Reply to reviewer #2

First of all we want to thank reviewer 2 for the constructive comments and the positive evaluation.
Overall – and in particular following advice by reviewer #1 – we reworked the manuscript substantially. Please find changes resulting from specific suggestions of the reviewers indicated in the annotated version of the resubmitted manuscript.

General Reply

Both reviewers raised a number of concerns about the interpretation of our findings, especially when compared to results from earlier studies. We can follow these concerns in many aspects, and we updated our manuscript accordingly (see detailed comments below). However, part of these concerns might be caused by misunderstandings, which should be briefly clarified in the following:

A) Main focus of the manuscript

The main purpose of our study (which was probably not very clear in the original version) is about introducing a **new technique** to a satellite data set of tropospheric pollution, and the **consideration of one new and important variable, i.e. wind direction**. This was pointed out in the original version of the manuscript, but we emphasize this in the revised version more clearly (changes in section 4 and the conclusions).

A1) New technique: Using generalized additive models provides new means for a highly adaptive model formulation in the rapid analysis of large spatio-temporal satellite data. Our study illustrates the great potential of a new approach which surely has to be further exploited in the future – formulating and testing hypotheses about the interaction between observable and other quantities is rather simple (and easy to interpret) when relying on a non-parametric additive model. Specific important new features of the method are that a) the functional relationships are determined from the algorithm itself, i.e., they are *not predefined* but can be *learned* from the data; b) that irregularly sampled data can be used, and c) that the temporal resolution is high (daily instead of monthly data), high enough to study the interaction of NO2 and a rapidly changing observable – the estimated local wind field.

A2) New spatio-temporal variable "wind direction": The focus of our study is on introducing a parameter which is of particular interest when evaluating the spatio-temporal distribution of short-lived atmospheric trace gases -- the wind direction. Including this information not only improves the accuracy of the results of the other influencing parameters, but yields new and exciting information on the influence of transportion processes on the local NO_2 values. Of course, other influencing parameters such as linear trend, seasonal and weekly cycle have to be included in the study as they represent well-known and relevant factors in the analysis of NO2 and - for this very reason – we include them into our model

formulation. But they are not the main focus of our work. We are well aware that several studies on these parameters exist, which partly include more complex data retrieval schemes and/or observations from sensors with higher spatial resolution. We discuss them nevertheless in the present work as we believe that emphasizing similarities and (few) differences between the traditional parametric model and the non-parametric model used here may help to uncover strengths and weaknesses of both approaches. In the revised version of the manuscript we point out more clearly what the main focus of our study is, and how it is related to previous work. We state more clearly that our results on the seasonal cycle should be treated with care (see below). However, we are confident, that the findings on the weekly cycle and the linear trends are sound.

B) Selection of the data set

By the time we designed this study, GOME observations provided the longest global record on tropospheric NO2. Compared to observations from SCIAMACHY, the global coverage of GOME data is better by a factor of two, i.e. the temporal resolution of the time series is twice as high. Today, also SCIAMACHY and OMI provide long time series, and have much higher spatial resolution. OMI data also have daily global coverage. These data sets have a high potential for the application of GAM. We already stated clearly in the conclusions of the original manuscript, that

> *'In the future, increasing satellite data sets with improved spatio-temporal coverage, higher spatial resolution and improved cloud correction will become available. Using such data sets will allow much more detailed studies…'*

Nevertheless, for a prototype study as presented in our manuscript, the GOME data already proved to be very useful. One peculiar advantage is that the spatial resolution (in west-east direction) is very similar to that of the wind data set. In the revised version of the manuscript we make more clear why the GOME data set was used (in the introduction). We also add more information on the potential of tropospheric NO2 data sets from other sensors.

C) Uncertainties and systematic biases of the used data set

As already stated in the original version of the manuscript, a rather simple retrieval scheme was used. In particular, the specialities of the atmospheric radiative transfer in the troposphere were not considered. This was done for simplicity, and because the main conclusions of our study are not critically affected by this procedure. The only aspect, which is substantially influenced by these simplifications is the investigation of the seasonal cycle. In contrast to the phenomena taking place at shorter time scales (such as weekly cycle, and wind influence) or at longer time scales (such as linear trends), many atmospheric parameters change systematically with season. Such influencing parameters are: layer height of the tropospheric NO2 (and aerosols), aerosol properties, surface albedo, and viewing geometry (e.g. SZA). Due their change with season, these parameters also affect the results of our tropospheric NO2 VCD, even if studied on a relative basis.

We addressed this important point in the revised version of our manuscript twofold: First, we added more information in the text. In particular, we make more clear that the results on the seasonal cycle might not only contain the signal of the tropospheric NO_2 VCD, but also that of other influencing parameters. Second, we introduced a new sub-section on the effects of atmospheric radiative transfer on the retrieved tropospheric NO2 VCD (section 2.1). We also quantify the systematic biases of our GOME data set on tropospheric NO2.

Reply to specific questions and remarks

R2.1: *Although I do not fully understand how a general additive model works (how it determines the functional form of each term), the results appear to be reasonable.*

Reply: We have added some information on spline fitting into the section 3.

R2.2. *p. 9370, lines 19-20. I would argue that there is more than a "little prior knowledge" available concerning NO2 – from previous analyses of satellite data and from chemical transport models. Please either remove or modify this statement.*

Reply: Has been changed.

R2.3: *p. 9373, lines 12-16. What about variation in the tropospheric NO2 profile shape? Shouldn't this be important for the AMF? Shouldn't some assumption from a chemical transport model be used in specifying these shapes?*

Reply: We agree with the suggestion of the reviewer that additional information (e.g. from a chemical transport model) could minimize these shortcomings. However, in this prototype study for the application of GAM on satellite data sets we decided not to include such a correction for simplicity. As we already pointed out in the last version of the manuscript, we only use stratospheric AMF. Of course, the radiative transfer in the troposphere is much more complex compared to that in the stratosphere, and in principle additional corrections would be required. Fortunately, most of the results of our study are not affected by this simplification, as we are on the lookout for relative changes (over time, with the wind). However, especially the results for the seasonal cycle might be influenced by other parameters, such as the NO_2 layer height, which also show a systematic seasonal variation. We added more information on these effects in the revised version of the manuscript. In particular we added a new subsection (2.1), which discusses the atmospheric radiative effects an the retrieved tropospheric NO_2 VCD.

R2.4: *p. 9374, line 24: What is R?*

Reply: R refers to the set of real numbers. In the manuscript, we now used the more conventional mathematical symbol (latex: \mathbb{R}) for this.

R2.5: *p. 9377, lines 21-24. Obviously, some knowledge of the behavior of NO2 is required for choosing these functions. The authors need to acknowledge that fact.*

Reply: Has been done

R2.6: *Somewhere in the manuscript the authors need to explain a bit more in layman's terms how the model "learns" the relationship between the data being analyzed and the explanatory variables.*

Reply: Has been done.

R2.7: *p. 9379, lines 21-22. Not sure what this sentence means.*

Reply: We reworked the sentence

R2.8: *p. 9379, line 24 through p. 9380 line 4. It is unclear from this paragraph as to whether the authors are advocating parametric or non-parametric approaches for the annual cycle.*

Reply: We reformulated this chapter, emphasizing (at the end of the chapter) that we prefer the nonparametric approach for modelling the annual cycle.

R2.9: *p. 9380, line 19. change high to long*

Reply: Has been done

R2.10: *General comment; In general the English is fine, but there are some sentences throughout the manuscript that are a bit awkward and would benefit being edited by a native English speaker.*

Reply: Has been done