Accounting for long-term serial correlation in a linear regression problem

Accounting for long-term serial correlation in a linear regression problem

A N Gruzdev

A. M. Obukhov Institute of Atmospheric Physics, Pyzhevsky per. 3, 119017 Moscow, Russia

E-mail: a.n.gruzdev@mail.ru

Abstract. The method is proposed for accounting for the long-term serial correlation (autocorrelation) of data in a linear regression problem. A serially correlated residual series is presented as an autoregressive (AR) process of an order, *k*, that can be much larger than 1, and the autocorrelation (AK) function of the processes is calculated by solving the system of the Yule-Walker equations. Given the AK function, the corresponding AK matrix is constructed which enters the formulas for estimates of regression coefficients and their errors. The efficiency of the method is demonstrated on the base of the multiple regression analysis of data of ground-based measurements of the column NO_2 content at the Zvenigorod Research Station, Russia, and overpass satellite total ozone measurements over the station. Estimates of regression coefficients and their errors depend on the AR order, *k*. At first the error increases with increasing *k*. Then it can approach its maximum and thereafter begin to decrease. In the case of NO_2 and O_3 at Zvenigorod the errors more than double in their maxima compared to the beginning values. The decrease in the error stops at larger *k* if *k* approaches the value such that the AR process of this order is able to account for important features of the AK function of the residual series. The multiple linear regression model for NO_2 and O_3 observations includes the seasonally dependent linear trends and the effects of the solar cycle, the quasi-biennial oscillation, the North Atlantic and Southern Oscillations, and the Pinatubo volcanic eruption. Annual and seasonally dependent estimates of these effects in NO_2 and O_3 have been obtained taking into account the serial correlation as long as 50 months.

1. Introduction

Memory is an inherent property of many geophysical processes. Long memory affects significantly the ability of prediction and/or detection of long term changes [1]. In a regression analysis, the memory can manifest itself as an autocorrelation (a serial correlation) of a residual (error) term, and the presence of the autocorrelation affects uncertainties of regression coefficients (see for example [2]).

A few practically effective methods are known for accounting for the serial correlation. They suppose the residual to be an autoregressive (AR) process of the first order. In an AR model of order *k*, a current value of the process is expressed in the form of a linear combination of *k* previous values of the process [3]. In an AR process of the 1-st order the current value is only the function of the immediately preceding value.

The autocorrelation leads to the reduction of the number of independent values, i.e. to the decrease in the number of degrees of freedom. In [4], the reduction is expressed in terms of the autocorrelation (AK) coefficient (which is equal to the AR coefficient in the 1-st order AR process) and a correction is

found to the variance of a residual series. This method was used, for example, to estimate the thend in ocean surface temperature [5]. A slightly different approximation of the residual series variance was proposed in [6], also expressed in terms of the AR coefficient of the 1-st order AR process. The method was applied, for example, in [7] for NO_2 trend estimates. Cochrane and Orcutt [8] proposed an iterative procedure of accounting for the serial correlation when the residual series could be an AR process of the first order (see also [9]). The procedure was also used under assumption of the residual series to be a 2-nd order AR process [10]. The same approach was used for accounting for the serial correlation in the resent ozone trend estimates [11].

The AK function of the 1-st order AR process decays exponentially [3]. However, the AK function of an actual residual series can differ significantly from the exponent. Therefore, in terms of the AR presentation of the residual, a higher AR order can be required to describe a more complicated behavior of the residual. It means that the longer scale memory is present in the residual series than presumed by the 1-st order AR process.

In this paper, a method is proposed to account for a serial correlation on large scales. It is demonstrated by the example of analysis of ground-based measurements of the column NO_2 content at the Zvenigorod Research Station, Russia, and data of satellite measurements of total ozone over the station.

2. Method

The linear regression model is convenient to write in matrix form (see e.g. [2]):

$$\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{Y}$ is a vector of observations (series under analysis) of length n, $\mathbf{X}$ is an $(n \times p)$ matrix of predictors (independent variables), p is the number of predictors, $\boldsymbol{\beta}$ is a vector of unknown coefficients (desired regression coefficients) of length p, $\boldsymbol{\varepsilon}$ is a vector of errors (residuals) of length n. The overestimated system of equations (1) ($n > p$) is solved by the least squares method. If the residual series, ε, obeys the Gaussian distribution with variance σ^2 and zero mean, then standard errors of desired coefficients can be found simultaneously.

2.1. Accounting for autocorrelation (serial correlation)

Following [2], we present the basic relations for the solution of problem (1), that is, expressions for the root-mean-square estimates of the regression coefficients and their standard errors. Let $\mathbf{V}(\boldsymbol{\varepsilon})$ be the variance matrix of residuals, ε, such that $\mathbf{V}(\boldsymbol{\varepsilon}) = \mathbf{V}\sigma^2$ where σ^2 is the variance of an uncorrelated component of $\boldsymbol{\varepsilon}$. If $\boldsymbol{\varepsilon}$ is uncorrelated, then $\mathbf{V}(\boldsymbol{\varepsilon}) = \mathbf{I}\sigma^2$ where $\mathbf{I}$ is the unit matrix. However it is typical for geophysical data that errors are serially correlated and at least some non-diagonal elements of matrix $\mathbf{V}(\boldsymbol{\varepsilon})$ are not zeros. According to [2], a nonsingular symmetric matrix $\mathbf{P}$ exists such that

$$\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{P}\mathbf{P} = \mathbf{P}^2 = \mathbf{V} \tag{2}$$

where $\mathbf{P}^{\mathrm{T}}$ is the transpose of matrix $\mathbf{P}$. If ε can be presented as an AR process, matrix $\mathbf{V}$ would be the AK matrix of the process [2]. For example, in the case of the 1-st order AR process, the associated AK function decays exponentially (modulo) and the element V_{ij} of $\mathbf{V}$ would be $\rho^{|i-j|}$ where ρ is the AK coefficient at unit lag [3] (see also Appendix A in [9]).

Denote $\mathbf{e}$ the vector of uncorrelated errors associated with the vector of correlated errors, $\boldsymbol{\varepsilon}$. Then [2]

$$\mathbf{e} = \mathbf{P}^{-1}\boldsymbol{\varepsilon} \tag{3}$$

and vector $\mathbf{b}$ of the generalized least squares estimator of $\boldsymbol{\beta}$ is

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{Y} \tag{4}$$

while the vector of squared standard errors, $\boldsymbol{\delta}_b^2$, of estimates $\mathbf{b}$ consists of the diagonal elements of matrix $(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\sigma^{\mathbf{2}}$:

$$\boldsymbol{\delta}_b^2 = \mathrm{diag}(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\sigma^2. \quad (5)$$

Formulas (4) and (5) describe the solution of the regression problem (1) generalized to the case of serial correlation. In the absence of serial correlation, $\mathrm{V}^{-1} = \mathrm{I}$ and formulas (4) and (5) take the standard form [2].

2.2. Method of solution

As shown in section 3, a low order of an AR presentation of the error term can be insufficient to account for the actual long term correlation of residuals. Instead of evaluation formulas [4–7] and iterative procedure [8–11] a method is proposed here for estimating regression coefficients and their standard errors using formulas (4) and (5). A key point is the construction of the AK matrix, **V**, of the AR process approximating the residual. It is known that the AK function of an AR process and the AR coefficients are in recurrent relationship forming a system of the Yule-Walker equations [3].

Two ways of finding the **V** matrix are proposed here. Given the ε series, its AK function can be calculated by usual way. Then solving the system of the Yule-Walker equations one can get the AR coefficients. Given these coefficients, the tail of the AK function is calculated using the aforementioned recurrent relations and therefore the **V** matrix is determined fully.

More complicated but more optimal way is based on the use of the Burg algorithm for calculation of the AR coefficients, which is based on the maximum entropy principle [12]. Given the AR coefficients, the associated AK coefficients are calculated by solving the system of the Yule-Walker equations with respect to the AK coefficients.

Thus, the solution of the linear regression problem (1) is found in two steps. At step 1, the solution of (1) is found by standard way supposing $\mathbf{V} = \mathbf{I}$. A residual series, ε, is then presented as an AR process of order *k* that can differ significantly from 1. In the Burg algorithm, *k* can be as large as one third of the length of ε. The AR coefficients and the associated AK function at lags changing from 1 to *k* are calculated by any of the two described methods. Then the tail of the AK function of the AR process and thereby the full **V** matrix are determined.

At step 2, the least squares estimates of the regression coefficients and their standard errors are calculated using formulas (4) and (5). If necessary, the vector of uncorrelated errors, **e**, can be found using (3). The matrix **P** is a solution of the quadratic matrix equation (2) which can be found with the help of transforming matrix **V** to the Jordan form.

In this work, the method using the Burg algorithm has been implemented to data under analysis.

3. Data used and regression model

For demonstration of the efficiency of the method proposed, the data of column NO_2 measurements at the Zvenigorod Research Station of the A. M. Obukhov Institute of Atmospheric Physics are used. The station is located in a rural area 50 km west to Moscow and is a member of the Network for the Detection of Atmospheric Composition Change (NDACC). Measurements are carried out since 1990 during morning and evening twilight by zenith viewing visible spectrometer. The method of observations and retrieval of the NO_2 content is described in [13–15]. The data of NO_2 measurements are freely available at ftp://ftp.cpc.ncep.noaa.gov/ndacc/station/zvenigor/.

Also used are the overpass data of satellite measurements of total ozone with TOMS instrument for 1996-2005 (ftp://jwocky.gsfc.nasa.gov/pub/eptoms/data/overpass/) and OMI (version TOMS-like) in 2005-2017 (https://avdc.gsfc.nasa.gov/index.php?site=1593048672&id=28). Previous TOMS data are not used because of the long gap. TOMS and OMI data are combined into a single series. The results of NO_2 and O_3 measurements have been statistically verified by the method of [16], and the monthly mean values were calculated. Monthly mean column NO_2 contents derived from evening measurements are shown in Fig. 1a.

The multiple linear regression model for NO_2 and O_3 includes the following predictors: the free member, the linear term, the equatorial zonal wind velocity at the 20 hPa level and the mean of the wind velocities at the 40 and 50 hPa levels (http://www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/index.html) as the proxies of the quasi-biennial oscillation

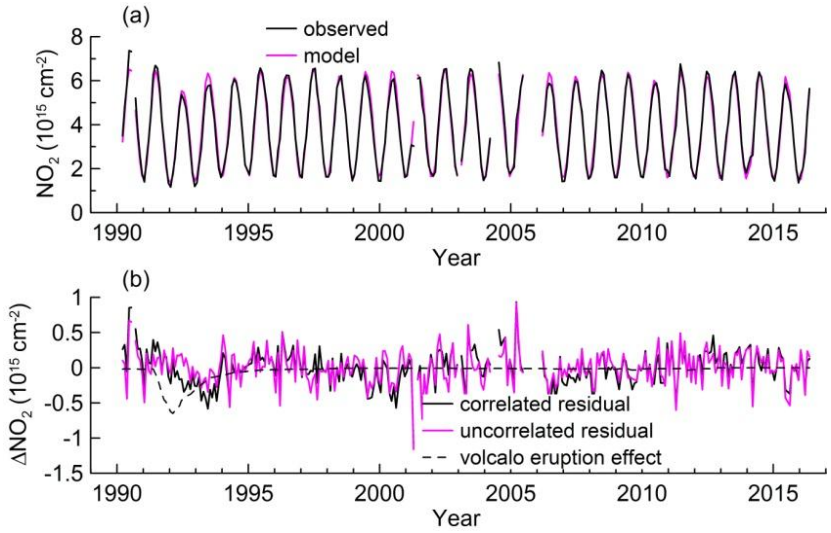

Figure 1. (a) Observed monthly mean column NO_2 contents and their approximation by the regression model. (b) Correlated and uncorrelated (at k=50) error time series and the annual mean Pinatubo volcano eruption effect.

(QBO) effect on NO_2 and O_3, the Nino 3-4 index (https://www.esrl.noaa.gov/psd/data/correlation/nina34.data) and lagged Nino 3-4 index as the proxies of the El Nino–Southern Oscillation (ENSO) effect, the North Atlantic Oscillation (NAO) index (https://crudata.uea.ac.uk/cru/data/nao/nao.dat) as the proxy of the NAO effect, the solar radio flux at 10.7 mm wavelength (https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-radio/) as the proxy of the solar activity effect, the Northern Hemisphere mean of the stratospheric aerosol optical depth at the 550 nm wavelength (https://data.giss.nasa.gov/modelforce/strataer/) as the proxy of the Pinatubo volcanic eruption effect. These predictors represent the major important processes in the regression model that influence the variability of the stratospheric species.

Two proxies of the QBO effect are used to account for lags of the NO_2 and O_3 responses to the equatorial QBO [17–20]. Variations of the wind velocities at the 20 and 40-50 hPa levels are orthogonal to each other to a high degree (the correlation coefficient is ~ –0.01). The use of two orthogonal QBO proxies is also typical for analysis of ozone trends (see e.g. [10, 11]). Similar idea is applied to accounting for the ENSO effect. The second ENSO proxy is constructed by shifting the original Nino 3-4 index by 14 months forward, so that the lagged and the original series are orthogonal to each other (the correlation coefficient is ~ 0.01).

All the required regression coefficients in the regression model are expanded into Fourier pairs corresponding to the annual and semi-annual harmonics to account for the seasonal cycles in NO_2 and O_3, the seasonality of trends and the seasonal dependence of the predictors' effects on NO_2 and O_3 (see also [10, 11]). Note that the NO_2 seasonal cycle at Zvenigorod is mainly described by a set of the annual and semiannual harmonics [13].

4. Method approbation on the example of NO_2 and O_3

Figure 1a shows that the NO_2 contents predicted by the regression model correspond well to observations. As explained by the dashed curve in Fig. 1b, the significant decrease in NO_2 in 1992 is the Pinatubo volcano effect caused by the heterogeneous chemistry on the surface of stratospheric sulfate aerosol formed by products of the volcano eruption (see e.g. [21]). Note however that the proxy used to account for the Pinatubo-related effect underestimates the effect in 1993 (compare two curves in Fig. 1a and see also the black curve in Fig. 1b).

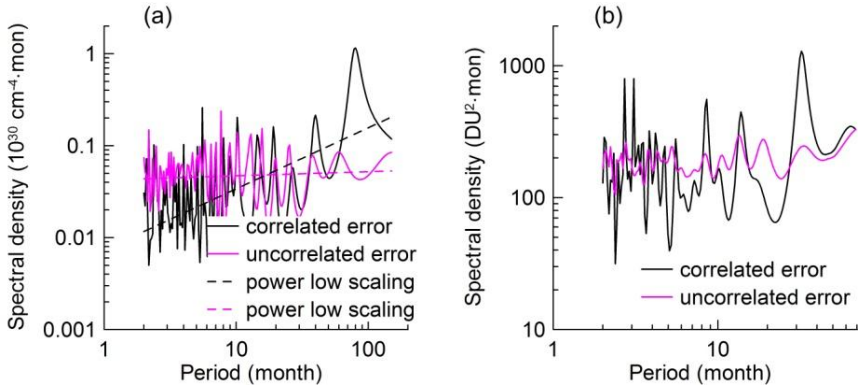

Figure 2. Power spectra of correlated and uncorrelated (at k=50) error time series for (a) NO_2 and (b) O_3.

The correlated error, ε, and the "uncorrelated" error, e, corresponding to the AR order of 50 are shown in Fig. 1b. Figure 2 compares power spectra of the correlated and uncorrelated errors for NO_2 and O_3 calculated with the help of the maximum entropy method [12]. The spectrum of the NO_2 correlated error points to the existence of variations of the residual time series, ε, with the period of about 80 months (Fig. 2a). Another feature of this spectrum is an approximate power-low scaling with a scaling exponent of about 0.7 (the black dashed line in Fig. 2a). Unlike this, the spectrum of the uncorrelated error follows the power low with the near-zero scaling exponent that is a theoretical value of the scaling exponent of the white noise. The spectrum of the correlated error for O_3 is characterized by the pronounced maxima on periods of about 3 and 40 months, while these features are absent in the spectrum of the uncorrelated error (Fig. 2b).

Figure 3 shows the AK functions of the AR processes approximating the correlated error under various values of the AR order. The AK function corresponding to the AR order k = 90 exhibits the oscillation with a period of about 80 months that is consistent with the period revealed by the spectrum analysis in Fig. 2a. At k = 1, the AK function decays exponentially from its beginning value, while at k = 10 a quasi-exponential decay occurs at lags exceeding 10 months. The value k = 30 allows the AK function to get an oscillation but its period is still lower than 80 months. The value k = 50 is sufficient to resolve the 80-month oscillation, and the corresponding AK function, coinciding in the 1–50 months lag range with the function corresponding to k = 90, follows it in general at larger lags too.

To demonstrate the influence of the value of the AR order on estimates of the regression coefficients, Figure 4 shows the annual estimates of the linear trends in NO_2 and O_3 and their 95% confidence intervals as functions of order, k, of the AR process. The value k = 0 corresponds to the case when no serial correlation is taken into account. Figure 4a shows that in the NO_2 case the k value influences both values and errors of the regression coefficients. The trend estimated at different k

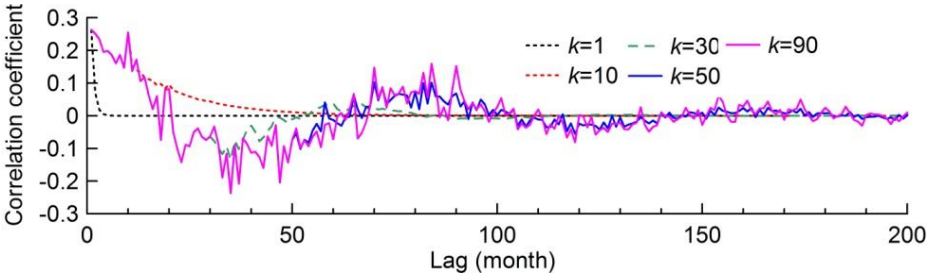

Figure 3. Autocorrelation function of the autoregressive process approximating the correlated error time series for NO_2 at the autoregressive order, k, equal to 1, 30, 50, and 90.

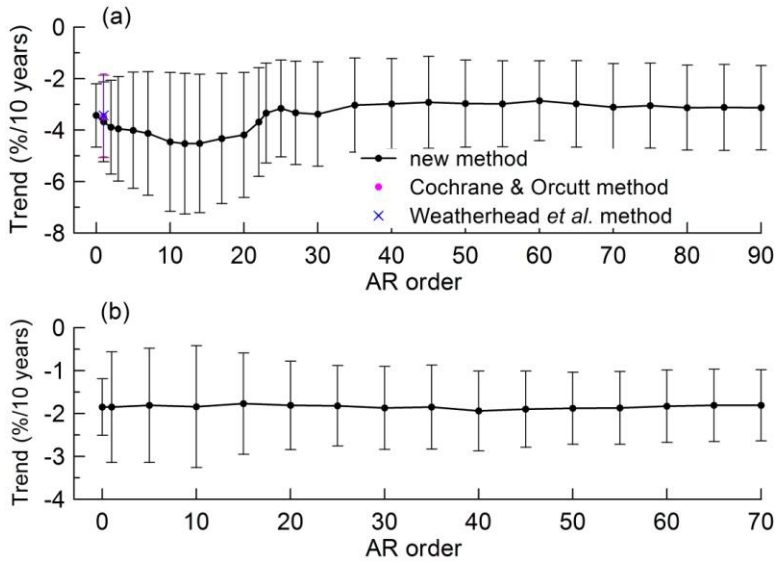

Figure 4. The annual linear trends in (a) NO_2 and (b) O_3 and thir 95% confidence intervals as function of the AR order.

varies by about one third of its value; however the discrepancies are within the 95% confidence intervals. The error of the trend estimate increases significantly, more than twice, with increasing k at small k, approaches the maximum at $k = 12 \div 15$, and gradually decreases until k approaches 50. At $k > 50$ the trend estimates and their errors are approximately constant. Figure 4a shows also the trend estimates calculated with the use of the methods of [6, 8].

In the O_3 case, the trend estimate does not depend but the trend error depends on the AR process order. The error increases with increase in k at small k but then decreases and stays approximately constant at $k>40$ (Fig. 4b).

Figure 5 shows the linear trend in NO_2 and the magnitudes of the predictors' effects on NO_2 for $k = 50$. The annual, monthly and seasonal mean values are plotted. The solar cycle effect denotes the associated NO_2 change when solar activity varies from the minimum to the maximum phase. The QBO effect corresponds to the changes in the equatorial zonal wind velocity from the maximum easterly to the maximum westerly phase. The Pinatubo-related effect means the NO_2 change corresponding to the maximum of the aerosol optical depth. The NAO and ENSO effects are the associated NO_2 changes per one standard deviation of the NAO and ENSO indices, respectively. The percent values of the effects are determined relative to the multi-year annual mean NO_2 content. Note that the ENSO effect is presented only for one ENSO proxy since the annual and most of the monthly values of the NO_2 response to the other, lagged, ENSO predictor are statistically insignificant at the 95% confidence level.

According to Fig. 5a, the linear trend in NO_2 is seasonally dependent with maximum (modulo) values in spring (up to –6% per decade) and autumn. The annual trend of about –3% per decade is also statistically significant.

While the estimated annual effect of solar activity changes on NO_2 is small and statistically insignificant, a significant effect is derived for the late winter–early-to-middle spring period (Fig. 5b). The NO_2 content in March during the minimum phase of solar activity is by 8% larger than during the maximum phase. Relative to the monthly mean NO_2 values, the solar cycle effect approaches –10% in February and March.

The maximum (modulo) Pinatubo effect on the evening NO_2 contents of about –30% is obtained for summer (Fig. 5c). If, however, the NO_2 changes are normalized by the monthly mean NO_2 contents, the Pinatubo effect is less seasonally dependent and varies around the value of –20%.

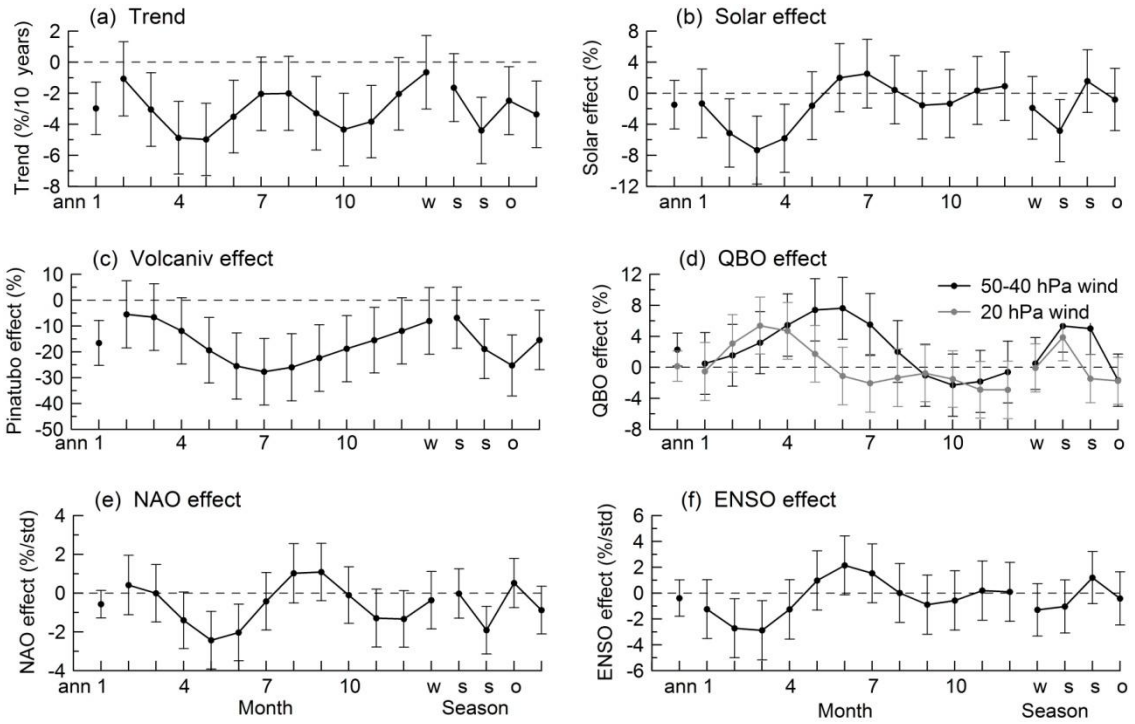

Figure 5. Annual (dots in the left parts of the plots), monthly (curves in the middles of the plots) and seasonally mean (curves in the right parts of the plots) estimates of (a) the linear trend in NO_2 and (b–f) the magnitudes of the effects of different predictors on NO_2 at k=50. Vertical bars are the 95% confidence intervals

The QBO effect on NO_2 is mostly pronounced in the spring–summer period (Fig. 5d). The magnitude of the effect in summer is about 8% relative to the annual mean NO_2 content. Different seasonal dependences of the two curves in Fig. 5d indicate that the phase shift of the QBO effect in NO_2, relative to the equatorial QBO, changes with season. For example, the QBO effect in NO_2 in March lags behind the QBO in the 20 hPa wind velocity by one eighth of the QBO period (~3.5 months), since the April values on the two curves in Fig. 5d are almost equal to each other and the phase of the equatorial wind QBO propagates downwards. The QBO effect on NO_2 in June is in phase with the 20 hPa wind velocity QBO. The summer NO_2 contents are in general larger during the westerly phase then during the easterly phase of the equatorial zonal wind at the 20 hPa level. The spring NO_2 contents are generally larger during the westerly phase compared to those during the easterly phase of the QBO in the equatorial zonal wind in the 20-50 hPa layer.

A small negative statistically significant effects of the NAO on the NO_2 contents are derived for the spring and, at the edge of statistical significance, for the autumn periods (Fig. 5e). The changes in NO_2 in spring are estimated as 2–3% if the NAO index changes by value of its standard deviation.

A negative estimate of ~3% has been obtained for the ENSO effect on NO_2 in the late winter–early spring, but an effect of opposite sign, at the edge of statistical significance, is seen in June (Fig. 5f).

It should be noted that estimates of the linear trends in NO_2, effects of the 11-year solar cycle and volcanic eruption on NO_2 at Zvenigorod obtained by the multiple linear regression method were reported in earlier works [16, 22–24]. The signs of the estimates from the present work and from the specified works coincide. The annual estimate of the NO_2 trend obtained by the new method is close to the latest estimate in [24] although somewhat smaller (modulo) than it, but two times smaller (imodulo) than the earlier estimates [16, 22]. The annual estimate of the volcanic effect in the present work is close to the earlier estimates in [16, 22, 23] but by a factor of 1.5 less than previous estimates obtained without the use of multiple regression [25, 26]. For the solar cycle effect on the evening NO_2

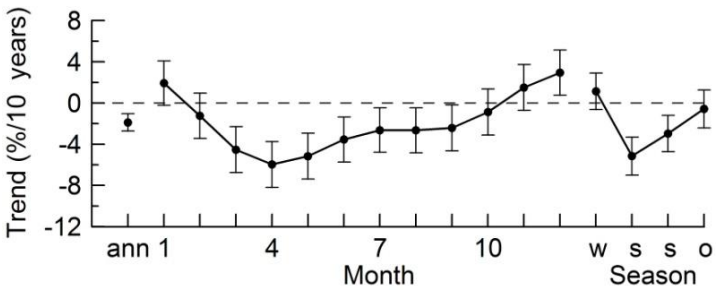

Figure 6. Annual, monthly, and seasonally mean estimates of the O_3 linear trend and thir 95% confidence intervals at k=50.

contents, a statistically significant negative value was obtained in [23] that almost twice as large modulo as the statistically insignificant annual estimate obtained in the present work. The reasons for the observed quantitative differences may be different. The length of the analyzed series as well as differences in the regression models, in particular, the choice of a volcanic aerosol proxy, may be important.

Figure 6 shows annual, monthly, and seasonal estimates of the O_3 linear trend. The annual trend estimate for the 1996–2007 period over Zvenigorod is about –2% per decade. The trend is maximum (modulo) in spring (on average about –3% per decade). The minter and autumn mean estimates of the trend are statistically insignificant, but the trend is positive in December, of about 3% per decade.

5. Conclusions

In this paper the method is proposed for taking into account long-term serial correlation of data in a linear regression problem. The efficiency of the method has been demonstrated on the base of the multiple linear regression analysis of data of column NO_2 measurements at the NDACC station of Zvenigorod, Russia, and data of the overpass satellite total ozone measurements over the station.

Some conclusions can be drawn which can be also valid for regression analysis of other data since the long term correlation is an inherent property of various geophysical parameters.

Estimates of regression coefficients and/or their errors depend on the order, k, of the AR processes approximating the correlated residual series. Generally the longer the scale of the autocorrelation of the residual, the larger AR order is required. At first the error increases with increasing k. Then it can approach its maximum and thereafter begin to decrease. In the case of column NO_2 and total ozone over Zvenigorod the corresponding errors more than double in their maxima compared to the beginning values.

The decrease in the error can stop at larger k if k approaches the value such that the AR process accounts for important features of the AK function of the residual series. If however the error does not achieve a maximum within the range of possible values of k, one may believe that the regression model used does not describe adequately observations and should be modified (see also [2]).

In the case of the NO_2 and O_3 data used, the errors maximize at k ~10 and then decreases until k approaches 50. For larger possible (up to about 1/3 of the length of the series) values of k the regression coefficients and their errors keep the values approximately constant.

The multiple linear regression model for the data of NO_2 and O_3 observations includes the seasonally dependent linear trend and the seasonally dependent effects of the solar cycle, the equatorial QBO, the NAO and ENSO, and the Pinatubo volcanic eruption. Annual and seasonally dependent estimates of these effects in NO_2 as well as the O_3 trend estimates have been obtained taking into account the autocorrelation of the error up to time scale of 50 months.

The method proposed is applicable to the analysis of a variety of data. The quantitative characteristics of the dependence of the regression coefficients and their errors on the order of the AR process for those data may be specific.

One important application of the method proposed could be the detection of the stratospheric ozone recovery (see e.g. [27]). As Fig. 4 shows, the confidence intervals at large AR orders are smaller than at k=1 (especially for O_3). This can help earlier detection of the ozone recovery compared to those when the serial correlation is only accounted for short time scale (k=1÷2).

Acknowledgments
The work was partly financed by the Russian Foundation for Basic Research (project No. 16-05-00663) and the Programs of the Russian Academy of Sciences. The data of NO_2 measurements are archived by the Network for the Detection of Atmospheric Composition Change (NDACC). The ozone data are presented for free access by the NASA's Goddard Space Flight Center. The data of the F10.7 solar activity index are freely provided by the NOAA National Geophysical Data Center. The data of the equatorial stratospheric zonal wind velocities are provided by Freie Universität Berlin. The Nino 3-4 index data are provided by the NOAA-ESRL Physical Sciences Division. The data of the NAO index are provided by the Climatic Research Unit, University of East Anglia. The data of the stratospheric aerosol optical depth are prepared by the NASA Goddard Institute for Space Studies.

References
[1] Beran J 1994 *Statistics for Long Memory Processes* (New York: Chapman & Hall)
[2] Draper N R and Smith H 1998 *Applied Regression Analysis* (New York: John Wiley & Sons)
[3] Box G E P and Jenkins G M 1970 *Time Series Analysis: Forecasting and Control* (San Francisco: Holden-Day)
[4] Bartlett M S 1935 Some aspects of the time correlation problem in regard to test of significance *J. Roy. Stat. Soc.* **98** 536–43
[5] Folland C K, Salinger M J, Jiang N and Rayner N A 1949 Trends and variations in South Pacific island and ocean surface temperatures *J. Climate* 2003 **16** 2859–74
[6] Weatherhead E C *et al.* 1998 Factors affecting the detection of trends: Statistical considerations and applications to environmental data *J. Geophys. Res.* 1998 **103** 17149–61
[7] Hendrick F *et al.* 2012 Analysis of stratospheric NO_2 trends above Jungfraujoch using ground-based UV-visible, FTIR, and satellite nadir observations *Atmos. Chem. Phys.* 2012 **12** 8851–64
[8] Cochrane D and Orcutt G H 1949 Application of least squares regression to relationships containing auto-correlated error terms *J. Amer. Statist. Assoc.* **44** 32–61
[9] Tiao G C, Reinsel J C, Xu D, Pedrick J H, Zhu X, Miller A J, DeLuisi J J, Mateer C L and Wuebbles D J 1990 . Effects of autocorrelation and temporal sampling schemes on estimates on trend and spatial correlation *J. Geophys. Res.* **95** 20507–17
[10] Bodeker G E, Boyd I S and Matthews W A 1998 Trends and variability in vertical ozone and temperature profiles measured by ozonesondes at Lauder, New Zealand: 1986-1996 *J. Geophys. Res.* **103** 28661–81
[11] Harris N R P *et al.* 2015 Past changes in the vertical distribution of ozone – Part 3: Analysis and interpretation of trends *Atmos. Chem. Phys.* **15** 9965–82
[12] Kay S M and Marple S L 1981 Spectrum analysis – A modern perspective *Proc. IEEE* **69** 1380–419
[13] Elokhov A S and Gruzdev A N 2000 Nitrogen dioxide column content and vertical profile measurements at the Zvenigorod Research Station *Izvestiya, Atmos. Oceanic Phys.* **36** 763–77
[14] Gruzdev A N and Elokhov A S 2010 Validation of Ozone Monitoring Instrument NO_2 measurements using ground based NO_2 measurements at Zvenigorod, Russia *Internat. J. Remote Sens.* **31** 497–511
[15] Gruzdev A N and Elokhov A S 2011 Variability of stratospheric and tropospheric nitrogen dioxide observed by visible spectrophotometer at Zvenigorod, Russia *Internat. J. Remote Sens.* **32** 3115–27

[16] Gruzdev A N 2008 Latitudinal dependence of variations in stratospheric NO_2 content *Izvestiya, Atmos. Oceanic Phys.* 2008 **44** 319–33
[17] Gruzdev A N 2011 Quasi-biennial variations in the total NO_2 content *Doklady Earth Sci.* **438** 678–82
[18] Ageyeva V Yu and Gruzdev A N 2017 Seasonal features of quasi-biennial variations of NO_2 stratospheric content derived from ground-based measurements *Izvestiya, Atmos. Oceanic Phys.* **53** 65–75
[19] Gruzdev A N and Bezverkhnii V A 2005 Quasi-biennial oscillation in the atmosphere over North America from ozonesonde data *Izvestiya, Atmos. Oceanic Phys.* 2005 **41** 29–42
[20] Gruzdev A N and Bezverkhnii V A 2006 Quasi-biennial variations in ozone and meteorological parameters over Eastern Europe from ozonesonde data *Izvestiya, Atmos. Oceanic Phys.* **42** 203–14.
[21] Koike M, Jones N B, Matthews W A, Johnston P V, McKenzie R L, Kinnison D and Rodriguez J 1994 Impact of Pinatubo aerosols on the partitioning between NO_2 and HNO_3 *Geophys. Res. Lett.* **21**: 597–600
[22] Gruzdev A N 2009 Latitudinal structure of variations and trends in stratospheric NO_2 *Internat. J. Remote Sens.* **30** 4227–46
[23] Gruzdev A N 2014 Estimate of the effects of Pinatubo eruption in stratospheric O_3 and NO_2 contents taking into account the variations in the solar activity *Atmos. Oceanic Optics* **27** 403–11
[24] Gruzdev A N, Arabov A Ya, Borovsky A N, Elansky N F, Elokhov A S, Golitsyn G S and Mokhov I I 2016 Trends in stratospheric column NO_2 in mid-latitudes of the European part of Russia *Quadr. Ozone Symp. 4-9 September 2016, Edinburgh, UK* http://presentations.copernicus.org/QOS2016-218_presentation.pdf
[25] Elokhov A S and Gruzdev A N 1998 Measurements of column contents and vertical distribution of NO_2 at Zvenigorod Scientific Station *Proc. SPIE* **3583** 547–54
[26] Gruzdev A N and Elokhov A S 2005 Ground-based spectrometric measurements of vertical distribution and column abundance of NO_2 at Zvenigorod, Russia *Proc. SPIE* **5832** 292–9
[27] Chipperfield M P, Bekki S, Dhomse S, Harris N R P, Hassler B, Hossaini R, Steinbrecht W, Thieblemont R and Weber M 2017 Detecting recovery of the stratospheric ozone layer *Nature* **549** 211–8