Response to Levi Golston:

We appreciate your comments (listed in *italics* hereafter). Our responses are in blue. New/modified text in the manuscript is in **bold**.

My comment pertains to the conclusion: "This physical oversampling is applied to OMI NO2 products and IASI NH3 products, showing substantially improved visualization of trace gas distribution and local gradients.", as well as the premise of producing a grid that is significantly finer than the observed satellite pixels. My main argument is that the satellite measurement is the convolution of the spatial response function on the real atmospheric gas distribution, and that performing sensitivity weighted averaging like described does not change that the effective resolution of the product is still limited (to at least some extent, which does not seem to be characterized at all in the existing oversampling literature) by the Level 2 satellite pixel resolutions. Increasing resolution is actually a complex inverse problem to solve for the estimated gas distribution using the lower resolution Level 2 pixels even in the absence of noise.

An important difference between resolution and sampling is brought up by these comments. This difference has been thoroughly discussed for one-dimensional data, such as satellite observed spectra, in the literature. One good example is Chance et al. (2005)[1]. The resolution of the one-dimensional spectra is constrained by the full width at half maximum (FWHM) of the Instrument Spectral Response Function, whereas the sampling is independently defined by the spectral interval of linear detector array. If the sampling interval is coarser than Nyquist sampling, the spectra are referred to as "undersampled"; if the sampling interval is (much) finer than Nyquist sampling, the spectra are referred to as "oversampled".

For two-dimensional spatial data, the resolution can be similarly constrained by the Instrument Spatial Response Function of the sensor. However, the difference between spatial resolution and the size of spatial sampling grid is often blurred. It has been a common practice to refer to the grid size of Level 3 data as "resolution" of the data. For example, "super-resolution" in the literature actually means "super fine spatial sampling", and strictly speaking, the spatial resolution stays the same as it is defined by the OMI pixel sizes that are independent of the target grid size. To avoid confusion, we replaced the term "**grid resolution**" in the manuscript by "**grid size**".

By way of example, the Comment Figure below shows a simulated 1 km by 1 km region of high gas values as 'oversampled' if there was a series of 7 km by 9 km pixels each with a 2D boxcar response function. Due to the limited resolving power of the underlying pixels, the result is seen to be spread over a much larger, 13 km by 17 km area. This is a tough test case because the gas distribution is discontinuous, nevertheless the paper shows point sources applications and a very similar pattern (distinct 'circular blur') is seen in the observational results (Figs. 8, 9, and 10) at sub-pixel scales.

- If the Comment Figure is interpreted directly, then one would conclude there are gradients around the 1 km by 1 km source, when this is purely an artifact - the algorithm displays far more information than there actually is. For this reason, caution should be used in applications where

there are apparent gradients around point sources, which could easily lead to mistaken conclusion about atmospheric transport or decay processes.

- Similarly, if this Comment Figure were used for a high resolution emission inventory, the emission should be proportional to the value at that specific 1 km by 1 km grid cell, which has a source strength here defined as 1. By my calculation, the corresponding oversampled grid cell has a value of only 1/63. The true source value can be retrieved - but only by integrating the whole 13 km by 17 km area which negates the high-resolution.

- One exception where higher-resolution can be used quantitatively is that that the center location of a source can be identified below Level 2 pixel resolution (exactly in the ideal case, but noise will of course play a role in real applications).

It is well-known to the community that the true spatial resolution (not the "resolution" of grid) of Level 3 maps is determined by the Level 2 pixel sizes. The apparent gradient due to finite Level 2 pixel sizes was taken into account in previous studies[2–5]. One subsection (section 4.3) and one figure (Fig. 8 of the revised manuscript) is added to clarify the definition of spatial resolution vs. spatial sampling:

"**The difference between resolution and sampling density for 1-D spectral data has been thoroughly discussed in the literature (e.g., Chance et al., 2005). However, for 2-D, spatially resolved data, it is common to refer to both the sizes of the Level 2 pixels and the size of the Level 3 grid as the spatial "resolution" of the data. To avoid confusion, it is emphasized here that the true spatial resolution is limited by the sizes of Level 2 pixels. The size of Level 3 grid only determines the density of spatial sampling, which does little to enhance the true resolving power of the data after reaching a certain point. For example, the oversampling results using synthetic OMI data at 1 km vs. 0.05 km grids are very similar (Fig. 6). Nonetheless, it is still beneficial to oversample, i.e., make Level 3 grid size significantly smaller than Level 2 pixel sizes, as demonstrated by Fig. 8. As the ground truth, an array of 2-D Gaussian functions are generated with FWHM ranging from 1 km to 16 km (the second column of Fig. 8) and peak height of unity, and this true field of concentration is measured by an imaginary sensor whose spatial response function is a 2-D super Gaussian (Eq. 8) with FWHM = 10 km and $k_1 = k_2 = 8$ (the first column and the white boxes inserted in the third column). The third column shows the oversampling results using 10000 randomly located observations. The fine structures in the ground truth are clearly smoothed, limited by the spatial resolution that is inherent to the Level 2 pixel sizes (10 km). However, by oversampling at a fine grid (0.2 km for the first row vs. 5 km for the second row), the spatial gradients are better recovered, and spatial features finer than individual Level 2 pixels can be identified. Additionally, the details in the spatial response function is better resolved with a finer target grid, which is particularly beneficial when collocating with higher resolution measurements (e.g., a cloud imager). As such, although the spatial resolving power is ultimately determined by the spatial extent of satellite pixels, the physical oversampling approach helps enhancing the visualization of spatial gradient and the identification of emission sources.**"

Figure 8 of the revised manuscript:

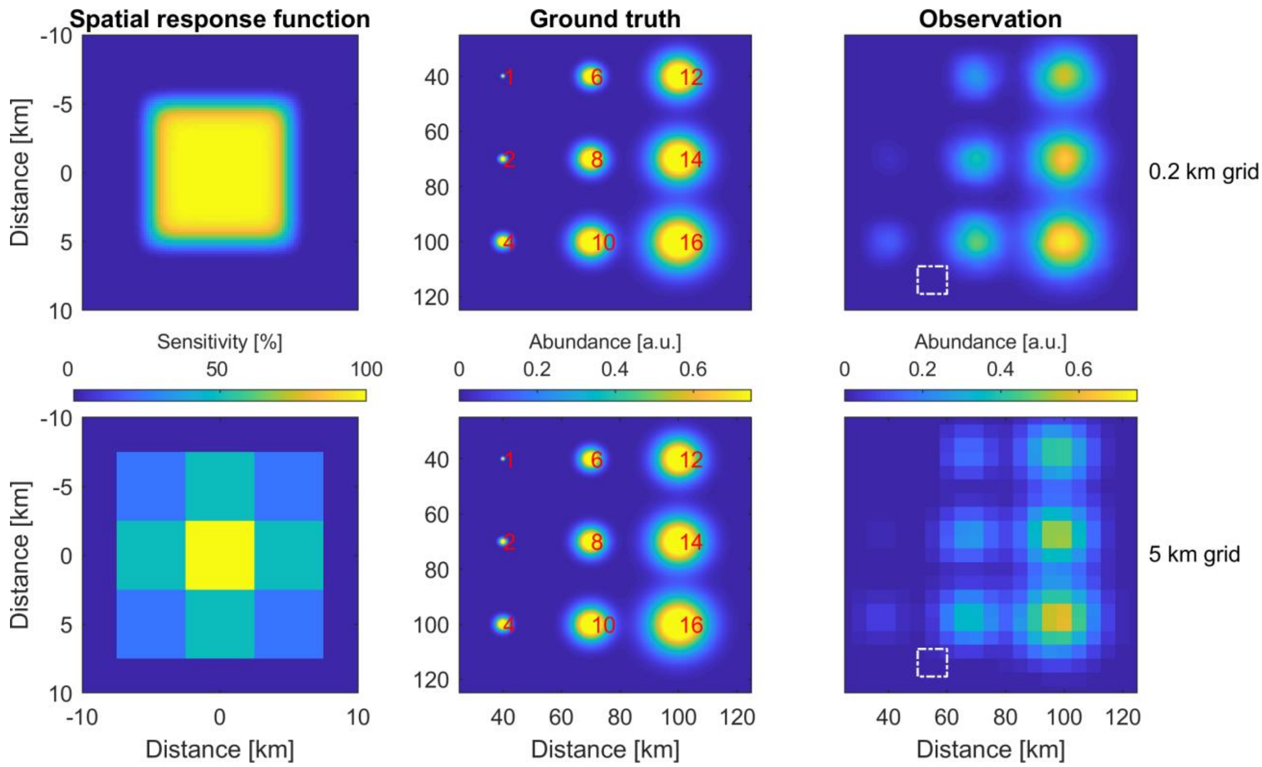

Figure 8. First column: spatial response function of an imaginary sensor discretized at 0.2 km (top) and 5 km (bottom) grid. Second column: ground truth spatial distribution generated as an array of 2-D Gaussian functions of same height (the top and bottom panels are the same). The FWHM of each Gaussian is labeled. Third column: physical oversampling results using 10000 randomly generated observations and discretized at 0.2 km (top) and 5 km (bottom) grid. The pixel size, which determines the spatial resolution, is labeled as the inserted white boxes.

In summary, besides the application just mentioned, it is unclear whether the much higher-than-pixel resolution output can be justified. In the past literature, Fioletov et al. 2011 (GRL) and Fioletov et al. 2013 (JGR:A) claims 'detailed "subpixel-resolution" spatial distribution' is possible and Streets et al. 2013 (Atmos. Environ.) refers to oversampling as achieving "super-resolution". To what extent is the enhanced resolution (such as 1 km, as proposed in this manuscript) physically real? I suggest adding discussion of the limitations of the approach and caveats in interpreting the results of oversampling. This will strengthen the manuscript and help the community who may use the described algorithm or other similar methods in the future.

The advantage of fine-grid output vs. coarse grid output can be easily seen by comparing the first row (0.2-km grid) with the second row (5-km grid) in Fig. 8 of the revised manuscript (this figure is included in the response to the previous comment) and by comparing the first column (10-km grid) with columns 2-4 (1-km grid) in Fig. 8 of the original manuscript (now Fig. 9 in the revised manuscript). See responses to previous comments for clarification of spatial resolution vs. spatial sampling.

A point on the nomenclature of oversampling: while the term has been used in some past satellite papers, nevertheless I find it problematic since it is quite different than how it used in signal processing, where there is a well-known and widely used meaning of natively sampling at high resolution and then converting to a lower one. The authors note instead that the presented algorithm is a type of interpolation. I believe that referring to the algorithm as 'gridding' rather than 'oversampling' is formally correct and gives a much better intuition of what the algorithm actually does.

As shown in the response to the first comment, we believe oversampling is the appropriate nomenclature.

A minor point - 'agile' is also used to describe the algorithm, can a few words be added to clarify what the intended meaning is?

It means that the algorithm works for different satellite sensors with quadrilateral/elliptical pixel shapes and different pixel sizes, and that the sensitivity distribution can be flexibly chosen by changing the parameters of 2-D super Gaussian function. See page 3, line 7 of the manuscript.

"**In this work, we present an agile, physics-based oversampling approach that represents each Level 2 satellite pixel as a sensitivity distribution on the Earth's surface (e.g., the spatial response function), instead of a point or a polygon as assumed in previous methods.**"

References:

1. Chance, K., Kurosu, T. P. & Sioris, C. E. Undersampling correction for array detector-based satellite spectrometers. *Appl. Opt.* **44,** 1296–1304 (2005).

2. de Foy, B., Lu, Z., Streets, D. G., Lamsal, L. N. & Duncan, B. N. Estimates of power plant NOx emissions and lifetimes from OMI NO2 satellite retrievals. *Atmos. Environ.* **116,** 1–11 (2015).

3. Liu, F. *et al.* NOx lifetimes and emissions of cities and power plants in polluted background estimated by satellite observations. *Atmos. Chem. Phys.* **16,** 5283–5298 (2016).

4. Valin, L. C., Russell, A. R. & Cohen, R. C. Variations of OH radical in an urban plume inferred from NO2 column measurements. *Geophys. Res. Lett.* **40,** 1856–1860 (2013).

5. Beirle, S., Boersma, K. F., Platt, U., Lawrence, M. G. & Wagner, T. Megacity emissions and lifetimes of nitrogen oxides probed from space. *Science* **333,** 1737–1739 (2011).