Answers to Referee #1 comments submitted 20 Oct 2022

Referee comment:

I found the final step to summarize the trends is somewhat disappointing and discouraging ("The absolute/relative trends are calculated as described in the first paragraph after Fig. 21, i.e. by just using the start and end values of the trend"). All those careful considerations and model settings regarding the nonlinearity, emissions, and meteorologies, but at the final step ONLY the start and end points (1 Jan 2005 and 31 Dec 2019) are used to summarize the trends. This is essentially implying no matter what sort of nonlinearities in the data, as long as the (fitted) start and end points are identical. they will have the same summarized trend values, regardless of the autocorrelation, data variability, sample size, etc... and all those factors deem to be important for trend detection.

"We agree that such absolute/relative trend values are most representative as a trend over the whole period when the trend is linear and less so when the trend changes abruptly near the end of the period."

This should not be a justification since the whole point of this paper is that GAM is adopted for handling the nonlinearity. As demonstrated in their results, the air quality analysts often need to deal with time series data from hundreds or thousands of monitoring stations at once. Under this circumstance it is not possible to inspect each time series in detail separately, the same model formulation will be specified for all stations and in the end merely the final linear trends will be reported (e.g., Fig 2 in the paper). How can the practitioners make sure the results are appropriate? Especially no standard errors associated with the relative trends are reported.

Author's response: In Section 3 of the main paper, we calculate and report the *relative change in the expected concentration level at each station from 2005 to 2019.* We do this by using the start and end values of the meteorology-adjusted non-linear trend curve over this period since these values represent the expected concentration levels at the start and end of the period. Section 2.2 in the main paper describes the precise definition of these expected concentrations in connection with the calculation of the physically interpretable trend curves. We acknowledges that the way we wrote this could easily be misunderstood to mean that we were reporting the overall trends themselves. We have therefore corrected the text in Section 3 to make this clearer.

The trends, as such, are non-linear curves over the period. These curves are shown for each station in the plot files

`<sname>_gam.trend_metadj_2005_2009.png`

where `<sname>` denotes the station name, and the trend curve values are given in the text files

`<sname>_gam.trend_metadj_2005_2009.txt`.

In addition, 2.5 and 97.5 percentiles are given in this latter file for each trend curve value, forming a 95% confidence interval for this value. Thus, they describe the uncertainty associated with the estimated trend pointwise through the period.

All these non-linear trend results can be found in the data repositories accompanying the paper. However, it isn't easy to summarize their non-linear nature concisely. This is why we decided on only reporting the overall absolute and relative changes in the expected concentrations based on the non-linear trend curve at each station over the period.

It is possible to use the 95% confidence intervals for the start and end trend values to define an approximate 95% confidence interval for the corresponding absolute or relative changes in the expected concentrations at each station. However, due to the very large number of stations, we decided to leave this out since it would have required almost doubling the number of plots.

Note that, based on the output from AirGAM it is perfectly possible to report several absolute and relative changes in expected concentrations over parts of the period, e.g., from say 2010 to 2014, or 2015 to 2019, etc. Such results over parts of the period can be easily calculated based on the output of the data for the non-linear trend curves. This would not have been possible if we only used regression with a linear trend.

That we use only the start and end of the trend curve is done just to simplify the presentation of the total change over the time period for the reader. If the non-linear trend estimated by AirGAM is undulating there is no simple way to summarize that quantitatively. One solution could be to estimate the total change over the period as done now and possibly combine that with a qualitative description of the trend. We admit that this could be a challenging task when studying a large number of stations, but we see this as a challenge for the data analyst in the post-processing due to the complexity of the trends themselves and should not be seen as a critic against the GAM method itself.

Author's changes: Rewritten parts of Sections 3.1.1 and 3.2-3.4 in the main paper to clarify that we are reporting the absolute and relative changes in expected concentrations based on the meteorology-adjusted non-linear trend at each station from 2005 to 2019. The changes are marked in red.

Referee comment: l.163 is a "linear" regression model linking... to one or more explanatory "nonlinear" variables.

Author's response: No, our GAM models are non-linear regression models. As explained above, we calculate the *absolute and relative change in the expected concentration level at each station from 2005 to 2019* by using the start and end points of the meteorology-adjusted non-linear trend. It does not mean that we assume in any way that the trend in itself is linear over the period. Thus, the absolute or relative change in expected concentration is not the result of performing regression assuming the trend to be linear.

Author's changes: No corrections made.