Typical Scenarios for the German Tropospheric Composition Based on Cluster Analyses

A contribution to subproject GLOREAM

M. Klein and J. Zimmermann

Deutscher Wetterdienst, FE25, Kaiserleistr. 42, 63067 Offenbach, Germany
monika.klein@dwd.de

Summary

A new project is presented which aims at determining a characteristic set of distributions of air constituents over Germany. This will be achieved with performing a cluster analysis on model results for the tropospheric composition in Europe for the summer periods since 1999. In order to reduce the data set a principal component analysis is performed on part of the data set. The results are promising. However, they are more dependent on the complete diurnal cycles of all parameters than was expected.

Introduction

Until recent times the validation of models for atmospheric chemistry was based on special cases. Concerning chemical mechanisms, smog chamber experiments were used with rather arbitrary compositions of the chamber air. Chemical transport models were usually compared with field experiments, which characterised just episodes. This limits the relevance of the used data sets. Thus there is a need for having more objectively defined scenarios of air compositions e.g. of the troposphere of Germany. Unfortunately long-term routine measurements of trace gases are performed for a very restricted set of species and mostly at the ground. In addition, the measurements are often influenced by local effects and measurement errors. Therefor long-term model calculations may be an alternative way for determining characteristic scenarios of the tropospheric composition. However, the reliability of the model can only be verified by the available measurements. The resolution of the model would be another limitation of this method.

When model runs are used to obtain scenarios for model evaluation, it is necessary to determine characteristic distributions of air pollutants over Europe from large collections of model simulations. Currently the Deutscher Wetterdienst proceeds with daily model simulations between April and October in the area of the LM, the local model for weather forecasts, which covers most of Europe between the Atlantic Ocean west of Ireland and west of Russia and between Scandinavia and Greece. The model resolution is 21 km. The chemical transport model is based on the EURAD model developed in Köln, while the emission modelling was set up by the IER Stuttgart. The model system was developed in the TFS research initiative funded by the BMBF in Germany. More details are published by Tilmes and Rißmann (1999). The comparison with measurements was discussed by Tilmes et al. (1999). A comparison of the model system with other models in Europe is given by Tilmes et al. (2002).

In the following a new research project will be introduced, which aims at the determination of a set of characteristic scenarios. It is part of a co-operation with with the Institute for Atmospheric Chemistry at the Forschungszentrum Jülich, where an atmospheric chamber is operated. In addition, results of this project will be input to research with scenarios at the EURAD group at the University of Köln.

P.M. Midgley, M. Reuther (Eds.)
Proceedings from the EUROTRAC-2 Symposium 2002

Objectives

The main goal of the project is the determination of a set of characteristic scenarios for usage for model evaluation, scenario calculations and for experiments in atmospheric chambers. Further more the aim is to match those scenarios to common European meteorological patterns known as the German term "Grosswetterlagen". This is to be achieved by means of cluster analyses of daily model results since 1999. Prior to the cluster analysis the daily data sets need to be chosen carefully. Therefore a selective procedure needs to be developed to test several types of data sets.

Activities

As a test data set the CTM model results from May 2000 are chosen because this period has 5 different types of Grosswetterlagen. Each day of this period will be classified with cluster analysis.

There is a variety of different types of cluster analysis methods. In this study two of them are to be tested. One method chosen here is a hierarchical method that gives a dendrogram as a result. The objects here are the 31 days of May. In the beginning the objects are in separate clusters, they are iteratively joined to one remaining cluster. The steps of joining the clusters give the information about similarity of the objects.

In a second step the results of the hierarchical method are validated with a K-Means cluster analysis method, which is characterised by the fact, that the number of clusters is given. In this method the objects are clustered in a way that the variance inside a cluster is smallest while the variance between objects of different clusters is maximised.

Furthermore with this method the number of clusters found in the hierarchical method can be matched to the amount of "Grosswetterlagen " found in the test period. This can be used as a direct test of whether and to which extend there are dependencies of chemical scenarios on the meteorological conditions.

The cluster analysis has computational limits that make a data reduction essential.

One important step is to look for dependencies of data in the characteristic chemistry scenarios with principal component analysis. The aim here is to reduce the data set as much as possible.

The days of the test period May 2000 representing a typical "Grosswetterlage" are selected for the test data set. The test data set is used for sensitivity studies with the principal component analysis. The days are the 6^{th} representing a south eastern pattern, the 15^{th} standing for a high over central Europe, the 18^{th} is characterised by a trough over central Europe, the 22^{nd} for a zonal western weather pattern and the 30 representing a cylonal south western pattern. The cluster analysis as well as the principal component analysis is performed separately for 4 height levels that are 700m, 1200m, 3000m and 5000m.

The concentrations of O_3, CO, NO, NO_2, HCHO, the RADM2 lumped hydrocarbon HC3, isoprene, OH, H_2O_2, HNO_3, a parameter that accounts for the age of an air mass referred as AIRAGE is calculated by the ratio of $[H_2O_2]/[HNO_3]$ (Sillman 1995) are selected as chemical parameters. The meteorological parameters temperature and wind components are included in the test data sets as well. The wind components have been taken out in some of the sensitivity studies.

The parameters are spatial distributions of the CTM model grid (109 x 109) for each level. An ideal data set should contain hourly fields of each parameter. This results in a large data set and therefore the aim is to reduce the data set but take the daily variability into account. Therefore 3 different data sets are put together. Two data sets are a day (2PM) and a night

(0 AM) data set. In those data sets each parameter has 4 spatial fields: the daily maxima and minima, the daily variation and the day, night field respectively. The 3rd test data set contains 5 hourly fields at 00 AM, 6 AM, 12 PM, 2 PM and 6 PM for each parameter.

Results

The results of the principal component analysis show a strong dependency of the daily variation of CO and NO_x in the lowest level. The impact of the wind parameters seems to be negligible at lower levels. At higher levels O_3 and CO seem to have the strongest impact. The wind parameters clearly need to be taken into account here. In the higher levels more Eigenvectors are needed to explain 95% and more of the variance which indicates that processes are more complex here and transport of air masses is becoming more significant with height. The night and day data sets do not differ so much in the principal component statistics but in the spatial patterns of the corresponding first principal component

The results are preliminary because it was found that the results strongly depend on the daily variability of some parameters. Therefore the composition of the data sets will be further optimised. The figures show the statistic results of the principal component analysis for the 22nd of May for the three different data sets and the 2 PM- data set without wind components.

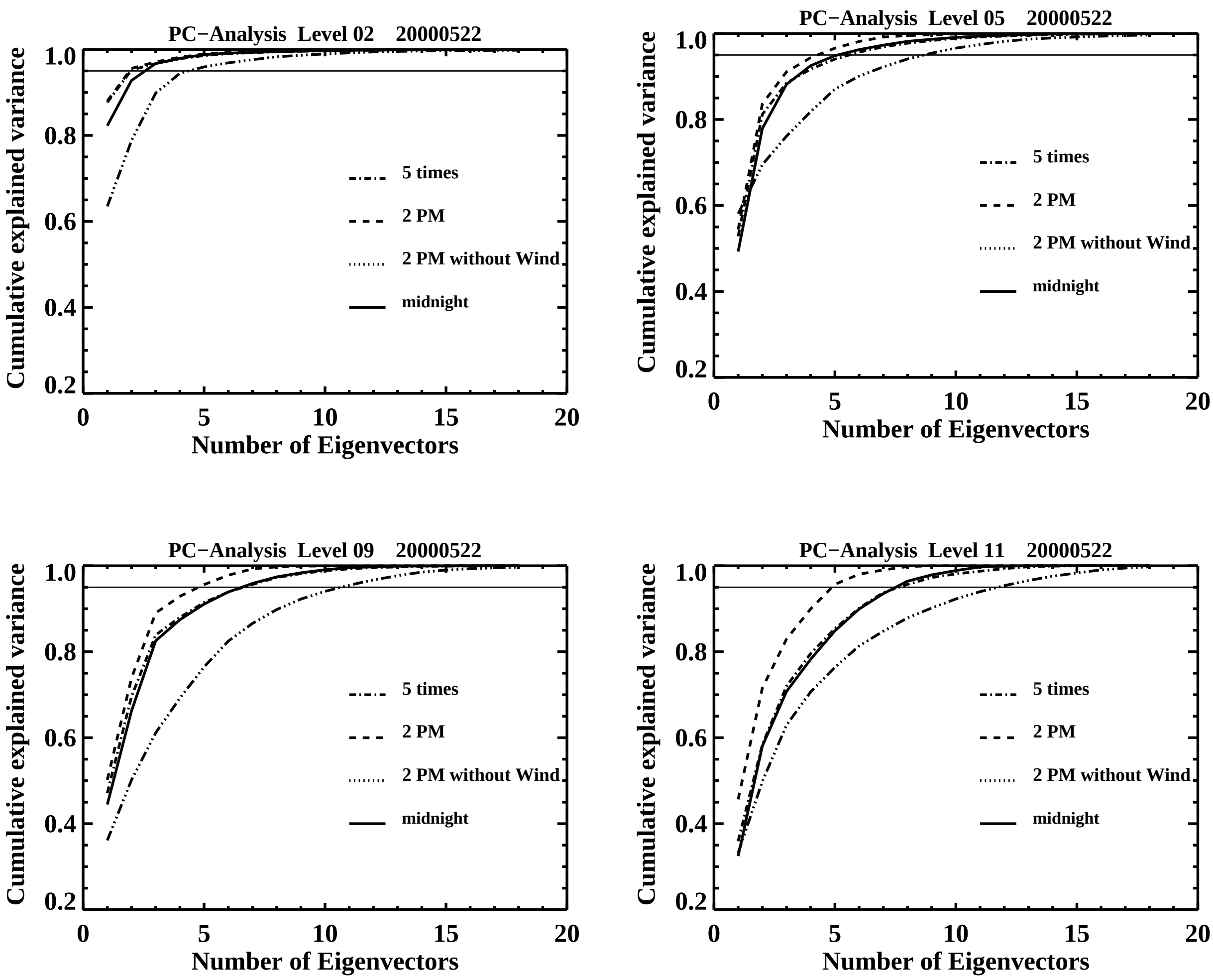

Figure 1. statistical results of the principal component analysis for level 2 = 700 m, level 5 = 1200m, level 9 = 3000m and level 11 = 5000m.

Discussion

The success of this study strongly depends on the selection of the appropriate data set used as input for the cluster analysis. The principal component analysis is a good tool for reducing the data sets but can't provide sufficient information about the significance of the parameters in the data set except for those that can be matched in the first principal component. An additional method of analysis needs to be employed here. Furthermore a high explained variance in the first eigenvectors does not necessarily indicate a good data set since important parameters may be still missing. One important aim is to optimise the selection of hourly fields that are necessary to account for the daily variations of the parameters.

Acknowledgements

We gratefully acknowledge funding by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie of Germany under grant 07ATF02-Nr. 371.

References

Sillman, S. 1995: The use of NO_y H_2O_2 and HNO_3 as indicators for ozone-NOx -hydrocarbon sensitivity in urban locations

Tilmes, S., J. Zimmermann, J. Rißmann, and I. Jacobsen, 1999: Comparability of measured and modelled concentrations of secondary air pollutants. *In Proceedings of EUROTRAC Symposium '98*, (editors: P. M. Borrell and P. Borrell), WITPress, Southampton, 493 – 497.

Tilmes, S., and J. Rißmann, 1999: Forecast of atmospheric trace gases. *Quart. Rep. of the Operational NWP-Models of the Deutscher Wetterdienst*, **20**, 9 – 11.

Tilmes, S., Brandt, J., Flatoy, F., Bergström, R., Flemming, J. Langner, J., Christensen, J.H., Frohn, L.M., Hov, O., Jacobsen, I., Reimer, E., Stern, R., and Zimmermann, J., 2002: Comparison of five Eulerian air pollution forecasting systems for the summer 1999 using the German ozone monitoring data, *J. Atmos. Chem.*, in print.