Responses to Anonymous Referee #1

The authors appreciate the comments made by this reviewer. They were considered and properly addressed. The following paragraphs provide response to each one.

This reviewer submitted his/her comments in a single and large paragraph. With the exception of the second and third comments, which were properly numbered, the other comments were split as best as possible.

This second version of the paper, "Finding candidate locations for aerosol pollution monitoring at street level using a data-driven methodology," has been edited in a minor way by the authors. However, certain fundamental, important issues have not been addressed.

The revised manuscript published as a discussion paper was revised and modified accordingly to the comments and recommendations made by the editor and one reviewer during the initial stage. The second reviewer did evaluate the original submission, but did not provide any comment.

Each comment was considered and properly addressed in the response letter submitted along the revised draft. Two comments stood out of that review. Firstly, the proposed data-driven modeling approach was confused with a source apportionment model. A paragraph in the revised manuscript was included to clarify that the proposed methodology did not correspond to any kind of receptor model. The second comment suggested that the article did not fit within the scope of *Atmospheric Chemistry and Physics*. An erroneous appreciation because the manuscript had been submitted to *Atmospheric Measuring Techniques*.

(1) The need to apply the data driven model to a range of conditions beyond 2 hour evening limits of the initial measurements to account for longer term averages that account for a range of emissions contributions and meteorological variability. Longer term averages, of course, are the manner in which air pollution regulations for health risk are written.

The main goal of this work was to evaluate the Self Organization Map (SOM) capability as a data-driven modeling method to approximate nonlinear relationships between urban parameters and air pollution data at ground level. The satisfying results predicting aerosol concentrations gave opportunity to interpolate concentrations in a complete gridded domain using data from a 10-day measurements campaign during the evening rush hour as a proof of concept.

By no mean the results obtained from the evaluation test can be used for air quality regulatory purposes or health risk assessment. The article states clearly that the approximated relationships can only be used for the particular pollutants monitored in the selected domain during the surveyed two hours. However, the methodology can be extended to other pollutants in a larger domain covering the whole diurnal course in future studies.

It is important to point out that *Atmospheric Measuring Techniques* aims to discuss advances in methodology and instrumentation for atmospheric research and monitoring, and not about new findings in atmospheric sciences or study cases as *Atmospheric Chemistry and Physics* does.

(2) The impact of conditions outside the local area which will affect air quality inside the area of interest. While the local land use variables for this very complex urban condition are accounted for, the model does not include any meteorological variability other than temperature and humidity. It is well known that conditions outside of a local area in a city are key to air quality conditions within that area. Local street canyon air flow variability as well as larger scale wind conditions averaged over a suitable period of time will be a strong factor in the monitoring data. The model being locally data driven may implicitly account for such influences, but this should be tested, perhaps with application of a grid-based conventional air quality model. This aspect of the modeling should be discussed qualitatively to provide for a broader critical discussion of the strengths and weaknesses of the methodology.

Depending on the urban setting and morphology, topography and climatology, advection could be an important issue to consider. However, as explained in the manuscript, the measurements to test SOM were conducted during days of typical meteorological conditions and not affected by transboundary pollution. Singapore is an island located in the tropics, where the climate is perennial. The monsoon seasons, defined by predominant wind directions, may have some impact on the concentration of pollutants at ambient level (i.e. above the urban canopy), but not really at ground, especially on days no affected by the so called smoke-haze produced by wildfires in neighboring islands. Figure 1 contrasts the significant difference between $PM_{2.5}$ concentrations at ambient and ground level during one set of measurements.

The proposed methodology targets the pollution at ground level, where the urban morphology and local emissions are the main players. The studied domain is located in the middle of the city and surrounded by neighborhoods with no major emission sources (e.g., large industries).

The application of a grid-based conventional air quality model as suggested by this reviewer to evaluate the spatial distribution of pollutants is beyond the scope of the work presented here. It is necessary to point out that those models are designed to forecast pollution at ambient level, but not at street level. In addition, at this time the application of a model like that in Singapore is not feasible. Singapore does not count with an emissions inventory of particles, let alone of any pollutant gas; and to evaluate the model's performance, only limited air quality data are publicly available (see Velasco & Roth, 2012; Velasco & Rastan, 2015).

The authors have not included a rationale for the pathway chosen for the measurements; the areal coverage of measurements is much smaller than the region modeled for relying on a data driven spatial extrapolation.

The method of SOM was proposed and tested as a tool for helping to design monitoring networks at ground level having in mind an inexpensive and practical approach. The length and route of the measurements along the streets and alleys were carefully designed to cover as much as possible the different land uses and urban topologies of the studied domain in two hours. The measurements were limited to two hours to meet the assumption that emissions and meteorological conditions were constant, and also because of the limited battery life of the portable monitors. This is explained in section 2.3 of the manuscript and the route is shown in Figure 2a.

The authors have not discussed the importance of integrating gas and particle data to represent air pollution conditions in the model. It is likely that measurements of different gases would yield different spatial extrapolations in the community compared with the aerosols.

The approach proposed here can be applied to both, trace gases and particles. Different relationships between urban parameters and air pollution species would be obtained, and therefore differences on the spatial extrapolations could be expected. The manuscript published as a discussion paper includes already a note about this in the conclusions (page 3329, lines 25-27).

Note that the article does not recommend to select locations for monitoring stations only on particles data. The test to evaluate the potential application of SOM was based on particles data because of two reasons: i) access to fast response and portable sensors, and ii) particles are responsible for driving the worst air quality conditions in Singapore.

The results shown for the clusters in Figure 10 are puzzling. The clusters appear to cover spatial regimes that are disconnected from one another. One would expect the conditions in such a complex topographical environment to be driven by concentrations gradients that are linked with one another or to similar emission sources. The authors should discuss the cluster results in more detail–trying to interpret their physical meaning from the model results. Since the link with emissions is not really discussed, the authors could add a short qualitative narrative that would link motor vehicle or cooking or other elevated emission densities to the cluster results to make sense of them. The revised paper provides a summary of this approach to a data based model for spatial extrapolation but does not extend insight for the reader beyond the initial draft.

This comment is not clear. Figure 10 shows the distribution of the grid cells divided in the four different clusters identified to form the studied domain. These clusters were obtained based on urban parameters and not on air pollution data. No concentration gradients, therefore, are shown in this figure. The red spots indicate the most representative cells of each cluster in terms of such urban parameters and maximum information gain over the whole domain. These cells were therefore selected as candidate locations for monitoring stations. Section 5 describes the clustering process and the last paragraph of section 6 (page 3339, lines 4-16) explains the cluster distribution and location of the selected cells shown in Fig. 10.

The discussion paper indicates that in a following article the features and roles of the urban parameters in the air quality at ground level of the studied domain will be analyzed. The aim of such article will be to understand the relationships between urban parameters and air pollution for a better urban planning, in particular when designing strategies to improve urban mobility promoting walking and cycling (pages 3338-3339, lines 27-3).

The siting options of the hypothetical monitoring sites is interesting, but the results appear unconvincing that the siting of three added stations for aerosols will improve exposure risk estimates beyond the single government station now in operation. I recommend that that authors look more closely at the strengths and weaknesses of the method at this stage of development and offer more insight about the methodology in the broader context of hypothetical air quality characterization specified by the model vs. current practice, and the needs for exposure assessment.

As mentioned in a previous response, by no mean the results obtained from the evaluation test presented here can be used for air quality regulatory purposes or health risk assessment. The objective of the work was to evaluate the capability of SOM as a data-driven modeling method to approximate nonlinear relationships between urban parameters and air pollution data at ground level. The satisfactory results from the evaluation suggest that SOM could be used to extrapolate air pollution data at fine resolution based on data of urban parameters as those used and cited in the manuscript. The work also demonstrated that in combination with a clustering algorithm SOM could be used to determine the optimum number and locations for monitoring stations to cover the different urban settings forming a defined neighborhood.

The siting of three of four ground level monitoring stations in a specific neighborhood, as wrongly considered to be the output of this work by this reviewer, may not represent a big contribution to the current air quality risk assessment of Singapore. However, the relevance of this work relies on the development and testing of a methodology that could represent a new tool for the air quality management in cities.

This work must be seen as a proof of concept. The manuscript states clearly the limitations of the methodology and do not use the results of the approach's evaluation to provide any health risk recommendation. The manuscript was prepared to fit the *Atmospheric Measuring Techniques* scope of discussing new methodologies in atmospheric sciences and not to discuss current needs for exposure assessment.

References

Velasco, E., and Roth, M.: Review of Singapore's air quality and greenhouse gas emissions: Current situation and opportunities. *Journal of the Air & Waste Management Association*, *62*(6), 625-641, 2012.

Velasco, E., and Rastan, S.: Air quality in Singapore during the 2013 smoke-haze episode over the Strait of Malacca: Lessons learned. *Sustainable Cities and Society*, *17*, 122-131, 2015.