Manuscript ID: amt-2022-37

Spectral replacement using machine learning methods for continuous mapping of Geostationary Environment Monitoring Spectrometer (GEMS)

Yeeun Lee, Myoung-Hwan Ahn*, Mina Kang, Mijin Eo

Reply to comments #2 from Dr. Glen Jaross

We would like to thank Glen Jaross for his insightful comments and suggestions, which are significant to reconsider the direction of our research and include some missed points not fully explained in the first draft. We have revised the manuscript considering the suggested comments and in this document, we will go over major concerns raised by the referee and describe how to have improved the manuscript accordingly. The referee's comments are presented in **bold** and the presented figures and tables have been included in the revised manuscript with identical numbering. The revised manuscript is presented in *italic.*

<General comments>

Comment – Part #1:

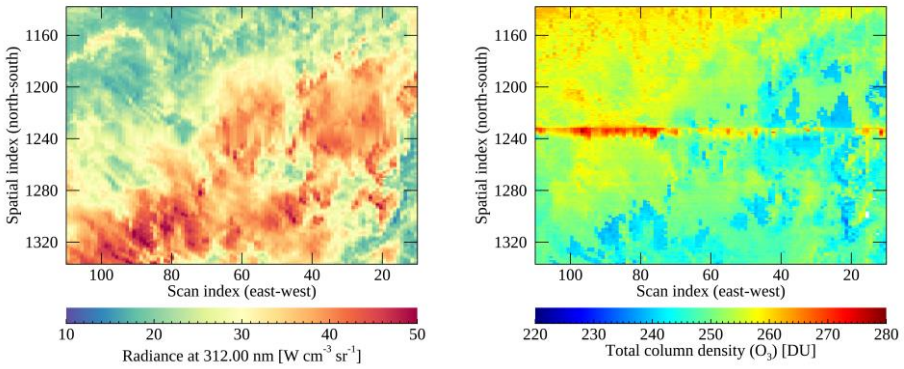
"I have difficulty understanding the goals of this paper. There seems to be a mismatch between what the authors are trying to achieve and the investigation they describe in the paper. The authors describe a problem on GEMS whereby "bad" pixels, i.e. missing pixel radiances, cause problems in the ensemble of measured radiances. … There is no mention of what criteria are used to assess the results.

Response #1:

We agree with the comments that the goals of this paper are not fully explained. Bad pixels of GEMS could cause distinct spatial discrepancy in GEMS Level 1 data and it definitely introduces errors in Level 2 products (or make impossible to derive the Level 2 products). Thus the primary goal of the current work is to evaluate the applicability of machine learning methods for the spectral gap filling, which could reduce bad pixel effects to both Level 1 and 2 products. For the purpose, bad pixel effects are discussed first (as suggested by the referee) and thus we have added the following part to Section 2.1, describing the effects of bad pixels to radiances and retrieved properties (i.e., ozone):

The interpolation error could seriously affect the Level 2 products of which the spectral fitting windows are overlapped with bad pixel areas. For instance, cloud properties and aerosol effective height (AEH) of GEMS are retrieved from O_2-O_2 absorption bands around 477 nm (Choi et al., 2021; Kim et al., 2021) where the cluster of bad pixels is located (Defect 3). During the IOT, Defect 3 caused spatial discontinuity to the retrieved cloud and AEH distribution, which made the fitting window of the products modified to avoid bad pixel effects. Ozone retrieval is also affected by Defect 2 (300-400 nm) as the spectral radiances within 300-380 nm provide major information for the ozone retrieval of GEMS (Bak et al., 2019). When specifying Defect 2 and its surrounding pixels, the bad pixel effects are clearly

shown in Fig. 2 which presents GEMS radiances at 312 nm and the retrieved ozone total column density. Even though radiances at a certain wavelength are homogeneous with its surroundings (see Fig. 2a), the spectral patterns are not properly reproduced with the existed method (spatial interpolation) causing the retrieval errors easily found in the Level 2 product. The detailed discussion of reproducing spectral patterns of bad pixels will be continued in Sect. 3.2.2 with PCA-based evaluation.



(a) (b)

Figure 1 *Spatial distribution of (a) the GEMS measured radiance at 312 nm and (b) ozone total column density by zooming in the bad pixel area for Defect 2 and its surrounding area.*

Comment – Part #2:

"From an academic perspective the question of how well ML techniques can describe Earth backscattered radiances is an interesting one. If the authors approached this paper from that perspective they might provide a useful contribution to our ability to characterize the atmosphere through numerical techniques. But in doing so they must take a more rigorous approach to evaluating their radiance predictions.

The first thing the authors should do is to forget about the GEMS pixel defects. These are of no use in evaluating the efficacy of the technique since the true radiances remain unknown for these regions. Instead, choose regions of the detector where there are good measurements and treat them as missing for the purpose of deriving errors. There can be a variety of region shapes and sizes, including ones that look very similar to Defects 1, 2, and 3."

Response 2:

The points are also raised by another referee and we appreciate the suggestions. Following the suggestions, we targeted a certain area including each defect (Defects 1-3) and its surroundings (100-indices toward both north and south directions) where actual measurements (regarded as 'true') could be obtained. The results are added to the 'Results and discussion' (Section 3) which has been re-organized with 'Model selection' (Section 3.1) and 'Evaluation' (Section 3.2). The following part has been inserted to Section 3.2:

3.2 Evaluation

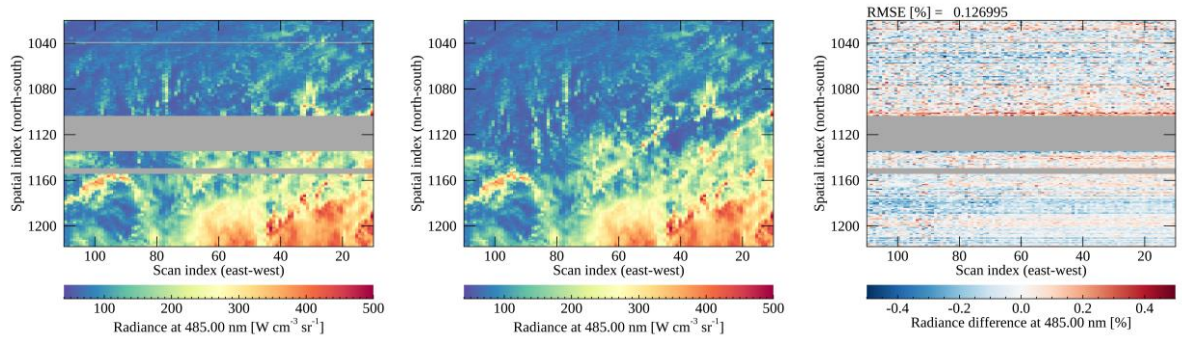
3.2.1 Spatial and spectral inspection

For the quantitative evaluation of the reproduced spectra, certain areas are targeted which include each defect (Defects 1-3) and its surroundings where actual measurements regarded as 'true' could be obtained. The evaluation is made with the data measured on 10 March 2021 (06 UTC), which are not used for the model training. The center longitude of the areas is set to 128° E, which is identical with the sub-nadir longitude of GK-2B. Along the spectral direction, we focus on the specific spectral range of the whole spectral gap of Defects 1-3, as shown in Table 3. Specifying the range helps to closely analyze the spectral patterns of absorption lines of trace gases and cloud properties. Table 3 presents spectral ranges of Defects 1-3 and the target wavelengths for the analysis.

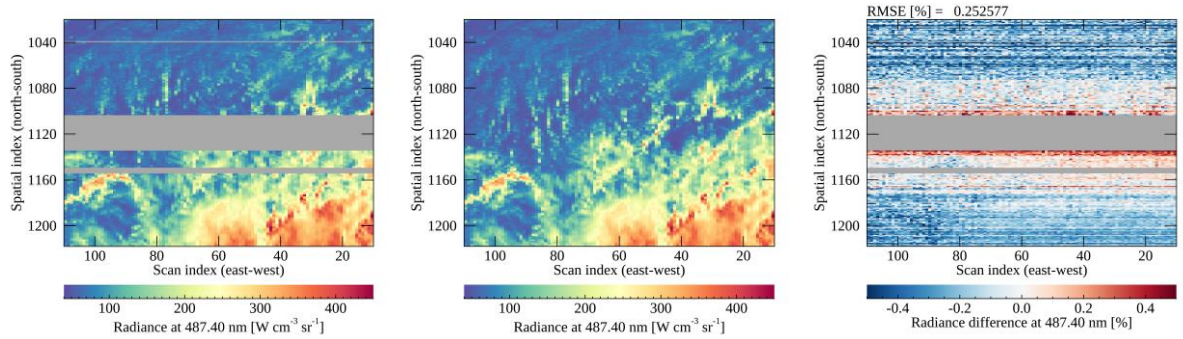
Table 1 *The spectral range of Defects 1-3 and target wavelengths for the analysis. The third column presents GEMS retrieval products of which each fitting window is overlapped with Defects 1-3.*

Defect	*Target wavelength*	*GEMS Level 2 product*	*Optimized model*
1 (400-500 nm)	*432-450 nm*	*CHOCHO, NO2*	*PCA-Linear*
2 (300-400 nm)	*312-360 nm*	*O3, HCHO, SO2, NO2, aerosol optical depth*	*PCA-ANN*
3 (484-491 nm)	*484-491 nm*	*Cloud, AEH*	*PCA-Linear*

For the evaluation, actual GEMS radiances and the reproduced radiances with machine learning methods are directly compared, hereafter called GEMS radiances and ML radiances, respectively. In Figs. 11-13, each column shows GEMS, ML radiances and the difference while the first and second rows show the representative wavelengths for the smallest and the largest difference, respectively. Figure 11 shows the comparison results of the Defect 3 area, which shows the best performance compared to the Defects 1-2 areas. The difference in Fig. 11 is close to zero (within ± 0.5%) because the spectral gap of Defect 3 is narrower than the counterparts of Defects 1-2. The narrower the spectral range of the output radiances is, the more abundant information could be obtained from the input radiances. For Defect 3, there is no scene dependence over the output wavelengths and the difference shows noise-like features except for the spatial dependence which might be originated from instrument artifacts.



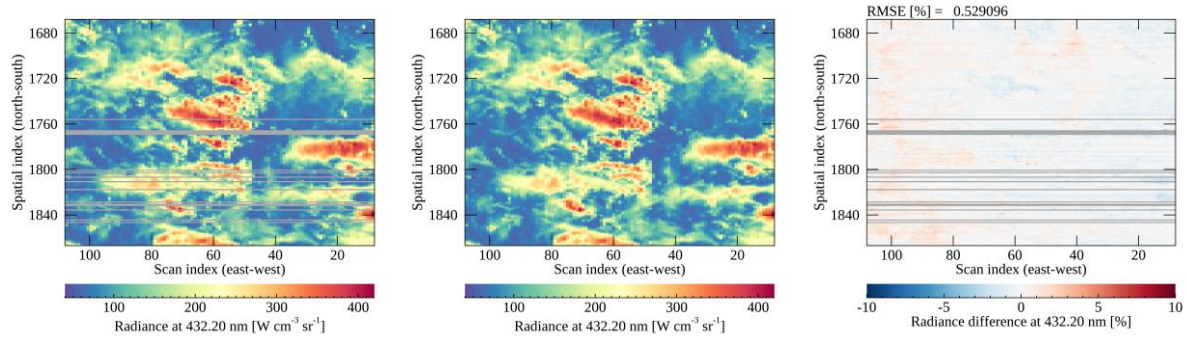
(a)



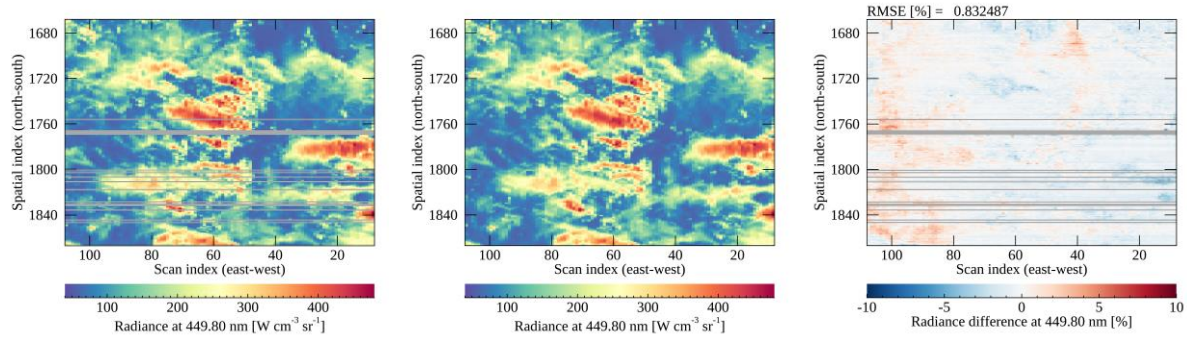
(b)

Figure 2 *The GEMS, ML radiances and the difference from left to right at the wavelengths presenting (a) the smallest and (b) the largest difference for the Defect 3 area. Bad pixels are marked in dark gray and the difference is calculated as (ML-GEMS)/GEMS in percent. The color bar range for the difference is* $\pm$ *0.5% and the unit of RMSE is in percent divided by the mean radiance.*

Figure 12 shows the Defect 1 area where the ML radiances are within about 5% of the GEMS radiances. It also shows that dark targets (clear sky with small radiance) show a positive difference while bright targets (mostly cloudy sky with large radiances) show an opposite tendency. The tendencies are also found from the ML radiances on the other dates for different angle conditions such as SZA and VZA. It seems the applied machine learning model (PCA-Linear) might not be fully trained to resolve the different atmospheric conditions and radiances which causes a certain bias depending on the scenes.



(a)



(b)

Figure 3 *Same as Fig. 11 for the Defect 1 area.*

For the Defect 2 area, it is clear that the information from valid radiances of wavelengths longer than 400 nm is insufficient to effectively reproduce the spectral features at shorter wavelengths (consistent results with Figs. 8-9). Both output spectral lengths of Defects 2-3 are nearly identical around 100 nm but it seems radiances near 300 nm need more information to be successfully reproduced.

The stripping feature found in Fig. 12b is significant at 312 nm for the ML radiances, while it doesn't at 357.2 nm in Fig.12a. The stripping feature seems to be added during the reproducing process especially for shorter wavelengths, and the reason is still unclear. Another distinct feature found in Fig. 12 is that the difference in northern parts is very large with the difference of 10%. We suspect that the reason might be the VZA effect considering that VZA increases at the northern parts in the area. Without angle conditions in the input parameters for the model, the difference becomes doubled at 312 nm presenting similar patterns with the difference in Fig. 12b. This indicates the angle effect can be emulated in the model by applying VZA and SZA as the input parameters, but it is not fully resolved especially for the radiances at shorter wavelengths.

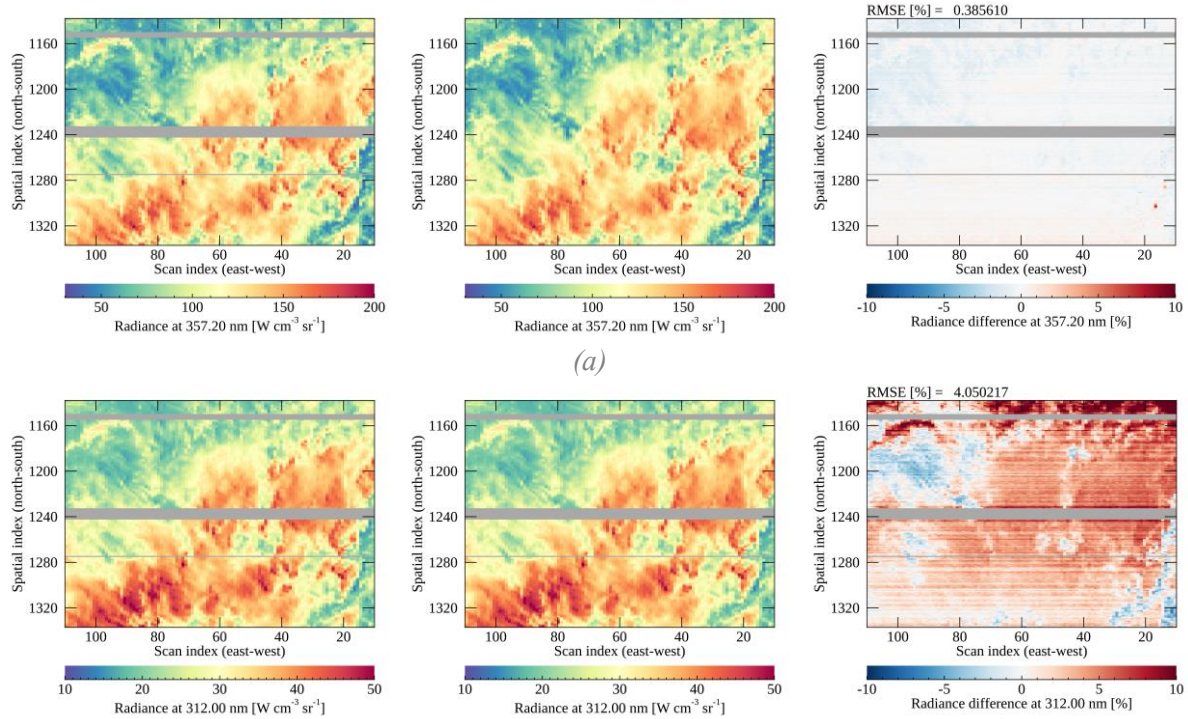
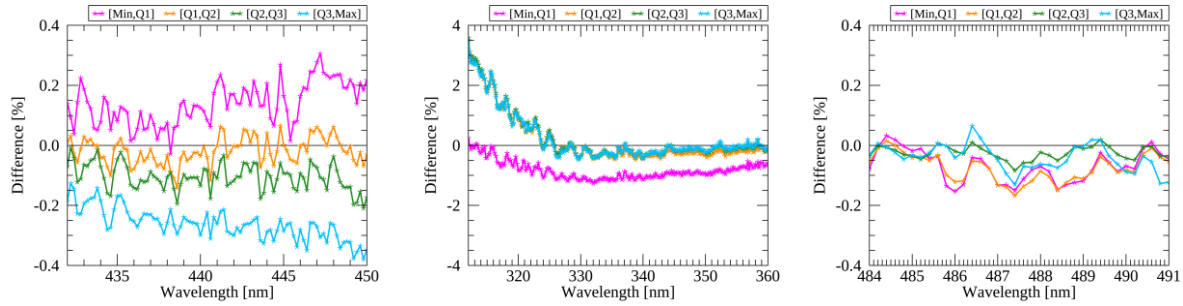


Figure 4 *Same as Fig. 11 for the Defect 2 area.*

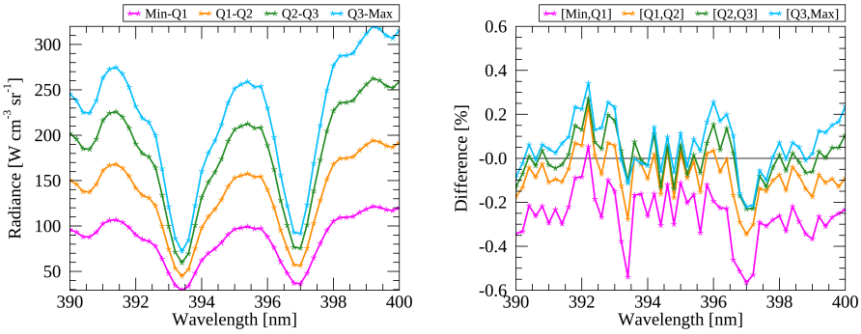
A closer inspection is performed to analyze the general spectral features over target wavelengths. For each defect area in Figs. 11-13, the collected spectra are divided into four groups depending on the scene brightness considering that ML radiances could have different systematic biases depending on the scenes. With the data, the mean difference is calculated for each wavelength. As found in Fig. 11, Fig. 14a shows that the ML radiances over dark scenes have the positive bias while brighter scenes have the negative bias. It is interesting that the scene dependence is only significantly found for Defect 1. Figure 14b indicates that the ML radiances are overestimated except for the very brighter scenes. It should be noted that the y-axis range of Fig.14b is wider than the figures for Defects 1 and 3. With the results, it can be deduced that the complicated atmospheric effects at the shorter wavelengths are difficult to be emulated and instrument artifacts such as stray light also would affect the reproducing process. Figure 14c shows relatively large difference at the spectral peaks, but generally the difference is smaller than 0.2%



(a) *(b)* *(c)*

Figure 5 *Mean difference between ML and GEMS radiances within the target area of (a) Defect 1, (b) Defect 2 and (c) Defect 3. Each color indicates the average for each quartile and Q1, Q2 and Q3 represent the first, second and third quartile, respectively. The difference is calculated as (ML-GEMS)/GEMS in percent.*

Besides the shorter wavelengths of Defect 2, mean ML radiance and the difference with GEMS radiances are presented by targeting Fraunhofer lines from 390 to 400 nm (see Fig. 14). The Ring effect caused by rotational Raman scattering can be found over the two peaks in Fig. 14a, which is generally known to be very small and largely affected by clouds (Joiner et al., 1995). Figure 14b shows that PCA-ANN reproduces the dominant features at the peaks very well on average within 0.6%, but it seems the difference increases with darker scenes where the Ring effect becomes stronger. This indicates that the ML radiances would need additional information to successfully reproduce the exact spectral features especially for the very small signals such as the Ring effect.



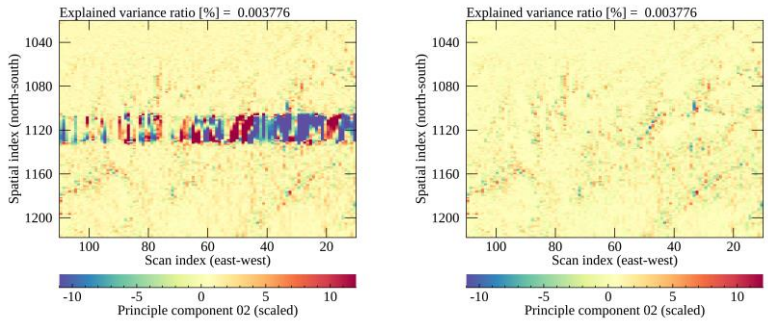
(a) *(b)*

Figure 6 *(a) Mean ML radiances (b) and the difference with GEMS raidances at the Fraunhofer lines for the Defect 2 area. Each color indicates the average for each quartile and Q1, Q2 and Q3 represent the first, second and third quartile, respectively. The difference is calculated as (ML-GEMS)/GEMS in percent.*

3.2.2 PCA-based spectral analysis

As applied in the pre-processing step in our research, PCA is a very useful tool to capture the meaningful variances along the spectral direction and it has been widely used to retrieve environmental and surface properties (Horler and Ahern, 1986; Joiner et al., 2016; Li et al., 2013, 2015). To investigate further the spectral patterns, we apply PCA to GEMS radiances (except for bad pixels) at the target wavelengths (see Table 3) collected within each area in Fig. 11-13. With PCA, various spectral patterns are compressed and a spectrum can be projected to PC subspaces by multiplying with the constructed PC matrix (eigenvector matrix). This indicates that if a spectrum has disparate spectral patterns, the projected PCs would also have distinct values when comparing with the PCs of GEMS radiances. Figure 15 presents the results when projecting both GEMS and ML radiances with PCA. For the inspection, the Defect 3 area is presented which has the wider defective width along the north-south

direction. Because the first PC scores represents mean radiances, the second PC are used for the analysis. As we assumed, bad pixels in Fig. 15a show disparate values because the spectral patterns of the interpolated spectra are inconsistent with GEMS radiances. The ML radiances in Fig. 15b show spatially homogenous PC scores which indicates that the machine learning methods could properly reproduce the dominant spectral patterns, in this case of the second PC.



(a) *(b)*

Figure 15. *The second PC of (a) actual measurements and (b) reproduced spectra on the target area for Defect 3. The PC is scaled for clarity of presentation.*

The dominant patterns for each PC are presented in Fig. 16 with GEMS radiances for the target wavelengths of Defects 1-3. Each color indicates the eigenvector of the first-sixth PCs which determines how each PC score of a spectrum contributes to the original spectrum. Li et al. (2015) verified that the leading PC scores from the UV/VIS backscattered radiation (shorter than 360 nm) are significantly correlated with dominant absorption features and surface properties. The trailing PC scores might be associated with instrument artifacts and other unresolved spectral features. Figure 16 shows that the first PC corresponds to the mean spectrum as discussed in Sect. 3.1.1 and the second-sixth PCs show dominant spectral patterns originated from absorption features of trace gases, surface properties and unresolved features. This indicates that the comparison of PC scores could provide the information on the similarity of the dominant patterns between ML and GEMS radiances as shown in Table 4 with the correlation coefficient. The results indicate that the mean spectral feature (the first PC) and some dominant patterns (the second and third PCs) could be well reproduced with the suggested models, but other spectral features such as the fourth PC for Defect 2 have difficulty obtaining valid information from input radiances for accurate reproduction. The magnitude of radiance from the major PCs except for the first PC might not be large considering that even the leading PCs have small explained variance ratio for hyperspectral data in UV/VIS spectrum. However, it would be enough to determine the exact spectral signals which are mostly related to the important information for the retrieval process.

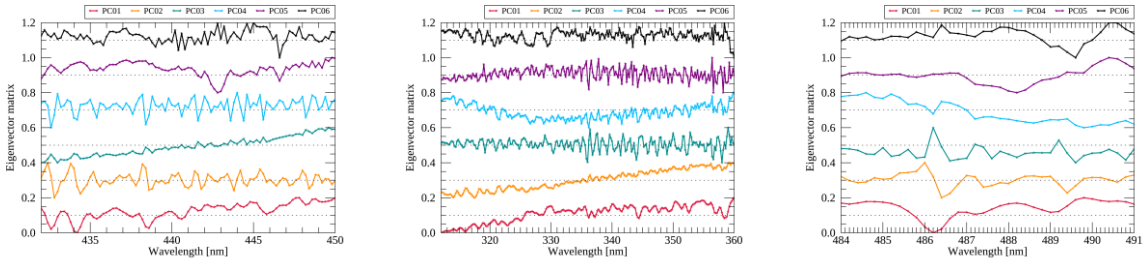


Figure 7 *Eigenvector of the first-sixth PCs applied to GEMS radiances for the target wavelengths of (a) Defects 1, (b) Defect 2 and (c) Defect 3. All eigenvectors are scaled (min-max scaling) and shifted for clarity of presentation.*

Table 2. *Correlation coefficient of PC scores of reproduced and actual measurements for Defects 1-3.*

Defects	*PC 01*	*PC 02*	*PC 03*	*PC 04*	*PC 05*	*PC 06*
Defect 1	*0.9999*	*0.9976*	*0.8172*	*0.9779*	*0.6846*	*0.6609*
Defect 2	*0.9999*	*0.8129*	*0.9876*	*0.4294*	*0.7035*	*0.5046*
Defect 3	*0.9999*	*0.9962*	*0.9787*	*0.6644*	*0.5399*	*0.2649*

Comment – Part #3:

"But what is the real problem that the authors are trying to solve? In an instrument such as GEMS, designed to measure trace gas composition of the atmosphere through hyperspectral measurements, the goal is probably to produce trace gas products without spatial gaps caused by missing radiances. The authors do not discuss the issue of trace gas retrievals or other products derived from their predicted radiances. Is there any improvement at all in those products? …

The authors must also devise evaluation criteria that are more robust and quantitative than "these spectra look realistic." Since the goal for GEMS radiances is to derive atmospheric products such as trace gases, perhaps these trace gas retrievals can be used as the metric. Merely stating that predicted radiances agree on average with measured radiances to within X% ignores the subtle spectroscopic sensitivity of trace gases such as NO2, where the exact relationship between wavelengths is of utmost importance."

Response 3:

As presented in the previous section, we have evaluated the reproduced spectra by comparing with actual measurements and applied PCA to analyze spectral features of reproduced spectra. With the analysis, it was found that the machine learning models properly reproduce dominant spectral patterns for Defects 1-3 with only radiances from the rest part of spectra and angle conditions. However, the exact spectral features (<1%) determined by small signal (the important information) may be accurately reproduced only if the spectral range of output radiances are closer to the input radiances enough to obtain sufficient information from the input radiances (such as Defect 3). Also, it seems additional information would be needed to reproduce exact spectral features when the output radiances are overlapped with strong absorption or scattering lines. Considering the ultimate goal of measuring hyperspectral data, we agree with the referee's suggestion to apply the retrieval algorithms for evaluating the reproduced spectra. However, as the initial approach reproducing missing radiance of GEMS, we hope to evaluate the applicability of machine learning methods for the GEMS measurements, which have meaningful information as well as instrument artifacts. As the referee pointed out, the effect of reproducing spectra for the retrieval process is a necessary step and based on the findings in this research, we hope to investigate further the step in a follow-up study.

Reference

Bak, J., Baek, K. H., Kim, J. H., Liu, X., Kim, J. and Chance, K.: Cross-evaluation of GEMS tropospheric ozone retrieval performance using OMI data and the use of an ozonesonde dataset over East Asia for validation, Atmos. Meas. Tech., 12(9), 5201–5215, doi:10.5194/amt-12-5201-2019, 2019.

Choi, H., Liu, X., Gonzalez Abad, G., Seo, J., Lee, K.-M. and Kim, J.: A Fast Retrieval of Cloud Parameters Using a Triplet of Wavelengths of Oxygen Dimer Band around 477 nm, Remote Sens., 13(1), 152, doi:10.3390/rs13010152, 2021.

Horler, D. N. and Ahern, F. J.: Forestry information content of thematic mapper data, Int. J. Remote

Sens., 7(3), 405–428, doi:10.1080/01431168608954695, 1986.

Joiner, J., Bhartia, P. K., Cebula, R. P., Hilsenrath, E., McPeters, R. D. and Park, H.: Rotational Raman scattering (Ring effect) in satellite backscatter ultraviolet measurements, Appl. Opt., 34(21), 4513, doi:10.1364/AO.34.004513, 1995.

Joiner, J., Yoshida, Y., Guanter, L. and Middleton, E. M.: New methods for the retrieval of chlorophyll red fluorescence from hyperspectral satellite instruments: Simulations and application to GOME-2 and SCIAMACHY, Atmos. Meas. Tech., 9(8), 3939–3967, doi:10.5194/amt-9-3939-2016, 2016.

Kim, G., Choi, Y. S., Park, S. S. and Kim, J.: Effect of solar zenith angle on satellite cloud retrievals based on O2–O2 absorption band, Int. J. Remote Sens., 42(11), 4224–4240, doi:10.1080/01431161.2021.1890267, 2021.

Li, C., Joiner, J., Krotkov, N. A. and Bhartia, P. K.: A fast and sensitive new satellite SO2 retrieval algorithm based on principal component analysis: Application to the ozone monitoring instrument, Geophys. Res. Lett., 40(23), 6314–6318, doi:10.1002/2013GL058134, 2013.

Li, C., Joiner, J., Krotkov, N. A. and Dunlap, L.: A new method for global retrievals of HCHO total columns from the Suomi National Polar-orbiting Partnership Ozone Mapping and Profiler Suite, Geophys. Res. Lett., 42(7), 2515–2522, doi:10.1002/2015GL063204, 2015.