

Interactive comment on “A climatology of surface ozone in the extra tropics: cluster analysis of observations and model results” by O. A. Tarasova et al.

Anonymous Referee #2

Received and published: 12 September 2007

General comments

"A climatology of surface ozone in extra tropics: cluster analysis of observations and model results" presents a useful new analysis of measured ozone concentrations and applies it to evaluate the performance of a global chemistry transport model. This is potentially a valuable contribution but a few points need further consideration.

Introduction

Some discussion should be included of previous uses of cluster analysis for the analysis of air pollution; both to cluster back-trajectories, (e.g. Cape et al 2000) and local air

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

pollution e.g. Gramsch et al. *Atm. Env.* 40 (28): 5464-5475

Data

Model output: was all the data taken from the lowest model grid box? This should be stated. If it was, it would have been more appropriate for some of the elevated stations e.g. Jungfraujoch to have use an elevated model box. The reason for this is that the orography at 2.8x2.8 degrees is far below the highest mountain stations. This would probably increase the number of output points for the model (there might be a lower and a higher box in the same column) and possibly allow the model clusters to capture the elevated cluster as well. Depending on the amount of work involved in rerunning the clustering this would be worth doing.

Statistical Analysis

I have 2 main questions on the clustering technique:

1) How were the clusters found? Was this done with a brute force method? How were the two clustering rules combined? Please give more details. 2) How stable is the technique to perturbations in the input data e.g. interannual variability, numbers of sites included, random error?

Results

It seems odd to me that 2 of the stations in Anatarctica shown on the map (South Pole and McMurdo) are in cluster 1 not cluster 3. Is there any reason for this?

In the model clusters section you state: "the maximum of the stratospheric contribution in absolute values is observed in spring.". However the peak is in February for most clusters i.e. in Winter. It also seems to precede the maximum in observational cluster 1 which suggests that either the observed spring maximum is not solely dynamical in origin, or the seasonality of STE is incorrect in the model. See also conclusions. It is also possible though that some of this is an artefact introduced by the fact that the lifetime of stratospheric ozone will be longer in winter than in spring as a result of lower

photolysis rates. It may also depend on how the photolysis of stratospheric ozone is treated.

"the model exceeds the observations by less than 8 nmol/mol". The model is lower than the observations. Therefore it is also wrong to conclude that the stratospheric contribution in the model may be overestimated.

Any discussion of the model's contribution from stratospheric ozone should include evidence from other model evaluation exercises. For example how does the annual STE compare to the average found in the ACCENT Photocomp Experiment (Stevenson et al, 2006, J. Geophys. Res., 111, D08301, doi:10.1029/2005JD006338.) and any other model evaluations performed.

Why would the stratospheric contribution be incorrect for one cluster (MC1) and not for another (MC2)?

You suggest that based on the results of model cluster 4 that the role of chemistry is overestimated. An alternative explanation would be that dry deposition in the model is too strong. Also to clarify this sentence it would be better to say that chemical destruction of ozone is overestimated.

Conclusions

I do not agree with the conclusion that because the spring maximum is the same for all times of day that this means its origin is purely dynamical. The modelled peak stratospheric ozone concentrations are earlier in the year than this and if the spring increase in ozone was not local to these sites it would not necessarily be seen in the diurnal cycle of the local ozone.

Technical corrections

Abstract:

The following phrases need to be clarified:

"no dependence of the maximum timing is observed" - This is unclear. It probably should be "no diurnal variation of the seasonal maximum is observed"

"The majority of cases are covered by the regimes with a spring seasonal maximum or with a broad spring-summer maximum (with prevailing summer)." I do not understand what is meant by "with prevailing summer".

Introduction:

"showed that the summer maximum changes stronger than the spring maximum." should be "showed that the summer maximum changes more strongly than the spring maximum."

Statistical analysis:

"For each particular location (measurement site or corresponding model grid cell) 24 averaged over measurement period seasonal cycles were derived representing each hour of the day for the whole measurements/simulation period." This sentence is confusing. "For each particular location (measurement site or corresponding model grid cell), for every month, 24 averages over the measurement period were derived representing each hour of the day for the whole measurements/simulation period."

I would suggest moving the paragraph:

"Unlike the hierarchical clustering procedure, applied here, non-hierarchical clustering (e.g., the k-means algorithm) supposes that the number of clusters is already known and that the objects are distributed between the discrete numbers of the groups (Moody et al., 1991). This algorithm is widely used in those cases where a priori information on the nature of the measurements is available. An example is the classification of aerosol types (Omar et al., 2005). Since we have no a priori information on the number of the particular patterns in our data this method is not applicable here."

to be before the sentence:

"The purpose of the hierarchical clustering is to join objects into successively larger clusters, using some measure of similarity or distance."

The equation for the distance between two objects (equation 1) is confusing as i , j , h and m are not defined. "In this paper a squared Euclidean distance is used as a measure of distance between the objects:"-> "In this paper a squared Euclidean distance is used as a measure of distance between two objects i and j :"

and add after the equation "Where the sum is over all 24 hours, h and all 12 months, m ."

The use of i and j to sometimes indicate cluster centres and sometimes indicate individual cluster members is also rather confusing, please clarify equations 2 and 3.

Results:

New Zealand -> New Zealand

(less average conditions) -> (lower average conditions)

MC3 is never exceeding 35% -> MC3 never exceeds 35%

In comparison with the other observation or model cluster -> In contrast to the other observation or model clusters

the relative contribution of the stratospheric source is reaching 100% -> the relative contribution of the stratospheric source reaches 100%

Figures:

Figure 1. I would rather see all of the contour plots use the same scale here instead of cluster 4 being on a slightly different scale - it makes it harder to compare this cluster to the others.

Figures 5 and 6 need a short explanation in the legend of what they show i.e. the diurnal cycle for every month in turn as this is not immediately obvious.

Interactive comment on Atmos. Chem. Phys. Discuss., 7, 12541, 2007.