

This is the response of L. Frossard and H.E. Rieder on behalf of all authors (AC, written in italic) to the comments of Referee 2 (R2).

First of all we thank the referee for his/her positive judgment and valuable suggestions that helped us to improve the manuscript. Most of the referee's comments were related to the structure and clarity of Section 3. We thank the referee for his/her critical review of this part and are especially grateful to him/her for having taken the time to understand this part of the paper, and for coming back with precise and therefore very valuable comments on it. We substantially edited this section and some related subsections with regard to improved clarity and readability. We hope that the material of the paper is now easier to access for a broad audience.

The individual points raised by the referee are addressed in the reply below. Comments that are not specifically mentioned either related to typos and minor reformulations and have all been addressed, or they related to lack of clarity in the mathematical part and should have been solved in the revised version of the manuscript.

Specific Comments:

R2: p 13163 L1: The term “model” has a different meaning in statistics than in atmospheric sciences. ...

AC: We adjusted according to the referee's suggestion to 'We use statistical models' Further, we added a footnote after 'Statistical modeling' on p.13164 L 26 reading 'In this paper about a statistical analysis of past ozone data, the only “models” considered are statistical ones, even if this is not always explicitly stated.'

R2: p 13163 Abstract: The abstract should summarize not only the method but also the results.

AC: We extended the abstract according to the referee's suggestion. We included a summary of the main results as given below.

'While results for the solar cycle, QBO and EESC are in good agreement with findings of earlier studies, unprecedented spatial fingerprints are retrieved for the dynamical covariates. Column ozone is enhanced over Labrador/Greenland, the North Atlantic sector and over the Norwegian Sea, but is reduced over Europe, Russia and the Eastern United States during the positive NAO phase, and vice-versa during the negative phase. The NAO's southern counterpart, the AAO, strongly influences column ozone at lower southern mid-latitudes, including the southern parts of South America and the Antarctic Peninsula, and the central southern mid-latitudes. Results for both NAO and AAO confirm the importance of atmospheric dynamics for ozone variability and changes from local/regional to global scales.'

R2: p 13164 bottom - p 13165 top: long sentence.

AC: We modified this sentence as follows:

'In particular, a better description of extreme events and an assessment of whether their distribution has changed over time are needed to make progress in answering important questions about ozone-climate interactions.'

R2: p 13165 L18: the acronym ARMA should be defined.

AC: *We defined ARMA accordingly at its first appearance in the main text.*

R2: p 13165 L18: ... Further, at the end of the introduction it would be advantageous to make a clear and simple statement which question shall be answered by which method.

AC: *During the revision of the manuscript we also modified the end of the introduction, where we now explain our approach.*

R2: p 131167 L1-4: This short motivation of the choice of methods would be useful already at the end of the introduction. Otherwise the reader might be at a loss with the technical terms at the end of the intro. ...

AC: *The end of the introduction now contains more details on the models we use, see above.*

R2: p 13167 L6: Please make a short statement here what the purpose of r-largest order statistics is. ... Such an introductory statement will help the reader to better understand the details listed below.

AC: *The revised Section 3.1 of the paper contains an introduction to extreme value theory models.*

R2: p 13169 Eq.1: There is a cascade of definitions, which extends down to line 8 on page 13170, which all are necessary to understand Eq. 1. Couldn't this part be reorganized from a top-down style to a bottom-up style to avoid that the reader encounters so many yet undefined quantities? I.e. go from Eq (3) to Eq (2) to Eq (1) and then to the z-test?

AC: *This has been clarified in the revised version, where section 3.1 consists of two subsections; the first about models for extremes in general contains Eq.1 and the z-test, whereas the second about the implementation for our ozone data contains Eqns. 2 and 3. We did not change the order of these two equations to emphasize the generality of the approach, but we hope that it is now more accessible.*

R2: p 13169 L12ff: "normality" can be anything. ... What does the MAXIMUM likelihood estimator mean here? ... To me it is not quite clear here what "the estimate" is. ...

AC: *We replaced 'normal' by 'Gaussian' throughout the paper to avoid any ambiguities. Further, the revised version includes an Appendix where some statistical concepts are briefly*

explained, among others the terms ‘estimate’, ‘maximum likelihood estimator’ and its asymptotic Gaussian distribution.

R2: p 13169 L12-16: This paragraph assumes that the reader knows all details on the z-test. ... Thus some explanation might be useful. Please state explicitly that the p-value is ...

AC: The terms ‘z-test’ and ‘p-value’ are also explained in the added Appendix.

R2: p 13169, L15: Calculation of the standard error may be difficult. Is there reason to assume that the elements of the sample are independent?

AC: We compute standard errors for the maximum likelihood estimates based on the asymptotic distribution of the maximum likelihood estimator (MLE). For this result to hold for the r-largest order statistics model, our assumption that the blocks are independent is sufficient. The asymptotic Gaussian distribution of the MLE for the ARMA model has been established in Hannan, Journal of Applied Probability 10, 130-45, 1973.

R2: p 13169, L20: This would all be clearer if it was stated somewhere what the order statistic predicts: ...

AC: The role of the r-largest order statistics model in extreme value theory is presented in the revised version.

R2: p 13169 Eq 2 and below: Is the design matrix $Z(t)$ really necessary? To me it seems to add a lot of unnecessary formalism and further technical terms which are not helpful for readability of the paper. I do not see where the design matrix is further used in the paper, since Eq. 3 provides directly the μ needed in Eq 1.

AC: We decided to keep the design matrix because of its generality and because it allows giving a covariate-dependent expression that is independent of any statistical model. We also make more use of it in the revised versions of this paper and of the companion paper by Rieder et al.

R2: p 13170 L14: “best”: Has a certain norm been minimized or is “best” a subjective choice?

AC: As stated in the lines preceding the one cited, it was infeasible to properly do model selection at every grid cell. On the other hand, our modeling approach required one “global” set of covariates. For these two reasons, we finally selected this set of covariates after having explored on different random subsets of 72 grid cells that all these covariates are important in at least some grid cells.

Since this not a fully objective criterion, the last sentence before Eq. 3 now reads ‘The form of Eq (2) finally used is ...’

R2: p 13170 Eq3 and above: ... The use of the term “model”, however, seems to clash with the term “model” elsewhere in the paper.

AC: *In the revised version we state already in the introduction that we only talk about statistical models in this paper (see also our reply to the first comment).*

R2: p 13172 L1-9: This text suggests that multiple linear regression models never consider autocorrelations. This is, however, not true, there exist multiple linear regression models which support the use of the full covariance matrix of the data errors (c.f. von Clarmann et al., Atmos. Chem. Phys., 10, 6737-6747, 2010) and there exist methods to determine the covariance matrices from the data (e.g. AR(1) method as used by Vyushin et al., JGR 112, D14307, 2007). Are linear regression models considering covariances not also called multiple linear regression models?

AC: *The studies of von Clarmann et al. (2010) and Vyushin et al. (2007) show ways to deal with autocorrelation when analyzing atmospheric data. Von Clarmann et al. (2010) show how to do trend estimation for groupwise correlated data and Vyushin et al (2007) outline the importance of addressing long-range correlation in total ozone time series. In addition both studies point out that autocorrelation is frequently not accounted for in such analyses.*

In our context we address the temporal correlation of the ozone data by an autoregressive moving average (ARMA) model, which is an easy and flexible model for correlated time series, and which contains the commonly-used autoregressive model, but can be more parsimonious.

In order to clarify the term ‘multiple linear regression model’ and to acknowledge the progress in previous work, we modified the paragraph as given below.

‘The multiple linear regression model, a standard tool for modeling and analyzing changes in a variable of interest (e.g., total ozone) and the contribution of individual covariates to these changes, is widely used in atmospheric science. In a first step we therefore fitted a multiple linear regression model in its simplest form, i.e., with independent errors, to the monthly total ozone mean values at both northern and southern mid-latitudes. However, the residuals of this model were significantly correlated, showing that the assumption of independent errors does not hold for our data. Autocorrelation in model residuals of total ozone has been reported in previous studies. Vyushin et al. (2007), for example, use a model with autocorrelated noise for total ozone, and von Clarmann et al. (2010) also suggest accounting for correlations when analyzing atmospheric data.’

R2: p 13172 L10ff: please describe shortly the main characteristics of ARMA methods. What is their advantage compared to other methods? ...

AC: *We added the sentence ‘The strengths of the ARMA model are its flexibility and parsimony, as a wide range of temporal correlation patterns can be expressed with only few parameters.’ before giving the definition of an ARMA process.*

R2: p 13172 Eq 4: What is the nature of Y ? How is it related to y ? ...

AC: We clarified notation in this part by using W in the definition of an ARMA process and using only Y for total ozone mean values. Capital letters are used because these quantities formally denote random variables.

R2: P 13172 L11: ... Now, in retrospective, pages 13169 and 13170 make more sense to me, but it would be most helpful to state already near p 13168 in easy words what the general rationale of the procedure is. ...

AC: The revised version should be clear in this respect.

R2: General question: Are the autocorrelations in the residuals only an issue in the analysis of mean values and not in the analysis of extremes?

AC: Autocorrelation is a more important issue for mean values than for extremes, because extreme events may be more distant and hence tend to be less correlated. Clustering of extremes might though be an issue. This is discussed on p 13168 L25ff.

R2: Which problem is solved by the ARMA approach? To determine the covariances of the residuals, or to consider these in the fit, or both?

AC: The ARMA model allows to model correlations in the data to which it is fitted. Its parameters describe these correlations, but they are not of interest in this study. Here, we just use an ARMA model to make sure that the correlations do not lead to artificial effects of the covariates.

R2: p 13173 L10: I doubt that the Akaike information criterion is known by the majority of the readers of ACP. Please explain.

p 13174 L10: The Bonferroni correction might not be known by the majority of readers of ACP. Please explain. ...

AC: These two terms are now explained in the new Appendix.

R2: p 13174 L15: Please guide the reader through the results by first stating what the purpose of the analysis is, what hypothesis is to be corroborated/falsified, which relation is to be demonstrated. Otherwise the reader might get lost in the quantitative details and lose the thread.

AC: In the revised version we also slightly reorganized Sect. 4.1 to make it better understandable.

R2: p 13182 L4: Is ozone really transported that far? According to Brasseur and Solomon, Aeronomy of the middle atmosphere, Springer, Sect 5.2.3., this view could be somewhat oversimplified.

AC: *We thank the referee for this comment. As stated in Brasseur and Solomon (2005) the primary region supplying ozone to the extra-tropics is a narrow transition zone, where dynamics and chemistry compete. In this transition zone air flows fast and retains relatively high mixing ratios and the pole-ward and down-ward meridional flow through that transition zone feeds ozone in the extra-tropics and dominates the extra-tropical column ozone. Therefore, we updated the statement to ‘... as resulting from the interaction between ozone production, ozone transport to mid- and high-latitudes (due to the Brewer-Dobson circulation, and enhanced ozone depletion in polar regions...’*

R2: Figures are generally too small. Axes labels are hardly readable.

AC: *We will check on this during the production phase of the revised version, assuming that the revision is acceptable.*

R2: Figures 2 and 3: panels b and e should be discussed in more detail. I wonder that correlations appear not even in tendency to be autoregressive. In fig. 2 there is a maximum at time lag 12 months. Does this suggest that there is a problem with the annual cycle?

AC: *The referee correctly points out that a high autocorrelation at lag 12, ρ_{12} , might indicate a problem with the annual cycle. In the present case, however, such a problem (if any) is not serious, because not only ρ_{12} is just slightly larger than the approximate asymptotic confidence bound for white noise, but also its magnitude does not stand out against the next largest autocorrelations at lags 1 and 11. Further, we recall that our model for extremes assumes that months are independent. Given that the monthly mean values are correlated, a slight correlation in the residuals of the model for extremes has therefore to be faced.*

To address the other aspects of this comment, we added a short discussion at the end of Sect. 4.1 in the revised version as given below.

‘In the residuals of both models, however, autocorrelation has practically gone (panels (b) and (e) in Fig. 2 and 3), even though the correlations at some lags are slightly outside the approximate confidence bounds. For the mean values this is expected, as modeling temporal correlation is the main purpose of the fitted ARMA model. The extreme events may arise from more distant observations and hence tend to be less correlated than the underlying daily values.’