

## Reply to reviewer #1

First of all we want to thank this reviewer 1 for the very detailed and constructive comments. They were extremely helpful in reworking and – hopefully - improving the manuscript.

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 NO<sub>2</sub> 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 transportation processes on the local NO<sub>2</sub> values. Of course, other factors 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 NO<sub>2</sub> 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 NO<sub>2</sub>. 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 NO<sub>2</sub> 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 NO<sub>2</sub> (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 NO<sub>2</sub> 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<sub>2</sub> 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 NO<sub>2</sub> VCD (section 2.1). We also quantify the systematic biases of our GOME data set on tropospheric NO<sub>2</sub>.

## Reply to specific questions and remarks

**R1.1** 1. *Abstract and Introduction: Authors refer to multiple sensors even though they only consider GOME observations. Please delete this misleading text.*

**Reply:** Although the integration of NO<sub>2</sub> measurements from multiple sensors may be a very interesting extension of the current work indeed, our intention was to point out that in recent years the observation of tropospheric trace gases using space-borne sensors has seen a very rapid development.

To avoid confusion, we reworked the sentence.

**R1.2** 2. *Introduction: first paragraph has old and outdated references, please update*

**Reply:** We updated the reference to the World Health Organization (now 2005) and removed the reference to Elsayed, 1994.

**R1.3** 3. *missing discussion of Boersma et al. 2008 and Boersma et al. 2008 (acpd) for diurnal and weekly cycle*

**Reply:** We now cite Boersma et al. (2008, 2009) in the introduction and discuss their results shortly in section 4.

**R1.4** 4. I. 9-13: *although it appears to be a crucial point, it is very unclear*

**Reply:** We expanded this paragraph significantly, now providing more details.

**R1.5** 5. *While the authors provide (too?) much detail about the GOME instrument, they talk very little about the quality of data, associated error, bias, loss of data due to cloud cover etc. etc. Much more information is needed here.*

**Reply:** We now provide more information on systematic and random errors of the GOME data set. We also added a new sub-section detailing on the atmospheric radiative transfer in the troposphere and cloud coverage (section 2.1).

Please also see point C of the general remarks above.

**R1.6** 6. *The resolution of ECMWF met fields is very coarse. Also, it is not clear why 24h average is better than, say 6 to 12. Again, more details and justification is needed.*

**Reply:** We now provide details on the ECMWF data in a separate sub-section (section 2.2) and we added more information/justification for our choice.

**R1.7** 7. *The use of  $R$  notation to denote time dimension that varies with each footprint is incorrect or at least unnecessary.*

**Reply:** We now only discuss the observations of the individual pixel, i.e. the individual time series, where this notation is valid.

**R1.8** 8. *The inconsistencies in using  $Y$  and  $X$  (and bold  $Y$  and  $X$ , bold  $y$  and  $x$ , bold italic  $y$  and  $x$ ) need to be fixed.*

**Reply:** Thank you for pointing out these inconsistencies. We reworked the notation.

**R1.9** 9. *Make equation  $y = \eta + \varepsilon$  into equation (1)*

**Reply:** Has been done.

**R1.10** 10.  *$x$  and  $f$  subscripts  $(i,j,k)$  are highly confused and at times incorrect. It is not clear if they correspond to  $n$ ,  $m$  or  $T$  ranges. Please clarify.*

**Reply:** Has been done.

**R1.11** 11. *Appreciating authors' efforts to be as general as possible, I find description of the functional form unnecessarily vague and unclear. I suggest starting with an example like  $f_{ann}$  etc. It would also be helpful to say what e.g.  $f_{rain}$  would correspond to in  $x$  space, amount of rain?*

**Reply:** We reworked this section. We now start with the  $f_{ann}$  example and detail on how to novel terms such as  $f_{rain}$ .

**R1.12** 12. *I. 23-24 but  $X$  is not observed (p. 9375)*

**Reply:** Has been changed to "data pairs".

**R1.13** 13. *I. 5-15 (p. 9375): it's not clear if this is all in GAM.*

**Reply:** We assume, the reviewer is referring to the part of the paragraph beginning with "a binomial distribution in a binary detection task, or a Poisson distribution when measuring rare events", then detailing on the additive components and ending with "The individual model terms  $f_j$  are assumed to be independent: a seasonal model component  $f_{ann}$ , for example, would be assumed to be independent from a weekly trend  $f_{week}$ , the wind direction  $f_{wind}$ , or the amount of rainfall  $f_{rain}$ ." ?

Generalized additive model can indeed be employed under the assumption of several noise statistics [e.g. Wood 2006, or <http://stat.ethz.ch/R-manual/R-patched/library/mgcv/html/gam.html> for short].

The reviewer is right, however, that ‘independence’ in its statistical definition is only assumed for the realization  $Y$  of the modeled process. Such a strong assumption is not required for the model terms. In (Hastie and Tibshirani, 1986), it is reported, that the sum of the model terms, resulting from the backfitting algorithm, is unique, but the model terms themselves are not, since “dependence among the covariates can lead to more than one representation for the same fitted surface.” We will describe this in this way in the manuscript and we will refer to (Hastie and Tibshirani, 1986).

Raising the question of ‘what is GAM’ we now also clarify on differences between the generalized additive model, nonparametric model terms, and cross-validations, which are part of the modeling process, but not necessary of the GAM itself.

**R1.14** 14. Equation on l. 3 (p. 9376) has very unclear dimensionality

**Reply:** Has been changed.

**R1.15** 15. l. 2 (p. 9377) “qualitative search” is too vague of a term to describe what the model does, please improve the description **and R1.16** 16. l. 4 (p. 9377) same as above “irrelevant variable” -- how is that determined? Small  $b$ ?

**Reply:** A ‘qualitative search’ for relevant model components is what we recommend the operator to do once he/she has the quantitative output of the model. This may be inspection of the p-values (Fig. 3), or by evaluating the model terms visually (Fig. 2).

Detail on how the p-values are determined from the data (by the algorithm) were given in lines 5-19 on the same page. An example on how to ‘read’ model terms is given at the beginning of the Result and Discussion section.

To clarify on this and to provide some guidance to other users of GAM, we reworked the paragraph.

**R1.17** 17. l. 18 (p. 9377) again, I thought observations were only of  $Y$ , not  $X$

**Reply:** Has been changed.

**R1.18** 18. l. 11-15 (p. 9378) Please elaborate on how the approximation of considering only surface winds affects the results, how much error does it introduce.

**Reply:** We detail on this now in the manuscript. We argue (in chapter 4.4) that the surface winds are in high correlation to winds up to altitude levels that are affected mainly by NO<sub>2</sub> from pollution sources.

**R1.19** 19. l. 16 (p. 9379) *Fig 3. does not show p-values*

**Reply:** Figure 3 shows “log<sub>10</sub> of the p-value”, i.e the *logarithm* of the p-value, as stated in the figure caption. We now state this verbatim.

**R1.20** 20. l. 5 (p. 9380) *“vary in space on large scales only”: what does that mean? RContinental scales? Here and elsewhere please be specific, otherwise the reader learns nothing.*

**Reply:** Has been changed to “vary in space on continental scales mainly”.

**R1.21** 21. Section 4.1: *I do not find this section convincing. Why would Indian Ocean region experience stratospheric correction errors, but not other places? We at least need a reference for that.*

**Reply:** The reviewer is right. A more convincing explanation has been presented by: Kunhikrishnan, T., M. G. Lawrence, R. von Kuhlmann, A. Richter, A. Ladstätter-Weissenmayer, and J. P. Burrows (2004), *Semiannual NO<sub>2</sub> plumes during the monsoon transition periods over the central Indian Ocean*, *Geophys. Res. Lett.*, 31, L08110, doi:10.1029/2003GL019269  
We now discuss this in the text.

**R1.22** *Does your further analysis of wind speed and direction confirm the findings in paragraph 4 of this section?*

**Reply:** Yes. We noted this in the original version on page 9380. We now emphasize this in the reworked manuscript.

**R1.23** *Paragraph 5 describes Tran Siberian railway pattern which is nowhere to be found in Figure 5.*

**Reply:** We have expanded the graphic northwards. Now the winter maximum is visible.

**R1.24** *Paragraph 6 brings no new science. On l. 2 (p. 9381) it is not clear what differences the authors are referring to. This section needs to be rewritten.*

**Reply:** As discussed at the beginning of this letter, the seasonal cycle is not the main

focus of this work and we are well aware of earlier studies addressing the annual cycle in full depth. Please see our general comments above (A2).

However, as there are some differences in our data processing, and the data processing of these earlier studies, we have rewritten this section and now discuss systematic errors. We also discuss systematic differences to the paper of van der A, 2008. (In our studies, the region in western part of the USA showing an annual signal with a maximum in summer is significantly more extended to the east than it is shown in the paper of van der A, 2008.)

**R1.25** 22. *Last paragraph of section 4.2: Authors claim agreement where they admit there is no agreement. If there is, please be more specific (region/time/trend etc). Also how about Stavrakou et al. 2008?*

We reworked the paragraph providing these details and citing the reference.

**R1.26** 23. *Section 4.3 does not mention the very obvious swath pattern in Figure 3 in the weekly plot. How is that affecting results?*

**Reply:** We now mention the swath pattern in the manuscript, and could explain it by a sampling artifact due to the 35 days repeat cycle of ERS-2 (see also Beirle et al., 2003).

**R1.27** *Third paragraph sounds particularly unconvincing in the light of the swath pattern and 3 day global coverage.*

**Reply:** The swath patterns the reviewer is referring to can be found primarily in regions of lower latitudes and with a coverage of about three days. Most of these regions are free of anthropogenic NO<sub>2</sub> sources, and the significance of weekly cycle – i.e. the significance of the artificial swath pattern -- is relatively low.

Most of this is different for the ‘weekend-dip’ of the West-European NO<sub>2</sub> plume moving eastward: The observed region is at relatively high latitudes, and the coverage is at a higher rate than three days. The region does not show a swath pattern, and the signal of the weekly is much more significant (Fig. 3). A further argument is the systematic behavior of the appearance of the minimum during the week and the fact that in between the swath pattern, we observe large regions without a contribution of the weekly term, different from the large connected region in Fig. 8 which is spanning over 1400km. We comment on this now in the manuscript.

From a more conceptual point of view one should note the very difference of tracking single “negative” plume from Sunday to Thursday – which may be in fact impossible given the temporal and spatial coverage -- and the averaging behavior of the GAM. So, Fig 8 represents the average pattern of several hundred plumes over the whole time of observations.



**R1.28** 24. *Section 4.3: Boersma et al. 2008 have showed this already.*

**Reply:** Please see our general discussion at the beginning.

**R1.29** 25. *Section 4.4: Inconsistent and wrong spelling of Hong Kong, please fix. Also not clear what region is “Arabia” referring to, Arabian Peninsula?*

**Reply:** Has been changed.

**R1.30** 26. *Section 4.4: Brings no new scientific insight.*

**Reply:** Given the interdisciplinary character of many studies in the atmospheric sciences – which is also reflected in the interdisciplinary background of the authors of the given study -- the definition of general “scientific interest” may be difficult.

Appreciating the reviewers detailed comments on the formal, algorithmic background of this study -- which greatly help us in improving the respective sections of the manuscript -- and his/her valid advice to expand on the description of data processing and accuracy – an aspect which is all too often neglected in the presentation of many studies -- we may have slightly different scientific interests here.

We may summarize the innovations reported in section 4.4. as follows:

- The wind term increases the accuracy of a model with (linear, annual, weekly term) as used by, for example (van der A et al. 2008),
- We find highly consistent flow fields spatially correlating with the NO<sub>2</sub> time series. Tracking these flow fields allows to estimate the areas influenced by sources. We may leave the question of how these areas correlate with average transport open for further (simulation) studies.
- We can uncouple the part of NO<sub>2</sub> signal correlating with wind and – possibly – with short-term transport processes from locally generated NO<sub>2</sub>.

Acknowledging the reviewers difficulties to appreciate these – from our perspective – relevant aspects, we now point them out more clearly in the revised manuscript.

**R1.31** 27. *Figure 7 is never mentioned*

**Reply:** We now mention and discuss the figure in section 4.4

**R1. 32** *Figure 4: What is the white area? Insignificant p values?*

**Reply:** Yes, it is. We now provide this information not only in Figure 3, but also in the caption of figures 4 to 7.

**R1.33** 12. *Figure 2: use actual date instead of day for x-axis*

**Reply:** Has been done.

**R1.34** 14. *Figure 5: Group colors more by seasons, e.g. all green shades = summer (JJA)*

**Reply:** We followed the reviewer's suggestion and indicate the seasons by shades of blue (Winter), green (Spring), yellow/red (Summer) and purple (Autumn).

**R1.35** 15. *Figures 1, 3, 4, 10: use fewer colors (e.g. 20) to allow easier identification of features*

**Reply:** We tested different color scales. We found that reducing the number of colors obscured the spatial features we were interested in. For this reason we decided to keep the old color scale.