EPA Public Access

Author manuscript

Atmos Environ (1994). Author manuscript; available in PMC 2020 January 01.

About author manuscripts

Submit a manuscript

Published in final edited form as:

Atmos Environ (1994). 2019 ; 218: 1–116988. doi:10.1016/j.atmosenv.2019.116988.

Time series analysis of wintertime O₃ and NO_x formation using vector autoregressions

David A. Olson*, Theran P. Riedel, Russell Long, John H. Offenberg, Michael Lewandowski, Tadeusz E. Kleindienst

for submission to Atmospheric Environment Office of Research and Development, United States Environmental Protection Agency, 109 T. W. Alexander Drive, Research Triangle Park, North Carolina 27711, United States

Abstract

Concentrations of 11 species are reported from continuous measurements taken during a wintertime field study in Utah. Time series data for measured species generally displayed strong diurnal patterns. Six species show a diurnal pattern of daytime maximums (NO, NO_y, O₃, H₂O₂, CH₂O₂, and Cl₂), while five species show a diurnal pattern of night time maximums (NO₂, HONO, ClNO₂, HNO₃, and N₂O₅). Vector autoregression analyses were completed to better understand important species influencing the formation of O₃ and NO_x. For the species studied, r^2 values of predicted versus measured concentrations ranged from 0.82–0.99. Fitting parameters for the autoregressive matrix, Π , indicated the importance of species precursors. In addition, values of fitting parameters for Π were relatively insensitive to data size, with variations generally <10%. Variable causation was quantified using the Granger causation method. Assuming O₃ and NO_x behave as chemical products, reactants (in order of importance) are as follows: H₂O₂, N₂O₅, HONO, and ClNO₂.

Keywords

ozone formation; hydrogen peroxide; nitrous acid; dinitrogen pentoxide; nitryl chloride

1 Introduction

Time series analysis has been used in a variety of fields, including business, economics, engineering, and physical sciences (Box et al., 2008). Time series analysis quantifies the dynamic responses of variables over time, thereby exploiting the pattern of serial correlation that is often intrinsic to time-dependent measurements. Box et al. (2008) noted several

*Corresponding author. Tel: +1-919-541-3190; olson.david@epa.gov (D.A. Olson).

Appendix. Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:doi.org/10.1016/j.atmosenv.2019.116988 and this material is available free of charge via the Internet at <http://www.sciencedirect.com>.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

possible motivations for this type of analysis, including forecasting future values and understanding interrelationships among multiple variables.

The formation of ozone (O_3) and nitrogen oxides (NO_x) in the atmosphere, being dependent on time, location, and chemical composition, is one such example involving time series measurements. In this case, the reaction chemistry of both O_3 and NO_x has been well-studied in the primary literature and in review articles (e.g., Atkinson and Arey, 2003; Atkinson, 2000). In addition, numerous studies have reported either experimental or computational results (e.g., Russell and Dennis, 2000; Sillman, 1999). Nonetheless, a number of complications exist in quantifying the formation of O_3 and NO_x from field measurements, such as identifying key intermediate reactions, reconciling the short duration of many field studies with longer-scaled model simulations, and prioritizing chemical intermediates of importance. These experimental and computational challenges are likely exacerbated for the case of wintertime measurements of O_3 and NO_x , where formation conditions can be markedly different from other time periods. Indeed, previous researchers (e.g., Rappenglück et al., 2014) have noted several wintertime conditions that can contribute to elevated chemical concentrations, including lower mixing heights, lower wind speeds, and higher photolysis rates (from high albedo on snow surfaces).

By using time series analysis, our aim is to apply an exclusively data-driven approach to understanding O_3 and NO_x formation. A time series analysis involving multiple variables can be used to identify causal relationships in the data, meaning that subjective decisions on species selection are minimized. Indeed, a major challenge to developing representative chemical markers from field measurements is that of species selection. By removing subjective decisions on species selection, time series analyses should complement existing approaches used to quantify chemical reactions in the atmosphere. In this way, both measurement campaigns and model formulations can focus on species that are most important to O_3 and NO_x formation.

The purpose of this paper is to employ time series analysis to better understand the formation of O_3 and NO_x . Despite numerous studies on the formation of O_3 and NO_x , none have examined their autoregressive properties to quantify the effect of precursors on formation. Existing studies involving time series analysis of O_3 and NO_x have generally used univariate approaches (e.g., Kumar and Jain, 2010; Kumar et al., 2004; Slini et al., 2002; Robeson and Steyn, 1990), meaning that precursors relevant to formation chemistry are excluded from the analysis. Likewise, studies on time series analysis from particulate matter (PM) mass and/or its components (e.g., Jian et al., 2012; Vicente et al., 2012; Diaz-Robles, 2008; Vana et al., 1999) have focused on a single variable approach. For this study, we completed real-time measurements of gas-phase organic and inorganic species during a winter field campaign in Utah, then applied a vector autoregression to examine precursor species that are important to the formation of O_3 and NO_x .

2 Methods

2.1 Site description

Gas-phase measurements were collected at one stationary outdoor location. The sampling site was located on the Utah State University campus in Logan, UT (41°45'31.95"N, 111°48'54.44"W) in the Cache Valley of northern Utah. Instrumentation for the study was located in a temperature-controlled shelter during the sampling period. Sampling dates for the measurements reported herein were from January 20, 2017 through February 15, 2017.

2.2 Analytical methods

Ozone concentrations were measured using a 2B Technologies Model 211 UV photometric analyzer. Concentrations of nitrogen oxides were measured using either a Teledyne API T200U or a Teledyne API T200U NO_y chemiluminescence analyzer. For both ozone and nitrogen oxide measurements, instrument calibrations were completed in accordance with manufacturers' operation manuals and applicable FRM requirements listed in 40 CFR part 50, Appendix F and D (40 Code of Federal Regulations § 50 Appendix F, 2012; 40 Code of Federal Regulations § 50 Appendix D, 2016). Automated zero and span checks were completed nightly to validate calibrations and correct for drift or variations in the span and/or zero response. A NIST-traceable and programmable dynamic dilution calibration system (Teledyne API Model T700U) was used for both calibration gas concentrations and the nightly zero and span gas concentrations. Reference gas standards ($\pm 2\%$ accuracy) and the T700U enclosed ozone generator (with NIST traceable photometer) were used for instrument calibration. NO span gas was titrated with ozone in the gas phase to generate NO₂ concentrations for calibration. Contaminant-free dilution air (zero air via a Teledyne T701H zero air generator) was used for dilution of all test gases. Particulate matter (PM) was removed from the inlets of all trace gas analyzers using particulate filters that were replaced on a weekly basis. The Teledyne API T200U NO_y analyzer was operated without a particle filter to allow sampling of particulate nitrate, a key component of NO_y. Concentrations of NO_x were calculated as the sum of NO and NO₂ concentrations. Concentrations of NO₂ were measured directly from the NO_x analyzer, where NO₂ reacted with heated molybdenum in the analyzer to form NO.

A high-resolution time-of-flight chemical ionization mass spectrometer (CIMS, Aerodyne Research Inc.) using iodide-adduct ion chemistry was used to measure oxidized gas-phase species (Lee et al., 2014). The CIMS co-sampled from the same manifold as the O₃ and NO_x measurements described above and operated in negative ion mode for the entire field study. Iodide ions (I⁻) were generated by passing a dilute mixture of methyl iodide in high-purity nitrogen gas through an α -emitting ²¹⁰Po ionization source (Staticmaster® P-2021, NRD Static Control). Ambient air was drawn directly into the instrument where analyte gases mixed with the reagent gas were detected as molecular clusters with iodide (Kercher et al., 2009). In general, iodide reagent ion chemistry is typically better optimized to detect oxidized and polar chemical species. Nitryl chloride (ClNO₂), molecular chlorine (Cl₂), nitric acid (HNO₃), dinitrogen pentoxide (N₂O₅), formic acid (CH₂O₂), hydrogen peroxide (H₂O₂), and nitrous acid (HONO) mixing ratios were recorded at 1-minute time intervals. The gas-phase species listed here included all species measured by the CIMS that were

above detection limits. Regular calibrations were performed as described in previous publications (Lee et al., 2014; Riedel et al., 2012; Lee et al., 2018). These targeted measurements were then used as variables in the model approach described below. For the model approach described below, 10-minute averages were used to represent the sample measurements ($n=3,765$). Across all species, sample completeness ranged from 98 to 99%.

2.3 Time series analysis and model description

Time series analysis involves quantifying the dynamic response of a variable with changes in time. A key feature of this approach is to use previous values of a variable, also called lagged values, to predict future values. Typically, the most recent lagged values have the strongest influence on the prediction of a future value.

In time series analysis, a single (scalar) variable is often expressed in the form of an autoregression, where future values are predicted based on the weighted sum of a prescribed number of prior (lagged) values. For the more general multivariate case, this variable can depend on both its own prior values as well as the prior values of other variables. A multivariate time series based on p lagged values, called a p^{th} -order vector autoregression, can be written as follows (Hamilton, 1994):

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \varepsilon_t \quad (1)$$

where y_t is an ($n \times 1$) vector of a given realization (at time t) of n variables, c is an ($n \times 1$) vector of constants, Φ_j is an ($n \times n$) matrix of autoregressive coefficients (with j having values from 1 to p), and ε_t is an ($n \times 1$) vector describing noise in the data. The noise vector (ε_t) is typically approximated as a normal distribution, i.e., $\varepsilon_t \approx N(0, \Omega)$ with a covariance-variance matrix described by Ω .

Thus, the vector autoregression in Equation 1 consists of measured values (y 's) and their corresponding fitting parameters (c and Φ_j). These can be written more compactly as:

$$x_t' \equiv [1 \ y_{t-1} \ y_{t-2} \ \dots \ y_{t-p}] \quad (2)$$

and

$$\Pi' \equiv [c \ \Phi_1 \ \Phi_2 \ \dots \ \Phi_p] \quad (3)$$

where x_t is an $[(np + 1) \times 1]$ vector, Π is an $[(np + 1) \times n]$ matrix, and the prime symbol ($'$) refers to a matrix transpose. The autoregressive fitting parameters in Π are typically estimated using maximum likelihood techniques; the maximum likelihood estimate is written as (Hamilton, 1994):

$$\Pi' = [\sum_{t=1}^T y_t x_t'] [\sum_{t=1}^T x_t x_t']^{-1} \quad (4)$$

where T is the number of data points. The covariance-variance matrix (Ω) from the formulation of y_t can also be estimated using maximum likelihood techniques, and its estimate is written as (Hamilton, 1994):

$$\Omega = (1/T) \sum_{t=1}^T (y_t - \Pi' x_t)(y_t - \Pi' x_t)' \quad (5)$$

The terms inside the summation sign from Equation 5 can be viewed as sample residuals, i.e., differences between measured values (y_t) and predicted values (estimated from Π).

Lastly, Hamilton (1994) describes a causality test from time series data that are formulated using vector autoregressions. A test that is valid under asymptotic conditions and for lagged dependent variables is defined as (Hamilton, 1994):

$$S = T (RSS_0 - RSS_1)/RSS_1 \quad (6)$$

where S is the test statistic and RSS_j is the sum of squared residuals for condition j . This test statistic follows a χ^2 distribution, where the critical value of $\chi^2(n)$ is used at a given statistical significance. This test for causality is based on a bivariate case originally developed by Granger (1969) and is commonly referred to as Granger causation. Thus, a variable is said to Granger cause a dependent variable for a decrease in predicted error (at a critical χ^2 value). This decreased error compares the case of using both variables in a vector autoregression versus using only the dependent variable. In other words, a given variable is not just explained through its own lagged values, but also through the lagged values of other variables.

All vector autoregressions used in this paper were based on measured concentrations averaged over 10-minute intervals. In addition, all species concentrations were normalized to range from 0 to 1. We made this adjustment because of the wide range of concentrations among the different species measured in this study. For example, typical concentrations of N_2O_5 are at least three orders of magnitude lower than either O_3 or NO_2 , meaning that N_2O_5 is unlikely to affect resulting autoregressions. Thus, known chemical precursor to O_3 and NO_x formation would be excluded from consideration for analyses that are not normalized. All vector autoregressions and maximum likelihood estimates for model parameters were coded in SAS version 9.4 (SAS Institute Inc., Cary, NC) using the Interactive Matrix Language (PROC IML).

3 Results and Discussion

3.1 Species concentrations

Figure 1 shows concentrations as a function of time for 3 species measured during the study (O_3 , NO_2 , and $ClNO_2$). Summary statistics for all species measured during this study is given in Table S1 in Supporting Information. Other than O_3 and NO_x species, the highest average concentrations were for H_2O_2 , HNO_3 , and CH_2O_2 , respectively (Table S1). As seen in Figure 1, species concentrations consistently show diurnal patterns. For example, O_3 concentrations show an expected pattern of reaching a maximum during the midday hours, while NO_2 reaches a maximum during night time hours. For both species, maximum daily concentrations were typically between 40 and 50 ppbv. Six species show a diurnal pattern of daytime maximums (NO , NO_y , O_3 , H_2O_2 , CH_2O_2 , and Cl_2), while five species show a diurnal pattern of night time maximums (NO_2 , $HONO$, $ClNO_2$, HNO_3 , and N_2O_5).

Table S2 in Supporting Information gives Pearson correlation coefficients combinations of species pairs for the 11 compounds measured in this study. The absolute value in r was less than 0.4 for most species pairs in Table S2. Moreover, only a few species pairs (4 out of 55 possible combinations) had values of $|r| > 0.8$, 3 of which were for different combinations of the NO_x and O_3 species. Species having the lowest r values included Cl_2 and N_2O_5 , where all species pairs involving Cl_2 and N_2O_5 had values of $|r| < 0.2$. Given the apparent low relationships between species pairs, standard statistical tests are unlikely to provide insight into important precursors of O_3 and NO_x formation. These findings, coupled with the clear diurnal patterns in species concentrations, suggest that alternate approaches are needed to identify relationships among compounds. For these reasons, we continue this discussion of O_3 and NO_x formation using vector autoregressions.

3.2 Time series analysis

Figure 2 shows measured and predicted values for both O_3 and H_2O_2 as a function of time. Predicted values, estimated using Equation 1, were based on 2 lagged values for all 11 species measured for this study. Lagged values for measured data were written explicitly for all time steps using Equation 2. Fitting parameters were optimized by maximum likelihood techniques as described in the Methods section. As seen in Figure 2, model predictions closely match measured data across the entire range of concentrations for both O_3 and H_2O_2 . As discussed below for the 10-minute averaging time (Table S3), the relatively high range of r^2 values for the 11 measured species suggests that model predictions were consistent across all species.

A full listing of r^2 values between predicted and measured concentrations is given in Table S3. Across the 11 species studied in this research, r^2 values for the 10-minute averaging time ranged from 0.82–0.99. Except for Cl_2 ($r^2 = 0.82$) and N_2O_5 ($r^2 = 0.87$), r^2 values ranged from 0.93–0.99. The strong correlations between predicted and measured values is an expected result, given that concentrations are predicted only one time step beyond the measured data where lagged differences are likely to be small. Nonetheless, predicted values, using fitting parameters from a single matrix equation (Equation 4), accurately reflect measured concentrations across the range of compounds studied. Using r^2 values as the sole measure of model performance is unlikely to reveal differences in model formulations, considering the narrow range of values and overall high correlation between predicted and measured values. For this reason, the root mean square error between predicted and measured values will be used as a measure of model performance. This statistical diagnostic is expected to show more variability than r^2 values, where lower residuals indicate a better approximation of the measured data.

The number of different model formulations is in this case limited, especially considering that fitting parameters can be described through a single matrix equation (Equation 4). Still, the effect from two such formulations are noteworthy: (1) the number of lags and (2) the number of data points.

First, we consider the number of lags. Figure 3 shows the root mean square error between predicted and measured values plotted as a function of the number of lags for all 11 species considered in this analysis. Across species, O_3 and NO_2 had the highest root mean square

error and were generally at least twice the remaining species. In fact, 4 of the 5 highest species in Figure 3 were for O₃ and NO_x species. The magnitude of species concentrations cannot explain this result because all species were normalized to their maximum value. One possible explanation is that O₃ and NO_x species, being secondary reaction products, are dependent on numerous precursor species and thereby prone to higher errors. While NO_x levels could be the result of primary sources of emissions, the clear pattern of night time maximums in Figure 1 suggests that these species are predominantly nonlocal in origin. Thus, for this analysis we have assumed that NO_x species behave as secondary reaction products. The remaining species in Figure 3, being mostly precursors to O₃ and NO_x formation, show lower deviations from measured values. Each additional lag results in improved model performance, although it should be noted that such improvements also entail added model complexity. Indeed, the only noticeable decrease in the error term comes from increasing the number of lags from 1 to 2. For the remaining lags in Figure 3 (2 to 3 lags, 3 to 4 lags, etc.), decreases in the error term were <3% for all species. Autogressions using 2 lagged values will be used for the remainder of this paper.

Second, we consider the number of data points. Figure 4 shows values of the first lag constant of Φ as a function of the number of data points. As seen in Figure 4, the first lag values of Φ show little variation with data size. The largest of these changes are generally for the smallest data sizes (e.g., from 500 to 1000 data points), while for the remaining data sizes these lag constants typically only show variations of <10%. Thus, a critical limitation of many field studies—that conclusions only apply to the short snapshot of time corresponding with the study duration—shows minimal effect here for the case of vector autoregressions.

Although the primary objective of this paper is to use time series analysis to quantify the effect of precursors on O₃ and NO_x formation, another component of time series analysis involves forecasting future values. This discussion on forecasting then becomes less focused on relationships among variables, and more on the accuracy of model predictions for longer time intervals. Figure 5 shows the root mean square error for the 11 species in this study; errors are given as a function of the time interval between lagged values. As expected, errors increase with increasing time interval between lagged values. Indeed, the typical error for a given species in Figure 5 more than doubles when the 10-min lag spacing increases to 1 hr. Another related result is that r^2 values between predicted and measured concentrations decrease with increasing time intervals (Table S3). For example, the autoregressions described in Figures 2–4 showed robust model fits for O₃ ($r^2 = 0.96$). These model fits steadily degrade as the time intervals increase to 1-hr ($r^2 = 0.76$), 2-hr (0.62), and 3-hr (0.53). Similar results can be seen for the remaining species (Table S3).

There are several possible factors that explain the poorer model fits with increased time interval. Clearly, the single points in time used as model inputs exhibit too much variability to allow for predicted values to be extrapolated to longer time intervals. These measurements could be averaged over longer time intervals, but this approach would likely result in a loss of the detail that describes the interactions among variables. Alternatively, more lagged values could be added to model formulations, resulting in improved predictions. However, this added model complexity comes with a cost of additional fitting parameters, such that

any one fitting parameter is likely to have less interpretive value. The remaining option to reduce the number of fitting parameters is to reduce the number of species, an approach that would similarly result in model formulations with less interpretive value. Both Breiman (2001) and Shmueli (2010) have previously noted that statistical models can at times have competing aims, particularly concerning the explanation versus prediction of data. While time series analysis could be used to forecast future values of O_3 and NO_x , such model formulations in this case would result in significant loss of information related to precursors most important to O_3 and NO_x formation. For this reason, we suggest that forecasting values from O_3 and NO_x formation is the topic of future research.

3.3 Identifying important precursors to O_3 and NO_x using vector autoregressions

In this section, results from vector autoregressions are used to identify important precursors to O_3 and NO_x formation. By using a data-driven approach, our aim is to avoid some of the limitations from a purely deterministic or purely experimental approach. For a purely deterministic model, experimental parameters like photolysis and kinetic rates are typically limited to a narrow range of experimental conditions, conditions that may not be representative of field measurements. For a purely experimental approach, field measurements are likely incomplete both in terms of the duration and range of environmental conditions. Both these approaches are further constrained for the case of wintertime formation, where environmental conditions are likely different from other time periods.

By identifying important variables from resulting autoregressions, we seek a uniform method for analyzing time series measurements, one that is driven both by measured data and by intrinsic species dependencies. Important precursors to O_3 and NO_x formation are considered in two ways: (1) examination of the lag constants, and (2) statistical tests of causation.

First, we consider the behavior of the lag constants from the matrix Π (Table 1). These first lag constants, coupled with the second lag constants (Table S4), give a complete representation of the measured data as discussed above for the time series shown in Figure 2. The first column in Table 1 represents the constant “c” (Equation 3), while the remaining values are the autoregressive coefficients from the matrix Φ . As expected, the strongest predictor of future values as shown in Table 1 is the first lag of that species. Important precursors to O_3 and NO_x compounds (aside from other O_3 and NO_x compounds) can then be identified from the non-diagonal elements in Φ . From the first 4 rows in Table 1, the highest non-diagonal lag constant from species other than O_3 and NO_x compounds is for H_2O_2 ($\Phi_j = 0.44$) followed by N_2O_5 (0.17) and then HONO (0.15). Another noteworthy feature to these fitting parameters is that the first lags are typically paired with a second lag that is similar in magnitude but with the opposite sign (see Table S4), meaning their effect should be more pronounced at higher concentrations. Regardless, the higher values for H_2O_2 , N_2O_5 , and HONO indicate that these compounds are the strongest precursors to O_3 and NO_x compounds.

Second, we consider statistical tests of causation to identify species important to O_3 and NO_x formation. Values of the test statistic S (Equation 6) can be calculated from any 2-

parameter combination of a time series. Higher order combinations of variables are described in Hamilton (1994) and are based on the same pairwise test for Granger causality, although in this case the test would examine combinations of variables which cause another combination of variables. Given the exponential increase in combinations with the number of variables, this approach is intractable for large numbers of variables. For this study, we assumed O_3 and NO_x are products (dependent variables) and then evaluated the remaining species using Equation 6 (see Table S5). For this set of dependent variables (H_2O_2 , HONO, $ClNO_2$, HNO_3 , CH_2O_2 , Cl_2 , and N_2O_5), the highest S value was for H_2O_2 ($S = 62.6$). Four compounds (H_2O_2 , N_2O_5 , HONO, and $ClNO_2$) had values of $S > 5.99$, which for this case is the critical value of $\chi^2(2)$ at a level of 5% significance. In fact, the three highest compounds (H_2O_2 , N_2O_5 , and HONO) had the same order of importance as our earlier discussion on first lag fitting parameters. Conversely, the lowest S values were for HNO_3 and CH_2O_2 , species that are both termination products in the formation of O_3 . Thus, the Granger test of causality can correctly identify both important precursors and termination products.

In short, important precursors to O_3 and NO_x formation, using both the examination of first lag fitting parameters and statistical tests of causation, included H_2O_2 , N_2O_5 , HONO, and $ClNO_2$. While numerous studies have reported concentrations of these species in the atmosphere (e.g., as reviewed in Lee et al., 2000; Reeves and Penkett, 2003; Spataro and Ianniello, 2011), fewer studies have focused on wintertime conditions. For example, several recent studies have reported atmospheric concentrations of H_2O_2 (Watanabe et al., 2016 and 2018; Wang et al. 2016; Liang et al., 2013; Hua et al., 2008), though none focused on wintertime field campaigns. More studies have reported wintertime concentrations of N_2O_5 , HONO, or $ClNO_2$ (McDuffie et al., 2018; Tsai et al., 2018; Rappenglück et al., 2014; Kim et al., 2014; Michoud et al., 2014; Brown et al., 2013; Riedel et al., 2013), though in general these studies reported a small number of species (thus precluding a comparison of OH production among species). Rappenglück et al. (2014) measured O_3 , reactive nitrogen 372 ($NO/NO_2/NO_x$) and several precursors (e.g., HNO_3 , HONO, HCHO) during wintertime in Wyoming, where the majority of OH production was attributed to HONO. Vertically resolved measurements of nitrogen oxides, aerosol composition, soluble gas-phase acids, and halogen species were reported by Brown et al. (2013) during wintertime at a site in Colorado. The authors noted that the radical reservoirs of HONO and $ClNO_2$ were consistently observed at night, and attributed 80% of OH production near the surface to HONO. Additional analyses from the same measurement campaign were completed by Kim et al. (2014) and Riedel et al. (2013). Similar findings were seen for the current study, in addition to the importance of H_2O_2 and N_2O_5 , from a thorough assessment of their autoregressive characteristics. Despite being an oversimplification of the formation chemistry for these precursor species (many of which originate from the decomposition of NO_x compounds), the results reported herein clearly demonstrate that time series autoregressions are an effective method to quantify the importance of atmospherically relevant compounds.

4 Conclusions

Despite numerous experimental and computational studies on O_3 and NO_x , little is known about their formation potential based on an interpretation of autoregressive properties. The

overall objective of this work is to use time series analysis to better understand wintertime formation of O₃ and NO_x. Specifically, a vector autoregression was optimized using maximum likelihood techniques, and important precursors to O₃ and NO_x formation were identified and ranked. Key observations resulting from this study include the following:

- Real-time measurements collected during a wintertime field study in Utah were well-approximated by a vector autoregression. For a vector autoregression using all 11 compounds measured for this study, including O₃, NO_x, H₂O₂, and HONO, r² values ranged from 0.82–0.99 (comparing predicted versus measured values by species).
- Fitting parameters for the autoregressive matrix Π indicated the importance of species precursors. For a 2-lag vector autoregression, the highest values of Π were for the first and second lags of the same species. Of the remaining species combinations (species lags predicting O₃ and NO_x), the highest values of Π were for H₂O₂, N₂O₅ and HONO.
- Values of fitting parameters for Π were relatively insensitive to data size, with variations generally <10%. A major limitation of many field studies, where conclusions are only applicable to the time period of the study, is thus minimized by using a vector autoregression.
- A statistical test of quantifying variable causation for vector autoregressions (Granger causation) was used to identify important precursor species. Assuming O₃ and NO_x behave as dependent variables, reactant species (in order of importance) are as follows: H₂O₂, N₂O₅, HONO, and ClNO₂.

Topics for future research include time series analyses from studies having longer sampling durations, as well as from studies measuring precursors of both gas- and particle-phase compounds.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The U.S. Environmental Protection Agency through its Office of Research and Development funded and collaborated in the research described here under Contract EP-C-15-008 to Jacobs Technology. The manuscript has been subjected to internal review and has been cleared for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

- 40 Code of Federal Regulations § 50 Appendix F, 2012 Measurement principle and calibration procedure for the measurement of nitrogen dioxide in the atmosphere (gas phase chemiluminescence).
- 40 Code of Federal Regulations § 50 Appendix D, 2016 Reference measurement principle and calibration procedure for the measurement of ozone in the atmosphere (chemiluminescence method).
- Atkinson R, 2000 Atmospheric chemistry of VOCs and NO_x. *Atmospheric Environment* 34, 2063–2101.

- Atkinson R, Arey J, 2003 Atmospheric degradation of volatile organic compounds. *Chemical Reviews* 103, 4605–4638. [PubMed: 14664626]
- Box GEP, Jenkins GM, Reinsel GC, 2008 *Time Series Analysis: Forecasting and Control*, 4th edition, John Wiley and Sons, Hoboken, New Jersey.
- Breiman L, 2001 Statistical modeling: The Two Cultures. *Statistical Science* 16, 199–231.
- Brown SS, Thornton JA, Keene WC, Pszenny AAP, Sive BC, Dubé WP, Wagner NL, Young CJ, Riedel TP, Roberts JM, VandenBoer TC, Bahreini R, Öztürk F, Middlebrook AM, Kim S, Hübner G, Wolfe DE, 2013 Nitrogen, Aerosol Composition, and Halogens on a Tall Tower (NACHTT): Overview of a wintertime air chemistry field study in the front range urban corridor of Colorado. *Journal of Geophysical Research: Atmospheres*, 118, 8067–8085.
- Diaz-Robles LA, Ortega JC, Fu JS, Reed GD, Chow JC, Watson JG, Moncada-Herrera JA, 2008 A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment* 42, 8331–8340.
- Granger CWJ, 1969 Investigating casual relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.
- Hamilton JD, 1994 *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.
- Hua W, Chen ZM, Jie CY, Kondo Y, Hofzumahaus A, Takegawa N, Chang CC, Lu KD, Miyazaki Y, Kita K, Wang HL, Zhang YH, Hu M, 2008 Atmospheric hydrogen peroxide and organic hydroperoxides during PRIDE-PRD'06, China: Their concentration, formation mechanism and contribution to secondary aerosols. *Atmospheric Chemistry and Physics* 8, 6755–6773.
- Jian L, Zhao Y, Zhu YP, Zhang MB, Bertolatti D, 2012 An application of ARIMA model to predict submicron particle concentrations from meteorological factors at a busy roadside in Hangzhou, China. *Science of the Total Environment* 426, 336–345. [PubMed: 22522077]
- Kercher JP, Riedel TP, Thornton JA, 2009 Chlorine activation by N_2O_5 : Simultaneous, in situ detection of ClNO_2 and N_2O_5 by chemical ionization mass spectrometry. *Atmospheric Measurement Techniques* 2(1), 193–204.
- Kim S, VandenBoer TC, Young CJ, Riedel TP, Thornton JA, Swarthout B, Sive B, Lerner B, Gilman JB, Warneke C, Roberts JM, Guenther A, Wagner NL, Dubé WP, Williams E, Brown SS, 2014 The primary and recycling sources of OH during the NACHTT-2011 campaign: HONO as an important OH primary source in the wintertime. *Journal of Geophysical Research: Atmospheres* 119, 6886–6896.
- Kumar U, Jain VK, 2010 ARIMA forecasting of ambient air pollutants (O_3 , NO, NO_2 and CO). *Stochastic Environmental Research and Risk Assessment* 24, 751–760.
- Kumar K, Yadav AK, Singh MP, Hassan H, Jain VK, 2004 Forecasting daily maximum surface ozone concentrations in Brunei Darussalam—An ARIMA modeling approach. *Journal of the Air and Waste Management Association* 54, 809–814. [PubMed: 15303293]
- Lee M, Heikes BG, O'Sullivan DW, 2000 Hydrogen peroxide and organic hydroperoxide in the troposphere: a review. *Atmospheric Environment* 34, 3475–3494.
- Lee BH, Lopez-Hilfiker FD, Mohr C, Kurtén T, Worsnop DR, Thornton JA, 2014 An iodide-adduct high-resolution time-of-flight chemical-ionization mass spectrometer: Application to atmospheric inorganic and organic compounds. *Environmental Science and Technology* 48, 6309–6317. [PubMed: 24800638]
- Lee BH, Lopez-Hilfiker FD, Veres PR, McDuffie EE, Fibiger DL, Sparks TL, Ebben CJ, Green JR, Schroder JC, Campuzano-Jost P, Iyer S, D'Ambro EL, Schobesberger S, Brown SS, Wooldridge PJ, Cohen RC, Fiddler MN, Bililign S, Jimenez JL, Kurtén T, Weinheimer AJ, Jaegle L, Thornton JA, 2018 Flight deployment of a high-resolution time-of-flight chemical ionization mass spectrometer: Observations of reactive halogen and nitrogen oxide species. *Journal of Geophysical Research: Atmospheres* 123, 7670–7686.
- Liang H, Chen ZM, Huang D, Zhao Y, Li ZY, 2013 Impacts of aerosols on the chemistry of atmospheric trace gases: A case study of peroxides and HO_2 radicals. *Atmospheric Chemistry and Physics* 13, 11259–11276.
- McDuffie EE, Fibiger DL, Dubé WP, Lopez-Hilfiker F, Lee BH, Thornton JA, Shah V, Jaegle L, Guo H, Weber RJ, Reeves JM, Weinheimer AJ, Schroder JC, Campuzano-Jost P, Jimenez JL, Dibb JE, Veres P, Ebben C, Sparks TL, Wooldridge PJ, Cohen RC, Hornbrook RS, Apel EC, Campos T,

- Hall SR, Ullmann K, Brown SS, 2018 Heterogeneous N_2O_5 uptake during winter: Aircraft measurements during the 2015 WINTER campaign and critical evaluation of current parameterizations. *Journal of Geophysical Research: Atmospheres* 123, 4345–4372.
- Michoud V, Colomb A, Borbon A, Miet K, Beekmann M, Camredon M, Aumont B, Perrier S, Zapf P, Siour G, Ait-Helal W, Afif C, Kukui A, Furger M, Dupont JC, Haeffelin M, Doussin JF, 2014 Study of the unknown HONO daytime source at a European suburban site during the MEGAPOLI summer and winter field campaigns. *Atmospheric Chemistry and Physics* 14, 2805–2822.
- Rappenglück B, Ackermann L, Alvarez S, Golovko J, Buhr M, Field RA, Soltis J, Montague DC, Hauze B, Adamson S, Risch D, Wilkerson G, Bush D, Stoeckenius T, Keslar C, 2014 Strong wintertime ozone events in the Upper Green River basin, Wyoming. *Atmospheric Chemistry and Physics* 14, 4909–4934.
- Reeves CE, Penkett SA, 2003 Measurements of peroxides and what they tell us. *Chemical Reviews* 103, 5199–5218. [PubMed: 14664648]
- Riedel TP, Bertram TH, Crisp TA, Williams EJ, Lerner BM, Vlasenko A, Li S-M, Gilman J, de Gouw J, Bon DM, Wagner NL, Brown SS, Thornton JA, 2012 Nitryl chloride and molecular chlorine in the coastal marine boundary layer. *Environmental Science & Technology* 46 (19), 10463–10470. [PubMed: 22443276]
- Riedel TP, Wagner NL, Dubé WP, Middlebrook AM, Young CJ, Öztürk F, Bahreini R, VandenBoer TC, Wolfe DE, Williams EJ, Roberts JM, Brown SS, Thornton JA, 2013 Chlorine activation within urban or power plant plumes: Vertically resolved ClNO_2 and Cl_2 measurements from a tall tower in a polluted continental setting. *Journal of Geophysical Research: Atmospheres* 118, 8702–8715.
- Robeson SM, Steyn DG, 1990 Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. *Atmospheric Environment* 24B, 303–312.
- Russell A, Dennis R, 2000 NARSTO critical review of photochemical models and modeling. *Atmospheric Environment* 34, 2283–2324.
- Shmueli G, 2010 To explain or predict? *Statistical Science* 25, 289–310.
- Sillman S, 1999 The relation between ozone, NO_x and hydrocarbons in urban and polluted rural environments. *Atmospheric Environment* 33, 1821–1845.
- Slini T, Karatzas K, Moussiopoulos N, 2002 Statistical analysis of environmental data as the basis of forecasting: an air quality application. *The Science of the Total Environment* 288, 227–237. [PubMed: 11991526]
- Spataro F, Ianniello A, 2014 Sources of atmospheric nitrous acid: State of the science, current research needs, and future prospects. *Journal of the Air & Waste Management Association* 64, 1232–1250. [PubMed: 25509545]
- Tsai C, Spolaor M, Colosimo SF, Pikelnaya O, Cheung R, Williams E, Gilman JB, Lerner BM, Zamora RJ, Warneke C, Roberts JM, Ahmadov R, de Gouw J, Bates T, Quinn PK, Stutz J, 2018 Nitrous acid formation in a snow-free wintertime polluted rural area. *Atmospheric Chemistry and Physics* 18, 1977–1996.
- Vana M, Tamm E, Viil M, 1999 Experimental study of the air pollution transport by synchronised monitoring of atmospheric aerosols. *Atmospheric Environment* 33, 4615–4628.
- Vicente AB, Jordan MM, Sanfeliu T, Sanchez A, Esteban MD, 2012 Air pollution prediction models of particles, As, Cd, Ni and Pb in a highly industrialized area in Castellon (NE, Spain). *Environmental Earth Sciences* 66, 879–888.
- Wang Y, Chen Z, Wu Q, Liang H, Huang L, Li H, Lu K, Wu Y, Dong H, Zeng L, Zhang Y, 2016 Observation of atmospheric peroxides during Wangdu Campaign 2014 at a rural site in the North China Plain. *Atmospheric Chemistry and Physics* 16, 10985–11000.
- Watanabe K, Yachi C, Nishibe M, Michigami S, Saito Y, Eda N, Yamazaki N, Hirai T, 2016 Measurements of atmospheric hydroperoxides over a rural site in central Japan during summers using a helicopter. *Atmospheric Environment* 146, 174–182.
- Watanabe K, Yachi C, Song XJ, Kakuyama S, Nishibe M, Jin SJ, 2018 Atmospheric hydroperoxides measured over a rural site in central Japan during spring: Helicopter-borne measurements. *Journal of Atmospheric Chemistry* 75, 141–153.

Highlights

- Measurements of 11 species were taken during a wintertime field study in Utah.
- A high-resolution time-of-flight chemical ionization mass spectrometer was used.
- Vector autoregression analyses were completed for all species.
- Important precursor species included H_2O_2 , N_2O_5 , HONO, and ClNO_2 .

Figure 1.

Time series of concentrations (in ppbv) for selected species measured during the study (O₃, NO₂, and ClNO₂). Concentrations shown in figure are 1-hr average values. Time series shown in figure (1/29/17 through 2/2/17 inclusive) is a subset of the field study duration.

Figure 2.

Time series of measured and predicted concentrations (in ppbv) for O₃ and H₂O₂. Time series shown in figure (1/29/17 through 1/31/17 inclusive) is a subset of the field study duration.

Figure 3. Root mean square error (predicted minus measured values) as a function of the number of lags.

Figure 4.

Values of the first lag, within species constants for Π , plotted as a function of the number of data points.

Figure 5.
Root mean square error (predicted minus measured values) as a function of the interval between lagged values (minutes).

Table 1.

Values of [II] for the first lag fitting parameters.

	const.	NO ₂	NO	NO _y	O ₃	H ₂ O ₂	HONO	CINO ₂	HNO ₃	CH ₂ O ₂	Cl ₂	N ₂ O ₅
NO ₂	0.03	1.07	0.58	-0.87	-0.09	-0.37	0.05	<0.01	-0.13	-0.03	0.08	0.17
NO	0.01	0.03	1.16	-0.22	0.01	-0.11	0.15	-0.14	-0.07	0.01	0.03	0.04
NO _y	0.01	0.11	0.39	0.36	-0.07	-0.19	0.15	-0.08	-0.09	-0.01	0.03	0.06
O ₃	0.01	-0.06	-0.62	0.91	1.07	0.44	-0.09	-0.06	0.11	0.10	-0.09	-0.18
H ₂ O ₂	0.01	<0.01	-0.09	0.16	-0.04	1.24	-0.03	<0.01	0.01	0.03	-0.04	-0.09
HONO	0.01	<0.01	0.13	-0.07	-0.02	-0.03	0.96	<0.01	-0.04	-0.09	<0.01	0.05
CINO ₂	<0.01	<0.01	0.03	-0.05	-0.02	<0.01	-0.02	1.11	0.05	-0.06	0.13	0.01
HNO ₃	<0.01	-0.02	-0.10	0.15	0.01	<0.01	-0.01	0.04	1.39	0.06	<0.01	-0.01
CH ₂ O ₂	0.01	-0.04	-0.10	0.18	0.02	0.02	<0.01	0.01	-0.04	1.24	<0.01	-0.04
Cl ₂	<0.01	-0.02	-0.04	0.05	<0.01	<0.01	<0.01	-0.06	0.01	-0.01	1.07	0.01
N ₂ O ₅	<0.01	0.02	0.01	-0.05	-0.02	<0.01	0.01	<0.01	-0.01	0.04	<0.01	0.97