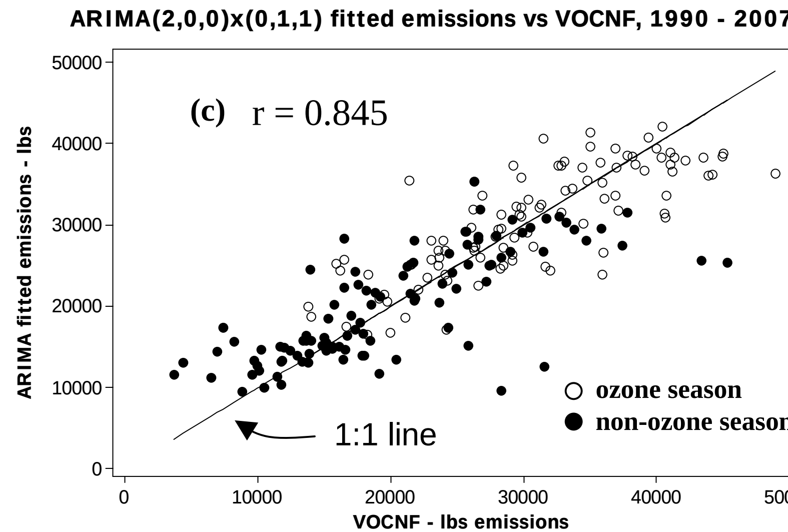
Lag	12	24	36	48
Chi-Square	5.4	19.4	35.5	43.2
DF	8	20	32	44
P-Value	0.709	0.494	0.309	0.506











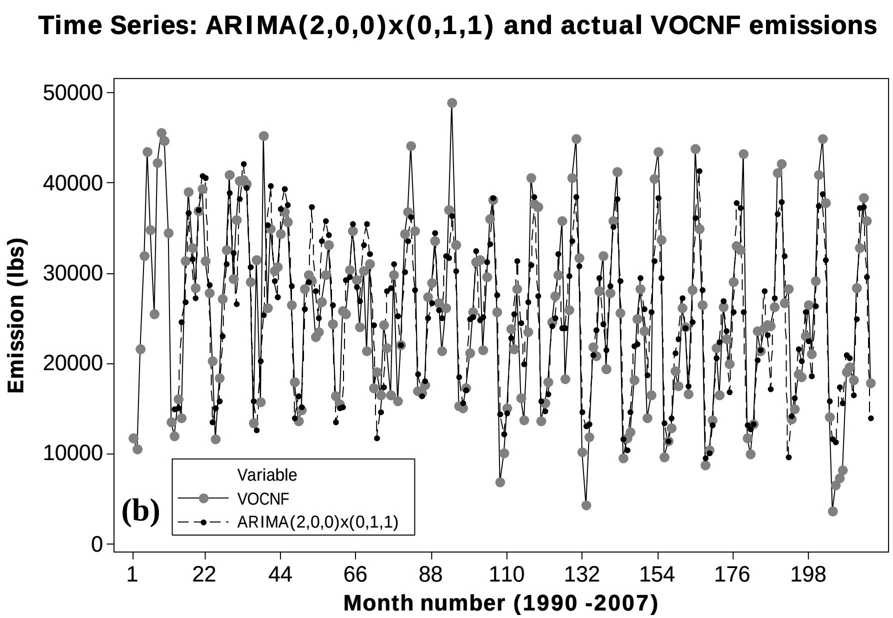
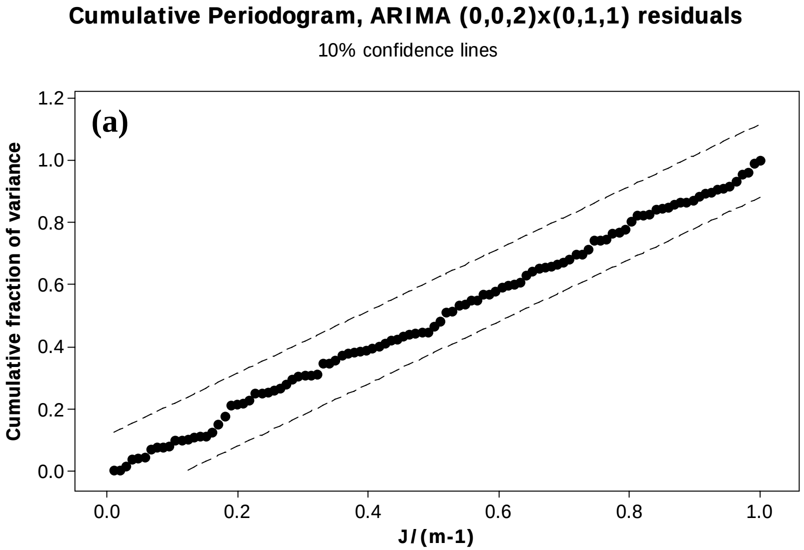
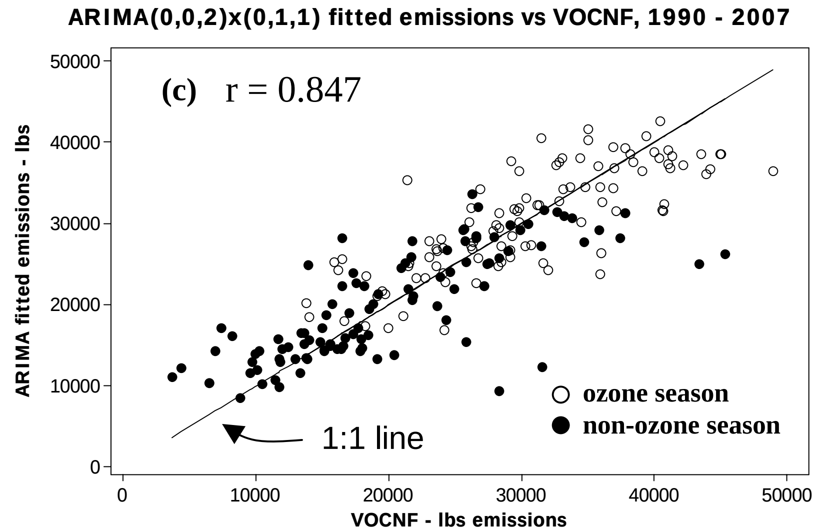


Figure 9. (a) Cumulative periodogram of N= 204 ARIMA (2,0,0)x$(0,1,1)_{12}$ residuals. In the x-axis, J is the harmonic number (e.g. J=1 represents the first harmonic frequency of the residual time series = J/N = 1/204). The number m is the largest integer strictly less than N/2. For a white noise (random) sequence, a plot of the cumulative periodogram vs. J/(m-1) should be approximately linear with slope = 1. The dotted lines are the p= 0.10 critical values of a test statistic for testing the hypothesis H_0: *the series is a white noise sequence* (Diggle, 1990, p. 55). The cumulative periodogram shows no deviation outside the critical lines, indicating insufficient evidence to reject the null hypothesis H_0. **(b)** Time series of ARIMA fitted vs. actual nonfumigant VOC emissions. **(c)** ARIMA fitted vs actual VOC emissions.





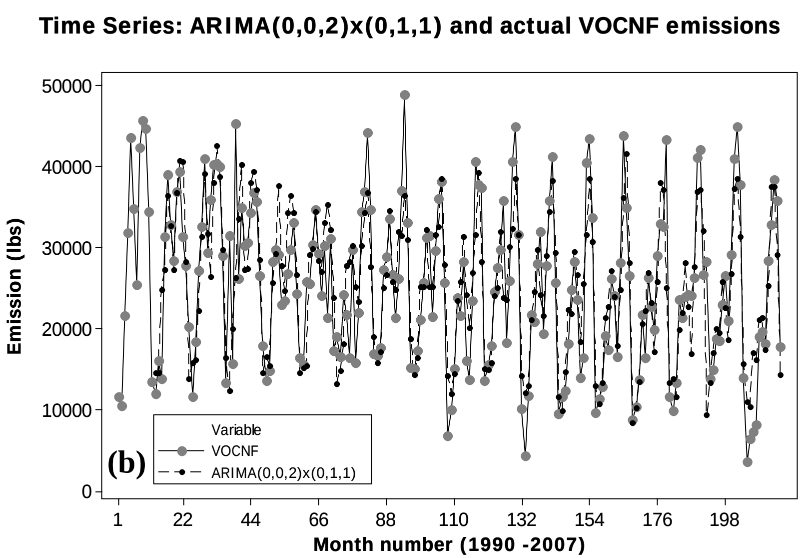
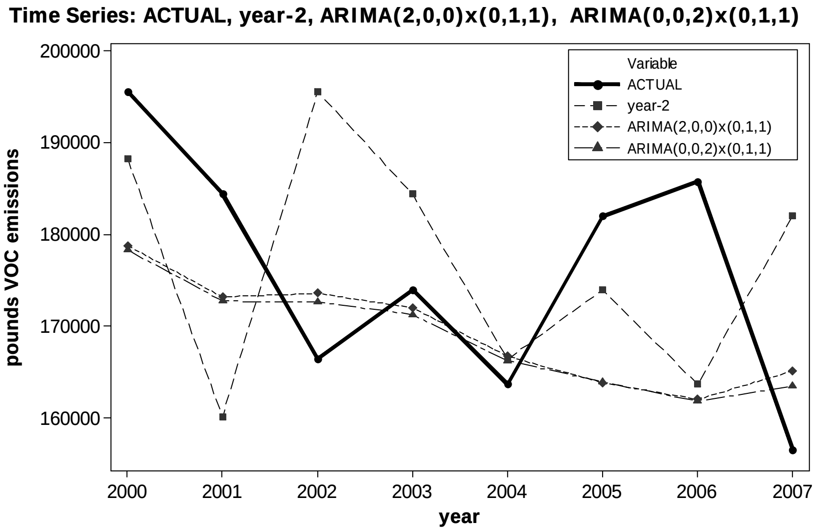


Figure 10. (a) Cumulative periodogram of N= 204 ARIMA $(0,0,2)x(0,1,1)_{12}$ residuals. In the x-axis, J is the harmonic number (e.g. J=1 represents the first harmonic frequency of the residual time series = J/N = 1/204). The number m is the largest integer strictly less than N/2. For a white noise (random) sequence, a plot of the cumulative periodogram vs. J/(m-1) should be approximately linear with slope = 1. The dotted lines are the p= 0.10 critical values of a test statistic for testing the hypothesis H_0: *the series is a white noise sequence* (Diggle, 1990, p. 55). The cumulative periodogram shows no deviation outside the critical lines, indicating insufficient evidence to reject the null hypothesis H_0. **(b)** Time series of ARIMA fitted vs. actual nonfumigant VOC emissions. **(c)** ARIMA fitted vs actual VOC emissions.

Figure 11. Comparison of "YEAR-2", ARIMA(0,0,2)x(0,1,1)$_{12}$ and ARIMA(2,0,0)x(0,1,1)$_{12}$ model forecasts to actual Ventura County ozone season VOC_{NF} data 2000 - 2007. All ARIMA model forecasts for any given *year* are based solely on that respective ARIMA model as fitted to actual 1990 through actual (*year*-2) VOC_{NF} data.



Correlations: ACTUAL, YEAR-2, AR_ARIMA, MA_ARIMA

	ACTUAL	Year-2	ARIMA(2,0,0)x(0,1,1)$_{12}$
Year-2	-0.197		
	0.640		
ARIMA(2,0,0)	0.338	0.502	
x(0,1,1)$_{12}$	0.413	0.205	
ARIMA(0,0,2)	0.406	0.461	0.996
x(0,1,1)$_{12}$	0.318	0.250	0.000

Cell Contents: Pearson correlation

P-Value