Dear colleague.

Thank you for your review received 24/12/20. I am of course very happy to respond to the points raised and hope the added justification and clarity of approach significantly improves the manuscript.

*Overall Comments: The analysis needs to be extended as the whole forecast workflow to understand the human impact on NO2 concentrations and forecasting seems to be procrustean to Facebook's Prophet model.The work that needs to be done for this contribution is major, as relevant forecast diagnostics need to be included, regardless of whether they are implemented by Facebook or not.The novelty of the paper is the workflow that involves resolving multiple data sources, that is then used in forecasting and clustering.*

*The workflow does not exist in the write-up (methodology section). In addition, exploratory data analysis to drive the type of model used (referred to as vanilla without any mathematical description) needs to be provided. Authors need to provide motivations behind models they select and this is not possible without exploratory data analysis. Missing are general approaches such as partial autocorrelation analysis of time series, methods used for resolving different spatial scales of input data,the type of forecast method used (including the trend and seasonality model), and why Prophet model is chosen.*
**Response:** I hope the additional material added to the manuscript, in response to the specific points also repeated below, are sufficient. It is unclear why the reviewer is referring to '*different spatial scales of the input data*'. All of the environmental and mobility data are from fixed sites, or point source measurements. The domain setup of the EMEP model is described. Whilst we add some background text on sources and behaviour of NO2 for ease of reading and to better understand the rationale, the definition of the AURN site type is pre-defined by local authorities in order to meet the requirements of the review and assessment process for Local Air Quality Management.

*A new section on exploratory data analysis with concrete mathematical motivations to use a specific model is a must for this paper. Prophet model needs to be explained in the context of the problem.*
**Response**: The new manuscript will include a description of the Prophet framework. It will also include a brief background summary of the known sources and sinks, thus seasonal nature of NO2 for the non-atmospheric reader. It will also include a presentation of auto-correlation plots from a subset of the sites studied, linking to an archive of plots for all sites.

*Comments on Sections:*
*2.1 needs to be extended. In its current form this section is not informative about the model being used. The prophet model has many models, pre and post processing steps. Models used in the study need to be clearly discussed.*
**Response.** Yes I agree, this was an oversight in referring to the Prophet documentation alone. I have added a description of the model used.

2.2 Similar problem as 2.1 This section contains an overly abridged description of the method followed by details regarding application. This section needs to be rewritten.The methodology requires a workflow section that describes the overall workflow built for this study.
**Response**: This has now been included.

Table 1 shows differing spatial resolutions, authors need to clearly present how this disparity is resolved (interpolation, sampling, etc.)
**Response:** It is unclear what additional analysis we could offer, or is to be expected, regarding Table 1, but perhaps additional text may help the reader. Table 1 lists air quality measurement sites managed by the automatic Urban and Rural Network (AURN) in the UK. The network has grown in response to legislative, scientific, technical and policy requirements over time. The classifications provided in Table 1 are not dictated by the authors of this manuscript, rather have been assigned by local authorities in order to meet the requirements of the review and assessment process for Local Air Quality Management. Indeed, In order to ensure the UK's compliance with the EU air quality Directives, the specific categories listed in table 1 are used and are based on known sources and dispersion of key pollutants. In the revised manuscript I suggest the following text is added in section 2:

"*Table 1 lists air quality measurement sites managed by the automatic Urban and Rural Network (AURN) in the UK. The network has grown in response to legislative, scientific, technical and policy requirements over time. The classifications provided in Table 1 are not dictated by the authors of this manuscript, rather have been assigned by local authorities in order to meet the requirements of the review and assessment process for Local Air Quality Management. Indeed, In order to ensure the UK's compliance with the EU air quality Directives, the specific categories listed in table 1 are used and are based on known sources and dispersion of key pollutants*."'

2.4 This section contains no information on data. If the only purpose of this section is state the changes to mobility patterns after March 12th, this can be done in the section on data sources.
**Response**: I agree. This has been changed.

The prophet model is a general additive model for time series, where seasonanlity, trend, and cyclicity are linearly superimposed. There is no motivation in the paper why such additive model will be appropriate for NO2 concentrations. It is customary to provide at least a partial auto corelation function plot of the signal in question and performing statistical tests on these components to understand whether an additive model would be appropriate to start with.
**Response:** The new manuscript will include a brief background summary of the known sources and sinks, thus seasonal nature of NO2 for the non-atmospheric reader. It will also include a presentation of auto-correlation plots from a subset of the sites studied, linking to an archive of plots for all sites.

Specific Comments:
Line 2: Missing comma after In this study
**Response:** This has been corrected.

Line 12: “the nature of local traffic”: artificial constructs do not possess natural qualities.
**Response**. This is irrelevant. To comment on the nature of something is to comment on its basic or inherent features, character, or qualities.

If you mean historic patterns, please write so.
**Response**: This has been clarified.

Line 13: Why is HGV abbreviation in square brackets?
**Response**: The bracket format is now consistent throughout the manuscript.

Line 16: Overall missing comma
**Response:** This has been corrected.

Line 56-60 belongs in the introduction section.
**Response**: The introduction section has been restructured to accommodate the requests on background and rationale.

Line 62: “The internal cross-validation methods provided by Prophet are used to arrive at a set of performance metrics applied across all sites.” This is not specific enough, you need a subsection under methodology for diagnostics and discuss why they are relevant.
**Response**: The introduction to the numerical framework Prophet is based on, and the use of cross-validation for time-series forecasting, is now explained in a new re-written section 2.

Line 63-64: ‘this process includes fitting to historical data over a specified period”, does this imply you are using an autoregressive forecasting method. If so, this does not imply novelty as these methods have been studied intensively for a long time.
**Response**: There is no claim the chosen method is novel anywhere in the manuscript. This paper was submitted as a model evaluation study.

Line 65 - 74: These descriptions have nothing to do with the Prophet model but the problem definition. Please distinguish the method from the application.
**Response**: As the reviewer has quite rightly raised in the preceding comments, it is important to clarify the rationale behind any given method according to the nature of the problem being solved. This is now improved significantly in a revised manuscript.

Line 82-91: Here application is discussed, this section is not related to the method.EMEP model is an integral piece of this work, please describe the system being solved with this model here.
**Response:** In the revised manuscript we now clarify how EMEP is used to simulate the concentrations of key pollutants, including NO2, across the UK with higher resolution nests focused over Manchester.

’Line 94: need to be carried over to notes or the appendix
**Response:** This has been changed.

Line 98: “meteorological data on wind speed, direction and temperature provided by the UK Met Office.” Please elaborate which model, data product, whether it is reanaly-sis data. In addition, please introduce
**Response**: This is a good point and will be added to the manuscript.

Line 116: “this method captures between 8 % – 15 % of vehicles making a journey between two sensors,”. Does this mean the traffic data captures up to 15% of actual traffic. If so the sampling bias, and portion of the traffic needs to be discussed in extenso.
**Response**: We do not have access to the statistics that would allow us to discuss the sampling bias in full [in extenso]. However we can ensure that, despite an open presentation on the assumptions made, this should be considered in how any further study builds on this.

Line 156-157: “The multi-modal behaviour of percentage deviation by site type may be indicative of local interventions not captured by the default change-points used during the fitting process.” This seems to be a speculation rather than a result. This can be an important point to bring up in the discussions.
**Response**: I agree. With a restructure in how the model results are presented, the site type variability is now part of a wider discussion.

Line 158 - 160: “One can alter the weighting given to such changes. A manual analysis on an individual site level might identify significant changes in local activity that would be expected to change the seasonality in measured NO2 and thus define change-points that need to be captured during the fitting process.” This is a use tip aboutProphet, I don’t think this relates to results.
**Response**: This is no longer included in the manuscript.

Line 164: Vanilla prophet is not discussed methodologically. This must be expanded in the methodology section.
**Response**: This has now been clarified in a rewrite of the methodologies used.

Line 169: Improvement in forecast from 18% to 10% deviation can be quite misleading. Authors need a transform bias correction scheme for the Yeo-Johnson transform to address this. Please see the seminal work of Beauchamp and Olson (1973) on pitfalls of using transformations. Secondly, a transformation is applied to data without being introduced in the methodology. There is no discussion of why this transformation is chosen over Box-Cox transform or a log transform.
**Response**: The new manuscript now includes a clear overview of the pre-processing steps used. In this instance, since we were transforming a log product, we used the Yeo-Johnson scheme.

Line 220: HYSPLIT model is introduced for the first time under results. This is a major part of the method yet not discussed in methodology.
**Response**: This has now been included in the revised manuscript.

Line 225: Ward's Method is introduced for the first time in results without discussion under methodology
**Response**: Apologies. This is standard practice. The purpose of back trajectory analysis is to confirm the arrival of different types of air masses at a given site. Whilst the specific clustering of trajectories will change with choice of linkage scheme, the usefulness of the results presented here was to confirm the change in air masses that coincided with significant increases in NO2 even during lockdown conditions.

Line 227- 229: "Discrepancies between forecasts that incorporate traffic and measuredvalues could arise from errors associated with a number of factors, but data on vehicle traffic type suggests this could also be due to an increase in the ratio of Heavy GoodsVehicles [HGVs]." This statement is not possible to make with current metrics and the level of exploratory analysis of variables. Bias residuals, in addition to very common forecast metrics such as MAPE and MASE are missing in the analysis. If NO2 concentrations have strong seasonality, then observing bimodal deviance (such Figure 1 is not surprising, that is just an indication that there are two regimes where model does significantly bad compared to others).
**Response**: Numerous studies of course confirm that different vehicles emit significantly different levels of NO2. This is core to discussions around clean-air zones and selective filtering schemes in urban areas. True it is not possible to statistically confirm this, but the evidence on a change in the nature of the local traffic could be significant. The new manuscript restructures the presentation of model performance.

Figure 4: On average, the model consistently over-estimates measured values. This needs to be discussed.
**Response**: This will now be included.