1 General Comments:

This promising paper studies sources of altitude-dependent variation in stratospheric ozone, based on 3+ decades of monthly averaged Umkehr vertical profiles at two sites - Boulder and Arosa, excluding years impacted by El Chichon and Pinatubo. Statistical methods used include functional principal components for dimension reduction, followed by spline-based models on the principal component scores, under Gaussian error assumptions. Spline-based Bayesian Confidence intervals provide uncertainty in a mixed model framework with heterogeneous variances. Combining these methods enables one to study changes in the altitude-dependent shape of the ozone profiles as a function of time-varying dynamical/chemical/other forcings/trends. I personally believe careful use of this approach has strong potential to provide additional scientific insight on ozone-related processes on a variety of space-time scales. (The authors cite my closely related ozonesonde work - Meiring, 2007 - henceforth M07). The main contributions in this paper above M07, include considering additional explanatory indices, and analyzing monthly Umkehr profile data in a single model with heterogeneous variances. I believe the paper under review has very valuable potential, after the authors address/clarify important questions about the impact of specific model/analysis choices, including methods used to study interactions. There are many possible analysis approaches - and I simply encourage the authors to study whether their results about interactions (and main effects in the absence of interaction) are sensitive to their choices (especially the type of mean filtering and smoothing parameter selection). I hope revision may help in signal detection of interactions etc - if they are present in the data. I agree with Referee 1 in all his comments - including that we are left hanging about the Arosa results. Some suggestions for edits/corrections are also given below.

2 Specific Comments:

In this section I concentrate on modeling choices/assumptions/fitting procedures, keeping in mind the famous idea that "all models are wrong but some are useful" (attributed to George Box). My questions are based on my own experience refining results in M07, which I am grateful the authors cite. I hope the authors will find my suggestions/questions constructive and helpful. There are infinitely many possible approaches, and I totally respect that the authors may have very valid reasons to justify their own choices.

Below I occasionally refer to your paper as PGP13 (using the authors' initials).

1. Was any other data processing performed, or any transformations tried - besides calculating monthly averages and eliminating years when the Umkehr data were affected by the Volcanic Eruptions? Would a log transformation help?

2. Am I correct in understanding that only an altitude dependent mean field (constant in time) was subtracted prior to principal component analysis? Any modes of variability not filtered out in the mean will be included in the covariance - and therefore in the resulting principal components and scores. If a mean seasonal cycle is not filtered out in the mean term, this strong seasonal cycle will dominate the variation in the principal component scores (as can be seen in PGP13 figures). If a seasonal and altitude dependent mean can be estimated with relatively high precision, and then filtered, modes of variability that are less dominant than the mean seasonal cycle might be clearer in the principal component scores. Interactions with seasonal cycle can still be examined in both approaches. Did the authors consider a more complex mean cycle? I recognize that there are pros/cons in both approaches.

3. I believe that the breadth of explanatory indices that PGP13 include in their score/coefficient models, is a strength of the paper with great potential. I suggest including plots of the AO and solar index used, also superimposing the time-periods which were omitted due to the volcanic activity. Would you expect any time-lags if there were effects of these indices? How do your results compare/ contrast with other literature using the AO and solar indices (in total column or other ozone studies)? Are the omitted volcanic time-periods also close to the peak of the solar cycles?

4. From the writing, I missed what methods are being used to determine whether interactions or other effects are "negligible". Are formal model selection methods being used (such as AIC/BIC/other methods if available and appropriate)? Which methods are being used? Are your results sensitive to the method of smoothing parameter choice?

5. The standard errors of the Bayesian confidence intervals depend on the smoothing parameter. As you mention, GCV smoothing parameter selection can be very sensitive to correlation. Did you try GML in the additive model, which can be fitted in a mixed model framework (see e.g., Wang 1998 a-b, Wang 2011)?

6. Like Referee 1, I am concerned about us being left hanging about the Arosa results. Is this an Umkehr data quality issue, or would different modeling choices have enabled detection of signals other than those found, or do you think these signals are not present at Arosa? I also wonder about sensitivity in the Boulder results.

7. Thank you for providing the reference to Marra and Wood's recent papers, of which I was not yet aware, related to coverage properties and GAM model selection. I have not yet had opportunity to read these - so some justification to my other questions may be based on that.

3 Technical Corrections/ suggestions for edits:

1. Please add the source for the Umkehr data set, as you did for the indices you used. Are the Umkehr data publicly available?

2. If I understand notation correctly, the authors make Normal error and random effects assumptions throughout. If so, I recommend replacing the GAM acronym by Additive Models (AM instead of GAM) and GAMM by additive mixed models (AMM instead of GAMM) throughout. This would avoid confusion if readers want to fit GAM and GAMM for other distributions.

3. I suggest editing (rewriting certain sections to improve flow and emphasize the primary goal of the paper): possibly reducing the AM section with GCV, if you prefer the additive mixed model (AMM) approach. GCV is very valuable in some situations, but is highly sensitive to correlation that may be present in these data even on monthly time-scales (although difficult to model). I believe there is some evidence that GML is less sensitive (See Yuedong Wang references below). AM's with either smoothing spline or penalized regression spline components could be fitted in a mixed model framework using GML - so the AM and AMM sections could possibly be combined unless there is strong reason not to do this.

4. For the spline based terms, are you using regression splines, penalized regression splines, or smoothing splines? Different terms are used in different parts of the paper - for example, the abstract mentions "knot-based regression cubic splines", but penalization/smoothing is later described in the paper. On page 12347 - do you mean "cyclic and noncyclic cubic smoothing splines instead of "cyclic and noncyclic regression cubic splines" when citing Wahba (1990)? Are you then using a penalized regression spline to approximate a smoothing spline, but with reduced basis function/knot selection? This could simply be a difference in terminology - but please check your descriptions and references for consistency.

5. Also, what basis functions were used for interactions in both the additive models (AM) and additive mixed models (AMM)?

6. Clarification on description of methods cited in M07:

(a) Due to the irregular measurements in the ozonesonde data, M07 initially interpolated the data to a fine grid of regular intervals, followed by multivariate PCA, and then cubic spline interpolation was used to get continuous principal component basis functions. (Due to the strength of the dominant modes of variation and smoothness of the resulting basis functions, no additional pre or post-smoothing was used in the pc's - unlike the current PGP13 description of M07). The irregularly spaced ozonesonde data from each flight were then regressed on the continuous principal component derived basis functions to get the vector of PC scores for each flight based on the original irregularly spaced data. Fortunately the Umkehr data you analyze is regularly spaced and you use monthly averages (smaller data set even though more years), so direct FPCA steps for the entire data set may be more computationally feasible to obtain the PC scores in your case.

(b) The SSANOVA models fitted to the principal component scores in M07 actually were initially fitted using GML (related to REML) for smoothing parameter estimation in a mixed model frame-work, as implemented in the assist package (Wang 1998, Ke and Wang 2002) which also depends on the nlme R library (Pinheiro and Bates - already mentioned in PGP13). Some stabilization was then done across months - described in M07.

(c) A few references are provided in M07 to other studies of ozone vertical profiles.- although not necessarily based on pc's.

7. Typographical errors/editing: (in addition to those mentioned by Referee 1)

(a) Add the assumption of independence to the normal distribution assumptions for ϵ_{li} at the top of page 12347 unless auto-correlation is considered.

(b) In equation (A3) and the subsequent paragraph: does each A_j include a row/column for a constant for each term? Is this specific to mgcv? How does this affect the length of α_l? Are the elements of α_l corresponding to zero rows/columns in the S_l matrix identifiable/constrained? Does B also have zero columns/rows? I haven't checked all the details with literature, but please double-check. Please also check the dimensions of α_l. S_l, and B in equations (A2-A4) and discussions, and also the non-singularity of $B^T B + S_l$ required in (A4) with the current definitions, and modify if needed.

(c) Minor: Space is needed within the title of section 4

(d) How is the intercept estimated (line before equation (A2)?

(e) Parentheses should be around everything after "min" in equation (A2).

(f) Also it may help to replace min by "$\underset{\alpha_l}{argmin}$" and indicate (A2) is for fixed lambda vector - possibly adding a lambda value in the subscript here and elsewhere if it will help.

(g) At the top of page 13551, should λ_l vector include λ_{l7} (currently only goes to 6)?

(h) Some sentence editing will help. Due to the length of my other comments I haven't checked all details. Please read carefully for other edits.

(i) Figures 12 and 13 could be placed in 2 adjacent columns. Similarly for Figures 14-15.

(j) Is Table 2 needed? What is added by this? Additional discussion may help if including this Table.

(k) The use of y_i for the normalized residuals may confuse on page 12353 - since y's are used elsewhere with different interpretation.

==========

4 References:

Yuedong Wang (1998), Mixed-Effects Smoothing Spline ANOVA, Journal of the Royal Statistical Society B, 60: 159-174.

Yuedong Wang (1998), Smoothing Spline Models With Correlated Random Errors, Journal of the American Statistical Association, 93: 341-348.

Chunkai Ke and Yuedong Wang - assist R package.

Yuedong Wang (2011), Smoothing Splines: Methods and Applications, Chapman & Hall/CRC Monographs on Statistics & Applied Probability.

5 Thank you for the opportunity to provide comments on this interesting paper.