The authors present an interesting study on the so called “climate penalty” i.e. the dependency of ozone concentration on temperature. They have analyzed a large observational dataset from different measurements station in Germany and estimated the changes of the climate penalty over time. Additionally they tried to interpret the data by using statistical method (general additive methods, GMAs). The manuscript is clear and well written.

Despite that, it is still unclear to me what the real goal of this manuscript is. If the authors are trying to explain the causes of climate penalty changes (at least for these stations), I believe the manuscript fall short of its objective. I have the impression that the authors overlooked (voluntarily or involuntarily) the importance of many factors for a clear explanation of the climate penalty.

Here below I have listed the main concerns, followed by more specific comments, as well as few technical corrections.

Main comments:

VOC contribution: I generally miss the consideration of VOC in this paper. Together with NOx, VOC determine the dominant chemical regime which is relevant for interpreting the changes of ΔO3 in response to decreasing NOx. For a specific time, temperature is likely a good approximation for VOC concentrations as described in the paper, but when comparing two different time periods it is important to consider the change in VOC and the resulting VOC/NOx ratio. I think some of the open questions in this paper could be solved with an evaluation of VOC data. If it is hard to find VOC data: I know most of the measurement sites in Germany measure benzene which could be analyzed as a surrogate for anthropogenic VOC for the relative development over time

At the urban sites it is pointed out that a VOC limited regime is dominant (at least for most NOx concentrations and temperatures). However, this doesn’t fit together with the abstract where it is said that ‘lowering NOx concentrations resulted in decreasing O3 production rates’. Decreasing NOx concentrations only lead to a decrease in O3 production in a NOx limited regime (Pusede et al. 2012).

I do additionally wonder if any changes in biogenic emissions could be expected. Although temperature (as mentioned before) can be used for VOC concentrations, I wonder if we can consider this dependency stable over the entire observed period (or even within the two periods in which the data are divided). Maybe the absolute changes are not significant, but almost certainly is their relative impact in comparison with anthropogenic VOC emissions.

Figure interpretation: Although it does make sense that the temperature sensivity of O3 production decreases for decreasing NOx, I don’t think the Figures show this. For the urban site in RP (Rhineland Palatinate) I get the impression that the slope in Figure 5 right (and therefore the climate penalty) is higher for the second time period, which would indicate the opposite.

Results representativeness: The title of the paper might be a bit too general considering that the analysis focuses on a small number of locations in Germany. Again, the response of O3 to NOx reductions can have different outcomes depending on VOC and the dominant chemical regime which varies strongly with the considered location

GAMs model: It would be great if a little more background information could be provided on how the calculations were made and how the model performance was tested. As mentioned later, the model description could be moved entirely to the electronic supplement and also enlarged to include a more exhaustive description, while on the manuscript only a small summary could be shown. Importantly, the code availability and the data source should be mentioned (best if the code could be uploaded as supplement as well).

Specific comments

Page 1, line 11 : nice that with the GAMs analysis you finds exactly the same results as before (see line 6) (i.e. „decreasing sensitivity of temperature" in the second period). Instead of repeating the same sentence as a new finding, you could add "the GAMs model confirm that…"

Page 1, Line 14 : Please consider the main comments: of course NOx concentration is not the only factor influencing the climate penalty. This sentence is indeed too general for explaining your 23 pages work!

Page 1, line 20 : (Sillman, 1999) : The authors of this paper shown in the references are Pusede et al. (2015)

Page 1, lines 23-24 : What about anthropogenic VOC? Many are temperature-dependent, too. Compare Pusede et al. 2014 (also relevant for the summary)

Page 2, line 33-34 : Jing et al. (2017) state a value of 0.43 ppb K^{-1} by which the climate penalty was lower in 1999-2007.

Page 3, line 24 : '… variability of O3 production can explain a considerable proportion…' Do you have any reference?

Page 4, equation 1 : Please explain all variables and what the function means. To me it seems a linear regression (as you mentioned a linear model) applied for the 2 periods. Do we really need to define a mask (P) and 2 equations to explain that? I have the impression that here the readability has been somehow compromised and made difficult to follow.

Page 6, line 9-31 : The description of the GAMs model is indeed difficult to follow. Most possibly is because the needed background in formations are quite large. Although the normal reader would just trust the results, I think that the results of this manuscript should be reproducible by anyone interested (I think that ACP has also a policy on that, if I'm not wrong). Therefore I would strongly suggest to move the GAMs model description to the supplement, with the goal to expand it and make it intelligible by anyone interested. Further, in the electronic supplement of the model the code used for the analysis (i.e. the GAMs algorithm) should be make publicly available, as well with an indication where to download (or obtain) the station's data.

Page 7, line 4 : Why did you need to fit the GAMs model if you build exactly from the dataset? I thought this information to be already used in the construction of the model.

Page 7, line 19 Figure S1: I find this plot hard to analyze. Maybe it could help to add a black stroke color to all data points.

Page 7, line 26 Figure 3: For Figure S3, I like that the differences between the two periods become clear on first sight, maybe this could be added to Figure 3 for a better understanding.

Page 8, line 11 Figure 4: Could you make a comparison plot similar to Figure S3 for Figure 4, too?

Page 8, line 4 : "A final model including three interaction terms" : Again, the description of the GAMs model was not exceptional. Probably would not hurt if you list again these terms here.

Page 9, line 8: '…implying a VOC-limited chemistry'

If I understand the plot correctly, for DERP025 for low NOx and high temperatures the chemistry is NOx-limited for example at T=24°C for the first time period 1999-2008, ΔO3 increases from 5 to

6 µg/m³ and then decreases again, so the change from VOC to NOx limited chemistry would occur at c(NOx)=20 µg/m³. For the second time period 2009-2018 and the same temperature T=24°C the regime change occurs already at c(NOx)=15 µg/m³. For DEBE034 it looks like an even smaller part of each plot shows a NOx limited regime. I think this could be a result from lower NOx at DEPR025 compared to DEPR034 and a more effective NOx reduction over time (as shown on the right). Maybe this could be interesting to mention or to further analyze.

Page 9, line 9 : 'We found a stronger temperature dependence of ΔO3 in the first period…'

In Figure 5 left, I have problems to identify this, e.g. for DERP025 for c(NOx)=20 µg/m³, for a temperature increase from T1=20°C to T2=24°C, ΔO3 increases from 2 to 6 µg/m³ for 1999-2008 and from 0 to 6 µg/m³ for 2009-2018, same for higher NOx. Figure 5 right, too, doesn't support the stronger temperature dependence in the first period: For Rhineland Palatinate I would even say that the correlation of temperature and ΔO3 is larger for the second period when looking at the ΔO3 for a set temperature interval, for example 20-25°C: For 1999-2008, ΔO3 increases by approx. 6.5 µg/m³ and for 2009-2018, ΔO3 increases by approx. 7.5 µg/m³.

For Berlin, the temperature dependence could be slightly stronger for the first period, but the difference is marginal and hard to tell from the graph.

Page 9, lines 21-27 : What about the influence of VOC? If you look at a certain time the influence of VOC can be well represented by temperature, but between the two different time periods emission controls have decreased VOC. For the shown temperature interval and the mean NOx concentration the dominant chemical regime is VOC limited. If NOx reductions exceed VOC reductions O3 increases which is shown by the prediction in Figure 5 left. However, if VOC reductions are larger than expected (and if I get this right the prediction only considers a change in NOx, not in VOC), this would counter run the O3 increase made by the prediction (because decreasing VOC decrease O3 in a VOC limited regime, Pusede et al. 2012) and would yield a lower curve than the predicted one. So maybe VOC reductions were more effective in Rhineland Palatinate than in Berlin – just an idea.

Page 10, line 2-3 : 'This indicates that the ΔO3 sensivity to higher NOx at moderate and high T is lower in the second period.'

…or indicates that VOC reductions at RP were more effective. (which would be consistent with the deviating prediction line for Rhineland Palatinate in Figure 5 right)

Page 10, line 5 : '… temperature dependence of ΔO3 is stronger in the first period…'

I do see this here for the rural sites, but not for the presented urban sites. Higher NOx in urban areas could be an explanation here.

Page 10, line 12 : '… could partially explain the decrease in ΔO3 estimates …'

In Berlin, for high temperatures a NOx limited regime is dominant for any NOx concentration. For lower temperatures a VOC limited regime is dominant. I think this is a good explanation for the observed plot course in Figure 8 left. The red and the blue plot intersect at around 17°C which is approximately the transition temperature for the observation of a NOx (above) and a VOC limited regime (below). For lower NOx during the second period ΔO3 should therefore be lower in the NOx limited regime at high temperatures but higher in the VOC limited regime at low temperatures.

Page 10, line 25 : Figure 10. For high temperatures, both periods in Berlin and the earlier period in RP are dominated by a NOx limited chemistry which is also shown by Figure 8 left. Consequently, ΔO3 increases with increasing NOx. For the second period in RP a VOC limited regime is dominant and ΔO3 decreases for increasing NOx. I find this surprising considering the shown decrease in mean NOx in

Figure 8 right from the first to the second period. Could there be an explanation? What about VOC concentrations at these sites?

Page 11, Section 4.3.2 : Which parameters could have changed over time so that the plots look different for the first and the second period particularly for RP and why not for Berlin? It might be better for the understanding of the results if the explanation of VPD was included a bit earlier, maybe at the beginning of this paragraph.

Technical corrections

Page 1, line 22	varies
Page 2, line 13,22,24	the United States
Page 2, line 21	photochemical
Page 3, line 7	maximum
Page 5, line 5	the latter
Page 9, line 10	the second period
Page 9, line 21	shown
Page 10, line 1/36,26	decrease
Page 10, line 2	indicates
Page 12, line 2	decrease
Page 12, line 11	lead